TECHNISCHE UNIVERSITÄT DARMSTADT

Digital-Fachtagung
MECHATRONIK 2021
Darmstadt, 24.-25.03.2021

VDI VDE



Veranstalter:

Technische Universität Darmstadt
DIGITAL

- **Mechatronische Produkte**
  (neue Funktionalitäten, Industrie 4.0, Kosteneffizienz, Zuverlässigkeit, …)
- **Serienanfertigung mechatronischer Produkte**
  (Fallbeispiele, Komponenten, Architektur, Qualitätsmanagement, …)
- **Ressourceneffizienz**
  (Energie, Material, …)
- **Nutzerfreundlichkeit und Akzeptanz**
  (Assistenzsysteme, Schnittstellen, Interaktion, Gesetzgebung, …)
- **Systems Engineering und Entwicklungsmanagement**
  (Prozesse, Verfahren, Software, …)
- **Innovative Konzepte und digitale Geschäftsmodelle**
  (Modelle, Regelung, Optimierung, Eco-Systeme, …)
- **Systemvernetzung und Systemintegration**
  (Konzepte, Verfahren, …)
- **Smarte Aktoren**
  (Konzepte, Beispiele, …)
- **Automatisierte Mobilität**
  (Konzepte, Modelle, Sensorik, …)
- **Robotik**
  (Kooperation, Kollaboration, Perzeption, …)

www.VDI-Mechatroniktagung.de

Digital-Fachtagung
MECHATRONIK 2021
Darmstadt, 24.-25.03.2021

Herausgeber: T. Bertram, B. Corves, K. Janschek, S. Rinderknecht











TECHNISCHE
UNIVERSITÄT
DARMSTADT

2. Auflage

www.VDI-Mechatroniktagung.de

Der Tagungsband, der die Vorträge der Fachtagung enthält, erscheint als nichtredigierter Manuskriptdruck. Die einzelnen Beiträge geben die auf persönliche Erkenntnisse beruhenden Ansichten und Erfahrungen der jeweiligen Vortragenden bzw. Autoren wieder.

Printed in Germany.

Veröffentlicht unter CC-BY 4.0 International
https://creativecommons.org/licenses

DIGITAL-FACHTAGUNG MECHATRONIK 2021

Tagungsleitung:

Univ.-Prof. Dr.-Ing. Prof. h.c. Dr. h.c. Torsten Bertram, Technische Universität Dortmund
Univ.-Prof. Dr.-Ing. Dr. h.c. Burkhard Corves Rheinisch-Westfälische Technische Hochschule Aachen
Univ.-Prof. Dr. techn. Klaus Janschek, Technische Universität Dresden
Univ.-Prof. Dr.-Ing. Stephan Rinderknecht, Technische Universität Darmstadt

Programmausschuss:

- Dipl.-Ing. Andreas Abel, ESI ITI GmbH, Dresden
- Prof. Dr.-Ing. Johann Bals, DLR e. V., Weßling
- Univ.-Prof. Dr.-Ing. Philipp Beckerle, Friedrich-Alexander-Universität Erlangen-Nürnberg
- Univ.-Prof. Dr.-Ing. Michael Beitelschmidt, Technische Universität Dresden
- Prof. Dr.-Ing. Martin Bothen, Technische Hochschule Aschaffenburg
- M.Sc. Mehmet Bozkurt, VDI, Düsseldorf
- Prof. Dr.-Ing. Thorsten Brandt, Hochschule Rhein Waal
- PD Dr.-Ing. Annerose Braune, Technische Universität Dresden
- Univ.-Prof. Dr. rer. nat. Debora Clever, Technische Universität Darmstadt & ABB AG
- Prof. Dr.-Ing. Michael Dattner, Beuth Hochschule für Technik, Berlin
- Dr.-Ing. Hans Theo Dorißen, HELLA GmbH & Co. KGaA., Lippstadt
- Dr.-Ing. Olaf Enge-Rosenblatt, Fraunhofer IIS/EAS, Dresden
- M.Sc. Christian Fischer, Technische Universität Darmstadt
- Univ.-Prof. Dr.-Ing. Iris Gräßler, Universität Paderborn
- Dr.-Ing. Sven Herold, Fraunhofer LBF, Darmstadt
- Univ.-Doz. DI Dr. Michael Hofbaur, Joanneum Research, Klagenfurt
- Univ.-Prof. Dr.-Ing. Dr. h.c. Rolf Isermann, Technische Universität Darmstadt
- Prof. Dr.-Ing. Jürgen Kiel, Hochschule Düsseldorf
- Univ.-Prof. Dr. techn. Andreas Kugi, Technische Universität Wien
- Univ.-Prof. Dr. rer. nat. Tim Lüth, Technische Universität München
- Univ.-Prof. Dr.-Ing. Jürgen Maas, Technische Universität Berlin
- Dr.-Ing. Jörn Malzahn, Istituto Italiano di Tecnologia (IIT), Genova
- Dr.-Ing. Paul Motzki, ZeMA gGmbH, Saarbrücken
- Univ.-Prof. Dr.-Ing. Rainer Müller, ZeMA gGmbH, Saarbrücken
- Dr.-Ing. Rüdiger Neumann, Festo SE & Co. KG, Esslingen
- Univ.-Prof. Dr.-Ing. Tobias Ortmaier, Leibniz Universität Hannover
- Univ.-Prof. Dr.-Ing. Annika Raatz, Leibniz Universität Hannover
- Dr.-Ing. Christoph Rösmann, Technische Universität Dortmund
- Univ.-Prof. Dr.-Ing. Thomas Sattel, Technische Universität Illmenau
- Dr.-Ing. Martin Schmidt, AVL, Bensheim
- Univ.-Prof. Dr.-Ing. Dr. h.c. Dieter Schramm, Universität Duisburg-Essen
- Univ.-Prof. Dr.-Ing. Stefan Seelecke, Universität des Saarlandes
- Dr.-Ing. Gregor Stengel, Aquaduna, Sternenfels
- Prof. Dr.-Ing. Guido Stollt, Smart Mechatronics GmbH, Dortmund
- PD Dr.-Ing. Tom Ströhla, Technische Universität Ilmenau
- Dr. rer. nat. Thomas Tentrup, DÜRR Assembly Products GmbH, Püttlingen
- Univ.-Prof. Dr.-Ing. Ansgar Trächtler, Universität Paderborn
- Univ.-Prof. Dr.-Ing. Jörg Wallaschek, Leibniz Universität Hannover
- Dr.-Ing. Marian Walter, RWTH Aachen

MECHATRONIK, SO AKTUELL WIE EH UND JE

Die Ursprünge der Mechatronik liegen bereits mehr als 50 Jahre zurück. Damals wurde dieses Kunstwort geprägt, welches eine Synthese aus Mechanik und Elektronik zum Ausdruck bringt. Schon bald hielten Computer-programmierbare Funktionen Einzug in die mechatronischen Systeme, so dass deren Umfang fortan nicht nur Hardware, sondern auch Software umfasste. Die Mechatronik als interdisziplinäre Fachwissenschaft führt die Teildisziplinen Maschinenbau, Elektrotechnik und Informationstechnik zusammen und ist seit langem bestimmend für die Produkte unterschiedlichster Branchen wie der Automatisierungstechnik, der Luft- und Raumfahrttechnik, der Fahrzeugtechnik, der Medizintechnik oder der Energietechnik, die bei dieser Tagung angesprochen werden.

Im nächsten Evolutionsschritt wurden mehrere mechatronische Systeme lokal miteinander vernetzt, wie dies bspw. bei dem um die Jahrtausendwende erstmals in Serienfahrzeugen eingesetzten Abstandsregeltempomaten der Fall ist. Etwa ein Jahrzehnt später wurde dann der Begriff des Cyber-physischen Systems geprägt, das vom Grundsatz als global vernetze Mechatronik anzusehen ist und eine Verbindung realer technischer Systeme mit der virtuellen Welt herstellt. In diesem Zusammenhang gewinnt auch die Einbeziehung des Menschen eine immer größere Bedeutung, die in der Mechatronik als Mensch-Maschine-Schnittstelle bzw. Mensch-Mechatronik-Interaktion Berücksichtigung findet.

Aktuell ist ein großer Hype um die digitale Transformation und den Einsatz Künstlicher Intelligenz zu beobachten. Wenn insbesondere Maschinelles Lernen nicht nur in der Produktentwicklung, sondern auch während der Produkt-nutzung eingesetzt wird, stellt dies einen Paradigmenwechsel dar: Wurden zuvor alle Produkteigenschaften in der Entwicklungsphase festgelegt, abgesichert und freigegeben, so sollen nun optimierende Änderungen der Produkt-eigenschaften während der Nutzungsphase erfolgen. Dieser weitere Evolutionsschritt kann als selbstlernende Mecha-tronik bezeichnet werden, deren Herausforderungen in der Absicherung sich ändernder Eigenschaften von Produkten in Kundenhand liegt. Die Mechatronik steht somit für innovative, digitalisierte, vernetzte und lernende Produkte sowie Prozesse und ist damit so aktuell wie eh und je.

In bewährter Form liegen die Tagungsleitung, Organisation und Durchführung der 14. VDI/VDE Mechatronik-Tagung 2021 in der akademischen Welt mit einer festen fachlichen Verankerung in der VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik (VDI/VDE-GMA) und der VDI-Gesellschaft Produkt- und Prozessgestaltung (VDI-GPP). Das Tagungsprogramm umfasst eine große Bandbreite von Beiträgen aus Industrie und Wissenschaft und deckt die Themenfelder Automatisierte Mobilität, Modellbildung & Regelung mechatronischer Lenkungen, Nutzerfreundlichkeit & Akzeptanz, Robotik, Smarte Aktoren & Sensoren, Smarte Materialien & Systeme, Systems Engineering & Entwicklungsmanagement sowie Systemvernetzung & Systemintegration ab.

Ursprünglich sollte die 14. VDI/VDE Mechatronik-Tagung 2021 in der Digitalstadt Darmstadt und den Räumlichkeiten der TU Darmstadt stattfinden. Die Corona-Pandemie hat bedauerlicherweise dazu geführt, dass in diesem Jahr auf reale Begegnungen verzichtet werden muss und die Tagung nur im virtuellen Raum ausgerichtet werden kann. Das gewählte digitale Format soll an zwei Nachmittagen mit einzügigen Live-Pitch-Sessions inklusive Fragerunden sowie fortlaufenden Diskussionsforen auch unter den gegebenen Umständen einen attraktiven Rahmen bieten. Die Veranstalter Torsten Bertram (TU Dortmund), Burkhard Corves (RWTH Aachen), Klaus Janschek (TU Dresden) und Stephan Rinderknecht (TU Darmstadt) hoffen, dass sich Ihre Erwartungen an eine digitale Tagung erfüllen und Sie zwei interessante und bereichernde Tage erleben.

Im Namen der Veranstalter das Darmstädter Team

Sabine Backhaus, Birgit Lampert, Benjamin Blat Belmonte, Christian Fischer, Timo Hopf, Daniel Schöneberger, Stephan Rinderknecht,

Darmstadt im März 2021

TAGUNGSBEITRÄGE

SEITE

Session 1 - Robotik 1 (Kooperation, Kollaboration, Perzeption, …)

Wallbot: Robotersystem zur automatisierten Mauerwerkserrichtung, **1**
Robert Zickler, Christian Richter und Frank Will,
TU Dresden, Institut für Mechatronischen Maschinenbau

Automatisierte 3D-Umfeldmodellierung für eine effiziente Trajektorien Planung kollaborativer Roboter, **7**
Maximilian Krämer und Torsten Bertram,
Technische Universität Dortmund, Lehrstuhl für Regelungssystemtechnik

Bots2ReC: Technische Herausforderungen bei der automatisierten Asbestentfernung in Wohnräumen mittels mobiler Manipulatoren, **14**
Tobias Haschke, Mathias Hüsing und Burkhard Corves,
RWTH Aachen, Institut für Getriebetechnik, Dynamik und Robotik

Entwicklung eines autonomiefokussierten hochmobilen Bodenrobotersystems für den Katastrophenschutz, **20**
Marius Schnaubelt, Tobias Ullrich, Jonas Diegelmann, Matthias Hoffmann, Moritz Torchalla und Oskar von Stryk,
Technische Universität Darmstadt, Fachgebiet Simulation, Systemoptimierung und Robotik

Session 2 - Smarte Aktoren und Sensoren 1 (Konzepte, Beispiele, …)

Nutzen von Zusatzsensorik für die Unterscheidbarkeit von Modellen für elektrische Antriebsstränge bei der Struktur- und Parameteridentifikation, **26**
Mathias Tantau und Mark Wielitzka, Leibniz Universität Hannover, Institut für Mechatronische Systeme;
Lars Perner, Lenze SE, Braunschweig

Mechatronik: Risiken und Herausforderungen bei der Entwicklung eines permanentmagnetisch vormagnetisierten aktiven Magnetlagers, **32**
Timo Hopf und Stephan Rinderknecht,
Technische Universität Darmstadt, Institut für Mechatronische Systeme im Maschinenbau

Eine mechatronische Lösung gegen Positionierfehler in Werkzeugmaschinen infolge von schwankenden Umgebungsbedingungen, **38**
Christian Brecher und Filippos Tzanetos,
Fraunhofer Institut für Produktionstechnologie (IPT), Aachen

Applikation und Test eines dedizierten Schaltungssystems für elektrische Mehrgang-Fahrzeugantriebe, **43**
Daniel Schöneberger, Guanlin Gao und Stephan Rinderknecht,
Technische Universität Darmstadt, Institut für Mechatronische Systeme im Maschinenbau

Session 3 - Systems Engineering und Entwicklungsmanagement (Prozesse, Verfahren, Software, …)

Spezifikationsmethodik für opto-mechanische Systeme in der Nanolithografie, **49**
Marwene Nefzi, Stefan Hembacher und Jens Kugler

Mechatronischer Entwicklungsprozess eines automatischen Sensorfahrzeugs für den Außeneinsatz, **56**
Harald Staab, ABB AG Forschungszentrum Deutschland,
Eduardo Botelho, ABB Ltd. Brasilien,
Will Eakins und Harshang Shah, ABB Inc. USA

Systematische Vorgehensweise zur Anwendung der Digitaler Zwilling-Technologie bei Spezialgetrieben, **62**

TAGUNGSBEITRÄGE

SEITE

Peter Hehenberger, Edwin Kimpl, Michael Stroi, Clemens Voglhuber und Kiyoung Kwak, Fachhochschule Oberösterreich, Forschungsgruppe Smart Mechatronics Engineering

Anwendbarkeit von Requirement Mining in Benutzerrezensionen für die Entwicklung mechatronischer Produkte im B2C-Markt, **68**
Iris Gräßler und Daniel Preuß,
Universität Paderborn, Heinz-Nixdorf-Institut

Session 4 - Systemvernetzung und Systemintegration (Konzepte, Verfahren, …)

Featureextraktion aus heterogenen multivariaten Zeitreihendaten mechatronischer Systeme mittels Autoencoder-Netzwerken, **74**
Karl-Philipp Kortmann, Moritz Fehsenfeld, Mark Wielitzka und Tobias Ortmaier

Vergleich struktureller Totzeitberücksichtigung und kausaler Datenvorverarbeitung bei der datenbasierten Modellierung von Zeitreihen, **80**
Johannes Zumsande, Mark Wielitzka und Tobias Ortmaier

Maximierung der Anlageneffizienz in der Automobilproduktion mit Hilfe web-basierter IIoT-Applikationen, **86**
Quynh Nhu Truong und Thomas Tentrup

Maschinelles Lernen für die Situationserkennung einer aktiven Kniegelenkorthese, **92**
Philipp Mattfeldt, Bastian Latsch, Jürgen Hielscher, Niklas Schäfer, Romol Chadda und Mario Kupnik

Session 5 - Modellbildung und Regelung mechatronischer Lenkungen (Modelle, Regelung, Optimierung, Eco-Systeme, …)

Formale semantische Repräsentation regelungstechnischer Methoden, **98**
Robert Heedt, Carsten Knoll und Klaus Röbenack

Einsatz eines detaillierten Modells eines Lenksystems in hochdynamischen HiL-Tests von EPS-Motoren, **104**
Markus Hemmersbach, Hermann Briese, Michael Haßenberg, Marcus Irmer und Hermann Henrichfreise

Linear-Quadratic-Gaussian Position Control of the Hand Wheel Actuator for a Steer-by-Wire Steering System, **110**
Robert Gonschorek und Torsten Bertram

Robuste Lenkgefühlvorgabe für elektromechanische Servolenkungen, **116**
Marcus Irmer und Hermann Henrichfreise

Session 6 - Automatisierte Mobilität (Konzepte, Modelle, Sensorik, …)

Nichtlineares Zweispurmodell eines LKW-Sattelaufliegers mit experimenteller Validierung der lateralen und vertikalen Reifenkräfte, **122**
Simon Friedrich Gerhard Ehlers, Zygimantas Ziaukas, Jan-Philipp Kobler, Mark Wielitzka und Tobias Ortmaier

Fein-Positionierung mit reflektierenden Markern, **128**
Stefan-Octavian Bezrucav, Tobias Haschke, Markus Schmitz, Simon Klamt, Burkhard Corves, Jascha Paris und Stefan Kurtenbach

Fahrbahnzustandskarten aus der cloudbasierten Informationsfusion von Kamera- und Fahrdynamikinformationen, **134**
Alexander Busch, Tim Ahrenhold, Mark Wielitzka und Tobias Ortmaier

TAGUNGSBEITRÄGE

SEITE

Systemidentifikation und modellbasierte prädiktive Pfadfolgeregelung eines automatisierten knickgelenkten Muldenkippers, **140**
Arash Shahirpour, Christian Framing, René Zweigel und Dirk Abel

Session 7 - Nutzerfreundlichkeit und Akzeptanz (Assistenzsysteme, Schnittstellen, Interaktion, Gesetzgebung, ...)

Mechatronisches System für die modellbasierte Datenfusion zur Ermittlung des psychologischen Fahrerzustands, **146**
Alexander Hugenroth, Alexander Warnecke und Torsten Bertram

Untersuchung charakteristischer Parameter des Spurfolgeverhaltens zur Generierung einer Referenz für hochautomatisiertes Fahren, **152**
Laurin Vasile und Dieter Schramm

Fahrstilanalyse zur Erhöhung der Nutzerakzeptanz automatisiert energieeffizienter Fahrzeuglängsführung, **158**
Daniel Fink, Mark Wielitzka, Tobias Ortmaier, Christoph Schweers und Ahmed Trabelsi

Robotergestützte Handhabung im kryogenen Arbeitsumfeld, **164**
Philipp Jahn, Frank Ihmig, Margit Biehl und Annika Raatz

Session 8 - Smarte Materialien und Systeme (Konzepte, Beispiele, …)

Konstruktion eines Festkörpergelenkgreifers mit bistabiler Formgedächtnisaktorik, **170**
Dominik Scholtes, Stefan Seelecke und Paul Motzki

Aktor- und Sensorelemente basierend auf Dielektrischen Elastomeren zur Verwendung in tragbaren Assistenztools, **175**
Sebastian Gratz-Kelly, Sophie Nalbach, Paul Motzki und Stefan Seelecke

In-Situ-Auswuchten rotierender Massen auf Basis bewegungsgesteuerter magnetischer Fluide, **181**
Valentin Schreiner und Jürgen Maas

Session 9 - Robotik 2 (Kooperation, Kollaboration, Perzeption, …)

Impedanzregelung eines Luftmanipulationssytems bestehend aus einem Multirotor und einem Sechs-Gelenk-Manipulator, **187**
Chao Yao, Stephan Schwarz und Klaus Janschek

Teach-In für das kraftsensitive automatische Wickeln eines Seiles im dreidimensionalen Raum, **193**
Michael Schwegel, Hanspeter Augschöll und Andreas Kugi

Ein Lösungsansatz zur Verbesserung der Genauigkeit von kraftgeregelten Applikationen basierend auf einem seriellen Roboter mit integrierter Sensorik, **199**
Rainer Müller, Ali Kanso und Marco Schneider

Kraftsensitive Kalibriermethode für Industrieroboter, **205**
Stefan Schütz, Nikolaus Elsner, Christian Henke und Ansgar Trächtler

Entwicklung einfacher modularer Seilroboter ohne Verwendung von Kraftsensorik, **211**
Jonas Bieber, David Bernstein, Micha Schuster, Michael Beitelschmidt

PLENARVORTRAG:

Robotics in conjunction with mechatronics /Mechano-Informatics for Humanoid Robots with Motion Intelligence

Prof. Dr.-Ing. Tamim Asfour
Wissenschaftlicher Sprecher KIT – Zentrum Information-Systeme (KCIST)

Abstract:
Humanoid robotics play a central role in robotics research as well as in understanding intelligence. Engineering humanoid robots that are able to learn from humans and sensorimotor experience, to predict the consequences of actions and exploit the interaction with the world to extend their cognitive horizon remains a research grand challenge. Currently, we are experiencing AI systems with superhuman performance in games, image and speech processing. However, the generation of robot behaviors with human-like motion intelligence and performance has yet to be achieved. In this talk, I will present recent progress towards engineering 24/7 humanoid robots that link perception and action to generate intelligent behavior. I will show the ARMAR humanoid robots as example for synergetic integration of informatics, artificial intelligence and mechatronics methods to create versatile systems that are able to perform complex grasping and manipulation tasks in kitchen and industrial environments, to learn actions from human observation and experience and to reason about object-action relations.

Curriculum vitae:

Tamim Asfour is full Professor of Humanoid Robotics at the Institute for Anthropomatics and Robotics at the Karlsruhe Institute of Technology (KIT). His research focuses on the engineering of high performance 24/7 humanoid robotics as well as on the mechano-informatics of humanoids as the synergetic integration of informatics, artificial intelligence and mechatronics into humanoid robot systems, which are able to predict, act and interact in the real world. In his research, he is reaching out and connecting to neighboring areas in large-scale national and European interdisciplinary projects at the intersection of robotics, machine learning and computer vision. Tamim is the developer of the ARMAR humanoid robot family. He is scientific spokesperson of the KIT Center “Information · Systems · Technologies (KCIST)”, president of the Executive Board of the German Robotics Society (DGR), the Founding Editor-in-Chief of the IEEE-RAS Humanoids Conference Editorial Board, deputy Editor-in-Chief and Editor of the Robotics and Automation Letters.
→www.humanoids.kit.edu



PLENARVORTRAG:

Auf dem Weg zu einer nachhaltigeren Luftfahrt – Digitalisierung und Elektrifizierung bei Rolls-Royce

Heiko Witte
Business Development Manager Digital, Rolls-Royce

Abstract:
Der Klimawandel stellt die Luftfahrt vor große Herausforderungen. Gleichzeitig eröffnen Technologien zur Elektrifizierung der Antriebssysteme ebenso wie die Digitalisierung Möglichkeiten für einen Wandel in der Luftfahrtindustrie. Unterstützt durch Förderprogramme werden bei Rolls-Royce Deutschland zusammen mit Partnern aus Forschung und Industrie entsprechende Projekte durchgeführt, bei denen durch digitale Integration zwischen Entwicklung und Produktion ein größeres Produktverständnis erreicht wird. Ebenso werden neuartige Antriebssysteme konzipiert.
In der Keynote wird darauf eingegangen, welche Lösungsansätze existieren, um die Luftfahrt nachhaltiger und klimafreundlicher zu machen, sowie Beispiele zur Einführung von digitalen Lösungen wie auch elektrischeren Antrieben diskutiert.

Curriculum vitae:

Heiko Witte hat Umfangreiche Berufs- und Führungserfahrung in Luft- und Raumfahrt, Projektmanagement, Business Transformation, Kontinuierlicher Verbesserung und Digitalisierung.
Seit 1997 ist er bei BMW Rolls-Royce / Rolls-Royce Deutschland tätig. Dort arbeitet er mit Erfahrung und Verantwortung in verschiedenen Bereichen, u.a. Triebwerksversuchen, Projektmanagement, Produktentwicklung, Kontinuierliche Verbesserung und Digitalisierung. Ebenso war er Projektleiter seitens Rolls-Royce bei der Produkteinführung des V2500Select-Triebwerks (Antrieb der Airbus 320 Familie).



Ab 2009 übernahm er den Head of Engineering Improvement and Quality, in dieser Rolle seit 2017 auch die Koordination und Projektleitung von Digitalisierungs- und Industrie 4.0 – Projekten. Seit Beginn des Jahres 2020 ist er Business Development Manager Digital für Rolls-Royce, zuständig für digitale Transformation bei laufenden und zukünftigen Produktentwicklungen.
Er weist weitreichende Erfahrungen bei Entwicklung und Umsetzung von Kanban (Flowline)- und Agiler Methodik in der Entwicklungsumgebung auf. Er hält Vorträge bei verschiedenen Konferenzen in Deutschland und Europa (u.a. Product Innovation – PI, London Tech Week).

Wallbot: Robotersystem zur automatisierten Mauerwerkserrichtung
Wallbot: Robotic system for the automated construction of masonry

Dipl.-Ing. Robert Zickler, TU Dresden, Institut für Mechatronischen Maschinenbau, Stiftungsprofessur für Baumaschinen, 01187 Dresden, Deutschland, robert.zickler@tu-dresden.de

Dipl.-Ing. Christian Richter, TU Dresden, Institut für Mechatronischen Maschinenbau, Stiftungsprofessur für Baumaschinen, 01187 Dresden, Deutschland, christian.richter1@tu-dresden.de

Prof. Dr.-Ing. Frank Will, TU Dresden, Institut für Mechatronischen Maschinenbau, Stiftungsprofessur für Baumaschinen, 01187 Dresden, Deutschland, frank.will@tu-dresden.de

Kurzfassung

Der Beruf des Maurers gehört bereits seit Jahren, aufgrund des fehlenden Nachwuchses, des demografischen Wandels und der körperlich schweren Arbeit, zu den am schwersten zu besetzenden Positionen im Baugewerbe. Für die betroffenen Bauunternehmen ergibt sich hieraus zunehmend der Druck, ihre Arbeitsprozesse zu automatisieren und die Arbeiter zu entlasten. Das größte Problem hierbei ist bisher das Fehlen marktfähiger technischer Lösungen. Im Rahmen eines Forschungsprojektes entwickelt die TU Dresden gemeinsam mit der SKM GmbH einen praxistauglichen Mauerroboter - Wallbot. Mit der geplanten Lösung soll neben gleichbleibender Qualität bei hoher Produktionsleistung eine Entlastung der Beschäftigten erzielt werden. Die wesentliche Innovation in dieser Entwicklungsaufgabe ist der ganzheitliche Ansatz der Lösung, bestehend aus Aufnahmeeinheit, Bindemittelbereitstellung, Materialverarbeitung sowie einem integrierten Versorgungs- und Sicherheitskonzept. Die Praxistauglichkeit und Funktionalität werden durch den Einsatz von Leichtbaukomponenten, eine autarke Energieversorgung und die Entwicklung eines neuartigen Positioniersystems erreicht.

Abstract

The job of bricklayer has been one of the hardest to fill in the construction industry for years due to the lack of young talent, demographic change and the physically difficult work. For the construction companies affected, this results in increasing pressure to automate their work processes and to relieve the workers. The biggest problem here is the lack of marketable technical solutions. The TU Dresden is developing a practical wall robot - Wallbot - together with the SKM GmbH as a research partner. With the planned solution, in addition to consistent quality with high production output, a relief for employees is to be achieved. The main novelty in this development task is the holistic approach of the solution, consisting of a handling unit, binding agent provision, material processing as well as an integrated supply and safety concept. The practicality and functionality are achieved through the use of lightweight components, an autonomous energy supply and the development of a new positioning system.

1 Einleitung

Das Ziel des bilateralen ZIM-Forschungsprojektes Wallbot (FKZ ZF4667402PO9) zwischen der Stiftungsprofessur für Baumaschinen der Technischen Universität Dresden und der SKM GmbH ist es, einen vollautomatisierten Mauerwerksroboter zu entwickeln, der alle Anforderungen eines praxistauglichen und wirtschaftlichen Baustelleneinsatzes erfüllt.
Insgesamt soll das System alle Prozessschritte des klassischen Mauerwerkbaus abdecken, d. h. auch Nebenprozesse, wie das Sägen und Bemörteln der Steine werden automatisch ausgeführt. Die Planungsdaten werden dabei aus einem digitalen Gebäudemodell abgeleitet und der aktuelle Baufortschritt aufgezeichnet.
Hauptaugenmerk auf Seiten der TU Dresden liegt auf der Entwicklung der Steuerung und Sensor-Aktor-Kommunikation sowie der Erprobung der Komponenten.

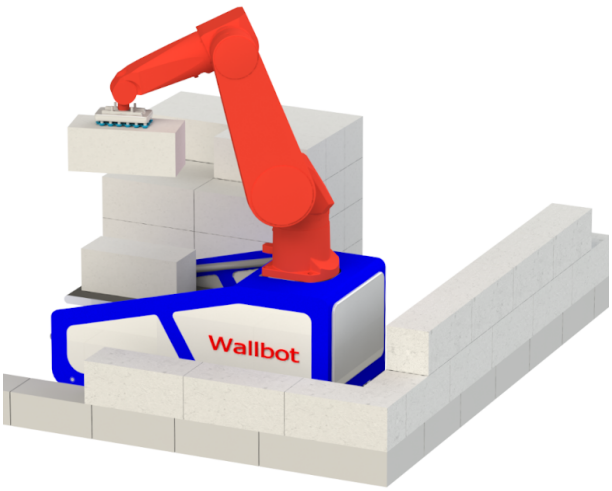


Abbildung 1 Wallbot-Konzeptentwurf

2 Stand der Technik

Ein etabliertes Werkzeug für den Mauerwerksbau ist das *Versetzgerät* (siehe **Bild 2**), welches sogenannte Steinstangen mit einer Breite von bis zu 2 m auf Deckenhöhe befördern und in einem Arbeitsradius von 5 m versetzen kann [1]. Es werden nur Steine verarbeitet, bei denen die vertikale Stoßfuge nicht vermörtelt wird, d. h. Steine mit Nut-Feder-System. Unterschiedliche Greifzangen ermöglichen das Greifen der im Optimalfall bereits in der geforderten Verarbeitungsbreite angelieferten Steine. Das Versetzgerät wird in einem Team von zwei Arbeitern bedient, wobei einer für den Mörtelauftrag sowie das Ausrichten der Steine verantwortlich ist und ein Zweiter das Gerät steuert und für den Materialnachschub sorgt.



Abbildung 2 Klassischer Mauervorgang mit Versetzgerät. [2]

Das Institut für Steuerungstechnik der Werkzeugmaschinen und Fertigungseinrichtungen sowie das Zentrum Fertigungstechnik der Universität Stuttgart entwickelten im Jahr 1997 den *BRONCO* (bricklaying robot for the use on the construction site; siehe **Bild 3**), welcher sich auf der Baustelle orientieren, Verlegepläne verarbeiten sowie Steine mit einer Masse von bis zu 50 kg bemörteln und setzen konnte. [3, 4]
Das System war dabei jedoch zu schwer um auf nicht durchgehärteten Geschossdecken zu arbeiten.

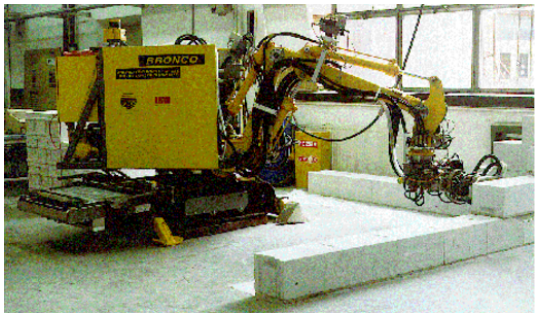


Abbildung 3 BRONCO beim Verlegen von Steinen. [4]

Die Professur für Architektur und Digitale Fabrikation der ETH Zürich beschäftigt sich u. a. mit dem Thema der robotergestützten Mauerwerkserstellung. Dabei wurden unter dem Titel *dimRob* untersucht, wie sich Unsicherheiten beim Erstellen von Mauerwerk mit toleranzbehafteten Werkstücken reduzieren lassen, als auch, wie die sichere Interaktion mit Menschen in diesem Umfeld gesteigert werden kann [5]. Dabei wurde ein Industriemanipulator auf einem Raupenfahrwerk montiert, der sich dadurch manuell im Raum neu positionieren lässt.
Der *In Situ Fabricator* (siehe **Bild 4**) ist ebenfalls ein auf einem Raupenfahrwerk montierter Knickarmroboter [6]. Hier ist zusätzlich die Steuerung auf der Plattform untergebracht und das Fahrwerk wird von der Anlage aktiv genutzt, um sich im Raum zu positionieren. Ein kleiner Steinvorrat wird in einem Magazin gespeichert und verfährt mit dem Roboter.



Abbildung 4 In Situ Fabricator der ETH Zürich. [7]

Ein weiteres System wird von der Firma *Fastbrick Robotics* mit dem *Hadrian X* (siehe **Bild 5**) entwickelt. Bei diesem werden Steine von einem Speicherwagen über einen teleskopierbaren Ausleger mit integriertem Fördersystem transportiert und in der Mauer platziert. Innerhalb des Wagens befindet sich neben einem Vorratsspeicher auch eine Säge zum Kürzen der Steine.
Das System arbeitet mit proprietären Betonhohlblocksteine und einem speziellen Kleber [8]. Das Erfassen der Endeffektorposition erfolgt mithilfe motorisierter Tachymeter, die vornherein eingemessen werden und zu denen immer eine direkte Sichtverbindung bestehen muss.



Abbildung 5 Hadrian X von Fastbrick Robotics. [9]

Was unterscheidet Wallbot? Wie die aufgeführten Systeme zeigen, ist der Mauerwerkprozesses prinzipiell automatisierbar, jedoch gelang der Schritt zum serienmäßigen Einsatz auf der Baustelle bisher noch nicht. Hohe Investitionskosten, ungenügendes Einhalten von Toleranzen, eingeschränkte Manövrierbarkeit und ein zu hohes Eigengewicht können dabei als Hauptargumente gegen den Einsatz bereits entwickelter Systeme genannt werden. Mit Wallbot wird eine praxisorientierte und gesamtheitliche Lösung für das Mauern von Wänden angestrebt.
Im aktuellen Entwicklungsstand erfolgt das Verlegen der untersten Schicht (Kimmschicht) weiterhin manuell (siehe **Bild 6**), da bei diesem Prozessschritt eine Bitumenabdichtbahn und Normalmörtel verarbeitet und die Toleranzen der Grundplatte beachtet werden müssen. Der weitere Aufwand für den Arbeiter beschränkt sich auf das Auffüllen der Verbrauchsmaterialien (Dünnbettmörtel, Wasser und Energieträger). Im Arbeitszustand der Maschine können vom Bediener andere Arbeiten durchgeführt werden. Das können Vor- oder Nacharbeiten für weitere Bauabschnitte sein, als auch das Überwachen und die Kontrolle des Mauerprozesses.



Abbildung 6 Die Kimmschicht wird manuelle verlegt.

3 Die Hardware

Für den Einsatz ist eine hohe Manövrierbarkeit zwischen den Paletten und dem Mauerwerk, als auch durch enge Gänge und Türen erforderlich. Des Weiteren führen die Notwendigkeit des Transports sowie die Versetzbarkeit mittels Ladekrans dazu, dass die Grundabmessungen von Wallbot der einer Europoolpalette entsprechen. Das Gesamtgewicht wird unter der zulässigen Belastungsgrenze für nicht vollständig durchgehärteter Decken gehalten.

3.1 Manipulator und Endeffektor

Als Manipulator wird ein TX2 160 von der Firma Stäubli verwendet (siehe **Bild 7**). Dieser besitzt ein Eigengewicht von 260 kg, eine Reichweite von 2 m und eine Traglast von bis zu 34 kg [10]. Nach Abzug der angestrebten Werkzeug-Eigenmasse ist der Manipulator immer noch in der Lage, Steine mit einer Masse von bis zu 30 kg zu versetzen. Gleichzeitig ist es durch die erhöhte Montageposition möglich, Steine sowohl auf dem Boden als auch auf Deckenhöhen bis 2,7 m zu platzieren.



Abbildung 7 TX2 160 von Stäubli. [10]

Für das sichere Erproben der Algorithmen wird vorerst der kompaktere Manipulator *UR10* von *Universal Robots* verwendet, mit dem jedoch noch nicht jedes Steinformat gehandhabt werden kann.
Der Endeffektor wird für die Aufnahme, Reinigung und Detektion der Porenbetonsteine verwendet. Soll ein Stein von der Palette aufgenommen werden, so beseitigt zuvor eine integrierte Ausblaseinrichtung eine mögliche Staubschicht, welche sich aufgrund von Reibung der Steine beim Transport bildet. Der staubbefreite Stein wird anschließend mittels Vakuumerzeugung durch Ejektoren angesaugt und am Ablageort abgelegt. Für Wallbot entstand dafür eine Eigenentwicklung, die es erlaubt, neben einer kompakten und leichten Bauform auch eine optimale Sensorpositionierung zu ermöglichen.

3.2 Steinverarbeitung

Das größte Unterscheidungsmerkmal zu anderen Mauerrobotern ist die integrierte Verarbeitungsstrecke, welche es Wallbot ermöglicht auf detektierte Werkstücktoleranzen reagieren zu können. Dadurch können auch Steine angepasst und gesetzt werden, die nicht dem Standardmaß entsprechen und vor allem an Maueröffnungen für Fenstern und Türen oder Mauerenden benötigt werden.
Jeder Stein wird vor dem Transport von der Palette vermessen und nach der Ablage in der Mauer mit den Mauer-Soll-Maßen verglichen. Mit der Säge kann daraufhin das benötigte Maß des Folgesteins und somit die geforderte Maueform hergestellt werden. Durch die Möglichkeit die Säge zu verdrehen, können auch Steine mit stufenförmigen Ausschnitten - z. B. für die Aufnahme von Fensterstürzen - oder diagonalen Schnittkanten, für nicht-rechtwinklige Raumecken oder Giebel, hergestellt werden.
Vom Mischen bis zum Aufbringen des Mörtels mit einer Düse, sind alle dafür benötigten Komponenten in dem System integriert. Eine verstellbare Düse ermöglicht den variablen Materialauftrag auf mehreren Steinseiten oder für unterschiedliche Abmessungen. Für Porenbetonsteine wird Dünnbettmörtel verwendet, der in einer Stärke von 1 mm aufgebracht wird. Im Gegensatz zum manuellen Arbeitsprozess geschieht der Auftrag jedoch auf den zu verlegenden Stein und nicht auf die bereits verlegten Mauersteine.

3.3 Leistungsversorgung und Antrieb

Für den Manipulator ist eine dreiphasige 400 V-Versorgung sowie für die weiteren Verbraucher, wie die Säge- und Räderantriebe, der Steuerungstechnik und Sensoren entsprechende weitere Energieniveaus notwendig. Im aktuellen Entwicklungsstand ist eine externe Energieversorgung mit einem Schleppkabel vorgesehen. Perspektivisch ist eine rein elektrische Versorgung über Akku oder Brennstoffzelle vorgesehen.
Um die erforderliche Manövrierbarkeit von Wallbot zu gewährleisten werden Mecanum-Räder eingesetzt, welche einzeln angesteuert eine omnidirektionale Bewegung erlauben. Somit lassen sich ohne eine mechanische Lenkung beliebige Bewegungsrichtungen erzielen und auch das Drehen am Platz ist möglich. Aktuell werden Tests an einem Modell (siehe **Bild 8**) durchgeführt um die Bewegungsplanung zu erproben. Durch den komplexeren Radaufbau ergeben sich Fahrfehler, welche mit einer geeigneten Regelung entgegengewirkt werden soll.

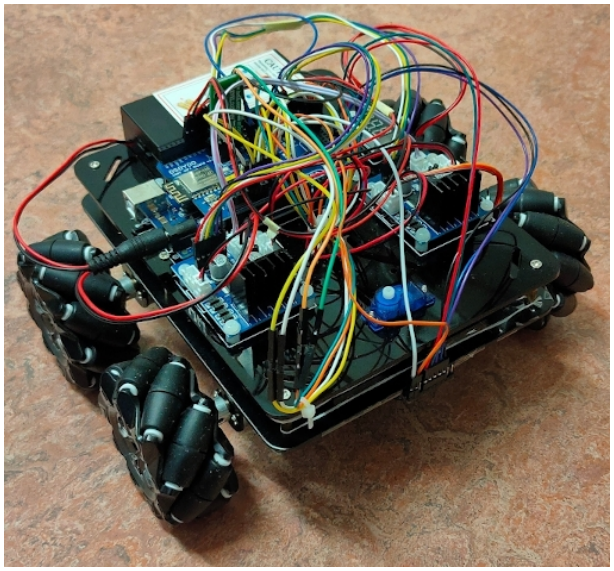


Abbildung 8 Modell zum Testen der Mercanum-Räder.

4 Sensorik

Aufgrund der großen Bewegungsfreiheit im Raum und der direkten Steininteraktion, sind am letzten Glied des Manipulators diverse Sensoren montiert um das System zu lokalisieren und Kollisionen vorzubeugen. Dabei liefert eine Tiefenbildkamera (siehe **Bild 9**) Informationen über die grobe räumliche Entfernung von Objekten. Die Farbwertinformationen der Kamera werden im aktuellen Entwicklungsstand zusätzlich genutzt, um Apriltag-Marker am Fuß der Paletten zu detektieren und auf deren Lage sowie die der Steine schlussfolgern zu können.
Um die Lage des dadurch grob georteten Steins besser zu bestimmen, wird anschließend ein Abstandssensor eingesetzt um über eine Messung von drei Seiten den Stein exakt lokalisieren und verarbeiten zu können. Die Auswahl der Sensoren wurde auf Grundlage von Laborversuchen hinsichtlich Störanfälligkeit, Kontrastfähigkeit, Gegenlichtempfindlichkeit und Genauigkeit durchgeführt.



Abbildung 9 Die Tiefenbildkamera *RealSense D435i* von *Intel* (links) sowie der Abstandssensor von *Micro-Epsilon* am Greifer.

Des Weiteren ist zwischen Manipulator und Endeffektor eine Kraft-Drehmoment-Messdose angebracht, mit der mögliche Kollisionen detektiert und das sichere Ablegen des Steins auf der Bearbeitungsplattform und dem Mauerwerk gewährleistet werden.
Um die horizontale Steinablage zu ermöglichen wird außerdem eine Inertiale Messeinheit (IMU) auf der Plattform untergebracht. Damit kann das System trotz möglicher Abweichung von der horizontalen Lage, die Steine ideal verlegen. Am Fahrmodul werden 2D-Lidarsensoren verbaut, die zusammen eine 360°-Überwachung der Wallbot-Umgebung ermöglichen.

5 Die Software

Der umfangreiche Bearbeitungsprozess erfordert ein reibungsloses Zusammenspiel der einzelnen Aktoren und Sensoren. Gleichzeitig soll die Möglichkeit der Modularisierung gegeben sein, um das System erweitern oder Komponenten austauschen zu können und dabei die erarbeiteten Programmcodes in anderen Entwicklungsprojekten wiederzuverwerten. Die Open Source Middleware *Robot Operation System* (ROS) erfüllt all diese Anforderungen und stellt für Wallbot die Software-Infrastruktur.

5.1 ROS als Middleware für die Daten- und Informationsverwaltung

Durch ROS wird eine Publisher-Subscriber-Infrastruktur aufgebaut, bei der anfallende Daten systemweit registriert, verwaltet und zur Verfügung gestellt werden. Jedem Systemteilnehmer ist es dabei gestattet Informationen zu abonnieren und zu veröffentlichen. Die Sensoren und Aktoren sind dabei sternförmig über digitale oder analoge Schnittstellen mit einem zentralen Industrie-PC verbunden, auf dem ROS unter dem Betriebssystem Ubuntu Linux läuft.

Daten und Informationen, die nicht für alle Teilnehmer relevant sind und eine unnötige Systembelastung mit sich bringen, werden auf spezielle Hardware-Knoten ausgelagert, die beispielsweise für die Steuerung des Fahrmoduls oder die Bildverarbeitung zuständig sind (siehe **Bild 10**). Für viele Sensoren und Aktoren bestehen bereits frei zugängliche ROS-Pakete, wodurch auf den Einsatz proprietärer Software verzichtet werden kann. Eine umfangreiche Bibliothek mit weit über einhundert Industriemanipulatoren vereinfacht deren Integration und Austauschbarkeit.
Gibt es für eine Komponente kein entsprechendes Paket, lässt sich dieses mit Python- oder C++-Skripten erstellen. Auch werden die Ablaufsteuerung, Datenverarbeitung und Visualisierung mit diesen Sprachen programmiert.

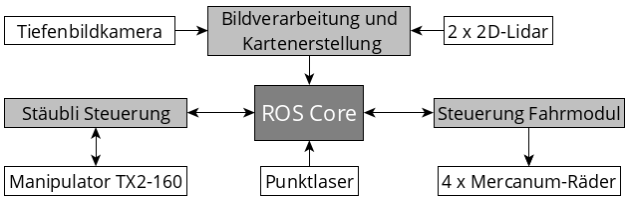


Abbildung 10 Vereinfachte Darstellung des Datenflusses. Der ROS Core koordiniert die einzelnen Knoten, welche wiederum die Kommunikation mit den Sensoren und Aktoren steuern.

5.2 Digitale Planungsdaten

Um Wallbot die zu fertigende Struktur mitzuteilen, wird ein *BIM*-Modell (building information modeling; siehe **Bild 11**) überspielt, welches im Dateiformat *IFC* (Industry Foundation Classes) beschrieben ist. Dieses wird zuvor aus dem Gebäudemodell mit den zu verwendenden Steinen automatisch generiert. In ROS wird das Modell verarbeitet und die Soll-Koordinaten der Steine sowie eine Bearbeitungsreihenfolge abgeleitet. Parallel wird zur Qualitätskontrolle ein Ist-Modell erstellt, welches im Bedienpanel kontinuierlich den aktuellen Bearbeitungsstand festhält.

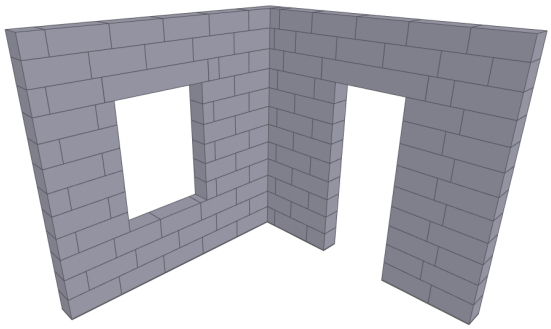


Abbildung 11 BIM-Modell mit Porenbetonsteinen.

5.3 Informationen aus Tiefenbildern

Die Tiefenbildkamera stellt neben Farb- auch Tiefeninformationen bereit, welche sich in eine Punktwolke überführen lassen (siehe **Bild 12**). Aus diesen Daten wird durch das ROS-Paket *RTAB-Map* (Real-Time Appearance-Based Mapping) eine 2D-Belegtheitskarte erstellt, in der Wallbot kollisionsfrei verfahren kann. Um zu gewährleisten, dass z. B. Türstürze den Bereich darunter nicht sperren, wird eine maximale Hindernishöhe definiert, bis zu der Objekte berücksichtigt werden.
Für das Überwachen der direkten Umgebung wird die 2D-Karte um eine diskretisierte 3D-Karte (Octomap) ergänzt. Neben dem Planer für das Fahrmodul nutzt diese Informationen auch das ROS-Paket *MoveIt* für die Pfadplanung des Manipulators. Um die Selbstdetektion zu vermeiden, welche eine fehlerhaft detektierte Kollision sowie ein Verharrungszustand nach sich zieht, wird die eigene Geometrie aus der Octomap herausgefiltert.

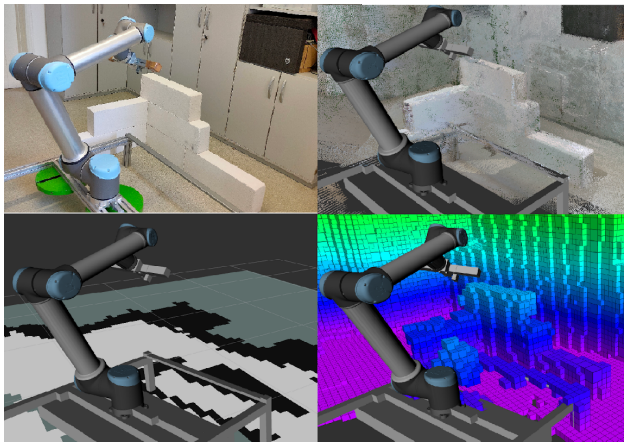


Abbildung 12 Die reale Situation sowie die sich aus den Tiefenbildern ableitende Umgebungsinformationen: Punktwolke (o. r.), 2D-Belegtheitskarte (u. l.) sowie Octomap (u. r.). Hier noch ohne das Herausfiltern der Robotergeometrie.

5.4 Pfadplanung mit MoveIt

Das ROS-Paket *MoveIt* bildet eine umfangreiche Erweiterung für die Pfadplanung und Steuerung von seriellen kinematischen Strukturen. Es werden dafür rotatorische und translatorische Glieder mit deren kinematischen und dynamischen Eigenschaften vorgeben. Dabei können die Glieder beliebig angeordnet werden und auch Verzweigungen sind möglich. Im Hintergrund nutzt MoveIt für die Pfadplanung die *Open Motion Planning Library* (OMPL) es kann jedoch auch ein vom Hersteller bereitgestellter oder eigener Planer eingebunden werden. Mit MoveIt ist es möglich sowohl einen als auch mehrere reale oder simulierte Manipulatoren oder beides gleichzeitig anzusteuern.

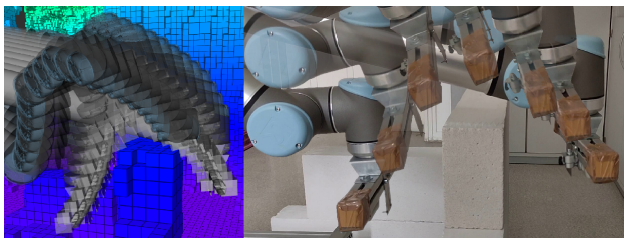


Abbildung 13 Erkannte Hindernisse (Octomap) werden mit dem Pfadplaner MoveIt automatisch umfahren.

Sobald Wallbot seine Pose auf der Baustelle kennt, ergeben sich aus den digitalen Planungsdaten die Sollposen der Mauersteine. Nach der Detektion einer oder mehrerer Paletten und der sich daraus ergebenden Steinposen, sind alle Soll-Werkzeugposen bekannt, wodurch MoveIt die benötigten Gelenksollwerte ermittelt und an die Manipulatorsteuerung überträgt. Detektierte oder vordefinierte Hindernisse (z. B. Sperrzonen) werden für die Pfadplanung berücksichtigt und entsprechend behandelt (siehe **Bild 13**). Wird ein Objekt - z. B. ein Stein - aufgenommen, so kann dies MoveIt mitgeteilt und für die Pfadplanung berücksichtigt werden.

6 Test des Greifersystems

Die meisten Systeme (siehe Kapitel 2) nutzen einen mechanischen Greifer um die Steine zu bewegen. Dies ermöglicht einen sicheren Griff, bringt aber Nachteile mit sich, wenn die Steine von der Palette separiert werden müssen. Auch ist für das Handhaben verschiedener Steinformate ein komplexer Mechanismus erforderlich. Es wurde sich deshalb für ein System entschieden, welches mittels Unterdruck den Stein ansaugt. Durch Aufteilen in mehrere Saugkammern und das automatische Abschalten von nicht belegten Kammern, können so auch Steine unterschiedlichsten Formates gegriffen werden.

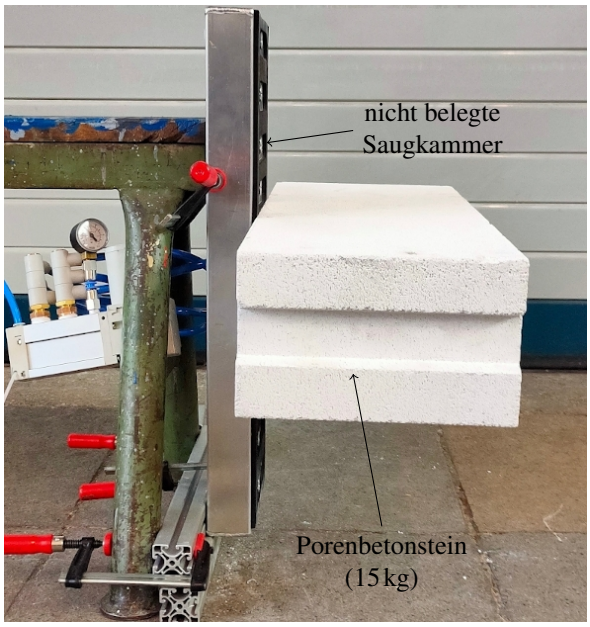


Abbildung 14 Ansaugversuch zum seitlich gegriffenen Stein bei nicht vollständig belegter Greiferoberfläche.

In ersten Versuchen konnte gezeigt werden, dass die raue Steinoberfläche den Evakuierprozess nicht nachteilig beeinflusst und nicht belegte Saugkammern sich ausreichend sperren lassen (siehe **Bild 14**). Die angesaugte Steinmasse überstieg die für den Manipulator maximal zulässige Traglast von 30 kg um das Vierfache (siehe **Bild 15**), wodurch sich weitere Optimierungen für des Greifersystems ergeben.



Abbildung 15 Der vom Ejektor erzeugte Unterdruck hält die mit Masse beaufschlage Platte, welche die Steinoberfläche nachbildet.

7 Zusammenfassung

Es kann bisher gezeigt werden, dass die generelle Machbarkeit eines Mauerroboters, der Steine vom Modell bis zur fertigen Wand verarbeitet, möglich ist. Da bisher nur die einzelnen Baugruppen des Systems erprobt wurden, sind Tests des Gesamtsystems erforderlich. Einzelkomponenten befinden sich derzeit in der Fertigung bzw. Beschaffung. Perspektivisch ist geplant, die gewonnenen Erkenntnisse für die Automatisierung weiterer Baustellenprozesse zu nutzen. Auch ist eine modulare Plattform denkbar, mit der verschiedene Prozesse bearbeitet werden können.

8 Literatur

[1] Wessig, J.: *KS-Maurerfibel*. Düsseldorf: Beton-Verlag GmbH, 1992.

[2] *Callena Bau: Neubau*. http://callena-bau.de/neubau.html (Zugriff am 15.1.2020)

[3] *Mauerroboter auf der BAUMA '98: BRONCO war der Messehit*. Stuttgarter unikurier Nr. 79 (Juni 1998).

[4] Rudlaff, T. Kurz, J. und Scheurer, F.: *Baustellengerechte Bedienung und wirtschaftlicher Einsatz von mobilen Robotern zur automatisierten Erstellung von Mauerwerk auf der Baustelle*. Zentrum Fertigungstechnik Stuttgart, 1997.

[5] Helm, V. et al.: *Mobile robotic fabrication on construction sites: DimRob*. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 7-12 Oktober 2012 Vilamoura. ISBN:978-1-4673-1736-8.

[6] Giftthaler, M., et al.: *Mobile robotic fabrication at 1:1 scale: the in situ fabricator*. Construction Robot 1. 2017. https://doi.org/10.1007/s41693-017-0003-5.

[7] NCCR Digital Fabrication (DFAB): *In-situ Fabricator: An autonomous construction robot*. Robohub.org, 2015.

[8] Smisek, P.: *Inside the Hadrian X Bricklaying Robot*. The B1M, 2019.

[9] Fastbrick Robotics: *FBR: Hadrian X* https://www.fbr.com.au/view/hadrian-x (Zugriff am 11.1.2021).

[10] Stäubli: *TX2-Roboterbaureihe, Robotics Experts in Man and Maschine*. https://www.staubli.com/de-de/file/21193.show (Zugriff am 11.1.2020)

Automatisierte 3D-Umfeldmodellierung für eine effiziente Trajektorienplanung kollaborativer Roboter

An Automated Approach for Modelling the 3D-Environment of Collaborative Robots for Efficient Trajectory Planning

Maximilian Krämer, Technische Universität Dortmund, Lehrstuhl für Regelungssystemtechnik, 44227 Dortmund, Deutschland, maximilian.kraemer@tu-dortmund.de

Univ.-Prof. Dr.-Ing. Prof. h.c. Dr. h.c. Torsten Bertram, Technische Universität Dortmund, Lehrstuhl für Regelungssystemtechnik, 44227 Dortmund, Deutschland, torsten.bertram@tu-dortmund.de

Kurzfassung

Ein Vorteil von kollaborativen Robotern sind ihre Sicherheitsfunktionen und geringen Bewegungsenergien weshalb sich auch Personen innerhalb des Arbeitsraums befinden dürfen und sie daher effizient an wechselnden Orten eingesetzt werden. Sowohl Personen, welche die Umgebung des Roboters verändern, als auch Ortswechsel erfordern, dass Umgebungsmodelle für die Trajektorienplanung aktualisiert oder neu erstellt werden müssen. Dieser Beitrag stellt ein dreistufiges Verfahren zur automatischen Rekonstruktion und Modellierung der statischen Umgebung mit einer robozentrischen 3D-Kamera in Auge-in-Hand-Konfiguration vor. Die Segmentierung kann erfolgreich feine Cluster ermitteln, welche der Granularität der Umgebung entsprechen. Aus den drei Verfahren zur Zuweisung der Hüllkörper zu Clustern liefert die Methode der algorithmischen Geometrie die robustesten Ergebnisse und agiert im Bereich von Millisekunden.

Abstract

One advantage of collaborative robots are their safety features and low kinetic energies, allowing people to be inside the workspace and thus enabling the robots to be deployed efficiently in different locations. Both, people changing the robot's environment and robot relocation require environment models to be updated or recreated for trajectory planning. The following paper presents a three-step method for automatic reconstruction and modeling of the static environment based on a robocentric 3D-camera in eye-in-hand configuration. The segmentation step can successfully obtain fine clusters that match the granularity of the environment. Out of three methods for assigning bounding volumes to clusters, the algorithmic geometry method provides the most robust results and operates in the range of milliseconds.

1 Einleitung

Für die Planung kollisionsfreier Robotertrajektorien werden Umgebungsmodelle benötigt. Die Umgebungen von kollaborativen Robotern sind so vielfältig wie ihre Einsatzszenarien selbst. Immer häufiger wechseln die Roboter ihren Einsatzort oder Personen sind Teil des Arbeitsraums und tragen zunehmend zu nichtdeterministischen Veränderungen der Umgebung bei. In diesen Fällen steigt die Häufigkeit mit welcher Modelle der quasi-statischen Umgebung aktualisiert oder neu erstellt werden müssen. Als quasi-statisch werden Umgebungsobjekte bezeichnet, die sich nur während der Standzeiten oder durch einen Ortswechsel des Roboters verändern und häufig den größten Teil des Arbeitsraums einnehmen.

Zur Repräsentation von globalen Umgebungsmodellen werden in der Literatur *Octo-Maps* [1–4] oder hüllkörperbasierte Darstellungen [5, 6] verwendet. *Octo-Maps* unterteilen die Umgebung in Würfel und weisen jedem eine Belegungswahrscheinlichkeit zu. Sie erlauben eine präzise Modellierung des belegten Raums und können stochastische Größen abbilden. Ihr Nachteil ist der Aufwand für Distanzberechnungen welche entweder für jeden Würfel oder über komplexe Polyeder erfolgt. Die hüllkörperbasierte Darstellung besitzt hier Vorteile, da der belegte Raum häufig über wenige, größere Hüllkörper modelliert werden kann. Zur Erstellung globaler Umgebungsmodelle bieten externe Sensorsysteme in Form von 3D-Kameras [1, 2] oder *Motion-Capture*-Systeme [5, 6] die umfangreichste Wahrnehmung, welche dazu auch eine laufende Aktualisierung begünstigen. Nachteile sind der hohe technische Aufwand, potentielle Verdeckungen sowie die mangelnde Flexibilität bei Änderungen des Arbeitsraums. Robozentrische Sensorsysteme treten meistens in Form von 3D-Kameras [3] oder *Laser-Scannern* [4] auf und umgehen diese Nachteile. Damit trotz ihrer perspektivisch limitierten Wahrnehmung ein für die Trajektorienplanung ausreichend großer Teil der Umgebung erfasst wird, werden entweder *Octo-Maps* als Gedächtnis eingesetzt oder spezielle Erkundungsbewegungen vom Roboter ausgeführt [7].

Der folgende Beitrag stellt ein dreistufiges Verfahren zur automatischen Rekonstruktion und Modellierung der quasi-statischen Umgebung vor. Es kombiniert die Vorteile der hüllkörperbasierten Darstellung mit einer robozentrischen 3D-Kamera in Auge-in-Hand-Konfiguration unter Verwendung von Aufnahmebewegungen. Die drei Stufen bestehend aus Datenaufnahme, Modellierung und Nachbearbeitung werden in den Abschnitten 2, 3 und 4 näher erläutert. Die Modularisierung ermöglicht effiziente Iterationen bei Änderungen der Umgebung, Hüllkörperformen oder Parameter,

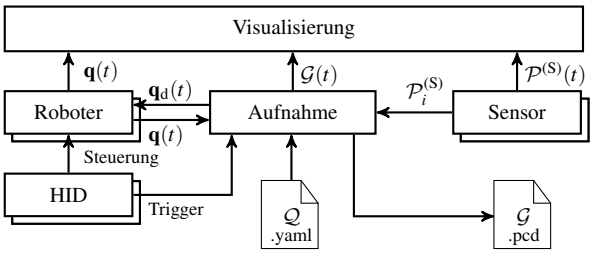


Abbildung 1 Die Aufnahme bildet die erste Stufe.

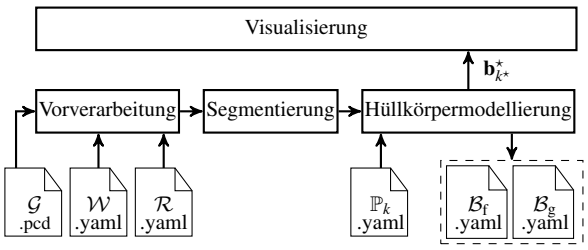


Abbildung 2 Die zweite Stufe umfasst Vorverarbeitung, Clustering und Hüllkörpermodellierung.

da Zwischenergebnisse exportiert und wiederverwertet werden können. Das Verfahren basiert auf der Annahme, dass ausreichend feine Cluster segmentiert werden, die jeweils durch einen Hüllkörper modelliert werden können. In Abschnitt 5 werden drei verschiedene Lösungsmethoden zur Zuordnung von Hüllkörpern zu Clustern vorgestellt. Die Evaluierung in Abschnitt 6 zielt auf die Gültigkeit dieser Annahme ab und analysiert die verschiedenen Stufen des Konzepts. Der Beitrag schließt mit einer Zusammenfassung inklusive Ausblick in Abschnitt 7 ab.

2 Aufnahme von 3D-Punktwolken

Die erste Stufe umfasst die Aufnahme und Registrierung von einzelnen Punktwolken $\mathcal{P}^{(\mathrm{S})} \subset \mathbb{R}^3$ relativ zum Sensorkoordinatensystem mit dem Ergebnis einer Gesamtpunktwolke $\mathcal{G} \subset \mathbb{R}^3$ im Weltkoordinatensystem. Abbildung 1 verdeutlicht die internen Schnittstellen und Komponenten welche im Folgenden erläutert werden.
Der Sensor liefert Punktwolken $\mathcal{P}^{(\mathrm{S})}(t)$ zum Zeitpunkt $t \in \mathbb{R}$ im Sensorkoordinatensystem, welche laufend der Visualisierung zugeführt werden. Die Aufnahme nimmt jeweils zum Zeitpunkt t_i eine von insgesamt $M \in \mathbb{N}$ Einzelpunktwolken $\mathcal{P}_i^{(\mathrm{S})}$ auf, mit $i = 1, 2, \ldots, M$. Der Roboter befindet sich zu diesem Zeitpunkt in der Konfiguration $\mathbf{q}_i \in \mathbb{R}^N$ für einen Roboter mit $N \in \mathbb{N}$ Freiheitsgraden. Die Punktwolke $\mathcal{P}_i^{(\mathrm{S})}$ wird anschließend mit Hilfe der direkten Kinematik in das Weltkoordinatensystem transformiert und als $\mathcal{P}_i$ bezeichnet. Das gemeinsame Referenzsystem ermöglicht die Vereinigung der Punktwolken $\mathcal{P}_i$ zur Gesamtpunktwolke $\mathcal{G}$, welche zum Zeitpunkt t wie folgt definiert ist:

$$\mathcal{G}(t) \coloneqq \bigcup_{j=1}^{i} \mathcal{P}_j \text{ mit } t \in \begin{bmatrix} t_i & t_{i+1} \end{bmatrix} \text{ und } t_{M+1} \coloneqq \infty. \tag{1}$$

Die finale Gesamtpunktwolke wird als `pcd`-Datei für die nächste Stufe abgespeichert. Die Aufnahmekonfigurationen $\mathbf{q}_i$ werden entweder fest vorgegeben oder durch Anwendungen zur Steuerung des Roboters erzeugt.

3 Modellierung der Hindernisse

Die zweite Stufe umfasst die Vorverarbeitung der Gesamtpunktwolke $\mathcal{G}$ aus der vorherigen Stufe mit anschließendem Clustering und der Zuordnung von Hüllkörpern. Abbildung 2 verdeutlicht den Zusammenhang dieser drei Komponenten. Das Resultat sind zwei Mengen $\mathcal{B}_\mathrm{f}$ und $\mathcal{B}_\mathrm{g}$ bestehend aus feinen und groben Hüllkörpern.

3.1 Vorverarbeitung

Die Vorverarbeitung hat das Ziel den Informationsgehalt auf das Wesentliche zu reduzieren, um eine effiziente Weiterverarbeitung zu ermöglichen. Die erste Filterstufe beschränkt $\mathcal{G}$ auf den Arbeitsraum $\mathcal{W} \subset \mathbb{R}^3$ des Roboters:

$$\mathcal{G}_1 = \mathcal{G} \cap \mathcal{W}. \tag{2}$$

Anschließend werden im Vorfeld bekannte Objekte $\mathcal{R} \subset \mathbb{R}^3$, wie beispielsweise die Roboterbasis, aus $\mathcal{G}_1$ entfernt:

$$\mathcal{G}_2 = \mathcal{G}_1 \setminus \mathcal{R}. \tag{3}$$

Ein dreidimensionales Gitter unterteilt das Volumen von $\mathcal{G}_2$ äquidistant in würfelförmige Voxel $\mathcal{V}_j \subset \mathbb{R}^3$ der Kantenlänge $a \in \mathbb{R}^+$ für $j = 1, 2, \ldots$. Ein Voxelfilter reduziert anschließend die Punkte innerhalb jedes Voxels auf deren Schwerpunkt und sorgt somit für eine homogene Dichte:

$$\mathcal{G}_3 = \bigcup_{\mathcal{V}_j \text{ mit } \mathcal{V}_j \cap \mathcal{G}_2 \neq \emptyset} \left(\frac{1}{|\mathcal{V}_j \cap \mathcal{G}_2|} \sum_{\mathbf{p} \in (\mathcal{V}_j \cap \mathcal{G}_2)} \mathbf{p} \right). \tag{4}$$

Die Menge $\mathcal{O} \subset \mathcal{G}_3$ an Ausreißern beinhaltet Punkte $\mathbf{p} \in \mathbb{R}^3$, dessen mittlerer Abstand $\bar{d}(\mathbf{p}, L)$ zu seinen $L \in \mathbb{N}$ nächsten Nachbarn größer als die gewichtete Varianz $\alpha\sigma(\mathbf{p}, L)$ mit $\alpha \in \mathbb{R}^+$ ist:

$$\mathcal{O} \coloneqq \left\{ \mathbf{p} \in \mathcal{G}_3 \mid \bar{d}(\mathbf{p}, L) > \alpha\sigma(\mathbf{p}, L) \right\}, \tag{5}$$

Nach Entnahme der Ausreißer endet die Vorverarbeitung:

$$\mathcal{G}_4 = \mathcal{G}_3 \setminus \mathcal{O}. \tag{6}$$

3.2 Segmentierung

Die Segmentierung unterteilt die Punktwolke $\mathcal{G}_4$ in $C \in \mathbb{N}$ grobe Cluster $\mathcal{C}_i \subseteq \mathcal{G}_4$ mit $i = 1, 2, \ldots, C$:

$$\mathcal{C}_i \coloneqq \{ \mathbf{p} \in \mathcal{G}_4 \mid ||\mathbf{p} - \mathbf{v}||_2 \geq \beta,\ \mathbf{v} \in \mathcal{G} \setminus \mathcal{C}_i \}, \tag{7}$$

sodass alle Punkte des Clusters mindestens den Abstand $\beta \in \mathbb{R}^+$ zu Punkten außerhalb des Clusters besitzen. Nach dieser Segmentierung wird beispielsweise ein Tisch samt darauf liegenden Objekten als zusammenhängendes Cluster erkannt, welches sich noch nicht für die Hüllkörpermodellierung eignet. Vor der zweiten Clusteranalyse werden daher gezielt Verbundstrukturen $\mathcal{S} \subset \mathbb{R}^3$ wie Ebenen aus $\mathcal{C}_i$ extrahiert, sodass eine weitere Unterteilung der Objekte in

$C_i \in \mathbb{N}_0$ Sub-Cluster $\mathcal{C}_{i,j}$ mit $j = 1,2,\ldots,C_i$ erfolgen kann:

$$\mathcal{C}_{i,j} := \left\{ \mathbf{p} \in \mathcal{C}_i \cap \mathcal{S} \mid \max_{\mathcal{S}} |\mathcal{C}_i \cap \mathcal{S}| \right\}. \tag{8}$$

Die Lage der Verbundstrukturen $\mathcal{S}$ wird über den RANSAC-Ansatz in $\mathcal{C}_i$ ermittelt, sodass sie möglichst viele Punkte einschließen. Die ersten $\hat{C}_i \in \mathbb{N}_0$ Sub-Cluster mit $j = 1,2,\ldots,\hat{C}_i \leq C_i$ entsprechen den extrahierten Verbundstrukturen $\mathcal{S}$. Erreicht $\mathcal{C}_{i,j}$ nicht die Mindestgröße $|\mathcal{C}_{i,j}| < T$, wird es verworfen und keine weiteren Sub-Cluster von $\mathcal{C}_i$ erstellt. Ist $\hat{C}_i \geq 1$, wird das übrige Cluster $\mathcal{C}_i \setminus \left(\bigcup_{j=1}^{\hat{C}_i} \mathcal{C}_{i,j}\right)$, wie in (7) angegeben, erneut mit dem euklidischen Verfahren in weitere Sub-Cluster $\mathcal{C}_{i,j}$ mit $j = \hat{C}_i + 1, \hat{C}_i + 2, \ldots, C_i$ unterteilt. Ist $\hat{C}_i = 0$ gilt $\mathcal{C}_{i,1} = \mathcal{C}_i$. Nach der zweiten Clusteranalyse ist für jedes grobe Cluster $\mathcal{C}_i$ eine Menge von feinen Sub-Clustern $\{\mathcal{C}_{i,1}, \mathcal{C}_{i,2}, \ldots, \mathcal{C}_{i,C_i}\}$ vorhanden.

3.3 Hüllkörpermodellierung

Hüllkörper werden über Vektoren $\mathbf{b}^\mathsf{T} = [\mathbf{c}\ \boldsymbol{\theta}\ \mathbf{t}]$ dargestellt, die aus einem Parametersatz $\mathbf{c} \in \mathbb{P}_k \subset \mathbb{R}^{P_k}$, einem Satz xyz-Eulerwinkeln $\boldsymbol{\theta} \in \mathbb{R}^3$ und einer Translation $\mathbf{t} \in \mathbb{R}^3$ bestehen. Insgesamt werden $k = 1,2,\ldots,H \in \mathbb{N}$ verschiedene Hüllkörperformen respektive Parameterträume $\mathbb{P}_k$ für jedes zuvor ermittelte Cluster $\mathcal{C}_i$ beziehungsweise $\mathcal{C}_{i,j}$ betrachtet, um anschließend folgendes Minimierungsproblem jeweils für alle H Formen:

$$\mathbf{b}_k^\star = \underset{\mathbf{c},\boldsymbol{\theta},\mathbf{t}}{\operatorname{argmin}} V(\mathbf{b}), \tag{9}$$

unter den Nebenbedingungen $\mathbf{c} \in \mathbb{P}_k$ sowie $\mathcal{C}_i \subset [\mathbf{b}]_\boxtimes$ beziehungsweise $\mathcal{C}_{i,j} \subset [\mathbf{b}]_\boxtimes$ zu lösen. Der Operator $[\cdot]_\boxtimes$ beschreibt das dreidimensionale Volumen eines Hüllkörpers, während die Funktion $V(\mathbf{b})$ das skalare Volumen berechnet. Zu jedem Cluster existieren nachher H optimierte Hüllkörper $\mathbf{b}_k^\star$, von denen $\mathbf{b}_{k^\star}^\star$ jenes mit dem kleinsten gewichtetem Volumen $\omega_{k^\star} V(\mathbf{b}_{k^\star}^\star)$ repräsentiert und dem Cluster $(\mathcal{C}_i,\ \mathbf{b}_{k^\star}^\star)$ beziehungsweise $(\mathcal{C}_{i,j},\ \mathbf{b}_{k^\star}^\star)$ zugeordnet wird. Die Gewichte $\omega_k \geq 1$ ermöglichen einen Kompromiss zwischen Komplexität der Form und Effizienz der Abdeckung. Die jeweiligen Hüllkörper der groben Cluster $\mathcal{C}_i$ werden in der Menge $\mathcal{B}_\mathrm{g}$ und die Hüllkörper der feinen Cluster $\mathcal{C}_{i,j}$ werden in der Menge $\mathcal{B}_\mathrm{f}$ zusammengefasst.

4 Manuelle Nachbearbeitung

Die Automatisierung der vorherigen Schritte beschleunigt zwar den gesamten Vorgang, kostet jedoch auch Flexibilität in der Breite der Anwendungsfälle, in denen Kontextwissen bezüglich des Anwendungsszenarios erforderlich ist. Die Nachbearbeitung über einen visuellen Editor bietet die Möglichkeit, dieses Kontextwissen einfließen zu lassen.

4.1 Editor

Der Benutzer kann auf Basis der Visualisierung von $\mathcal{B}_\mathrm{g}$ und $\mathcal{B}_\mathrm{f}$ verschiedene Operationen durchführen. Er kann zwischen einem groben Hüllkörper und seinen zugehörigen feinen Hüllkörpern umschalten, sowie deren Position und Orientierung verändern. Darüber hinaus können weitere Hüllkörper hinzugefügt oder entfernt werden. Nach der manuellen Bearbeitung wird die finale Menge der Hüllkörper $\hat{\mathcal{B}}$ ausgegeben, dabei können optional komplexe Hüllkörper durch mehrere einfachere substituiert werden.

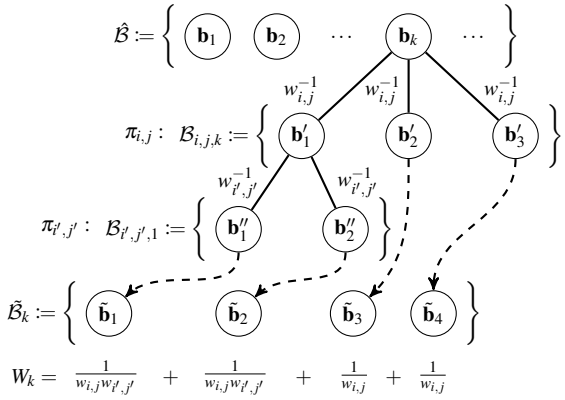


Abbildung 3 Substitution des Hüllkörpers $\mathbf{b}_k$ in die Menge $\tilde{\mathcal{B}}_k$.

4.2 Substitution

Die Substitution eines Hüllkörpers $\mathbf{b}_k \in \hat{\mathcal{B}}$ mit $k = 1,2,\ldots,|\hat{\mathcal{B}}|$ durch eine Menge $\tilde{\mathcal{B}}_k$ erfolgt durch die Anwendung verschiedener Regeln $\pi_{i,j}$. Abbildung 3 verdeutlicht die Vorgehensweise für einen Hüllkörper $\mathbf{b}_k$. Die Regel $\pi_{i,j}$ beschreibt wie ein Hüllkörper $\mathbf{b}_k$ der Form $\mathbb{P}_i$ in eine Menge $\mathcal{B}_{i,j,k}$ von Hüllkörpern der Form $\mathbb{P}_j$ überführt wird, mit $i,j = 1,2,\ldots,H$ und $i \neq j$. Würde die Menge eine maximale Anzahl $|\mathcal{B}_{i,j,k}| > w_{i,j} \geq 1$ an Hüllkörpern überschreiten, wird die Regel nicht angewendet. Hüllkörper, für die keine Regel existiert, werden nicht ersetzt. Nach Anwendung einer Regel werden, sofern möglich, weitere Regeln auf die Hüllkörper aus $\mathcal{B}_{i,j,k}$ angewendet, bis es keine (gültigen) Regeln mehr gibt. Jeder Kante des Graphen aus Abbildung 3 wird nach Anwendung einer Regel $\pi_{i,j}$ ein Gewicht $w_{i,j}^{-1} \in (0\ 1]$ zugeordnet. Die Blätter des Graphen bilden die Menge $\tilde{\mathcal{B}}_k$ von Substituten von $\mathbf{b}_k$. Für jedes Substitut wird das Produkt seiner Kantengewichte berechnet und als Summe $W_k \in \mathbb{R}^+$ über alle Substitute dem Hüllkörper $\mathbf{b}_k$ zugeordnet. Bei Hüllkörpern ohne Substitution gilt $\tilde{\mathcal{B}}_k = \{\mathbf{b}_k\}$ und $W_k = \infty$. Die finale Auswahl von Hüllkörpern erfolgt über einen *Greedy*-Algorithmus. Die Tupel $(\mathbf{b}_k, \tilde{\mathcal{B}}_k, W_k)$ werden aufsteigend nach W_k sortiert und alle $\mathbf{b}_k$ geordnet der Menge $\hat{\mathcal{B}}'$ hinzugefügt. Dadurch werden Substitutionen mit den größten Vorteilen zuerst berücksichtigt. Es werden anschließend so lange die Hüllkörper $\mathbf{b}_k$ mit $\tilde{\mathcal{B}}_k$ ersetzt, bis alle $\mathbf{b}_k$ ausgetauscht sind oder die Obergrenze $|\hat{\mathcal{B}}| \leq F$ erreicht wird.
Die Substitutionsregeln $\pi_{i,j}$ ergeben sich anhand der konkreten Formen. Der Schwellwert $w_{i,j}$ beschreibt die Anzahl von Substituten bei welcher die Komplexitätsreduzierung ihre Wirkung verliert. Er wird anhand des Rechenaufwands der zugehörigen Distanzfunktionen ermittelt.

5 Lösungsverfahren

Die konkreten Lösungsverfahren hängen von den Hüllkörperformen ab, welche nachfolgend vorgestellt werden. Zur

Lösung von (9) werden anschließend in den darauffolgenden Abschnitten drei verschiedene Methoden erläutert.

5.1 Hüllkörperformen

Die folgenden $H = 4$ Formen werden betrachtet:

- Kugel: $\mathbb{P}_1 \subset \mathbb{R}$ mit Radius $r \in \mathbb{R}^+$
- Rundzylinder: $\mathbb{P}_2 \subset \mathbb{R}^2$ mit Radius $r \in \mathbb{R}^+$ und x-Länge $l_\mathrm{x} \in \mathbb{R}^+$
- Rundquader: $\mathbb{P}_3 \subset \mathbb{R}^3$ mit Radius $r \in \mathbb{R}^+$, x-Länge $l_\mathrm{x} \in \mathbb{R}^+$ und y-Länge $l_\mathrm{y} \in \mathbb{R}^+$
- Ausgedehntes Rechteck[1]: $\mathbb{P}_4 \subset \mathbb{R}^2$ mit x-Länge $l_\mathrm{x} \in \mathbb{R}^+$ und y-Länge $l_\mathrm{y} \in \mathbb{R}^+$

Das Konzept unterstützt weitere Formen und erfordert dafür nur die Integration in eines der folgenden Verfahren.

5.2 Evolutionäre Optimierung (EVO)

Eine Population besteht aus $\Theta \in \mathbb{N}$ Hüllkörpern $\mathbf{b}$ derselben Form und wird mit zufälligen Parametern $\mathbf{c}$, $\mathbf{t}$ und θ initialisiert. Die Position $\mathbf{t}$ wird jeweils limitiert über ein virtuelles Quader der Punktwolke $\mathcal{C}_i$ beziehungsweise $\mathcal{C}_{i,j}$, das am Weltkoordinatensystem ausgerichtet ist. Für Hüllkörper mit einem Radius, ist dieser zwischen 0 und der Länge $d = ||\mathbf{p}_\mathrm{max} - \mathbf{p}_\mathrm{min}||_2$ der dreidimensionalen Diagonalen limitiert. Es sind $\mathbf{p}_\mathrm{min} \in \mathbb{R}^3$ und $\mathbf{p}_\mathrm{max} \in \mathbb{R}^3$ die äußersten Ecken des virtuellen Quaders. Die Rekombination $\mathbf{b}_\mathrm{neu} = \mathbf{Z}\mathbf{b}_1 + (\mathbf{I} - \mathbf{Z})\mathbf{b}_2$ wird über die anteilige Addition zweier Individuen mit der diagonalen Zufallsmatrix $\mathbf{Z} \in [0\,1]^{6+P}$ und der Einheitsmatrix $\mathbf{I}$ realisiert. Die Mutation $\mathbf{b}_\mathrm{neu} = \mathbf{b}_{alt} + m\tilde{\mathbf{b}}$ entsteht durch die Addition eines zufälligen Hüllkörpers $\tilde{\mathbf{b}}$ zum bisherigen Individuum. Mit dem Mutationsfaktor $m \in \mathbb{R}^+$ wird das Ausmaß der Mutation für folgende Generationen reduziert. Mutation und Rekombination werden solange durchgeführt, bis das Individuum die Parametergrenzen und Nebenbedingungen einhält beziehungsweise angegebene Grenzen für den Algorithmus erreicht werden. Als Kostenfunktion dient (9).

5.3 Hauptkomponentenanalyse (PCA)

Für die Formen Rundzylinder und Rundquader werden zwei normierte Hauptkomponenten $\mathbf{e}_\mathrm{x}, \mathbf{e}_\mathrm{y} \in \mathbb{R}^3$ berechnet und eine dritte Komponente $\mathbf{e}_z = \mathbf{e}_\mathrm{x} \times \mathbf{e}_\mathrm{y}$ erzeugt. Die Rotation θ wird über die Hauptkomponenten bestimmt und transformiert das Cluster in das neue Koordinatensystem. Entlang des neuen Koordinatensystems werden wieder die Eckpunkte $\mathbf{p}_\mathrm{max}$ und $\mathbf{p}_\mathrm{min}$ eines virtuellen Quaders definiert. Die Position $\mathbf{t}$ ergibt sich über die halbe Diagonale des virtuellen Quaders:

$$\mathbf{t} = \frac{1}{2}\left(\mathbf{p}_\mathrm{max} - \mathbf{p}_\mathrm{min}\right). \tag{10}$$

Der Radius r wird über den größten Abstand zwischen den Punkten des Clusters und der beschränkten Längsachse des virtuellen Quaders bestimmt:

$$r = \max_{\mathbf{p}_k \in \mathcal{C}_i} d(\mathbf{p}_k, \mathbf{t} + p_\mathrm{x,min}\mathbf{e}_\mathrm{x}, \mathbf{t} + p_\mathrm{x,max}\mathbf{e}_\mathrm{x}), \tag{11}$$

[1] Nur die z-Rotation wird betrachtet.

mit $d(\cdot,\cdot,\cdot)$ als Distanzfunktion. Der Rundquader verwendet hierfür die halbe Höhe des virtuellen Quaders:

$$r = \frac{p_\mathrm{z,max} - p_\mathrm{z,min}}{2}. \tag{12}$$

Die Längen l_x und Breite (nur Rundquader) l_y werden der Länge und Breite des virtuellen Quaders entnommen:

$$l_\mathrm{x} = p_\mathrm{x,max} - p_\mathrm{x,min} - 2\bar{r} \tag{13}$$

$$l_\mathrm{y} = p_\mathrm{y,max} - p_\mathrm{y,min} - 2\bar{r}. \tag{14}$$

Hiervon wird über $\bar{r} \in [0\ r]$ jeweils maximal das zweifache von r abgezogen, sodass alle Punkte des Clusters noch im Hüllkörper liegen.
Für das ausgedehnte Rechteck, werden die Punkte des Clusters zunächst auf die x-y-Ebene projiziert, ehe die PCA zwei zueinander orthogonale Komponenten $\mathbf{e}_\mathrm{x}$ und $\mathbf{e}_\mathrm{y}$ ermittelt. Position $\mathbf{t}$ und Orientierung θ sowie Länge l_x und Breite l_y werden wie zuvor bestimmt. Die Höhe $t_\mathrm{z} = p_\mathrm{z,max}$ entspricht dem höchsten Punkt des Clusters.
Für die Kugel kann direkt ein virtueller Quader und damit $\mathbf{t}$, wie zuvor beschrieben, bestimmt werden, ohne eine PCA durchzuführen. Der Radius r ist bei der Kugel der größte Abstand eines Punktes aus dem Cluster zum Mittelpunkt $\mathbf{t}$:

$$r = \max_{\mathbf{p}_k \in \mathcal{C}_i} d(\mathbf{p}_k, \mathbf{t}), \tag{15}$$

mit $d(\cdot,\cdot)$ als Distanzfunktion.

5.4 Algorithmische Geometrie (ALG)

Dieses Verfahren unterscheidet sich zu der PCA hauptsächlich in der Berechnung des virtuellen Quaders. Während die PCA über die Kovarianzmatrix die Orientierung der Punkte bestimmt, wird beim alternativen Verfahren explizit nach einem Quader mit minimalem Volumen, das alle Punkte einschließt, gesucht.
Beim Rundzylinder und Rundquader nutzt der Algorithmus die Eigenschaft, dass der virtuelle Quader zwei rechtwinklig angrenzende Flächen besitzt, die bündig mit Kanten der konvexen Hülle der Punktwolke abschließen [8]. Eine iterative Suche über alle Paare von angrenzenden Flächen des virtuellen Quaders ermittelt dann jenen mit dem geringsten Volumen. Der Radius r, sowie die Länge l_x und Breite l_y werden wie zuvor berechnet. Beim ausgedehnten Rechteck, wird die zweidimensionale Variante auf den projizierten Punkten des Clusters angewendet. Die Länge l_x und Breite l_y sowie Höhe t_z werden wie zuvor berechnet.
Für die Kugel wird ein anderer Algorithmus verwendet, der aus jeweils 4 Punkten eine potentielle Kugel bestimmt und für alle übrigen Punkte prüft, ob diese innerhalb des Volumens liegen. Die Menge an 4 bildenden Punkten wird schrittweise aktualisiert, sodass immer mehr Punkte von der Kugel umschlossen werden und das Volumen minimal ist.

6 Evaluation

Die Evaluierung umfasst die Aufnahme, die Vorverarbeitung, die Hüllkörpermodellierung und die Substitution. Untersucht werden insbesondere die Vor- und Nachteile der verschiedenen Varianten zur Zuordnung und Parametrisierung der Formen hinsichtlich des Rechenaufwandes und der

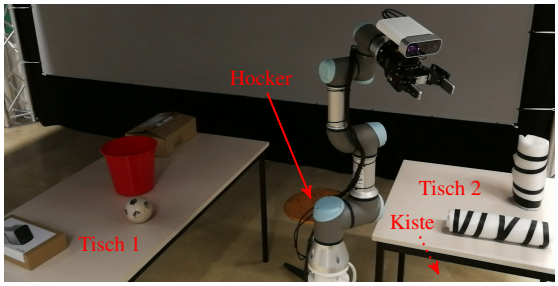


Abbildung 4 Roboter mit 3D-Kamera in der Testumgebung.

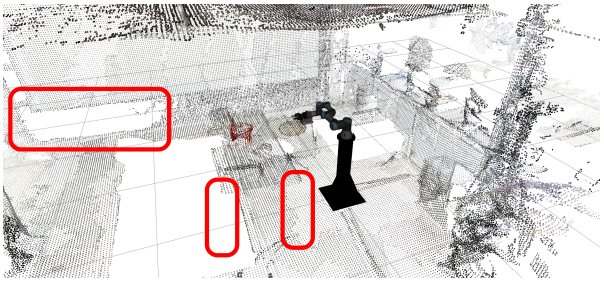


Abbildung 5 Gesamtpunktwolke $\mathcal{G}$ nach der Aufnahme.

räumlichen Effizienz des Umgebungsmodells. Der Ablauf von der Aufnahme bis zur Hüllkörpermodellierung ist auch im verlinkten Video[2] zu sehen.

6.1 Experimentalsystem

Als Roboter kommt ein UR3 (*Universal Robot* 3) zum Einsatz ($N = 6$), an dessen Endeffektor eine 3D-Kamera des Typs *Azure Kinect* befestigt ist. Der Arbeitsraum befindet sich in einer Laborumgebung mit zwei Tischen in direkter Nähe, siehe dazu Abbildung 4. Auf dem ersten, größeren Tisch befinden sich zwei quaderförmige Kartons sowie ein Ball und ein roter Eimer. Auf dem zweiten Tisch befinden sich zwei zylindrische Objekte, eins in liegender und eins in stehender Orientierung. Zwischen den Tischen steht ein Hocker. Zusätzlich ragt eine auf dem Boden stehende Kiste in den Arbeitsraum (nicht in der Abbildung).
Für die Auswertung kommt ein Rechner mit einem *Intel* i7-8700K @4,8 GHz und 32 GB RAM @3,2 GHz zum Einsatz. Die Implementierungen erfolgen in C++ unter *Ubuntu* 18.04 und basieren auf der *Point Cloud Library* [9]. Als evolutionärer Optimierer wird *Open GA* verwendet [10]. Es werden maximal 10.000 Generationen betrachtet und eine Elite von 200 Individuen bei der Rekombination geschützt. Die ALG-Methode verwendet die *Geometric Tools Engine* [11]. Die Visualisierung findet in *RViZ* aus dem *Robot Operating System* (Melodic) statt. Die verwendeten Parameter, welche nicht im weiteren Verlauf angegeben werden, sind Tabelle 1 zu entnehmen.

6.2 Aufnahme

Zu Beginn der Aufnahme befindet sich der Roboter in der Startkonfiguration $(-170\ -90\ 0\ -28\ -90\ 0)°$. Von dort wird das erste Gelenk in 36° Schritten bis 143° bewegt und das vierte Gelenk in 23° Schritten bis −172°. Daraus ergeben sich $M = 72$ Aufnahmekonfigurationen. Die resultierende Gesamtpunktwolke $\mathcal{G}$ ist in Abbildung 5 abgebildet und besteht aus 58.060.800 Punkten.
Die Markierungen verdeutlichen exemplarisch fehlende Strukturen, wie die Beine des Hockers und der Tische sowie einen Teil der Wand auf der linken Seite. Die 3D-Kamera verwendet eine infrarotbasierte Methode zur Tiefenschätzung, die auf die Reflektion eines Infrarotmusters angewiesen ist. Der Grad der Reflektion ist bei dunklen und matten Oberflächen geringer und wird durch einen spitzen Betrachtungswinkel weiter reduziert. Sowohl die Beine, als auch der untere Teil der Wand sind in mattem schwarz und besitzen einen solchen spitzen Betrachtungswinkel.

6.3 Vorverarbeitung

Der Arbeitsraum wurde künstlich auf $x \in [-0{,}5\ 1{,}5]\,\mathrm{m}$, $y \in [-1\ 1{,}5]\,\mathrm{m}$ und $z \in [0{,}1\ 2]\,\mathrm{m}$ relativ zum Ursprung im Zentrum des Sockels erweitert, da der Arbeitsraum des UR3 aufgrund seiner geringen Größe wenig Aussagekraft hätte. Das Entfernen von Punkten außerhalb des Arbeitsraums benötigt 920 ms und reduziert die Punktwolke von 58.060.800 auf 39.054.324 Punkte. Da damit auch der Boden entfernt wird ist die Roboterbasis bereits nicht mehr in $\mathcal{G}_1$ enthalten, sodass (3) keine weiteren Punkte entfernt aber dennoch 620 ms benötigt. Der Voxelfilter ist mit 1.307 ms der aufwändigste Schritt und reduziert die Punktwolke weiter auf 40.624 Punkte. Die Entfernung von Ausreißern benötigt 160 ms und beendet die Vorverarbeitung mit Punktwolke $\mathcal{G}_4$ bestehend aus 40.472 Punkten. Abbildung 6 stellt $\mathcal{G}_4$ mit $\mathcal{G}$

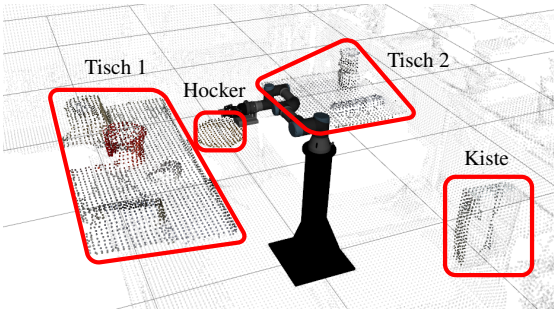


Abbildung 6 Punktwolke $\mathcal{G}_4$ nach der Vorverarbeitung.

[2] https://youtu.be/kfRwKf7axiM

Tabelle 1 Parameter

Parameter	Wert	Bedeutung
a	0,01 m	Voxelgröße
L	8	Nächste-Nachbarn
α	1,2	Gewicht der Varianz
β	0,05 m	Cluster-Schwellwert
T	5.000	Minimale Strukturgröße
ω_k	1 ($\forall k$)	Volumengewichte
Θ	2.000	Populationsgröße
m	0,2	Mutationsfaktor

(halbtransparent) gegenüber und verdeutlicht die Objekte des Arbeitsraums.

6.4 Hüllkörpermodellierung

Die erste Segmentierung hat $C = 4$ grobe Cluster $\mathcal{C}_1$ bis $\mathcal{C}_4$ ermittelt, welche die in $\mathcal{G}_4$ freigestellten Objekte umfassen. Die zweite Segmentierung hat für die ersten beiden groben Cluster erfolgreich eine feinere Unterteilung in $\mathcal{C}_{1,1}$ bis $\mathcal{C}_{1,5}$ ($C_1 = 5$) beziehungsweise $\mathcal{C}_{2,1}$ bis $\mathcal{C}_{2,3}$ ($C_2 = 3$) gefunden. Abbildung 7 und 8 zeigen die jeweiligen Cluster in Form ihrer Hüllkörper gemäß der ALG-Methode.
Tabelle 2 umfasst die besten Formen jeder Lösungsvariante und für jedes Cluster. Der EVO-Ansatz bestimmt viermal das kleinste Volumen (fett), allerdings werden auch unrealistisch hohe Volumina wie beispielsweise bei $\mathcal{C}_{1,1}$ ermittelt. Dies kommt durch ungünstige Mutationen und einem sinkenden Einfluss der Mutation ($m < 1$) bei steigenden Generationen, womit der Algorithmus in einer suboptimalen Lösung stagniert. Die PCA stellt zweimal das kleinste Volumen mit ansonsten teils höheren, aber plausiblen Volumina. Das Verfahren berechnet die Orientierung über die Kovarianzmatrix auf Basis der Verteilung der Punktwolke. Somit werden teils Orientierungen ermittelt, die nicht ideal im Sinne des Volumens zur Form des Clusters passen. Die ALG-Methode generiert siebenmal das kleinste Volumen und auch sonst die stabilsten Ergebnisse. Es berechnet die Ausrichtung des Hüllkörpers basierend auf minimalen Quadern beziehungsweise Rechtecken und kann daher die Hüllkörper besser an die Cluster anpassen. Der Schritt vom Hilfsquader zum Rundquader oder Rundzylinder verändert nachträglich die Form, sodass dadurch das Volumen suboptimal sein kann. Hierunter leiden sowohl PCA als auch die ALG-Methode, nicht aber der EVO-Ansatz.
Tabelle 3 gibt die jeweils mittlere und maximale Laufzeit der sequentiellen Lösung von (9) pro Form über alle Cluster an. Der EVO-Ansatz benötigt für den Rundquader und das unendliche Rechteck länger als für die übrigen Formen. Grundsätzlich ist der EVO-Ansatz im Bereich von Sekunden im Vergleich am langsamsten. Die PCA besitzt Laufzeiten im Bereich von Mikrosekunden und ist damit am schnellsten. Da die Kugel keine Orientierung besitzt, terminiert der Algorithmus in diesem Fall schneller. Bei den übrigen Formen ist der aufwändigste Teil des Algorithmus derselbe,

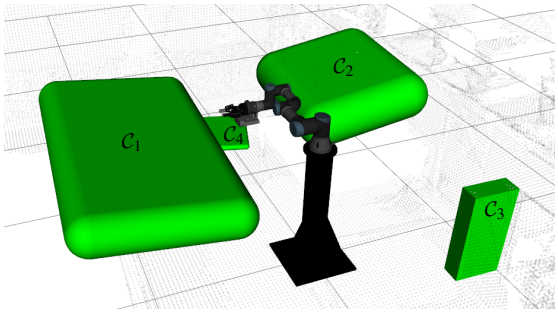


Abbildung 7 Modellierung nach der ALG-Methode auf Basis der groben Cluster.

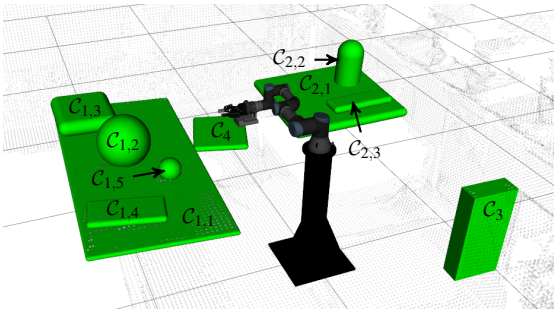


Abbildung 8 Modellierung nach der ALG-Methode auf Basis der feinen Cluster.

womit auch die Laufzeiten ähnlich ausfallen. Die ALG-Methode besitzt Laufzeiten im Bereich von Millisekunden und liegt damit zwischen den beiden vorherigen Verfahren. Insbesondere der Rundzylinder und der Rundquader stechen gemeinsam heraus, da beide denselben Algorithmus zur Bestimmung des Hilfsquaders verwenden.
Eine alternative parallele Implementierung zur Berechnung aller $H = 4$ Formen pro Cluster gleichzeitig, benötigt mit Ausnahme der PCA in etwa die Laufzeit der langsamsten Form. Bei der PCA dauert die Zuweisung der Berechnung auf die vorinitialisierten *threads* länger als die Summe der sequentiellen Variante.

6.5 Substitution

Die Substitution wird anhand von Cluster $\mathcal{C}_{1,3}$ untersucht. Hierfür befindet es sich allein im Arbeitsraum an Position $\left(0\ 0{,}5\ 0{,}8\right)$ m, sodass die Bewegungsplanung [5] ausweichen muss. Der simulierte Roboter startet in der Nullkonfiguration und erhält den Auftrag in die Zielkonfiguration $\left(180\ 0\ 0\ 0\ 0\ 0\right)^\circ$ zu fahren. Die mittlere Laufzeit während der siebensekündigen Bewegung beträgt 47,8 ms. Die erste Substitutionsregel $\pi_{3,2}$ zerlegt einen Rundquader in $\left\lceil \frac{l_y}{2(1-\gamma)r} \right\rceil$ äquidistante Rundzylinder mit Überlappungsfaktor $\gamma = 0{,}3$. Die zweite Substitutionsregel $\pi_{2,1}$ zerlegt einen Rundzylinder in $\left\lceil \frac{l_x}{2(1-\gamma)r} \right\rceil$ äquidistante Kugeln. Die Obergrenzen sind $w_{3,2} = w_{2,1} = 5$. Nach der ersten Regel substituieren

Tabelle 2 Paare aus jeweils bester Form und Volumen (dm^3). Formen sind als Parametersätze angegeben. Kleinstes Volumen je Cluster ist fettgedruckt. Clustergröße in Klammern.

	EVO	PCA	ALG
$\mathcal{C}_1$ (22.356)	$\mathbb{P}_2$, 783,8	$\mathbb{P}_3$, 365,9	$\mathbb{P}_3$, **302,2**
$\mathcal{C}_{1,1}$ (15.647)	$\mathbb{P}_2$, 792,3	$\mathbb{P}_3$, 31,0	$\mathbb{P}_3$, **26,3**
$\mathcal{C}_{1,2}$ (2.841)	$\mathbb{P}_3$, **11,7**	$\mathbb{P}_2$, 14,5	$\mathbb{P}_1$, 13,7
$\mathcal{C}_{1,3}$ (2.062)	$\mathbb{P}_2$, 24,1	$\mathbb{P}_3$, **11,6**	$\mathbb{P}_3$, 12,1
$\mathcal{C}_{1,4}$ (1.347)	$\mathbb{P}_2$, 8,4	$\mathbb{P}_3$, 3,7	$\mathbb{P}_3$, **2,6**
$\mathcal{C}_{1,5}$ (395)	$\mathbb{P}_2$, **0,9**	$\mathbb{P}_3$, 1,3	$\mathbb{P}_1$, 1,1
$\mathcal{C}_2$ (12.361)	$\mathbb{P}_3$, 414,5	$\mathbb{P}_3$, 312,6	$\mathbb{P}_3$, **218,3**
$\mathcal{C}_{2,1}$ (9.424)	$\mathbb{P}_4$, 404,5	$\mathbb{P}_3$, 56,5	$\mathbb{P}_3$, **33,0**
$\mathcal{C}_{2,2}$ (1.810)	$\mathbb{P}_2$, **7,2**	$\mathbb{P}_2$, 8,5	$\mathbb{P}_2$, 7,7
$\mathcal{C}_{2,3}$ (1.118)	$\mathbb{P}_2$, 4,2	$\mathbb{P}_3$, **2,7**	$\mathbb{P}_3$, **2,7**
$\mathcal{C}_3$ (3.208)	$\mathbb{P}_4$, **19,7**	$\mathbb{P}_4$, 20,9	$\mathbb{P}_4$, 20,8
$\mathcal{C}_4$ (2.547)	$\mathbb{P}_3$, 23,4	$\mathbb{P}_3$, 5,6	$\mathbb{P}_3$, **5,2**

Tabelle 3 Mittlere und maximale Dauer der sequentiellen Berechnung einzelner Formen. Mittelwert und Maximum pro Form über alle Cluster.

	EVO (s)		PCA (µs)		ALG (ms)	
	∅	Max	∅	Max	∅	Max
$\mathbb{P}_1$	0,19	0,20	0,94	1,50	0,02	0,02
$\mathbb{P}_2$	0,68	3,30	9,44	15,60	14,68	65,88
$\mathbb{P}_3$	4,64	17,80	6,49	11,20	14,68	65,75
$\mathbb{P}_4$	12,02	48,20	8,43	14,10	0,02	0,05

4 Rundzylinder den Rundquader. Die mittlere Laufzeit der Bewegungsplanung beträgt jetzt 44,9 ms und konnte damit reduziert werden. Durch die zweite Regel werden die Rundzylinder durch je 5 Kugeln substituiert. Die mittlere Laufzeit der Bewegungsplanung beträgt jetzt 66,6 ms. Obwohl die Kugel die einfachste Distanzfunktion besitzt, sorgt ihre Anzahl (20) schließlich für eine Erhöhung der Laufzeit.

7 Zusammenfassung

Der Beitrag stellt ein dreistufiges Verfahren zur Aufnahme, Modellierung und manuellen Nachbearbeitung vor, mit dem Ziel, den quasi-statischen Teil eines Arbeitsraums effizient über Hüllkörper zu modellieren. Die Ergebnisse zeigen, dass das verwendete Sensorsystem Schwächen bei mattschwarzen Oberflächen und spitzen Betrachtungswinkeln besitzt, weshalb Objekte fehlen können. Die Vorverarbeitung reduziert die Gesamtpunktwolke auf das Wesentliche und erhält gleichzeitig genügend Informationen, sodass eine effiziente Weiterverarbeitung möglich ist. Die zweischrittige Segmentierung kann einfache Verbundstrukturen auflösen und damit feinere Cluster ermitteln, welche der Granularität der tatsächlichen Umgebung entsprechen. Aus den drei Verfahren zur Zuweisung der Hüllkörper liefert die ALG-Methode die robustesten Ergebnisse und agiert am zweitschnellsten. Der EVO-Ansatz ist am langsamsten und besitzt in der gewählten Einstellung wenig Robustheit. Die PCA ist am schnellsten, besitzt allerdings leichte Schwächen bei der Orientierung der Hüllkörper. Die Substitution erzielt kleine Verbesserungen der Laufzeit, stößt durch das exponentielle Wachstum von Hindernissen jedoch schnell an Grenzen.
Verbesserungen beim Sensorsystem können in Zukunft durch Hinzunahme weiterer Systeme wie beispielsweise Laserscanner untersucht werden. Die Granularität der zweischrittigen Segmentierung kann durch weitere Schritte, welche die Farbinformationen nutzen oder die Form und Struktur analysieren, profitieren, um auch in komplexeren Umgebungen die Annahme, jedes Cluster kann durch eine Form beschrieben werden, zu erfüllen. Die Initialisierung des EVO-Ansatzes über die Lösung der PCA reduziert möglicherweise die Laufzeit und erhöht die Robustheit, sodass beide Ansätze gegebenenfalls ihre Schwächen gegenseitig kompensieren können.

Danksagung

Die Autoren bedanken sich bei Jens Püttschneider für die Unterstützung bei der Implementierung.

Literatur

[1] H. Dermot Doran and S. Marti. “Robot Workspace Monitoring with Redundant Structured Light Cameras - A Preliminary Investigation”. In: *International Conference on Informatics in Control, Automation and Robotics*. 2014, pp. 611–617.

[2] A. Fetzner, C. Frese, and C. Frey. “A 3D representation of obstacles in the robots reachable area considering occlusions”. In: *International Symposium on Robotics*. 2014, pp. 1–8.

[3] M. Pomarlan and I. A. Sucan. “Motion planning for manipulators in dynamically changing environments using real-time mapping of free workspace”. In: *International Symposium on Computational Intelligence and Informatics*. Ed. by A. Szakál. 2013, pp. 483–487.

[4] Y. Yu and K. Gupta. “Sensor-based roadmaps for motion planning for articulated robots in unknown environments: some experiments with an eye-in-hand system”. In: *Human and environment friendly robots with high intelligence and emotional quotients*. 1999, pp. 1707–1714.

[5] M. Krämer et al. “Model predictive control of a collaborative manipulator considering dynamic obstacles”. In: *Optimal Control Applications and Methods* (2020).

[6] J. A. Corrales, F. A. Candelas, and F. Torres. “Safe human–robot interaction based on dynamic sphere-swept line bounding volumes”. In: *Robotics and Computer-Integrated Manufacturing* 27.1 (2011), pp. 177–185.

[7] D. Riedelbauch and D. Henrich. “Exploiting a Human-Aware World Model for Dynamic Task Allocation in Flexible Human-Robot Teams”. In: *International Conference on Robotics and Automation*. 2019, pp. 6511–6517.

[8] J. O’Rourke. “Finding minimal enclosing boxes”. In: *International Journal of Computer & Information Sciences* 14.3 (1985), pp. 183–199.

[9] R. B. Rusu and S. Cousins. “3D is here: Point Cloud Library (PCL)”. In: *International Conference on Robotics and Automation*. 2011, pp. 1–4.

[10] A. Mohammadi et al. “OpenGA, a C++ Genetic Algorithm Library”. In: *International Conference on Systems, Man, and Cybernetics*. 2017, pp. 2051–2056.

[11] D. Eberly. *Geometric Tools Engine, Version 5.4*. `github.com/davideberly/GeometricTools`. Besucht am 18.11.2020.

Bots2ReC: Technische Herausforderungen bei der automatisierten Asbestentfernung in Wohnräumen mittels mobiler Manipulatoren

Bots2ReC: Technical Challenges for Automated Asbestos Removal from Residential Buildings with Mobile Manipulators

Tobias Haschke*, M. Sc., haschke@igmr.rwth-aachen.de
Prof. Dr.-Ing. Mathias Hüsing*, huesing@igmr.rwth-aachen.de
Univ.-Prof. Dr.-Ing. Dr. h. c. Burkhard Corves*, corves@igmr.rwth-aachen.de

*RWTH Aachen University, Institut für Getriebetechnik, Maschinendynamik und Robotik, 52064 Aachen, Deutschland

Kurzfassung

Die Baubranche zählt zu einer der größten und gleichermaßen vielseitigsten Branchen überhaupt. Dabei zeichnet sie sich aus der Perspektive der Automatisierungstechnik insbesondere durch die stark variierenden Umgebungszustände und die hohe Variabilität an Prozessschritten und Teilaufgaben aus. [1] Als Ergebnis sind in dieser Branche zahlreiche Robotersysteme, sog. Single Task Construction Robots (STCRs) entstanden, die einzelne dedizierte und häufig repetitive Prozessschritte übernehmen [2]. Ein Beispiel für ein solches System ist das im Forschungsprojekt Bots2ReC entwickelte Tandem aus mobilem Manipulator und mobiler Unterstützungseinheit (siehe Abbildung 1) zur Entfernung von Asbest aus Wohnräumen für das Abschleifen von asbesthaltigem Putz. [3]
Das System wurde in einer ersten Designstudie als einzelner mobiler Manipulator mit differentiellem Radantrieb ausgelegt und der 6-DOF Manipulator um eine siebte Linearachse zur Vergrößerung des Arbeitsraumes erweitert. Zur Realisierung der hohen Leistungsaufnahme durch das Entfernungswerkzeug und die mitgeführte Absaugung für den Abraum, wurde eine mitgeführte aktuierte Kabeltrommel entwickelt, die den autonomen Betrieb für einen kabelgebundenen mobilen Roboter ermöglichte. Im Rahmen von Tests konnte das System und seine Teilkomponenten validiert und Schlussfolgerungen für eine Weiterentwicklung gezogen werden. Ein entscheidendes Ergebnis aus der ersten Designstudie war die deutlich höhere notwendige Absaugleistung für den automatisierten Entfernungsprozess, sodass der Systemwechsel auf eine zentrale Absaugung notwendig wurde. Unter Beibehaltung der Agilität des Systems entstand das vorliegende Konzept des Tandems, um die Mitführung des Absaugschlauches sowie des Stromkabels auf die Begleiteinheit auszulagern.



Abbildung 1: *Mobiler Manipulator zur automatisierten Asbestentfernung*

Dieses System stellt durch seine Tandembauweise eine innovative, neuartige Systemtopologie von kooperierenden mobilen Robotern dar, die neue Herausforderungen in der Navigation, Steuerung, Regelung sowie der Aufgabenplanung eröffnen. Aus diesem Grund stellt dieser Beitrag den zweistufigen Entwicklungsprozess dieses Systems im Detail vor und arbeitet dabei die technischen Entscheidungsgrundlagen sowie die zugehörigen Rahmenbedingungen auf. Damit wird die zukünftige Entscheidungsfindung bei der Entwicklung von mobilen Baustellenrobotern systematisch unterstützt und die potentiellen Einflüsse auf andere Entwicklungsteile offengelegt.

Einleitung

Die Arbeit am Forschungsprojekt Bots2ReC ist durch die Kontamination von mehreren Millionen Wohnungen und Gebäuden mit Asbest in ganz Europa motiviert. Für Frankreich werden durch die L'Union sociale pour l'habitat insgesamt 15 Millionen Wohnungen gemeldet, die in der Hochphase der Asbestnutzung von 1960 bis 1990 gebaut wurden. [4] Ähnliche Studien liegen für viele weitere EU-Staaten vor. Die Erkenntnis über die gesundheitsschädlichen Eigenschaften von Asbestfasern für den Menschen, erfordert eine Sanierung der kontaminierten Gebäude unter strengen Sicherheitsauflagen. Dabei sind insbesondere lungengängige Fasern hochgradig gefährlich, da sie über längere Zeiträume die Krankheit der Asbestose hervorrufen können. [5] Unter diesen Umständen sind insbesondere das Entfernen von asbesthaltigem Putz und Fliesenkleber für den Menschen besonders gefährlich, da diese idR. mit abrasiven Prozessen vorgenommen werden und der dabei entstehende Staub eine große Gefahr darstellt. Aus diesem Grund hat sich das Bots2ReC Forschungsprojekt zum Ziel gesetzt, einen mobilen Manipulator zu entwickeln, der in Wohnungen das Abschleifen von kontaminierten Wänden und Decken ohne direkte Einwirkung des Menschen ermöglicht.

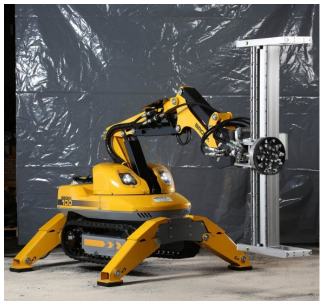
Dieser Beitrag ist im Weiteren wie folgt aufgebaut: Im *Stand der Technik* werden kurz die existierenden maschinenbasierten Schleif- und Asbestentfernungsansätze vorgestellt. Im Anschluss werden die *Entwicklungsstufen des Bots2ReC Baustellenroboters* vorgestellt und im Detail die *Herausforderungen eines gekoppelten Tandems mobiler Roboter im Bauwesen* aufgezeigt. Die *Fallstudie* für das vorliegende robotische System zeigt abschließend den Anwendungsbezug auf und mündet in der *Zusammenfassung*.

Stand der Technik

Der Großteil der maschinellen Ansätze zur Asbestentfernung basieren auf einem Schleifprozess mit einer speziellen Schleifscheibenkontur, welche für die jeweilige Oberfläche optimiert ist. Die industriellen Komponenten von Brokk und Schwamborn stellen dafür gute Beispiele dar. Brokk kombiniert das Schleifwerkzeug mit einer Linearachse, um den te-leoperierten Einsatz zu vereinfachen und ermöglicht bei einer Schleiflänge von 1530 mm eine maximale Schleiftiefe von 5 mm. [6] Der WDS 530 von Schwamborn verzichtet auf die Führung durch eine Linearachse und ermöglicht durch den Einsatz eines Kardangelenks eine gleichbleibende Prozesskraft bei Oberflächen mit einer Neigung von bis zu 6 %. [7] Beide Schleifwerkzeuge sind mit einem Gewicht jenseits der 200 kg zu schwer für normale Wohnungen, wo die Geschosstragfähigkeit im Mittel auf 300 kg pro Quadratmeter limitiert ist. Aus diesem Grund haben sich für Wohnhäuser kleinere und leichtere Verfahren entwickelt. Ebenfalls abrasiv arbeiten die Prototypen von Eco-Amiante und Takenaka. Der Eco-Amiante schleift bei einem Gesamtgewicht von 190 kg mittels einer Schleifscheibe von 250 mm Durchmesser bis zu 50 mm tief pro Arbeitsgang und erlaubt sowohl eine teleoperierte Fernsteuerung als auch einen Lidar-basierten Automatikmodus. Dabei zielt der Prototyp insbesondere auf das Entfernen von asbest-



AS PROTEK
(Quelle: Gilles Rambaud | Le Moniteur)



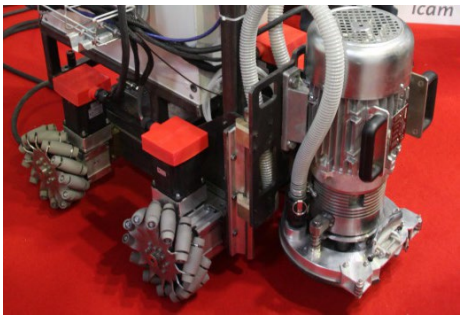
BROKK 100 mit PMP 250
(Quelle: www.brokk.fr)



Roboter zur Entfernung von Asbest von Doppel-T-Trägern (Takenaka)
(Quelle:Arai und Hoshino)



Schwamborn WDS 530
(Quelle: www.schwamborn.com)



Eco-Amiante
(Quelle: H.Y. | Construction Cayola)

Abbildung 2: *Bisherige Lösungen der maschinellen Asbestentfernung*

haltigem Fliesenkleber ab und ist durch die Nutzung von Mecanumrädern und seinen Dimensionen eigens für Wohnungen entwickelt. [8] Ebenso speziell auf ein Problem zugeschnitten ist der von Arai und Hoshino im Auftrag der Takenaka Corporation entwickelte Roboter. Dieser ist dafür ausgelegt, eine spezielle asbesthaltige Brandschutzfarbe von Doppel-T-Trägern in Hochhäusern zu entfernen. Dazu ist auf einer mobilen Plattform mit Ackermann-Lenkung eine vertikale Hubsäule installiert, die sich zum einen an der Decke abstützt und zum anderen einen 7-DOF Manipulator inklusive Peripherie dem Träger zustellt. Das Werkzeug führt in diesem Fall einen schabenden und keinen schleifenden Prozess durch. Der siebte Freiheitsgrad des Manipulators ist für die Bearbeitung der Hinterschnitte eines Doppel-T-Trägers erforderlich. [9] Der Prozess der AS PROTEK unterscheidet sich deutlich von den bisher vorgestellten Ansätzen, da das Verfahren auf Basis eines Hochdruckwasserstrahls arbeitet. Dafür wird ein Rahmen zur Positionierung von Linearachsen über dem zu säubernden Flächenelement aufgestellt und anschließend mit Hilfe einer CNC-Steuerung abgefahren. Der große Vorteil dieses Verfahrens liegt darin, dass keine kontaminierten Stäube entstehen. [10] Jedoch gestaltet sich die anschließende Filterung und Aufbereitung des kontaminierten Wassers als so kostspielig, dass sich die Ersparnisse bei der reduzierten Sicherheitstechnik (weniger Luftkammern und Schutzfolien) durch die zusätzlichen Aufwände der Aufbereitung gegenheben. Alle vorgestellten Prozesse und Maschinen sind in **Abbildung 2** dargestellt.

Entwicklungsstufen des Bots2ReC Baustellenroboters

Bei der Entwicklung des Bots2ReC Roboters haben sich zwei Schlüsselherausforderungen ergeben, welche maßgeblich Einfluss auf die weiteren Entwicklungsaspekte ausgeübt haben. Diese Herausforderungen lauten wie folgt:

1. Gewährleistung der erforderlichen Mobilität in Wohnräumen bei gleichzeitiger Integration aller prozessrelevanten Komponenten innerhalb des robotischen Systems (sowohl Aktuatorik/Kinematik als auch Peripherie)
2. Gewährleistung eines stabilen Entfernungsprozesses, der in allen relevanten Kombinationen aus Umwelt und Roboter den hohen Sicherheitsauflagen gerecht wird

Die erste Herausforderung stellt dabei in erster Linie eine geometrische, bauraumspezifische Entwicklungsproblematik im Wohnungskontext dar, während die Zweite die Prozesssicherheit auf Steuerungs-, Regelungs- und Komponentenebene umfasst. Beide Herausforderungen beeinflussen sich gegenseitig, sodass ihr Wechselspiel berücksichtigt werden muss. Im Rahmen der Entwicklung wurde für die geometrische Auslegung des Roboters auf die Methoden der CAD-Konstruktion sowie der physikalischen Simulation der Kinematik sowie der Aktorik zurückgegriffen. Die Überprüfung von Erreichbarkeiten, Kollisionsproblemen und verfügbaren Arbeitsräumen war dadurch gegeben. Gleichermaßen konnten daraus auch Prozesseigenschaften für die Systemstabilität oder die zu erwartende Endeffektorgeschwindigkeit während des Prozess sichergestellt werden. Eine umfassende simulative Überprüfung der Prozesssicherheit war allerdings durch die stark wechselnden Bedingungen im Bauwesen ausgeschlossen. Die Variabilität von Oberflächen, deren Zusammensetzung sowie Dicken und Rauheiten erforderten experimentelle Verfahren. Aus diesem Grund wurde ein erster Prototyp entwickelt, um schnellstmöglich Wissen und Daten über die Prozessbereiche aufzubauen, welche für den Erfolg des Vorhabens unerlässlich waren: Das geregelte Abschleifen von verputzen Wänden. Dieser erste Prototyp ist in **Abbildung 3** in der linken Spalte gezeigt und war bewusst nicht mit allen zuvor ermittelten Anforderungen konform, da der Erkenntnisgewinn über den Prozess im Vordergrund stand.
Dieser bewusste Schritt zu einem zweistufigen Entwick-

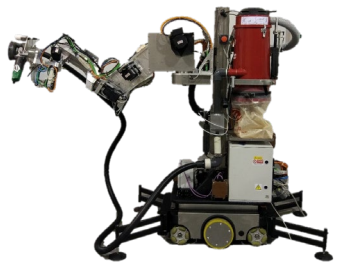
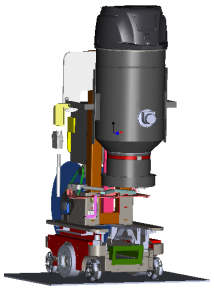
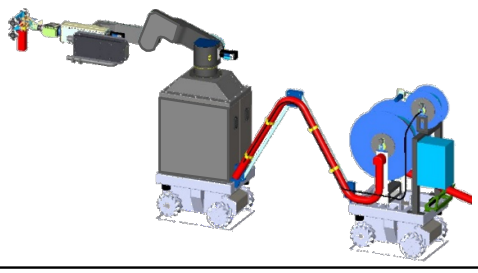
Erster Prototyp	Konzeptstudie Absaugung	Zweiter Prototyp
		
Anzahl mobiler Einheiten: 1 Antrieb mobile Einheit: Differential Manipulator: 6-DOF + Siebte Achse Absaugung: Mitgeführt Stromversorgung: Zentral		Anzahl mobiler Einheiten: 2 Antrieb mobile Einheiten: Omnidirektional Manipulator: 7-DOF Absaugung: Zentral Stromversorgung: Zentral

Abbildung 3: *Entwicklungsstufen des Bots2ReC Roboters*

lungsprozess hat sich rückwirkend als richtig erwiesen, da essentielle Veränderungen hin zum zweiten Prototypen vorgenommen werden mussten. Diese Erkenntnisse werden im Folgenden vorgestellt. Eine starke Einschränkung für das System wurde durch das geforderte Maximalgewicht von 300 kg vorgegeben. Im Zuge dessen wurde ein Akkumulator für einen mehrstündigen Betrieb aus Gewichtsgründen ausgeschlossen und eine dauerhafte externe Stromzufuhr erforderlich. Um Personen aus dem mit Staub kontaminierten Bereich zu halten, wurde diese zentrale Stromzufuhr durch eine mitgeführte aktuierte Kabeltrommel realisiert. Auf diese Weise war es dem Roboter möglich, eigenständig auf Basis seiner durchgeführten Bewegung entsprechend Kabel ab- oder aufzuwickeln. Eine zugrundeliegende Grundrissüberprüfung ergab, dass eine Kabellänge von 20m bis 25m für die anvisierten Wohnungsgrößen ausreichend war. Mit der zentralen Stromzufuhr wurde es möglich, den leistungsintensiven Absaugprozess auf den Roboter zu verlegen. Auf diese Weise konnte auf das Handling eines Absaugschlauches verzichtet werden, indem der Roboter abgesaugten Putz in Säcken verschweißt in der Wohnung deponiert. Um das System weiterhin simpel und leicht zu halten, wurde ein Differentialantrieb für die mobile Grundplattform vorgesehen, welcher im Einklang mit der Plattformgröße auf die Wohnungseigenschaften abgestimmt war. Für das Handling des Schleifwerkzeuges wurde ein eigens entwickelter 6-DOF Manipulator auf einer vertikalen Linearachse entwickelt. Die Eigenentwicklung des Manipulators war durch das hohe Gewicht des Werkzeugs bei gleichzeitig erforderlichem geringem Eigengewicht des Manipulators begründet und wurde auf Basis von strukturoptimierten Aluminiumprofilen realisiert. Der erste Prototyp mit dieser Systemarchitektur erlaubte die experimentelle Durchführung von Versuchsreihen und lieferte unter anderem folgende Erkenntnisse:

- Trotz herstellergerechten Auslegung der Absaugung, war die Absaugleistung zu gering, um über 95% des Putzes abzusaugen. Im Mittelwert leistete die gewählte Absaugeinheit eine Rate von 75%.
- Die Standfestigkeit des Roboters war im zentralen Arbeitsraum unmittelbar zwischen Roboter und Wand ausreichend. Bei seitlich ausladenden Bewegungen, führte die Prozesskraft zu einer Verdrehung des Gesamtsystems, da die Haftreibung der differentiell angetriebenen Räder auf dem Boden überschritten wurde. Die Installation von stabilisierenden Auslegern konnte kurzfristig helfen (siehe **Abbildung 3**).
- Die Kabelführung entlang des Manipulators war unzureichend und schränkte den kinematischen Arbeitsraum so ein, dass viele kompakte Konfigurationen des Manipulators nicht möglich waren. Entsprechend erhöhte sich der minimale Prozessabstand zur bearbeiteten Wand um 20cm.

Insbesondere diese drei Erkenntnisse führten zu einer Veränderung der Systemarchitektur. Das Konzept einer leistungsgerechten Absaugung ist in **Abbildung 3** (Mitte) gezeigt und überschreitet die geometrischen Anforderungen an den Hüllkörper des Roboters. Entsprechend wurde das Konzept eines begleitenden Roboters entwickelt. Im Zuge dessen wurde der Konzeptwechsel von einer mitgeführten Absaugung zu einer zentralen Absaugung durchgeführt, um evtl. notwendige Leistungsanforderungen in Zukunft besser umsetzen zu können. Entsprechend wurde neben dem Kabelhandling auch das Handling des Absaugschlauches auf dem begleitenden Roboter umgesetzt. Bei der anschließenden Validierung dieses Konzepts wurde klar, dass die notwendige Beweglichkeit eines solchen Tandems aus zwei mobilen Robotern nicht auf der Basis eines differentiellen Antriebs umgesetzt werden kann. Aus diesem Grund wurde der Plattformantrieb auf Mecanumräder umgestellt, welche gleichermaßen das Problem der Standfestigkeit lösen sollten, da nun vier aktuierte Aufstandspunkte in der Nähe der Eckpunkte des Roboters existierten. Aus Sicherheitsgründen wurden die stabilisierenden Ausleger für Entwicklungszwecke beibehalten. Um das Problem der Kabelführung und eines potentiell wachsenden Bearbeitungsabstandes zur Wand zu lösen, wurde ein kinematisch redundanter 7-DOF Manipulator entwickelt, der eine besonders platzsparende Konfiguration zwischen Roboter und Wand erlaubte und dadurch das Gewicht der vertikalen Achse einsparte. Bei der Konstruktion wurde auf geschlossene Aluminiumprofile zurückgegriffen, welche innerhalb der Globalstruktur eine interne Verkabelung erlaubten. Dieses Gesamtsystem ist rechts in **Abbildung 3** dargestellt.

Herausforderungen eines gekoppelten Tandems mobiler Roboter im Bauwesen

Im Folgenden wird die Nutzung eines fest gekoppelten Tandems in geschlossenen Räumen beleuchtet, um die neuartigen Anforderungen an die zugehörige Pfadplanung aufzuzeigen. Die Kopplung beider mobiler Einheiten erfolgt sowohl für die ununterbrochene Daten- und Stromverbindung als auch die Bereitstellung der Absaugleistung. Während die Daten- und Stromleitung verhältnismäßig flexibel sind, liegt das Hauptaugenmerk auf dem Absaugschlauch. Durch die erforderliche Wandstärke sowie den Durchmesser, ist er relativ steif und somit nur in großen Radien biegsam. Dies hat zur Folge, dass der Begleitroboter nicht alle beliebigen Positionen und Orientierungen um den Entfernungsroboter annehmen kann. Während der Entwicklung wurde für diese Zustände ein diskretes Konfigurationsschema eingeführt, um entsprechende Konfigurationswechsel besser planen und durchführen zu können. Dieses Schema basiert auf den Himmelsrichtungen eines Kompasses und trifft die Aussage darüber, in welcher der acht Positionen sich der Begleitroboter im Verhältnis zum Entfernungsroboter befindet. Die Orientierung zueinander wird zusätzlich durch die Stellungen der x-Achsen beider Roboter dargestellt. Dabei wird der Entfernungsroboter stets als Referenzsystem verwendet und darin die Orientierung der x-Achse des Begleitroboters ausgedrückt. Die Angabe „W→“ bedeutet in diesem

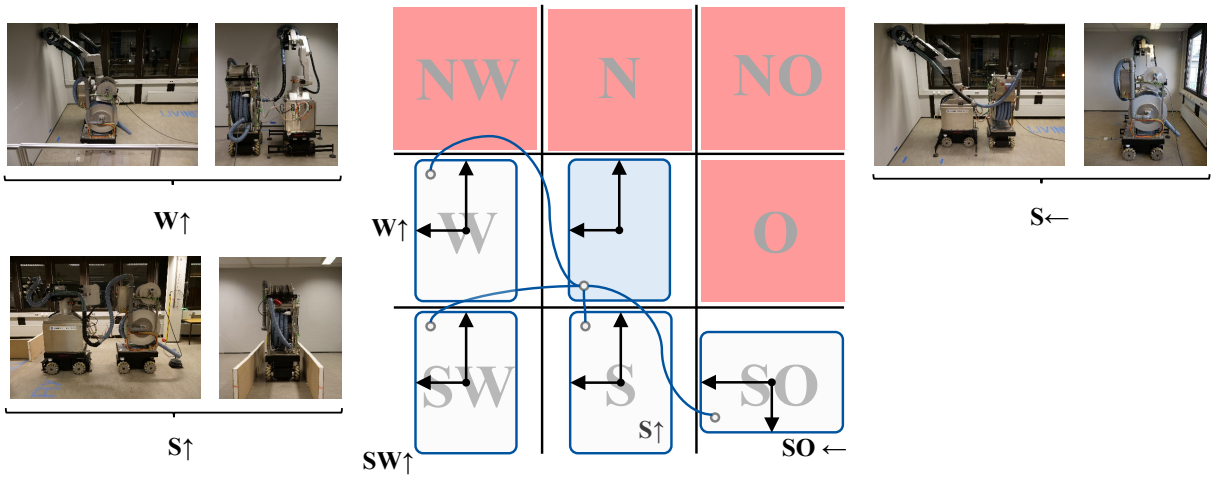


Abbildung 4: *Konfigurationsmöglichkeiten des Bots2ReC Tandems mit Szenendarstellung*

Schema also, dass sich der Begleitroboter links neben dem Entfernungsroboter befindet und die x-Achse dem Entfernungsroboter normal zugerichtet ist. Ziel dieser stark vereinfachten Angabe von Konfigurationen ist es, für den teleoperierten Betrieb durch den Bediener leicht verständliche Vorgaben machen zu können, die anschließend durch die Roboter automatisch umgesetzt werden können. Die Besonderheit beim Bots2ReC Tandem ist die asymmetrische Beschränkung der möglichen Konfiguration. Dies bedeutet, dass durch den Absaugschlauch als Kopplungselement zwar die Konfiguration „W↑“ möglich ist, jedoch die gespiegelte Konfiguration „O↑“ nicht eingenommen werden kann, da der Absaugschlauch aus konstruktiven Gründen links am Manipulator vorbeigeführt werden muss. Eine Darstellung der einnehmbaren Konfigurationen inklusive des beschriebenen Schemas sowie einiger realen Beispiele ist in **Abbildung 4** gegeben.

Fallstudie

Um die notwendigen Eigenschaften eines automatischen Pfadplaners zu ermitteln, wird anhand eines realen Wohnungsgrundrisses der Einsatz des Bots2ReC Tandems analysiert. Der Grundriss entstammt einem realen Sanierungsprojekt aus Frankreich und stellt eine typische Wohnung für den anvisierten Anwendungsfall dar. In **Abbildung 5** ist der Roboter im Realmaßstab in diesem Grundriss in unterschiedlichen Beispielkonfigurationen dargestellt. Während die annähernd quadratischen Räume „Wohnen“, „Essen“ und „Schlafen“ genug Platz für eine Rekonfiguration des Bots2ReC Tandems bieten, eignen sich verwinkelte Räume wie das „Bad“ oder eher schmale Räume wie die „Küche“ oder der „Flur“ schlecht oder überhaupt nicht für die Durchführung von Rekonfigurationen. Im allgemeinen Fall bedeutet dies, dass der Roboter nicht alle Raumseiten in dieser Konfiguration bearbeiten kann. Dies hat weiterhin zu Folge, dass die Startkonfiguration des Tandems, mit der es in einen schmalen oder verwinkelten Raum hineinfährt, automatisch die Flächen vorgibt, welche bearbeitet werden können. Es ist also notwendig, mehrfach in diese Räume aus und einzufahren, um im nächstgelegenen größeren Raum die Konfiguration herzustellen, mit der anschließend die fehlenden Wände bearbeitet werden können. Ein automatischer Pfad- und Konfigurationsplaner für das Bots2ReC Tandem muss entsprechend über folgende Eigenschaften verfügen:

- Berücksichtigung der Topologie des Grundrisses bzw. der Räumlichkeiten
- Berücksichtigung der Geometrie der Räumlichkeiten
  - Eine Klassifikation in Räume mit und ohne Rekonfigurationsvermögen erscheint naheliegend
- Berücksichtigung der Kopplungs- und somit der Rekonfigurationsbedingungen zwischen den Tandempartnern
- Zielfunktion zur Lösung des Problems sollte die Minimierung der Bearbeitungszeit sein
  - Eine Optimierung oder eine zugeschnittene Lösungsheuristik ist entsprechend naheliegend

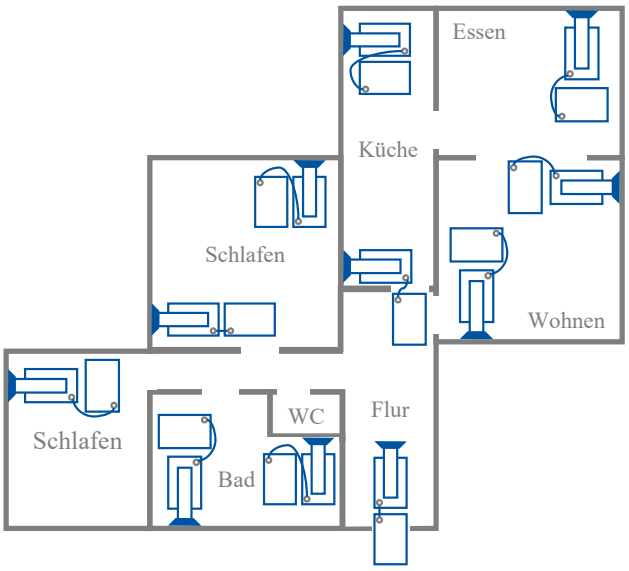


Abbildung 5: *Einsatz verschiedener Konfigurationen innerhalb einer Wohnung*

Zusammenfassung

Dieser Beitrag zeigt die bekannten maschinellen Lösungen für das Entfernen von Asbest im Baubereich auf und stellt auf Basis dessen, den Entwicklungsprozess eines neuen mobilen Manipulators zur Asbestentfernung in Wohnräumen vor. Dabei wird der zweistufige Entwicklungsprozess beschrieben und die getroffenen Entscheidungen begründet und untereinander in Beziehung gesetzt. Dabei werden insbesondere die Beweggründe für die Entwicklung eines gekoppelten Robotertandems vorgestellt. Die zugehörigen eingeführten Nomenklaturen zur Entwicklung dieses Tandems und auf die damit einhergehenden Herausforderungen werden präsentiert. In Form einer Fallstudie wird aufgezeigt, dass der Wohnungskontext keine freie Bearbeitung durch ein Robotertandem dieser Größe erlaubt und ein zugeschnittener Pfad- und Konfigurationsplaner erforderlich ist. Die Eigenschaften eines solchen Planers werden hergeleitet und legen damit die grundlegenden Anforderungen für dessen Entwicklung nahe. Die Entwicklung dieses automatischen Pfad- und Konfigurationsplaners ist aktuell aktives Forschungsthema am IGMR der RWTH Aachen University.

Danksagung

Dieser Beitrag wurde auf Basis des *Robots-to-ReConstruction* (Bots2ReC) Projektes erarbeitet, welches unter der Grant Agreement Nr. 687593 von der Europäischen Union im Rahmen des *Horizon 2020 Research and Innovation Program* gefördert wurde.

Literatur

[1] Bock, T.: *Construction robotics*. In: Autonomous Robots (22) 2006. Springer Science+Business Media. DOI: 10.1007/s10514-006-9008-5, S. 201-209

[2] Linner, T.: *Automated and Robotic Construction: Integrated Automated Construction Sites*. 2013 Technische Universität München

[3] Detert, T., Charaf Eddine, S. Fauroux J.-C., Haschke, T., Becchi, F., Corves, B., Guzman, R., Herb, F., Linéatte, B., Martin, D.: *Bots2ReC: introducing mobile robotic units on construction sites for asbestos rehabilitation*. In: Construction Robotics 2017. DOI: 10.1007/s41693-017-0007-1. S. 29-37.

[4] L'Union sociale pourl'habitat: Amiante: *l'Union sociale pour l'habitat réagit*. April 2014. Zuletzt aufgerufen am 14.01.2021, von https://www.union-habitat.org/sites/default/files/communiques/documents/2018-01/amiante%20-%20l%27USH%20r%C3%A9agit_02.04.14.pdf

[5] Karges, W. und Dahouk, S. A.: *Innere Medizin in 5 Tagen* 2014. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-41618-7

[6] Brokk: Planers – Technical Specifications 2020. Zuletzt aufgerufen am 14.01.2021, von https://www.brokk.com/product-attachment/planers/

[7] Schwamborn: *The New WDS 530,* 2019. Zuletzt aufgerufen am 14.01.2021, von https://www.schwamborn.com/content/1-produkte/1-wds-530/schwa-19-03-002_wds-prospekt_en.pdf

[8] H.Y.: *Robotic Raboteuse : le robot d'Eco-Amiante* 2016. Zuletzt aufgerufen am 14.01.2021, von https://www.constructioncayola.com/environnement/article/2016/09/19/107550/robotic-raboteuse-robot-ecoamiante

[9] Arai, M. und Hoshino, H.: *Robot Systems for Removing Asbestos Sprayed on Beams* In: Proceedings of 28th International Symposium on Automation and Robotics in Construction 2011. DOI: 10.22260/ISARC2011/0265. S. 1409-1414.

[10] Ramboud, G.: *Les professionnels du désamiantage priment l'eau sous pression* 2015. Zuletzt aufgerufen am 14.01.2021, von https://www.lemoniteur.fr/article/les-professionnels-du-desamiantage-priment-l-eau-sous-pression.687254

Entwicklung eines autonomiefokussierten hochmobilen Bodenrobotersystems für den Katastrophenschutz

Development of an autonomy-focused highly mobile rescue robot

Marius Schnaubelt, Tobias Ullrich, Moritz Torchalla, Jonas Diegelmann, Matthias Hoffmann, Prof. Dr. rer. nat. Oskar von Stryk, Technische Universität Darmstadt, Fachgebiet Simulation, Systemoptimierung und Robotik, 64289 Darmstadt, Deutschland, {schnaubelt, ullrich, stryk}@sim.tu-darmstadt.de

Kurzfassung

Mobile Rettungsroboter ermöglichen den menschlichen Bedienern die Bearbeitung von Aufgaben aus sicherer Entfernung in risikoreichen Umgebungen. Die aktuell übliche Teleoperation der Robotersysteme verursacht aufgrund der unstrukturierten Umgebung der komplexen und vorab unbekannten Einsatzszenarien eine hohe kognitive Belastung für den Roboteroperator, was schnell zur Ermüdung führt. Durch intelligente autonome Assistenzfunktionen können die Operatoren entlastet werden, wodurch die Wahrscheinlichkeit von Bedienfehlern reduziert und die Effizienz des Robotereinsatzes erhöht werden kann. Diese innovativen Assistenzfunktionen benötigen jedoch ein mechatronisches Design, dessen Anforderungen an Hard- und Software für ein effektives Gesamtsystem eng aufeinander abgestimmt und umgesetzt werden müssen. Die Entwicklung eines hochmobilen autonomiefokussierten Bodenroboters mit modularen Sensornutzlasten ermöglicht dem Operator ein umfassendes Situationsbewusstsein und bietet Unterstützung bei Navigation und Manipulation. Die Evaluation des Gesamtsystems und von Einzelkomponenten analysiert die Erfüllung des Anforderungskatalogs und demonstriert so die Eignung für (semi-)autonome Rettungsrobotikeinsätze.

Abstract

Mobile rescue robots enable human operators to perform tasks from a safe distance in hostile environments. Due to the unstructured environment of the mostly complex and a priori unknown mission scenarios, the currently common teleoperation of robotic systems causes a high cognitive load for the robot operator, which quickly leads to fatigue. Intelligent autonomous assistance functions can relieve the operators increasing reliability and efficiency. However, autonomous assistance functions require a mechatronic system design whose hardware and software requirements are considered and implemented in close connection. The development of a highly mobile autonomy-focused ground robot with modular sensor payloads provides comprehensive situational awareness to the operator and assistance in navigation and manipulation. The evaluation of the overall system and individual components analyzes the fulfillment of the requirements catalog and thus demonstrates the suitability for (semi-)autonomous rescue robotics missions.

1 Einleitung

Mobile Bodenroboter können in Katastrophengebieten den menschlichen Bedienern ein verbessertes Situationsbewusstsein basierend auf Sensordaten (z.B. Kameras) aus sicherer Entfernung bereitstellen [1]. Dies ermöglicht den Rettungskräften die sichere Bearbeitung von Aufgaben in menschenfeindlicher Umgebung, zum Beispiel die Erkundung in unstrukturiertem Gelände oder das Steuern des Roboterarms für Manipulationsaufgaben.

Casper u. a. [2] nutzten Rettungsroboter, um nach dem Einsturz der Türme des World Trade Centers durch die Terroranschläge am 11. September 2001 nach menschlichen Überlebenden zu suchen und um die Struktur der Ruinen zu inspizieren. Nach der Nuklearkatastrophe von Fukushima im Jahre 2011 wurden von Nagatani u. a. [3] Bodenroboter eingesetzt, um das Innere der Reaktorgebäude zur Lagebilderstellung zu erkunden.

Die Umgebung der Einsätze ist in der Regel unstrukturiert, im Voraus unbekannt und enthält oft Hindernisse, welche überwunden werden müssen, um das Missionsziel zu erreichen. Durch die unvorhersehbare Natur von Katastrophen ist die vielseitige Einsetzbarkeit von Rettungsrobotern von hoher Relevanz.

Intelligente autonome Flugroboter [4] können dank schneller Fortbewegung in der Luft und der Fähigkeit, Hindernisse überfliegen zu können, ein großflächiges Lagebild liefern. Jedoch sind die mögliche Sensornutzlast sowie die maximale Einsatzdauer ebenso wie witterungsabhängige Einsatzmöglichkeiten vergleichsweise eingeschränkt. Außerdem sind für die Erkundung von beengten Umgebungen sehr präzise Kollisionsvermeidungsfähigkeiten oder kollisionstolerante Konstruktionen notwendig, was die Effektivität von Flugrobotern in solchen Umgebungen einschränkt [1].

Bodenroboter mit Kettenantrieb werden dank einer relativ einfachen Regelung, höherer Traktion in unwegsamen Gelände im Vergleich zu Rädern und ihrer Transportkapazität für schwere Nutzlasten häufig als Rettungsroboter eingesetzt. Kommerziell erhältliche Systeme wie zum Beispiel die Telerob Telemax-Familie oder der iRobot PackBot sind vor allem auf teleoperierte Erkundung oder Manipulation ausgelegt und haben nur vereinzelt erste Assistenzfunktionen, zum Beispiel Inverskinematik (IK) und Tool-Center-Point-Steuerungen für den Manipulator sowie simultane Lokalisierung und Kartierung (SLAM) in 2D.

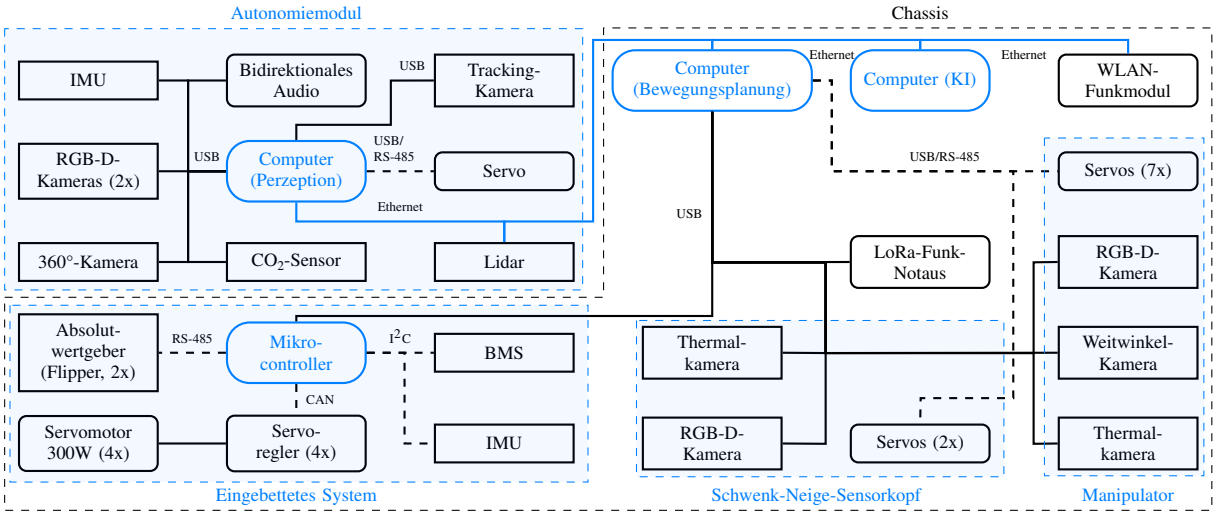


Abbildung 1 Überblick über die Architektur des Bodenrobotersystems mit einem Autonomiemodul als Beispiel für eine modulare Sensor- und Funktionsnutzlast.

Bodenroboter aus der Forschung [3, 5–7] fokussieren sich häufig auf hohe Mobilität und gute Manipulationsfähigkeiten, aber bislang noch weniger auf die Anforderungen autonomer Assistenzfunktionen (siehe Abschnitt 2.1). Schneider u. a. [8] präsentierten ein für Assistenzfunktionen nutzbares Modularisierungskonzept für die Nachrüstung von großen Bodenrobotern mit großen Tragflächen und -lasten. Dieses lässt sich jedoch nicht direkt auf kompakte und hochmobile Bodenroboter übertragen, die nur wenig Fläche für zusätzliche Sensor- und Computing-Module bieten. Die Teleoperation der aktuellen kompakten Bodenrobotersysteme in komplexen Einsatzszenarien erzeugt eine hohe kognitive Belastung für den Roboteroperator [9]. Dadurch ist nicht nur die Wahrscheinlichkeit von Bedienfehlern hoch, es wird zudem eine zuverlässige Datenverbindung mit niedriger Latenz und hoher Bandbreite zwischen Roboter und Bediener benötigt. Rettungsroboter mit integrierten Assistenz- und KI-Funktionen zur Entlastung des Bedieners können hier helfen, die kognitive Last des Bedieners und somit auch die Wahrscheinlichkeit von Bedienfehlern zu reduzieren. Außerdem können die benötigten Anforderungen an die Datenverbindung reduziert werden, was wiederum die Einsatzmöglichkeiten erweitert.

Die Nachrüstung von bisher rein teleoperierten Robotern durch Sensorik und Onboard-Computer liefert suboptimale Ergebnisse, da häufig die Umfeldsensorik abgeschattet wird und zudem die interne Sensorik und Gelenkregelung nicht mit adäquaten Taktraten und Genauigkeiten angesprochen werden kann. Durch einen erhöhten Roboterschwerpunkt und mögliche Selbstkollisionen mit den neuen Aufbauten wird die Mobilität des Roboters zudem eingeschränkt. Für ein hochperformantes intelligentes mechatronisches System müssen die Hard- und Softwarekomponenten des Roboters daher eng aufeinander abgestimmt werden.

Basierend auf mehrjähriger Forschung im Bereich künstlicher Intelligenz und autonomen Fähigkeiten für mobile Rettungsroboter [10, 11] wird die Entwicklung eines kettenbasierten, hochmobilen und modularen Bodenroboters vorgestellt, der dank Erfüllung der Anforderungen an intelligente autonome Assistenzfunktionen optimale Voraussetzungen für Entwicklung und Einsatz leistungsfähiger Algorithmen bietet.

2 Design- und Entwicklungsprozess

Der Fokus des zu entwickelnden Robotersystems liegt auf der autonomen Exploration von unbekanntem Gebiet bei gleichzeitig hoher Mobilität und Manipulationsfähigkeit, auch bei schwacher oder nicht vorhandener Funkverbindung. Hierfür soll das System mit modularen Sensornutzlasten ausgerüstet werden können, welche verschiedene Funktionalitäten bieten.

2.1 Anforderungen für intelligente Assistenzfunktionen

Intelligente Assistenzfunktionen für Navigation und Manipulation benötigen eine akkurate und vollständige Erfassung des Umfelds und des eigenen Roboterzustands. Die Umfeldsensorik sollte möglichst das komplette Gebiet rund um den Roboter abdecken, um eine robuste Lokalisierung und Hindernisvermeidung gewährleisten zu können, die Nahbereiche direkt vor und hinter dem Roboter sind dabei jedoch am wichtigsten. Für die Generierung eines qualitativ hochwertigen Umgebungsmodells ist die Positionierung der externen Sensorik essenziell, um Abschattungen durch die eigene Robotergeometrie zu vermeiden. Hieraus entsteht allerdings ein Zielkonflikt zwischen Umfeldsensorik, Mobilität und Manipulationsfähigkeit des Roboters. Neben der möglichst guten Umfeldsensorik mit ausreichend leistungsfähigem Onboard-Computing zur Verarbeitung der Daten sollte ein Rettungsroboter möglichst klein, leicht und geländegängig sein. Die für eine erhöhte Mobilität – zum Beispiel die Überwindung von Hindernissen – benötigten Flipper (verstellbare Kettenerweiterun-

gen) mit möglichst großer Länge sorgen jedoch für eine Sensorabschattung. Daher sollten die Flipper kompakt einfahrbar sein, wenn diese nicht benötigt werden. Für vielseitige Manipulationsfähigkeiten wird ein Manipulator mit großer Reichweite und hoher Nutzlast benötigt, um zum Beispiel Türen öffnen zu können. Dieser beeinträchtigt wiederum die Umfeldsensorik, weshalb sich der Manipulator in eine kompakte Konfiguration fahren lassen sollte, wenn er nicht benötigt wird. Für zeiteffiziente Explorationen sollte sich der Roboter in der Ebene mit mindestens $1\,\mathrm{m\,s^{-1}}$ fortbewegen können, sowie Treppen mit bis zu 45° Steigung überwinden können. Für die Hindernisüberwindung sowie für vielseitige Möglichkeiten bei der Manipulation sollte sich der Roboter aus eigener Kraft auf die Flipper stellen können. Da der menschliche Operator im autonomen Betrieb keine kleineren Regelungsfehler von Manipulator, Flipper und Sensorkopf ausgleichen kann, sind präzise Positionsregelungen der Robotergelenke notwendig, welche über Absolutwertgeber die aktuellen Gelenkpositionen zurückmelden können. Die verfügbaren Funktionen der Hardware müssen über geeignete und erweiterbare Softwareinterfaces für die Algorithmik ansprechbar und überwachbar sein. Beispiele hierfür sind das Anfahren von Arm- und Flippersollpositionen für Manipulation und Hindernisüberwindung oder die Geschwindigkeitsregelung der Roboterbasis mit rückgeführter Odometrie zum Abfahren von geplanten Trajektorien bei der autonomen Navigation. Während der Überwindung von unwegsamen Gelände sind hochfrequente, zeitsynchronisierte und genaue Messdaten von der inertialen Messeinheit (IMU) und dem Lidar notwendig, um auch bei hochdynamischen Bewegungen z.B. dem Kippen des Roboters, eine robuste 3D-SLAM-Posenschätzung berechnen zu können. Für eine effektive Anpassung an vorliegende Situationen sollte sich die Umfeldsensorik schnell in Form von modularen Aufbauten, vergleichbar zu [8], allerdings auf wesentlich kleinerer Fläche, wechseln lassen. Zudem sollte sich das System durch einfach austauschbare Modulkomponenten einfach warten lassen.

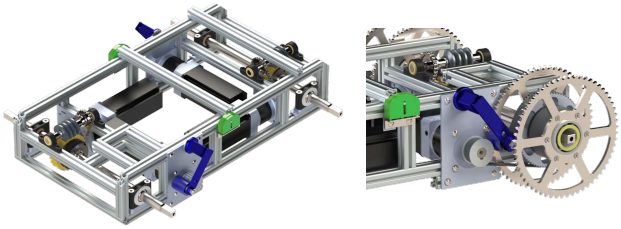
2.2 Prototypentwicklung

Basierend auf den in Abschnitt 2.1 genannten Kriterien und den gesammelten Erfahrungen mit dem zuvor entwickelten Bodenrobotersystem ohne Flipper [11] wurde ein autonomiefokussiertes und hochmobiles Bodenrobotersystem entwickelt. **Abb. 1** gibt einen Gesamtüberblick über dessen Architektur, die inklusive des eingebetteten Systems auf einer quelloffenen Middleware, dem Robot Operating System (ROS) [12] (Version Noetic) aufbaut.

2.2.1 Digitaler Zwilling

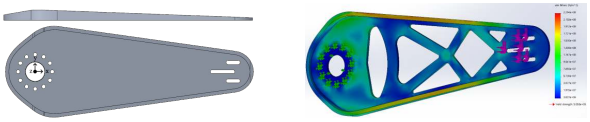
Zur Planung der Umfeldsensorikplatzierung sowie der allgemeinen Robotergeometrie wurde parallel zum realen Prototyp ein digitaler Zwilling von Roboterbewegungen, Sensorik und Interaktion mit der Umwelt in der physikbasierten Echtzeitsimulation Gazebo[1] entwickelt.

[1] `http://gazebosim.org/`



(a) Der Profilrahmen mit installierten Hauptkomponenten. **(b)** Der Kettenantrieb für Haupt- und Flipperkette

Abbildung 2 CAD-Renderings der Chassiskonstruktion.



(a) Vorder- und Seitenansicht der initialen Flipperstruktur. **(b)** Die gewichtsoptimierte Flipperstruktur.

Abbildung 3 Die FEM-basierte Topologie-Optimierung der Flipperstruktur.

2.2.2 Chassis

Der Rahmen des Roboters besteht aus mit Dibondplatten verkleideten Aluminiumprofilen und stellt die Befestigungspunkte für die Kernkomponenten wie Motoren, Achslager, Manipulator und externe Sensormodule bereit, siehe **Abb. 2a**. Die unteren Längsprofile des Rahmens sind nicht durchgehend, so dass die rotationssymmetrisch angeordneten Motoren alle in der gleichen Ebene liegen können, was zu einem niedrigeren Schwerpunkt, einer symmetrischeren Gewichtsverteilung und einer besseren Wärmeableitung durch die Aluminium-Bodenplatte bei nur geringen Stabilitätseinbußen führt.

Kettenantrieb Der Roboter wird von zwei bürstenlosen 300 W AC-Servomotoren von LS Mecapion mit einem Nennmoment von 0,95 N m bei einer Nenndrehzahl von $3000\,\mathrm{min^{-1}}$ angetrieben. Das Spitzendrehmoment der Motoren beträgt 2,85 N m, die Motorregelung erfolgt unter Nutzung der im Motor integrierten Inkrementalgeber über Elmo Motion Control Whistle 10/60 Motorcontroller, welche mit 48 V versorgt werden. Zur Bereitstellung des notwendigen Drehmoments werden die beiden auf der Flipperwelle gelagerten Rollenkettenräder über ein zweistufiges Getriebe mit einem Gesamtübersetzungsverhältnis von 1 : 30 angetrieben (siehe **Abb. 2b**). Die erste Stufe ist ein PLQE-Planetengetriebe von Neugart mit einem Übersetzungsverhältnis von 1 : 10. Die nächste Stufe besteht aus einem außerhalb des Roboters liegenden Zahnriemengetriebe mit einem Übersetzungsverhältnis von 1 : 3. Die gewählte Übersetzung führt zu einer Translationsgeschwindigkeit der Ketten von knapp über $1\,\mathrm{m\,s^{-1}}$ bei Nenndrehzahl in der Ebene und ausreichend Drehmoment zum Befahren von Treppen mit 45° Steigung. Die starr verbundene Kettenradbaugruppe treibt die Haupt- und Flipperkette an. Zur Gewichtsersparnis sind die Kettenräder mit Aussparungen versehen und aus unlegiertem Titan gefertigt.

Flipperbaugruppe Zur Erhöhung der Mobilität erhält die Roboterplattform vier Flipper, von denen die vorderen und hinteren Flipper über eine gemeinsame Welle gekoppelt sind. Die beiden Flipperwellen werden jeweils über ein zweistufiges Getriebe (Planetengetriebe und selbsthemmendes Schneckengetriebe) mit einer Gesamtübersetzung von 1 : 168 durch einen 300 W AC-Servomotor angetrieben (siehe Abschnitt 2.2.2). Das Schneckenrad ist über eine Rutschkupplung mit der kontinuierlich rotierbaren Flipperwelle verbunden, welche die 20 mm starke Flipperwelle aus unlegiertem Titan vor Übermoment schützt. Die Positionsregelung der Flipper wird über kapazitative Absolutwertgeber (CUI AMT21) mit 12 Bit Auflösung ermöglicht, welche über Zahnriemen mit der Flipperwelle gekoppelt sind. Die Hauptstruktur des Flippers besteht aus Aluminium 7075 T6, welcher über eine Vierkantverbindung formschlüssig mit der Flipperwelle verbunden ist. **Abb. 3** zeigt die per Finite-Elemente-Methode (FEM) topologieoptimierte Flipperstruktur zur Gewichtsreduktion. Die Flipperlänge wurde so gewählt, dass sich ein möglichst lange, jedoch einseitig vollständig einklappbare Flipper ergeben, um die Sichteinschränkung der Umfeldsensorik zu minimieren.

Eingebettetes System Der Teensy 4.0 Mikrocontroller, welcher als Basis des eingebetteten Systems genutzt wird, ist über USB-Rosserial[2] an das ROS-System angeschlossen. Dieser liest die Absolutwertgeber aus und kommandiert die Servoregler der Flipper- und Antriebsmotoren über CAN (Controller Area Network) mit 100 Hz Regelfrequenz. Zusätzlich ist ein Batteriemanagementsystem (BMS) integriert, welches die Akkumulatorzellspannungen überwacht. Eine zusätzliche IMU dient der verbesserten differentiellen Geschwindigkeitsregelung der Roboterbasis. Die Ansteuerung des eingebetteten Systems und die Rückmeldung von Sensor- und Diagnosedaten erfolgt direkt über das ROS-System.

2.2.3 Modulare Sensornutzlasten

Als Basis für die vielfältigen Assistenzfunktionen stehen zwei modulare Sensornutzlasten zur Verfügung, das Autonomiemodul sowie das Navigationsmodul für gestörte Sichtbedingungen. Die modularen Sensornutzlasten werden über Ethernet und 24 V Versorgungsspannung an die Roboterbasis angebunden.

Autonomiemodul Für 3D-SLAM und 3D-Perzeption ist das Autonomiemodul mit einem kontinuierlich rotierenden Velodyne VLP-16 3D-Lidar ausgestattet, der über ein Stirnradgetriebe von einem Robotis Dynamixel XM430-W350 Servo angetrieben wird. Die Übertragung der Daten und Stromversorgung ist mittels eines Schleifrings realisiert. Mit diesem Aufbau wird eine nahezu vollständige Abdeckung aller Richtungen mit hochgenauen und dichten Punktwolkendaten erreicht. Im Nahbereich des Roboters werden die Tiefendaten des Lidars durch zwei nach vorne und hinten gerichtete Intel Realsense D435 RGB-D-Kameras ergänzt. Oberhalb des Lidars ist eine kompakte, 360°-Kamera (Insta360 Air) montiert, welche für die omnidirektionale Erfassung visueller Informationen verwendet wird. Das 360°-Bild kann dazu benutzt werden, Objekte in allen Richtungen gleichzeitig zu erkennen, ohne dass eine Kamera mechanisch bewegt werden muss. Um die Lage der Plattform zu messen, ist das Fahrzeug mit leistungsfähiger Inertialsensorik (Microstrain 3DM-GX3-45 IMU) ausgestattet. Diese misst die Beschleunigungen und Winkelgeschwindigkeiten und schätzt die Orientierung des Sensors. Eine Intel Realsense T265 Tracking-Kamera wird zur Verbesserung der Geschwindigkeitsregelung des Roboters genutzt. Zur Onboard-Verarbeitung der Daten wird ein Intel Hexacore NUC verwendet.

Navigationsmodul für gestörte Sichtbedingungen Visuelle Sensoren wie Lidar und Kameras liefern unter Einfluss von Rauch, Feuer, aufgewirbeltem Staub oder Nebel nur ungenügende und unzuverlässige Messwerte. Um weiterhin eine robuste Lokalisierung und Navigation unter diesen Bedingungen gewährleisten zu können, wurde das Modul mit Radar- und Ultraschallsensoren ausgestattet, die von diesen Aerosolen nicht signifikant beeinflusst werden. Die Rohdaten des Texas Instruments (TI) IWR1443BOOST Radarsensors werden über ein TI DCA1000EVM Modul in Echtzeit über Ethernet an einen kleinen Einplatinencomputer gestreamt. Die Radarmessungen werden für das 2D- SLAM-Verfahren durch eine IMU und drei in 45° zueinander angewinkelten Ultraschallsensoren (MaxBotix MB1300 XL-MaxSonar-AE0) ergänzt, welche durch einen Teensy 4.0 Mikrokontroller ausgelesen werden. Für eine höhere Datendichte wird das Radar durchgehend durch einen Robotis Dynamixel XM430-W210 Servo rotiert, Gigabit-Ethernet und Versorgungsspannung werden über einen Schleifring übertragen. Für einen direkten qualitativen Vergleich des für Radarsensoren angepassten SLAM-Verfahrens von [13] sowie für die multimodale Sensordatenfusion ist das Modul mit einem Hokuyo UTM-30LX-EW 2D-Lidar ausgestattet.

2.2.4 Manipulator

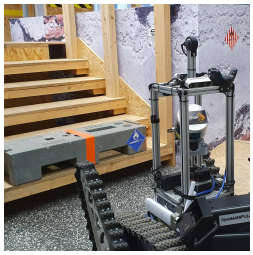
Der eingesetzte RM-P60-RNH Manipulator von Robotis hat die Kinematik eines industriellen Vertikalknickarmroboters mit sechs Achsen. Auf dem Robotis RH-P12-RN Endeffektor sind eine Seek Thermal Wärmebildkamera, eine Intel Realsense D435 RGB-D-Kamera sowie eine Logitech C920 Kamera mit Weitwinkelobjektiv montiert. Der Arm lässt sich per MoveIt[3] über ROS ansprechen.

2.2.5 Schwenk-Neige-Sensorkopf

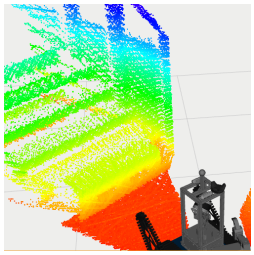
Für die ergänzende Umfeldwahrnehmung und die Endeffektorverfolgung ist der aus zwei Robotis Dynamixel XM430-W350 aufgebaute Schwenk-Neige-Sensorkopf mit einer Seek Thermal und einer Intel Realsense D435 RGB-D-Kamera ausgestattet.

[2] `http://wiki.ros.org/rosserial`

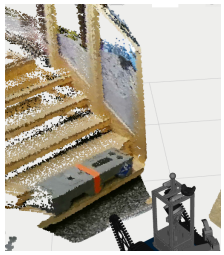
[3] `https://moveit.ros.org/`



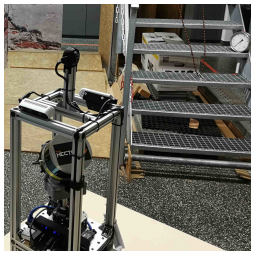
(a) Überblick der ersten Szene.



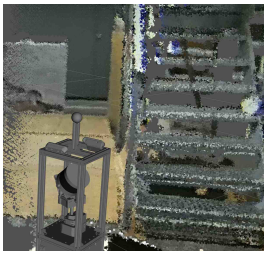
(b) Die nach Höhe über dem Boden eingefärbte dichte Punktwolke.



(c) Die mit der 360°-Kamera kolorierte Punktwolke.



(d) Überblick der zweiten Szene.



(e) Die mit der 360°-Kamera kolorierte Punktwolke.

Abbildung 4 Die kolorierten Lidar-Punktwolken [11] bieten ein verbessertes Situationsbewusstsein für den Roboteroperator.

3 Experimentelle Validierung und Ergebnisse

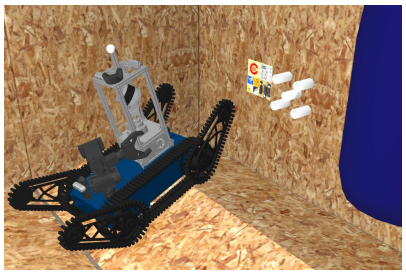
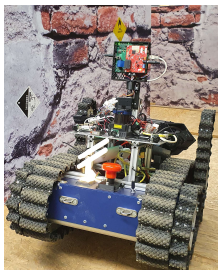


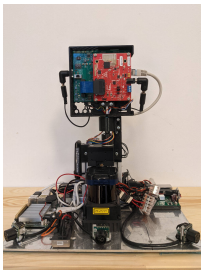
Abbildung 5 Gazebo-Simulation der Roboterplattform.

Digitaler Zwilling Der in **Abb. 5** gezeigte digitale Zwilling kann durch vollständig simulierte Sensorik und Physik mit zum realen Roboter kompatiblen Softwareinterfaces zur Entwicklung und ersten Evaluation von intelligenten Assistenzfunktionen genutzt werden.

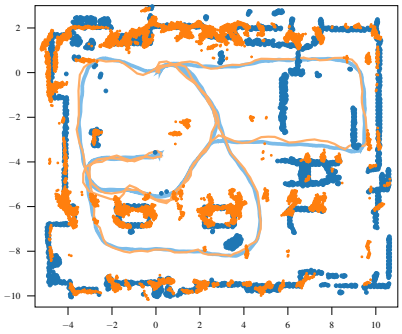
Modulare Sensornutzlasten Das System integriert zahlreiche Sensoren in das Autonomiemodul, welche die Entwicklung und Anwendung umfassender Operatorassistenzfunktionen ermöglichen. **Bild 4b** zeigt die durch den drehenden Lidar erstellte dichte Punktwolke der Umgebung (**Bild 4a**), welche nach Höhe über dem Boden eingefärbt ist und so räumliche Informationen liefert. Durch die Kolorierung der Punktwolke (**Bild 4c**) mit den Bilddaten der 360°-Kamera [11] wird ein verbessertes Situationsbewusstsein für den Roboteroperator ermöglicht.
Dank dem in **Abb. 6** gezeigten Modul ist auch die Lokalisierung und Navigation in Umgebungen mit gestörten Sichtbedingungen möglich. Während die durch das Radar-SLAM erstellte Karte im Vergleich zur mit [13] erstellten Lidarkarte noch weniger dicht ist, liegt der Lokalisierungsfehler der Trajektorie mit einem durchschnittlichen Translationsfehler von $8{,}6 \pm 5{,}9$ cm und einem durchschnittlichen Rotationsfehler von $0{,}27 \pm 0{,}29°$ schon in der gleichen Größenordnung. Durch den Einsatz eines hochwertigeren Radarmoduls würden sich dichtere Karten erzeugen lassen.

Mobilität Zum Test der Mobilität des Gesamtsystems wurden verschiedene, an [14] angelehnte Tests durchgeführt, siehe **Abb. 7**. Die daraus resultierenden Spezifikationen der hochmobilen Plattform sind in Tabelle 1 gelistet. Die spezielle Konstruktionsweise der Roboterkette sorgt für hohe Traktion auf Flächen sowie an Kanten und bietet eine passive Dämpfung. Treppen mit 45° Steigung lassen sich zwar unter Ausnutzung von Leistungsreserven der Servoregler befahren, für eine durchgehende Leistung muss der Kettenantrieb nochmals optimiert werden. Der Roboter kann sich aus eigener Kraft auf beide Flipper stellen, um so Hindernisse überwinden zu können. Somit erfüllt das entwickelte Robotersystem die gestellten Mobilitätsanforderungen.





(a) Roboterplattform mit installiertem Modul. **(b)** Nahaufnahme des Moduls.

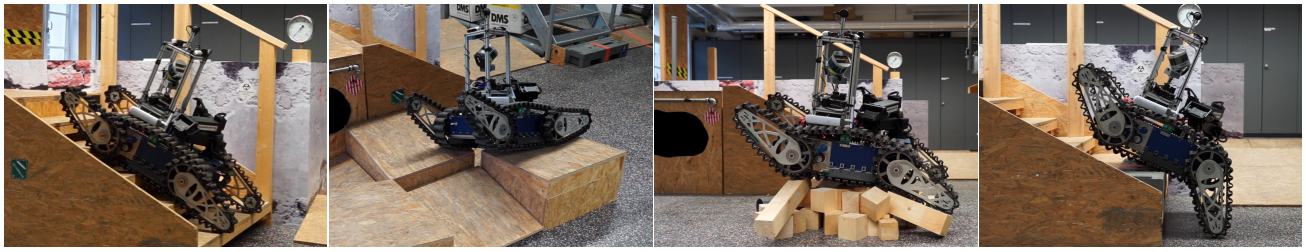


(c) Die vom Radar (orange) erstellte Karte und Trajektorie im Vergleich mit der des Lidars (blau).

Abbildung 6 Vergleich der vom Navigationsmodul für gestörte Sichtbedingungen mit Radar und Lidar erstellten 2D-Karten und Lokalisierungen.

4 Zusammenfassung und Ausblick

In diesem Beitrag wurden Design und Entwicklung eines kettenbasierten, hochmobilen und modularen Bodenroboters vorgestellt, der die Anforderungen an intelligen-



(a) Treppe mit 35 Grad Steigung. **(b)** Erhöhte angewinkelte Rampen. **(c)** Ein Geröllhaufen. **(d)** 32 cm Stufenhöhe.

Abbildung 7 Tests zur Überprüfung der Mobilität des neuen Bodenrobotersystems.

Tabelle 1 Spezifikationen der Robotikplattform.

Spezifikation	*Wert*
Kompakte Betriebsgröße	0,71 m Länge, 0,518 m Breite, 0,692 m Höhe
Maximale Betriebsgröße	1,21 m Länge, 0,518 m Breite, 0,692 m Höhe
Gesamtgewicht	57,3 kg
Maximale Steigung	45°
Maximale Stufenhöhe	30 cm
Höchstgeschwindigkeit (Ebene)	1,1 $m\,s^{-1}$
Manipulatorreichweite	0,85 m
Manipulatornutzlast	2,5 kg
Stromquelle	LiPo-Akkumulatoren (24 V&48 V, 377,4 W h)
Typ. Laufzeit	1,5 h
Bedienmodi	Semi- & Voll-Autonomie, Teleoperation

te autonome Assistenzfunktionen erfüllt. Die entwickelte mechatronische Plattform bietet dank der autonomiefokussierten Ausrichtung optimale Voraussetzungen für die Entwicklung komplexer Assistenz- und KI-Funktionen zur Bedienerentlastung. Basierend auf künftigen Evaluationen in Rettungsrobotikwettbewerben [14] sollen Hard- und Software fortlaufend iterativ verbessert werden.

5 Danksagung

Diese Arbeit wurde teilweise durch die LOEWE Initiative des Landes Hessen im Rahmen des LOEWE-Zentrums emergenCITY gefördert. Teile dieses Beitrags wurden durch Mittel des Bundesministeriums für Bildung und Forschung (BMBF) im Rahmen des Teilvorhabens „Autonome Assistenzfunktionen für Bodenroboter" des Verbundprojekts A-DRZ (Förderkennzeichen 13N14861) gefördert.

Literatur

[1] J. Delmerico u. a. „The current state and future outlook of rescue robotics". In: *Journal of Field Robotics* 36.7 (2019), S. 1171–1191.

[2] J. Casper u. a. „Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center". In: *IEEE Transactions on Systems, Man, and Cybernetics* 33.3 (2003), S. 367–385.

[3] K. Nagatani u. a. „Redesign of rescue mobile robot Quince". In: *2011 IEEE Int. Symp. on Safety, Security, and Rescue Robotics*. 2011, S. 13–18.

[4] C. Sampedro u. a. „A fully-autonomous aerial robot for search and rescue applications in indoor environments using learning-based techniques". In: *J Intell Robot Syst* 95.2 (2019), S. 601–627.

[5] R. Edlinger u. a. „MARC - Modular Autonomous Adaptable Robot Concept". In: *2019 IEEE Int. Symp. on Safety, Security, and Rescue Robotics (SSRR)*. 2019, S. 1–7.

[6] T. Takemori u. a. „Development of the multifunctional rescue robot FUHGA2 and evaluation at the world robot summit 2018". In: *Advanced Robotics* 34.2 (2020), S. 119–131.

[7] S. Habibian u. a. „Design and implementation of a maxi-sized mobile robot (Karo) for rescue missions". In: *ROBOMECH Journal* 8.1 (Jan. 2021).

[8] F. E. Schneider u. a. „Using robots for firefighters and first responders: Scenario specification and exemplary system description". In: *2017 18th Int. Carpathian Control Conference (ICCC)*. 2017, S. 216–221.

[9] R. R. Murphy. „Human-robot interaction in rescue robotics". In: *IEEE Transactions on Systems, Man, and Cybernetics* 34.2 (2004), S. 138–153.

[10] S. Kohlbrecher u. a. „Hector open source modules for autonomous mapping and navigation with rescue robots". In: *Robot Soccer World Cup*. Springer. 2013, S. 624–631.

[11] S. Kohlbrecher u. a. *RoboCup Rescue 2019 Team Description Paper Hector Darmstadt*. Techn. Ber. Technical University of Darmstadt.

[12] M. Quigley u. a. „ROS: an open-source Robot Operating System". In: *ICRA workshop on open source software*. 2009.

[13] K. Daun u. a. „Large Scale 2D Laser SLAM using Truncated Signed Distance Functions". In: *2019 IEEE Int. Symp. on Safety, Security, and Rescue Robotics (SSRR)*. 2019, S. 222–228.

[14] J. Pellenz u. a. „Novel rule set for the robocup rescue robot league". In: *2016 IEEE Int. Symp. on Safety, Security, and Rescue Robotics (SSRR)*. IEEE. 2016, S. 282–283.

Nutzen von Zusatzsensorik für die Unterscheidbarkeit von Modellen für elektrische Antriebsstränge bei der Struktur- und Parameteridentifikation

Benefit of extra sensors for distinguishability of models of electric power trains in structure und parameter identification

M. Sc. Mathias Tantau; Dr.-Ing. Mark Wielitzka, Institut für Mechatronische Systeme, Leibniz Universität Hannover, An der Universität 1, 30823 Garbsen, Deutschland, mathias.tantau@imes.uni-hannover.de

Dipl.-Ing. Lars Perner, Lenze SE, Am Alten Bahnhof 11, 38122 Braunschweig, Deutschland, lars.perner@lenze.com

Kurzfassung

Für viele Fragestellungen aus Reglerauslegung, Vorsteuerung und Zustandsüberwachung werden Prozessmodelle mit korrekter und physikalisch interpretierbarer innerer Struktur benötigt (phenomenologische Modelle). Die modellbasierten Ansätze werden in der Industrie vielfach noch nicht angewandt, weil die Modellerstellung ein hohes Maß an Expertenwissen und die langwierige Programmierung von Experimenten erfordert. Eine automatischen Struktur- und Parameteridentifikation ist dadurch bebrenzt, dass anhand des Ein-/Ausgangsverhaltens häufig die Unterscheidbarkeit von Modellen nicht gegeben ist. In dieser Veröffentlichung liegt der Fokus auf industriellen Anlagen mit elektrischem Antriebsstrang und einfacher Kinematik wie Regalbediengeräten, Werkzeugmaschinen und Positionierantrieben. Diese Systeme haben häufig nur einen Positions- und einen Stromsensor. Es wird in Experimenten gezeigt, dass durch Hinzunahme von einfach zu installierenden Zusatzsensoren wie Beschleunigungssensoren oder Drehratensensoren in einigen Fällen eine eindeutige Strukturidentifikation ermöglicht wird, auch wenn nur wenig Vorwissen über den Sensorort vorliegt.

Abstract

For many questions related to controller design, feed-forward and condition monitoring, process models with correct and physically meaningful internal structure are required (phenomenological models). But model-based approaches have not yet arrived ubiquitously in industry because the creation of models requires a high degree of expert knowledge and the lengthy programming of experiments. An automatic structure and parameter identification is limited by the fact that based on the input / output behavior, it is often not possible to distinguish between models. In this publication, the focus lies on industrial systems with electrical drive trains and simple kinematics such as storage and retrieval units, machine tools and positioning stages. Often, these systems have only one position and one current sensor. It is shown that by adding easily integrable, additional sensors, such as acceleration sensors or gyroscopes, a unique structure identification can be achieved in many cases, even if only little prior knowledge of the sensor location is available.

1 Einleitung

Für viele Problemstellungen aus den Bereichen Reglerauslegung, Vorsteuerung und Zustandsüberwachung werden Prozessmodelle mit bekannter innerer Struktur und physikalisch interpretierbaren Parametern benötigt (phenomenologische Modelle) [9, 17]. Modellbasierte Ansätze sind in der Industrie jedoch noch nicht sehr verbreitet, weil die Modellerstellung viel Expertenwissen und zumeist eine langwierige Entwicklung von Identifikationsexperimenten erfordert.

Vereinzelte Ansätze zur automatischen Modellerstellung durch Struktur- und Parameteridentifikation existieren [2, 8, 12], welche mehrere Modelle identifizieren und im Anschluss das Modell mit dem geringsten Residuum wählen, zzgl. eventuell weiterer Kriterien. Jedoch ist die Unterscheidbarkeit von Modellstrukturen anhand des Ein-/Ausgangsverhaltens häufig nicht gegeben, gerade bei elektrischen Antriebssträngen mit nur einem Positions- und einem Stromsensor, korrespondierend zum Drehmoment. Das Modell, das die beste Übereinstimmung zeigt, muss nicht physikalisch korrekt sein. Dadurch erzielt z.B. die flachheitsbasierte Vorsteuerung nicht die gewünschte Genauigkeitsverbesserung am Endeffektor oder der Zweimassenschwinger-Regler optimiert nicht das Einschwingverhalten des Endeffektors, sondern das einer unwichtigen Strukturstelle. Auch eine Lokalisation von Schwachstellen in der Struktur, die zu schlechtem regelungstechnischem Verhalten führen, ist nicht möglich. Ohne weitere Untersuchungen ist eine existierende Mehrdeutigkeit eventuell nicht einmal als solche bekannt.

Der Struktur- und Parameteridentifikation sollte daher eine Untersuchung der strukturellen Unterscheidbarkeit auf Basis der symbolischen Definition des Systemverhaltens vorangehen. Dabei bedeutet die Aussage, dass Modelle $M(\cdot)$ und $\hat{M}(\cdot)$ mit den Parametervektoren $\boldsymbol{p}_0 \in \Omega$ und $\hat{\boldsymbol{p}}_0 \in \Omega$ *strukturell unterscheidbar* sind folgendes: Für fast keine Parametrierung $\boldsymbol{p}_0$ existiert eine Parametrierung $\hat{\boldsymbol{p}}_0$, sodass beide Modelle das gleiche Ausgangsverhalten zeigen und umgekehrt [13]. Nur für eine begrenzte Menge

von Parametrierungen $\boldsymbol{p}_0$, die keinen Raum im Parameterraum einnimmt, darf es Ausnahmen geben, z.B Symmetrien und Pol-/Nullstellenkompensation. Im Folgenden wird der Begriff *strukturell ununterscheidbar* als Gegenteil verwendet, obwohl differenziertere Definitionen existieren [1, 3, 10, 13, 15].
Methoden zur Untersuchung der Unterscheidbarkeit benötigen eine symbolische Beschreibung beider dynamischer Systeme, genannt Strukturinvariantenvektor [13] oder *exhaustive summary* [15], welche dann mithilfe von verschieden Kriterien auf Unterscheidbarkeit untersucht werden [4–6, 19]. Daneben gibt es den Ansatz über die Ähnlichkeitstransformation von linearen Modellen [1], doch dieser wird hier wegen des hohen Rechenaufwandes nicht weiter verfolgt.
Bisherige Arbeiten beschränken sich hauptsächlich auf biomedizinische Modelle, die den Übergang von Stoffen zwischen verschiedenen Kompartimenten beschreiben [3, 5, 7, 16, 18, 19]. Der Beitrag dieser Veröffentlichung ist eine Untersuchung zum Nutzen von Zusatzsensorik (Beschleunigungs- oder Drehratensensoren) in Bezug auf die strukturelle Unterscheidbarkeit von Mehrmassenschwingermodellen, wie sie zur Darstellung von elektrischen Antriebssträngen in Frage kommen. Die Zusatzsensoren können mit geringem Aufwand an unterschiedlichen Stellen der Struktur temporär für die Modellbildung und Inbetriebnahme montiert werden. Die Besonderheit ist, dass vor der Analyse noch nicht die genaue Zuordnung von konzentrierten Massen des Modells zu physikalischen Bauteilen des Aufbaus bekannt ist und so auch nur wenig Vorwissen über die Positionen der Zusatzsensoren in Bezug auf die Modelle vorausgesetzt werden kann.

2 Unterscheidbarkeit mit Zusatzsensorik

In diesem Abschnitt wird die Methodik zur Untersuchung der strukturellen Unterscheidbarkeit vorgestellt und auf die Einbeziehung der Zusatzsensorik eingegangen. Dazu sind kombinatorische Überlegungen zu Mehrmassenschwingermodellen erforderlich, wie sie für die Modellierung der Antriebsstränge mit angekoppelter Mechanik verwendet werden.

2.1 Strukturinvarianten

Der Strukturinvarianten-Vektor $\boldsymbol{\Phi}$ zur Definition des dynamischen Systemverhaltens kann ohne Berücksichtigung der Anfangsgrößen über die Übertragungsfunktionsmatrix $\boldsymbol{G}$ des MIMO-Systems definiert werden. Der Eintrag jl hat die allgemeine Form:

$$G_{jl}(s) = \frac{b_0 + b_1 s^1 + \cdots + b_{n-1} s^{n-1}}{a_0 + a_1 s^1 + \cdots + a_{m-1} s^{m-1} + s^m}. \tag{1}$$

Die Koeffizienten sind analytische Funktionen der Modellparameter $\boldsymbol{p}$. Sie definieren die Menge der Strukturinvarianten, wenn einer der Koeffizienten fest ist, beispielsweise der letzte Nennerkoeffizient.
Bei MIMO Übertragungsfunktionen, die aus der linearen Zustandsraumdarstellung berechnet wurden, ist bekannt, dass der Nenner für alle Ein-/Ausgabepaare gleich ist. Folglich müssen die Nennerkoeffizienten nicht mehrmals ausgewertet werden, sodass die Anzahl der Strukturinvarianten unterproportional mit der Anzahl der Übertragungsfunktionselemente zunimmt [4].
Zusätzlich können die Markov-Parameter eines dynamischen Systems als Strukturinvarianten verwendet werden. Davon wird hier jedoch abgesehen, weil der Rechenaufwand hoch ist und nur in wenigen Fällen ein Zusatznutzen entsteht [12].

2.2 Unterscheidbarkeitskriterien

Die folgenden notwendigen und hinreichenden Kriterien für Ununterscheidbarkeit basierend auf den Koeffizienten als Strukturinvarianten helfen zu bestimmen, ob eine Lösung existiert, ohne direkt nach Unbekannten aufzulösen. Dies würde einen noch höheren Rechenaufwand bedeuten und häufig nicht zu einer expliziten Lösung führen. Bei Verwendung der Markov-Parameter wären teilweise leichte Abwandlungen erforderlich.
Notwendige Bedingungen:

- Bed. 1: Die gleiche Anzahl von Momentinvarianten N_{MI} existiert.
- Bed. 2: In Anzahl der Zähler- und Nennerkoeffizienten stimmt überein.
- Bed. 3: Beide Sätze von Strukturinvarianten haben dieselbe symbolische Form. Die symbolische Form ist eine Repräsentation mit Nullen für konstante Elemente und Einsen für nicht-konstante (abhängig von $\boldsymbol{p}$ bzw. $\hat{\boldsymbol{p}}$) Elementen [14].
- Bed. 4: Der Rang N_{R} der Jacobi-Matrix $\boldsymbol{J} = \partial\boldsymbol{\Phi}/\partial\boldsymbol{p}$ ist gleich [19].
- Bed. 5: Es existieren die gleichen linearen Abhängigkeiten zwischen Strukturinvarianten [19]. Diese werden bestimmt, indem alle möglichen Kombinationen von N_{R} Zeilen von $\boldsymbol{J}$ erstellt werden. Die Kombinationen von Zeilen, die in symbolischer Form zum vollen Zeilenrang führen, werden notiert und müssen für die verglichenen Systeme identisch sein [4].

Die Bedingungen sollten in der angegebenen Reihenfolge ausgewertet werden, da sie nach zunehmender Rechenkomplexität sortiert sind und Abhängigkeiten existieren.
Hinreichende Bedingungen:

- Bed. 6: Die Übertragungsfunktionen sind nach Umbenennung und Permutation der Variablen in symbolischer Form identisch. Dies kann dann hilfreich sein, wenn Systeme spiegelsymmetrisch sind o.Ä.
- Bed. 7: In beiden Modellen entspricht die Anzahl der bestimmbaren Parameter N_{R} gerade der Anzahl nicht-konstanter Strukturinvarianten und die konstanten Strukturinvarianten stimmen überein [14].

Genau genommen berücksichtigen diese hinreichenden Bedingungen nicht den genauen Bereich der Parameter. Es kann vorkommen, dass zwei Modelle gleichwertig sind, die entsprechenden Lösungen für die Parameter jedoch unzulässig, z.B. negativ sind.

2.3 Kombinatorik der Mehrmassenschwingermodelle

Bei den Unterscheidbarkeitsuntersuchungen liegt der Fokus in dieser Arbeit auf linearen Mehrmassenschwingermodellen, die die strukturelastischen Eigenschaften des elektrischen Antriebsstrangs repräsentieren sollen. Bei gegebener Massenzahl existieren mehrere mögliche Ausprägungen des Mehrmassenschwingers, je nach genauer Anordnung der Elastiziäten, des Aktors und Sensors. Die Wahl des richtigen Modells entspricht der Frage, an welchen Stellen des Systems Elastizitäten und Massen berücksichtigt werden müssen, bzw. vernachlässigt werden können. Die kombinatorischen Überlegungen beschränken sich zunächst auf den Fall, dass es genau einen Aktor und einen Sensor gibt, weil dies den industriell häufigsten Fall darstellt. Danach werden weitere Sensoren hinzugenommen, sodass der Nutzen der zusätzlichen Sensoren untersucht werden kann.
In **Abb.** 1 ist ein translatorisches Beispiel für eine Dreimassenschwinger (DMS)-Kette gegeben; für rein rotatorische Systeme gelten die gleichen kombinatorischen Überlegungen. Vor der linken Masse fehlt das Feder-/Dämpferelement, was gleichsam die Führung, also den Freiheitsgrad des Systems darstellt. Dieser muss bei Antriebssträngen immer vorhanden sein, seine Position kann aber variieren. Aktor und Sensor überspannen diese Lücke.

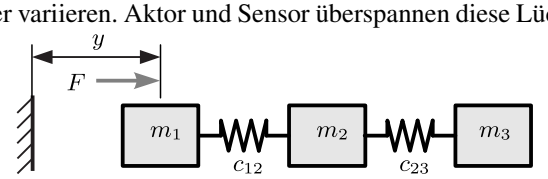


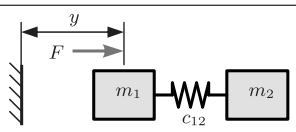
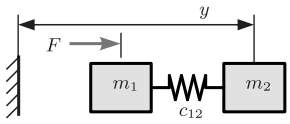
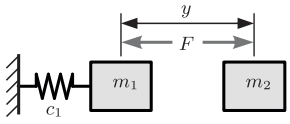
Abbildung 1 DMS Modell DMS_A1

In [11] werden die resultierenden Kombinatoren für den Fall linearer Ketten ohne Schleifen oder Verzweigungen allgemein hergeleitet. Hier werden nur die drei translatorischen Zweimassenschwinger (ZMS)-Modelle mit einem Sensor gezeigt (**Tabelle** 1). Bei einer Masse gibt es nur eine Ausprägung und beim Dreimassenschwinger ergeben sich elf Modelle.

2.4 Berücksichtigung von Zusatzsensorik

Für die Zusatzsensoren wird angenommen, dass diese stets die absolute Geschwindigkeit oder Beschleunigung einer Masse messen und keine Differenzen, denn dies entspricht der Verwendung von einfach erhältlichen MEMS Sensoren. Durch die zusätzliche Sensorik vervielfachen sich die kombinatorischen Möglichkeiten und das Problem der Strukturidentifikation wird eher schwieriger als einfacher. Daher muss Vorwissen mit berücksichtigt werden. Andererseits kann nicht davon ausgegangen werden, dass die Zuordnung von Strukturelementen zu Massen des Modells a priori bekannt ist. Untersucht werden sollen daher wenn

Tabelle 1 Kompletter Satz von zu berücksichtigenden ZMS-Modellen. Kollokierte Systeme sind mit einem Sternchen * gekennzeichnet.

Name	Skizze
ZMS_A1*	
ZMS_A2	
ZMS_B1*	

möglich die drei folgenden Fragestellungen, die ein gewisses, aber realistisches Maß an Vorwissen voraussetzen:

I Können Ununterscheidbarkeitsprobleme durch zusätzliche Sensoren am linken und / oder rechten Kraftangriffspunkt behoben werden? Es sollte möglich sein, den Ort auszumachen, wo der Aktor angreift. Dazu muss das Mehrmassenschwingerverhalten des Systems noch nicht vollständig verstanden sein.

II Können Ununterscheidbarkeitsprobleme durch einen zusätzlichen Sensor an einem komplett unbekannten Ort auf einer der Massen behoben werden? Hierbei wird wenig Vorwissen vorausgesetzt, jedoch muss gewährleistet sein, dass der Sensor wirklich auf einer Masse des Modells sitzt und nicht beispielsweise auf einer Feder zwischen zwei Massen.

III Kann mit einem bekannten Modell die Masse bestimmt werden, auf der sich ein zusätzlicher Sensor befindet? Nun geht also darum, die Zuordnung zwischen einem Bauteil und der zugehörigen Masse des Mehrmassenschwingers herzustellen.

Bei der Durchführung der Strukturidentifikation mit mehreren Sensoren mit verschiedenen physikalischen Eigenschaften bietet es sich an, zunächst ausschließlich den Positionssensor im Motor zur Identifikation der Parameter zu berücksichtigen und anschließend bei Mehrdeutigkeiten das Modell mit der besten Übereinstimmung zu den Beschleunigungssensor-Signale zu bestimmen. Alternativ kann von vornherein das Residuum über alle Sensorsignale minimiert werden. Um dabei den Vergleich von unterschiedlichen physikalischen Einheiten zu vermeiden, werden Phase und logarithmische Amplitude separat berechnet:

$$J = \left|\left| \mathbf{\Lambda}_{\mathrm{A/P}} \begin{bmatrix} \log_{10}\{|\boldsymbol{G}(\boldsymbol{p})|\} - \log_{10}\{|\boldsymbol{G}_{\mathrm{M}}|\} \\ \angle\{\boldsymbol{G}(\boldsymbol{p})\} - \angle\{\boldsymbol{G}_{\mathrm{M}}\} \end{bmatrix} \right|\right|_2^2. \quad (2)$$

Unterschiede in der Größenordnung von Beschleunigung und Position spielen so eine geringere Rolle.
Vor der Durchführung der Optimierungen sollte in jedem Fall mithilfe der Unterscheidbarkeitskriterien geprüft wer-

den, ob eine eindeutige Lösung gegeben ist.

3 Versuchsdurchführung

Im Folgenden soll der Nutzen der Strukturidentifikation und insb. der Zusatzsensorik an einem Prüfstand untersucht werden.
Der verwendete Prüfstand ist das Regalbediengerät aus **Abb.** 2. Es verfügt über drei orthogonale Achsen, die jeweils mit einem Getriebemotor der Firma Lenze mit Riemenantrieb verfahren werden. Der Mast hat eine Höhe von 5,6 m, korrespondierend zur y-Achse und die x-Achse entlang dem Regal ist 5 m lang. Die z-Achse fährt in das Regal hinein. In dieser Veröffentlichung sollen die x- und die z-Achse untersucht werden. Die Antriebssysteme sind mit Resolver und Stromsensor ausgestattet, sodass die ungefähre Übertragungsfunktion von Antriebsmoment zu Motorposition mit Seriensensorik bestimmt werden kann.

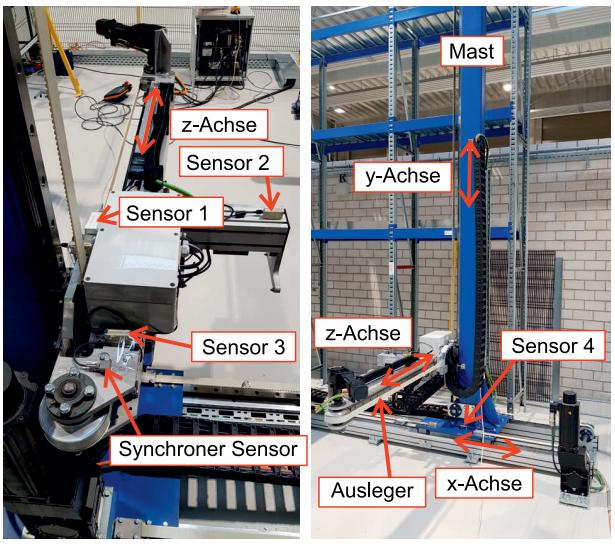


Abbildung 2 Prüfstand: Regalbediengerät

Zusätzlich sind batteriebetriebene, funkbasierte Beschleunigungs- und Drehratensensoren an unterschiedlichen Stellen montiert, siehe auch **Abb.** 3. Diese basieren auf dem Arduino Board ESP32 NodeMCU, sowie dem MEMS-Sensor LSM6DS3 von STMicroelectronics. Die Daten werden asynchron per WLAN an ein Empfangsmodul übertragen. Die FreeRTOS Implementierung erlaubt das gleichzeitige Messen von etwa sechs Beschleunigungssignalen mit 1 kHz Abtastrate in drei Raumrichtungen. Zur Synchronisation der Signale der MEMS-Sensoren mit den Motorsignalen dient der synchron abgetastete Beschleunigungssensor, dessen Daten über AD-Wandler zusammen mit den Motorsignalen aufgezeichnet werden. Die zeitliche Synchronisation erfolgt durch Maximierung der Kreuzkorrelation mit Sensor 3. Trotzdem wird die Phaseninformation der Beschleunigungs- und Drehratensensoren nicht verwendet, nur die Amplitude.

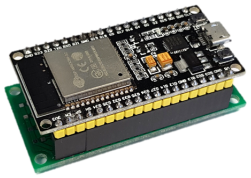


Abbildung 3 Mobiler Zusatzsensor auf Basis des Arduino Boards ESP32 NodeMCU

4 Versuchsergebnisse

In diesem Abschnitt sollen die genauen zu untersuchenden Fragestellungen erläutert und die Ergebnisse gezeigt werden. Dabei ist vor jedem Identifikationsschritt zu testen ob die Unterscheidbarkeit der Modelle theoretisch gegeben ist. Die drei Fragestellungen aus Abschnitt 2.4 dienen als Anhaltspunkt, wenn die Unterscheidbarkeit ohne Zusatzsensorik nicht gegeben ist. Es sind jedoch Anpassungen erforderlich, weil der Aufbau des Prüfstands nicht genau den translatorischen Grundmodellen, siehe Abb. 1 und Tabelle 1 entspricht.
Beginnend soll für die x-Achse das korrekte Modell bestimmt werden. Diese Achse ist durch einen rotatorischen Aktor mit umlaufendem Riemen angetrieben ist. Die Festlegung der Modellordnung (ZMS oder DMS) ist nicht der Schwerpunkt dieser Veröffentlichung, aber es ist anhand der Messung in **Abb.** 4 zu erkennen, dass ein DMS angemessen ist.

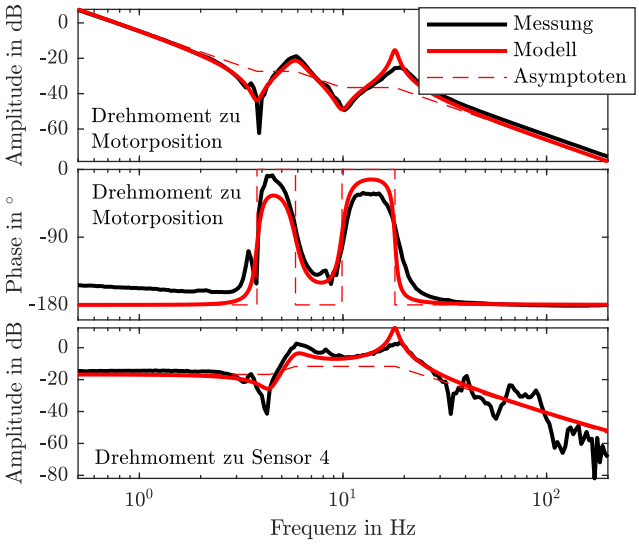
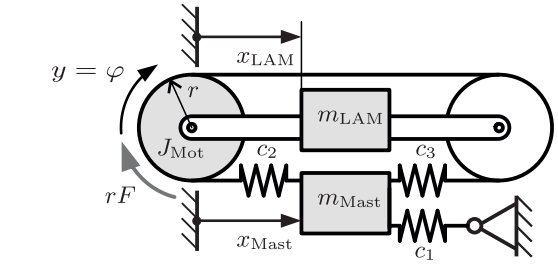


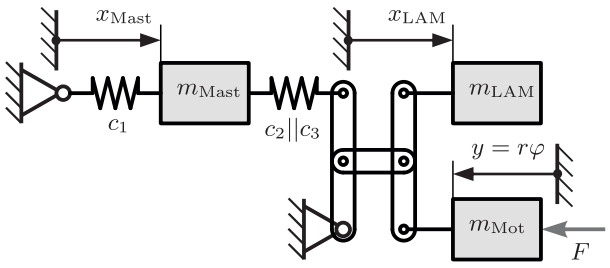
Abbildung 4 x-Achse: gemessene Frequenzgänge, inklusive Zusatzsensor vs. resultierendes Modell DMS_A1

Wird nun die Strukturidentifikation ohne Zusatzsensorik durchgeführt, ergibt sich Modell DMS_A1 als Lösung, siehe Abb. 1, aber vier weitere kollokierte Modelle kommen ebenso infrage, weil sie nicht unterscheidbar sind, abgesehen von teilweise leichten Unterschieden bei Berücksichtigung von Dämpfern parallel zu den Federn. Doch der Unterschied durch Dämpfung ist klein und die Verlässlichkeit der Strukturidentifikation somit begrenzt. Um diese Mehrdeutigkeit aufzulösen, werden die obigen Fragen untersucht. Frage I ist nicht direkt anwendbar, denn an dem bewegten Teil des Aktors (Rotor des Motors) wäre die Anbringung eines Beschleunigungssensors nicht sinnvoll

und ein Drehratensensor ist aufgrund des Bauraums auch ausgeschlossen. Frage II kann jedoch umgesetzt werden. Sensor 4 am Mastfuß wird zusätzlich bei der Parameteridentifikation der Modelle verwendet. Es ergibt sich wieder Modell DMS_A1 und nur ein zweites ununterscheidbares Modell für den Fall ohne Dämpfung (nicht dargestellt). Unter Berücksichtigung von Dämpfung kommt nur noch DMS_A1 in Frage. Der Modell-Frequenzgang inklusive Sensors 4 ist ebenfalls in der Abbildung zu sehen (rot). Für das nächstbeste Modell ergäbe sich ein etwas höheres Residuum (321 vs. 272). Der Sensor hat also einen Nutzen erbracht und die Auswahl weiter eingeschränkt. Wird Modell DMS_A1 als korrektes Modell vorausgesetzt, ergibt sich zusätzlich gemäß Frage 3 eindeutig, dass sich Sensor 4 auf der zweiten Masse befindet. Dies zeigt, dass Riemen und Mastschwingung beide zusammen den DMS definieren. Als nächstes werden die Sensoren auf dem Ausleger und weitere Sensoren am Mast ausgewertet, um die bestmögliche Zuordnung zu Masse 3 zu bestimmen. Jedoch ist dieses Mal auch die Position mit dem kleinsten Residuum nicht in guter Übereinstimmung mit dem berechneten Frequenzgang. Die Zuordnung der drei Massen zu Bauteilen war also aufgrund der Komplexität des Systems nur für die erste Masse (Rotor des Motors) und die zweite Masse (Mastfuß), jedoch nicht für die dritte Masse möglich.
Als nächstes wird die z-Achse untersucht, deren erwartetes Modell in **Abb.** 5, a) gezeigt ist. Es handelt sich um einen umlaufenden Riemen mit mitfahrendem Motor. Federsteifigkeit c_1 repräsentiert den Mast, die anderen Federsteifigkeiten den Riemen. Der Frequenzgang ist in **Abb.** 6 gezeigt. Aufgrund des rotatorischen Aktors und der hohen Modellkomplexität kann die beschriebene Methodik nicht direkt angewandt werden. Auch das translatorische Ersatzmodell, siehe Abb. 5, b) ist zu komplex.



(a) Realitätsnahes Modell



(b) Translatorisches Ersatzmodell

Abbildung 5 Modell der z-Achse. $c_2||c_3$ steht für die Parallelschaltung der Federn.

Zur weiteren Analyse wird das translatorische DMS-Modell daher auf verschiedene Weise auf ZMS-Modelle

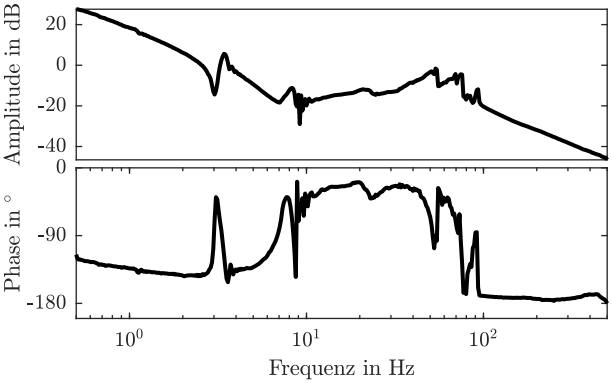
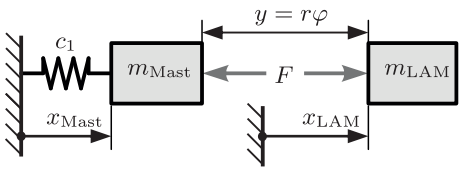
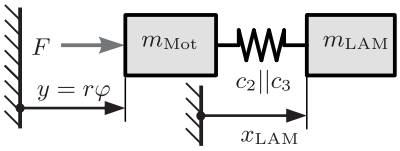


Abbildung 6 z-Achse: gemessener Frequenzgang von Drehmoment zu Motorposition

reduziert, siehe **Abb.** 7. Im ersten Fall werden J_{Mot} und c_2, c_3 vernachlässigt, im zweiten Fall c_1 und somit auch m_{Mast}. Es entstehen zwei mögliche ZMS-Modelle, die in Tabelle 1 als ZMS_A1, ZMS_B1 wiederzufinden sind. Mit Seriensensorik sind diese nicht unterscheidbar. Doch mit Zusatzsensorik kann im Sinne von Frage II eine Strukturidentifikation für die drei ZMS Modelle aus Tabelle 1 durchgeführt werden.



(a) J_{Mot} und c_2, c_3 vernachlässigt



(b) c_1 vernachlässigt und somit auch m_{Mast}

Abbildung 7 Zwei mögliche Modellvereinfachungen

Dabei kommt Sensor 1 zum Einsatz, dessen Befestigungsort den Fußpunkt der z-Achse darstellt, siehe Messung in **Abb.** 8. Der gemessene Frequenzgang wird im Sinne der Modellvereinfachung auf ZMS nur bis 7 Hz ausgewertet. Jenseits von 7 Hz würden sich Dynamiken höherer Ordnung bemerkbar machen, insb. die Riemenschwingung und die Lose im Antrieb. Das Residuum wird gemäß (2) über beide Sensoren minimiert.
Die resultierenden Residuen der drei Modelle und der zwei verschiedenen angenommenen Positionen von Sensor 1 sind in **Tabelle** 2 angegeben. Erwartungsgemäß ergibt sich das höchste Residuum bei dem nichtkollokierten Modell ZMS_A2. Interessant ist, dass Modell ZMS_B1 bei weitem die beste Übereinstimmung zeigt. Es hat sich somit als Ergebnis der Untersuchung ergeben, dass es sich bei der z-Achse in der ZMS Näherung um eine elastische Basis handelt und der Riemen sich erst bei höheren Frequenzen bemerkbar macht. Der Frequenzgang von Modell ZMS_B1 ist ebenfalls in Abb. 8 zu sehen.

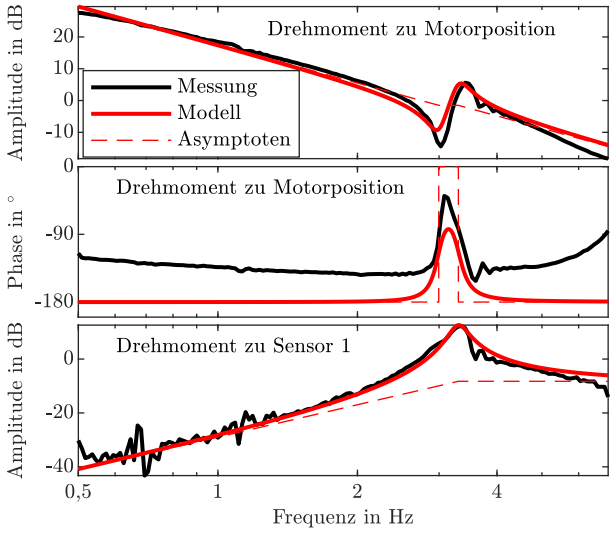


Abbildung 8 z-Achse: gemessene Frequenzgänge vs. resultierendes Modell ZMS_B1

Tabelle 2 Residuen der drei ZMS-Modelle in Abhängigkeit der Sensorposition

Masse	*ZMS_A1*	*ZMS_A2*	*ZMS_B1*
1	1528	4756	145
2	1563	12342	1456

5 Schlussfolgerung

Um den Nutzen von Zusatzsensorik für die eindeutige Struktur- und Parameteridentifikation zu zeigen, wurden Versuche mit mehreren funkbasierten Sensoren an zwei Achsen eines Regalbediengerätes mit industrieller Hardware durchgeführt. Die Notwendigkeit und Sinnhaftigkeit der einzelnen Teilexperimente wurde mithilfe von symbolischen Untersuchungen zur Unterscheidbarkeit von Strukturen sichergestellt, wobei aufgrund der speziellen Systemstruktur leichte Abweichungen vom Standardschema erforderlich waren.

Es hat sich gezeigt, dass die häufigen Mehrdeutigkeitsprobleme durch die Zusatzsensorik theoretisch eingedämmt werden können und dass auch in den konkret untersuchten Fallstudien ein besseres Systemverständnis für dominierende, niederfrequente Elastizitäten generiert wird. Für die Identifikation von komplexeren Modellen sind allerdings noch Anpassungen der hier verwendeten generischen Menge an Modellkandidaten nötig.

6 Literatur

[1] TV Avdeenko und SA Kargin: *A computer algebra method for testing structural distinguishability of state space models*. In: *5th Korea-Russia International Symposium on Science and Technology (KORUS)*, Band 1, Seiten 77–80. IEEE, 2001.

[2] K. P. Burnham und D. R. Anderson: *Model selection and inference - a practical use of the information-theoretic approach*. Springer, New York, 1998.

[3] M. J. Chapman und K. R. Godfrey: *A methodology for compartmental model indistinguishability*. Mathematical biosciences, 96(2):141–164, 1989.

[4] N. R. Davidson, K. R. Godfrey, F. Alquaddoomi, D. Nola und J. J. DiStefano III: *DISTING: A web application for fast algorithmic computation of alternative indistinguishable linear compartmental models*. Computer Methods and Programs in Biomedicine, 143:129–135, 2017.

[5] N. D. Evans, M. J. Chappell, M. J. Chapman und K. R. Godfrey: *Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems*. Automatica, 40(11):1947–1953, 2004.

[6] K. R. Godfrey und M. J. Chapman: *Identifiability and indistinguishability of linear compartmental models*. Mathematics and Computers in Simulation, 32(3):273–295, 1990.

[7] K. R. Godfrey, M. J. Chapman und S. Vajda: *Identifiability and indistinguishability of nonlinear pharmacokinetic models*. Journal of pharmacokinetics and biopharmaceutics, 22(3):229–251, 1994.

[8] G. J. Gray, Y. Li, D. J. Murray-Smith und K. C. Sharman: *Structural system identification using genetic programming and a block diagram oriented simulation tool*. Electronics Letters, 32(15):1422–1424, 1996.

[9] G. Massonis, J. R. Banga und A. F. Villaverde: *Repairing dynamic models: a method to obtain identifiable and observable reparameterizations with mechanistic insights*. arXiv preprint arXiv:2012.09826, 2020.

[10] K. M. D. Motchon, K. M. Pekpe, J. P. Cassar und S. De Bièvre: *On the input-output distinguishability of single output continuous linear time-invariant systems*. IEEE Transactions on Automatic Control, 61(7):1906–1911, 2015.

[11] M. Tantau, C. Helmke, L. Perner und M. Wielitzka: *Distinguishability study of 3-Mass Models for Electromechanical Motion Systems*. International Journal of Modelling, Identification and Control, 2020. Im Druck.

[12] M. Tantau, L. Perner und M. Wielitzka: *Model Selection for Servo Control Systems*. International Journal of Mechatronics and Automation, 2020. Im Druck.

[13] S. Vajda: *Structural equivalence of linear systems and compartmental models*. Mathematical Biosciences, 55(1-2):39–64, 1981.

[14] S. Vajda: *Structural equivalence and exhaustive compartmental modeling*. Mathematical biosciences, 69(1):57–75, 1984.

[15] E. Walter, Y. Lecourtier und J. Happel: *On the structural output distinguishability of parametric models, and its relations with structural identifiability*. IEEE Transactions on Automatic Control, 29(1):56–57, 1984.

[16] E. Walter und L. Pronzato: *On the identifiability and distinguishability of nonlinear parametric models*. Mathematics and computers in simulation, 42(2-3):125–134, 1996.

[17] M. Witczak, A. Obuchowicz und J. Korbicz: *Genetic programming based approaches to identification and fault diagnosis of non-linear dynamic systems*. International Journal of Control, 75(13):1012–1031, 2002.

[18] J. W. T. Yates, R. D. O. Jones, M. Walker und S. Y. A. Cheung: *Structural identifiability and indistinguishability of compartmental models*. Expert opinion on drug metabolism & toxicology, 5(3):295–302, 2009.

[19] L. Q. Zhang, J. C. Collins und P. H. King: *Indistinguishability and identifiability analysis of linear compartmental models*. Mathematical biosciences, 103(1):77–95, 1991.

Mechatronik: Risiken und Herausforderungen bei der Entwicklung eines permanentmagnetisch vormagnetisierten, aktiven Magnetlagers
Mechatronics: Risks and challenges in the development of a permanent-magnet-biased active magnetic bearing

Timo Hopf*, timo.hopf@tu-darmstadt.de
Prof. Dr.-Ing. Stephan Rinderknecht*, rinderknecht@ims.tu-darmstadt.de
*TU Darmstadt, Institut für Mechatronische Systeme, 64287 Darmstadt, Deutschland.

Kurzfassung

Durch die vermehrte Verwendung erneuerbarer Energieträger hat die Energiespeicherung im Stromnetz an Bedeutung gewonnen. Hier können Kinetische Energiespeicher (KES) zur Stabilisierung eingesetzt werden. Bei KES in Außenläuferbauform ist das Schwungrad als Hohlzylinder ausgeführt, wodurch der KES im Vergleich zu einem Innenläufer eine höhere Energiedichte aufweist. Um die Energieeffizienz des prototypischen KES ETA-290 zu steigern, wird ein verlustarmes, permanentmagnetisch vormagnetisiertes, aktives Magnetlager eingesetzt. Dessen Entwicklung wird im Folgenden beschrieben. Dabei werden zunächst Anforderungen an die aktiven Magnetlager definiert und Maßnahmen zur Verlustreduzierung erläutert. Anschließend werden die Herausforderungen und Risiken bei der Entwicklung einer permanentmagnetischen Vormagnetisierung diskutiert. Die Funktionsfähigkeit des entwickelten Lagers und eine zuvor numerisch abgeschätzte Effizienzsteigerung werden durch Messungen validiert.

Abstract

Due to the increased use of renewable energy sources, the storage of energy has gained importance. Kinetic energy storage (KES) can be used to stabilize the power grid. Outer rotor KES, in which the flywheel is designed as a hollow cylinder, have a higher energy density compared to internal rotor KES. To increase the energy efficiency of the prototypical KES ETA-290, a low-loss, permanent-magnet-biased active magnetic bearing is used. Its development is described below. At first, requirements for the active magnetic bearings are defined and opportunities for loss reduction are explained. Subsequently, challenges and risks in the development of a permanent-magnet-bias are discussed. The functionality of the developed active magnetic bearing and a numerically estimated efficiency increase are validated by measurements.

1 Einleitung

Um CO_2-Emissionen zu reduzieren, wird der Fokus der Energieversorgung immer mehr auf die Verwendung erneuerbarer Energieträger gelegt. Die dadurch gewonnene Energie sowie deren Nutzung sind häufig starken Schwankungen unterlegen. Die überschüssige Energie muss daher gespeichert und im Bedarfsfall wieder bereitgestellt werden [1]. Kinetische Energiespeicher (KES) eignen sich in diesem Kontext als dynamische Speicher zur Lastspitzenglättung. Die Hauptvorteile von KES gegenüber Lithium-Ionen-Batterien sind ihre hohe Lebensdauer und hohe Dynamik. Nachteilig sind ihre hohen Standby-Verluste sowie die vergleichsweise geringe Energiedichte [2]. Am Institut für Mechatronische Systeme (IMS) der TU Darmstadt wird an KES in Außenläuferbauform geforscht. Bei dieser speziellen Bauform dreht der Rotor in Form eines Hohlzylinders um den Stator, wodurch diese Systeme im Vergleich zu Innenläufern eine höhere Energiedichte aufweisen. Bisher wurden drei solcher Speicher am IMS entwickelt, aufgebaut und getestet [3]. Um die Wettbewerbsfähigkeit der Systeme zu erhöhen, muss die Gesamtenergieeffizienz gesteigert werden. Die Verluste des Systems setzen sich aus Verlusten der Motor-Generator-Einheit (MGE), der aktiven Magnetlager (engl. Active Magnetic Bearing, AMB), der Luftreibung sowie der Peripherie zusammen. Da die Speicher im Standby-Betrieb nicht stillstehen, können die AMB nicht abgeschaltet werden. Daher weist eine Verlustreduzierung in den AMB wichtiges Potential zur Effizienzsteigerung der Systeme auf. In diesem Artikel werden Maßnahmen zur Verlustreduktion der AMB diskutiert und die Herausforderungen der Entwicklung eines permanentmagnetisch vormagnetisierten aktiven Magnetlagers (PM-AMB) für den am IMS entwickelten KES ETA-290 aufgezeigt. Durch Messungen am Realsystem wird die Verlustreduktion der PM-AMB validiert.

2 Aufbau des ETA-290.

Der Aufbau des ETA-290 mit den wichtigsten Komponenten ist in Bild 1 anhand des CAD-Modells zu sehen.

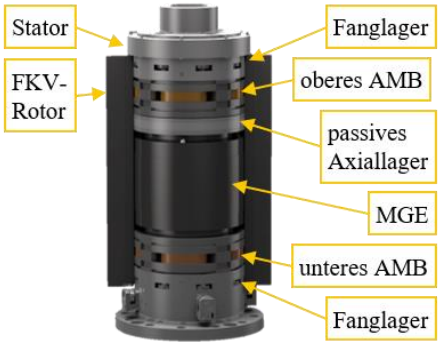


Bild 1 Teilschnittdarstellung des ETA-290 [angelehnt an 1].

Der Rotor besteht aus einem Wickelkörper aus Faserkunststoffverbund (FKV), der die funktionalen Einbauten der aktiven Komponenten umgibt. Die in den Rotor integrierten funktionalen Einbauten sind in Umfangsrichtung segmentiert ausgeführt, um die bei hohen Umfangsgeschwindigkeiten wirkenden Spannungen in den Einbauten zu reduzieren und ein Versagen des Materials zu verhindern. Gelagert wird der Rotor axial durch ein passives Magnetlager und radial durch zwei AMB. Als Rückfallebene im Falle des Versagens der AMB sind mechanische Fanglager verbaut. Die MGE ist als permanentmagnetische Synchronmaschine ausgeführt und dient der Energiewandlung. Der Betriebsbereich des ETA-290 liegt zwischen 7.500 U/min und 15.000 U/min, da bei niedrigeren Drehzahlen keine nennenswerte Leistungsstellung mehr möglich ist. Um die Luftreibungsverluste zu minimieren, wird das System im Vakuum betrieben. Aufgrund des Vakuums und der berührungslosen Lagerung kann die aus verschiedenen Verlustmechanismen entstehende Wärme im Rotor nur durch Wärmestrahlung abgeführt werden. Daher müssen die Verluste im Rotor minimiert werden, um ein Überhitzen des Rotors zu verhindern.
Im ursprünglichen Design sind die AMB als Homopolarlager in Differenzwicklung mit Biasspule ausgeführt und werden im Folgenden mit BS-AMB bezeichnet. In Bild 2 ist das CAD-Modell mit den funktionalen Einbauten des Rotors (Anker) zu sehen.

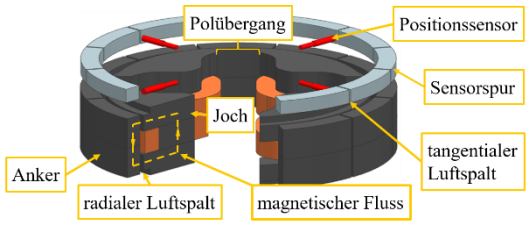


Bild 2 CAD-Modell des BS-AMB [angelehnt an 1].

Das Joch und der Anker im Rotor leiten den magnetischen Fluss und sind aus einem weichmagnetischen Verbundwerkstoff (engl. Soft Magnetic Composite, SMC) gefertigt. Zur Erfassung der Rotorposition werden pro Lager vier induktive Abstandssensoren verwendet.

3 Anforderungen an die AMB

Bei der Überarbeitung des AMB müssen, um den Rotor im gesamten Betriebsbereich des KES stabil zu lagern, mechanische, magnetische und elektrische Randbedingungen beachtet werden. Daneben steht für die neuen AMB nur ein begrenzter Bauraum zur Verfügung, da diese das BS-AMB ersetzen sollen.

3.1 Mechanische Anforderungen

Das Lager muss die negativen Steifigkeiten des Axiallagers und der MGE sowie die im Betrieb wirkende Unwuchtskraft kompensieren. Die linearisierte Kraft eines AMB kann, unter Vernachlässigung von Streuflüssen und magnetischen Widerständen des Jochs und des Ankers, in Abhängigkeit des Steuerstroms i_s und der Rotorauslenkung x zu

$$F_{\mathrm{AMB}} = k_i i_s + k_s x$$
$$\text{mit } k_i = N^2 A \mu_0 \frac{i_0}{s_0^2} \text{ und } k_s = -N^2 A \mu_0 \frac{i_0^2}{s_0^3} \quad (1)$$

abgeschätzt werden [4]. Die Konstante k_i ist der Kraft-Strom-Faktor und k_s der Kraft-Weg-Faktor. Beide sind von der Windungszahl der Spulen N, der Polfläche A, der magnetischen Feldkonstante μ_0 sowie dem Vormagnetisierungsstrom i_0 und dem Nominalluftspalt s_0 abhängig. Anhand von Gleichung (1) zeigt sich eine Herausforderung, die speziell bei den Außenläufern entsteht. Durch die hohen Drehzahlen des Systems kommt es zur Aufweitung des Rotors, wodurch sich der Nominalluftspalt s_0 mit steigender Drehzahl vergrößert und die Maximalkraft des AMB sinkt
Neben der Kraft ist die Kraftdynamik für die Stabilisierung des Rotors im Betrieb entscheidend. Die zeitliche Änderung der Kraft ergibt sich, unter Vernachlässigung der Aufweitung, aus (1) zu

$$\frac{dF_{\mathrm{AMB}}}{dt} = k_i \frac{di_s}{dt} + k_s \frac{dx}{dt}. \quad (2)$$

Unter der Annahme eines idealen Lagers, kann die zeitliche Änderung des Weges vernachlässigt werden. Die Kraftdynamik ist dann nur von der zeitlichen Stromänderung in den Steuerspulen abhängig. Diese wird zu

$$\frac{di_s}{dt} = \frac{u_{\mathrm{Amp}} - R \cdot i_s}{L} \text{ mit } L = \frac{\mu_0 N^2 A}{s_0} \quad (3)$$

bestimmt. Dabei ist u_{Amp} die Ausgangsspannung des Verstärkers und R der Spulenwiderstand. Anhand der Spuleninduktivität L zeigt sich dabei der erste Zielkonflikt bei der Auslegung des AMB. Bei festem i_0 und s_0 können Kraft und Kraftdynamik des Aktors nicht gleichzeitig maximiert werden, da eine kleinere Induktivität L ebenfalls zu einem kleineren Kraft-Strom-Faktor k_i führt. Bei der Auslegung muss daher ein Kompromiss zwischen Maximalkraft und Kraftdynamik gefunden werden. Die Verstärkerleistung hat ebenfalls einen Einfluss auf die Aktordynamik. Da der Leistungsverstärker des Systems nicht angepasst werden soll, wird dies hier nicht berücksichtigt. Eine Beschreibung dieses Einflusses kann [4] entnommen werden.

3.2 Magnetische Anforderungen

Um eine möglichst lineare Lagerkraft und ein effizientes AMB zu entwickeln, müssen die Materialsättigung und Streuflüsse minimiert werden. Die Materialsättigung beschreibt das Abflachen der B-H-Kurve des Materials bei Erreichen der Sättigungsflussdichte. In diesem Bereich wird die Aktorkraft stark nichtlinear. Zusätzlich ist der Betrieb im Sättigungsbereich ineffizient, da für eine Kraferhöhung eine überproportionale Stromerhöhung notwendig ist. Daher wird in der Auslegung gefordert, dass die maximale Flussdichte im Eisen kleiner als dessen Sättigungsflussdichte bleibt. Einen weiteren Einfluss auf die Effizienz des AMB haben Streuflüsse. Dies sind magnetische Flusskomponenten, die nicht im Arbeitsluftspalt verlaufen und dadurch nicht zur Kraftstellung beitragen. Der

Streuflussanteil hängt vom magnetischen Widerstand des Streuflusspfades und damit von der Geometrie des magnetischen Kreises ab. Für ein effizientes Lager sollte der magnetische Widerstand des Streuflusspfades daher möglichst groß sein.

4 Verlusteffekte

Für die Entwicklung effizienter AMB sind, neben den Anforderungen an die stabile Lagerung des Rotors, die Verluste bei der Auslegung zu berücksichtigen. Diese werden in Ummagnetisierungsverluste des Materials und ohmsche Verluste der Spulen eingeteilt.

4.1 Ummagnetisierungsverluste

Ummagnetisierungsverluste entstehen aufgrund von Flussdichteänderungen im Material. Nach [5] werden die Verluste in Hysterese-, Wirbelstrom- und Anomalieverluste eingeteilt. Diese können über

$$P_{\mathrm{Fe}} = V_{\mathrm{Fe}} \cdot \left(k_{\mathrm{h}} f \hat{B}^{\alpha} + k_{\mathrm{w}} f^2 \hat{B}^2 + k_{\mathrm{a}} f^{1.5} \hat{B}^{1.5}\right) \qquad (4)$$

berechnet werden. Hierbei ist V_{Fe} das ummagnetisierte Eisenvolumen, α, k_{h}, k_{w} und k_{a} sind Materialparameter, f ist die Ummagnetisierungsfrequenz und $\hat{B}$ ist die Amplitude der Änderung der magnetischen Flussdichte.
Das Materialvolumen V_{Fe} kann durch eine kompakte Aktorik und einen hohen Ausnutzungsgrad des Materials mit hoher Sättigungsflussdichte minimiert werden. Die Ummagnetisierungsfrequenz f ist von der Drehzahl des Systems, der Stellaktivität des Reglers und der Schaltfrequenz des Magnetlagerverstärkers abhängig. Geringere Drehzahlen führen zu einer geringeren Energiedichte des Systems und eine geringere Schaltfrequenz zu einer kleineren Reglerdynamik, wodurch eine Beeinflussung der Verluste durch diese Stellgrößen kaum möglich ist. Die Verlustkoeffizienten α, k_{h}, k_{w} und k_{a} sind materialabhängig, weshalb meistens verlustarme Elektrobleche verwendet werden. Der Einsatz von Elektroblechen ist auf zweidimensionale Flussverläufe beschränkt. Im Übergangsbereich vom radialen zum axialen Flussdichteverlauf treten allerdings dreidimensionale Flussverläufe auf (vgl. Bild 2). Um hohe Wirbelstromverluste in diesen Bereichen zu verhindern, wird SMC verwendet.
Zur Beurteilung von $\hat{B}$ in (4) müssen Rotor und Stator getrennt betrachtet werden. Unter der Annahme, dass die Vormagnetisierung konstant ist, führt diese im Stator zu einem konstanten Flussdichteanteil. Dieser wird vom Wechselfeld der Steuerspulen überlagert, wodurch $\hat{B}$ im Statorbereich nur vom Steuerstrom abhängig ist.
Die Ummagnetisierung im Rotor entsteht maßgeblich durch den Flussdichteeinbruch in den Polübergängen zwischen den einzelnen Aktoren. Dieser ist von der Gesamtflussdichte im Rotor und damit von der Vormagnetisierung und den Steuerströmen abhängig. Eine Verlustreduktion kann durch Verkleinern des Polübergangs erreicht werden, da dadurch der Flussdichteverlauf gleichmäßiger wird [3, 6]. Grenzen sind dabei durch die Koppelinduktivität zwischen den Aktoren gesetzt, da diese zu einer ungewollten Kraftkopplung zwischen den Aktoren und damit den Lagerachsen führen. Bild 3 zeigt den Flussdichteverlauf im Rotor anhand einer Simulation mit ANSYS Electronics Desktop für einen Polabstand von d=23 mm und d=1 mm.

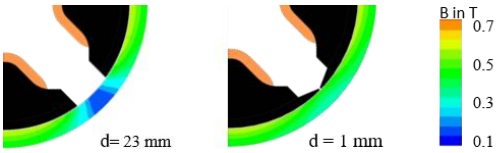


Bild 3 Flussdichteverlauf im Rotor bei verschiedenen Polabständen.

4.2 Ohmsche Verluste

Der zweite Verlustanteil, die ohmschen Verluste, entsteht in den Spulen der Elektromagnete und ist von dem elektrischen Widerstand R_{Spule} und dem Spulenstrom i abhängig. Die Steuerstromamplituden des BS-AMB liegen im Betrieb bei $i_{\mathrm{s}} \approx 1{,}3$ A. Der Vormagnetisierungsstrom ist auf $i_0 = 6$ A festgelegt. Der ohmsche Widerstand der Steuer- und Vormagnetisierungsspulen beträgt 1,1 Ω. Mit

$$P_{\mathrm{Cu}} = 4 \cdot R_{\mathrm{Spule}} \cdot i_0^2 + 2 \cdot R_{\mathrm{Spule}} \cdot i_{\mathrm{s}}^2 \qquad (5)$$

folgt $P_{\mathrm{Cu}} = 162{,}1$ W. Davon entstehen 158,4 W durch die Vormagnetisierung. Um die Verluste zu reduzieren, kann die Vormagnetisierung verringert werden. Da k_{i} und k_{s} von der Vormagnetisierung abhängen (vgl. Gleichung (1)), führt dies zu einer kleineren Lagerdynamik und einer kleineren maximalen Lagerkraft. Gleichzeitig sind bei kleinerer Vormagnetisierung größere Steuerströme notwendig, um identische Kräfte zu stellen. Dadurch steigen die Verluste der Steuerströme an. Eine Alternative zur Vormagnetisierungsreduktion stellt die Vormagnetisierung durch Permanentmagnete (PM) dar. Hierdurch werden keine ohmschen Verluste durch die Vormagnetisierung erzeugt.

5 Herausforderungen PM-Vormagnetisierung

Die irreversible Entmagnetisierung und die niedrige Permeabilität des Materials sowie Wirbelstromeffekte stellen den Entwickler bei der Verwendung von PM vor Herausforderungen. Die geringe Permeabilität wird häufig genutzt, um den magnetischen Kreis des PM vom magnetischen Kreis der Steuerspulen zu trennen. Damit wird der PM außerhalb des magnetischen Wechselfeldes platziert, sodass keine Wirbelstromeffekte im PM auftreten und die Gefahr der Entmagnetisierung durch Gegenfelder reduziert wird. In der Literatur gibt es verschiedene Ansätze, wie dies bei AMB erreicht werden kann. In [7, 8] wird eine homopolare Vormagnetisierung mit einem heteropolaren Steuerfluss kombiniert. Dadurch teilt sich der magnetische Fluss im Rotor in eine axiale und eine radiale Komponente auf. Aufgrund der Aufweitung des Rotors entstehen tangentialen Luftspalte zwischen den Ankersegmenten (vgl. Bild 2), daher ist dieses Prinzip für die Außenläufer nicht umsetzbar.
In [3] wird ein Konzept gezeigt, bei dem der Steuer- und der Vormagnetisierungsfluss im Rotor homopolar ausgeführt werden. Die Aufteilung erfolgt im Stator, indem die

magnetischen Flüsse der Steuerspulen der x- bzw. y-Achse gekoppelt werden. Um Kopplungen zwischen x- und y-Achsen zu verhindern, ist eine komplexe Geometrie und ein erhöhter Bauraum erforderlich. Daher ist auch diese Variante im ETA-290 nur schwer umsetzbar.

Eine weitere Möglichkeit ist die Platzierung des PM im magnetischen Kreis der Steuerspule. Die niedrige Permeabilität des Materials erhöht dabei den magnetischen Widerstand, wodurch die Maximalkraft und die Kraftdynamik des Lagers reduziert werden. Für die Berechnung der Kraft und der Kraftdynamik werden Reluktanznetzwerke verwendet. Hierzu wird der PM als Reihenschaltung aus Durchflutungsquelle und magnetischer Widerstand abstrahiert. Um die erzeugte Durchflutung des PM Θ_{PM} zu bestimmen, wird ein magnetischer Kurzschluss des magnetischen Kreises betrachtet. Das heißt, dass der Luftspalt des Magnetkreises geschlossen wird. Unter der Annahme, dass die Querschnittsfläche des PM der des magnetischen Kreises entspricht und Streuflüsse sowie der Eisenwiderstand vernachlässigbar sind, gilt nach [9]:

$$\Theta_{\mathrm{PM}} = \frac{l_{\mathrm{PM}} B_{\mathrm{r,PM}}}{\mu_{\mathrm{r,PM}} \mu_0}. \tag{6}$$

Dabei ist l_{PM} die Länge des PM in Flussrichtung, $B_{\mathrm{r,PM}}$ die Remanenzflussdichte und $\mu_{\mathrm{r,PM}} \approx 1$ die relative Permeabilität des PM. Die ideale Länge l_{PM} entspricht nach [9, S. 29 ff.] der doppelten Luftspaltlänge. Unter Vernachlässigung der Streuflüsse und des magnetischen Widerstands des Eisens im magnetischen Kreis, bedeutet dies, dass der magnetische Luftspalt des Lagers verdoppelt wird. Bei identischer Vormagnetisierung ist für eine identische Kraftstellung die doppelte Windungszahl der Steuerspule notwendig (vgl. Gleichung (1)). Da die Induktivität proportional zu N^2/s_0 ist, führt dies bei doppelter Windungszahl und doppeltem Luftspalt zu einer größeren Spuleninduktivität (vgl. Gleichung (3)). Dadurch wird die Dynamik des Lagers reduziert. Somit sind die Kraft und/oder Kraftdynamik des Lagers im Vergleich zu einem Lager mit Vormagnetisierungsspule verringert.

Des Weiteren werden durch Wirbelströme Gegenfelder induziert, die zu einer Schwächung des Feldes und zu Verlusten führen. Aus der Schwächung des Feldes resultiert eine Reduktion der Kraft und der Kraftdynamik. Um die Einflüsse durch Wirbelströme zu verringern, kann der PM, analog zu Elektroblechen, segmentiert werden. Je feiner die Segmentierung gewählt wird, desto geringer sind die Wirbelstromverluste. Da zwischen den Segmenten allerdings immer ein kleiner Luftspalt bleibt, sinkt mit steigender Segmentierung auch die erzeugte Durchflutung, da weniger Magnetmaterial auf einer identischen Fläche vorhanden ist. Ebenfalls steigt der Montageaufwand des Lagers. Daher ist bei der Wahl der Segmentierung ein Kompromiss aus Verlustreduktion, Montageaufwand und erzeugter Durchflutung zu finden.

Für eine Abschätzung der Wirbelstromverluste nach Gleichung (4), ist der Materialparameter k_{w} notwendig. Dieser wird bei PM nicht in den Materialdaten angegeben. Daher wird k_{w} für eine grobe Abschätzung der Verluste anhand des Verlustfaktors für Wirbelströme in Blechen nach [10, S. 87] zu

$$k_{\mathrm{w}} = \frac{\pi^2 d_{\mathrm{PM}}^2}{6\rho_{\mathrm{PM}}} \tag{7}$$

abgeschätzt. Der Parameter d_{PM} beschreibt die Breite der Segmente und ρ_{PM} die spezifische Leitfähigkeit des Materials. Diese Abschätzung der Verluste ist kritisch zu betrachten, da die PM-Segmente keinem Blechprofil entsprechen. Dennoch kann der berechnete Wert zur Abschätzung der Größenordnung der entstehenden Verluste genutzt werden.

Die Feldschwächung durch die Wirbelströme kann mittels der Eindringtiefe des Wechselfelds in das Material abgeschätzt werden. Für den Flussdichteabfall im PM gilt:

$$B = B_0 e^{-\frac{l_{\mathrm{PM}}}{\delta}} \text{ mit } \delta = \sqrt{\frac{2\rho_{\mathrm{PM}}}{2\pi f \mu_{\mathrm{PM}}}}. \tag{8}$$

Hierbei ist δ die Eindringtiefe in das Material, bei der die Flussdichte B_0 auf $1/e \cong 36\,\%$ abgefallen ist [11, S. 107]. Abschließend wird die irreversible Entmagnetisierung der PM betrachtet, die durch zu hohe Temperaturen und Gegenfelder im PM-Material hervorgerufen wird. Bild 4 zeigt den Einfluss von Gegenfeldern auf den Arbeitspunkt (P) des PM.

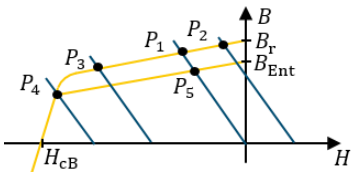


Bild 4 Typische Entmagnetisierungskurve des PM.

Der Arbeitspunkt, in dem der PM betrieben wird, wird näherungsweise durch die Entmagnetisierungsgerade (gelb) und die Arbeitspunktgerade (blau) bestimmt. Unter vereinfachter Annahme (keine Streuflüsse, $A_{\mathrm{Luft}} = A_{\mathrm{PM}}$) gilt für die Arbeitspunktgerade [12, S. 607 ff.]:

$$B_{\mathrm{PM}} = -\mu_0 \frac{l_{\mathrm{PM}}}{s} \left(H_{\mathrm{PM}} - H_{\mathrm{Gegen}}\right). \tag{9}$$

Ohne Gegenfeld, das heißt ohne Steuerstrom, beginnt die Arbeitspunktgerade im Ursprung. Der Schnittpunkt zwischen Arbeitspunktgerade und Entmagnetisierungsgerade ergibt den Arbeitspunkt (P_1). Für eine Krafterhöhung wird durch die Steuerspule das Feld des PM verstärkt, wodurch die Arbeitspunktgerade parallel zu P_2 verschoben wird. Um die Kraft zu reduzieren, wird ein Gegenfeld aufgebracht, womit P_1 zu P_3 verschoben wird. Ist das Gegenfeld zu groß, wird P_4 unterhalb des Abknickpunkts der Entmagnetisierungsgerade erreicht. Bei Wegnahme des Gegenfeldes wird nicht mehr P_1, sondern P_5 erreicht, wodurch B_{r} dauerhaft auf B_{Ent} reduziert wird. Der Verlauf der Entmagnetisierungskurve ist material- und temperaturabhängig. Bei höheren Temperaturen sind die Remanenzflussdichte und die Koerzitivfeldstärke kleiner. Für die Auslegung des PM muss daher sowohl die maximale Temperatur als auch die maximale Gegenfeldstärke beachtet werden.

Für die Anwendung im ETA-290 sind Samarium Cobalt Magnete (SmCo-Magnete) und Neodym-Magnete (NdFeB-Magnete) geeignet. Beide Materialien zeichnen sich durch hohe Remanenzflussdichten (0,8 T-1,4 T) aus, wobei diese bei Neodym-Magneten größer ist. Dafür können SmCo-Magnete bei höheren Temperaturen eingesetzt werden.

6 Vergleich der AMB

Anhand der zuvor beschriebenen Einflüsse auf die Verluste der AMB wird ein verlustarmes PM-AMB für den ETA-290 entwickelt. Durch numerische Simulation in ANSYS Electronics Desktop und analytische Berechnungen werden Kraft, Kraftdynamik und Verluste abgeschätzt. Das entwickelte PM-AMB ist in Bild 5 zu sehen.

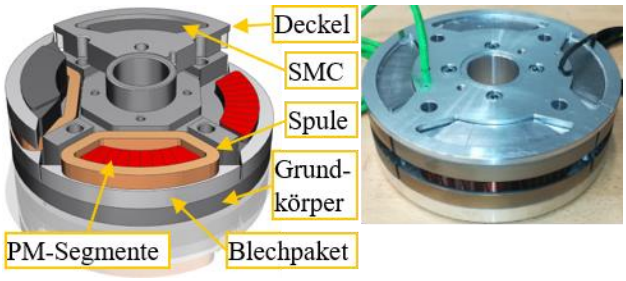


Bild 5 CAD-Modell des verlustarmen PM-AMB (links); Aufgebautes PM-AMB (rechts).

Die Aktoren des PM-AMB sind, wie bei der Variante des BS-AMB (vgl. Bild 2), C-förmig und homopolar ausgeführt. Die Polabstände betragen nur noch 4 mm statt 23 mm, wodurch die Rotorverluste bei 15.000 U/min von 42 W auf 5 W reduziert werden. Die Vormagnetisierung wird durch Neodym-Magnete (NdFeB-33EH) erzeugt. Diese sind je Aktor in zehn Segmente unterteilt. Die Wirbelstromverluste in den PM liegen nach Gleichung (4) und Gleichung (7) unter 10 W. Die Feldschwächung wird in der Auslegung der Lagerkraft berücksichtigt. Im Vergleich zum BS-AMB ist ein Teil des Eisenkreises aus Elektroblechen gefertigt. Anhand von analytischen und numerischen Abschätzungen können die Verluste bestimmt werden. Die Gesamtverluste des BS-AMB liegen bei 204 W. Im Vergleich dazu hat das PM-AMB nur 28 W Gesamtverluste. In Tabelle 1 sind die Eckdaten der beiden AMB zusammengefasst.

Tabelle 1 Vergleich der Lagerdaten.

		BS-AMB	PM-AMB
Statordurchmesser		288,6 mm	
Höhe		93 mm	
Polfläche A		3.200 mm²	4.100 mm²
Luftspalt s_0		1 mm	
Windungszahl	N_{Steuer}	170	200
	N_{Bias}	170	-
Vormagnetisierung i_0		6 A	-
Remanenzflussdichte B_r		-	1,15 T
Kraftdynamikfaktor $k_\text{i}(s_0)/L(s_0)$		7,5 kN/V	7 kN/V
Kraft $F_{\max}(s_0)$		2.250 N	1.870 N
Verluste $P_{\text{Ges}}(15.000\ \text{U/min})$		204 W	28 W

7 Experimentelle Validierung

Um die zuvor gezeigten Ergebnisse zu validieren, werden Verlustmessungen am Realsystem in Form von Auslaufversuchen durchgeführt. Dazu wird das System zunächst beschleunigt und bei Erreichen der Solldrehzahl der Motorwechselrichter abgeschaltet. Durch die Schleppverluste der MGE und der AMB sowie die Luftreibungsverluste, wird der Rotor bis 384 U/min abgebremst. Bei der ersten Messung (V 1) wird der Rotor mit einer Kombination der beiden AMB-Designs gelagert (unten das BS-AMB und oben das PM-AMB, vgl. **Bild 1**). Während der Messung werden die Ströme und Temperaturen der AMB sowie die Drehzahl des Systems erfasst. Das Ergebnis des ersten Versuchs zeigt Bild 6.

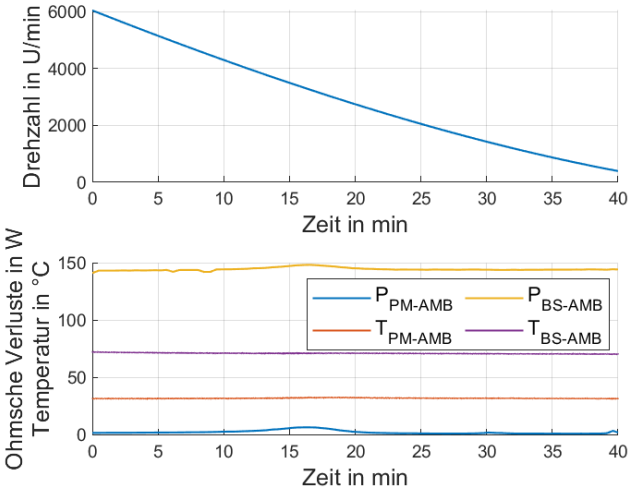


Bild 6 V 1: Rotordrehzahl (oben); Ohmsche Verluste P und AMB-Temperaturen T jeweils für das BS-AMB und PM-AMB (unten).

Zu Beginn des Versuchs hat der Rotor eine Drehzahl von 6.000 U/min. Diese nimmt mit der Versuchsdauer ab und erreicht nach 40 min eine Enddrehzahl von 384 U/min. Anhand der gemessenen Lagerströme werden die ohmschen Verluste nach Gleichung (5) bestimmt. Diese sind im unteren Bereich von Bild 6 zu sehen. Das BS-AMB hat ohmsche Verluste im Bereich von $P_{\text{BS-AMB}} = 145$ W. Den Hauptverlustanteil erzeugen die Vormagnetisierungsspulen mit 140 W. Die ohmschen Verluste des PM-AMB $P_{\text{PM-AMB}}$ liegen unter 5 W und sind vergleichbar mit den Verlusten der Steuerspulen des BS-AMB. Die geringeren ohmschen Verluste sind auch an den gemessenen Temperaturen in Bild 6 zu sehen. Das PM-AMB hat mit maximal 32 °C eine deutlich geringere Wärmeentwicklung als das BS-AMB mit 71 °C.

Die Rotorverluste der Magnetlager können nicht direkt gemessen werden. Um diese zu quantifizieren wird die mechanische Leistung des Systems

$$P_{\text{Mech}} = -P_\text{S} = \Theta_{\text{Rotor}}\omega\dot{\omega} \qquad (10)$$

aus der Winkelgeschwindigkeit ω, -beschleunigung $\dot{\omega}$ und der Massenträgheit des Rotors Θ_{Rotor} bestimmt. Da keine elektrische Leistung in das System geführt wird, entspricht die mechanische Leistung den Schleppverlusten des

Systems. Um zu zeigen, dass durch die neue Lagergeometrie eine Verlustminderung der Rotorverluste erzeugt wird, wird ein zweiter Auslaufversuch (V 2) durchgeführt. Bei diesem sind beide Lagerebenen mit den neuen PM-AMB ausgestattet. Aus dem Vergleich der beiden Messungen kann die Verlustminderung bestimmt werden. Dies ist allerdings nur unter der Annahme möglich, dass sich die Schleppverluste der MGE nicht verändern. Diese sind von der Temperatur der PM der MGE abhängig, da sich die Remanenzflussdichte und damit die Ummagnetisierungsverluste mit der Temperatur ändern [12, S. 605 ff.]. Daher müssen die PM der MGE bei den Auslaufversuchen die gleiche Temperatur haben. Bild 7 zeigt den Vergleich der beiden Messungen.

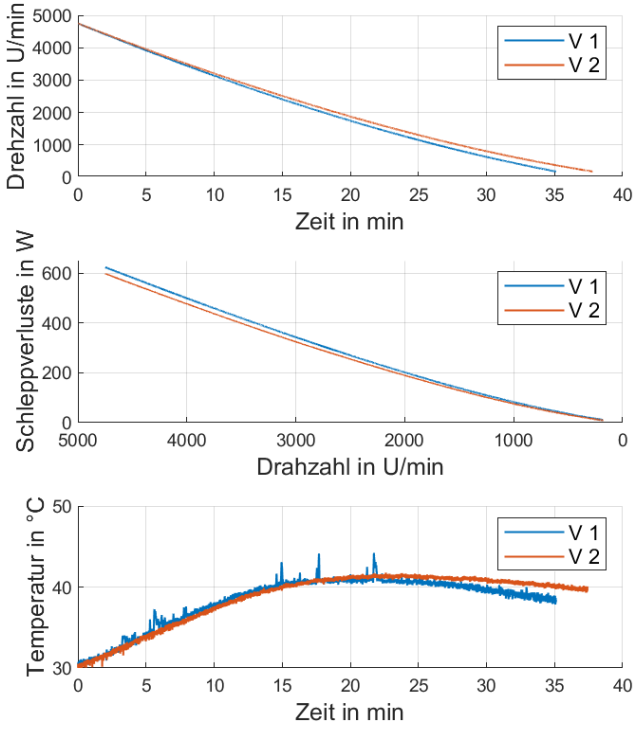


Bild 7 Vergleich der Auslaufversuche: Drehzahlen (oben); Schleppverluste (mitte), Temperatur der PM der MGE (unten).

Beim zweiten Versuch (V 2) läuft der Rotor 2,5 min länger aus, bis er die Enddrehzahl von 384 U/min erreicht. Damit sind die Schleppverluste beim zweiten Versuch kleiner, womit auf kleinere Rotorverluste der PM-AMB geschlossen werden kann. Dies zeigt auch der Vergleich der Schleppverluste in Bild 7. Bei 4.800 U/min sind die Verluste von V 2 ca. 23 W geringer als von V 1. Die Temperaturen der PM der MGE sind in beiden Versuchen vergleichbar, wodurch die Annahme, dass sich die Schleppverluste der MGE nicht ändern, gerechtfertigt ist. Die Reduktion der Rotorverluste durch das Anpassen der Polgeometrie wird damit durch die Messungen bestätigt.
Insgesamt werden durch das PM-AMB die Gesamtsystemverluste um ca. 340 W verringert. In Bezug auf die Gesamtsystemverluste im Standbybetrieb, die bei ca. 2000 W liegen, bedeutet dies eine Effizienzsteigerung von 17 %.

8 Zusammenfassung

Durch die Verwendung von PM zur Vormagnetisierung und die angepasste Polgeometrie, werden die Verluste der AMB um 80-90 % verringert. Die Funktion des entwickelten PM-AMB wird experimentell gezeigt, wobei eine Effizienzsteigerung des Gesamtsystems von 17 % im Standbybetrieb erreicht wird. Nachteilig bei dem Design ist der erhöhte Entwicklungs- und Montageaufwand der AMB. Ein weiterer Nachteil ist die Verwendung von Seltenerdmetallen. Die Menge an verwendeten Seltenerdmetallen kann z. B. durch eine Hybridlösung aus PM und Spule zur Vormagnetisierung reduziert werden.

Literatur

[1] L Quurck, M Richter, M Schneider, D Franz, and S Rinderknecht, “Design and practical realization of an innovative flywheel concept for industrial applications,” *Technische Mechanik*, vol. 37, 2-5, 2017.

[2] D. Franz, M. Schneider, M. Richter, and S. Rinderknecht, “Thermal Behavior of a Magnetically Levitated Spindle for Fatigue Testing of Fiber Reinforced Plastic,” *Actuators*, vol. 8, no. 2, p. 37, 2019.

[3] D. Franz, M. Richter, M. Schneider, and S. Rinderknecht, “Homopolar Active Magnetic Bearing Design for Outer Rotor Kinetic Energy Storages,” in *2019 IEEE International Electric Machines & Drives Conference (IEMDC)*, San Diego, CA, USA, 2019, pp. 774–778.

[4] G. Schweitzer, Ed., *Magnetic bearings: Theory, design, and application to rotating machinery*. Berlin: Springer, 2009.

[5] G. Bertotti, “General properties of power losses in soft ferromagnetic materials,” *IEEE Transactions on Magnetics*, vol. 24, no. 1, pp. 621–630, 1988.

[6] M. A. Pichot and M. D. Driga, “Loss reduction strategies in design of magnetic bearing actuators for vehicle applications,” *IEEE Trans. Magn.*, vol. 41, no. 1, pp. 492–496, 2005.

[7] A. Filatov, L. Hawkins, and P. McMullen, “Homopolar Permanent-Magnet-Biased Actuators and Their Application in Rotational Active Magnetic Bearing Systems,” *Actuators*, vol. 5, no. 4, p. 26, 2016.

[8] Brian T. Murphy, Hamid Ouroua, Matthew T. Caprio, John D. Herbst, *Permanent Magnet Bias, Homopolar Magnetic Bearings for a 130 kW-hr Composite Flywheel*. 9th International Symposium on Magnetic Bearings, Lexington KY, USA,2004.

[9] A. Chiba, *Magnetic bearings and bearingless drives*. Amsterdam: Elsevier, 2005.

[10] A. Traxler, *Eigenschaften und Auslegung von berührungsfreien elektromagnetischen Lagern*. Zürich, Eidgenössische Techn. Hochsch., Diss., 1985.

[11] R. Boll, *Weichmagnetische Werkstoffe: Einführung in den Magnetismus ; VAC-Werkstoffe und ihre Anwendungen*. Berlin [u.a.]: Siemens-Aktienges, Abt. Verl., 1990.

[12] A. Binder, *Elektrische Maschinen und Antriebe,* 1st ed. Berlin: Springer, 2012.

Eine mechatronische Lösung gegen Positionierfehler in Werkzeugmaschinen infolge von schwankenden Umgebungsbedingungen
A mechatronic solution against machine tool positioning errors due to fluctuating environmental conditions

Prof. Dr.-Ing. Christian Brecher, Fraunhofer Institut für Produktionstechnologie (IPT), 52074 Aachen, Deutschland
M.Sc. Filippos Tzanetos, Fraunhofer Institut für Produktionstechnologie (IPT), 52074 Aachen, Deutschland, filippos.tzanetos@ipt.fraunhofer.de

Kurzfassung

Die Umgebungsbedingungen am Ort der Installation einer Werkzeugmaschine sind ein wichtiger Bestandteil des Kaufvertrags zwischen Hersteller und Anwender der Maschine. Die Mehrheit der Anwender ergreifen noch zusätzlichen nichtmechatronischen Maßnahmen gegen die schwankenden Umgebungsbedingungen im Nachhinein. Standardmäßig werden Werkzeugmaschinen vor Produktionsstart warmgelaufen und während des Betriebs von erfahrenen Bedienern überwacht. Parallel greift jeder zweite Maschinenanwender auch zu Hallenklimatisierung.
Die Modellierung der komplexen Umgebungsbedingungen mit phänomenologischen Methoden setzt die Gegenmessung der Positionierfehler mit umfangreichen Versuchen voraus. Die individuellen Umgebungsbedingungen einer Produktionshalle sind schwer nachzubilden. Eine physikalische Vorhersage der Positionierfehler anhand von Messungen der thermisch bedingten Verformungen in der Maschinenstruktur bietet praktische Vorteile für den Einsatz einer automatisierten und robusten Korrektur an. Dieses Prinzip wurde unter realitätsnahen Bedingungen in Werkzeugmaschinen validiert.

Abstract

The ambient conditions at the place of installation of a machine tool are an important part of the sales contract between the manufacturers and the users of machine tools. The majority of users take additional, non-mechatronic measures against the fluctuating ambient conditions after the commissioning of the machine. Usually, machine users warm-up the machine before production starts and experienced operators monitor the positioning accuracy between operation intervals. At the same time, every second machine user also installs an indoor air conditioning system.
The modelling of the complex ambient conditions with phenomenological methods requires the counter-measurement of the positioning errors with extensive tests. The individual ambient conditions of a production hall are difficult to reproduce. A physical-principle based prediction of the positioning errors using measurements of the thermally induced deformations in the machine tool structure offers practical advantages for the use of an automated and robust correction. This principle has been validated in machine tools under realistic conditions.

1 Einleitung

Die Umgebungsbedingungen am Ort der Installation einer Werkzeugmaschine sind ein wichtiger Bestandteil des Kaufvertrags zwischen Hersteller und Anwender der Maschine [1]. Die Umfragen von Putz et al. deuten darauf hin, dass ca. 80% der Anwender noch zusätzlichen Maßnahmen gegen die schwankenden Umgebungsbedingungen im Nachhinein ergreifen. Dies validiert die hohe Bedeutung dieser Fehlerursache für die Industrie. Trotzdem werden hierbei hauptsächlich nicht-mechatronische und somit nicht-automatisierbare Lösungen eingesetzt. Standardmäßig werden Werkzeugmaschinen vor Produktionsstart warmgelaufen und während des Betriebs von erfahrenen Bedienern überwacht. Diese Überwachung beinhaltet eine regelmäßige Abnahme der Fehler und den Einsatz von intern entwickelten, empirischen Strategien. Auf diesen Strategien stützen Maschinenanwender ihre Qualitätsvorteile im Wettbewerb. Parallel greift jeder zweite Maschinenanwender auch zu Hallenklimatisierung.

Den hohen Energie- und Personal-Bedarf und die Produktivitätsverluste des aktuellen Stands der Technik in der Industrie kann eine mechatronische Lösung deutlich reduzieren, wenn sie steuerungstechnisch die Positionierfehler korrigiert. Die Herausforderung im Kontext der komplexen Umgebungsbedingungen ist, geeignete Hilfsvariablen zu erfassen, um den thermo-mechanischen Zustand der Werkzeugmaschine automatisiert und zuverlässig zu quantifizieren. In der Forschung wurden bisher meistens Lösungen untersucht, die Temperatursensoren an numerisch ermittelten, optimalen Messstellen anbringen und den Zusammenhang mit den Positionierfehlern phänomenologisch bilden. Dies setzt die Gegenmessung der Positionierfehler mit geeigneten Versuchen voraus, die am Installationsort beim Maschinenanwender wegen des Produktionsdrucks nicht möglich ist. Die individuellen Umgebungsbedingungen einer Produktionshalle sind aber schwer nachzubilden.
Eine physikalische Vorhersage der Positionierfehler anhand von Temperatursensoren erfordert komplexe FEM-

und CFD-Simulationen, die unterschiedliche Modellparameter schätzen sollen [2]. Die Arbeit von Glänzel et al. [3] stellt ein charakteristisches Beispiel für den nötigen Aufwand an Daten, Experimenten und Simulation, um den Wärmeübergangskoeffizient der Maschine zur Umgebung zu ermitteln. In diesem Fall werden mehrere CFD-Simulationen mit theoretischen oder experimentell ermittelten Umgebungstemperaturen und Luftströmungsprofilen als Eingangsparameter durchgeführt. Um diese Simulation im Online-Betrieb nicht wiederholen zu müssen, werden die simulativ ermittelten Wärmeübergangskoeffizienten der Maschinenoberflächen in Form von Kennfeldern gespeichert. Diese Kennfelder dienen als Modellparamater für eine reduzierte, numerisch effiziente, thermo-elastische FEM-Simulation, die das Verformungsfeld anhand von gemessenen Temperaturen prozess-parallel simulieren kann [4].

Neben den Strukturtemperaturen werden entweder im Online-Betrieb oder in Voruntersuchungen viele weitere Hilfsvariablen benötigt, die schwer zu erfassen sind, wie zum Beispiel die Luftströmung, die Luftfeuchtigkeit, der Luftdruck und das Temperaturfeld der Umgebung im angrenzenden Volumen zur Maschine. Alle diese Informationen beeinflussen den Wärmeübergang zwischen Maschine und Umgebung. Stattdessen werden diese Hilfsvariablen empirisch ermittelt [2] oder räumlich interpoliert [3]. Dies führt zu ausgeprägten Unsicherheiten in der Modellierungskette, sodass die Ergebnisse dieser physikalischen Lösungsansätze an mangelnder Zuverlässigkeit leiden. Insbesondere dann, wenn die Modellparameter nicht an den Umgebungsbedingungen am Installationsort der behandelten Werkzeugmaschine ermittelt wurden.

2 Lösungsansatz

2.1 Interpretation der integralen Verformungsmessung

In dieser Arbeit wird eine neue Lösung vorgestellt, die auf die Messung thermisch bedingter Verformungen in der Werkzeugmaschinenstruktur beruht [5]. Das eingesetzte Messprinzip basiert auf Referenzstäben aus kohlenstofffaserverstärktem Kunststoff (CFK), die an den Oberflächen der Maschinenstruktur installiert werden. Mit Wegaufnehmern kann die relative Dehnung zwischen zwei Punkten in Richtung des Referenzstabs gemessen werden. Der Referenzstab dehnt sich gegenüber der Maschinenstruktur unter thermischer Belastung viel weniger, da der thermische Ausdehnungskoeffizient des CFK mit -0,1 µm/m/K wesentlich kleiner ist als der Koeffizient üblicher Werkstoffe für Werkzeugmaschinengestelle von ca. 11 µm/m/K.

Unter der Annahme linearer Materialeigenschaften dehnt sich eine Maschinenkomponente proportional zur Temperaturdifferenz in Längsrichtung aus (siehe **Bild 1**). Jedoch kann der Temperaturgradient einem linearen, quadratischen oder beliebigen Verlauf folgen, sodass das Temperaturprofil über die Komponentenlänge integriert werden muss:

$$u_{IDS}(t,x_1) = \int_{x_A}^{x_E} \varepsilon_1(t,x_1)dx_1$$
$$u_{IDS}(t,x_1) = \int_{x_A}^{x_E} \alpha \cdot \Delta T_1(t,x_1)dx_1 \quad (1)$$

Das eingesetzte Messprinzip misst direkt das Ergebnis dieses Integrals, was weiterhin als „integrale Verformung“ (Integral Deformation Sensors – IDS) bezeichnet wird, denn diese Gleichung wird zur Interpretation der IDS-Messungen herangezogen.

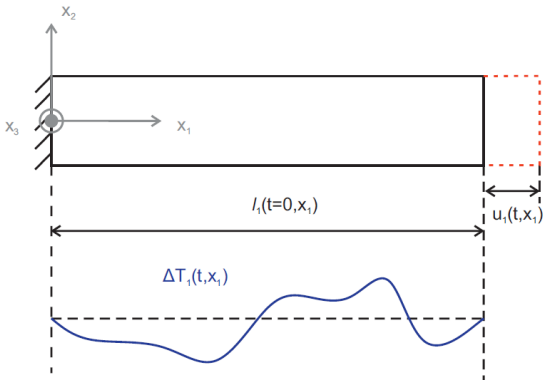


Bild 1 Thermische Ausdehnung am Beispiel eines fest eingespannten Balkens

Wie es in Gleichung (1) ersichtlich ist, misst der IDS die momentane, thermisch bedingte, integrale Verformung entlang seiner Länge, da beide Seiten der Gleichung für einen Zeitpunkt t definiert sind. Ein physikalisches Verformungsmodell basierend auf diesen Sensoren braucht daher keine Dynamik zu beschreiben, sondern lediglich den momentanen Zusammenhang zwischen IDS-Messung und Verformungsfeld der Maschinenstruktur.

Im Gegensatz zur eindimensionalen Ausdehnung würde ein Temperaturprofil über den Querschnitt einer Maschinenkomponente zu Biegung, Schub oder Torsion führen. Werden also mehrere Sensorapplikationen räumlich angeordnet, so lassen sich auch solche Verformungen anhand von Gesetzen der Mechanik abbilden, wie die Euler-Bernoulli-Balkentheorie und die Love-Kirchhof-Plattentheorie.

2.2 Physikalisches Vorhersage-Modell

Diese mechanischen Modelle beschreiben das Verformungsfeld in Hinsicht von Starrkörperverschiebungen der Balken- oder Platten-Querschnitte der Struktur, denn die Annahmen dieser Theorien setzen einen konstanten oder apriori bekannten Verlauf der Dehnungen über den Querschnitt voraus. Die Verformung in x_1-Richtung eines Euler-Bernoulli Balkens lässt sich zum Beispiel hinsichtlich der Längsdehnung $\bar{u}_1$ und der Verdrehungen $\bar{\varphi}_2$ und $\bar{\varphi}_3$ seiner eben bleibenden Querschnitten beschreiben:

$$u_1(t,x_1) = \bar{u}_1(t,x_1) + x_3 \cdot \bar{\varphi}_2(t,x_1) - x_2 \cdot \bar{\varphi}_3(t,x_1) \quad (2)$$

Der momentane Zusammenhang dieser Verformung mit einem in x_1-Richtung gelegenen IDS ergibt sich aus der Substitution der Gleichung (2) in Gleichung (1):

$$u_{IDS}(t,x_1) = \int_{x_A}^{x_E} \varepsilon_1(t,x_1)dx_1$$

$$= \int_{x_A}^{x_E} \left[\frac{\partial \bar{u}_1(t,x_1)}{\partial x_1} + x_3 \cdot \frac{\partial \bar{\varphi}_2(t,x_1)}{\partial x_1} - x_2 \cdot \frac{\partial \bar{\varphi}_3(t,x_1)}{\partial x_1}\right] dx_1 \quad (2)$$

$$= \bar{u}_1(t,x_1)|_{x_A}^{x_E} + x_3 \cdot \bar{\varphi}_2(t,x_1)|_{x_A}^{x_E} - x_2 \cdot \bar{\varphi}_3(t,x_1)|_{x_A}^{x_E}$$

Im Falle eines einseitig eingespannten Balkens sollte der IDS in der Nähe der Einspannung installiert werden, sodass die unbekannten Querschnittsfreiheitsgrade bei $x_1 = x_A$ entfallen. Somit ergeben sich drei Unbekannten zur Beschreibung der Längsdehnung sowie der zwei Biegungsrichtungen des Balkens. Diese sind durch drei IDS in x_1–Richtung bestimmbar.

Um das gesamte Verformungsfeld für jeden Punkt der Struktur (x_1, x_2, x_3) zu ermitteln, werden zusätzlich die mechanischen Randbedingungen herangezogen. Die Kombination mit den Querschnittsfreiheitsgraden bei $x_1 = x_E$ führt dann zu einer eindeutigen Beschreibung des Verformungsfeldes. Diese Beschreibung lässt sich im einfachsten Fall durch globalen Polynomen oder im industriellen Einsatz durch abschnittsweise definierten Ansatzfunktionen und geeigneten Finiten Elementen implementieren.

Neben der geeigneten Anordnung der IDS an den Maschinenoberflächen ist zudem deutlich, dass die momentanen Lagen der Wärmequellen auch am Eingang dieses Modells herangezogen werden müssen, um das Gesamtverformungsfeld richtig abzubilden. Diese Herausforderung ist zur Beschreibung der thermisch bedingten Verformung infolge von Achsbewegungen und vor allem bei Komponenten mit bewegenden Maschinenteilen von Bedeutung. Unter Berücksichtigung der aktuellen Positionen der Achsen lässt sich der momentane Mittelpunkt der Wärmequelle in thermischer Echtzeit ableiten. Ergebnisse in diesem Kontext sind in [5] veröffentlicht.

Der Fokus dieser Arbeit liegt aber in der Beschreibung der thermisch bedingten Verformungen infolge von schwankenden Umgebungsbedingungen. Für diesen Lastfall reichen die mechanischen Randbedingungen und angemessen positionierten IDS aus.

3 Versuchsbedingungen

Zur Validierung des Lösungsansatzes wurde ein Drift- oder ETVE-Test gemäß den Normen [6] an einem Bearbeitungszentrum in Fahrständerbauweise mit Dreh-Schwenk-Kopf und YbXZABC Kinematik gemäß der Konvention aus den Normen [7]. Um ein Aufwärmen der Maschinenachsen während der Untersuchung zu vermeiden, wurden alle Maschinenachsen sechs Stunden vor Beginn des Experiments ausgeschaltet (Regelung aus). Die Maschine stand in den letzten zwei Tagen vor dem Experiment still und wurde in dieser Periode lediglich von Montagearbeiten und kurzzeitigen Achsbewegungen thermisch belastet. Der Drift-Test ist sechs Stunden nach der Installation des Messdorns an der Maschine durchgeführt, damit Maschine und Messsystem sich in thermischem Gleichgewicht befinden. Die einzige thermische Last auf die Maschine während des Experiments war die Schwankung der Umgebungstemperatur in der Halle, nachdem die Hallenklimatisierung um 20:00 ausgeschaltet wurde. Das Experiment lief bis 6:00 am Tag danach, wenn die Hallenklimatisierung wieder eingeschaltet wurde.

Mithilfe eines Messdorns gemäß den Normen [6] konnten drei translatorischen und zwei rotatorischen Drift-Komponenten gemessen werden. Die kombinierte Messunsicherheit des Messdorns beträgt 3,6 µm. Der Messdorn lag bei X -1257, Y -1294, Z -1288 mm in Bezug auf das Maschinenkoordinatensystem (siehe **Bild 2**).

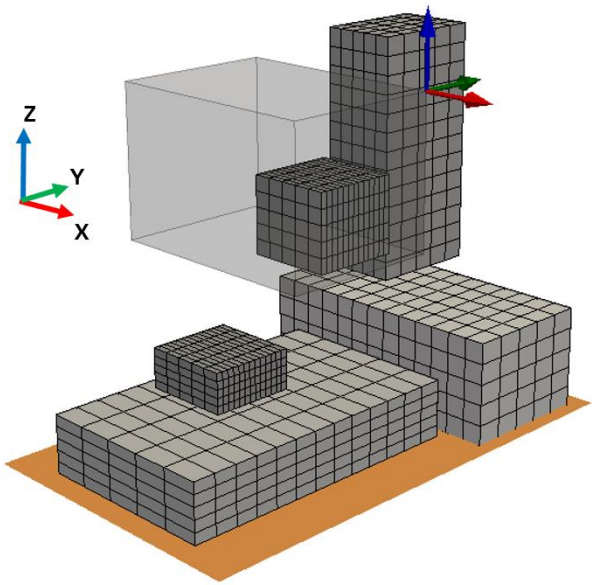


Bild 2 Räumliches Modell der Versuchsmaschine mit Darstellung des Maschinenkoordinatensystems und des Arbeitsraums

Die Maschinentür blieb geschlossen während des Experiments. Das Bearbeitungszentrum ist von allen Seiten verkleidet, außer von oben, sodass der Arbeitsraum und alle naheliegenden Maschinenkomponenten unmittelbar beeinflusst sind. Die hintere Seite des Maschinenbetts und der Fahrständer stehen nicht frei zur Umgebung.

Die Strukturbauteile des Bearbeitungszentrums wurden anhand ihrer Querschnittsgeometrie und Dimensionen folgendermaßen modelliert:

- Hinterer Teil des Maschinenbetts als Reissner-Mindlin-Platte in Kombination mit volumetrischer Dehnung in drei Raumrichtungen

- Vorderer Teil des Maschinenbetts als Euler-Bernoulli-Balken in Kombination mit Längsdehnung
- Werkstücktisch als Love-Kirchhoff-Platte in Kombination mit volumetrischer Dehnung in drei Raumrichtungen
- Fahrständer als Euler-Bernoulli-Balken in Kombination mit volumetrischer Dehnung in drei Raumrichtungen
- Spindelstock als dreidimensionales Element (nur volumetrische Dehnung in drei Raumrichtungen wurde berücksichtigt)

4 Ergebnisse

Bild 3 zeigt die Entwicklung der IDS-Messungen während des Experiments. Die installierten IDS sind:

- Drei IDS am hinteren Teil des Maschinenbetts (RB01, RB02 und RB03)
- Ein IDS am vorderen Teil des Maschinenbetts (FB01)
- Drei IDS am Werkstücktisch (FB02, WT01 und WT02)
- Vier IDS am Fahrständer (MC01, MC02, MC03H und MC04)
- Drei IDS am Spindelstock (HS01, HS02 und HS03H)

Somit konnten alle Strukturbauteile der Versuchsmaschine berücksichtigt werden.

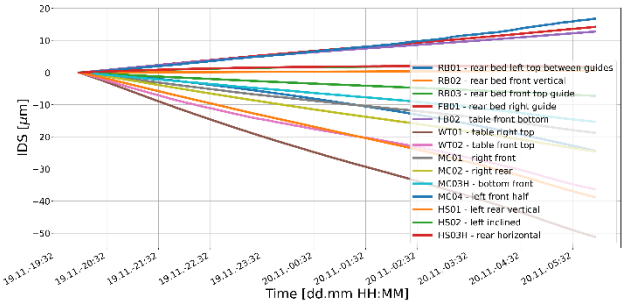


Bild 3 Entwicklung der IDS-Messungen während des Versuchs

Um die Angemessenheit der Modellergebnisse zu beurteilen, kann man das Verformungsfeld aller Strukturbauteil in einem beliebigen Zeitpunkt visualisieren. **Bild 4** zeigt entsprechend das Gesamtverformungsfeld der Versuchsmaschine im Zustand des maximalen Drifts am Ende des Versuchs.

Man erkennt deutlich die volumetrische Stauchung aller Strukturbauteilen sowie die ausgeprägte Verbiegung des Werkstücktisches und des hinteren Teils des Maschinenbetts. Diese Strukturbauteile verbiegen sich lediglich, weil ihre mechanischen Randbedingungen keine freie Stauchung zulassen. Zum Beispiel lassen die Führungsschienen der Y-Achse keine freie Stauchung des Werkstückstiches in X-Richtung. Dies verursacht Biegung um die Y-Achse. Auf der anderen Seite kann sich der Werkstücktisch in Y-Richtung frei stauchen. Aus diesem Grund erkennt man keine Biegung um die X-Achse am Werkstücktisch.

Der gleiche Prozess kann für alle Strukturbauteile gefolgt werden, um ihre Modelle zu verbessern, indem man die Annahmen der eingesetzten Theorie der Mechanik prüft, oder um die thermischen Schwachstellen der Maschinenkonstruktion zu identifizieren.

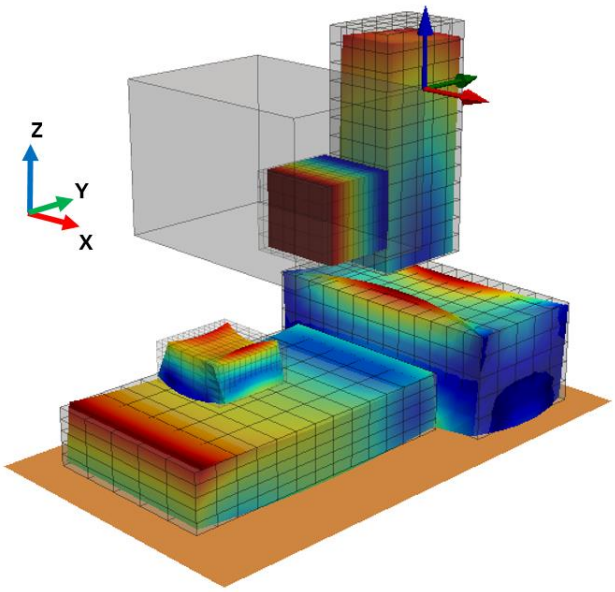


Bild 4 Thermo-elastischer Zustand der Versuchsmaschine am Ende des Drift-Tests

Bild 5 gibt einen Vergleich des translatorischen Drifts der Versuchsmaschine im korrigierten und im nicht korrigierten Fall. Da die Achsen während des Versuchs nicht in Regelung standen, wurde die Korrektur Offline an den gemessenen TCP-Daten angewendet. Für den Maschinenanwender ist aber der korrigierte Zustand von größter Bedeutung.

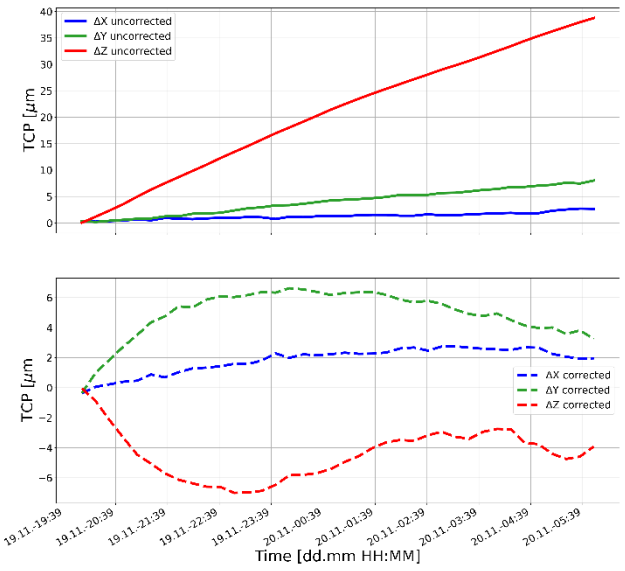


Bild 5 Vergleich der Entwicklung des translatorischen Drifts im nicht korrigierten (oben) und im korrigierten (unten) Fall

Der Drift in X- und Y-Richtung liegt betragsmäßig in der gleichen Größenordnung mit der kombinierten Messunsicherheit des Messdorns und mit der kombinierten Unsicherheit der Vorhersage. In diesem Fall ist es nicht möglich, den Fehler der Vorhersage von der Messunsicherheit zu unterscheiden. Vielmehr ist die Vorhersagegenauigkeit des Lösungsansatzes in Z-Richtung ersichtlich, da der Drift in diese Richtung deutlich größer ist.
Zudem kann der ETVE-Test nur an einer Position im Arbeitsraum durchgeführt werden. Aus diesem Grund kann noch keine Aussage getroffen werden, wie die Vorhersage-Genauigkeit und–Unsicherheit sich im Arbeitsraum entwickeln. Je nach Sensoranordnung besitzt dieser Ansatz für eine Maschinenkinematik einen definierten Verlauf über die drei Raumrichtungen. Es ist möglich, dass eine Seite des Arbeitsraums genauer abgebildet werden kann, als die andere. Dieser Verlauf kann mithilfe von statistischen Simulationen ermittelt werden. Ein Beispiel hierfür ist die Monte-Carlo-Simulation mit Normalverteilungen der Messunsicherheit aller eingesetzten IDS als Eingabeparameter. Solche Untersuchungen werden in künftigen Arbeiten genauer betrachtet.

5 Zusammenfassung

Ein physikalisch basierter Ansatz für die automatisierte Korrektur thermisch bedingter Fehler in Werkzeugmaschinen wurde unter realitätsnahen schwankenden Umgebungsbedingungen validiert. Der praktische Vorteil dieses Ansatzes ist, dass man eine zuverlässige Vorhersage der thermisch bedingten Positionierfehler ohne aufwendiges Daten-Training erreichen kann. Im Gegensatz hängt die Qualität und Zuverlässigkeit der Methode von der Sensorplatzierung, den angenommenen Modellvereinfachungen (zum Beispiel Ebenbleiben der Querschnitte nach der Euler-Bernoulli-Balkentheorie) und von der Ermittlung des momentanen Mittelpunkts einer bewegenden Wärmequelle ab. Diese Abhängigkeiten sollen in künftigen Arbeiten näher untersucht werden. Der Einfluss der Sensorplatzierung wurde bereits für die gleiche Maschinenkinematik in [8] untersucht.
Die Ergebnisse zeigen eine gute Übereinstimmung der gemessenen mit den ermittelten Fehlern unter schwankenden Umgebungsbedingungen. Eine charakteristische Anwendung mit diesen Bedingungen ist, dass der Maschinenbediener der Frühschicht einen Überblick des thermischen Verzuges der Maschine seit der letzten Schicht bekommt und im idealen Fall die Maschine ohne Warmlaufen in Betrieb nimmt. Dies wäre aber nur dann möglich, wenn die Unsicherheit der Vorhersage für die Produktion ausreichend ist. Aus diesem Grund ist die a-priori Schätzung der Unsicherheit von großer praktischer Bedeutung und soll in künftigen Arbeiten behandelt werden.
Eine weitere praxis-relevante Anwendung ist der Einfluss des unvorhersehbaren Öffnens der Hallentoren. Der Ansatz soll für diese Anwendung ebenfalls validiert werden. In diesem Fall ist der Mittelpunkt der signifikanten Wärmesenke (Luftströmung aus Hallentor) bekannt und nicht beweglich. Diese Anwendung stellt daher keine weitere Herausforderung für den Ansatz, wird aber die Unsicherheit der Ermittlung dieses Mittelpunktes auch in Betracht ziehen.
Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft – Projekt-ID 174223256 – Sonderforschungsbereich TRR 96 – Teilprojekt C03 gefördert.

6 Literatur

[1] Putz, M.; Richter, C.; Regel, J.; Bräunig, M. (2018): *Industrial consideration of thermal issues in machine tools*. In: Prod. Eng. Res. Devel. 12 (6), S. 723–736. DOI: 10.1007/s11740-018-0848-6.

[2] Da Silva, P.; Pena-Gonzalez, L. E.; Tanabe, I.; Takahashi, S. (2018): *Machine tool distortion estimation due to environmental thermal fluctuations — A focus on heat transfer coefficient*. In: Journal of Machine Engineering 18 (2), S. 17–30. DOI: 10.5604/01.3001.0012.0921.

[3] Glänzel, J.; Ihlenfeldt, S.; Naumann, C.; Putz, M. (2017): *Decoupling of Fluid and Thermo-elastic Simulations on Machine tools Using Characteristic Diagrams*. In: Procedia CIRP 62, S. 340–345. DOI: 10.1016/J.PROCIR.2016.06.068.

[4] Hernández-Becerro, P.; Purtschert, J.; Konvicka, J.; Buesser, C.; Schranz, D.; Mayr, J.; Wegener, K. (2021): Reduced-Order Model of the Environmental Variation Error of a Precision Five-Axis Machine Tool. In: Journal of Manufacturing Science and Engineering 143 (2), S. 660. DOI: 10.1115/1.4047739.

[5] Brecher, C.; Klatte, M.; Lee, T. H.; Tzanetos, F. (2018): *Metrological analysis of a mechatronic system based on novel deformation sensors for thermal issues in machine tools*. In: Procedia CIRP 77, S. 517–520. DOI: 10.1016/j.procir.2018.08.245.

[6] ISO 230-3 *Test code for machine tools*. - 2007. - Bd. Part 3: Determination of thermal effects.

[7] ISO 10791-6 *Test conditions for machining centres*. - 1998. - Bde. Part 6: Accuracy of feeds, speeds and interpolations.

[8] Brecher C., Herzog, R., Spierling, R., Tzanetos, F., Naumann, A. (2020): *Evaluation of optimal positioning methods of integral deformation sensors for the correction of thermal errors in machine tools*. In: Proceedings of the euspen Special Interest Group Thermal Issues, S. 26-29. ISBN: 978-0-9957751-6-9

Applikation und Test eines dedizierten Schaltungssystems für elektrische Mehrgang-Fahrzeugantriebe

Daniel Schöneberger, M.Sc., Guanlin Gao, M.Sc., Prof. Dr.-Ing. Stephan Rinderknecht, Technische Universität Darmstadt, Institut für Mechatronische Systeme im Maschinenbau, 64287 Darmstadt, Deutschland

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Kurzfassung

Mehrmotorige und mehrgängige elektrische Fahrzeugantriebe können bei gleicher Systemleistung bessere Fahrleistungen (Anfahrmoment und Höchstgeschwindigkeit) und eine höhere Effizienz - gleichbedeutend mit einer größeren Reichweite bei gleicher Batteriekapazität - gegenüber Festgang-Antrieben erzielen. [1]
Im Rahmen des öffentlich geförderten Forschungsprojekts „Speed4E“ wird ein solches innovatives Antriebskonzept entwickelt. Am IMS der TU Darmstadt liegt dabei ein Fokus auf der Entwicklung eines dedizierten mechatronischen Schaltungssystems, das gegenüber konventionellen Schaltungssystemen erhebliche Vorteile im Hinblick auf Kosten, Komplexität, Bauraum und Schaltzeiten aufweist.
In diesem Beitrag wird das innovative Schaltungssystem vorgestellt und erste Systemtests an einem Schaltungsprüfstand diskutiert. Im Ausblick werden die kommenden Systemtests, bis hin zum Einsatz in einem Versuchsfahrzeug, skizziert.

Abstract

Multiple-motor and multiple-speed electric powertrains can increase the launch torque and maximum vehicle speed for the same system power and furthermore lead to increased efficiency and thus, higher range for a given battery capacity compared to a fixed-speed transmission. [1]
In the scope of the publicly funded research project “Speed4E”, such an innovative powertrain is developed with a dedicated mechatronic shifting system. Thereby, complexity, costs, installation space and shifting time can be reduced significantly compared to conventional shifting systems.
This paper introduces the innovative shifting system and discusses the first system tests at a shifting test bench. The upcoming system tests including tests with a demonstrator vehicle are outlined in the outlook.

1. Einleitung

In konventionellen automatisierten Stirnradstufengetrieben, wie automatisierten Schaltgetrieben (ASG), oder Doppelkupplungsgetrieben (DKG), werden Sperrsynchronisierungen durch pneumatische, hydraulische oder elektromechanische Aktoriksysteme betätigt. Die Synchronisierungszeit ist unter anderem von der Betätigungskraft abhängig. Um kurze Schaltzeiten zu erzielen ist also eine hohe Aktorkraft notwendig. [2]
Auch das Einlegen eines Ganges bei Fahrzeugstillstand kann bei konventionellen Getrieben hohe Stellkräfte erfordern. In elektrischen Antrieben kann durch eine aktiv geregelte elektrische Synchronisierung zum einen auf die mechanische Synchronisierung verzichtet werden und zum anderen das Einspuren auch bei Fahrzeugstillstand erleichtert werden. Die dadurch deutlich reduzierten Anforderungen an die notwendige Schaltkraft führt dazu, dass auch alternative zu den am Markt befindlichen Aktorkonzepten in Betracht kommen. Hierzu zählen insbesondere lineare Aktoren, wie Tauchspulenaktoren und Elektromagnete.
Im Rahmen des Speed4E-Projektes wird daher ein Schaltungssystem, bestehend aus Klauenkupplung, Schaltaktorik (inklusive deren Ansteuerung), und der übergeordneten Schaltablaufsteuerung entwickelt, in dem die Teilsysteme domänenübergreifend aufeinander und auf den Antrieb selbst abgestimmt sind. Dadurch werden die Komplexität - damit einhergehend die Kosten - und gleichzeitig die Schaltzeiten erheblich reduziert. Übergeordnet

wird so die elektrische Mehrgängigkeit sowohl für Fahrer als auch für Fahrzeughersteller attraktiver.
In diesem Beitrag wird zunächst in Kapitel 2 der Speed4E-Antrieb und das Schaltungssystem beschrieben. Anschließend werden Systemtests des Funktionsmusters am Schaltungsprüfstand „InnoShift“ diskutiert. Abschließend werden eine Zusammenfassung und ein Ausblick auf die noch ausstehenden Untersuchungen gegeben.

2. Systembeschreibung

Wie in der Einleitung bereits beschrieben sind alle Komponenten des Speed4E-Schaltungssystems aufeinander und auf den Antriebsstrang selbst abgestimmt. Aus diesem Grund ist das System auch als dedizierte Schaltung zu verstehen. Für eine bessere Nachvollziehbarkeit wird zunächst der Speed4E-Antriebsstrang vorgestellt. Anschließend wird die Klauenkupplung und der smarte Schaltktor vorgestellt.

2.1 Speed4E-Antriebssystem

Wie bereits bei dem in [3] beschriebenen Vorgängerprojekt „Speed2E“, handelt es sich beim Speed4E-Antriebsstrang um einen Antriebsstrang mit zwei elektrischen Maschinen und zwei Teilgetrieben, wovon eines mehrgängig ausgeführt ist. **Bild 1** zeigt das schematische Layout des Antriebsstrangs. EM1 ist hierbei eine Asynchronmaschine, die über Teilgetriebe 1 (STM1) mit einer konstanten Übersetzung fest an das Differential angebunden ist. EM2 ist eine permanenterregte Synchronmaschine, die über das schaltbare Teilgetriebe 2 (STM2) verfügt.

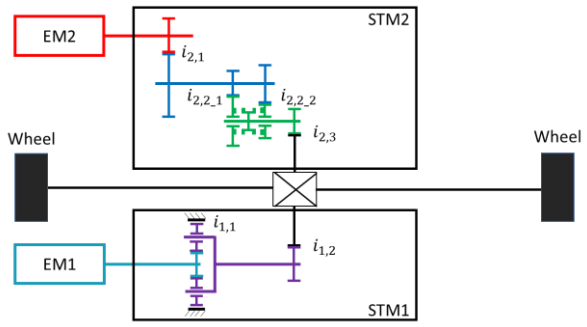


Bild 1 Layout des Speed4E-Antriebsstrangs [4]

Die als Klauenkupplung ausgeführte Schalteinheit ermöglicht die Schaltstellungen „Gang 1“, „Gang 2“ und „Neutral“. Die Betriebsstrategie entscheidet, welcher Anteil am Antriebsmoment durch die jeweilige Maschine gestellt wird und welcher Gang eingelegt werden soll.
Wie in [4] beschrieben ermöglicht die in **Bild 2** dargestellte Steuergerätearchitektur nicht nur eine aktive Differenzdrehzahlregelung, sondern darüber hinaus eine winkelgenaue Regelung der Klauenkupplung. Das bedeutet, dass beim Einlegen des Gangs die Differenzdrehzahl in der Klauenkupplung nahezu Null ist und eine Klaue-Lücke-Stellung aktiv eingeregelt wird. Dadurch wird beim Einlegen des Gangs ein Kontakt der Klauendächer, oder -Flanken zuverlässig verhindert.

Dies führt nicht nur zu schnellen und ruckfreien Einlegevorgängen, sondern trägt auch dazu bei, die Kraftanforderung an die Aktorik zu reduzieren.

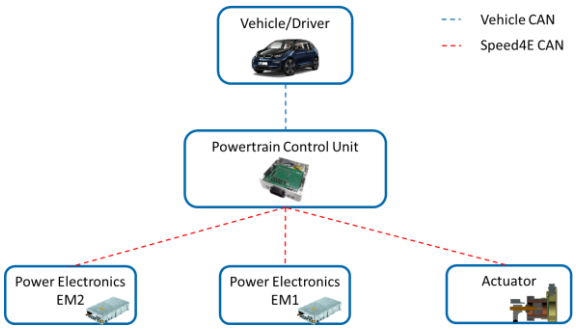


Bild 2 Steuergerätearchitektur im Speed4E-Fahrzeug [4]

2.2 Klauenkupplung

Klauenkupplungen überzeugen nach [2] insbesondere durch ihre Kompaktheit, geringe Anzahl an Bauteilen und die erreichbaren Schaltgeschwindigkeiten.
Die Speed4E-Klauenkupplungen verfügt über eine optimierte Zahngeometrie, die an die Anforderungen der aktiven Synchronisierung angepasst ist. Hierbei ist insbesondere das Einspurspiel ein relevanter Parameter, da dieses so klein wie möglich sein soll und gleichzeitig ausreichend groß sein muss, um ein sicheres Einlegen zu ermöglichen.
Zusätzlich ist die Schalteinheit durch eine Reduktion der bewegten Masse und Anpassung der Arretierungskräfte auf den im folgenden Kapitel vorgestellten Schaltaktor abgestimmt.

2.3 Schaltaktor

In **Bild 3** ist der dedizierte, smarte Speed4E-Schaltaktor als CAD-Schnittdarstellung inklusive der Schaltstange, Schaltgabel und des Positionssensor dargestellt.

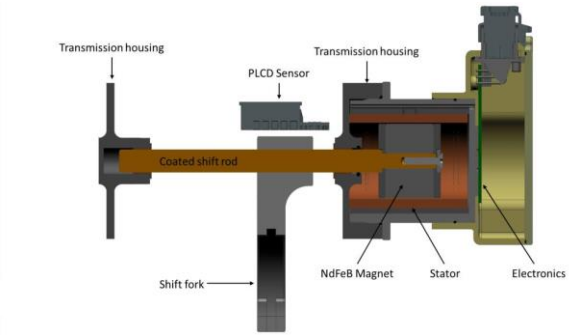


Bild 3 CAD-Schnittdarstellung der integrierten Hardware [5]

Wie in [5] ausführlich beschrieben, handelt es sich um einen Zweispulen, permanenterregten Tauchspulenaktor. Er arbeitet nach dem elektrodynamischen Prinzip, ist damit bidirektional wirkend und hat ein sehr lineares Kraft-Weg-Verhalten. Insbesondere im Vergleich zu Aktoren, die nach dem Reluktanzprinzip arbeiten (Elektromagneten) zeichnet er sich damit durch eine gute Regelbarkeit und einfachen mechanischen Aufbau aus. Die Aktivteile

sind durch eine Pareto-Optimierung auf die Anforderungen und Randbedingungen des Speed4E-Antriebs angepasst. [5]
Die im folgenden Unterkapitel beschriebene Elektronik ist direkt an den Aktor montiert, was mechanische und elektrische Schnittstellen reduziert.

2.3.1 Elektronik und Software

Durch die Integration der in **Bild 5** dargestellten Elektronik in den Aktor wird ein sogenannter „Smarter Aktor" realisiert. Gangschaltungen werden von diesem eigenständig durchgeführt, sobald dies von einer übergeordneten Schaltablaufsteuerung gefordert wird.

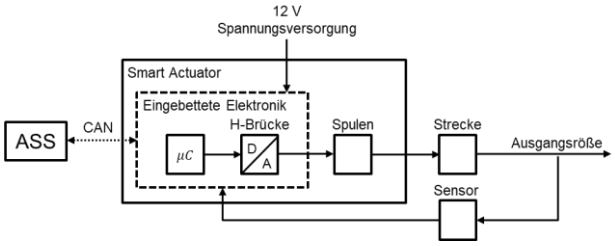


Bild 4 Blockschaltbild der eingebetteten Elektronik

Bild 4 zeigt das Blockschaltbild des Regelkreises zur Positionsregelung der Schaltgabel. Kernelemente der eingebetteten Elektronik sind der Microcontroller und die H-Brücke zur Bestromung der Spulen. Der Spulenstrom führt zu einer Aktorkraft, welche zu einer Bewegung der Schaltgabel führt. Diese wird von dem Positionssensor gemessen und an den Microcontroller zurückgeführt, in dem der Regelkreis geschlossen wird. Der Positionssensor wird direkt von der Elektronik des Aktors mit Spannung versorgt. Die einzigen Schnittstellen nach außen sind damit eine CAN-Schnittstelle zur Antriebssystemsteuerung (ASS) und die Spannungsversorgung.
Die fertige, manuell bestückte Platine ist in **Bild 5** dargestellt und deren Aufbau im Folgenden beschrieben.

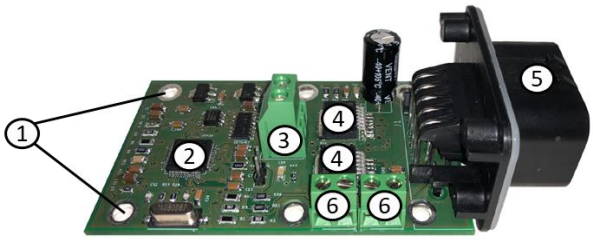


Bild 5 Elektronikplatine des Aktors

Über die Verschraubungspunkte (1) wird die Platine am Aktor befestigt und gleichzeitig eine Masseverbindung zu diesem und damit zur Fahrzeugmasse hergestellt. Dies reduziert nicht nur die elektrischen Schnittstellen weiter, sondern verhindert auch eine Masseschleife über die Aktor-Elektronik. Ein Microcontroller (2) der Baureihe STM32F4 ist möglichst weit entfernt von den Halbbrücken (4) positioniert, um elektromagnetische Störungen gering zu halten. Die Spulen des Aktors werden über den Klemmblock (3) an die Platine angeschlossen. Die zusätzlichen Klemmblöcke (6) sind für Temperatursensoren vorgesehen. Ein zentraler, wasserdichter Stecker (5) wird für die Spannungsversorgung und die CAN-Kommunikation genutzt. Er stellt außerdem die Schnittstelle zum Sensor dar und wird zur Programmierung des Microcontrollers genutzt.

3. Systemtests

Das in Kapitel 2 beschriebene Schaltungssystem wird im Rahmen des Speed4E-Projektes einer Vielzahl von Systemtests unterzogen. Deren Ziel ist es Verbesserungspotentiale auf dem Weg zur realen Anwendung im Versuchsfahrzeug aufzudecken. In diesem Beitrag werden Systemtests an einem Schaltungsprüfstand am Institut für Mechatronische Systeme im Maschinenbau (IMS) der TU Darmstadt beschrieben. Ausblickend wird das System an Antriebssystemprüfständen und im Fahrzeug getestet, worauf in Kapitel 4 abschließend eingegangen wird.

3.1 Schaltungsprüfstand InnoShift

Zur detaillierten Untersuchung des Schaltungssystems wird ein eigener Prüfstand, der sogenannte „InnoShift", aufgebaut. Unabhängig von anderen Komponenten des Antriebsstrangs kann auf diesem schon in einer frühen Entwicklungsphase begonnen werden das reale Schaltungssystem zu untersuchen und die Regelung des Schaltaktors zu applizieren.

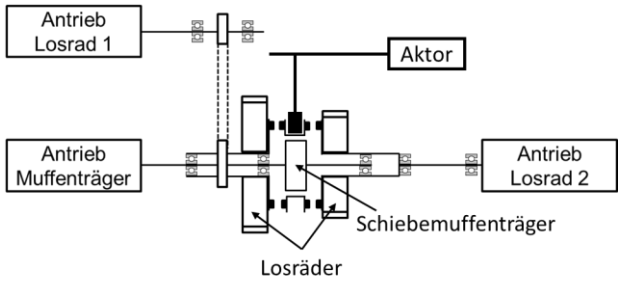


Bild 6 Vereinfachte Darstellung des InnoShift

Bild 6 zeigt stark vereinfacht den Aufbau des InnoShift. Anstelle der Verwendung einer kompletten Getriebestufe werden die Losräder und der Schiebemuffenträger der Schalteinheit individuell angetrieben, sodass die Differenzdrehzahl unabhängig von der Absolutdrehzahl vorgegeben werden kann.
Der Aufbau des Prüfstands ist so modular gestaltet, dass neben dem in dieser Arbeit beschriebenen Schaltungssystem auch andere Systeme getestet werden können. Für Untersuchungen von Schaltvorgängen unter Differenzdrehzahl lassen sich Ersatzträgheiten anbringen, um ein realistisches Schaltverhalten, insbesondere im Hinblick auf die Einlegewahrscheinlichkeit, zu erzielen.
Zur Untersuchung der winkelgenauen Synchronisierung verfügen alle Antriebe über Inkrementalgeber mit denen der Winkel und die Drehzahl bestimmt werden kann. Im späteren Fahrzeugantriebsstrang werden diese über die Resolver der elektrischen Maschinen berechnet. Alle Antriebe bestehen aus Getriebemotoren mit eigenen Motorsteuerungen. Diese kommunizieren über CAN mit einem zentralen Echtzeitsystem, über das der Aktor auch direkt

angesteuert werden kann. Auf diese Weise lassen sich beliebige Schaltmanöver durchführen, wovon in den folgenden Kapiteln ausgewählte vorgestellt werden.
Alternativ kann auch die ASS eingebunden werden und das Echtzeitsystem für eine Restbussimulation genutzt werden. So kann auch die Schaltablaufsteuerung im Verbund mit dem Schaltungssystem vorab untersucht werden. **Bild 7** zeigt das Schaltungssystem montiert am InnoShift.



Bild 7 Dedizierter Schaltaktor am InnoShift

3.2 Applikation und Test im Stillstand

Der Schaltaktor wird zunächst bei Stillstand der Klauenkupplung und einer manuell ausgerichteten Klaue-auf-Lücke-Stellung in Betrieb genommen.
Die mit Hilfe eines Labormusters entworfene Programmstruktur kann in weiten Teilen übernommen werden. Lediglich Reglerparameter und die Parameter der neuen Klauenkupplung, wie etwa der reduzierte Schaltweg, müssen angepasst werden.
Aufgrund der guten Regelbarkeit des Gesamtsystems kann mit einem einfachen PDT1-Regler bereits ein gutes Schaltergebnis im Stillstand erzielt werden. Beispielhaft hierfür ist die Messung eines Schaltvorgangs in den ersten Gang in **Bild 8** dargestellt.

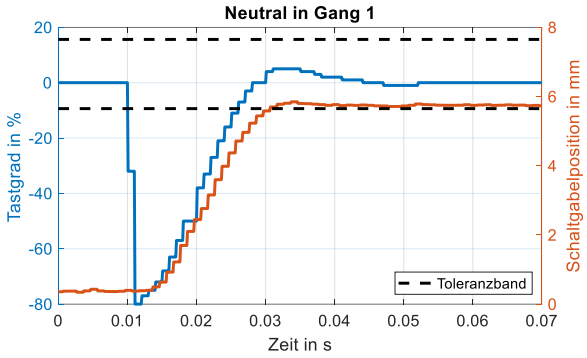


Bild 8 Schaltgabelposition und Tastgrad beim Gangwechsel von Neutral in den 1. Gang im Stillstand

Die Stufenform ist durch die digitale Übertragung über CAN mit einer Zykluszeit von 1 ms an die Auswerteeinheit begründet. Die Regelfrequenz der Positionsregelung ist mit 10 kHz jedoch deutlich höher, wodurch die kurzen Schaltzeiten ermöglicht werden können.
Eine Analyse der Schaltvorgänge im Stillstand zeigt bereits das Potential des entwickelten Schaltungssystems. Die Schaltdauer kann gegenüber der konventionellen elektromechanischen Schaltwalzenaktorik des Speed2E-Projektes von 500 ms [3] auf 20 ms reduziert werden. Um zu untersuchen, ob die Reglerparameter robust gewählt wurden müssen weitere Systemtests durchgeführt werden, die im Folgenden beschrieben werden.

3.3 Applikation und Test des winkelgenauen Einlegens

Im Speed4E-Antrieb wird bei der aktiven Synchronisierung nicht nur die Drehzahl synchronisiert, sondern die Klauenkupplung gezielt auf eine Klaue-Lücke-Stellung gebracht, sodass ohne Kontakt der Zahnflanken oder der Zahndächer der Gang eingelegt werden kann. Die Regelung dieser Synchronisierung wird im Speed4E-Antriebsstrang direkt auf der Leistungselektronik der Traktionsmaschinen EM2 umgesetzt, um Kommunikationslatenzen gering zu halten und eine hohe Regelfrequenz zu erzielen. [4]
Am InnoShift wird die Methodik der winkelgenauen Synchronisierung zunächst bei niedrigen Absolutdrehzahlen untersucht, um Beschädigungen der Klauenkupplung bei Schaltfehlern zu vermeiden. Nachdem zunächst die Synchronisierung am InnoShift vorgestellt wird, werden anschließend winkelsynchrone Ein- und Auslegevorgänge diskutiert.

3.3.1 Drehzahl- und Winkelsynchronisierung

Wie in **Bild 9** dargestellt, wird für die Untersuchung der Synchronisierung zunächst eine konstante Drehzahl des Abtriebs (Schiebemuffenträger) vorgegeben, was einer Konstantfahrt des Fahrzeugs entspricht. Die Drehzahl des Losrades von Gang 1 ist zunächst Null, was einem Stillstand der EM2 im Speed4E-Antrieb entspricht. Bei etwa 0,25 s wird die aktive Synchronisierung gestartet. Die Zieldrehzahl der Drehzahlregelung entspricht der aktuellen Drehzahl des Abtriebs.

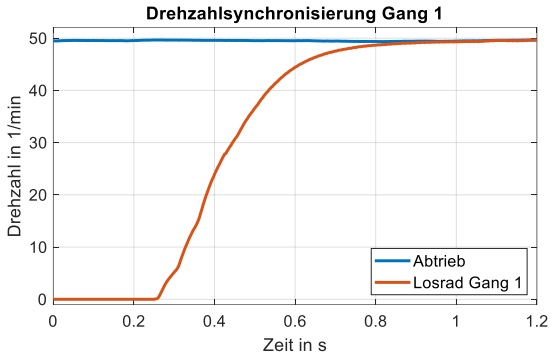


Bild 9 Drehzahlverlauf bei aktiver Synchronisierung des Abtriebs (Muffenträger) und des Losrades von Gang 1

In **Bild 10** ist zusätzlich der Differenzwinkel zwischen dem Zahn der Muffe und dem Kupplungskörper des ersten Gangs dargestellt, was ausführlich ebenfalls in [4] erläutert ist. Diese Synchronisierung auf die nächstmögliche Zahnlücke führt im Vergleich zu der in [6] gezeigten Methoden zu einer Verkürzung des Synchronisierungsvorgangs. Das in **Bild 10** gekennzeichnete Toleranzband entspricht gerade dem Einspurspiel der Klauenkupplung. Sowohl die Drehzahl- als auch die Winkelsynchronität

wird mit Hilfe eines Algorithmus der Leistungselektronik festgestellt. Hierzu darf insbesondere der Differenzwinkel ein Toleranzband für eine bestimmte Zeit nicht verlassen werden. Erst dann meldet die Leistungselektronik der Antriebssystemsteuerung „Synchronität“ zurück. Daraufhin fordert diese vom Schaltaktor das Einlegen des Gangs. Damit ist der Einlegevorgang aus Sicht des Schaltaktors bei Synchronität identisch zu dem im Stillstand.
Nur bei mangelnder Synchronität kommt es zu einem Flankenkontakt und damit zu einem höheren Kraftbedarf, oder schlimmstenfalls zu einer Dach-auf-Dach-Stellung.

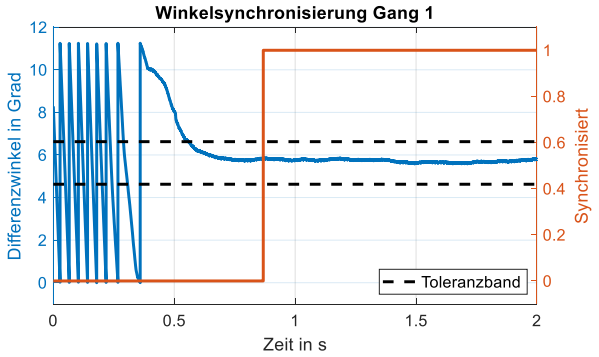


Bild 10 Zeitlicher Verlauf des Differenzwinkels beim Synchronisieren von Gang 1

3.3.2 Winkelsynchroner Einlegevorgang

Ein winkelsynchroner Einlegevorgang ist in **Bild 11** dargestellt. Zu Beginn des Einlegevorgangs bei 0,01 s wird die maximale Stellgröße, die über den Tastgrad der H-Brücke ausgedrückt wird und maximal ±100 % betragen kann, fast vollständig ausgenutzt. Dadurch werden die Spulen kurzzeitig stark bestromt und die maximale Aktorkraft wird ausgeschöpft. Da kein Flankenkontakt stattfindet, tritt die Schaltgabelposition, wie auch bei dem Versuch im Stillstand, bereits nach rund 20 ms in das schwarz gestrichelt dargestellte Toleranzband ein.

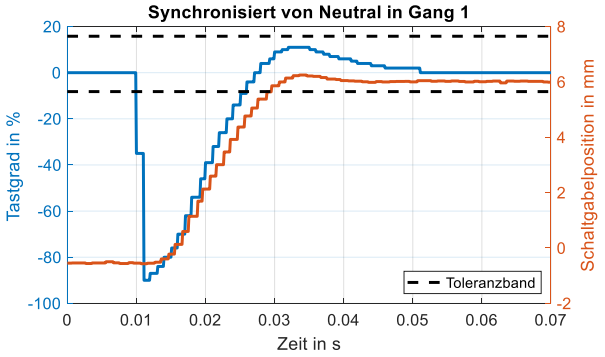


Bild 11 Winkelsynchroner Einlegevorgang

Wenn sich die Schaltgabel innerhalb dieses Toleranzbandes befindet, ist der Zielgang sicher eingelegt, da dann die Schiebemuffe weit genug eingerückt ist, um das volle Moment übertragen zu können. Die Schiebemuffe wird nicht durch den Aktor an den Endanschlag gedrückt, was zu hohen Geräuschemissionen und einer Belastung der Schaltgabel, Schaltstange und Lager führen würde. Sie wird lediglich durch Arretierungs- und Axialkräfte, hervorgerufen durch den Klauenhinterschnitt, an den Endanschlag gezogen. Das Toleranzband darf für 20 ms nicht verlassen werden, bevor der Aktor „Gang eingelegt“ an die Antriebssystemsteuerung zurückmeldet.

3.3.3 Winkelsynchroner Auslegevorgang

Der Auslegevorgang findet auf ähnliche Weise statt. Sobald die Klauenkupplung im Idealfall vollständig momentenfrei ist, wird der Gang durch den Aktor ausgelegt. Auch hier wird, wie in **Bild 12** dargestellt, bereits nach rund 20 ms in das Toleranzband eingetreten.

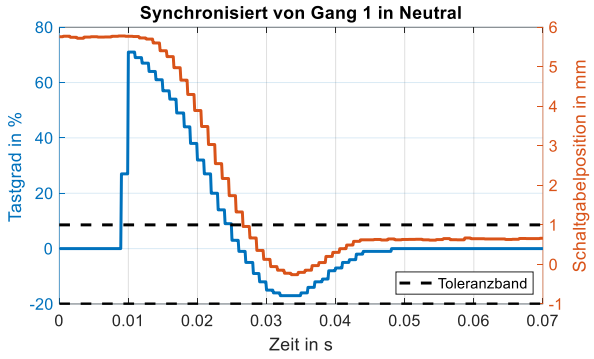


Bild 12 Winkelsynchroner Auslegevorgang

Das leichte Überschwingen führt hier zu keinen Geräuschemission, da kein mechanischer Anschlag vorhanden ist. Ein Überschwingen von über 2 mm würde jedoch zu einem Zahnkontakt mit dem nicht synchronisierten Gang führen, was unbedingt vermieden werden muss. Der bidirektional wirkende Aktor, zusammen mit der mechanischen Mittenarretierung der Klauenkupplung, ermöglicht dies zuverlässig mit dem hier genutzten PDT1-Regler.

3.4 Auslegevorgang bei Restmoment

Ziel der aktiven Synchronisierung ist es, dass vor jedem Auslegevorgang die Klauenkupplung vollständig entlastet wird. Ist dies der Fall wirkt, wie im Stillstand, lediglich die Arretierungskraft auf die Klauenkupplung. Der Schaltakor ist jedoch so dimensioniert, dass bei einem gewissen Kupplungsmoment der Gang trotzdem ausgelegt werden kann. In dieser Situation ist zu vermeiden, dass das Überschwingen im Auslegevorgang so groß wird, dass ein Zahnkontakt zwischen Muffe und Kupplungskörper des noch nicht synchronisierten Gangs auftritt. Dies wäre insbesondere bei der Verwendung eines linearen Reglers mit I-Anteil zu befürchten, kann aber auch durch integrierende oder verzögernde Anteile der Strecke z. B. durch die Induktivität der Spule hervorgerufen werden.
In **Bild 13** ist der Tastgrad und die Position der Schaltgabel für zwei Messungen dargestellt. Messung 1 zeigt einen gerade noch erfolgreichen Auslegevorgang. Bei Messung 2 ist das Kupplungsmoment zu hoch und der Gang wird nicht vollständig ausgelegt, da der Auslegeversuch zum Schutz der Elektronik nach einer gewissen Zeit abgebrochen wird.

Der gewählte PDT1-Regler führt also nicht zu dem befürchteten starken Überschwingen bei nicht vollständiger Momentenfreiheit der Klauenkupplung. Er führt jedoch auch nicht zu einem robusten Auslegen des Gangs bei Restmoment der Klauenkupplung. Bei Bedarf ist hierfür ein alternativer Regler zu entwickeln.

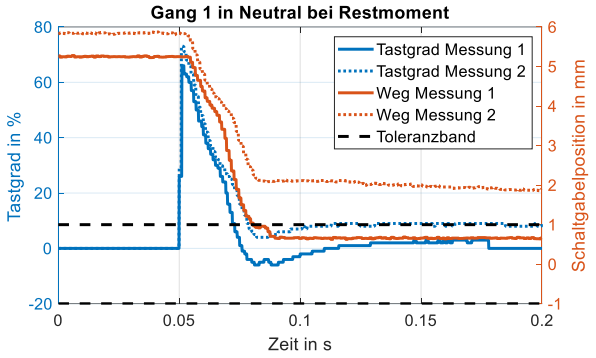


Bild 13 Erfolgreicher und nicht erfolgreicher Auslegevorgang bei Restmoment

4. Zusammenfassung und Ausblick

Im Speed4E-Projekt wird ein neuartiges Schaltungssystem dediziert für mehrgängige elektrische Antriebe entwickelt. Mit einem Funktionsmuster werden an einem Schaltungsprüfstand Systemtests durchgeführt und analysiert. Hierbei kann die Funktionsfähigkeit des Systems unter Laborbedingungen nachgewiesen werden. Insbesondere im Hinblick auf die erreichbaren Ein- und Auslegezeiten kann das System konventionelle Schaltungssysteme deutlich unterbieten.

Alle bisherigen Untersuchungen wurden bei Raumtemperatur durchgeführt. In einem nächsten Schritt wird am IMS das Verhalten des Aktors bei anderen Temperaturen untersucht. Bei einer Erhöhung der Temperatur ist ein höherer Spulenwiderstand und damit eine niedrigere Aktorkraft bei gleicher Spulenspannung zu erwarten. Um dies zu untersuchen wird ein kompakter Stellvertreterprüfstand, der nur Ein- und Auslegen der Gänge bei Stillstand ermöglicht, genutzt. Dieser kann vollständig temperiert werden.

Im Anschluss wird das Schaltungssystem im Verbund mit dem Speed4E-Antrieb untersucht. Hierzu werden zunächst Tests an Antriebssystemprüfständen der Leibniz Universität Hannover und der TU München durchgeführt.

Abschließend wird das Antriebssystem in ein Versuchsfahrzeug von BMW integriert, um mit diesem Untersuchungen auf dem Testfeld der TU Darmstadt durchzuführen.

5. Literatur

[1] Viehmann, A.; Schleiffer, J.-E.; Rinderknecht, S.: *Evaluation of the Dedicated Range-Extender Transmission Powertrain Concept DE-REX regarding Efficiency, Costs and Complexity*. In: 19. Internationaler VDI-Kongress, “Getriebe in Fahrzeugen 2019“, Bonn 2019

[2] R. Fischer, *Das Getriebebuch*. Wien: Springer Verlag, 2012.

[3] König, R.; Reul, A.; Rinderknecht, S.: *Abschlussbericht des Projekts Speed2E*, Darmstadt 2017

[4] Schöneberger, D.; Rinderknecht, S.; Reitmeier, D.; Mertens, A.: *Development of a Transmission Control for an Innovative High-speed Powertrain using Motor-related Control*. In: 20. Internationaler VDI-Kongress, “Getriebe in Fahrzeugen 2020“, Bonn 2020

[5] Schöneberger, D.; Mileti, M.; Stahl, K.; Rinderknecht, S.: *Development of an Innovative Shift Actuator for Electrified Multispeed Transmissions*. In: International Conference on Advanced Vehicle Powertrains, Hefei China 2019

[6] Taylor, R., Fracchia, M.: *Integrated Multispeed 48V EV Powertrain*. In: 16th International CTI Symposium Automotive Transmission, Berlin 2017

Spezifikationsmethodik für optomechanische Systeme in der Nanolithografie

Dr.-Ing. Dipl.-Wirt. Ing. Marwène Nefzi, Dr. rer. nat. Stefan Hembacher, Dr.-Ing. Jens Kugler, Carl Zeiss SMT GmbH, Rudolf-Eber-Strasse 2, 73447 Oberkochen

Kurzfassung

Die Nanolithografie ist das Verfahren zur Herstellung von Mikrochips für Laptops, Smartphones und sonstige moderne Industrie- und Haushaltsgeräte. Die Funktionsgruppen dieser Mikrochips bestehen aus elektronischen Bauelementen wie Transistoren oder Kondensatoren. Diese werden von einem sogenannten Reticle mithilfe eines Objektivs auf eine mit Photolack beschichtete Siliziumscheibe, den Wafer, projiziert. Schnellere Prozessoren und größere Speicherkapazitäten bedingen eine größere Packungsdichte der elektronischen Bauelemente und damit immer kleinere Strukturgrößen. Um Strukturgrößen < 20 nm ausreichend scharf abbilden zu können, arbeiten moderne Lithographiesysteme mit extrem Ultraviolettem-Licht (EUV) bei einer Wellenlänge von 13,5 nm [1].
Abbildung 1 zeigt den prinzipiellen Aufbau eines solchen EUV-Scanners. Das in der Quelle (1) erzeugte EUV-Licht beleuchtet über das Beleuchtungssystem (2) das Reticle (3). Die dort vorhandene Struktur wird mithilfe der Spiegel im Objektiv (4) auf den Wafer (5) abgebildet.
Neben der Einhaltung von hohen Stabilitätsanforderungen [2-5] ist eine multidisziplinäre Spezifikationsmethodik, die die Lücke zwischen Produkt- und Komponentenspezifikationen schließt, entscheidend für den erfolgreichen Einsatz von optomechanischen Systemen in der EUV-Lithografie. In diesem Zusammenhang hat sich der iterative Top-Down/Bottom-Up Ansatz durchgesetzt. Er besteht unter anderem darin, ausgehend von einem Gesamtfehlerbudget für die zulässigen optischen Aberrationen auf Systemlevel eine Fehlerbudgetierung für verschiedene Fachbereiche und Schlüsselkomponenten vorzunehmen. Dabei ist das zu entwickelnde System noch nicht vollständig bekannt, so dass sich der Systemingenieur auf Skalierungen und einfachen Modellen verlassen muss. In diesem Beitrag soll ausgehend von der Anforderung nach minimalen Abbildungsfehlern auf dem Wafer die Herleitung von Spezifikationen an Manipulatoren von optischen Elementen verdeutlicht werden.

1 Einführung

Die Nanolithografie ist das Verfahren zur Herstellung von Mikrochips für Laptops, Smartphones und sonstige moderne Industrie- und Haushaltsgeräte. Die Funktionsgruppen dieser Mikrochips bestehen aus elektronischen Bauelementen wie Transistoren oder Kondensatoren. Diese werden von einem sogenannten Reticle mithilfe eines Objektivs auf eine mit Photolack beschichtete Siliziumscheibe, den Wafer, projiziert. Schnellere Prozessoren und größere Speicherkapazität bedingen eine größere Packungsdichte der elektronischen Bauelemente. Um Strukturgrößen < 20 nm ausreichend scharf abbilden zu können, arbeiten moderne Lithographiesysteme mit extrem Ultraviolettem-Licht (EUV) bei einer Wellenlänge von 13,5 nm [1-5]. **Bild 1** zeigt den prinzipiellen Aufbau eines solchen EUV-Scanners. Das in der Quelle (1) erzeugte EUV-Licht beleuchtet über das Beleuchtungssystem (2) das Reticle (3). Die dort vorhandene Struktur wird mithilfe der Spiegel im Objektiv (4) auf den Wafer (5) abgebildet.



Bild 1 Prinzipieller Aufbau eines EUV Lithographie-Scanners. Quelle: www.asml.com

EUV-Objektive, die für eine realistische Serienfertigung von Chips ausgelegt sind, sind typischerweise 1 bis 2 m groß und 1 t schwer. Die abzubildenden Strukturen sind jedoch im Nanometerbereich. Der Fehler in der Abbildung darf davon nur einen Bruchteil betragen.
Die Genauigkeitsanforderungen an die optischen Elemente lassen sich am besten anhand von **Bild 2** verdeutlichen. Über die Entfernung Erde-Mond kann ein Punkt mit einer Positionsungenauigkeit von 10 cm getroffen werden

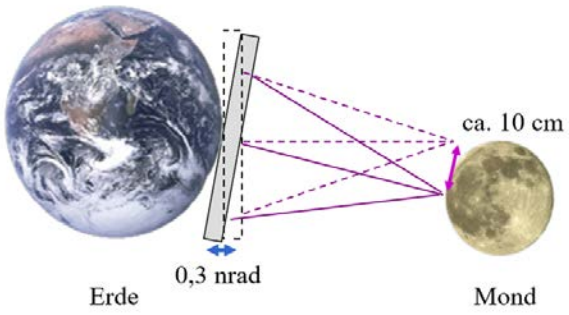


Bild 2 Maximal zulässiger Abbildungsfehler hochskaliert auf den Abstand Erde-Mond

Darüber hinaus zeigt **Bild 3** anhand eines refraktiven Systems wie deformierte optische Elemente die Wellenfront verändern und optische Aberrationen, die zu minimieren gilt, verursachen.

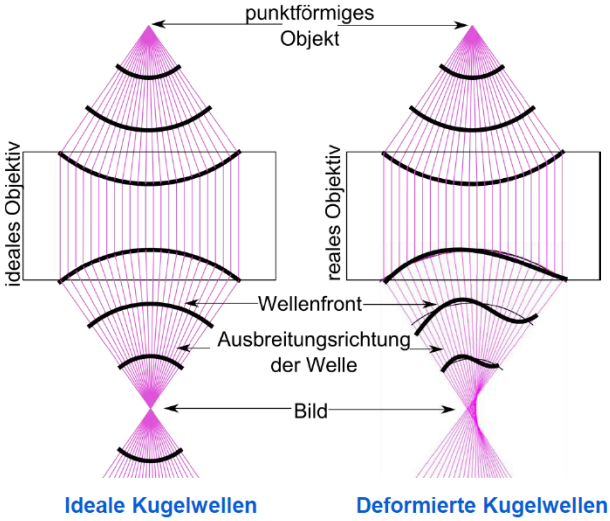


Bild 3 Deformierte optische Elemente verursachen Aberrationen

Anhand dieser einfachen Erläuterung der Genauigkeitsanforderungen an fotolithografische Systeme lassen sich bereits geeignete Beispiele ableiten, die im dritten Abschnitt besprochen werden. Im zweiten Abschnitt wird vorab der iterative Top-Down/Bottom-Up Ansatz erläutert. Es ist hervorzuheben, dass die in diesem Beitrag erläuterte Vorgehensweise nicht EUV spezifisch ist. Sie lässt sich prinzipiell auf alle optomechanischen Systeme erweitern.

2 Iterativer Top-Down/Bottom-Up Ansatz

Der iterative Top-Down/Bottom-Up Ansatz rückt den Spezifikationsprozess und die übergeordnete Rolle des Systemingenieurs in den Vordergrund. Unter Systemingenieur wird in diesem Zusammenhang ein Ingenieur verstanden, der auf technischer Ebene das Hauptziel der Einhaltung der Produktspezifikationen verfolgt, ohne dabei Kosten und Zeitplan völlig außer Acht zu lassen. Im Gegensatz zu einem Projektleiter stehen bei ihm technische Spezifikationen im Fokus. Fachingenieure helfen mit ihrem speziellen Fachwissen, die Komponenten des Produkts entsprechend den Regeln der jeweiligen Fachdisziplinen auszulegen.

Bei der Entwicklung eines komplexen Produkts ist es kaum möglich, dass der Systemingenieur über tiefe Fachkenntnisse in allen involvierten Fachdisziplinen und Schlüsselkomponenten verfügt. Nichtsdestotrotz muss er die Entwicklung mitgestalten, damit die Einhaltung der technischen Spezifikationen gelingt. In diesem Beitrag wird ein iterativer Ansatz vorgeschlagen, um diese Aufgabe zu bewältigen. Diese Methodik wird im Folgenden iterativer Top-Down/Bottom-Up Ansatz genannt. Dabei werden Top-Down Werte aus den Produktspezifikationen heruntergebrochen, während Bottom-Up Werte mit Hilfe der Fachingenieure und anhand eines konkreten Designentwurfs ermittelt werden müssen. Die Top-Dwon Werte sind in diesem Kontext oft ein Fehlerbudget für optische Aberrationen auf dem Wafer. Darunter werden in diesem Beitrag die maximal zulässigen Abbildungsfehler inkl. Overlay-Versatz, der später näher erläutert wird.

Der Systemingenieur gestaltet den Entwicklungsprozess aktiv mit, wenn er zwischen allen Fachdisziplinen ausbalancierte Spezifikationen aufstellt. Das kann nur iterativ erfolgen, siehe **Bild 4**.

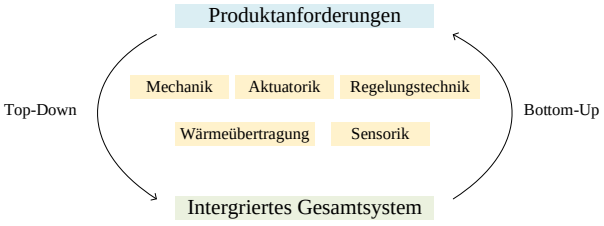


Bild 4 Iterativer Top-Down/Bottom-Up Ansatz

Der erste Schritt dieses Prozesses besteht darin, die bereits in einem frühen Stadium bekannten Produktspezifikation in Form von maximal zulässigen Abbildungsfehlern auf verschiedene Fachdisziplinen und Schlüsselkomponenten herunterzubrechen. Wie oben erwähnt spricht man in diesem Zusammenhang oft von einer Subbudgetierung des Gesamtfehlers. Das Ergebnis sind Subbudgets für verschiedene Komponenten des Produkts. In der Regel erfolgt dieser Schritt basierend auf Expertenwissen, Heuristiken, Vorgängerprodukten und idealerweise auf einer fundierten Abschätzung des zukünftigen Potenzials von verschiedenen Technologien. Bei einem komplexen Produkt ist es außerdem davon auszugehen, dass die Subbudgets im Laufe des Entwicklungsprozesses angepasst werden müssen. Hierfür muss der Systemingenieur die Entwicklung, Konstruktion, Integration, Fertigung usw. sehr nah mitverfolgen und ggfs. eingreifen, wenn klar wird, dass bestimmte Subbudgets nicht einzuhalten sind. Zwei Eingriffsmöglichkeiten sind vorstellbar. Als Erstes kann der Systemingenieur die Subbudgets neu aufstellen und die Fachdisziplinen mit einer deutlichen Überschreitung

der Subbudgets begünstigen, indem Spezifikationen relaxiert werden. Diese Lösung setzt voraus, dass andere Fachdisziplinen die vorgesehenen Subbudgets nicht völlig ausschöpfen. Andernfalls muss der Systemingenieur Verbesserungsfaktoren in den betroffenen Fachdisziplinen festlegen und idealerweise Technologieverbesserungen in Zusammenarbeit mit den Fachingenieuren anstoßen. Darüber hinaus muss er diesen Schritt gegenüber der Projektleitung plausibel darstellen und durchsetzen.
Wichtig hierbei sind neben einem umfassenden Wissen über verschiedene Technologien fundierte Kenntnisse der Methoden und physikalischen Modelle der verschiedenen Fachdisziplinen, mit deren Hilfe schnelle Plausibilitätsüberprüfungen und Anpassungen vorgenommen werden können. Somit wird deutlich, dass der Systemingenieur eine zentrale Rolle bei diesem Ansatz spielt, denn er muss anhand der Subbugetierung erkennen, in welchen Fachdisziplinen und für welche Schlüsselkomponenten Verbesserungen unabdingbar sind. Um die Projektleitung zu überzeugen, muss er ebenfalls in der Lage sein, die Konsequenzen auf die Kosten und den Zeitplan zu benennen.

Im Folgenden wird anhand der Anforderungen nach minimalen Abbildungsfehlern gezeigt, wie der iterative Top-Down/Bottom-Up Ansatz die Lücke zwischen Produktspezifikation und Komponentenspezifikation schließt.

3 Ein Fallbeispiel zur Erläuterung der Vorgehensweise

Der Ausgangspunkt der Entwicklung ist oft ein Budget für optische Abbildungsfehler, das nicht überschritten werden darf. Zwei entscheidende Spezifikationen werden hier umrissen: zum einen ein nicht zu überschreitender Overlay-Versatz und zum anderen Wellenfrontfehler. Dabei bezeichnet der Overlay-Versatz die Überdeckungsgenauigkeit von Strukturen aus verschiedenen fotolithografischen Fertigungsschritten. In der Regel wird ein Overlay-Versatz durch eine statische oder dynamische Starrkörperbewegung der optischen Elemente verursacht. Wellenfrontfehler entstehen hingegen, wenn die optischen Elemente verformt werden, siehe Bild 3.
Im Rahmen des iterativen Top-Down/Bottom-Up Ansatzes erfolgt nun eine Subbudgetierung dieser beiden Spezifikationen, indem das Gesamtfehlerbudget auf alle relevanten Verursacher im Objektiv verteilt wird. **Bild 5** verdeutlicht bildlich, dass nur ein Bruchteil des Fehlerbudgets für die Optomechanik übrigbleibt. Andere Verursacher sind beispielsweise Wärmeausdehnungen, Regelfehler, Beschichtung usw. Eine ausbalancierte Top-Down Subbudgetierung ist entscheidend für ein erfolgreiches Produkt. Zu enge Subbudgets in einer Fachdisziplin führen zu komplexen und kostenintensiven Lösungen. Der Systemingenieur wird dazu angehalten multidisziplinär zu arbeiten, um ausbalancierte Spezifikationen abzuleiten.

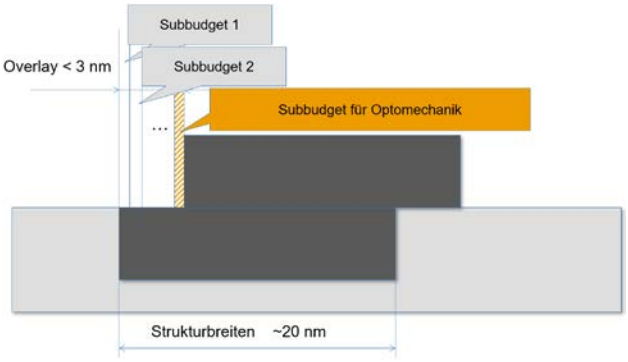


Bild 5 Subbudgetierung des Overlay-Versatzes

Die Herleitung der Subbudgets ist in der Anfangsphase herausfordernd. Das zu entwickelnde System ist noch nicht vollständig bekannt, so dass sich der Systemingenieur wie oben erwähnt auf Heuristiken und Vorgängerprodukten verlassen muss. Zusätzlich sollten einfache multidisziplinäre Modelle entwicklungsbegleitend aufgebaut und gepflegt werden, um Skalierungen und Plausibilitätsprüfungen durchzuführen.
Um einen Overlay-Versatz zu korrigieren und Wellenfrontfehler zu minimieren, müssen Objektive optisch justiert werden. Hierfür bedarf es mindestens der Starrkörperbewegung der optischen Elemente. Eine weitere Möglichkeit zur optischen Korrektur besteht darin, die optischen Elemente zu verformen. Allerdings wird diese Möglichkeit in diesem Beitrag nicht behandelt. Zur Realisierung von Starrkörperbewegungen haben sich 6 DoF (Degree of Freedom für Freiheitsgrad) Manipulatoren durchgesetzt.
An dieser Stelle muss man sich vergegenwärtigen, dass sechs Aktuatoreinheiten notwendig sind, um sechs Starrkörperfreiheiten verstellen zu können. Im Hinblick auf eine Minimierung der Verformung der optischen Elemente und somit der Wellenfrontfehler ist eine Lagerung des optischen Elements an drei Stellen günstig. Aus diesen Überlegungen ergibt sich die schematische Darstellung in **Bild 6**. Auf drei an den Lagerstellen aufgepannten Ebenen werden jeweils zwei Kräfte benötigt, um das optische Element zu halten oder zu verstellen. Alle anderen Kräfte und Momente, die ins optische Element eingeleitet werden, werden parasitär genannt, denn sie verformen das optische Element und sind somit unerwünscht.
Die Auslegung eines Manipulators für ein optisches Element muss mehreren Anforderungen gerecht werden. Basierend auf den Vorüberlegungen in den letzten Abschnitten müssen unter anderem vorgegebene Eigenfrequenzen nicht unterschritten werden, um Verkippungen der optischen Elemente und somit einen Overlay-Versatz zu verhindern. Darüber hinaus müssen die Manipulatoren das optische Element bewegen, ohne es zu verformen. Die Verformung resultiert aus den oben angesprochenen parasitären Kräften und Momenten, die an den Lagerstellen wirken.
Parasitäre Kräfte und Momente entstehen, weil eine ideale Starrkörperbewegung des optischen Elements kaum möglich ist. Eine ideale Bewegung würde mindestens bedin-

gen, dass jede Stützeinheit an jeder Lagerstelle keine Reaktionskräfte und -momente in die parasitären Richtungen verursacht. Das ist in diesem Kontext deshalb nicht realisierbar, weil die Stützeinheiten monolithische Konstruktionen mit Festkörpergelenken sein müssen.

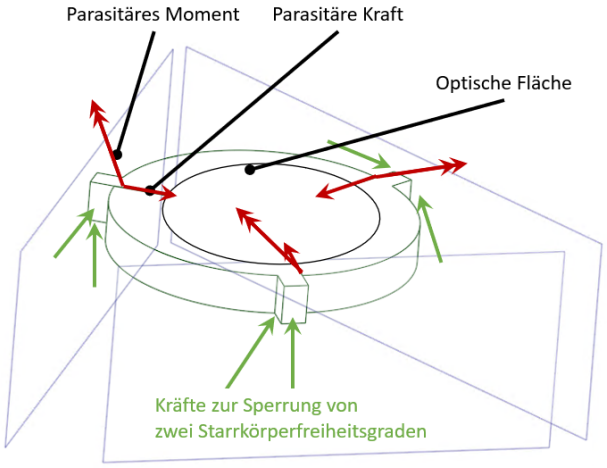


Bild 6 Schematische Darstellung der Steifigkeiten zur Lagerung eines optischen Elements

Die Hauptaufgabe des Systemingenieurs besteht nun darin, ausgehend vom optischen Fehlerbudget für den Overlay-Versatz und Aberrationen höherer Ordnungen (Top-Down Werte) abzuschätzen, welche Spezifikationen an den Manipulator zu stellen sind. Zusätzlich muss der Systemingenieur in der Lage sein, die vorläufigen Manipulatorentwürfe zu evaluieren (Bottom-Up Werte). In beide Richtungen (Top-Down und Bottom-Up) werden einfache Modelle benötigt.
Zuerst wird mit Hilfe der modalen Reduktion ein Modellierungsansatz vorgestellt, der eine Bewertung des Overlay-Versatzes ermöglicht. Anschließend werden Zernike Polynome erläutert, um die optischen Auswirkungen von verformten optischen Flächen infolge von parasitären Kräften und Momenten zu beurteilen.

Modell zur Bestimmung des Overlay-Versatzes

Im letzten Abschnitt ist deutlich geworden, dass eine Minimierung des Overlay-Versatzes mit einer Erhöhung der Eigenfrequenzen des optischen Elements inkl. Manipulator einhergeht. Zur Verdeutlichung wird in **Bild 7** eine externe Anregung auf die Tragstruktur eingeprägt, die zu Schwingungen der optischen Elemente und somit zu einem Overlay-Versatz auf dem Wafer führt. Hierbei wird davon ausgegangen, dass der Wafer der Tragstruktur mit einer hohen Regelbandbreite folgt.

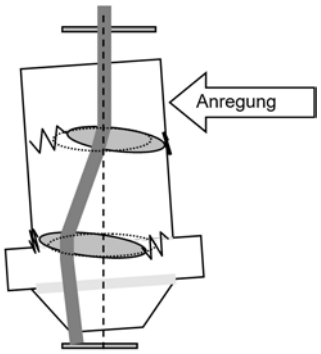


Bild 7: Schwingungen der optischen Elemente infolge einer Anregung der Tragstruktur

Wenn die optischen Elemente zu schwingen anfangen, verschiebt sich das Bild auf dem Wafer. Daraus resultiert ein Versatz zur darunterliegenden Struktur, also ein Overlay-Fehler, siehe **Bild 8**. Die verschiedenen Schichten (Layers), aus denen die Struktur eines Chips besteht, müssen exakt genug übereinander liegen. Es wird nun deutlich, dass eine Erhöhung der Eigenfrequenzen der optischen Elemente entscheidend für die Einhaltung der Produktspezifkationen ist.

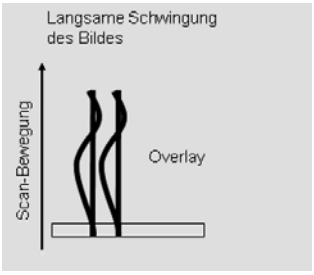


Bild 8: Overlay-Versatz infolge von Schwingungen der optischen Elemente gegenüber dem Wafer

Mit Hilfe eines elastischen Mehrkörpersystems lassen sich die Schwingamplituden der optischen Elemente ausgehend von den Amplituden der Anregungen bestimmen. Ein elastisches Mehrkörpersystem weist typischerweise die folgende Bewegungsgleichung auf

$$M_{ff}\ddot{u}_f + D_{ff}\dot{u}_f + K_{ff}u_f = P_f\,. \qquad \text{(Gl. 1)}$$

Die Massenmatrix M_{ff}, die Dämpfungsmatrix D_{ff} und die Steifigkeitsmatrix K_{ff} in (Gl. 1) ergeben sich aus der Geometrie und den Materialparametern des implementierten Mehrkörpersystems. Sie besitzen die Dimension $f \times f$, wobei f die Anzahl der Freiheitsgrade ist. u_f stellt den Vektor der generalisierten Koordinaten dar. Ähnlich kann der Anregungsvektor P_f als generalisierte Kraft interpretiert werden, die in Koordinatenrichtung wirkt. Die Indizes der Matrizen und Vektoren entsprechen deren Dimensionen, wobei der erste Index als die Anzahl der Zeilen, und der zweite als die Anzahl der Spalten zu verstehen ist.
Mit Hilfe von (Gl. 1) lässt sich die Dynamik der optischen Elemente inkl. Manipulatoren abbilden. Vor allem in der Anfangsphase des Entwicklungsprozesses lassen sich somit die Amplituden und den daraus resultierenden Over-

lay-Versatz relativ schnell ermitteln. Im Laufe des Entwicklungsprozesses steigt der Detaillierungsgrad. (Gl. 1) wird zunehmend komplex und unübersichtlich.
Moderne FEM-Berechnungsprogramme können Hilfe verschaffen. Sie bieten oft die Möglichkeit eine modale Reduktion durchzuführen, so dass der Benutzer lediglich die Vorgehensweise verstehen muss, um diese erfolgreich anzuwenden. Im Folgenden werden die Gleichungen der modalen Berechnung aufgegriffen und an den Kontext angepasst.
Die Anzahl der Freiheitsgrade f in (Gl. 1) kann über fünf Millionen betragen, wenn ein FEM Programm zur modalen Reduktion herangezogen wird.
Die Laplace Transformation der Differentialgleichung (Gl. 1) führt unter Vernachlässigung von Dämpfung zu einem Eigenwertproblem, dessen Lösung reelle Eigenwerte und Eigenvektoren ergibt, die paarweise sowohl zur Massen- als auch zur Steifigkeitsmatrix orthogonal sind. Diese Eigenschaft der Eigenvektoren ermöglicht die Diagonalisierung der letztgenannten Matrizen. Nimmt man zusätzlich an, dass die Dämpfungsmatrix gemäß der Bequemlichkeitshypothese eine Linearkombination der Massen- und Steifigkeitsmatrix ist, so kann auch diese mit Hilfe der Eigenvektoren diagonalisiert werden. Das Ergebnis dieser Diagonalisierung sind f entkoppelte Differentialgleichungen, die f verschiedene Eigenformen des implementierten Mehrkörpersystems beschreiben.
Die modale Reduktion besteht nun darin, die Anzahl der relevanten Eigenschwingungsformen unter Berücksichtigung der erwünschten Modellgenauigkeit zu reduzieren. Die modale Reduktion nimmt somit die Form

$$u_f \approx \Phi_{fh}\gamma_h = \sum_{i=1}^{h} \Phi_{fi}\gamma_i \,, \qquad \text{(Gl. 2)}$$

wobei Φ_{fi} einen Eigenvektor darstellt, der f Komponenten aufweist. Der Vektor γ_h beinhaltet die Modalkoordinaten γ_i. Diese Gleichung besagt, dass der ursprüngliche Koordinatenvektor u_f durch die Summe von $h < f$ unterschiedlich skalierten Eigenvektoren angenähert werden kann.
Setzt man (Gl. 2) in (Gl. 1) ein und multipliziert man letztere mit der Matrix Φ_{fh}^T, die h Eigenvektoren beinhaltet, so erhält man

$$\Phi_{fh}^T M_{ff}\Phi_{fh}\ddot{\gamma}_h + \Phi_{fh}^T D_{ff}\Phi_{fh}\dot{\gamma}_h + \Phi_{fh}^T K_{ff}\Phi_{fh}\gamma_h = \Phi_{fh}^T P_f \,. \qquad \text{(Gl. 3)}$$

Entscheidet man sich für eine kompakte Schreibweise, so lauten die entkoppelten Differentialgleichungen

$$m_{hh}\ddot{\gamma}_h + d_{hh}\dot{\gamma}_h + k_{hh}\gamma_h = P_h, \qquad \text{(Gl. 4)}$$

wobei m_{hh}, d_{hh}, und k_{hh} Diagonalmatrizen sind, die sich aus den modalen Massen, Dämpfungen und Steifigkeiten zusammensetzen. Wenn man nur an den Auswirkungen von bestimmten m Komponenten (Eingangsvariablen) des Anregungsvektors P_h auf bestimmte n Komponenten (Ausgangsvariablen) des Koordinatenvektors u_f interessiert ist, so bietet sich die Zustandsraumdarstellung, die folgendermaßen formuliert werden kann

$$\begin{aligned} \dot{\zeta}_{2h} &= A_{2h\times 2h}\zeta_{2h} + B_{2h\times m}P_m \\ u_n &= C_{n\times 2h}\zeta_{2h} + D_{n\times m}P_m \end{aligned}, \qquad \text{(Gl. 5)}$$

wobei $\zeta_{2h} = \begin{pmatrix}\gamma_h \\ \dot{\gamma}_h\end{pmatrix}$ den Zustandsvektor darstellt. Die Systemmatrix

$$A_{hh} = \begin{pmatrix} 0_{hh} & E_{hh} \\ -m_{hh}^{-1}k_{hh} & -m_{hh}^{-1}d_{hh}\end{pmatrix}$$

setzt sich aus den zuvor eingeführten Matrizen m_{hh}, d_{hh}, und k_{hh} sowie der Einheitsmatrix E_{hh} zusammen. Ähnlich erhält man die Eingangsmatrix

$$B_{2h\times m} = \begin{pmatrix} 0_{hm} \\ \Phi_{fh}^T R_{fm}\end{pmatrix},$$

die Ausgangsmatrix $C_{n\times 2h} = (S_{nf}\Phi_{fh} \quad 0_{nh})$ und die Durchgangsmatrix $D_{n\times m} = 0_{n\times m}$. Die Matrizen R_{fm}, S_{nf} sind binäre Matrizen, die zur Auswahl der *m* Eingangs- und *n* Ausgangsvariablen dienen. Auf diese Weise lässt das zugrundeliegende FE-Modell mit ca. fünf Millionen Freiheitsgraden auf ein übersichtliches Zustandsraummodell reduzieren, dessen Eingangs- und Ausgangsvariablen der gewünschten Anzahl an Aktuatoren, Sensoren und Störgrößenfunktionen entsprechen.
Diese Vorgehensweise bietet den Vorteil, dass die Systemmatrix A_{hh} aufgrund der vorherigen Diagonalisierung der Massen-, Dämpfungs- und Steifigkeitsmatrizen, m_{hh}, d_{hh} und k_{hh}, dünn besetzt ist, so dass deren Weiterverarbeitung, etwa bei der Berechnung der Übertragungsfunktionen

$$H_{n\times m}(s) = C_{n\times 2h}(sE_{hh} - A_{hh})^{-1}B_{2h\times m}, \qquad \text{(Gl. 6)}$$

nicht rechenzeitintensiv ist.
Der Übersicht halber wird (Gl. 6) vereinfacht zu

$$H = C(sE - A)^{-1}B$$

Wenn H der (komplexe) Frequenzgang des Objektivmodells von der Anregung am Freiheitsgrad zur Antwort am Freiheitsgrad (z.B. Verschiebung der Mitte eines optischen Elements) ist, dann berechnet sich die spektrale Leistungsdichte beispielsweise der Schwingung diesen optischen Elements zu

$$PSD_{optischesElement} = |H|^2 PSD_{Anregung}$$

Nun ist vom Schwingungsverhalten der optischen Elemente auf das des Bildes zu schließen. Über die Gesetze der geometrischen Optik kann die (statische) Verschiebung des Bildes, verursacht von einer (statischen) Verschiebung eines einzelnen optischen Elements aus seiner Solllage, berechnet werden. Wegen der Linearität des Modells ist der Effekt von Kombinationen von Verschiebungen optischer Elemente die Summe der Einzeleffekte. Damit lässt sich jedem optischen Element eine optische Sensitivität zuordnen und der Frequenzgang der Verschiebung des Bildes $H_{Bild,j}(\omega)$ als Antwort auf eine Anre-

gung am Freiheitsgrad j lässt sich als Linearkombination der Frequenzgänge der Verschiebungen der optischen Elemente, multipliziert mit ihrer optischen Sensitivität darstellen:

$$H_{Bild,j}(\omega) = \sum_k s_k H_{kj}(\omega) \text{, mit } k = 1...N \text{ .}$$

N ist die Anzahl der Freiheitsgrade der optischen Elemente und s_k ist die optische Sensitivität eines Linsenfreiheitsgrads.
Im Zeitbereich lässt sich der Overlay-Fehler als Moving Average (MA) der Bildschwingung bestimmen. Im Frequenzbereich lässt sich der MA-Wert ermitteln, indem man den Frequenzgang $H_{Bild,j}(\omega)$ zunächst mit einem Low Pass Filter multipliziert und anschließend den 3σ -Wert berechnet.
Nun verfügt der Systemingenieur über ein Modell, mit Hilfe dessen nicht nur Bottom-Up Werte ermittelt werden können sondern auch eine Verbindung zwischen optischen Produktspezifikationen und Komponentenspezifikationen hergestellt werden können. Ein weiterer Vorteil dieser Vorgehensweise ist die Möglichkeit, das Modell im Laufe des Entwicklungsprozesses kontinuierlich anzupassen und zu aktualisieren, um Iterationen zu ermöglichen.

Anforderung an die parasitären Steifigkeiten

Wie oben erwähnt entstehen bei einer Starrkörperbewegung zwecks einer optischen Justage parasitäre Kräfte und Momente, die das optische Element verformen. An jeder der drei Lagerstellen sind eine radiale Kraft und drei Momente parasitär. Vorteilhaft bei der Bestimmung der Wellenfrontfehler und der Herleitung der Spezifikationen an die parasitären Steifigkeiten ist die Kenntnis der Reaktion des optischen Elements auf jede parasitäre Kraft und jedes parasitäre Moment. Auf diese Weise lassen sich verschiedene Kombinationen an Kräften und Momenten aus Grundlastfällen durch lineare Superpositionen rekonstruieren. Bei kleinen Bewegungen der optischen Elemente im Submillimeterbereich trifft die Linearitätsannahme oft zu. Der Systemingenieur hat somit lediglich die Aufgabe für jede parasitäre Richtung eine Spezifikation abzuleiten. Zuvor ist allerdings eine Verbindung zu den optischen Produktspezifikationen herzustellen.
Zernike Polynome werden oft zur Untersuchung von Wellenfrontfehlern herangezogen. Sie sind gerade und ungerade orthogonale Polynome. Gerade Polynome sind durch die folgende Formel

$$Z_n^m(\rho, \phi) = R_n^m(\rho)\ \cos(m\,\phi)$$

und die ungeraden durch die Formel

$$Z_n^{-m}(\rho, \phi) = R_n^m(\rho)\ \sin(m\,\phi)$$

definiert. Dabei sind ϕ der azimutale Winkel und ρ der normierte radiale Abstand. n und m sind nichtnegative ganze Zahlen, für die gilt n≥m. Die Radialpolynome sind folgendermaßen definiert:

$$R_n^m(\rho) = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k\,(n-k)!}{k!\,((n+m)/2-k)!\,((n-m)/2-k)!}\,\rho^{n-2k}, \text{wenn n − m gerade ist}$$

$$R_n^m(\rho) = 0, \text{wenn n-m ungerade ist}$$

In **Bild 9** sind die Zernike Polynome nach Art der resultierenden optischen Abbildungsfehlern sortiert.

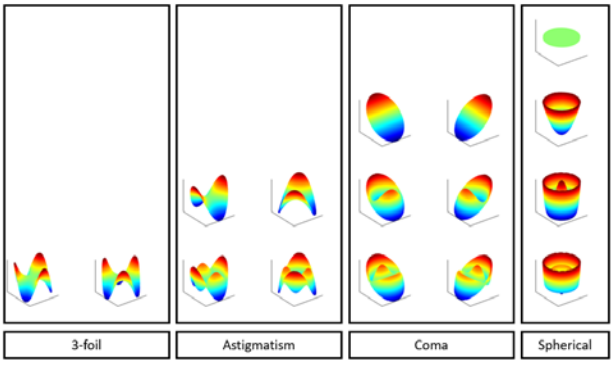


Bild 9: Zernike Polynome sortiert nach optischen Aberrationen

Mit Hilfe der Zernike Polynome lässt sich jede beliebige Fläche durch eine Summe von Zernike Polynomen beschreiben, siehe **Bild 10**. Insbesondere lassen sich verformte optische Elemente damit untersuchen.

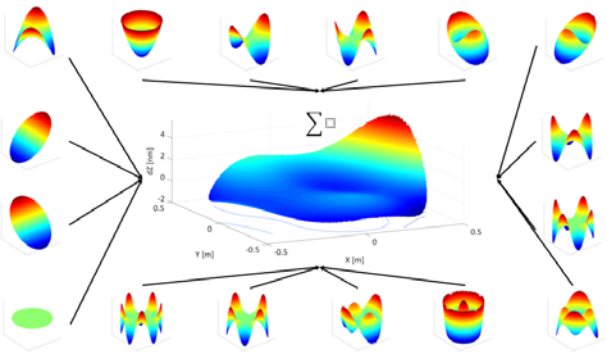


Bild 10 Beschreibung einer Fläche durch Zernike Polynome

Bild 11 zeigt die Verformung eines optischen Elements aufgrund eines Tangentialmoments an einer Lagerstelle. Diese Verformung kann nun als eine Summe von Zernike Polynomen beschrieben werden. Kennt man zusätzlich die Sensitivitäten auf die Abweichungen von den idealen Flächen der optischen Elemente, so gelingt es, die resultierenden optischen Aberrationen infolge eines Tangentialmoments an der optischen Fläche zu bestimmen. Diese Vorgehensweise lässt sich auf alle Grundlastfälle (radiale parasitäre Kraft, radiales parasitäres Moment und axiales parasitäres Moment) erweitern.

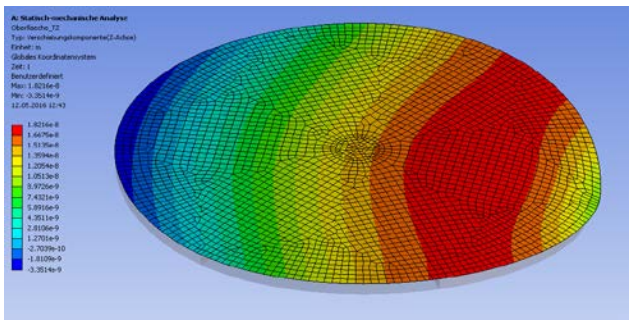


Bild 11: Verformtes optisches Element infolge eines Tangentialmoments

Auf diese Weise lassen sich sowohl Top-Down Subbudgets (ausgehend von optischen Produktspezifikationen) als auch Bottom-Up Werte (ausgehend von einem bekannten Manipulatorentwurf) ermitteln. Verschiedene Arten von optischen Aberrationen wie Defokus, Astigmatismus, sphärische Aberrationen usw. können nun separiert werden und den einzelnen Verursachern zugeordnet werden. Diese Methodik vereinfacht den Spezifikationsprozess erheblich. Vor allem wird mit diesem Werkzeug eine Brücke zwischen der Mechanik und Optik hergestellt, die verschiedene Top-Down/Bottom-Up Iterationen ermöglicht.
Dieses Werkzeug kann also die iterative Vorgehensweise unterstützen und dabei helfen, ausbalancierte Spezifikationen aufzustellen und im Laufe der Entwicklung anzupassen.

5 Zusammenfassung

In diesem Beitrag wird eine auf optomechanische Systeme zugeschnittene Spezifikationsmethodik vorgeschlagen. Diese Methodik ist iterativ und ermöglicht eine Verbindung zwischen Optik und Mechanik herzustellen, so dass der Systemingenieur den Überblick über das neu zu entwickelnde Produkt nicht verliert, ohne dabei sehr tief in die Detailebene einzudringen.
Zur Anwendung dieser Methodik werden zwei Verfahren aufgegriffen und im Kontext der Lithografie näher erläutert. Zum einen kann man mit Hilfe der modalen Reduktion den Overlay-Versatz bestimmen, zum anderen werden Zernike Polynome herangezogen, um die infolge von parasitären Kräften und Momenten entstandenen Wellenfrontfehler zu bewerten. Beide Verfahren ermöglichen Iterationen zwischen Top-Down und Bottom-Up Untersuchungen, um zwischen allen Fachdisziplinen ausbalancierte Spezifikationen und ein optimales Produkt zu gewährleisten.

6 Literatur

[1] Tilmann Heil, Martin Lowisch: *EUV lithography – the future of chip manufacturing*. Photonik international, 2008/2, S. 70-73

[2] Wengert, N.; Nefzi, M.; Eberhard, P.; Geuppert, B.: *Dynamics in lithographic projection objectives*. Multibody System Dynamics 30, 233–245 (2013).

[3] Geuppert, B.; Nefzi, M.; Wagner, F.: *Modellierung des dynamischen Systemverhaltens von Objektiven für die EUV-Mikrolithographie*. In: VDI-Fachtagung "Schwingungsanalyse und Identifikation 2013"

[4] Nefzi, M.; Wagner, F.; Baumann, K.: *A design approach to mechanical shock mitigation in optomechanical systems*. In: Proceedings of ISMA 2014

[5] Nefzi, M.: *The Demanding Implications of Nano-Lithography for High-Precision Opto-Mechanical Systems*. In: New Advances in Mechanisms, Mechanical Transmissions and Robotics, 2016.

Mechatronischer Entwicklungsprozess eines automatischen Sensorfahrzeugs für den Außeneinsatz

Mechatronic Development Process of a Robotic Sensor Carrier in Harsh Environments

Dr.-Ing. Harald Staab, ABB AG, Forschungszentrum Deutschland, harald.staab@de.abb.com; Eduardo Botelho ABB Ltd. Brasilien; Will Eakins, Harshang Shah, ABB Inc. USA

Kurzfassung

Der Bedarf an Automatisierung wächst, insbesondere in den Branchen Bergbau und Öl- und Gasindustrie, wo Standorte abgelegen sind, Arbeitsumgebungen rau und gefährlich sind und qualifizierte Arbeitskräfte knapp sind oder bald in den Ruhestand treten. Hochentwickelte Sensoren wie Kameras (thermisch, multispektral, visuell), Gassensoren usw. liefern reichhaltige Informationen für die Zustandsüberwachung und Fehlererkennung. Wenn die ortsfeste Installation vieler Sensoren in einer Anlage nicht wirtschaftlich ist, dann kann der Transport auf einem Trägerfahrzeug eine Alternative sein. In diesem Beitrag wird der spezifische mechatronische Entwicklungsprozess eines solchen Fahrzeugsystems auf einer Schiene für den Einsatz an Bandförderanlagen im Bergbau vorgestellt. Kern des Entwicklungsprozesses waren enge Iterationszyklen maßgeblicher Fahrzeugelemente inklusive prototypischem Aufbau und Test. Der zweite Teil dieses Beitrags setzt diesen Prozess in Beziehung zu bewährten Entwicklungsmethoden aus der industriellen Praxis.

Abstract

The need for automation is growing, especially in mining and oil & gas industries where sites are remote, work environments are harsh and hazardous, and skilled workers are scarce or soon to retire. Advanced sensors such as cameras (thermal, multi-spectral, visual), gas sensors, etc. provide rich information for condition monitoring and fault detection. If fixed installation of many sensors in a plant is not economical, then transport on a carrier vehicle can be an alternative. This paper presents the specific mechatronic development process of such a vehicle system on a rail for use at belt conveyors in mining. The core of the development process was tight iteration cycles of relevant vehicle elements including prototypical assembly and testing. The second part of this paper relates this process to proven development methods from industrial practice.

1 Einleitung

1.1 Allgemeine Anforderungen

In den vergangenen Jahren wurde der Technologieprototyp eines Sensorfahrzeugs für den Außeneinsatz entwickelt. Speziell wird dieses Sensorfahrzeug an Gurtbandförderern im Bergbau eingesetzt, um den Zustand der Tragrollen zu überwachen. Am Anfang dieses Entwicklungsprozesses standen der Wunsch, genau diese Aufgabe automatisiert durchzuführen, sowie eine erste Liste von Schlüsselanforderungen, die von Experten aus dem Bergbaugeschäft formuliert wurden. Diese allgemeinen Anforderungen waren:

a) Funktionale Zuverlässigkeit: Betriebsstörungen müssen die Ausnahme sein und als solche zuverlässig erkannt und an den Betreiber gemeldet werden, um sie zeitnah beheben zu können.
b) Raue Umgebung: Im Normalbetrieb gibt es eine hohe Staubbelastung aus der Fördergut mit bleibenden Ablagerungen. Je nach Einsatzort gibt es häufig oder gelegentlich Regen, Spritzwasser oder Schlamm und herabfallendes Schüttgut aus dem Förderband.
c) Je nach Einsatzort liegt der Temperaturbereich bei -20…50°C, im Extremfall bei -40...65°C und die rel. Luftfeuchtigkeit bei 20..100 %, sowie typische Tageszyklen in diesen Wertebereichen. Weiterhin gibt es direkte Sonneneinstrahlung mit hoher UV-Exposition und Einsatzorte bis 3000 m ü. NN.
d) Funktionalität: Sensordaten müssen vollautomatisiert und mit ausreichend hoher Qualität und Konsistenz aufgezeichnet werden.
e) Sicherheit: Im Normalbetrieb und bei einfachen Fehlfunktionen soll keine Gefährdung für in der Nähe befindliche Personen ausgehen
f) Kosten: Aufbau und Betrieb sollen geringstmögliche Kosten verursachen und im Verhältnis zum Nutzen für den Anwender stehen

Diese Anforderungen weisen zunächst nicht auf ein Sensorfahrzeug als mögliche Lösung hin, sondern sind allgemeiner. Erst aus der Tatsache, dass die zu überwachenden Tragrollen örtlich verteilt sind und es hunderte bis tausende davon an einer Bandanlage gibt, legt es nahe, die Sensoren mit einem Trägerfahrzeug zu bewegen.

1.2 Beschreibung des Ergebnisses

Der Technologieprototyp besteht aus 2 Fahrzeugen, die sich hängend an einer Schiene fortbewegen. Die Fahrzeuge sind ein Triebwagen und ein Sensorwagen von je 280 mm Länge. Der Triebwagen enthält den Traktionsantrieb und

Batterien und wiegt 5 kg, der Sensorwagen wird gezogen oder geschoben, enthält eine Schwenk-Neigeeinheit mit den Sensoren und wiegt 2 kg. Die Maximalgeschwindigkeit ist < 1 m/s, daher ist die übertragene Energie und damit die Gefährdung bei einer versehentlichen Kollision mit Körperteilen sehr gering. Weiterhin sind die Fahrzeuge mechanisch gegen Entgleisen gesichert. Somit ist Anforderung e) durch konstruktive Maßnahmen weitgehend erfüllt. Die Schiene ist ein gleichschenkliges L-Profil mit 35 mm oder 1,25“ Schenkellänge. Sie ist um 45° geneigt, um Ablagerungen zu minimieren. In regelmäßigen Abständen ist sie mit Halterungen am Bandgerüst montiert. Für eine detailliertere Beschreibung des Systems sei auf [1] und [2] verwiesen.



Bild 1 Übersicht über die Bestandteile des Technologieprototypen während eines Pilotbetriebs (2019-2020).



Bild 2 Übersicht über weitere Bestandteile des Technologieprototypen und konstruktive Maßnahmen gegen raue Umgebungsbedingungen (Anforderungen a) und b))

Bilder 1 und 2 zeigen die Hauptbestandteile des Systems. Ein weiterer nicht gezeigter Hauptbestandteil ist eine „Basisstation“ mit Ladekontakten für die Batterien, einem Schaltschrank für elektrische Baugruppen und einem geschlossenen und beheizbaren Wetterschutz für das Fahrzeug. Dieser ermöglicht ein Laden der Batterien bei > 0°C, was bei der derzeitigen Lithium-Batterietechnologie notwendig ist.

Dieser Technologieprototyp ist das Ergebnis eines mechatronischen Entwicklungsprozesses, der im folgenden Abschnitt beschrieben wird.

2 Mechatronischer Entwicklungsprozess

2.1 Funktionale Anforderungen

Wie in 1.1 beschrieben waren die allgemeinen Anforderungen frei von der Wahl des Trägerprinzips für die Sensoren. Um diese Wahl einzugrenzen, wurden zunächst die Anforderungen an die Sensoren detailliert, die sich aus der Aufgabenstellung und aus Anforderung d) ergeben:

i) Es wird eine Thermografiekamera benötigt, die Bildaufnahmen von den Lagerschilden (Stirnseiten) der Tragrollen machen kann, was eine Absolutwertbestimmung der Oberflächentemperatur ermöglichen soll.

ii) Es wird mindestens ein Mikrofon benötigt, das auf Lagerschilde ausgerichtet kurze Tonaufnahmen erlaubt, die im Rohformat gespeichert werden. Der Frequenzbereich soll bis min. 40 kHz reichen. Tonaufnahmen müssen einen Vergleich der Schallleistung zwischen verschiedenen Rollen und derselben Rolle zu verschiedenen Inspektionszeiten erlauben.

iii) Es wird eine visuelle Kamera benötigt, die bei Tageslicht oder in völliger Dunkelheit Bildaufnahmen der gleichen Bildszene macht, wie die Thermografiekamera. Dies dient dazu, um Objekte zu erkennen, zu denen dann Thermografiewerte zugeordnet werden können, und es dient der visuellen Inspektion auf mechanische Probleme an der Anlage.

Diese Anforderungen sind spezifisch für die eigentliche Aufgabe, die den Nutzen in der Anwendung generiert. Sie sind der Grund, Sensoren in der Anlage zu bewegen und zu positionieren. Interessant ist, dass es keine Anforderungen an die Positioniergenauigkeit der Sensoren gibt, sondern mit d) nur „ausreichende Qualität und Konsistenz“ gefordert ist und mit ii) entweder ein gleichbleibender Abstand oder eine Korrektur wegen unterschiedlicher Abstände gefordert ist. Wegen der optischen Sensoren bedeutet das prinzipiell, dass für die gleiche Bildqualität umso genauer positioniert werden muss, je weiter die Sensoren von der Szene entfernt sind.

In den Anforderungen ist jedoch nicht gefordert, ob d) und i) - iii) bei der Rohdatenaufnahme oder nach einer Datenvorverarbeitung erfüllt sein müssen. Für die Wahl des Sensorträgers bedeutet das, dass er eine Nutzlast in Größe und Gewicht der Sensorenbaugruppe ausreichend genau positionieren muss und die Position halten muss (i-iii).



Bild 3 Vorversuche zur Wahl der Trägerplattform, oben: Drohne (UAV), unten: Unmanned Ground Vehicle (UGV) während Versuchsfahrten zur Datenaufnahme; ein UGV benötigt neben der Sensorenbaugruppe (Kreis) eine umfangreiche Mobilitätsausstattung, um sich sowohl im Gelände wie auch auf Treppen und Laufstegen an der Bandanlage entlang zu bewegen; das abgebildete UGV könnte dies nicht leisten. Ein UAV müsste sehr tief fliegen, um geeignete Blickwinkel auf die Tragrollen zu bekommen.

2.2 Wahl des Sensorträgers

Prinzipiell kann der Sensorträger also 1. an den Boden, 2. die Luft, oder 3. die Anlage gebunden sein. In Variante 3. gibt es prinzipiell die Optionen, das laufende Band als Sensorträger zu verwenden oder ein eigenes Fahrzeug zu benutzen. Eine Bewertung, welche Variante am besten die Anforderungen a) bis f) erfüllt, kam zunächst zu keinem eindeutigen Schluss, so dass Vorversuche durchgeführt wurden (Bild 3). Erst danach war die Bewertung aller Beteiligten deutlich zugunsten eines an die Anlage gebundenen und selbstfahrenden Fahrzeugs.
Hier standen nun mehrere Konzepte für die Ausführung des Trägers und der Führung zur Auswahl (flexibel als Seil oder starr als Schiene, Kanal, Rohr, etc., und Mischformen daraus). Die Wahl fiel auf eine starre Trag- und Führungsschiene, an der das Fahrzeug hängend und im Reibschluss sich fortbewegt. Aus Platzgründen wird hier auf eine ausführliche Darstellung der anderen Varianten verzichtet, welche systematisch gegen den Anforderungen a) bis f) gespiegelt wurden. Ein Schienenfahrzeug kann offensichtlich die Anforderungen a), d) und e) deutlich besser erfüllen als Fahrzeuge am Boden oder in der Luft, weil der Bewegungsraum an die Schiene gebunden ist.
Anzumerken ist auch, dass bei den Entscheidungen zum Typ und zur Variante des Trägerfahrzeugs die derzeit verfügbaren Technologiekomponenten eine entscheidende Rolle spielten: z.B. impliziert „selbstfahrend“ eine eigene Energieversorgung, die mit Li-Batterietechnologie derzeit bewährt, variantenreich, und kostengünstig zur Verfügung steht. 10 oder 20 Jahre in der Vergangenheit oder in der Zukunft würden dieselben Fragen wahrscheinlich zu anderen Ergebnissen führen.

2.3 Schneller Prototyp

Nach der Entscheidung zum Konzept des Trägerfahrzeugs waren nun weitere Detailfragen zu klären, wofür im nächsten Schritt ein schneller Prototyp realisiert wurde. Dieser bestand aus einer fest installierten Schiene und einem Sensorfahrzeug mit Traktionsantrieb, Batterie, Elektronik, und Funkmodulen, sowie Sensoren für i) – iii), wie in Bild 4 gezeigt.
Der schnelle Prototyp erfüllt mehrere Zwecke und kann als Alternative zu einer Reihe anderer Methode gesehen werden, um die prinzipielle Machbarkeit zu prüfen und um weitere funktionale Anforderungen zu entwickeln. Weil es sich um ein im Bergbau neues Konzept zur Anlagenüberwachung handelt, diente der schnelle Prototyp außerdem dazu, das Konzept zu veranschaulichen und bei allen Beteiligten – auch mit nicht-technischem Hintergrund – die Vorstellung dieser neuen Lösung zu entwickeln.

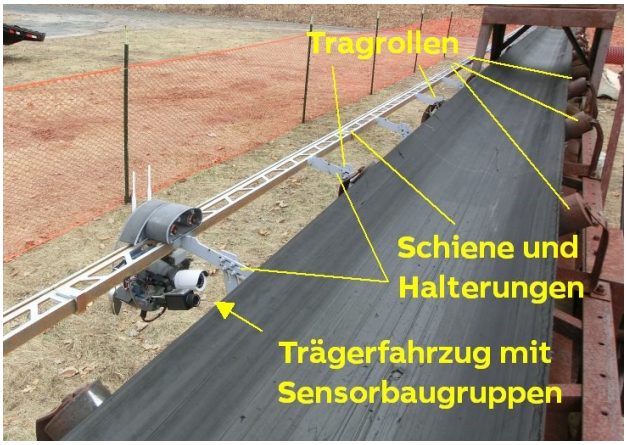


Bild 4 Schneller Prototyp eines Trägerfahrzeugs mit Schiene, um die prinzipielle Machbarkeit zu prüfen und um die funktionalen Anforderungen zu detaillieren.

2.4 Detaillierung der Anforderungen

Tests mit dem schnellen Prototypen führten u.a. zu folgenden Erkenntnissen:

1. Die Sensoren benötigen unterschiedliche Orientierung, um aus der Perspektive der Schiene alle Tragrollen erfassen zu können
2. Größe und Gewicht der Sensorbaugruppen bestimmen wesentlich Größe und Gewicht des Trägerfahrzeugs

3. Bei Tonaufnahmen darf das Fahrzeug keine Störgeräusche verursachen und sollte möglichst stillstehen
4. Die Schiene muss an unterschiedlichen Positionen seitlich oder innerhalb des Gerüsts der Bandanlage installierbar sein, um für möglichst alle bestehenden Bandanlagenkonstruktionen geeignet zu sein
5. Obwohl Bandanlagen aus sich wiederholenden Elementen bestehen, gibt es häufig unregelmäßige Zusätze, z.B. elektrische Verteilerkästen, Tragrollen mit Zusätzen (z.B. um Bandschieflauf zu korrigieren), Messvorrichtungen, etc. Diese können ein „Umfahren“ erfordern, daher sind Kurven zur Seite und nach oben und unten in engem Radius (c.a. 1 m) erforderlich und Steigungen bis c.a. 30° (67 %).
6. Ein hängendes Fahrzeug neigt zu Rollschwingungen, wenn der Schwerpunkt mittig unterhalb der Schiene liegt (wie in Bild 4)
7. Automatikbetrieb bedeutet, dass das System immer eingeschaltet ist und über genug eigene Logik verfügen muss, um sinnvoll auf Störungen zu reagieren. Dazu gehören u.a. ein Erkennen von Hindernissen, die die Weiterfahrt beeinträchtigen, oder die rechtzeitige Rückkehr zur Ladestation bei niedrigem Batteriestand, auch über Steigungen hinweg, sowie viele weitere Zusammenhänge zwischen möglichen Situation und einem sinnvollen „Verhalten“ des Systems.
8. Drahtlose Kommunikation muss mindestens in der Nähe einer Ladestation oder Basisstation vorhanden sein, um Datenaustausch mit dem Sensorfahrzeug und Programmierung von Inspektionsfahrten zu ermöglichen

Dies ist nur eine kleine Auswahl von Erkenntnissen aus dem schnellen Prototypen, die sich direkt in Anforderungen übersetzen ließen. Tatsächlich war der Aufwand, alle Erkenntnisse zu strukturieren und mit unterschiedlichen Experten zu diskutieren, um daraus detaillierte Anforderungen abzuleiten, um ein Vielfaches höher, als den schnellen Prototypen zu planen, zu bauen, und zu betreiben.

2.5 Reifegrad von Entscheidungen zu Funktionsbaugruppen

Der Detaillierungsgrad und Umfang der Anforderungen war nun groß genug, um Konzepte zur Umsetzung zu entwickeln. Dazu wurden die Anforderungen nach Funktionsbereichen gruppiert und Wechselbeziehungen ermittelt. Zum Beispiel hat der Funktionsbereich Elektronik und Verkabelung Wechselbeziehungen zu Umgebungsbedingungen, mechanische Konstruktion, Betriebsbedingungen (z.B. Vibrationen), usw. oder zum Beispiel hat der Funktionsbereich Fahrwerk/Antriebsstrang Wechselbeziehungen zu Schienenkonstruktion, Steuerungssoftware, Traktionssteuerung, Elektronik und Verkabelung, usw. Diese vielen und tiefen Wechselbeziehungen sind kennzeichnend für mechatronische Systeme. Daraufhin wurden Realisierungskonzepte für verschiedene Funktionsbereiche entwickelt und die Realisierbarkeit bewertet.

Für z.B. Elektronik und Verkabelung wurden mehrere Baugruppen identifiziert, z.B. Mikrocontroller, Treiber für den/die Traktionsmotoren, Funkmodul(e), Elektronik für die Sensoren, Art und Ausführung von Kabeln und Steckverbindern. Dafür wurden Platzhalter für die mechanische Konstruktion geschätzt, deren Größe während der weiteren Entwicklung an die tatsächlich ausgewählten Bauteile angepasst wurde. Wegen der Umgebungsbedingungen und Betriebsbedingungen müssen gedruckte Schaltungen in entsprechender Qualität hergestellt werden und zusätzlich versiegelt werden, wobei ausreichende Kühlflächen erhalten bleiben müssen. Entsprechend müssen auch interne Steckverbinder ausgewählt werden. Für all diese Maßnahmen gibt es fertige Lösungen und Produkte bis hin zur Umsetzung im Prototypen, deshalb gibt es an dieser Stelle keine weiteren Fragen an die Realisierbarkeit.

Anders ist es bei Fahrwerk/Antriebsstrang, wo die mechanische Ausführung völlig offen und zusammen mit der Schienenkonstruktion erfolgen muss und viele grundlegende Fragen erst geklärt werden müssen. Neben kreativen Diskussionen zwischen Experten, in denen mögliche Konstruktionen entwickelt und bezüglich der Anforderungen bewertet wurden, wurden auch hier für kritische Fragen schelle Prototypen gebaut und getestet.

Zwei wesentliche Ergebnisse waren, wie in [1] näher beschrieben, 1) die Wahl separater Fahrzeuge für Traktion und für Sensorik und die Möglichkeit, weitere Fahrzeuge anzuhängen. 2) die Wahl des seitlich versetzten Schwerpunkts, wodurch Rollneigung eliminiert und gleichzeitig die Normalkraft auf die Traktionsrollen rein konstruktiv und groß genug eingestellt werden kann, so dass auch im Reibschluss ausreichend Traktionskraft bei Steigungen, kombiniert mit ungünstigen Randbedingungen (z.B. rutschige Schienenoberfläche durch Regen oder Schlamm) vorhanden ist. Die Fragen an geeignete Konstruktion für diese komplexen Zusammenhänge ließ sich im Weiterennur durch prototypischen Aufbau und Tests klären.



Bild 5 Verschiedene Profilformen und Größen in der engeren Auswahl; links: kommerziell erhältlich, rechts: Spezialanfertigung

Tabelle 1 Bewertung einiger Schienenmaterialien bezüglich einiger Kernanforderungen durch die beteiligten Experten

Comparison of Rail Materials and Coatings	Steel, blank	Steel, powder coated	Steel, galvanized	Stainless steel	Aluminium, anodized
Material cost	$	$$	$$$	$$$	$$
Modification costs (curves,..)	+	-	-	-	++
Ease of installation and repair	++	+	0	0	-
Traction	+	-	++	0	0
Low wear on wheels	0	++	-	+	+
Robustness	+	+	+	+	-

Vor dem Aufbau weiterer Prototypen wurden zuerst Varianten der Schienenkonstruktion bewertet. Bild 5 zeigt eine Auswahl kommerziell erhältlicher Profile und eine Spezialanfertigung, und Tabelle 1 zeigt einen Ausschnitt aus der Bewertung verschiedener Materialien, die in die engere Auswahl gezogen wurden. Bei der Profilwahl ist mindestens ein versteifend wirkender Winkel nötig, daher fiel die Wahl auf die kostengünstigste Variante des L-Profils, wobei andere ähnliche Profile nur wenig geringer bewertet wurden und weiter in der engeren Auswahl blieben.
Es ist eine Eigenschaft mechatronischer Entwicklungen, dass besonders in der mechanischen Konstruktion die größten Unsicherheiten und damit auch die größten Freiheiten liegen. Die Entwicklung von Elektronik wird hingegen meistens auf die Erstellung gedruckter Schaltungen und Verkabelungen hinauslaufen, welches aus Systemsicht zu voraussagbaren Ergebnissen führt. Gerade bei konzeptionell neuen Geräten, wie auch hier, sollte auf die mechanische Konstruktion besonders viel Aufmerksamkeit gelegt werden, um dem Ziel einer optimalen Konstruktionslösung möglichst nahe zu kommen.

2.6 Entwicklung weiterer Prototypen

Mit diesem Grobkonzept von Schiene und Fahrzeug wurden erste Konstruktionskonzepte entwickelt und bewertet, um Größe und Anordnung der größten und schwersten Baugruppen zu ermitteln und erste Vorstellungen von Form und Abmessungen der Fahrzeuge zu erhalten. Bild 6 zeigt das wesentliche Ergebnis für das Traktionsfahrzeug.

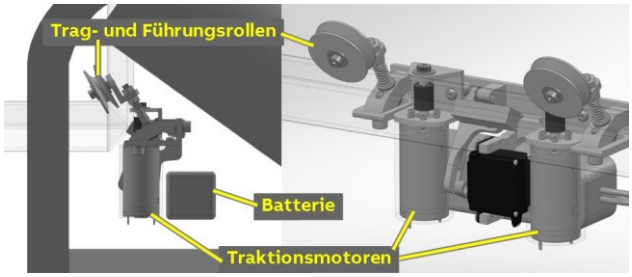


Bild 6 Konstruktionskonzept nach Bestimmung wesentlicher Eigenschaften von Schiene und Fahrzeug; links: Profilansicht, rechts: isometrische Ansicht

Die genauen Lagen und Abmessungen der Baugruppen sind erst ungefähr bestimmt, weil z.B. die Batteriekapazität oder Motorleistung nur geschätzt sind. Wesentliche Eigenschaften dieses Konzepts sind:

a) die geringe Fahrzeuglänge erlaubt enge Kurvenfahrten ohne wesentliche Ausladung
b) Trag- und Führungsrollen und Traktionsrollen laufen auf Schienenkanten, weil sich dort keine Ablagerungen bilden bzw. von den Traktionsrollen oder einfachen konstruktiven Zusätzen („Schneepflug“) zuverlässig beseitigt werden können
c) Es wurden 2 baugleiche Traktionsbaugruppen gewählt, um bei Ausfall einer noch fahrtüchtig zu sein (allg. Anforderung a)
d) die Fahrzeughöhe wird wesentlich von den Traktionsmotoren inkl. Untersetzung bestimmt
e) die Lage der Batterie als schwerste Baugruppe bestimmt wesentlich die Einstellung der Normalkraft auf die Traktionsrollen

Um das Funktionsprinzip und Fahrverhalten dieses Konzepts zu überprüfen, wurden Schienen mit Kurven und Steigungen aufgebaut und ein erstes Fahrzeug mit den in Bild 6 gezeigten Kernbaugruppen aufgebaut. Es enthält zunächst auch Elemente bekannter Plastikbausteine und konnte daher in wenigen Stunden fertiggestellt werden (Bild 7). Mit ersten Tests auf Schienen konnte stabiles Fahrverhalten gezeigt und damit das Grobkonzept der Fahrzeugkonstruktion verifiziert werden.

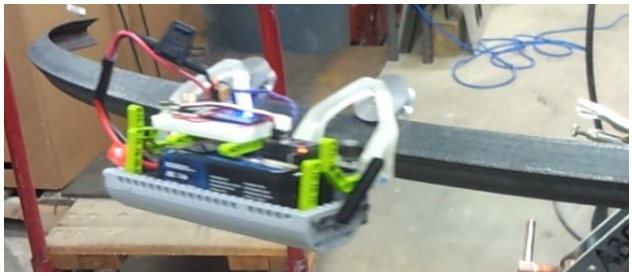


Bild 7 Test eines schnellen Prototypen an Schienenkurve zur Evaluierung der konzeptionellen Fahrzeugkonstruktion.

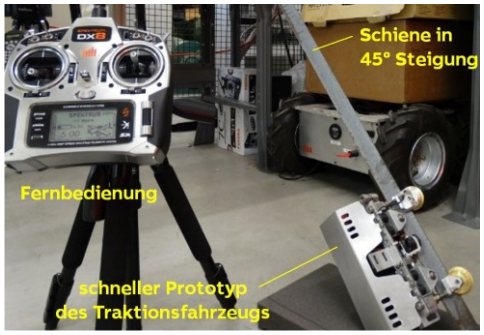


Bild 8 Test eines schnellen Prototypen an Schienenkurve zur Evaluierung der konzeptionellen Fahrzeugkonstruktion

Die nächsten Fragen an die Fahrzeugkonstruktion betreffen die notwendige Antriebsleistung und Normalkraft der Antriebsrollen bzw. Schlupfneigung. Auch dafür wurde ein schneller Prototyp konstruiert und aufgebaut. Dieser wurde auch mit zwei verschiedenen Drehgebervarianten ausgestattet, die für die Bestimmung und Regelung der Fahrzeugposition auf der Schiene benötigt werden. Bild 8 zeigt diesen Prototypen bei einem Steigungstest von 45°, der erfolgreich absolviert wurde und womit quantitative Anforderungen an Traktionsmotoren und Antriebsstrang ermittelt wurden, und wichtige Erkenntnisse bezüglich der Traktionsrollen und der Oberfläche der Schienenkante gewonnen wurden, die in die nächsten Prototypen einflossen.
Der nächste Schritt war der Versuch eines Fahrzeugprototypen, der alle funktionalen Eigenschaften enthält. Dazu wurden in den Triebwagen alle wesentlichen Merkmale des Prototypen aus Bild 8 übernommen bzw. weiter verbessert, sowie weitere Elektronik und Steuerungssoftware hinzugefügt. Zusätzlich wurde ein Sensorwagen aufgebaut, der viele gleiche Baugruppen enthält, wobei aber anstelle

von Antriebsstrang und Batterie eine erste Sensorenbaugruppe eingebaut wurde. Bild 9 zeigt das Ergebnis.

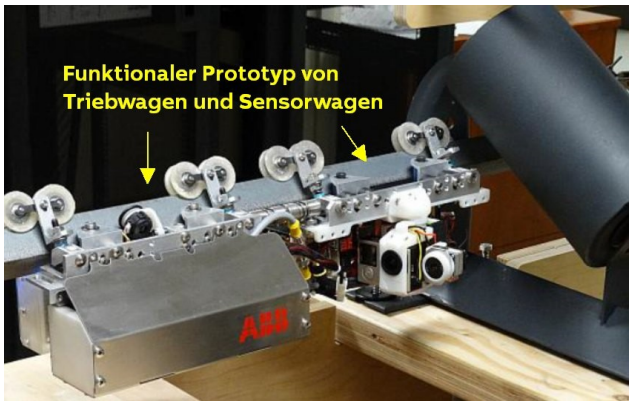


Bild 9 Funktionaler Prototyp von Schiene und Sensorfahrzeug an einem Laborprüfstand mit Tragrollen

Zwischen diesem und dem in 1.2. beschriebenen finalen Prototypen stand ein weiterer Prototyp, der Schutzmaßnahmen gegen die rauen Umweltbedingungen enthielt und durch viele Umwelttests gefahren wurde. Details hierzu sind in [1] beschrieben. Erst auf der Grundlage dieses Prototyps wurde der finale Prototyp entwickelt.

3 Merkmale des mechatronischen Entwicklungsprozesses

3.1 Beschreibung dieses Entwicklungsprozesses

Der in Kapitel 2 skizzierte Entwicklungsprozess enthält mehrere Iterationen von Prototypen, anhand deren die jeweils auftretenden Kernfragen geklärt wurden. Zusätzlich wurden wo nötig Teilkomponenten integriert, zu denen es offene Fragen oder mehrere aussichtsreiche Varianten gab.

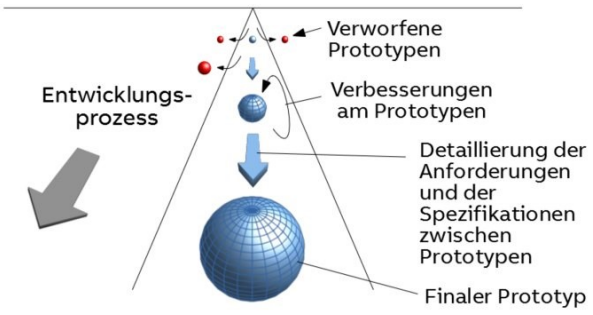


Bild 10 Schematische Darstellung des Vorgehens dieser Mechatronikentwicklung, bevor der finale und ausdetaillierte Prototyp beim Nutzer „ankommt“; bei vorigen, in der Abbildung weiter entfernt liegenden Prototypen sind weniger Details in Anforderungen, Spezifikationen, und Ausgestaltungen erkennbar

Dieses Vorgehen mehrerer Hardwareaufbauten kann kritisch gesehen werden, weil mutmaßlich mit hohem Aufwand und Kosten verbunden. Im Falle solcher kleiner mechatronischer Entwicklungen ist jedoch die benötigte Ausstattung von Werkstatt und Laboren sehr überschaubar. Weiterhin können bewährte Baugruppen voriger Prototypen wiederverwendet werden. Tatsächlich ist sind die Aufwände zu nacheinander folgenden Prototypen deutlich geringer, weil auch das Entwicklungsteam sich auf diese Vorgehensweise einspielt. Für diese Entwicklung war in der Rückschau nach Meinung aller Beteiligten dieser Ansatz der schnellen Prototypen deutlich effizienter und zielführender als Alternativen, die sich mehr auf Modellierung und Simulation stützen. Bild 10 versucht eine schematische Beschreibung dieses Vorgehens.

3.2 Vergleich dieses Entwicklungsprozesses mit Prozessmodellen

Der skizzierte Entwicklungsprozess ist in vielen Zügen vergleichbar mit bekannten Modellen, z.B. dem Drei-Ebenen-Modell [4], wobei die Prototypen (vgl. Bild 10, Kugeln) vergleichbar mit der Ergebnisebene und die Schritte dazwischen (blaue Pfeile) vergleichbar mit der Feinplanungsebene sind. Bild 10 ist insofern einschränkender und ggf. illustrativer, weil es herausstellt, dass in der Regel nur jeweils *ein* Prototyp auf dem *einen* eingeschlagenen Lösungsweg liegt.
Auch mit dem Münchner Vorgehensmodell [3] gibt es Gemeinsamkeiten, wenn man iteratives und rekursives Vorgehen annimmt. Hingegen wäre für diese Entwicklung ein reines V-Modell nach VDI 2206 nicht zielführend gewesen, da zu Beginn auch die Anforderungen allenfalls nur grob entwickelt waren, sondern stattdessen die Anforderungsentwicklung Teil des Prozesses war.

4 Zusammenfassung

Dieser Beitrag beschreibt den mechatronischen Entwicklungsprozess für ein neuartiges Sensorfahrzeug für den Außeneinsatz. Eine Besonderheit des Prozesses waren die durchweg experimentelle Ausdetaillierung sowohl von Anforderungen als auch von Baugruppen und Funktionen. Dies wurde am Beispiel der mechanischen Konstruktionslösung von Schiene und wichtigen Fahrzeugelementen im Detail beschrieben.

Literatur

[1] H. Staab, E. Botelho, D. T. Lasko, H. Shah, W. Eakins and U. Richter, "A Robotic Vehicle System for Conveyor Inspection in Mining," *2019 IEEE International Conference on Mechatronics (ICM)*, Ilmenau, Germany, 2019, pp. 352-357

[2] E. Botelho, H. Staab, „ABB Ability Conveyor Roller Inspection Servces“, *ABB Review 04/2018*, ABB Switzerland Ltd., Baden/Switzerland 2018, pp. 44-49.

[3] Udo Lindemann: Methodische Entwicklung technischer Produkte, Springer Berlin Heidelberg, 2009

[4] Giapoulis, A.: Modelle für effektive Konstruktionsprozesse. (Konstruktionstechnik München, Band 27) Aachen: Shaker 1998. Zugl. München: TU, Diss. 1996

Systematische Vorgehensweise zur Anwendung der Digitaler Zwilling-Technologie bei Spezialgetrieben
Systematic approach for the application of the digital twin technology within special gearboxes

Peter Hehenberger[1], Edwin Kimpl[2], Michael Stroi[2], Voglhuber Clemens[2], Kiyoung Kwak[2]
[1]Fachhochschule Oberösterreich, Forschungsgruppe Smart Mechatronics Engineering, 4600 Wels, Österreich, peter.hehenberger@fh-wels.at
[2]Eisenbeiss GmbH, 4470 Enns, Österreich, edwin.kimpl@eisenbeiss.at

Kurzfassung

Dieser Beitrag stellt eine systematische Vorgehensweise zur Erstellung eines digitalen Zwillings für Spezialgetriebe vor. Dabei steht die Phase des Betriebs im Hauptfokus um genaue Aussagen zu Abnutzung, Wartungszyklen und Lebensdauer auf Basis von begleitenden Messungen und Online-Vorausberechnungen treffen zu können und damit u.a. auch die Auslegungsberechnungen verbessern zu können. Die Eigenschaften von verschiedenen Komponenten sowie Betriebszuständen von Öl, usw. werden dabei in eine übergeordnete Modellbeschreibung übergeführt.
Die entwickelte methodische Vorgehensweise zur Erstellung eines digitalen Zwillings für diesen Anwendungsfall besteht aus mehreren Phasen. In einem ersten Schritt werden mit matrixbasierten Methoden die Einflüsse der Parameter auf die Systemeigenschaften analysiert und daraus die Anforderungen für die Modellbeschreibung bzw. die Messung erstellt. Im nächsten Schritt werden detaillierte physikalische Modelle über die Lebensdauer der einzelnen Komponenten erstellt, welche in weiterer Folge zu einem Gesamtlebensdauersystemmodell zusammengeführt werden.

Abstract

This paper presents a systematic approach for the creation of a digital twin for special gear units. The main focus is on the operating phase in order to be able to make precise statements about wear, maintenance cycles and lifetime on the basis of measurements and online forecasts and thus to be able to improve the design calculations. The properties of different components as well as operating conditions of oil, etc. are thereby transferred to a superordinate model description.
The developed methodic approach to create a digital twin for this use case consists of several phases. In a first step, matrix-based methods are used to analyse the influences of the parameters on the system properties and from this the requirements for the model description and the measurement are created. In the next step, detailed physical models are developed for the current durability of the individual components, which are subsequently merged into an overall lifetime system model.

1 Einleitung

Digitalisierung schreitet in vielen Bereichen voran und stellt nun die meisten Unternehmen vor zusätzliche Arbeitsaufgaben. Diese werden meist ausgelöst durch zusätzliche, neue Anforderungen der Kunden, welche nicht nur das Produkt erwerben wollen, sondern zusätzlich auch noch „Digitale Dienstleistungen“, wie virtuelle Modelle zur Einbindung/Verwendung in der Entwicklung um Aussagen über die Gesamtperformance einer Gesamtanlage geben zu können oder Unterstützung zum Bau und Inbetriebnahme (Virtuelle Inbetriebnahme) oder zusätzliche Hilfsmittel zur Steuerung der Wartungsaktivitäten wie z.B. zuverlässige Aussagen über die Restlaufzeit/-lebensdauer.
Alle diese Aspekte können unter dem Überbegriff „Digitaler Zwilling“ zusammengefasst werden und stellt nun vor allem die Produzenten mechatronischer Systeme vor zusätzliche Herausforderungen. Einerseits stehen meist aus der Entwicklungsphase eine Vielzahl von virtuellen Modellen (Geometrie, Auslegungsberechnungen, Simulationen, …) zur Verfügung, andererseits können aus Messungen während des Betriebs eine Vielzahl von Daten gewonnen werden. Diese zur Verfügung stehende Datenbasis ist aber nicht immer einheitlich und konsistent (z.B. unterschiedliche Programme, Datenformate, Vollständigkeit, Genauigkeiten). Darum ist es nicht immer in einem einzigen Schritt möglich einen Mehrwert zu schaffen.
Genau dieser Weg zur Erstellung eines digitalen Zwillings auf Basis von unterschiedlichen Anfangsgegebenheiten steht im Hauptfokus dieses Beitrags. Es wird eine Vorgehensweise vorgestellt, welches es Unternehmen erlaubt zielgerichtet in der Erstellung vorzugehen. Hauptelemente der Vorgehensweise sind die systematische Analyse des aktuellen IST-Zustands (Ausgangslage), Betrachtung der (Parameter)-Zusammenhänge auf einer Gesamtsystemebene mit Ableitung deren Relevanz, Ableitung von maßgeschneiderten Modellen in verschiedenen Detaillierungsgraden sowie die anschließende Implementierung des digitalen Zwillings am mechatronischen System.

Der Artikel gliedert sich wie folgt: Nach einer Diskussion der relevanten Grundlagen, Definitionen zur Thematik Digitaler Zwilling in Kapitel 2 wird das entwickelte Vorgehensmodell vorgestellt. Kapitel 4 stellt die Anwendung anhand von Spezialgetrieben (welche als mechatronisches System gesehen werden können) dar. Den Abschluss bilden die kritische Reflexion und das Aufzeigen der Potentiale.

2 Grundlagen und Trends

2.1 Der Übergang von mechatronischen über cyber-physikalische Systeme zum Internet der Dinge

In den letzten Jahren haben sich die Konzepte der „cyber-physikalischen Systeme“ (Cyber Physical Systems CPS) und das „Internet der Dinge“ (IoT, Internet of Things) entwickelt, die sich beide in ähnlicher, aber unterschiedlicher Weise auf den Entwurf, die Entwicklung und die Implementierung mechatronischer Komponenten und Systeme auswirken, sowie auf komplexere Systeme, die sich aus der Kombination und Integration dieser ergeben.

Im Fall von cyber-physikalischen Systemen werden einzelne Komponenten, oft, aber nicht immer mechatronischer Natur, durch den Einsatz fortschrittlicher intelligenter Software kombiniert und integriert. In der VDI-Richtlinie 2206 [1] wird in diesem Zusammenhang von cyber-physikalischen, mechatronischen Systemen (CPMS) gesprochen. Das Internet der Dinge stellt eine Domäne dar, die um den Zugang zu und den Austausch von Informationen erweitert ist, um sowohl mechatronische als auch cyber-physikalische Systeme zu integrieren um neuartige Systeme zu schaffen, die auf den Benutzer ausgerichtet sind [2, 3].

Beginnend mit der Mechatronik, betrachtet dieser Abschnitt die damit verbundenen Entwurfsfragen und erörtert den Produktentwicklungsprozess unter Verwendung von modellbasierten Ansätzen. In diesem Zusammenhang ist eine wichtige Anforderung die Interaktion zwischen Systemen auf verschiedenen hierarchischen Ebenen. Dies betrifft die Modellbildung als auch die Simulation.

2.2 Digitaler Zwilling

Das Konzept des „Digitalen Zwillings“ geht auf das Apollo-Programm der NASA zurück, bei dem mindestens zwei identische Raumfahrzeuge gebaut wurden, die es den Ingenieuren ermöglichten, die Bedingungen des Raumfahrzeugs während der Mission widerzuspiegeln, wobei das auf der Erde verbleibende Fahrzeug als Zwilling bezeichnet wird [3]. Heutzutage bezeichnet ein digitaler Zwilling ein computerunterstütztes Modell eines Produkts, welches für verschiedene Zwecke bzw. Produktlebensphasen verwendet wird.

Aktuell gibt es sehr viele Definitionen und Erläuterungen für den Begriff „Digitaler Zwilling“ (Digital Twin) in der Literatur. Zum Beispiel definiert die Gesellschaft für Informatik den digitalen Zwilling als „digitale Repräsentanz von Dingen aus der realen Welt“. Je nach Informationsfluss zwischen physikalischem und digitalem Objekt wird zwischen „Digitalem Modell“ (Digital Master), „Digitalem Schatten“ (Digital Shadow) und „Digitalem Zwilling“ (Digital Twin) unterschieden [4]. Im Rahmen dieses Artikels wird eine Festlegung in Anlehnung an [4] verwendet:

Ein digitaler Zwilling ist eine virtuelle dynamische Repräsentation eines physischen Systems, welcher mit diesem über das gesamte Produktleben für einen bidirektionalen Datenaustausch verbunden ist.

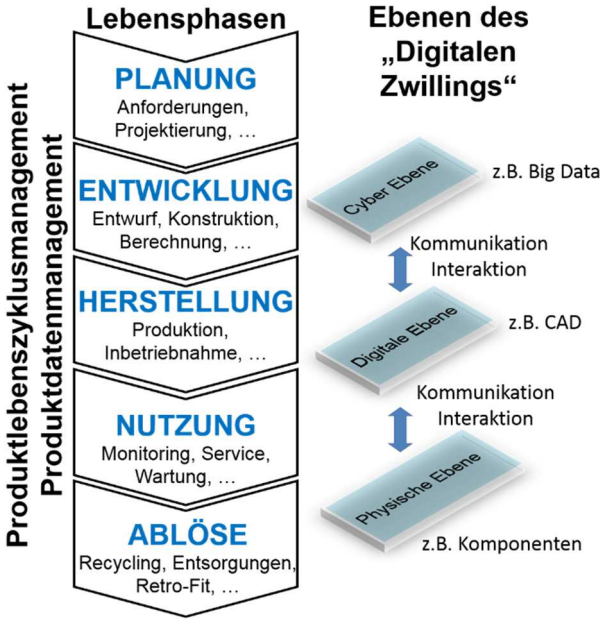


Bild 1 Ebenen des digitalen Zwillings

Bild 1 zeigt die typischen Phasen eines Produktlebens, wobei sich die Betrachtung auf die Phasen Entwicklung, Produktion und Nutzung fokussiert. Es kann zwischen der physischen, digitalen und cyber Ebene unterschieden werden.

2.3 Digitaler Zwilling als Dienstleistungskomponente

In den letzten Jahren zeigte sich, dass bei produzierenden Unternehmen die Bedeutung von Services in den nächsten Jahren steigen wird. Die steigende Verfügbarkeit von Daten eröffnet neue Möglichkeiten für technologiegetriebene Dienstleistungen, die vor allem produzierende Unternehmen maßgeblich betreffen (z.B. Predictive Maintenance oder Augmented Reality Support Systems). Der Übergang von rein physischen Produkten zu zunehmend komplexen Systemen wie Produkt-Service-Systemen [5] vollzieht sich in allen Branchen. Als Produkt-Service-System (PSS) werden Leistungsangebote mit signifikanten Anteilen sowohl an Sachgütern als auch an Dienstleistung verstanden (hybride Leistungsbündel). Die Dienstleistungen werden nicht nur zu Beginn oder bei einigen wenigen Anlässen (z.B.

Wartungsintervallen) erbracht, sondern stellen einen integralen Bestandteil des PSS dar.
Digitale Zwillinge können diese Dienstleistungskomponente unterstützen und ein zusätzliches Geschäftsmodell für die Entwickler und Hersteller mechatronischer Systeme bilden. Bei der Betrachtung von Systemeigenschaften in den verschiedenen Phasen des Produktlebens sind physikalische (prototypische, virtuelle) oder hybride Modelle ein wichtiges Hilfsmittel, welche den beteiligten Stakeholdern in unterschiedlichen Abstraktionsebenen und Detaillierungsgraden zur Verfügung gestellt werden müssen. Daraus ergibt sich ein Übergang zu IT-basierten Lösungen (wie z.B. cloudbasierten). Der digitale Zwilling soll in den verschiedenen Detaillierungsgraden Kunden zur Verfügung gestellt werden. Dabei sind Aspekte wie Software-Architektur und Integration, Nutzungsmodelle, Kosten, Datenschutz sowie Verfügbarkeit zu berücksichtigen [6,7].

3. Methodische Vorgehensweise: Digital Twin-3Phasen-Prozess DT-3PP

Zielsetzung ist die Erstellung einer methodischen Vorgehensweise, welche einen Leitfaden darstellt, der Unternehmen bei der Umsetzung von „Digitalen Zwilling"-Projekten unterstützt. In Bild 2 sind die drei Phasen mit den zugeordneten möglichen Arbeitsschritten, ausgehend von der definierten Zielsetzung des Projektes dargestellt. Im Folgenden werden die einzelnen Phasen allgemein beschrieben.

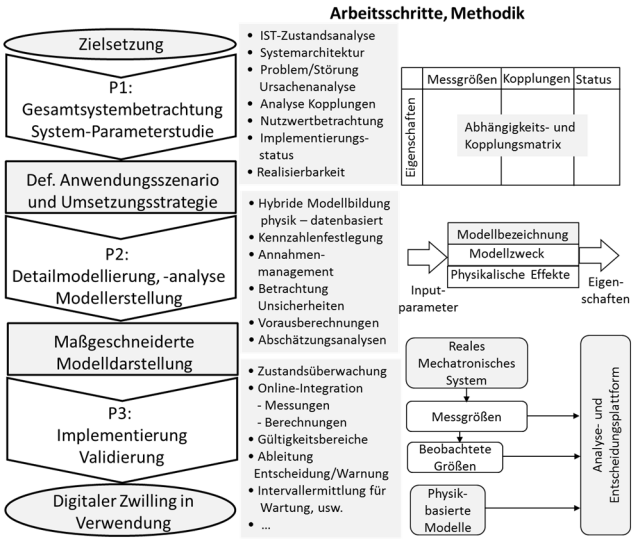


Bild 2 Digital Twin-3Phasen-Prozess DT-3PP

3.1 P1: Gesamtsystembetrachtung

In einem ersten Schritt ist die Aufnahme des IST-Zustands notwendig (welche Modelle, Daten und Informationen sind vorhanden). Weiters sind die Rahmenbedingungen in Hinblick welche Daten aus der Entwicklung oder schon aus dem Betrieb zur Verfügung stehen.
Mit Hilfe einer matrixbasierten Vorgehensweise werden die Eigenschaften des untersuchten Systems in Abhängigkeit der Einflussparameter ermittelt und daraus Prioritäten für die Umsetzung (anhand Nutzen, Implementierungsstatus, Realisierbarkeit) abgeleitet.

3.2 P2: Detailmodellierung

Eine Vorgehensweise stellt hier die hybride Modellierung dar, eine Kombination zwischen physikbasierter und datenbasierter Betrachtung. Je nach Anwendungsszenario und bestehender Ist-Situation muss der Schwerpunkt erst gelegt werden, da nicht immer beide Optionen in bestmöglicher Qualität realisiert werden können (z.B. Nicht zur Verfügung stehende, umfangreiche Messserien für datenbasierte Methoden [8, 9]). Kriterien und Kennzahlenfestlegung zur Bewertungen der Ergebnisse sind dabei von großer Relevanz.
Zwei Trendthemen, der physikbasierte digitale Zwilling und die Künstliche Intelligenz (KI) sind im Bereich Condition Monitoring von Bedeutung, haben aber gegensätzliche Anforderungen und Stärken, wodurch es eine Grundsatzentscheidung ist welche Strategie für die Realisierung verfolgt wird.
Die heute eingesetzten Verfahren der künstlichen Intelligenz beruhen auf angelernten neuronalen Netzen (schwache KI) und erfordern eine entsprechend große und repräsentative Menge an Datensätzen [13]. Diese sind in der Regel nur bei Produkten von Großserien vorhanden und begrenzen sich dabei meist auf den typischen Schaden bzw. die typischen Schäden.
Dem entgegengesetzt stehen die Anforderungen von Einzelgetrieben und Klein- bis Mittelserien. Getriebeschäden sind in der Regel sehr seltene Ereignisse und können unterschiedliche Ursachen haben. Die Generierung von Trainingsdaten für eine umfassende Überwachung ist hier nicht praktikabel. Selbst bei Großserien wird man sich dabei nur auf produkttypische Schäden oder auf eine Detektion von Abweichungen von typischen Messdaten beschränken.

3.3 P3: Implementierung

Als zentrale Elemente werden hier die Integrationsmöglichkeiten der entwickelten Modelle sowie bestehenden Messdaten aus der Zustandsüberwachung untersucht (siehe auch [10, 11]). Die Qualität der Vorhersagen ist von Bedeutung, da daraus Entscheidungen für Handlungsempfehlungen für die Nutzer des mechatronischen Systems abgeleitet werden (z.B. Intervallermittlung für Service- und Wartungsarbeiten). Die Implementierung sollte idealerweise auf einer gemeinsamen Plattform erfolgen, wobei alle notwendigen Teilinformationen zusammengeführt werden.

4. Digitaler Zwilling bei Spezialgetrieben

Industriegetriebe von Eisenbeiss kommen vor allem in der Schwerindustrie, der Extrusionsindustrie und der Energietechnik zum Einsatz. In diesen Einsatzgebieten stehen Themen wie Verfügbarkeit, Zuverlässigkeit und die Vermeidung von ungeplanten Stillständen an oberster Stelle. Wartungs- und Instandhaltungstätigkeiten müssen daher bedarfsorientiert und planbar sein um ein Optimum aus Verfügbarkeit und Zuverlässigkeit zu erzielen. (siehe Bild 3)



Bild 3 Spezialgetriebe (Eisenbeiss)

4.1 P1: Gesamtsystembetrachtung: Getriebezustandsüberwachung

Spezialgetriebe für eine bestimmte Anwendung werden vom Getriebehersteller aufgrund der Anforderungen im Lastenheft dimensioniert. In der Regel stellen diese Anforderungen keine klare Definition von Betriebspunkten, sondern weitgefasste Betriebsbereiche bei gleichzeitig langer Nutzungsdauer sowie hoher Zuverlässigkeit und Verfügbarkeit dar. Oft ist es deshalb nur möglich, das Getriebe auf die Maximalwerte der genannten Anforderungen zu dimensionieren. Höchste Wirkungsgrade, niedrige Schwingungsamplituden und optimales Lasttragevermögen werden jedoch nur im optimierten Auslegungsbereich des Getriebes erreicht. Ungünstige Betriebszustände und unvorhersehbare Belastungsspitzen, Wartungsmängel und Bedienungsfehler können jedoch nie ganz ausgeschlossen werden [14].
Moderne Condition Monitoring Systeme werden neben der Signalisierung von Alarmzuständen zur Abwehr von Maschinenschäden auch zur Festlegung von zustandsorientierten Wartungs-, Inspektions- und Ölwechselintervallen verwendet [12.]
Getriebe (auch Einzelanfertigungen) sind sehr gut bekannte Systeme und lassen einen, wie in Kap. 3.2 beschriebenen, physikbasierten digitaler Zwilling zu. Lebensdauern von Wälzlagern, erforderliche Schmiermengen, Schmierstoff-Viskosität und Öl-Drücke werden bereits in der Konstruktionsphase aufgrund von physikalischen Modellen definiert bzw. berechnet. Mögliche Schäden von Verzahnungen, Lager und Wellen als auch deren Schadensbilder sind eingehend erforscht.
Demzufolge ist es die technisch beste Methode jene, die dieselben physikalischen Modelle die auch in der Entwicklungsphase verwendet werden für die Zustandsüberwachung wiederverwendet. Sie stellt damit sicher, dass die selben Berechnungsansätze verwendet werden und Schwellwerte einen physikalisch fundierten Bezug haben. Der Grundgedanke dabei ist die in [17] angeführte Beziehung (1).

$$\frac{N_{LA}}{N_{LB}} = \left(\frac{p_{HA}}{p_{HB}}\right)^{-7} \cdot \left(\frac{\mu_A}{\mu_B}\right)^{-3} \cdot \left(\frac{\vartheta_{flaA}}{\vartheta_{flaB}}\right)^{-2} \cdot \left(\frac{\nu_{2A}}{\nu_{2B}}\right)^{1,2} \cdot \left(\frac{\lambda_A}{\lambda_B}\right)^{2,2} \quad (1)$$

Lebensdauer bzw. Anzahl der Lastwechsel N_L, Hertzsche Pressung p_H, Reibungszahl μ, Temperatur ϑ, Umfangsgeschwindigkeit ν und spezifische Schmierfilmdicke λ

Dabei steht Index A für die rechnerischen Ergebnisse der Auslegung und Index B für die rechnerischen Ergebnisse aufgrund der aus Messdaten ermittelten momentanen Beanspruchung.
Damit ist die Übertragung von im Betrieb kontinuierlich gesammelten Messdaten auf die Belastungs- und Lebensdauerberechnung eines Getriebes oder einzelner Komponenten beschrieben.
Die Kombination dieser verfügbaren Messdaten mit den Berechnungsmodellen in einem digitalen Zwilling verringert die Wahrscheinlichkeit für einen Schadensfall.
Für ein Überwachungs-, Diagnose-, Prognose-System sind modellgestützte Analyse- und Bewertungsverfahren notwendig, welche den Zusammenhang zwischen Getriebekinematik, Tribologiesystem und Bauteilsicherheiten sowie den zeitlich veränderlichen Messgrößen beinhalten. Bei diesen Verfahren werden, unter Berücksichtigung der aktuellen Belastungsdaten, Veränderungen einzelner Messdaten auf Zustandsänderungen einzelner (nicht direkt messbarer) Parameter zurückgeführt und einzelne Komponenten einer roulierenden Zustandsbewertung und (Rest-) Lebensdauerberechnung unterzogen. Durch diese Rückkopplung von laufend gewonnenen Mess- und Zustandsdaten auf die Getriebeauslegungsdaten zum Zeitpunkt des Engineerings können Bauteilsicherheiten, Lebensdauer und erforderliche Wartungsintervalle sowie die Verfügbarkeit und Zuverlässigkeit der Getriebe ständig aktuell und genauer ermittelt werden.

4.2 P2: Detailmodellierung: Lagerlebensdauer

Exemplarisch soll in diesem Abschnitt die Detailmodellierung einer Komponente des Getriebes gezeigt werden.
Die Lebensdauerbestimmung von Lagern in Industriegetrieben basiert in der Praxis auf der nachfolgenden Gleichung (2). Dabei findet die Belastung, die Lagerform bzw. -größe und die Drehzahl Eingang. Andere Einflüsse werden nicht berücksichtigt.

$$L_{10h} = \left(\frac{C}{P}\right)^p \cdot \frac{10^6}{n \cdot 60} \quad (2)$$

nominelle Lebensdauer L_{10h}, dynamische Tragzahl C, äquivalente dynamische Lagerbelastung P und Exponent p (Punktkontakt am Wälzkörper = 3, Linienkontakt am Wälzkörper = 10/3)

Abgesehen von dieser Ungenauigkeit erfolgt die Berechnung nur für einen vermuteten, ggf. gar nicht real auftretenden Betriebsfall. In Glg. (3) werden nun die rechnerischen Ergebnisse der Getriebeauslegung aufgrund angenommener Beanspruchung und die rechnerischen Ergebnisse aufgrund der aus den Messdaten ermittelten, momentanen Beanspruchung gegenübergestellt.

$$\frac{L_{10h-A}}{L_{10h-B}} = \left(\frac{P_B}{P_A}\right)^p \cdot \frac{n_B}{n_A} \quad (3)$$

Unter Verwendung der tatsächlichen Betriebslagerlast und der Betriebsdrehzahl lassen sich bereits mit der Umsetzung der Beziehung aus [17] gemäß Gleichung (3) Verbesserungen gegenüber dem Status quo erzielen.
Dabei erkennt man deutlich, dass wenn etwa die Auslegungslast um 25% unterschätzt wird, es bereits bei 50% der erwarteten Lebensdauer wahrscheinlich zum Ausfall führt (typischerweise p=10/3). Wird die Belastung laufend bewertet, kann frühzeitig auf diesen ungünstigeren Betriebszustand hingewiesen werden.
In Abschnitt 4.3 wird gezeigt, dass das Betriebsdrehmoment und die Eingangsdrehzahl gemessen werden. Über die mechanischen Zusammenhänge kann damit direkt auf die Größen P_B und n_B des jeweiligen Lagers geschlossen werden. Der Betriebszustand wird durch die Berücksichtigung weiterer Faktoren genauer bewertet. Dazu zählen die Schmierungsbedingung und der Schmiermittelzustand welche laut [15] im a_{ISO}-Faktor zusammengefasst sind.

$$L_{10mh} = a_{ISO} \cdot \left(\frac{C}{P}\right)^p \cdot \frac{10^6}{n \cdot 60} \quad (4)$$

Der a_{ISO}-Faktor ist eine Verbesserung der herkömmlichen Lebensdauerbestimmung. Darin werden die Einflüsse der Last P, die Schmierungsbedingung κ und der Schmierstoffverunreinigung e_C zusammengefasst, siehe Glg. (5).

$$a_{ISO} = f(\kappa, P, e_C) \quad (5)$$

Die Schmierungsbedingung κ ist von den Betriebsparametern Drehzahl n und Temperatur ϑ, sowie vom Öltyp (genauer: Nennviskosität v_{40} und der Richtungskonstante) abhängig. Die Last P wird vom Drehmoment T bestimmt. e_C wird von der Ölreinheitsklasse und von κ bestimmt. Die Ölreinheit wiederum ergibt sich aus dem Partikeleintrag. Dieser wird basierend auf einem von der FVA entwickelten Modell [16] berechnet. Darin finden neben dem initialen Partikelgehalt ebenso die vom jeweiligen Betriebszustand abhängigen Messgrößen Temperatur, Drehzahl und Drehmoment Eingang. Damit ist a_{ISO} als Funktion der Messgrößen (Glg 6).

$$a_{ISO} = f(n, T, \vartheta, e_C) \quad (6)$$

Der Schmiermittelzustand wird umfassend mit GearControl Oil® (siehe Abschnitt 4.3) überwacht. Damit kann beispielsweise der Wassergehalt im Schmiermittel zusätzlich bewertet werden. Glg. (4) kann nun auf eine analoge Form zu Glg. (3) gebracht werden, Glg. (7). Es erfolgt somit eine laufende Beurteilung der Lagerlebensdauer auf Basis der Mess- und Berechnungsgrößen im Vergleich zu den statischen Auslegungsdaten.

$$\frac{L_{10h-A}}{L_{10h-Bi}} = \frac{a_{ISO-A}}{a_{ISO-Bi}} \cdot \left(\frac{P_{Bi}}{P_A}\right)^p \cdot \frac{n_{Bi}}{n_A} \quad (7)$$

Die laufende Online-Lagerlebensdauerberechnung (bzw. Restlebensdauer) erfolgt nun auf Basis der Auslegungsberechnungen, der entwickelten Modelle und der tatsächlichen Messgrößen am Getriebe (Ergebnisse siehe in Bild 4). Durch die zielgerichtete Verwendung der Messgrößen und die dabei erfolgenden Transformationen von Messdaten in Information bzw. bewertbarer Lebensdauereinfluss sowie durch die sequentielle Beurteilung und Aufsummierung der Lagerschädigung ergibt sich eine neue Qualität der Lagerbeurteilung. Der Anlagenbetreiber kann durch diese besseren Informationen ggf. Betriebsstillstände bedingt durch unerwartete Lagerausfälle bzw. unnötige Lagertausche vermeiden, aber auch durch gezielte Schonung der Anlage gezielt produktionsfreie Zeiten für geplante Stillstände ansteuern.

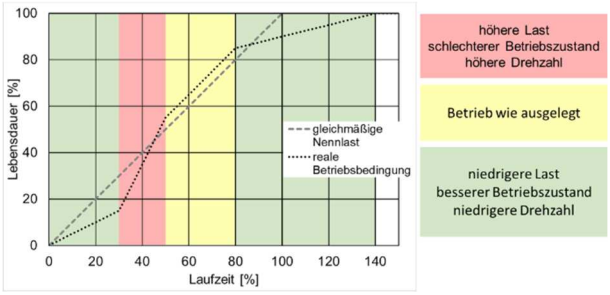


Bild 4 Laufende Lebensdauerprognose

4.3 P3: Implementierung

Die physikalischen Modelle aus der Getriebeentwicklung wurden bisher nur in der Konstruktionsphase mit den Auslegungsdaten verwendet. In Form von digitalen Zwillingen werden diese nun auf der Steuerung von GearControl® implementiert. Sie können somit physikalisch fundierte Informationen zu Öl-Alterung, Lagerlebensdauer, Schmierung und dem Getriebezustand auf Basis der tatsächlichen Betriebsdaten bereitstellen. Die Auswertung der Messdaten erfolgt in einer dezentralen Steuereinheit direkt auf dem Getriebe (siehe Bild 5). Mit oder ohne Cloud Anbindung ist damit das System immer betriebsbereit und voll funktionsfähig.
Basis für die Aufnahme der IST-Daten am Getriebe bildet ein von Eisenbeiss entwickeltes Getriebebegleitsystem GearControl®, welches im Gegensatz zu Zustandsüberwachungssystemen nicht nur dann warnt, wenn bereits ein

Problem aufgetreten ist, sondern Schäden vermeidet und somit die Lebensdauer als auch die Effizienz der Anlagen maximiert [14]. Es implementiert neben der Schwingungsüberwachung zusätzlich die Überwachung der Schmierölversorgung, der Schmierölqualität bzw. des Ölzustands sowie der Temperaturentwicklungen. Solche Multiparameteranalysen lassen bessere Aussagen über den aktuellen Getriebezustand bzw. einzelner Komponenten zu.

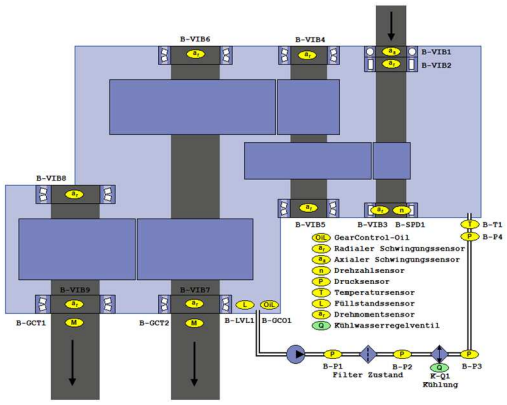


Bild 5 Blockdiagramm i:GEAR 4.0®

5. Zusammenfassung und Schlussfolgerungen

Die vorgestellte Vorgehensweise stellt ein einfach handhabbares Hilfsmittel/Leitfaden dar, dass vor allem Klein- und Mittelunternehmen beim Aufbau der firmenspezifischen Digitaler Zwilling Technologie unterstützt. Ein wichtiger Aspekt ist die Definition der Rahmenbedingungen und Festlegung der Umsetzungsstrategie. Die erste Validierung erfolgte bei einem Unternehmen das Spezialgetriebe entwickelt und herstellt. Der digitale Zwilling soll genutzt werden, um den Betriebszustand analysieren und daraus die Restlebensdauer bzw. den Wartungszeitpunkt ermitteln zu können.

6 Literatur

[1] VDI-Richtlinie 2206: *Entwicklung cyber-physischer mechatronischer Systeme (CPMS) Entwurf*. Verein Deutscher Ingenieure e.V., Düsseldorf 2020.

[2] Hehenberger, P.; Bradley, D.; Dehghani, A.; Traxler, P.: *Mechatronic and Cyber-Physical Systems within the Domain of the Internet of Things*. In: Systems Engineering in Research and Industrial Prac-tice (Editoren: Stjepandić J., Wognum N., J. C. Verhagen W.), London, Berlin, Heidelberg: Springer-Verlag, 2020. ISBN 978-3-030-33311-9.

[3] Hehenberger, P.; Bradley, D.: *Mechatronic Futures – Challenges and Solutions for future Mechatronic Systems and Designers*. London, Berlin, Heidelberg: Springer-Verlag, 2016. ISBN 978-3-319-32154-7.

[4] Trauer, J.; Schweigert-Recksiek, S.; Engel, C.; Spreitzer, K.; Zimmermann M.: *What is a digital twin? Definition and insights from an industrial case study in technical product development*. In: In-ternational Design Conference- DESIGN 2020. https://doi.org/10.1017/dsd.2020.15.

[5] Vajna, S.; Weber, C.; Zeman, K.; Hehenberger, P.; Gerhard, D.; Wartzack, S.: *CAx für Ingenieure, Eine praxisbezogene Einführung*. Berlin, Heidelberg: Springer-Verlag, 2018.

[6] Aheleroff; S.; Xu, X.; Zhong, RY.; Lu, Y.: *Digital Twin as a Service (DTaaS) in Industry 4.0: An Architecture Reference Model*. Advanced Engineering Informatics 47 (2021).

[7] Zheng, C.;Wang, Z.; Qin, X.; Eynard, B.; Hehenberger, P.; Li, J.; Bai, J.; Zhang, Y.: *Integrated design for product–service systems: a focus on multidisciplinary interface.* International Journal of Production Research (2020), DOI: 10.1080/00207543.2020.1794077

[8] Gonzalez, M.; Salgado, O.; Croes, J.; Pluymers, B.; Desmet, W.: *A Digital Twin for Operational Evaluation of Vertical Transportation Systems*. IEEE Access (2020)

[9] Hehenberger, P.; Follmer, M.; Geirhofer, R.; Zeman, K.: *Model-based system design of annealing simulators*. Mechatronics (2013) 23, ISSN 0957-4158, S. 247–256,

[10] Kaul, T.; Hentze, J.; Sextro, W.; Gräßler, I.*: Integration von Verlässlichkeitsmodellen der Entwicklung in einen Digitalen Zwilling zur Umsetzung einer vorausschauenden Instandhaltung*. In: Tagung Mechatronik 2019. Paderborn. VDI Wissensforum Düsseldorf 2019.

[11] Zheng, C.; Bricogne, M.; Le Duigou, J.; Hehenberger, P.; Eynard, B.: *Knowledge-Based Engineering for Multidisciplinary Systems: Integrated Design based on Interface Model*. Concurrent Engineering: Research and Applications (2018), DOI: 10.1177/1063293X17734591

[12] Kimpl, E; Stroi, M.: *Moderne Getriebezustandsüberwachung / Eisenbeiss GmbH*. In: Tribologie- und Schmierstoffkongress GETLUB, Mannheim 2014.

[13] Figge, M.M.: *Embedded KI für Predictive Maintenance* FVA 885 I, Forschungsvereinigung Antriebstechnik, Frankfurt/M 2020, 2020.

[14] Kimpl, E.; Stroi, M.: *GearControl – Condition Monitoring für Industriegetriebe*. stahl und eisen 135 (2015) Nr. 6/7, S. 236-242.

[15] ISO 281:2007, Rolling bearings - Dynamic load ratings and rating life.

[16] FVA: *Heft 888; Partikelgehalt I*, 2009.

[17] Niemann, G.; Winter, H.: *Maschinenelemente, Band 2*, Berlin, Heidelberg: Springer-Verlag, 2003.

Anwendbarkeit von Requirement Mining in Benutzerrezensionen für die Entwicklung mechatronischer Produkte im B2C-Markt
Applicability of Requirement Mining in User Reviews for Developing Mechatronic Products in B2C Markets

Prof. Dr.-Ing. Iris Gräßler, M. Sc. Daniel Preuß, Universität Paderborn, Heinz-Nixdorf-Institut, 33102 Paderborn, Deutschland, iris.graessler@hni.upb.de, daniel.preuss@hni.upb.de

Kurzfassung

In der Softwareentwicklung werden Requirement Mining-Ansätze eingesetzt, um in der Ermittlung von Anforderungen Informationen aus Benutzerrezensionen zu extrahieren. Auf der Online-Plattform Amazon werden verschiedene mechatronische Produkte des B2C-Markts vertrieben und deren Benutzerrezensionen gesammelt. Aufgrund der Komplexität mechatronischer Produkte werden Defizite selten nachvollziehbar im Benutzer-Feedback dokumentiert. In diesem Beitrag wird die Forschungsfrage beantwortet, ob Requirement Mining-Ansätze auch in der Produktgenerationenentwicklung mechatronischer Produkte des B2C-Markts anwendbar sind. Etablierte Requirement Mining-Ansätze werden in einer Literaturstudie identifiziert und die Anwendbarkeit initial bewertet. Der vielversprechendste Ansatz wird für eine weitergehende Analyse ausgewählt. Hierzu wird ein Fallbeispiel untersucht, indem Benutzeranforderungen eines Staubsauger-Roboters ausgewertet werden. Die Ergebnisse zeigen, dass die Informationsqualität hoch ist und Anforderungsingenieure diese Informationen zur Dokumentation der Bedürfnisse der Benutzer verwenden können.

Abstract

In software development, requirement mining approaches are used to extract information from user reviews in requirements elicitation. On the online platform Amazon, different mechatronic products of the B2C market are offered and their user reviews are collected. Due to the complexity of mechatronic products, deficits are rarely documented in user feedback. In this paper the research question is answered, whether requirement mining approaches are also applicable in the product generation development of mechatronic products of the B2C market. Established requirement mining approaches are identified in a literature review and their applicability is initially evaluated. The most promising approach is selected for further analysis. For this purpose, a case study is investigated by evaluating user requirements of a robot vacuum cleaner. The results show that the information quality is high and requirements engineers can use this information to document user needs.

1 Motivation und Problemstellung

Im Vergleich zu anderen Methoden zur Ermittlung von Anforderungen ist die Analyse von Benutzer-Feedback für Unternehmen besonders kosteneffizient [1]. Benutzerrezensionen sind auf Online-Plattformen dokumentiert und liefern in Abhängigkeit von Plattform-Community, Anreizsystemen und Produktkategorien eine Menge an unstrukturiertem Feedback. Beispielsweise bietet das Unternehmen iRobot verschiedene Produktgenerationen von Staubsauger-Robotern auf der Online-Plattform Amazon an. Die manuelle Extraktion von Informationen zur Ermittlung von Anforderungen aus verschiedenen Benutzerrezensionen ist zeitaufwändig und führt zu hohen Kosten für Unternehmen [2]. Requirement Mining bezeichnet die Anwendung von Data-Mining-Techniken für Requirements Engineering-Aufgaben. Dies gilt insbesondere für die Identifikation und Klassifizierung von Anforderungen [3]. Diese Techniken werden in der Softwareentwicklung eingesetzt, um die Extraktion von Informationen zur Ermittlung von Anforderungen aus Benutzerrezensionen zu automatisieren und die Effizienz zu steigern [2, 3]. Bei der Produktgenerationenentwicklung mechatronischer Produkte im B2C-Markt wird Requirement Mining aktuell nicht angewandt.

Auswirkungen der Komplexität mechatronischer Produkte auf die Informationsqualität von Benutzer-Feedback müssen berücksichtigt werden: Produktdefizite aufgrund diverser Wechselwirkungen von Komponenten und von äußeren Einflussfaktoren sind selten nachvollziehbar im Benutzer-Feedback dokumentiert. Die Ursachen von Fehlverhalten des Produkts sind für den Benutzer nicht zwangsläufig erkennbar (beispielsweise Defekte der Sensorik oder elektronische Defekte der Steuerplatine). Außerdem unterscheiden sich die Quellen von Benutzer-Feedback mechatronischer Produkte im Vergleich zu Software hinsichtlich Meta-Daten und der Menge der verfügbaren Daten. Da Requirement Mining-Ansätze in der Softwareentwicklung zu Effizienzsteigerungen führen [3] und solche Potenziale auch in der Entwicklung mechatronischer Produkte zu erwarten sind, wird in diesem Beitrag die folgende Forschungsfrage beantwortet: *„Sind automatisierte Requirement Mining-Ansätze in der Produktgenerationenentwicklung mechatronischer Produkte im B2C-Markt anwendbar?“*. Die Anwendbarkeit wird anhand der Informationsqualität der der Requirement Mining-Ergebnisse bewertet.

2 Methodik

Zur Beantwortung der Forschungsfrage wird zunächst die Identifikation von Requirement Mining-Ansätzen anhand einer Literaturstudie nach MACHI und MCEVOY [4] durchgeführt. Es werden Kriterien (z. B. „Gruppierung von Aussagen“) abgeleitet und die Anwendbarkeit der Ansätze initial bewertet. Ziel dieser initialen Bewertung ist es, einen Ansatz mit besonders hohem Anwendungspotenzial zu identifizieren. Zur weiteren Untersuchung der Anwendbarkeit wird ein Fallbeispiel ausgewertet. Der identifizierte Ansatz wird implementiert und für Benutzerrezensionen eines mechatronischen Produkts angewendet. Zur Bewertung der Ergebnisse werden die Merkmale der Informationsqualität (z. B. „Interpretierbarkeit“) der generierten Informationen – beispielsweise die Priorisierung einer Benutzer-Aussage – bewertet.

3 Requirement Mining – Literaturstudie

Für die Literaturstudie zum Requirement Mining werden relevante Suchbegriffe direkt aus der Forschungsfrage abgeleitet und mit Booleschen-Operatoren verknüpft. Diese Suchbegriffe werden auf den Online Forschungsrepositorien Google Scholar und Web of Science verwendet. In einer Tally-Matrix [4] werden die Informationen (Grundlegende Konzepte, Referenzen etc.) zu den verschiedenen Beiträgen in einer einheitlichen Struktur erfasst. Diese werden genutzt, um die Vorgehensweisen der einzelnen Ansätze zu vergleichen.

3.1 Vorgehen bei der Literaturanalyse

Die genutzten Suchbegriffe sind in **Tabelle 1** dargestellt. Diese Suchbegriffe werden zu einem Suchstring kombiniert: „(A_1 OR A_2 OR A_3 OR A_4) AND B_1 AND C_1“. Hieraus resultieren 34 Ergebnisse. Die Relevanz der Ergebnisse wird schrittweise anhand der Titel, der Kurzfassungen und des Textkörpers bewertet.

Tabelle 1 Suchbegriffe zur Identifikation von Requirement Mining-Ansätzen

Kategorie	*Bereich*	*Suchbegriff*
A	Requirement Mining	A_1: “Requirement Mining”
		A_2: “Data Mining”
		A_3: “Automated Requirement Elicitation”
		A_4: “Requirement Discovery”
B	Anforderungsart	B_1: “User Requirement”
C	Quelle	C_1: “User Review”

Beiträge werden aus verschiedenen Gründen ausgeschlossen, die mit dem Format, Wiederholungen, fehlendem Peer-Review, fehlender Darstellung eines entwickelten Requirement Mining-Ansatzes für Benutzerrezensionen, Redundanz und mangelnder Qualität der Dokumentation zusammenhängen. Insgesamt werden elf relevante Requirement Mining-Ansätze identifiziert und systematisch auf Anwendbarkeit im B2C-Markt mechatronischer Produkte bewertet.

3.2 Requirement Mining Ansätze

Im Folgenden wird das grundlegende Vorgehen von Requirement Mining-Ansätzen beschrieben. Es werden zunächst Benutzerrezensionen gesammelt (**Schritt 1**). Hierzu werden, je nach Ansatz, verschiedene Crawler eingesetzt [3, 5]. Darauf folgt die Vorbereitung der extrahierten Daten (**Schritt 2**): es werden Sätze aufgeteilt, Wörter in Tokens umgewandelt und Wörter klein geschrieben. Die Ansätze unterscheiden sich in den einzelnen Schritten der Datenvorbereitung [6]. Im nächsten Schritt werden Zielwörter identifiziert (**Schritt 3**). Diese sind für bestimmte Funktionalitäten des Produkts charakteristisch. Beispielsweise wird bei einem Staubsauger-Roboter das Wort „Filter“ als Zielwort identifiziert (s. **Bild 2**, Gruppe 1). Es werden Regeln angewendet, die auf Part Of Speech (POS) Tags – Identifikation und Bezeichnung von Satzbestandteilen wie Nomen, Verben etc. – basieren [2, 3, 7, 8]. Zusätzlich werden Sätze hinsichtlich des Inhalts klassifiziert, beispielsweise nach „Aspect Evaluation“, „Praise“, „Feature Request“, „Bug-Reports“ oder „Other“ [3, 9–12]. Die Zielwörter werden segmentiert (**Schritt 4**), indem bestimmte Kennzahlen hinsichtlich Relevanz berechnet werden [13]. Beispielsweise wird die „Term Frequency – Inverse Document Frequency“ (TF–IDF) berechnet [2]. Dadurch werden Wörter identifiziert, die besonders häufig im Text-Korpus auftreten. Anschließend wird eine Sentiment-Analyse durchgeführt (**Schritt 5**), um die Gesinnung der Benutzer zu den jeweiligen Zielwörtern zu erfassen [5–9, 13]. Hierzu muss zunächst der Bezug des Sentiment-Wortes – diese ermöglichen Rückschlüsse auf die Gesinnung des Benutzers, z. B. „lieben“ – auf das Ziel-Wort erfasst werden. Dazu werden semantische Regeln, die auf POS Tags, bzw. „Dependency Trees“ basieren, angewendet [6–9]. In einem „Dependency Tree“ werden Abhängigkeiten von Wörtern innerhalb eines Satzes als Baumstruktur dargestellt. Die Zielwörter werden priorisiert, indem Kennzahlen wie die TF–IDF oder das Sentiment der Benutzer verwendet werden (**Schritt 6**). Außerdem werden weitere Daten, wie die Sterne-Bewertung der Benutzer hinzugezogen [2, 5, 7, 8, 13]. Es werden entweder Prioritätskennzahlen zu einzelnen Zielwörtern [7, 9] oder zu Gruppen von Zielwörtern [2, 5, 10–12] gebildet. Benutzeraussagen, die hoch-priorisierte Zielwörter enthalten, sind für die Erhebung von Anforderungen besonders relevant.

4 Anwendungsfall in der Entwicklung mechatronischer Produkte

Um Kriterien für die Anwendbarkeit von Requirement Mining in der Produktgenerationenentwicklung aufzunehmen, muss das Requirements Engineering hinsichtlich Restriktionen analysiert werden. Die Entwicklung mechatronischer Produkte wird durch Richtlinien wie die VDI 2206 unterstützt [14]. Das dort zugrunde gelegte V-Modell beschreibt eine sachlogische Verknüpfung von Aufgaben der interdisziplinären Produktentwicklung [15]. Hauptaktivitäten des Requirements Engineering sind die Ermittlung, Dokumentation, Validierung und Verhandlung sowie das Management von Anforderungen [16, 17].

4.1 Requirements Engineering

Nach IEEE 610.12-1990 [18] ist eine Anforderung definiert als:

1. *„eine Bedingung oder Fähigkeit, die von einem Benutzer benötigt wird, um ein Problem zu lösen oder ein Ziel zu erreichen,*
2. *eine Bedingung oder Fähigkeit, die ein System oder eine Systemkomponente erfüllen oder besitzen muss, um einen Vertrag, einen Standard, eine Spezifikation oder andere formal auferlegte Dokumente zu erfüllen*
3. *und eine dokumentierte Darstellung einer Bedingung oder Fähigkeit wie in 1. oder 2. beschrieben.“*

Nach dieser Definition sind Anforderungen immer lösungsorientiert. Im Gegensatz dazu wird der Begriff "Bedürfnisse" für problemorientierte Aussagen verwendet [19]. Benutzeranforderungen (lösungsunabhängig) können von Systemanforderungen (lösungsspezifisch bzw. den Lösungsraum einschränkend) unterschieden werden [20]. Bei der Entwicklung eines Produkts müssen mehrere Aufgaben durchgeführt werden, um Benutzeranforderungen zu ermitteln [19, 21]. Bedürfnisse der Benutzer müssen verstanden und Wissen sozialisiert werden. Gleichzeitig ist es notwendig, den Anwendungskontext des Problems zu identifizieren. Auf Grundlage der Analyse der Bedürfnisse und des Anwendungskontexts werden die Benutzeranforderungen spezifiziert. Durch das Requirement Mining wird die Spezifizierung der Benutzeranforderungen unterstützt, indem Benutzerrezensionen ausgewertet werden. Es existieren verschiedene Quellen von Benutzerrezensionen für Software und mechatronische Produkte.

4.2 Benutzerrezensionen mechatronischer Produkte

Benutzerrezensionen mechatronischer Produkte für den B2C-Markt werden auf Online-Marktplätzen wie denen von Amazon oder Reichelt dokumentiert. Amazon ist der umsatzstärkste Online-Marktplatz im Elektronik- und Mediensegment, denen mechatronische Produkte für den B2C-Markt zugeordnet sind [22]. Dies ist ein Indikator dafür, dass besonders viele Kunden bei Amazon mechatronische Produkte des B2C-Markts erwerben und damit die Anzahl an Benutzerrezensionen höher als bei anderen Online-Marktplätzen ist. Da mit dem Umfang der Datengrundlage auch die Vollständigkeit der Anforderungserhebung steigt, wird auf diese Quelle hier der Fokus gelegt. Die Metadaten von Amazon-Benutzerrezensionen bestehen aus dem Titel der Benutzerrezension, Name des Verfassers, Datum, Anzahl der Benutzer, welche die Benutzerrezension als hilfreich empfinden, Bewertung (eins bis fünf Sterne) und dem Text der Benutzerrezension. Die Metadaten der Benutzerrezensionen können mit Crawlern (z. B. das Browser Plug-in „Amazon Review Scraper“) akquiriert und in einer CSV-Datei gespeichert werden.

5 Analyse der Anwendbarkeit von Requirement Mining

Auf Basis der Analyse von Requirement Mining-Ansätzen (Abschnitt 3) und des Entwicklungskontexts mechatronischer Produkte (Abschnitt 4) werden Kriterien definiert, welche die Bewertung der Anwendbarkeit existierender Requirement Mining-Ansätze erlauben.

5.1 Anforderungen an das Requirement Mining

Es werden sieben Kriterien definiert und in die Kategorien *Datengrundlage*, *Ergebnisse* und *Aufwand* eingeordnet.

Datengrundlage: Für den vorliegenden Anwendungsfall muss der Ansatz Amazon-Rezensionen (mit ihren Metadaten) verarbeiten **(K1)** (s. Abschnitt 4.2). Durch die Berücksichtigung des Ortes sollen geografische Besonderheiten (z. B. Netzspannung, kabellose Datenübertragung) und durch die zeitliche Dimension Ereignisse des Entwicklungs-Umfelds (z. B. Veröffentlichung eines Konkurrenz-Produktes) berücksichtigt werden. Der Ansatz muss innerhalb einer Benutzerrezension mehrere Ziel-Wörter erkennen können **(K2)**.

Ergebnis: Der Ansatz muss für mechatronische Produkte angewendet werden können, daher müssen Funktionen und physische Komponenten als Ziel-Wörter erkannt werden **(K3)**. Der Ansatz muss die Ziel-Wörter, bzw. Aussagen der Benutzer anhand ihrer Bedeutung für den Benutzer priorisieren **(K4)**. Die Informationen zur Ermittlung von Anforderungen müssen für den Entwickler verständlich sein **(K5)**. Dies betrifft die Gruppierung und Priorisierung von Aussagen.

Aufwand: Zur Reduzierung des Aufwands muss der Ansatz Aussagen mit relevanten Ziel-Wörtern gruppieren und Redundanzen entfernen. **(K6)**. Der Ansatz soll mit möglichst wenig Anpassungen auch für weitere mechatronische Produkte angewendet werden **(K7)**. Beispielsweise sollen keine Wissensbasen (Ontologien, Wort-Datenbanken etc.) manuell für eine weitere Anwendung angepasst werden müssen.

5.2 Bewertung der Anwendbarkeit

Die Verarbeitbarkeit von Amazon-Rezensionen **(K1)** ist nur in dem Ansatz von FU ET AL. [5] vollständig gegeben, weil das Verfassungsdatum der Rezensionen berücksichtigt und damit auf Ereignisse des Umfelds eingegangen wird. Sämtliche Ansätze berücksichtigen mehrere Ziel-Wörter in Rezensionen **(K2)**. Ebenfalls können sämtliche Ansätze physische Komponenten als Ziel-Wörter identifizieren **(K3)**. Nur die Ansätze von CHEN ET AL. [2] und JIANG ET AL. [7] ermöglichen eine aussagekräftige Priorisierung von Aussagen **(K4)**, weil Prioritätskennzahlen direkt zu Aussagen zugeordnet werden. Die Informationen sind bei sämtlichen Ansätzen nur teilweise nachvollziehbar, weil Ausschnitte aus Benutzerrezensionen präsentiert werden, die häufig vom Verfasser nicht verständlich formuliert worden sind **(K5)**. Im Gegensatz zu JIANG ET AL. werden von CHEN ET AL. zusätzlich Gruppen relevanter Ziel-Wörter gebildet, um die Anforderungsaufnahme zu verbessern **(K6)**. Zur Anwendung der Ansätze ist bei allen Ansätzen der Aufwand teilweise gering **(K7),** weil manuelle Schritte zur Datenvorbereitung notwendig sind. Die Anforderungserfüllung der Ansätze wird in **Bild 1** dargestellt.

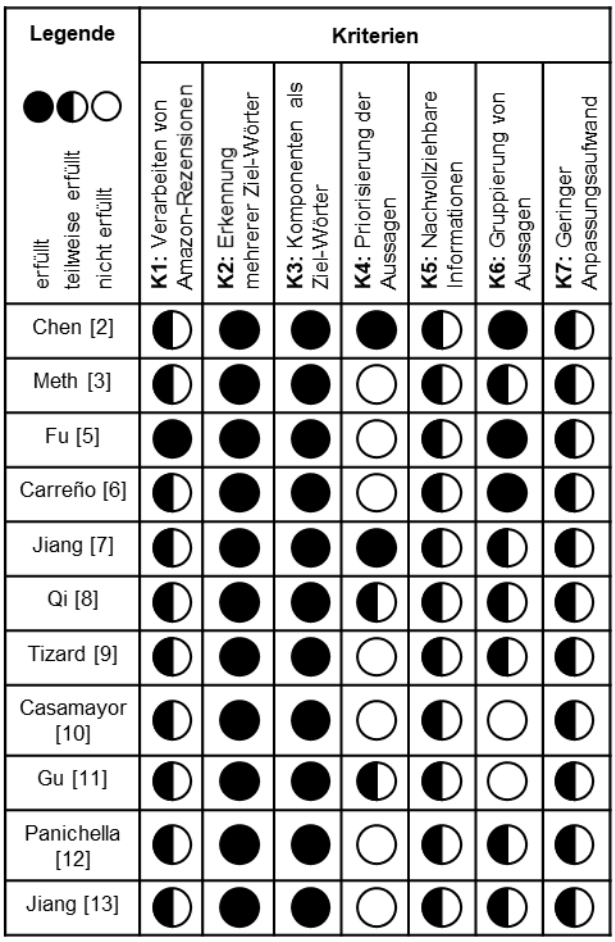


Bild 1 Erfüllung der Kriterien durch die Ansätze

Der Ansatz von CHEN ET AL. ist im Hinblick auf die Erfüllung der Kriterien besonders geeignet für die Analyse von Benutzerrezensionen im Kontext mechatronischer Produkte. Die Kriterien **K1**, **K5** und **K7** werden nicht vollständig erfüllt und darüber hinaus ist für die Einschätzung der Anwendbarkeit die Anwendung des Ansatzes im Praxiskontext zu untersuchen. Dafür wird ein Fallbeispiel ausgewertet.

6 Fallbeispiel

Die Anwendbarkeit des Ansatzes von CHEN ET AL. wird im Rahmen eines Fallbeispiels untersucht. Es werden Amazon-Benutzerrezensionen von einem mechatronischen Produkt analysiert, um Informationen zur Ermittlung von Benutzeranforderungen abzuleiten. Der Staubsauger-Roboter Noisz S5 wird als Produkt ausgewählt, weil ausreichend Benutzerrezensionen (s. Abschnitt 6.4, Merkmal „Umfang der Daten“) zur Anwendung des Requirement Mining Ansatzes vorliegen. Die Merkmale der Informationsqualität der Ergebnisse werden bewertet.

6.1 Qualitätsmerkmale von Informationen

Informationsqualität wird von der Gesellschaft für Informatik beschrieben als eine „*Menge von Qualitätsmerkmalen […], wobei die Auswahl und die genaue Definition der Merkmale den Experten der jeweiligen Anwendungsdomäne vorbehalten bleiben*“ [23]. Zur Messung der Informationsqualität werden verschiedene numerische Metriken zu Qualitätsmerkmalen aggregiert oder durch Experten eingeschätzt. Es werden die Qualitätsmerkmale von REITMEIER und PAETZOLD verwendet, weil diese einen direkten Bezug zu Informationen in der Produktentwicklung besitzen [24]. Die Merkmale sind in **Tabelle 2** dargestellt.

Tabelle 2 Qualitätsmerkmale von Informationen [24]

Kategorie	*Qualitätsmerkmal*
Intrinsische Informationsqualität (IQ)	Glaubwürdigkeit, Genauigkeit, Objektivität, Reputation
Kontextuelle IQ	Wertschöpfung, Relevanz, Aktualität, Vollständigkeit, Umfang der Daten
IQ der Repräsentation	Interpretierbarkeit, Verständlichkeit, konsistente und kompakte Repräsentation
IQ hinsichtlich Zugang	Verfügbarkeit, Zugänglichkeit hinsichtlich Sicherheit

6.2 Beschreibung des Ansatzes

Der Ansatz von CHEN ET AL. folgt dem in Abschnitt 3.2 erläuterten Vorgehen. Benutzerrezensionen von Amazon werden mit der Amazon API extrahiert. Die Zielwörter werden selektiert, indem eine Punktzahl hinsichtlich der Domänenrelevanz der Zielwörter berechnet wird. Zielwörter, die häufig in domänenfremden Benutzerrezensionen auftreten, werden entfernt. Es werden Gruppen von Zielwörtern gebildet, indem drei Ähnlichkeitsmaße berechnet werden: „Rating Similarity“, „Character Similarity“ und „Context Similarity“. Es wird eine Gruppen-Punktzahl berechnet, um die Wichtigkeit der Aussagen, in denen Zielwörter der jeweiligen Gruppe auftreten, wiederzugeben. Die Wichtigkeit der Aussage korrespondiert mit der jeweiligen Gruppen-Punktzahl. Diese berechnet sich aus dem Quotienten der Anzahl der Rezensionen, in denen Zielwörter der Gruppe auftreten und die durchschnittliche Bewertung der jeweiligen Rezensionen.

6.3 Darstellung der Ergebnisse

Bei der Anwendung des Ansatzes von CHEN ET AL. auf die Benutzerrezensionen des Noisz S5 werden 155 Benutzerrezensionen mit einer durchschnittlichen Bewertung von 4,05 (Skala von eins bis fünf) betrachtet. Insgesamt wurden 23 relevante Zielwörter identifiziert, die in zwölf Gruppen eingeordnet sind. Für die einzelnen Gruppen werden die Informationen dem Anwender analog zu **Bild 2** ausgegeben. Pro Gruppe werden durchschnittlich 1,92 Zielwörter mit einer Standardabweichung von 1,68 Worten zugeordnet. Die durchschnittliche Punktzahl der Gruppen beträgt 6,85 mit einer Standardabweichung von 8,36 Punkten.

6.4 Bewertung der Informationsqualität

Das Ergebnis des Ansatzes sind bewertete und selektierte Aussagen von Benutzern über Aspekte des Produkts, die durch Zielwörter beschrieben werden. Diese spiegeln die Meinung des individuellen Benutzers über den Aspekt des Produkts wider. Es können jedoch auch „Fake-Rezensionen“ erstellt werden, die nicht den Meinungen der Benutzer entsprechen [25] **(Glaubwürdigkeit: teilweise erfüllt)**. Bei diesem Ansatz werden dem Anwender Aussagen präsentiert, die für die Aufnahme von Anforderungen

teilweise keine hohe Relevanz besitzen (z. B. Aussage 2 in Gruppe 1 in Fehler! Verweisquelle konnte nicht gefunden werden.). Um dieses Problem zu lösen, kann eine Klassifikation im Post-Processing der Ergebnisse angewendet werden, um beispielsweise Aussagen mit einem „Feature Request“ zu erkennen [9] **(Genauigkeit: teilweise erfüllt)**. Die Bewertung der Relevanz der Benutzer-Aussagen sind formalisiert **(Objektivität: erfüllt)**. Die Aussagen aus den Rezensionen spiegeln direkt die Meinung der Nutzer wider. Jedoch können sich Meinungen individueller Nutzer widersprechen (z. B. Aussage 1 und 3 in Gruppe 2) **(Reputation: teilweise erfüllt)**.

Die Auswertung der Rezensionen wird durch die Priorisierung und Selektion relevanter Aussagen unterstützt. Dadurch spart der Entwickler Zeit im Vergleich zu einer manuellen Analyse und erlangt wichtige Erkenntnisse hinsichtlich der Wichtigkeit bestimmter Aussagen **(Wertschöpfung: erfüllt)**. Die Benutzerrezensionen enthalten relevante Informationen zur Aufnahme von Anforderungen. Weitere relevante Informationen können durch eine anschließende Klassifikation hervorgehoben werden (s. Merkmal Genauigkeit) **(Relevanz: erfüllt)**. Bei diesem Ansatz wird keine Selektion der Rezensionen nach Datum unterstützt, um Ereignisse des Umfelds zu berücksichtigen. Die Rezensionen müssen manuell selektiert werden **(Aktualität: teilweise erfüllt)**. CHEN ET AL. zeigen, dass die Selektion mit einer Präzision von 89 % durchgeführt wird, sodass der Großteil der Informationen korrekt ausgewählt wird [2] **(Vollständigkeit: erfüllt)**. Durch die Anwendung des Ansatzes werden viele Benutzer-Meinungen berücksichtigt (hier: 155). Gleichzeitig wird der Umfang der Informationen auf die anstehende Aufgabe durch die Priorisierung und Selektion beschränkt, sodass dem Anwender nur besonders relevante Informationen angezeigt werden. Die Aussagekraft der Gruppen-Punktzahl nimmt proportional mit der Anzahl der Benutzerrezensionen ab. Ein Potenzial zur Erweiterung der Datengrundlage bietet hier die Betrachtung von Referenzprodukten nach dem „Benchmarking“ [26] **(Umfang der Daten: teilweise erfüllt)**.

Aussagen der Benutzer sind in natürlicher Sprache verfasst, die Qualitätsdefizite z. B. hinsichtlich Eindeutigkeit der Aussagen aufweisen **(Interpretierbarkeit: teilweise erfüllt)**. Die Punktzahl, welche die Priorisierung darstellt, fluktuiert stark in den einzelnen Gruppen (s. Abschnitt 6.2). Es existieren deutliche Unterschiede in den einzelnen Gruppen, sodass dem Anwender klar wird, welche Gruppen von Aussagen besonders wichtig sind, um Anforderungen zu erheben. Allerdings ist die Bildung der Gruppen für den Anwender nicht nachvollziehbar. Außerdem ist die Punktzahl singulär betrachtet nicht aussagekräftig **(Verständlichkeit: teilweise erfüllt)**. Die Zielwörter in den Gruppen sind nicht konsistent. Es werden Zielwörter dargestellt, die sich auf Komponenten (z. B. „filter“) oder Funktionen (z. B. „get stuck“ oder „cleaning“) des Produkts beziehen **(Konsistente Repräsentation: teilweise erfüllt)**. Es werden relevante Aussagen hervorgehoben. Teilweise werden relevante Aussagen aus verschiedenen Rezensionen extrahiert, welche eine inhaltlich ähnliche Aussage treffen. Diese müssen für eine kompakte Darstellung weiter selektiert werden (z. B. durch die Berechnung der semantischen Ähnlichkeit der Aussagen [27]) **(Kompakte Repräsentation: erfüllt)**.

Die Benutzerrezensionen sind auf Amazon frei verfügbar **(Verfügbarkeit: erfüllt)**. Aktuell verstößt die Nutzung der Rezensionen nicht gegen die Geschäftsbedingungen von Amazon **(Zugänglichkeit hinsichtlich Sicherheit: erfüllt)**.

Gruppen / Informationen	Gruppe 1	Gruppe 2	Gruppe 3
Anzahl Benutzerrezensionen	9	32	83
Durchschnittliche Bewertung	4,22	3,91	3,5
Gruppen-Punktzahl	2,13	8,18	23,72
Repräsentative Ziel-Wörter	„filter“	„hair“, „area“, „dog“, „kitchen“	„works“, „get stuck“, „buy“, „base“, „goes back“
Repräsentative Aussagen	„Only reason I gave it a 4 star rating is because I believe the filter should be bigger.“ „This model has a filter that needs to be removed every time the dirt bin is emptied. “I also like that the filter and vacuum dirt can be cleaned out from lifting up the top and emptying out the middle.”	„It picks up fine dirt and animal hair great.” „It doesnt hold enough dirt and hair so I have to change it out often.” „It’s not good with dog hair.This is awful, its not my first robo vac, so I know it can be better.”	“It does a wonderful job at picking stuff up but it is constantly stuck.” „I love that it’s sweeps and mops for me I come home to a nice clean floor every day, I love it!Works, but gets stuck bumping into the same spots alot.”

Bild 2 Ausschnitt aus den Ergebnissen des Fallbeispiels: Es werden drei von zwölf Gruppen von Ziel-Wörtern dargestellt. Die Gruppen werden aufgrund der niedrigen, bzw. hohen Gruppen-Punktzahl (in Relation zum Mittelwert) ausgewählt. Die Aussagen in Gruppe 2 sind teilweise konfliktär. Die Ziel-Wörter in Gruppe 3 sind auf Funktionen bezogen, z. B. „get stuck“

7 Fazit und Ausblick

Die IQ-Merkmale werden als teilweise oder voll erfüllt bewertet. Die Ergebnisse zeigen demnach, dass Requirement Mining-Ansätze für die Entwicklung mechatronischer Produkte angewendet werden können. Der Beleg für die Anwendbarkeit von Requirement Mining eröffnet eine neue Forschungsperspektive hinsichtlich eines intelligenten Requirement Engineering und der Verwendung von Künstlicher Intelligenz in der mechatronischen Produktentwicklung. Zur weiteren Verbesserung der Ergebnisse des Ansatzes von CHEN ET AL. hinsichtlich der IQ-Merkmale können zusätzliche Methoden wie beispielsweise eine Klassifikation der Aussagen zur weiteren Reduzierung hinsichtlich Relevanz für den Anwender angewandt werden. Außerdem kann durch die Erkennung von Anforderungsabhängigkeiten Konflikte in den aufgenommenen Anforderungen aufgezeigt werden.

8 Literatur

[1] Wellsandt, S.; Hribernik, K. A.; Thoben, K.-D.: "*Qualitative Comparison of Requirements Elicitation Techniques that are Used to Collect Feedback Information about Product Use".* In: Procedia CIRP, Jg. 21, 2014, S. 212-217.

[2] Chen, R.; Wang, Q.; Xu, W.: "*Mining user requirements to facilitate mobile app quality upgrades with big data".* In: Electronic Commerce Research and Applications, Jg. 38, 2019.

[3] Meth, H.; Mueller, B.; Maedche, A.: "*Designing a Requirement Mining System".* In: Journal of the Association for Information Systems, Jg. 16, Nr. 9, 2015, S. 799-837.

[4] Machi, L. A.; McEvoy, B. T.: "*The literature review*". Corwin, Thousand Oaks, Calif., 2012.

[5] Fu, B.; Lin, J.; Li, L.; Faloutsos, C.; Hong, J.; Sadeh, N.: "*Why people hate your app".* In: Mining Data Semantics in Heterogeneous Information Networks Workshop (MDS' 2013, ACM Press, New York, USA, 2013, S. 1276.

[6] Carreño, L. V. G.; Winbladh, K.: "*Analysis of user comments: An approach for software requirements evolution*". In: *35th International Conference on Software Engineering (ICSE),* 2013, S. 582-591.

[7] Jiang, W.; Ruan, W.; Zhang, L.; Lew, P.; Jiang, J.: "*LNAI 8444 - For User-Driven Software Evolution: Requirements Elicitation Derived from Mining Online Reviews".* In: Advances in Knowledge Discovery and Data Mining, 2014

[8] Qi, J.; Zhang, Z.; Jeon, S.; Zhou, Y.: "*Mining customer requirements from online reviews: A product improvement perspective".* In: Information & Management, Jg. 53, Nr. 8, 2016, S. 951-963.

[9] Gu, X.; Kim, S.: *"What Parts of Your Apps are Loved by Users?".* In: 30th IEEE/ACM International Conference on Automated Software Engineering, IEEE, Piscataway, NJ, 2015, S. 760-770.

[10] Casamayor, A.; Godoy, D.; Campo, M.: "*Identification of non-functional requirements in textual specifications: A semi-supervised learning approach".* In: Information and Software Technology, Jg. 52, Nr. 4, 2010, S. 436-445.

[11] Tizard, J.; Wang, H.; Yohannes, L.; Blincoe, K.: "*Can a Conversation Paint a Picture? Mining Requirements In Software Forums*". In: 27th International Requirements Engineering Conference (RE), IEEE, 2019, S. 17-27.

[12] Panichella, S.; Di Sorbo, A.; Guzman, E.; Visaggio, C. A.; Gall, H. C.: "*ARdoc: App Reviews Development Oriented Classifier*", In: 24th ACM SIGSOFT International Symposium on the Fundations of Software Engineering, 2016.

[13] Jiang, C.; Liu, Y.; Ding, Y.; Liang, K.; Duan, R.: "*Capturing helpful reviews from social media for product quality improvement: a multi-class classification approach".* In: International Journal of Production Research, Jg. 55, Nr. 12, 2017,

[14] VDI-Fachbereich Produktentwicklung und Mechatronik: „*VDI 2206*". Beuth Verlag, 2020.

[15] Graessler, I.; Hentze, J.: "*The new V-Model of VDI 2206 and its validation". In:* at - Automatisierungstechnik, Jg. 68, Nr. 5, 2020, S. 312-324.

[16] Pohl, K.: „*Requirements engineering*". dpunkt-Verl., Heidelberg, 2008.

[17] Grande, M.: „*100 Minuten für Anforderungsmanagement*". Vieweg+Teubner Verlag / Springer Fachmedien Wiesbaden GmbH, Wiesbaden, 2011.

[18] *IEEE Standard Glossary of Software Engineering Terminology".* In: IEEE Std 610.12-1990.

[19] Walden, D. D.; Roedler, G. J.; Forsberg, K.: "*INCOSE Systems Engineering Handbook Version 4: Updating the Reference for Practitioners".* In: INCOSE International Symposium, Jg. 25, Nr. 1, 2015, S. 678-686.

[20] Janzen, A.; Hoffmann, A.; Hoffmann, H.: „*Anforderungsmuster im Requirements Engineering*". In: *Working Paper Series*, Kassel, Deutschland, 2013.

[21] Ehrlenspiel, K.: „*Integrierte Produktentwicklung*". Hanser, München, 2007.

[22] Statista: „*Umsatzstärkste Online-Shops im Elektronik- und Medien-Segment weltweit 2018".* URL: https://de.statista.com/prognosen/860274/top-online-shops-weltweit-elektronik-medien-ecommercedb, Abrufdatum 21.12.2020.

[23] Naumann, F.: „*Datenqualität".* URL: https://gi.de/informatiklexikon/datenqualitaet, 2007, Abrufdatum 27.01.2020.

[24] J. Reitmeier; K. Paetzold: "*Evaluation of Data Quality in the Context of Continous Product Validation throughout the Development Process".* In: Proceedings of the 18th International Conference on Engineering Design (ICED 11), 2011.

[25] He, S.; Hollenbeck, B.; Proserpio, D.: "*The Market for Fake Reviews,* 2020.

[26] Sabisch, H.; Tintelnot, C.: „*Integriertes Benchmarking*". Springer, Berlin, Heidelberg, 1997.

[27] Kusner, M.; Sun, Y.; Kolkin, N.; Weinberger, K.: "*From Word Embeddings To Document Distances". In:* International Conference on Machine Learning, 2015, S. 957-966.

Featureextraktion aus heterogenen multivariaten Zeitreihendaten mechatronischer Systeme mittels Autoencoder-Netzwerken

Autoencoder-based Feature Extraction from Heterogeneous Multivariate Time Series Data of Mechatronic Systems

Karl-Philipp Kortmann, M. Sc.[†‡], Moritz Fehsenfeld, M. Sc.[†] und Dr.-Ing. Mark Wielitzka[†]

[†] Leibniz Universität Hannover, Institut für Mechatronische Systeme, An der Universität 1, 30823 Garbsen, Deutschland
[‡] kortmann@imes.uni-hannover.de

Kurzfassung

Sensor- und Steuerungsdaten moderner mechatronischer Systeme liegen häufig als heterogene Zeitreihen vor und differieren in Abtastrate und Wertebereich. Für prädiktive Aufgaben, etwa im Rahmen eines *Condition Monitorings*, existieren bereits geeignete Klassifikations- und Regressionsverfahren aus dem Bereich des überwachten maschinellen Lernens, deren Leistungsfähigkeit allerdings stark mit der Anzahl gelabelter Trainingsdaten skaliert. Deren Bereitstellung ist oft mit hohem Aufwand in Form von Personenstunden oder Zusatzsensorik verbunden. In dieser Arbeit wird ein Verfahren zur unüberwachten Featureextraktion mittels Autoencoder-Netzwerken vorgestellt, das speziell dem heterogenen Charakter einer Datenbasis Rechnung trägt und die notwendige Menge an gelabelten Trainingsdaten gegenüber bestehenden Verfahren reduziert. Zur Validierung der Ergebnisse werden drei öffentliche Datensätze mechatronischer Systeme aus unterschiedlichen Anwendungsfeldern verwendet.

Abstract

Sensor and control data of modern mechatronic systems are often available as heterogeneous time series with different sampling rates and value ranges. Suitable classification and regression methods from the field of supervised machine learning already exist for predictive tasks, for example in the context of condition monitoring, but their performance scales strongly with the number of labeled training data. Their provision is often associated with high effort in the form of person-hours or additional sensors. In this paper, we present a method for unsupervised feature extraction using autoencoder networks that specifically addresses the heterogeneous nature of the database and reduces the amount of labeled training data required compared to existing methods. Three public datasets of mechatronic systems from different application domains are used to validate the results.

1 Einleitung

Moderne mechatronische Systeme verfügen über eine Vielzahl interner Sensor- und Steuerungsgrößen, auf die etwa über Feldbusschnittstellen zugegriffen werden kann. Zusätzlich lassen sich diese Systeme um externe Messsysteme erweitern, was in Summe zu einem heterogenen Bestand an multivariaten Zeitreihensignalen (*multivariate time series, MTS*) führt, wobei heterogen hierbei divergente Abtastzeiten, Messauflösungen oder Skalenniveaus der einzelnen univariaten Signalen meint.

Im Zuge einer fortschreitenden Digitalisierung werden solche Anlagensignale im zunehmenden Maße für Klassifikationsaufgaben (etwa *Alarm-* oder *Condition Monitoring*) oder zur Schätzung nicht direkt messbarer Zielgrößen (z. B. die Prozessstabilität oder -qualität) eingesetzt. Im Folgenden werden diese zusammenfassend als *Prädiktionsaufgaben* bezeichnet.

Die hierfür primär genutzten Prädiktionsverfahren aus dem Bereich des statistischen beziehungsweise maschinellen Lernens sind gerade bei einer hochdimensionalen, heterogenen Datenbasis auf eine umfangreiche Featureextraktion angewiesen [14]. Dies setzt im Fall einer manuellen Featureextraktion eine hohe fachliche Expertise bezüglich des vorliegenden Systems voraus und geht folglich mit einem hohen Aufwand während der Implementierung einher. Alternativ kommen überwachte End-to-End-Prädiktionsverfahren aus dem Bereich des Deep Learnings zwar explizit ohne zuvor extrahierte Features aus, benötigen aber eine große Anzahl bereits gelabelten Daten für das Training [15].

Methoden des *Representation Learning* haben sich in jüngster Zeit im Bereich der Bild- und Sprachverarbeitung etabliert und ermöglichen eine unüberwachte Featureextraktion, welche die Prädiktionsgüte gängiger Verfahren der manuellen und automatisierten Featureextraktion im Fall von nur wenigen vorhandenen gelabelten Daten übertrifft [1]. Die Anwendung auf mechatronische Systeme beschränkt sich bisher auf spezialisierte Einzellösungen [2], die sich nicht unmittelbar auf beliebige Systeme übertragen lassen.

2 Stand der Forschung

2.1 Representation Learning

Unter dem Begriff des Representation- oder Feature Learnings werden Methoden zusammengefasst, die eine automatische Extraktion relevanter Features ermöglichen und

den Schritt des manuellen Feature Engineerings somit prinzipiell erübrigen. Im Kontext dieser Arbeit liegt der Fokus auf einer unüberwacht arbeitenden Methode, die also ohne Kenntnis der Zielvariable und somit nur anhand der Eingangszeitreihen eine möglichst kompakte Repräsentation der MTS lernt.
Ürsprünglich aus dem Bereich des maschinellen Sehens (*computer vision*) stammend, werden Verfahren des Feature Learnings zunehmend auch für Prädiktionsaufgaben auf Basis von Zeitreihendaten adaptiert; so verwendeten etwa Chen et al. [2] einen Autoencoder zur Featureextraktion aus den Momentensignalen eines 6-Achs-Industrieroboters zur Prädiktion von Kollisionen im Arbeitsraum. Li et al. [18] und Jiang et al. [17] verwenden jeweils ein *Generative Adverssarial Netwerk* (*GAN*) zur Featureextraktion aus industriellen Zeitreihendaten, während Franceschi et al. [1] ein rein Encoder-basiertes Netzwerk in Kombination mit einer sogenannten *Triplet-Loss-Funktion* für die Klassifikation auf diversen Referenzzeitreihen nutzen.
In dieser Arbeit konzentrieren wir uns auf einen Autoencoder-basierten Ansatz, da diese im Allgemeinen im Gegensatz etwa zu GANs eine bessere Repräsentation der Grundgesamtheit der Trainingsdaten ermöglichen, häufig unter Inkaufnahme einer schlechteren Performance als rein generative Modelle [19] (also beim Erzeugen realistischer neuer Zeitreihen), was aber in dieser Arbeit nicht im Fokus steht.

Autoencoder Ein Autoencoder ist ein künstliches Neuronales Netz, dessen primäres Ziel die Rekonstruktion eines Eingangssignales $\mathbf{x}$ darstellt (siehe Abbildung 1). Die dimensionsreduzierte latente Variable (auch *bottelneck* oder *latent space*) $\mathbf{z} = \phi(\mathbf{x}; \theta_{\mathrm{En}})$ ist das Ergebnis einer Encoderfunktion mit den Parametern (Gewichten) θ_{En} und wird im Rahmen des Representation Learnings in dieser Arbeit als Feature herangezogen. Die latente Variable durchläuft während des Trainings anschließend einen Decoder $\tilde{\mathbf{x}} = \psi(\mathbf{z}; \theta_{\mathrm{De}})$ mit den Parametern θ_{De}, der eine Rekonstruktion $\tilde{\mathbf{x}}$ des Eingagnssignales erzeugt. Im vorliegenden Fall von reellen Eingangswerten, wird in der Regel der mittlere quadratische Fehler (*mean squared error, MSE*) als Rekonstruktionsfehler $L_{\mathrm{r}} = \mathrm{MSE}(\mathbf{x}, \tilde{\mathbf{x}})$ verwendet und dient als Verlustfunktion während der Optimierung.

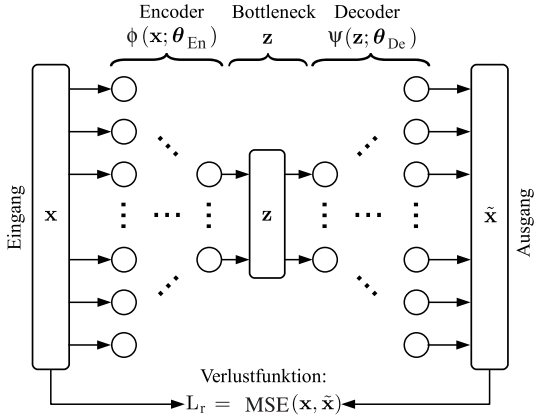


Abbildung 1 Netzwerkdiagramm eines simplen Autoencoders mit eindimensionalem Eingang und angedeuteten FC-Layern in Encoder und Decoder.

Neben vollständig vernetzten Schichten (*fully connected, FC* oder *dense layers*) mit nichtlinearen Aktivierungsfunktionen, finden auch komplexere neuronale Schichten Verwendung. So nutzen Bianchi et al. rekurrente (*RNN*), bidirektionale Schichten in Kombination mit einer zusätzlichen Kernel-Verlustfunktion für die Featureextraktion aus MTS mit fehlenden Werten [7].

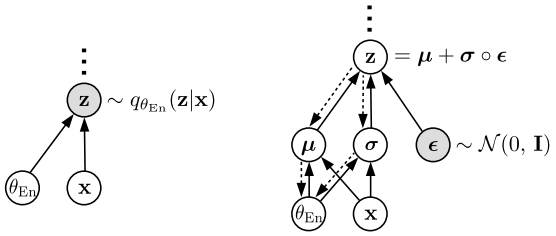
Variational Autoencoder Eine Erweiterung des herkömmlichen Autoencoders stellt der *Variational Autoencoder (VAE)* dar, der hauptsächlich als generatives Modell im Bereich des maschinellen Sehens Anwendung findet. Als Beispiel eines *Variational-Bayes*-Modells modelliert der VAE die unbekannte Verteilungsfunktion der Eingangsdaten $\mathbf{x} \sim p^*(\mathbf{x})$ mittels einer Modellverteilung $p_\theta(\mathbf{x}) \approx p^*(\mathbf{x})$ [12]. Der stochastische Decoder kann somit als bedingte Wahrscheinlichkeitsverteilung $p_{\theta_{\mathrm{De}}}(\mathbf{x}|\mathbf{z})$ aufgefasst werden, die zusammen mit der Prior-Verteilung der latenten Variable $p_\theta(\mathbf{z})$ ein generatives Modell als multivariate Verteilung

$$p_\theta(\mathbf{x}, \mathbf{z}) = p_\theta(\mathbf{z}) p_{\theta_{\mathrm{De}}}(\mathbf{x}|\mathbf{z}) \tag{1}$$

faktorisiert. Analog dazu stellt der Encoder ein Inferenzmodell dar, das als bedingte Wahrscheinlichkeitsverteilung der latenten Variablen bei gegebenen Eingangsdaten $q_{\theta_{\mathrm{En}}}(\mathbf{z}|\mathbf{x})$ aufgefasst werden kann.
Dieser Ansatz führt über die Anwendung der *evidence lower bound, ELBO* zu der abgewandelten Verlustfunktion [12]

$$L_{\theta_{\mathrm{En}}, \theta_{\mathrm{De}}}(\mathbf{x}) = D_{\mathrm{KL}}(q_{\theta_{\mathrm{En}}}(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})) - \mathbb{E}_{q_{\theta_{\mathrm{En}}}(\mathbf{z}|\mathbf{x})}\left(\log p_{\theta_{\mathrm{De}}}(\mathbf{x}|\mathbf{z})\right), \tag{2}$$

mit der Kullback-Leibler-Divergenz D_{KL}, welche die Abweichung zwischen vorgegebener Prior-Verteilung der latenten Variable $\mathbf{z}$ und tatsächlicher Verteilung durch den Encoder bestraft. Der zweite Term repräsentiert den Rekonstruktionsfehler. Als Prior für die latente Variable wird in der Regel die multivariate Standardnormalverteilung $p_\theta(\mathbf{z}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ genutzt.
Für das Training mittels des üblichen stochastischen Gradientabstiegsverfahren (*stochastic gradient descent, SGD*) wird der Gradient der Verlustfunktion $\nabla L_{\theta_{\mathrm{En}}, \theta_{\mathrm{De}}}(\mathbf{x}_{\mathrm{mb}})$ für jedes Mini-Batch an Trainingsdaten $\mathbf{x}_{\mathrm{mb}}$ gebildet, um eine Minimierung der Verlustfunktion in Abhängigkeit der Netzwerkparameter θ_{En} und θ_{De} durchzuführen (*Backpropagation*). Wie Abbildung 2a zu entnehmen ist, ist dies bei einem direkten Ziehen von $\mathbf{z} \sim q_{\theta_{\mathrm{En}}}(\mathbf{z}|\mathbf{x})$ nicht möglich, da die Backpropagation durch die Zufallsvariable $\mathbf{z}$ unterbrochen wird [16]. Erst mit einer mathematisch äquivalente Reparametrisierung durch Auslagerung der Zufallsvariation in die Zufallsvariable ϵ, die in der Regel als normalverteilt modelliert wird, ist eine Rückführung des Fehlers durch den Encoder möglich (sogenannter *reparameterization trick*, siehe Abbildung 2b).



(a) **(b)**

Abbildung 2 **(a)**: Das direkte Sampling von $\mathbf{z}$ macht die Backpropagation (⇢) der Verlustfunktion unmöglich. **(b)**: Erst mit der Reparametrisierung von $\mathbf{z}$ wird die Optimierung der Encoderparameter möglich. Unabhängige Zufallsvariablen sind grau hinterlegt. Basierend auf [13].

3 Methodik

In diesem Abschnitt wird das eigene Autoencoder-basierte Modell, sowie die einbettende Pipeline vorgestellt. Es folgen Erläuterungen zur Versuchsdurchführung, insbesondere zu den gewählten Vergleichsmethoden und den gewählten öffentlich verfügbaren Datensätzen.

3.1 Featureextraktion

Die unüberwachte Extraktion der Features aus einer MTS mit der Anzahl an univariaten Signalen n_{sig} erfolgt mithilfe mehrere VAE, die separat für jede univariate Zeitreihe trainiert werden. Dies ermöglicht im Gegensatz zu gängigen zweidimensional-faltenden Netzarchitekturen (*convolutional neural networks, CNN*) eine Parallelisierung des Trainings und eine individuelle Architektur der Netzwerke, je nach Abtastrate der heterogenen Signale.
Als zusammengefasstes Feature wurden während der späteren Inferenz die 1D-Verkettung der einzelnen latenten Variablen $\mathbf{x}_i,\ i \in \{1, ..., n_{\text{sig}}\}$ genutzt. In Tabelle 1 sind die gewählten (Hyper-)Parameter des entwickelten Modells angegeben. Diese wurden über alle Datensätze konstant gehalten, um eine Allgemeingültigkeit bestmöglich zu demonstrieren.

Anstatt eine konstante Dimension des latenten Raumes vorzugeben, wurde eine konstante Kompressionsrate κ gewählt, sodass die Anzahl an extrahierten Features dynamisch mit der Abtastrate und Signaldauer skaliert. Des Weiteren wurde während Training und Inferenz eine *Instance-Normalization* für jede gefensterte Eingangszeitreihe durchgeführt, um divergenten Wertebereichen der Signale Rechung zu tragen.

Tabelle 1 Zusammenfassung der wichtigsten (Hyper-)Parameter jedes univariaten VAE-Modells

(Hyper-)Parameter	Wert	Bemerkung
Eingangsdimension	$1 \times *$	$*$: maximale Fensterlänge
Kompressionsrate κ	25	Quotient aus $*$ und dim($\mathbf{z}$)
Hidden Layer	$[*/2, */2]$	zwei Hidden Layer
Aktivierungsfunk.	-	tanh bzw. lin. am Ausgang
Regularisierung	-	Early stopping und L_2-Norm
Optimierer	Adam	SGD-Optimierer [20]
Batch-size	$64 - 512$	$\sim n_{\text{train}}$
Normalisierung	-	Instance-/Layer Norm.
Lernrate	1×10^{-4}	kein Lernrate-Scheduler
max. Epochen	1×10^{3}	beachte: Early stopping

3.2 Pipeline

Der Ablauf von Modellierungs-/Trainings- und Inferenzphase ist typisch für ein sogenanntes *semi-supervised* Trainingsverfahren, bestehend aus unüberwachtem Representation Learning und überwachtem Training eines Prädiktors. Nach dem konsekutiven Training beider Modelle werden diese während der Inferenz (hier auf unabhängigen Testdatensätzen) unverändert angewendet. Die Menge an gelabelten Trainingsdaten wird während der Versuchsreihe variiert (logarithmische Skalierung).
Die Feature Learning Methode (VAE) bekommt zuvor sämtliche Trainingsdaten ohne Label zur Verfügung gestellt um Repräsentationen zu lernen. Dies entspricht dem realitätsnahen Anwendungsfall, dass eine Vielzahl an Rohdaten vorliegt, aber nur ein gewisser Anteil davon gelabelt wurde. Für jeden Datensatz, jede Methode und Menge gelabelter Trainingsdaten wurden $n_{\text{repeat}} = 10$ Wiederholungen durchgeführt, sodass empirischer Mittelwert und empirische Standardabweichung der jeweiligen Metrik berechnet werden können. Die Training-/Testaufteilung der Datensätze wurde von den Autoren selbiger übernommen.
Als Entwicklungsumgebung wurde Python 3.8[1] auf einem Rechner mit GPU-Unterstützung (CUDA 7.5) verwendet. Der Quelltext sowie die Links zu den verwendeten Datensätzen werden öffentlich zur Verfügung gestellt[2].

3.3 Vergleichsmethoden

Das vorgestellte Modell sowie die Pipeline werden mit einer Methode der Featureextraktion aus dem aktuellen Stand der Forschung verglichen. Zusätzlich dient die Hauptkomponentenanalyse als deterministische Vergleichsmethode. Die Python-Bibliothek *tsfresh* ermöglicht eine automatische Extraktion und Auswahl statistischer Zeit- und Frequenzbereichsfeatures und stellt somit eine Vergleichsmethode aus dem Bereich der manuellen Featureextraktion dar.
Bei allen Vergleichsmethoden wird ein Ridge-Regression-basierter Prädiktor in derjenigen Parameterisierung verwendet, da sie insbesondere von Autoren mehrerer aktueller Featureextraktoren als favorisierter Prädiktor empfohlen wird. [10, 11]. Für den VAE wird eine *support vector machine (SVM)* mit Gauß-Kern genutzt, um dem erzwungenen Verteilungscharakter der latenten Variable besser Rechnung zu tragen[3].

PCA Mithilfe der Hauptkomponentenanalyse (*principle component analysis, PCA*) lassen sich Zeitreihendaten durch Singulärwertzerlegung in ihrer Dimension reduzieren. Die PCA erzeugt diejenige orthogonale lineare Transformation der Eingangsdaten, welche die Varianz der Kom-

[1] Hierbei insbesondere: torch 1.7.1, sktime 0.4.3 und sklearn 0.24.1.
[2] `https://github.com/MrPr3ntice/vae_rep_learn_mts`
[3] Insbesondere können aufgrund des symmetrischen Kerns im Falle einer Klassifikation geschlossene Intervalle auf einer Variablen separiert werden.

Tabelle 2 Übersicht der ausgewählten Datensätze; $n_{\max}$ gibt die maximale Anzahl an Datenpunkten einer Zeitreihe an.

Datensatz	Trainings-samples	Test-samples	Anzahl Signale	Dauer pro Sample ($n_{\max}$)	Zielgröße (Art)	Referenz
Wälzlagerschäden	1440	800	7 Signale	0.2 s (800)	Wälzlagerzustand (3 Klassen)	[3]
Schrittmotoren	70152	23384	7 Signale	6 ms (60)	Betriebsmodus (4 Klassen)	[4]
Hydraulik	1544	661	17 Signale	60 s (1200)	Kühlerzustand (3 Klassen), Druck im Hydraulikspeicher (Regr.)	[5]

ponenten des Zielunterraums absteigend maximiert. Die somit gewonnene Abbildung kann als Feature für eine Klassifikation oder Regression genutzt werden. Die PCA kann insbesondere als Spezialfall eines linearen Autoencoders ohne Hidden Layer betrachtet werden [12], weshalb sie hier als Vergleichsmethode dienen soll.

Statistische Feature Die Extraktion von manuellen, statistischen Features aus Zeitreihen für ML-Anwendungen ist weit verbeitet. Die Python-Bibliothek *tsfresh* bietet eine Sammlung etablierter Extraktionsmethoden, um eine Vielzahl dieser Features automatisch aus Zeitreihendaten zu generieren und auszuwählen. Die Sammlung der Features enthält z. B. statistische Kennzahlen und Korrelationen im Zeit- und Frequenzbereich. Eine vollständige Übersicht über die Features kann der Dokumentation entnommen werden [8]. Im vorliegenden Fall wurden die Standardeinstellungen[4] genutzt.

ROCKET ***R**and**O**m **C**onvolutional **KE**rnel **T**ransform* ist ein State-of-the-Art-Verfahren, das mithilfe von zufällig gesampleten Faltungskernen Features aus Zeitreihen extrahiert. Anschließend wird eine logistische Regression (Ridge-Regression) mit den erzeugten Features trainiert. Durch diesen vergleichsweise einfachen Aufbau ist der Rechenaufwand von *ROCKET* geringer als bei vergleichbar performanten Verfahren [11]. Zum Stand dieser Arbeit erzeugt diese Methode die durchschnittlich besten Klassifikationsergebnisse auf den Datensätzen des UCR und UEA Zeitreihenarchives [21].

3.4 Datensätze

Die Anzahl der öffentlich zugänglicher Datensätze ist insbesondere im Bereich mechanischer und elektronischer Systemen beschränkt. Zur Validierung der Methode wurden aus den verfügbaren Datensätzen solche ausgewählt, die ein breites Spektrum mechatronischer Anwendungen abgedeckt. Alle Datensätze bestehen aus Messdaten mehrerer Sensoren, die sich z. B. in den Abtastraten unterscheiden. Daraus resultieren drei multivariate Datensätze aus heterogenen Zeitreihen, aus denen drei Klassifikations- und eine Regressionsaufgabe abgeleitet werden. Eine Übersicht der ausgewählten Datensätze findet sich in Tabelle 2.

Wälzlagerschäden Ein weitverbreiteter Anwendungsfall für Verfahren des maschinellen Lernens ist die Erkennung von Wälzlagerschäden. Ein umfassender Referenzdatensatz, der diesen Anwendungsfall abbildet, wurde von Leissmeier et al. [3] veröffentlicht. Zur Schadensklassifikation stehen neben hochfrequent abgetasteten Messungen der Motorströme und Gehäusevibration ($f_s = 64\,\text{kHz}$), zusätzliche, niederfrequentere Daten wie z. B. die Radialkraft ($f_s = 4\,\text{kHz}$) und die Temperatur ($f_s = 1\,\text{Hz}$) zur Verfügung. Der Datensatz umfasst unterschiedliche Schadensarten von Wälzlagern am Außen- und Innenring sowie die zugehörigen Gutdaten unter verschiedenen Betriebsbedingungen. Weitergehende Informationen zum Datensatz sind [3] zu entnehmen.

Schrittmotoren Für die Überwachung von Schrittmotoren existiert ein von Goubeaud et al. [4] publizierter Datensatz. Dieser beinhaltet Messungen des Stroms, der Spannung und der Vibration (Beschleunigung). Die Zielgröße ist der Betriebsmodus des Schrittmotors, der sich zum einen in einen Betrieb im- und gegen den Uhrzeigersinn unterscheidet sowie in einen Betrieb im Normalbereich und jenseits des mechanischen Anschlags. Eine ausführliche Beschreibung des Versuchsaufbaus und der Messdatenaufnahme findet sich in der Erstveröffentlichung [4].

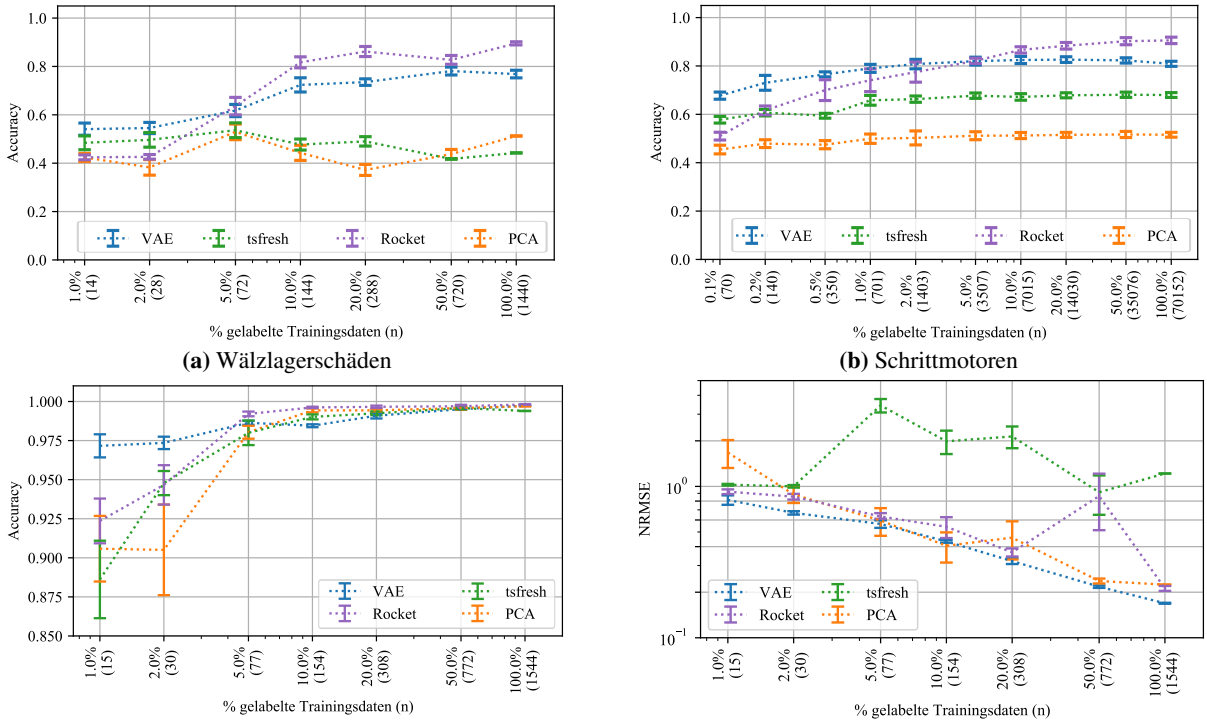
Hydraulik Das in [5] beschriebene hydraulische System ist mit einer Vielzahl verschiedener Sensoren ausgestattet. Neben Messungen des Drucks und der Volumenströme werden zusätzlich Temperatur, Strom und Vibration erfasst. Die Abtastraten reichen von $f_s = 1\,\text{Hz}$ (Temperatur) bis zu $f_s = 20\,\text{Hz}$[5] (Druck), woraus sich ein heterogener Datensatz mit unterschiedlichsten Sensordaten ergibt. Als Zielgrößen sollen an dieser Stelle der Zustand des Kühlers (Klassifikation) und der Druck im Hydraulikspeicher (Regression) betrachtet werden.

4 Ergebnisse

Die Ergebnisse sind in Abbildung 3 (**a**) - (**d**) dargestellt. Dateilierte Ergebnisse können den Tabellen 3 und 4 entnommen werden.
Für die ersten drei Fälle erreicht die Vergleichsmethode ROCKET die höchsten Testergebnisse hinsichtlich der mittleren Accuracy bzw. die niedrigste mittlere Fehlerrate. Lediglich im Fall der Regression des Druckes im Hydraulikspeicher zeigt der VAE fast durchgehend den niedrigsten quadratischen Fehler. Hierzu sei angemerkt, dass ROCKET für Klassifikationsaufgaben entwickelt und bisher mehrheitlich angewendet wurde. Die Hauptkomponentenanalyse und die durch tsfresh ermittelten statistischen Feature sind für die Prädiktionsaufgaben der ersten beiden Datensätze nicht konkurrenzfähig; bei der Klassifikation auf den

[4] Extraktion: `efficient`, Selektion: `extract_relevant_features()`

[5] Hier wurde gegenüber des Originaldatensatzes von 100 Hz auf 20 Hz zugunsten der Rechendauer heruntergesampelt



(a) Wälzlagerschäden

(b) Schrittmotoren

(c) Hydraulik Klassifikation

(d) Hydraulik Regression

Abbildung 3 Ergebnisse der vier verglichenen Featureextraktoren auf den drei Testdatensätzen für variable Anteile an gelabelten Trainingsdaten (logarithmische Skalierung). Die Diagramme **(a) - (c)** zeigen die Accuracy für die Klassifikationsaufgaben; **(d)** zeigt den normalisierten RMSE (log. Skalierung). Dargestellt sind jeweils empirische Mittelwert und Standardabweichung für $n_{\text{repeat}} = 10$.

Hydraulikdaten erreichen alle Methoden für einen höheren Anteil gelabelter Trainingsdaten ähnlich niedrige Fehlerraten.

Bei Betrachtung der Ergebnisse für eine geringe Anzahl gelabelter Trainingsdaten zeigt sich für alle Datensätze, dass die vorgestellte Implementierung des VAEs im Bereich bis mindestens 2 % eine höhere Prädiktionsgüte auf Basis der extrahierten Features bildet. Dies stützt die Forschungshypothese, dass Autoencoder-basierte Methoden des Representation Learnings zur Extraktion von Features aus heterogenen, multivariaten Zeitreihen besonders für sehr kleine Mengen an vorhandenen gelabelten Trainingsdaten geeignet sind.

5 Fazit

In dieser Arbeit wurde ein unüberwachter, Autoencoder-basierter Featureextraktor für Prädiktionsaufgaben auf heterogenen, multivariaten Messzeitreihen mechatronischen Ursprungs vorgestellt und auf drei Datensätzen aus der Forschungsgemeinschaft validiert. Auch wenn die Prädiktionsgüte auf Basis der gesamten Trainigsdaten nicht an aktuelle State-of-the-Art-Extraktoren heranreicht, so konnte doch eine erhöhte Güte für den Fall eines geringen Bestandes an gelabelten Trainingsdaten bei hoher Verfügbarkeit ungelabelter Daten gezeigt werden.

Die hier gezeigten Ergebnisse wurden ausschließlich mit aufbereiteten (Referenz-)Datensätzen erzielt. Ein nächster logischer Schritt hin zu automatisierten Prädiktionsmethoden für technisch-industrielle Anwendungen ist somit die Anpassung der Methode für nicht vorverarbeitete Rohdaten. Hierbei bietet sich etwa das Verfahren des *Denoising Autoencoders* als mögliche Erweiterung an.

6 Literatur

[1] Franceschi, J.-Y.; Dieuleveut, A.; Jaggi, M.: *Unsupervised Scalable Representation Learning for Multivariate Time Series*. Advances in Neural Information Processing Systems. (2019) 32, ISSN 1049-5258.

[2] Chen, T.; Liu, X.; Xia, B.; Wang, W.; Lai, Y.: *Unsupervised Anomaly Detection of Industrial Robots Using Sliding-Window Convolutional Variational Autoencoder*. IEEE Access. (2020) 8, ISSN 2169-3536, S. 47072–47081.

[3] Lessmeier, C.; Kimotho, J.; Zimmer,D.; Sextro, W.: *Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors*. European Conference of the Prognostics and Health Management Society. (2016).

[4] Goubeaud, M.; Grunert, T.; Lützenkirchen, J.; Joußen, P.; Ghorban, F.; Kummert, A.: *Introducing a New Benchmarked Dataset for Mechanical Stop Detection of Stepper Motors*. 2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS). (2020), S. 1–4.

[5] Helwig, N.; Pignanelli, E.; Schütze, A.: *Condition monitoring of a complex hydraulic system using multivariate statistics*. 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings. (2015) 24, ISSN, S. 210–215.

[6] Dong, Y.; Qin, S. J.: *A novel dynamic PCA algorithm for dynamic data modeling and process monitoring*. Journal of Process Control. (2018) 67, S. 1–11.

[7] Bianchi, F. M.; Livi L.; Mikalsen, K. Ø.; Kampffmeyer M.;

Tabelle 3 Testergebnisse der Klassifikationsaufgaben mit allen drei Datensätzen. Angegeben ist jeweils Mittelwert und Standardabweichung der Accuracy für unterschiedliche Anteile an gelabelten Trainingsdaten.

% gelabelter Trainingssamples (n_{train})	VAE + SVM	PCA + Ridge	tsfresh + Ridge	Rocket + Ridge
Datensatz Wälzlagerschäden				
1.0% (14)	**.541 ± .023**	.424 ± .015	.484 ± .026	.424 ± .009
2.0% (28)	**.545 ± .021**	.383 ± .029	.496 ± .027	.426 ± .008
5.0% (72)	.618 ± .023	.529 ± .029	.536 ± .027	**.636 ± .032**
10.0% (144)	.724 ± .026	.443 ± .028	.477 ± .020	**.817 ± .020**
20.0% (288)	.735 ± .012	.372 ± .020	.490 ± .017	**.862 ± .019**
50.0% (720)	.781 ± .015	.438 ± .017	.417 ± .001	**.827 ± .016**
100.0% (1440)	.768 ± .014	.512 ± .000	.443 ± .000	**.895 ± .006**
Datensatz Schrittmotoren				
0.1% (70)	**.677 ± .013**	.454 ± .016	.578 ± .013	.510 ± .014
0.2% (140)	**.730 ± .025**	.479 ± .014	.607 ± .012	.616 ± .016
0.5% (350)	**.765 ± .010**	.474 ± .015	.594 ± .010	.700 ± .033
1.0% (701)	**.790 ± .015**	.499 ± .017	.658 ± .017	.741 ± .036
2.0% (1403)	**.808 ± .017**	.502 ± .023	.663 ± .012	.774 ± .032
5.0% (3507)	.820 ± .014	.512 ± .015	.677 ± .011	**.821 ± .011**
10.0% (7015)	.825 ± .013	.512 ± .012	.672 ± .013	**.866 ± .012**
20.0% (14030)	.826 ± .011	.515 ± .010	.679 ± .010	**.883 ± .013**
50.0% (35076)	.823 ± .011	.517 ± .012	.681 ± .011	**.902 ± .013**
100.0% (70152)	.809 ± .010	.516 ± .010	.680 ± .010	**.906 ± .012**
Datensatz Hydraulik				
1.0% (15)	**.972 ± .006**	.906 ± .018	.886 ± .021	.924 ± .012
2.0% (30)	**.974 ± .003**	.905 ± .025	.948 ± .007	.947 ± .011
5.0% (77)	.986 ± .001	.980 ± .004	.980 ± .007	**.992 ± .001**
10.0% (154)	.984 ± .001	.994 ± .001	.990 ± .001	**.996 ± .000**
20.0% (308)	.991 ± .002	.994 ± .001	.992 ± .001	**.997 ± .001**
50.0% (772)	.995 ± .000	.996 ± .001	.996 ± .001	**.997 ± .001**
100.0% (1544)	.997 ± .000	.997 ± .000	.994 ± .000	**.998 ± .000**

Tabelle 4 Testergebnisse auf dem Hydraulikdatensatz. Angegeben ist jeweils Mittelwert sowie Standardabweichung des normalisierten RMSE für die Prädiktion des Druckes im Hydraulikspeicher für unterschiedliche Anteile an gelabelten Trainingsdaten.

% gelabelter Trainingssamples (n_{train})	VAE + SVR	PCA + Ridge	tsfresh + Ridge	Rocket + Ridge
1.0% (15)	**.814 ± .058**	1.674 ± .350	1.026 ± .013	.923 ± .033
2.0% (30)	**.669 ± .018**	.885 ± .107	1.002 ± .019	.856 ± .037
5.0% (77)	**.565 ± .033**	.595 ± .123	3.429 ± .350	.633 ± .022
10.0% (154)	.433 ± .009	**.405 ± .091**	1.986 ± .350	.541 ± .085
20.0% (308)	**.321 ± .014**	.458 ± .129	2.138 ± .350	.367 ± .023
50.0% (772)	**.217 ± .004**	.237 ± .009	.915 ± .266	.864 ± .350
100.0% (1544)	**.169 ± .001**	.224 ± .000	1.217 ± .000	.212 ± .008

Jenssen, R.: *Learning representations of multivariate time series with missing data*. Pattern Recognition. (2019) 96, ISSN 0031-3203.

[8] Christ, M.; Braun, N.; Neuffer, J.; Kempa-Liehr A.W.: *Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)*. Neurocomputing. (2018) 307, ISSN 1573-756X, S. 72–77.

[9] Dau, H. A. and Bagnall, A. and Kamgar, K. and Yeh, C. M. and Zhu, Y. and Gharghabi, S. and Ratanamahatana, C. A. and Keogh, E.: *The UCR time series archive*. IEEE/CAA Journal of Automatica Sinica. (2019) 6, S. 1293–1305.

[10] Le Nguyen, T.; Gsponer, S.; Ilie, I.; O'Reilly, M.; Ifrim, G.: *Interpretable time series classification using linear models and multi-resolution multi-domain symbolic representations*. Data Mining and Knowledge Discovery. (2019) 33, ISSN 1573-756X, S. 1183–1222.

[11] Dempster, A.; Petitjean, F.; Webb, G. I.: *ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels*. Data Mining and Knowledge Discovery. (2020) 34, ISSN 1573-756X, S. 1454–1495.

[12] Kingma, D. P.; Welling, M.: *An Introduction to Variational Autoencoders* . Foundations and Trends in Machine Learning. (2019) 12, No. 4, S. 307–392.

[13] Kingma, D. P.; Welling, M.: *Auto-encoding Variational Bayes*. (2013), arXiv preprint arXiv:1312.6114.

[14] Fawaz, H. I.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P. A.: *Deep learning for time series classification: a review*. Data Mining and Knowledge Discovery. (2019), 33(4), S. 917–963.

[15] Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S. X.: *An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data*. IEEE Transactions on Industrial Electronics. (2016), 63(5), S. 3137–3147.

[16] Rezende, D. J.; Mohamed, S., Wierstra, D.: *Stochastic back- propagation and approximate inference in deep generative models*. In: International Conference on Machine Learning. (2014), S. 1278—1286.

[17] Jiang, W.; Cheng, C.; Zhou, B.; Ma, G.; Yuan, Y.: *A novel GAN-based fault diagnosis approach for imbalanced industrial time series*. (2019), arXiv preprint arXiv:1904.00575.

[18] Li, D.; Chen, D.; Jin, B.; Shi, L.; Goh, J.; Ng, S. K.: *MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks*. In: International Conference on Artificial Neural Networks. (2019), Springer, Cham, S. 703–716.

[19] Grover, A.; Dhar, M.; Ermon, S.: *Flow-GAN: Combining maximum likelihood and adversarial learning in generative models*. In: Proceedings of the AAAI Conference on Artificial Intelligence. (2018) 32, 1.

[20] Kingma, D. P.; Ba, J.: *Adam: A method for stochastic optimization*. (2014), arXiv preprint arXiv:1412.6980.

[21] Ruiz, A. P.; Flynn, M.; Large, J.; Middlehurst, M.; Bagnall, A.: *The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances*. Data Mining and Knowledge Discovery. (2020), Springer OA, ISSN: 1573-756X, S. 1–49.

Vergleich struktureller Totzeitberücksichtigung und kausaler Datenvorverarbeitung bei der datenbasierten Modellierung von Zeitreihen

Comparison of Structural Time Lag Consideration and Causal Data Preprocessing in Data-based Modeling of Time Series

Johannes Zumsande, Prof. Dr.-Ing. Tobias Ortmaier, Dr.-Ing. Mark Wielitzka, Insitut für Mechatronische Systeme, Leibniz Universität Hannover, 30823 Garbsen, Deutschland, johannes.zumsande@imes.uni-hannover.de

Kurzfassung

Im Zuge der Digitalisierung moderner industrieller Produktionsanlagen finden datenbasierte Methoden verstärkt Anwendung. Die Daten weisen jedoch in der Regel einen Zeitversatz, insbesondere zwischen den einzelnen Stationen, auf. Obwohl Long Short-Term Memory (LSTM) prinzipiell fähig sind beliebig große Zeitverzüge zu berücksichtigen, versagen sie bei den großen Zeithorizonten der industriellen Produktion. Daher werden in dieser Veröffentlichung LSTM, bei denen die Zeitreihen zur Berücksichtigung einer stochastischen Totzeit verschoben wurden, und Feed-forward Neural Network (FNN), deren Datensatz zur Eliminierung der Totzeit vorverarbeitet wurde, untersucht. Die Methoden werden sowohl anhand simulierter als auch realer Zeitreihen eines Altsandkühlers verglichen. Die Ergebnisse zeigen, dass die Berücksichtigung von Zeitverzügen das Modellierungsergebnis durchweg verbessert. Das LSTM kann Informationen aus der Reihenfolge behalten, während das FNN für die Rückführung des Modellausganges besser geeignet ist.

Abstract

Due to the increasing digitalization in modern industrial production, data-based methods are increasingly used. However, the data relation is usually shifted by a time lag, especially between the individual stations. Although Long Short-Term Memory (LSTM) are theoretically capable considering arbitrarily large time lags between the time series, they fail in industrial production's reality for large time horizons. Therefore, this paper examines LSTM, where the time series have been shifted to consider a stochastic time lag, and Feed-forward Neural Network (FNN), whose dataset has been preprocessed to eliminate the time lag. The methods are compared using both simulated and real time series of a return sand cooler. The results show that the consideration of time delay improves the modeling result. The LSTM can consider information from the sequence, while the FNN is better for analyzing the model's output.

1 Einleitung

Getrieben von den Anforderungen nach immer verfügbaren, individuell gefertigten und qualitativ hochwertigen Produkten befindet sich die industrielle Produktion im Wandel. Die engen qualitativen Toleranzmaße bei einer gleichzeitig hohen Produktivität können nur durch einen gesteigerten Automatisierungsgrad und komplexe Prozessarchitekturen mit vielen Verarbeitungsschritten erreicht werden. Als Nebenprodukt der Automatisierung fallen große Mengen zugrundeliegender Daten an, die in Kombination mit weiteren heterogenen Datenquellen (z.B. Laborberichte, Logistikdaten, Störberichte) eine holistische Datengrundlage [12] für systematische Untersuchungen und Analysen, dem Data Mining, bilden [6, 14]. Aufgrund der zunehmenden Komplexität und Dimension der Prozesstopologien, müssen hierfür die Informationen einander zugeordnet werden. Zeitgleich erhobene Daten stehen nicht zwingend in direktem Zusammenhang.
Insbesondere die Zustands- und Qualitätsüberwachung [9, 18] profitieren von einer großen Datenmenge des Normalzustandes. Hierbei werden seit einigen Jahren die klassische statistischen Analysen [13] zunehmen durch Methoden des maschinellen Lernens ergänzt [10, 15, 16]. Dabei treten bei der Betrachtung von Zeitreihen, wie bei Daten mit einer natürlichen Reihenfolge (z. B. Bilder oder Text) im Allgemeinen, Effekte in Abhängigkeit der umgebenen Werte, dem Kontext, auf [1]. Häufig ist der Kontext leicht bestimmbar, da die betrachteten Systeme hinreichend klein sind. Auch industriellen Prozessdaten betreffend ist die Aussage bei Betrachtung einzelner, kleiner, abgeschlossener Subsysteme gültig [2, 8]. Seit einigen Jahren hat sich hier vor allem das LSTM (Long Short Term Memory) etabliert, das in der Lage ist, gezielt Informationen aus der Vergangenheit auch über größere Zeitverzüge zu behalten. Trotz der strukturellen Eignung des LSTM theoretisch beliebig lange Zeitverzüge zu berücksichtigen [4], führt dies nicht zwangsläufig zu einem entsprechenden Verhalten in der Anwendung. Vor allem große Totzeiten [19] und rauschbehaftete Daten erhöhen die Gefahr eines schlechten Konvergenzverhalten während des Trainings, da Zusammenhänge nur zwischen sehr kurzen Abschnitten sehr langer Sequenzen existieren. Die genannten großen Totzeiten treten unter anderem auf, wenn subprozessübergreifende Analysen in ausgedehnten Produktionsbetrieben durchge-

Diese Arbeit wird durch den Europäischen Fond für regionale Entwicklung unter der ZW 6-85018381 gefördert. Wir danken unseren industriellen Partner Heinrich Meier Eisengießerei GmbH & Co. KG, IAV GmbH und KÜNKEL WAGNER Germany GmbH für ihre Unterstützung.

führt werden. Für die Bestimmung des Zeitverzuges bietet sich die Messung mit z.B. RFID-Chips [3], die Ausnutzung deterministischer Zusammenhänge [7] oder die Schätzung basierend auf korrelierenden Signalen [11, 20] an.
Im Rahmen dieser Publikation werden zwei Ansätze zur Berücksichtigung von Zeitverzügen bzw. Totzeiten bei der datenbasierten Modellierung vorgestellt: die Generierung zeitkausaler Datensätze für das Training von Feed-forward Neural Network (FNN) und die Reduktion des Zeitverzugs vor dem Training von LSTM-Netzen. Der Vergleich mit dem Stand der Forschung, einem LSTM-Netz mit den unverschobenen Zeitreihen, erfolgt simulativ sowie an einem Altsandkühler einer Grünsandgussanlage.

2 Simulative Untersuchung

Zunächst soll der Einfluss eines stochastischen Zeitverzuges zwischen Ein- und Ausgangszeitreihe auf das Konvergenzverhalten während des Trainings datenbasierter Modelle simulativ untersucht werden. Hierbei wird eine numerische Eingangszeitreihe erzeugt, aus der sich ein Ausgangssignal berechnet. Für verschiedene Zeitverzugsverteilungen werden verglichen: FNN mit kausalem Datensatz, LSTM mit reduziertem Zeitverzug im Datensatz und LSTM mit dem ursprünglichen Datensatz.

2.1 Simulation der Zeitreihen

Um eine zufällige Eingangszeitreihe $\tilde{x}(k)$ mit einstellbarem Rauschverhalten, Periodizität und Glattheit zu simulieren, wird sie über einen Gauß-Prozess definiert. Die Eingangszeitreihe ist demzufolge eine Stichprobe des stochastischen Prozesses. Ein Gauß-Prozess $\mathcal{G}(m(k), \kappa(k,k'))$ ist definiert durch eine Mittelwertfunktion $m(k)$ und eine Kovarianzfunktion $\kappa(k,k')$, welche die Kovarianz zwischen zwei Zeitschritten k und k' vorgibt. Wir beschränken uns zur Vereinfachung auf einen mittelwertfreien

$$m(k) = 0 \qquad \forall k \in K \tag{1}$$

Gauß-Prozess mit radialsymmterischer Kovarianzfunktion

$$\kappa(k,k') = \kappa(r), \quad r = |k-k'|, \tag{2}$$

die diesen vollständig definiert. Die Kovarianzfunktion setzt sich aus einem rauen, periodischen Anteil

$$\kappa_\Pi(r) = \exp\left(-\frac{|\sin\left(\frac{\pi r}{T}\right)|}{l_c}\right) \tag{3}$$

mit Periodenlänge T und der charakteristischen Länge l_c, einem glatten Anteil

$$\kappa_S(r) = \exp\left(-\frac{r^2}{2l_c^2}\right) \tag{4}$$

und weißem Rauschen

$$\kappa_G(r) = \begin{cases} \varepsilon^2 & r = 0 \\ 0 & \text{else} \end{cases} \tag{5}$$

mit der Varianz ε^2 zusammen. Die Kovarianzfunktion ergibt sich somit zu

$$\kappa(r) = \kappa_\Pi(r)\,\kappa_S(r) + \kappa_G(r) \tag{6}$$

mit $T = 100$, $l_c = 10$ und $\varepsilon^2 = 0.03$. Bei der Generierung langer Zeitreihen steigt der Rechenaufwand aufgrund der Invertierung einer quadratisch anwachsenden Matrix, stark an [17]. Daher wird der Gaußprozess, nach [20], für ein kleines Fenster der Gesamtzeitreihe definiert. Dieses muss ein Vielfaches der Periodendauer lang und hinreichend groß sein, dass alle Einflüsse innerhalb der halben Fensterbreite abgeklungen sind. Die Abklingzeit ist allein von der charakteristischen Länge l_c abhängig. Nun kann der erste Teil der Zeitreihe generiert bzw. gezogen werden. Anschließend wird die erste Hälfte des Fensters mit den letzten Werten der simulierten Zeitreihe gefüllt und die zweite Hälfte über die bedingte Wahrscheinlichkeit (unter Kenntnis der ersten Hälfte) bestimmt und der simulierten Zeitreihe angehängt. Auf diese Weise können beliebig lange Zeitreihen simuliert werden.
Die Ausgangszeitreihe $\tilde{y}(k)$ wird aus der Eingangszeitreihe $\tilde{x}(k)$ und dem stochastischen Zeitverzug $\Lambda \sim \mathrm{N}(\mu_\Lambda, \sigma_\Lambda^2)$ mit der Wahrscheinlichkeitsdichtefunktion p_Λ durch die Berechnungsvorschrift

$$\tilde{y}(k) = \exp\left[\tanh\left(\frac{1}{Z}\sum_{i=\mu_\Lambda-3\sigma_\Lambda}^{\mu_\Lambda+3\sigma_\Lambda} p_\Lambda(i)\,\tilde{x}(k-i)\right)\right] \tag{7}$$

$$\tag{8}$$

mit der Normalisierungskonstante

$$Z = \sum_{i=\mu_\Lambda-3\sigma_\Lambda}^{\mu_\Lambda+3\sigma_\Lambda} p_\Lambda(i) \tag{9}$$

bestimmt. Die Berechnungsvorschrift beschreibt die diskrete Faltung des Eingangssignals mit einem Gauß'schen Fenster, beschränkt auf das 3σ-Intervall, und die nichtlineare Transformation durch die e- und tanh-Funktion. Um Messrauschen zu simulieren, wird weißes Rauschen $\Psi \sim \mathrm{N}(0, \varepsilon^2)$ mit der Varianz $\varepsilon^2 = 0.03$ zu dem Ein- und Ausgangssignal addiert

$$\tilde{x}_\mathrm{M}(k) = \tilde{x}(k) + \psi_\mathrm{x}(k) \quad \tilde{y}_\mathrm{M}(k) = \tilde{y}(k) + \psi_\mathrm{y}(k). \tag{10}$$

In **Bild 1** ist ein Ausschnitt der simulierten Zeitreihen beispielhaft zu sehen.

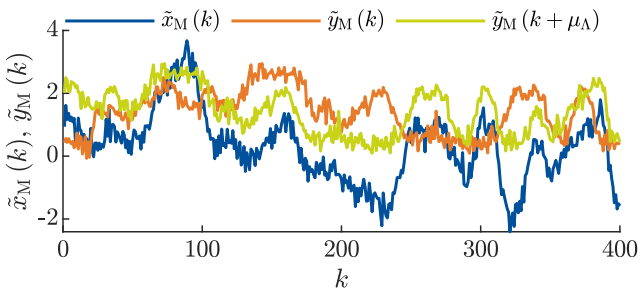


Abbildung 1 Simulierte Zeitreihen des Ein- und Ausgangsmesssignal $\tilde{x}_\mathrm{M}(k)$ und $\tilde{y}_\mathrm{M}(k)$, sowie das um den Mittelwert des Zeitverzugs verschobene Ausgangsmesssignal $\tilde{y}_\mathrm{M}(k+\mu_\Lambda)$. Die nichtlineare Transformation beschränkt den Wertebereich der Ausgangszeitreihe auf $\left(\frac{1}{\mathrm{e}}, \mathrm{e}\right)$.

	Architektur	Datensatz		Training	
		Beispiele	*Länge*	η	*Batchsize*
LSTM sim.	IN(1), LSTM(32), linear(1)	200	50000	0,001	160
FNN sim.	IN(1), ELU(6), ELU(6), linear1)	10×10^6		0,005	1×10^5
LSTM bin.	IN(1), LSTM(32), sigmoid(1)	7500	600	0,001	512
FNN bin.	IN(1), ELU(6), ELU(6), sigmoid(1)	4.5×10^6		0,001	512
LSTM kont.	IN(10), LSTM(32), linear(1)	32000	600	0,0005	128
FNN kont.	IN(10), PReLU(128), PReLU(128), linear(1)	11×10^6		0,001	512

Tabelle 1 Die Architektur, die Eigenschaften des Datensatzes, sowie die Trainingseinstellungen mit der Lernrate η für die im Rahmen dieser Veröffentlichung verwendeten Modelle. Die Architektur ist von links nach rechts zu lesen mit der Ausgangsdimension des jeweiligen Layers in Klammern. Die Aufteilung des Datensatzes in Trainings-, Validierungs- und Testdatensatz war 0,7 : 0,15 : 0,15, die Beispiele wurden vor jeder Epoche zufällig neu sortiert und die Optimierung erfolgte mit Adam [5].

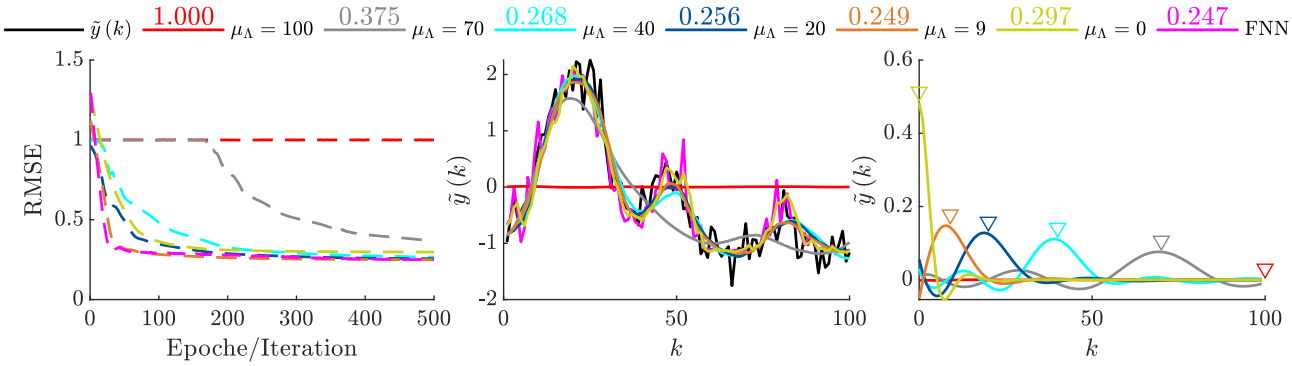


Abbildung 2 Root mean squared error (RMSE) während des Trainings (links), Güte der Modelle im Zeitbereich (mitte, Labels in schwarz) sowie die Antwort der LSTM-Modelle bei Impulsanregung der Höhe 1 zum Zeitpunkt $k = 0$ (rechts) für LSTM mit verschiedenen Zeitverzugsmittelwerten μ_Λ und dem FNN mit kausaler Datenvorverarbeitung. Der RMSE auf dem Testdatensatz ist in der Legende aufgeführt. Auf der x-Achse ist für das FNN die Iteration aufgetragen, 100 Iterationen entsprechen einer Epoche. Mit wachsendem Zeitverzug verschlechtert sich die Trainingskonvergenz und die Modellgüte. Im Verlauf ist zu sehen, dass die Ausschläge der abzubildenden Zeitreihe zunehmend unscharf abgebildet werden. Dies gilt auch für Zeitverzug bei Impulsanregung (rechts). Ist der Zeitverzug 0 verschlechtert sich aufgrund nicht-kausaler Einflüsse das Trainingsergebnis wieder.

2.2 Einfluss des Zeitverzugs

In diesem Kapitel wird das Trainingsergebnis von LSTM-Netzen für verschiedene Mittelwerte der Zeitverzugsverteilung $\mu_\Lambda \in \{0, 9, 20, 40, 70, 100\}$ und $\sigma_\Lambda = 3$ sowie von einem FNN mit kausalem Trainingsdatensatz verglichen. Ein Zeitverzugsmittelwert von 0 führt aufgrund der Faltung in (8) zu zeitlich nicht-kausalen Einflüssen. Zukünftige Werte von $\tilde{x}(k)$ haben einen Einfluss auf $\tilde{y}(k)$. Ein Zeitverzugsmittelwert von 9 führt zu dem Grenzfall, dass gerade alle Einflüsse zeitlich kausal sind, also der Zeitverzug so weit wie möglich reduziert ist, ohne die Kausalität zu verletzen. Die kausale Datenvorverarbeitung gewichtet jeden Wert zum Zeitpunkt k' mit der Wahrscheinlichkeit, dass der Einfluss kausal auf einen bestimmten Ausgangszeitpunkt k ist. Hierbei werden Zeitpunkte, bei denen der Einfluss den Zeitverzug zurückliegend auftrat, und Zeiträume, bei denen der Einfluss zwischen zwei Zeitverzugsverteilungen $p_{\text{start}}(k-k')$ und $p_{\text{ende}}(k-k')$ auftrat (z. B. Umgebungseinflüsse), unterschieden. Die Wahrscheinlichkeit, dass ein punktuelles Ereignis kausal auf einen Ausgangszeitpunkt ist, ergibt sich aus der Wahrscheinlichkeitsdichte der Zeitverzugsverteilung. Für über einen Zeitraum wirkende Größen kann für jeden Zeitpunkt k' die Wahrscheinlichkeit, dass der Zeitpunkt kausal auf den Zeitpunkt k ist

$$P(k' \mid k) = P_{\text{start}}(k-k')\left(1 - P_{\text{ende}}(k-k')\right) \tag{11}$$

basierend auf den Verteilungsfunktionen $P_{\text{start}}(k-k')$ und $P_{\text{ende}}(k-k')$ der begrenzenden Zeiten bestimmt werden. Die Formel bestimmt die Wahrscheinlichkeit, dass der betrachtete Zeitpunkt k' größer als der Start- und nicht größer als der Endzeitpunkt des bzgl. k kausalen Fensters ist. Der simulierte Datensatz ist um die Zeitverzugsverteilung verschoben, womit sich zur Generierung des kausalen Datensatzes die gleiche Faltung wie bei dem funktionalen Zusammenhang zwischen Ein- und Ausgangszeitreihe (8) ergibt. Nur wird das verrauschte Eingangssignal $\tilde{x}_{\text{M}}(k)$ anstatt $\tilde{x}(k)$ verwendet. Das FNN hat also den Vorteil der perfekten Kenntnis des zeitlichen Zusammenhanges und die Ergebnisse stellen den optimalen Fall dar.

Die Architektur des LSTM-Netzes und des FNN, der Umfang und die Aufteilung der Datensätze sowie die Trainingseinstellungen sind der **Tabelle 1** unter LSTM sim. und FFN sim. zu entnehmen. Wie das Trainingsergebnis in **Bild 2** zeigt, verbessert sich die Trainingskonvergenz der LSTM-Netze mit sinkendem Mittelwert des Zeitver-

zuges dahingehend, dass sowohl eine schnellere Konvergenz stattfindet als auch eine höhere Modellgüte realisiert werden kann. Auch die Berücksichtigung des Zeitverzuges erfolgt mit geringerer Unschärfe. Aufgrund der Berechnungsvorschrift, treten für einen Zeitverzugsmittelwert von $\mu_\Lambda = 0$ nicht-kausale Effekte auf, welche die Genauigkeit des Modells stark negativ beeinflussen. Es sollte daher angestrebt werden, den Zeitverzug in realen Systemen nur bis zu den ersten relevanten Einflüssen zu reduzieren und nicht bspw. den Mittelwert zu wählen.

3 Der Altsandkühler

Die Aussagen des vorherigen Kapitels werden ebenfalls anhand eines realen Systems, einem Altsandkühler, überprüft. Der Altsandkühler, siehe **Bild 3**, ist die erste Station in der Sandaufbereitung eines Grünsandgussprozesses, dessen verlorenen Formen aus bentonitgebundenem Formstoff bestehen. Zur Kühlung des zurückgeführten Formstoffes (Altsand) sowie zur Reaktivierung der unverbrannten (aktivierbaren) Bentonitplättchen wird einem kontinuierlichen Altsandstrom Wasser und Umgebungsluft zugeführt. Die Zugabe des Sandes wird durch das schaltbare Eingangsförderband gesteuert und eine Sandsonde am Ausgang detektiert, ob Sand vorhanden ist. Sowohl am Ein- als auch am Ausgang sind Temperatur- und Feuchtesonden für Schüttgüter angebracht. Die Durchlaufzeit des Sandes folgt aufgrund von Umverteilungen während der Durchmischung einer nicht-parametrischen Verteilung, die zusätzlich abhängig vom Füllgrad des Altsandkühlers ist. Zur Vereinfachung wird eine konstante Durchlaufzeitverteilung angenommen, die dem Zeitverzug zwischen den gemessenen Größen des Ein- und Ausgangsförderband entspricht. Neben der signalbasierten Schätzung des Zeitverzuges durch die Kreuzkorrelationsfunktion sowie die Transferentropie, sind Validierungsmessungen mit fluoreszierenden Farbstoffen durchgeführt worden [20]. Basierend auf den Mess- und den Schätzergebnissen wird eine Sanddurchlaufverteilung $\Lambda \sim \mathrm{N}\left(90\,\mathrm{s}, 100\,\mathrm{s}^2\right)$ angenommen.

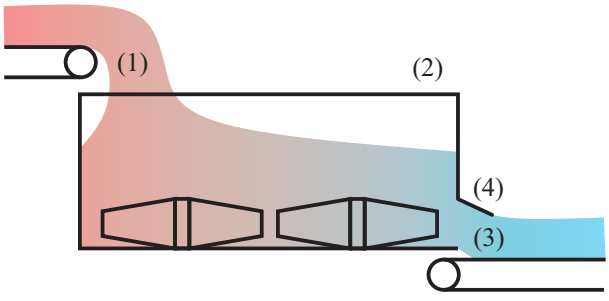


Abbildung 3 Vereinfachte Darstellung des Altsandkühlers mit Eingangsförderband (1), Altsandkühler (2) mit zwei Rotoren zur Sandbeförderung und Ausgangsförderband (3) mit Auslassklappe (4).

4 Modellierung des Altsandkühlers

In diesem Kapitel wird der Einfluss der Zeitverzugsberücksichtigung auf das Modellierungsergebnis anhand des Altsandkühlers bewertet. Hierbei werden ebenfalls LSTM mit und ohne Zeitverzugsberücksichtigung sowie ein FNN mit kausaler Datenvorverarbeitung verglichen. Zunächst wird ein vereinfachtes Subproblem mit einer Ein- und einer Ausgangszeitreihe untersucht: die Vorhersage von Sand auf dem Ausgangsförderband bei gegebener Aktivität des Eingangsförderbandes während der Anlaufphase des Altsandkühlers. Anschließend wird die Ausgangstemperatur des Sandes basierend auf der Eingangstemperatur und -feuchte des Sandes, der Temperatur und Feuchte der ein- und ausströmenden Luft, der Wasserzugabe, der Position der Auslassklappe, dem Motorstrom und der Aktivität des zuführenden Förderbandes bestimmt. Die Signale wurden mit einer Auflösung von 500 ms erfasst. Da die Taktzeit der Maschine 20 ms beträgt, wurden die binären Signale durch die Wahl des nächsten Nachbarn und kontinuierliche Signale durch lineare Interpolation mit additivem Messrauschen gestreckt. Die Validierung erfolgte anhand beider Datensätze. Das Training wurde auf einer AMD Radeon RX580 8GB mit TensorFlow-ROCm durchgeführt.

4.1 Sanddetektion am Ausgang

Der Datensatz mit Startvorgängen des Altsandkühlers wird in 7500 Sequenzen mit jeweils 600 Zeitschritten für den gemessenen Datensatz bzw. 14976 Zeitschritten für den gestreckten Datensatz aufgeteilt. Die Architektur der Netze, die Aufteilung des Datensatzes und die Einstellungen zur Optimierung sind in **Tabelle 1** unter LSTM bin. Und FNN bin. gelistet. Da am Ende das Modell mit dem besten Validierungsergebnis ausgewählt wurde, beziehen sich die gezeigten Ergebnisse auf den ungesehenen Testdatensatz. In **Bild 4** ist das Ergebnis auf dem Testdatensatz für LSTM mit variierender Zeitverschiebung und das FNN mit kausaler Datenvorverarbeitung gezeigt. Wie erwartet, verbessert sich das Trainingsergebnis mit der Verringerung der Zeitdifferenz zwischen Ein- und Ausgangszeitreihe. Sobald aufgrund der Verschiebung kausale Daten der Eingangszeitreihe nicht mehr Berücksichtigung finden können, verschlechtert sich die Modellgüte wieder. Werden die auf 20 ms Abtastzeit gestreckten Zeitreihen nicht verschoben, findet aufgrund der Zunahme des Zeitverzugs (gemessen in Zeitschritten) keine Konvergenz während des Trainings mehr statt. Bei einer Verschiebung der Eingangszeitreihe um $\Delta t = 60\,\mathrm{s}$ nach hinten kann das Modell wieder erfolgreich trainiert werden. Unerwartet ist, dass sich keine Verlagerung des Optimums hin zu größeren zeitlichen Verschiebungen Δt für die auf 20 ms Abtastzeit gestreckten Zeitreihen ergibt. Aufgrund der Zunahme des Zeitverzuges hatten wir hier einen größeren Benefit von stärkeren Verschiebungen erwartet. Zur Vorhersage des ersten Sandaufkommens am Ausgang des Altsandkühlers während des Startvorganges zeigt das FNN basierend auf der binären Kreuzentropie das beste Ergebnis. Im Gegensatz zu den

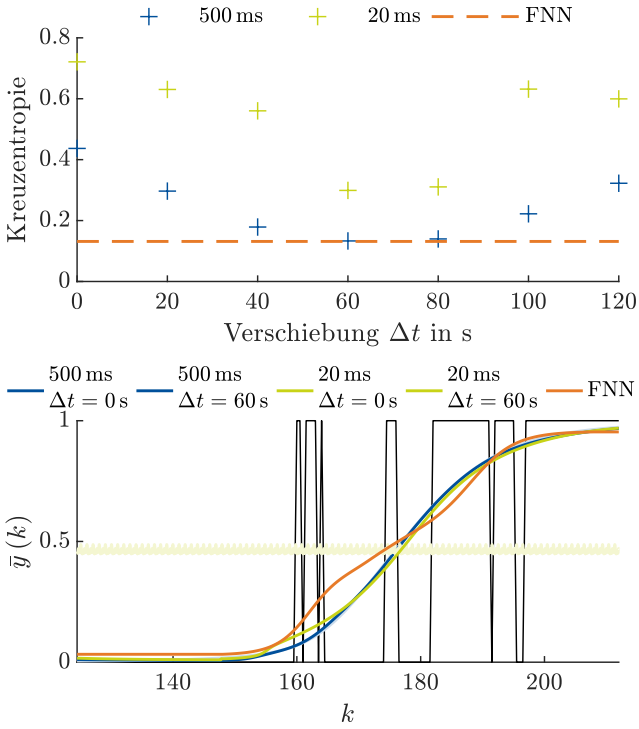


Abbildung 4 Validierungsergebnis auf dem Testdatensatz (oben) und der Vergleich zwischen dem Messsignal und den Modellen für ausgewählte Zeitverzüge $\Delta t \in \{0, 60\}$ s (unten) und einer Signalabtastung von 500 ms und 20 ms. Das FNN zeigt das beste Ergebnis. Insbesondere wird der häufigen Sanddetektion am Start und am Ende sowie der selteneren Detektion im Mittelteil des Überganges Rechnung getragen.

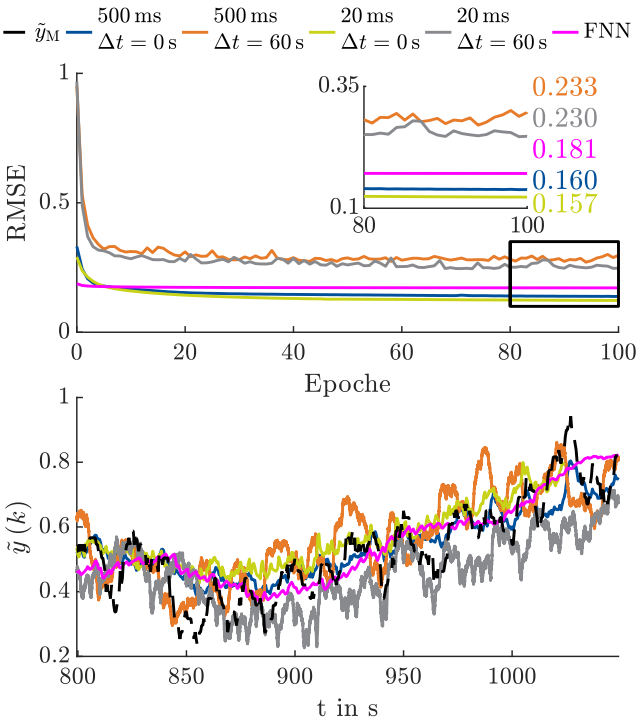


Abbildung 5 Konvergenz beim Training mit farblich kodiertem Validierungsergebnis auf dem Testdatensatz (oben) und der Vergleich zwischen dem Messsignal und den Modellen (unten) für die Zeitverzüge $\Delta t \in \{0, 60\}$ s und einer Signalabtastung von 500 ms und 20 ms. Als Relation sei hier genannt, dass alle Validierungsergebnisse auf dem Testdatensatz einen RMSE von weniger als 1 K aufweisen. Die Rekursions- und Eingangsgewichte der LSTM mussten bei den Sequenzen mit einer Abtastungen von 20 ms mit dem Faktor 0,01 L1-regularisiert werden, um ein Überlaufen der Variablen beim Training zu verhindern. Daher ist der RMSE während des Trainings höher als auf dem Testdatensatz.

LSTM gestaltet sich der Übergang weniger glatt, sodass die unterschiedlichen Bereiche abgebildet werden können.

4.2 Sandtemperatur am Ausgang

Die Architektur der Netze, die Aufteilung des Datensatzes und die Einstellungen zur Optimierung sind in **Tabelle 1** unter LSTM kont. und FNN kont. gelistet. Alle Daten sind z-standardisiert. Aufgrund der Ergebnisse aus dem vorigen Kapitel werden die Eingangszeitreihen für das LSTM mit Zeitverzugsberücksichtigung um 50 s verschoben. Der kausale Datensatz für das FNN umfasst hauptsächlich Einflüsse, welche währen der gesamten Durchlaufzeit des Sandes auf diesen einwirken und werden daher nach (11) gewichtet. Lediglich die Eingangstemperatur und -feuchte werden als punktuelle Einflüsse angenommen und entsprechend behandelt. Die Konvergenz während des Trainings und ein Ausschnitt der Sequenzen sind in **Bild 5** dargestellt. Dabei wird die Modellgüte nur in solchen Datenpunkten validiert, bei denen die Sandsonde am Ausgang auch Sand detektiert. Bei den LSTM zeigt sich eine schnellere Trainingskonvergenz sowie ein besseres Validierungsergebnis, wenn der Zeitverzug der Zeitreihen durch Verschiebung reduziert wird. Die Unterschiede sind jedoch weniger stark ausgeprägt, wie bei den binären Signalen, was vermutlich auf eine starke Autokorrelation der Signale sowie verstärkter Einflüsse ohne Zeitverzug (z.B. die aktuellen Umgebungsbedingungen) zurückzuführen ist. Das FNN zeigt gegenüber den LSTM mit 500 ms Abtastung schlechtere und gegenüber den LSTM mit 20 ms Abtastung bessere Validierungsergebnisse. Wie schon in der simulativen Untersuchung tritt die Konvergenz jedoch deutlich schneller auf. Wie in den Verläufen zu sehen ist, weisen alle LSTM-Modelle Schwingungen auf, die in der Größenordnung der Schwingung des Messsignals liegen. Das FNN kann diesen Schwingungen aus zwei Gründen nicht folgen. Erstens ist das Fenster so groß, dass Schwingungsphänomene in der vorliegenden Frequenz von ca. 0,05 Hz nicht abgebildet werden können. Außerdem sind Schwingungsphänomene prinzipbedingt nicht abbildbar, wenn sie nicht in den Eingangsdaten auftreten und z. B. der Funktionsweise der Maschine entspringen. Das LSTM hingegen ist in der Lage auch in diesen Fällen die Form der Zeitreihe durch die Berücksichtigung der Reihenfolge zu erlernen. Dieser Vorteil des LSTM gerät zum Nachteil, wenn eine Ursachenforschung anhand des Modells durchgeführt werden soll. Bei den LSTM kann das Ergebnis nur anhand der gesamten Zeitreihen zurückgeführt werden, was Aussagen über den Einfluss einzelner Eingänge erschwert.

5 Zusammenfassung

Um die auftretenden stochastischen Zeitverzüge zwischen den Daten in industriellen Produktionsbetrieben bei der datenbasierten Modellierung zu berücksichtigen sind zwei Verfahren vorgestellt worden. Neben der Verschiebung der Zeitreihen und anschließendes Training von LSTM-Netzen wurde die Generierung eines zeitkausalen Datensatzes für das Training von FNN vorgestellt, bei dem jeder Datenpunkt mit der Wahrscheinlichkeit gewichtet wird, dass er kausal auf einen bestimmten Ausgangszeitpunkt ist. Der Vergleich mit einem LSTM ohne Verschiebung der Zeitreihen zeigte sowohl für simulierte Zeitreihen als auch bei der Modellierung des Altsandkühlers eine durchweg verbesserte Modellgüte, wenn der Zeitverzug berücksichtig wurde. Aufgrund der Kausalität der LSTM darf eine Verschiebung jedoch nicht über die ersten relevanten Einflüsse hinausgehen. Das FNN zeigte eine vergleichbare Güte wie das LSTM, kann aber keine Informationen aus sich wiederholenden Reihenfolgen des Ausgangssignals (z. B. Schwingungen oder gleiche Zyklen) berücksichtigen. Dies wiederum ist bei der Rückführung des Modellausganges vorteilhaft. Es können die einzelnen Einflüsse separat betrachtet werden. Weitere Untersuchungen sind notwendig, um die Berücksichtigung des Zeitverzuges beim subprozessübergreifenden Monitoring zu ermöglichen und zu bewerten. Neben den dort auftretenden Zeitverzügen, die um ein Vielfaches größer sind, weisen die Signale in der Regel auch stark unterschiedliche zeitliche Auflösungen auf.

6 Literatur

[1] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection. *ACM Computing Surveys*, 41(3):1–58, 2009.

[2] S. E. Guttormsson, R. J. Marks, M. A. El-Sharkawi, and I. Kerszenbaum. Elliptical novelty grouping for on-line short-turn detection of excited running rotors. *IEEE Transactions on Energy Conversion*, 14(1):16–22, 1999.

[3] B. Hameed, I. Khan, F. Durr, and K. Rothermel. An rfid based consistency management framework for production monitoring in a smart real-time factory. In *2010 Internet of Things (IOT)*, pages 1–8. IEEE, 29.11.2010 - 01.12.2010.

[4] S. Hochreiter and J. Schmidhuber. Lstm can solve hard long time lag problems. In *Advances in neural information processing systems*, pages 473–479, 1997.

[5] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[6] J. Lee, E. Lapira, B. Bagheri, and H.-a. Kao. Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, 1(1):38–41, 2013.

[7] D. Lieber, M. Stolpe, B. Konrad, J. Deuse, and K. Morik. Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning. *Procedia Cirp*, 7:193–198, 2013.

[8] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. *arXiv preprint arXiv:1607.00148*, 2016.

[9] I. Maurer, M. Riva, C. Hansen, and T. Ortmaier. Cloud-based plant and process monitoring based on a modular and scalable data analytics infrastructure. In T. Schüppstuhl, J. Franke, and K. Tracht, editors, *Tagungsband des 2. Kongresses Montage Handhabung Industrieroboter*, pages 33–42, Berlin, Heidelberg, 2017. Springer Berlin Heidelberg.

[10] P. O'Donovan, K. Leahy, K. Bruton, and D. T. J. O'Sullivan. An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *Journal of Big Data*, 2(1):25, 2015.

[11] J. D. Olden and B. D. Neff. Cross-correlation bias in lag analysis of aquatic time series. *Marine Biology*, 138(5):1063–1070, 2001.

[12] Q. Qi and F. Tao. Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *IEEE Access*, 6:3585–3593, 2018.

[13] S. J. Qin. Process data analytics in the era of big data. *AIChE Journal*, 60(9):3092–3100, 2014.

[14] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen. About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine*, 48(3):567–572, 2015.

[15] F. Tao, Q. Qi, A. Liu, and A. Kusiak. Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48:157–169, 2018.

[16] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben. Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1):23–45, 2016.

[17] W. Yan, H. Qiu, and Y. Xue. Gaussian process for long-term time-series forecasting. In *2009 International Joint Conference on Neural Networks*, pages 3420–3427, 2009.

[18] D. Zhou, G. Li, and S. J. Qin. Total projection to latent structures for process monitoring. *AIChE Journal*, pages NA–NA, 2009.

[19] J. Zumsande, K.-P. Kortmann, M. Wielitzka, and T. Ortmaier. Probabilistic simulation and determination of sojourn time distribution in manufacturing processes. In *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 2094–2099, 2019.

[20] J. Zumsande, K.-P. Kortmann, M. Wielitzka, and T. Ortmaier. Sojourn time identification in bulk material processing based on cross-correlated signals. In *Proceedings of the IEEE International Conference on Industrial Technology*. 2021.

Maximierung der Anlageneffizienz in der Automobilproduktion mit Hilfe web-basierter IIoT-Applikationen
Improvement of daily production efficiency using web-based IIoT applications

Dr. oec. Linda Quynh Nhu Truong, Dr. rer. nat. Thomas Tentrup
Dürr Assembly Products GmbH, Köllner Str. 122-128, 66346 Püttlingen, Deutschland; linda.truong@durr.com

Kurzfassung

Es gibt kaum Lebensbereiche, auf die die Digitalisierung keine Auswirkungen hat. Auch die Automobilindustrie befindet sich seit Jahren in einer digitalen Transformation. Die Vernetzung der Fabrikhallen sowie die stärkere Nutzung von Produktionsdaten sollen die Automobilproduktion noch effizienter machen. In diesem Beitrag beschreiben wir den aktuellen Vernetzungsgrad des Prüfbereichs in der Automobilproduktion und den bisherigen Umgang mit Produktionsdaten. Gerade bei der heute noch üblichen manuellen Vorverarbeitung der Daten für eine spätere Interpretation und Berichterstellung gibt es Optimierungspotenziale. Als Konsequenz haben wir – einer der führenden Anlagenhersteller auf dem Gebiet der automobilen Prüftechnik – eigens Software-Lösungen auf Basis von Industrial Internet of Things (IIoT) entwickelt, die dieses Problem adressieren und gleichzeitig (i) Transparenz und Komfort bei der Anlagenbedienung erhöhen, (ii) Prüfabläufe effizienter machen und (iii) Betreiberkosten deutlich reduzieren.

Abstract

There is hardly any domain that has not been touched by digitalization. The automotive sector is no exception. Higher connectivity of sites and an increasing use of data are two key drivers for a more efficient automobile production. Against this backdrop, we describe the current state of connectivity and data use for the final assembly testing area in an automotive production setting and identify manual data pre-processing steps prior to data interpretation and reporting as a major area for improvement. Consequently, we – as a leading company for automotive testing systems – have developed software solutions based on the Industrial Internet of Things (IIoT) to address these issues, while (i) increasing transparency and comfort during machine operation, (ii) making testing processes more efficient, and (iii) significantly reducing overall operating costs.

1 Einleitung

Produktionsstillstände in der Automobilproduktion sind nicht nur ärgerlich, sondern kosten Hersteller bares Geld. Ein zehnminütiger Bandstopp kann bereits Kosten in niedriger fünfstelliger Summe auslösen [1, 2]. Aber auch fehlerhaft produzierte Fahrzeuge können neben Effizienzverlusten in der Produktion einen nicht unbedeutenden Reputationsschaden verursachen, wenn das fehlerhafte Auto an Endkunden ausgeliefert wird. Dies ist insbesondere der Fall, wenn es sich um fahr- oder sicherheitsrelevante Aspekte handelt. Laut einer Studie werden weltweit pro Minute 16,5 Produktionsfehler am Fahrzeug durch Endkunden beanstandet, die noch in den Gewährleistungszeitraum fallen. Die nachträgliche Behebung solcher Beschwerden allein verursacht 3.600 Euro pro Minute – dies entspricht pro Jahr weltweiten Kosten in Höhe von 1,9 Milliarden Euro für die Automobilhersteller [3].

Soweit muss es allerdings nicht kommen. Durch die steigende Vernetzung von Mensch, Maschine, Prozessen und Materialien werden Produktionsverantwortliche heute viel schneller auf reduzierte Anlagenperformance, geringere Anlagenverfügbarkeit sowie mangelnde Fahrzeugqualität während der einzelnen Prozessschritte aufmerksam gemacht, bevor Fahrzeuge die Werkshallen überhaupt verlassen. Seit die Automobilindustrie Milliarden in die Digitalisierung ihrer Fertigungsstraßen investiert [4], hat sich auch die Nachverfolgbarkeit von Produktionsvorgängen erheblich verbessert. Um diese Transparenz zu gewährleisten, werden unter anderem auch mehr Produktionsdaten gespeichert. Doch diese exponentiell wachsende Menge an Datenpunkten bloß aufzuzeichnen und abzulegen bringt im Falle eines Produktionsproblems noch keinen Mehrwert. Vielmehr müssen Daten in Echtzeit verfügbar gemacht und laufend mit entsprechendem Prozesswissen interpretiert werden, um einen Nutzen aus den Daten zu generieren.

In diesem Beitrag zeigen wir, wie wir als Anlagenlieferant unsere Kunden bei diesem Schritt unterstützen.

2 Stand der Technik

2.1 Der Prüfbereich eines Automobilwerks

Eine Automobilfertigung besteht im Wesentlichen aus drei Abschnitten: dem Rohbau, der Lackiererei und der Endmontage. Der letzte Abschnitt der Endmontage (auch bekannt als Bandende) dient auch der Qualitätskontrolle des gesamten Fahrzeugzusammenbaus. In diesem vor der Auslieferung stattfindenden Schritt werden fertig montierte Fahrzeuge auf Fahrwerkseinstellständen und Rollenprüfständen bezüglich ihrer Fahrtüchtigkeit geprüft; weiterhin werden auch Scheinwerfer und Fahrassistenzsysteme eingestellt [5]. Die Dürr Assembly Products GmbH ist seit mehr als 25 Jahren einer der weltweit führenden Hersteller in diesem Bereich.

Während des Prüfvorgangs eines Fahrzeugs an der Anlage werden eine Vielzahl von prüfungsrelevanten Daten generiert, die von der Automatisierungssoftware am jeweiligen Anlagen-PC heute schon erzeugt und gespeichert werden (**Abb. 1**). Ein Großteil der Prüfsysteme aus dem Hause Dürr werden mit der eigens für Prüfabläufe entwickelten Automatisierungssoftware x-line ausgestattet.[1] Doch nur ein kleiner Teil dieser generierten Daten wird nach Beendigung eines Prüfablaufs über das Werksnetz an Produktionsleitsysteme gesendet. Um detaillierte Informationen zu den geprüften Fahrzeugen oder der Anlage zu erhalten, müssen Interessenten die Daten aus der Anlage zunächst extrahieren, bevor diese überhaupt ausgewertet werden können.

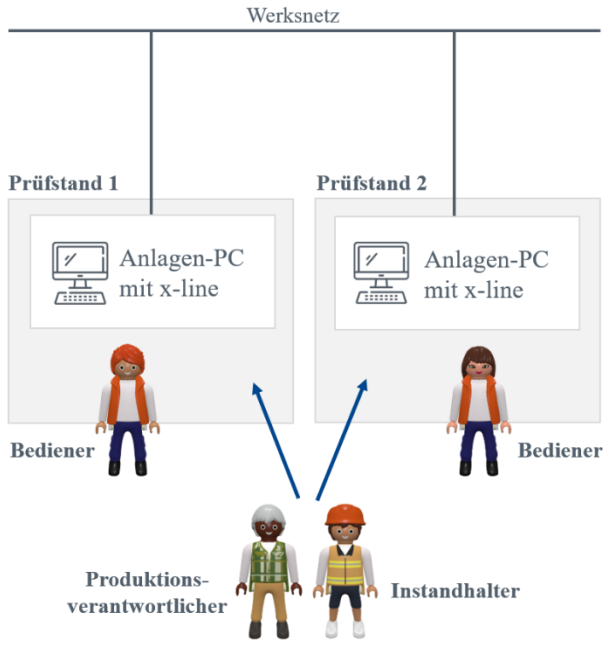


Abb. 1 Darstellung eines Bandendes ohne IIoT-Anbindung

[1] Alternativ ist auch eine SPS-Steuerung möglich.

Folgendes Beispiel: Als Verantwortlicher für den Bandende-Bereich einer Endmontagehalle stellen Sie fest, dass einer Ihrer Fahrwerkstände eine um zehn Prozent langsamere Taktzeit aufweist als die anderen Fahrwerkstände im selben Zeitraum. Hierdurch wird der Durchsatz reduziert, was einen Verlust in der Höhe der nicht produzierten Fahrzeuge erzeugt. Sie müssen also so schnell wie möglich die Ursache für die Planabweichung finden und beheben.

Doch der Problemanalyseprozess ist aufwendig und zeitraubend. Genaue Details zu einzelnen Prüf- und Einstellschritten, die Taktzeitverluste ausgelöst haben könnten, sind für Sie als Produktionsverantwortlicher auf den ersten Blick nicht ersichtlich – ähnlich wie bei einem Eisberg, wo auch zunächst nur die Spitze zu sehen ist (**Abb. 2**). Die vollständige Menge der Produktionsdaten ist nicht erkennbar. Bei einem Eisberg ist dies der Teil, der sich unter der Wasseroberfläche befindet.

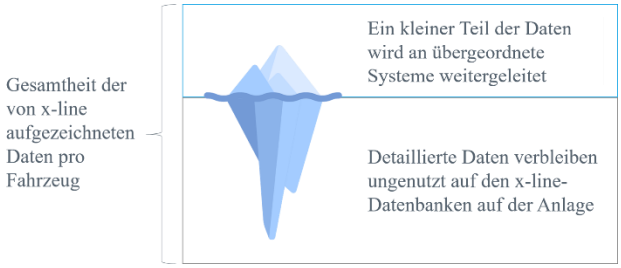


Abb. 2: Die Eisberg-Analogie – sichtbare vs. detaillierte Daten

Um nun für die Problemanalyse an die vollständigen Datensets eines Prüfstands zu gelangen, müssen diese manuell aus den jeweiligen Anlagen-PC extrahiert werden, sofern keine intelligente Vernetzungslösung integriert ist. In der Regel müssen die relevanten Dateien auf dem Rechner zunächst lokalisiert und dann auf einen USB-Stick zur Weiterverarbeitung kopiert werden. Bei Dürr-Anlagen können die relevanten Datensätze heute schon mit wenigen Klicks zusammengetragen und als .zip-Datei extrahiert werden. Doch werden zur Fehleranalyse oft mehrere Dateien unterschiedlicher Anlagen benötigt, um beispielsweise einen Vergleich unter Anlagen des gleichen Typs zu machen. Diese Dateien, die in unterschiedlichen Formaten vorliegen können, müssen von Hand und mit dem entsprechenden Know-how in das gleiche Datenformat konvertiert, zusammengeführt, von Redundanzen befreit und methodisch so aufbereitet werden, dass eine Interpretation möglich ist. Erst dann können die Erkenntnisse in einem Bericht zusammengefasst werden. In Summe stellt dieser Prozess einen enormen Zeitaufwand dar (**Abb. 3**). Mit steigender Menge an verfügbaren Daten oder der Frequenz, in

der berichtet werden muss, kann sich dieser Aufwand schnell potenzieren.

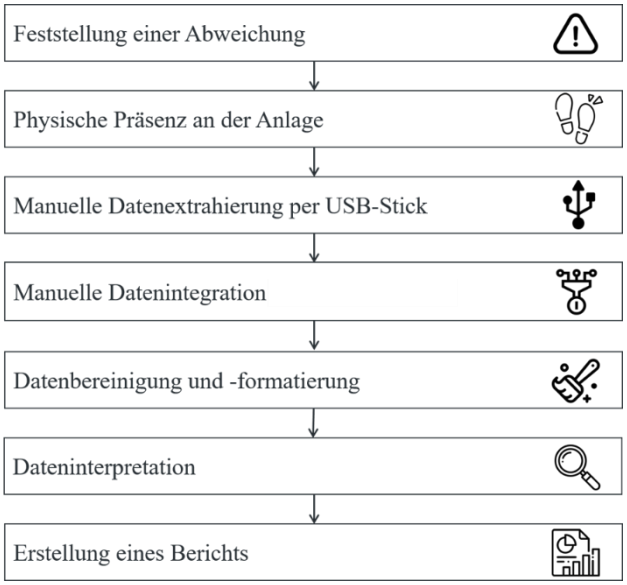


Abb. 3: Beispielhafter Ablauf einer Berichterstellung

2.2 Defizite des Analyseverfahrens im Prüfbereich

Zunächst werden auf die beschriebene Weise nur historische Daten ausgewertet. Eine Überwachung in Echtzeit – und damit Grundlage für eine zukünftige Voraussage von Vorgängen – ist noch nicht möglich. Zweitens sind Ingenieure noch gezwungen, sich für eine Datenextraktion physisch in die Montagehalle zu begeben. Ein direkter Remote-Zugang auf die Anlagen oder den Produktionsbereich ist meist nicht etabliert. Nicht zuletzt sind auch wesentliche Informationen, die für Betreiber, Instandhaltung oder Qualitätsabteilung notwendig sind (wie z. B. Wartungspläne, Kalibrier- oder Reparaturanleitungen), nicht auf Anhieb oder an zentraler Stelle verfügbar. Gerade bei zuverlässigen Anlagen, die nur sehr selten ausfallen, oder wenn die Arbeit an den Anlagen auf jahrelangem Wissen einzelner erfahrener Mitarbeiter statt auf schriftlichen Dokumentationen beruht, kann dies im Fehlerfall zu Verzögerungen bei der Problemlösung führen. Treten Anlagenstörungen auf, die der Automobilhersteller nicht selbst beheben kann, wird oft auch der Anlagenlieferant gerufen, um gemeinsam einen Blick auf die gesammelten Produktionsdaten zu werfen. Dazu werden meist Log-Dateien oder vollständige Datenbanken mit dem Anlagenlieferanten geteilt. Allerdings kommt es in der Praxis auch vor, dass Kunden gar nicht wissen, wo sie die Dateien mit den notwendigen Informationen zur Problemlösung finden. Häufig geht deswegen Zeit verloren, weil anfangs falsche oder unvollständige Datensätze miteinander ausgetauscht werden. Währenddessen steht die betroffene Anlage still und produziert nicht.

3 IIoT als Lösungsansatz

Bei den genannten Schwachpunkten kann das Internet der Dinge oder Internet of Things (IoT) Abhilfe schaffen. Die Grundidee ist einfach: Gegenstände (Things) werden mit dem Internet verbunden und so kann man über das Netzwerk mit den Gegenständen kommunizieren. Beim Industrial Internet of Things (IIoT) werden die Prinzipien des IoT vom Konsumentenbereich auf den Produktionssektor übertragen. Produktionsdaten sollen über vernetzte Sensorik, Komponenten, Anlagen, Prozesse, Werkshallen und Standorte hinweg gesammelt und ausgewertet werden können, um die Produktion optimieren zu können. Der Wunsch nach einer Anlage, die möglichst wartungs- sowie störungsarm läuft und sich bereits vor Eintritt eines Problems automatisch beim Betreiber meldet, ist der Traum eines jeden Produktionsplaners oder Anlagenbetreibers.

Gerade während der globalen COVID-19-Pandemie gewinnen solche Technologie an Bedeutung. Geschäftsreisen sind auf ein Minimum zurückgefahren worden und Firmen müssen ihre Produktionsprozesse aufgrund von geringerer Personalauslastung (durch Erkrankungen, Familienbetreuung, Kurzarbeit oder als Beitrag zum Infektionsschutz) noch effizienter gestalten. Nicht nur die Automobilfirmen, sondern alle produzierenden Betriebe sind daher verstärkt auf digitalisierte, ortsunabhängig durchführbare Aktivitäten und Vorgänge angewiesen. Durch den vermehrten Einsatz von IIoT kann nicht nur eine stabile Produktion gewährleistet werden, sie wird sogar gesteigert.

4 Praktische Umsetzung

Als Antwort auf diese technologische Transformation begann der Dürr-Konzern 2016 mit der Umsetzung von „digital@DURR“. Diese Strategie umfasst die Entwicklung von intelligenten Anlagen, Services und Prozessen, die in einer smarten Fabrik eingesetzt werden sollen. Gleichzeitig wurden interne Prozesse digitalisiert und verstärkt agile Arbeitsmethoden in den Entwicklungsabteilungen genutzt. Ein eigener Bereich wurde ins Leben gerufen, um digitale Lösungen zu entwickeln (vgl. auch „Software-as-a-Service“, SaaS [6]). Auch für den Bereich der Prüftechnik wurden so zügig marktreife, web-basierte IIoT-Applikationen entwickelt.

4.1. Digitale Lösungen für die Prüftechnik

Zum aktuellen Zeitpunkt sind drei Applikationen für den automobilen Prüfbereich verfügbar. Es handelt sich hierbei um standardisierte Softwareprodukte, die pro Werk installiert und passend auf die Produktionsabläufe vor Ort konfiguriert werden. Alle Applikationen können sowohl einzeln als auch im Verbund verwendet werden.

Die Applikationen **DXQ**equipment.monitoring und **DXQ**equipment.maintenance konzentrieren sich auf die kontinuierliche, zeit- und ortsunabhängige Zustandsüberwachung von Prüfanlagen. Hiermit kann von jedem zugriffsberechtigten PC, Laptop, Tablet oder Handy ein digitaler Zwilling der Anlage aufgerufen werden, um zu prüfen, in welchem Betriebsmodus sich die Anlage gerade befindet, welche Meldungen bisher geloggt wurden oder ob Wartungsaufgaben fällig sind. Sollte die Anlage aufgrund eines Fehlers in Störung gehen, wird das Instandhaltungsteam sofort per SMS oder E-Mail benachrichtigt. So lassen sich mögliche Reparaturvorbereitungen (z. B. Werkzeug- oder Ersatzteilbeschaffung) in Gang setzen, bevor lange Laufwege in einer Montagehalle zurückgelegt werden müssen und deswegen wertvolle Zeit verloren geht. Das Zusatzmodul „Troubleshooter" trägt im Falle eines Anlagenstillstands mögliche Fehlerursachen zusammen und unterbreitet Hilfestellungen zur Problembeseitigung als Text, Bild oder in Form eines Video-Tutorials. So kann die Anlage schneller wieder in Betrieb genommen werden. Kann das Problem nicht initial auf Kundenseite gelöst werden, besteht die Möglichkeit einen Servicemitarbeiter der Firma Dürr zu Hilfe rufen, der sich ohne Zeitverlust durch Reisen sofort per Fernzugriff auf die Anlage schalten kann. Relevante Produktionsdaten wurden durch die IIoT-Applikationen schon vorverarbeitet und beschleunigen den Reparaturvorgang.

Des Weiteren wird bei **DXQ**equipment.maintenance eine intelligente Liste aller anstehenden Wartungsbedarfe des Anlagenparks nach Fälligkeitsdatum geführt. Erledigte Aufgaben können abgehakt und automatisch dokumentiert werden, sodass Anlagenbetreiber im Falle eines Zertifizierungsaudits bestens vorbereitet sind. Bei der Wartungsdurchführung können hier auch konkrete Anleitungen als Text, Bild oder Video angezeigt werden, um eine einwandfreie Ausführung zu garantieren.

Während die zwei erstgenannten Lösungen die Erhaltung einer hohen Anlagenverfügbarkeit unterstützen, wird bei **DXQ**product.monitoring das Werkstück (sprich: das produzierte Fahrzeug) und seine Qualität in den Mittelpunkt gerückt. In dieser Applikation kann anhand der eindeutigen Fahrzeugidentifikationsnummern in Echtzeit nachvollzogen werden, welches Fahrzeug in welcher Montagelinie, von welchem Prüfstand und mit welchen konkreten Prüf- und Messergebnissen sowie Taktzeit getestet worden ist. Sollte hier ein Fahrzeug einen Prüfstand mit einem „n.i.O." (nicht in Ordnung) oder überzogener Prüfzeit verlassen, kann über das komfortable Tool sofort nachvollzogen werden, welche konkreten Einzelwerte zu diesem Ergebnis geführt haben, ohne wie bisher manuell Daten auslesen und bereinigen zu müssen. Zahlreiche Filter- und Analysefunktionen unterstützen die Qualitätsingenieure bei der Ursachenforschung und zeigen auf, ob es systematische Unregelmäßigkeiten gibt.

4.2. Aufbau und Funktionsweise

Obwohl die Anlagen der Firma Dürr Assembly Products eine Reihe von Daten bereits über x-line automatisch generieren und speichern, ist es sinnvoll, Daten und Datenbanken über Anlagen hinweg miteinander zu verknüpfen, um sinnvolle Schlüsse daraus ziehen zu können. Diese Daten müssten idealerweise so automatisiert aufbereitet werden, dass in kurzer Zeit und mit wenigen Klicks die notwendigen Informationen bereitgestellt werden können.

Grundausstattung für eine solche IIoT-Anwendung (**Abb. 4**) ist meist die Installation eines IIoT-Edge-Servers, an den die Daten aus den Anlagen geschickt werden. Es handelt sich hier um einen zusätzlichen Industrie-PC, der in die Nähe der zu vernetzenden Anlagen aufgebaut (daher der englische Begriff *edge*, also am Rand von etwas) und mit den Anlagen-PCs physisch per Netzwerkkabel verbunden wird. Auf dem Edge-Rechner wird ein sogenannter „Agent" – ein eigens für die IIoT-Anwendung mit den Dürr-Anlagen konzipierter Software-Service – aufgespielt, der die Daten aus den Anlagen automatisch sammelt, zusammenführt und verarbeitet, damit diese über die Benutzeroberfläche der Applikationen dargestellt werden können. So wird eine Art „single point of truth" gebildet, wo alle wesentlichen Daten zusammenlaufen und abgerufen werden können. Betreiber können auch über die Benutzeroberfläche der Applikationen Dokumente wie Handbücher, Spezifikationen, Parametrierinformationen oder Anlagen-Backups zentral ablegen. So werden wichtige Informationen zuverlässig an einem Ort hinterlegt und können im Bedarfsfall auch zentral aktualisiert werden.

Wenn zum Beispiel ein neuer Fahrzeugtyp in die Produktion eingebracht wird, ist eine Änderung des bisherigen Prüfablaufs notwendig. Hierfür müssen alle betroffenen Anlagen diese neuen Prüfanweisungen erhalten. Über eine IIoT-basierte Lösung kann dies zentral erfolgen, ohne Eingabe von Hand an jedem Anlagen-PC. Durch die Anwendung von IIoT-Applikationen entfällt die bisher manuelle, zeitaufwändige Dateneingabe, Datenextraktion und Datenaufbereitung. Stattdessen kann über die Benutzeroberfläche direkt und in Echtzeit auf die Daten zugegriffen werden, ohne dass der Benutzer seinen Arbeitsplatz verlassen muss. In Summe unterstützen diese Lösungen dabei, die Anlagen mit höherem Bedienkomfort zu betreiben und letztendlich die für Automobilisten so wichtigen Kennzahlen wie produzierte Stückzahl und maximale Direktläuferquote auf einem kontinuierlich hohen Niveau zu halten oder gar zu verbessern.

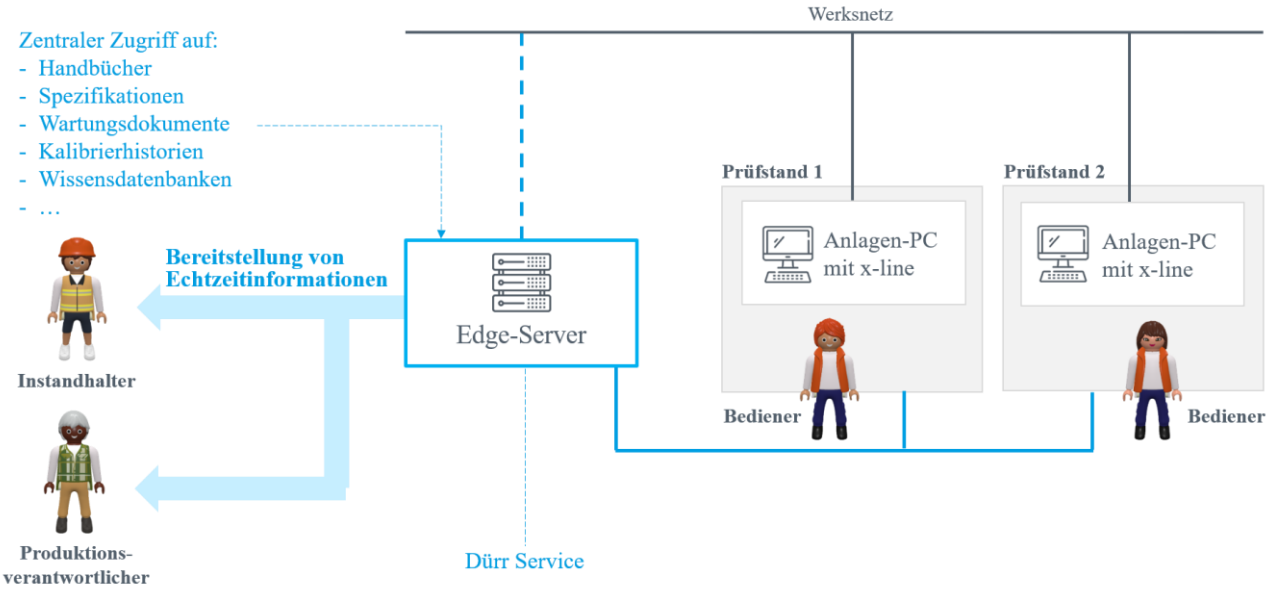


Abb. 4 Darstellung eines Bandendes mit IIoT-Anbindung

5 Zusammenfassung & Ausblick

Die Automobilbranche ist einer der Vorreiterindustrien in Sachen Automatisierung und Digitalisierung. Nichtsdestotrotz ist es nicht ausreichend, bloß Produktionsanlagen zu vernetzen und beliebig Daten zu sammeln. Stattdessen benötigt es zusätzlich das Wissen über die Ablageorte von Handbüchern und Schulungsunterlagen sowie internes und externes Expertenwissen, um Abweichungen in einer Produktion vollends verstehen und vermeiden zu können.

Die Dürr Assembly Products GmbH unterstützt ihre Kunden bei der Zentralisierung und Digitalisierung von wichtigen Anlagendokumenten und stellt gleichzeitig web-basierte IIoT-Applikationen zur Verfügung, die anlagenspezifische Informationen eines Dürr-Prüfstandes übersetzt und analysiert, damit Produktionsdaten schnell und einfach interpretiert, Probleme diagnostiziert und Aktionen eingeleitet werden können. Dadurch wird erstens der Durchsatz auf den Anlagen und zweitens die Qualität der produzierten Fahrzeuge auf einem gleichbleibend hohen Niveau gehalten, was die Produktionskosten pro Fahrzeug minimiert.

Solche Digitalisierungsvorhaben erhöhen die Transparenz, sparen Zeit auf der operativen Ebene und senken auf Dauer die Betriebskosten von Anlagen und Systemen, die bisher zu selten bei der Neuplanung von Produktionsstraßen in Betracht gezogen werden. Viel zu häufig konzentrieren sich planende und einkaufende Abteilungen darauf, Investitionskosten für Neuanlagen gering zu halten – nur selten werden Lebenszykluskosten in Betracht gezogen, wo neben Anschaffungs- auch Betriebs- und Verschrottungskosten über den gesamten Lebenszyklus einer Anlage oder eines Systems Beachtung finden. Denn neben Stückzahlen und Direktläuferquote will jedes produzierende Unternehmen die Produktionskosten reduzieren; das bedeutet im Umkehrschluss ungeplante Anlagenausfälle, Taktzeit- und Qualitätsverluste zu minimieren oder ganz zu vermeiden. IIoT-Lösungen aus dem Hause Dürr können unterstützen, Produktionen durch die sinnvolle Sammlung und Auswertung von Daten intelligenter und damit effizienter, robuster und kostengünstiger zu betreiben.

Heutzutage sind Lösungen für Vernetzung, Datenvisualisierung, Dateninterpretation und Vorhersage von Ereignissen sind im Markt gut etabliert. Doch einige Schritte sind noch zu gehen bis zur Realisierung einer „selbstoptimierenden Maschine“ [7], die sich selbst regelt und wo der Mensch vor allem als Überwacher und Korrektor fungiert. Allerdings hat die Industrie, allen voran der Dürr-Konzern, in den letzten Jahren neben dem existierenden Anlagenwissen das nötige Software-Know-how aufgebaut, damit eine solche selbstoptimierende Dürr-Anlage in naher Zukunft Wirklichkeit werden kann.

6 Literatur

[1] Fromm, L. (2015). Gerüstet für den Blackout. *Deutsche Verkehrs-Zeitung*, 12.03.2015. Verfügbar unter: https://www.dvz.de/rubriken/logistik/it-in-der-logistik/detail/news/geruestet-fuer-den-blackout.html, abgerufen am 05.01.2021.

[2] Süddeutsche Zeitung (2017). Betrunken und high: BMW-Mitarbeiter verursachen Stillstand am Band. *Süddeutsche.de*, 20.03.17. Verfügbar unter: https://www.sueddeutsche.de/muenchen/bmw-werk-betrunken-und-bekifft-bmw-mitarbeiter-verursachen-stillstand-am-band-1.3427597, abgerufen am 05.01.21.

[3] Kimberly Clark Professionals. (o.J.). Versteckte Kosten von Nacharbeit in der Automobilindustrie: Warum interne Nacharbeit mehr kostet, als Sie denken und was Sie tun können, um sie zu vermeiden. https://www.kcprofessional.de/media/144520592/Versteckte_Kosten_von_Nacharbeit_in_der_Automobilindutrie_von-Kimberly_Clark_Professional.pdf, abgerufen am 07.01.2021.

[4] Sedran, T. & Gissler, A. (2015). Wie die Autoindustrie die Chance der Digitalisierung richtig nutzt. *Accenture* / Sonderbeilage in Automobilwoche. Verfügbar unter: https://www.accenture.com/_acnmedia/accenture/conversion-assets/dotcom/documents/global/pdf/industries_18/accenture-automobilwoche-beilage-2015-german.pdf, aufgerufen am 07.01.2021.

[5] Kropik, M. (2009). *Produktionsleitsysteme in der Automobilfertigung*. Berlin: Springer-Verlag.

[6] Brodtmann, T., Schnittler, V., Becker, M., Herring, D., Mayer-Haug, K., & Schleyer, T. (2020). Kundenzentrierung als Chance für den digitalen Durchbruch – Was sich die Endkundenindustrien vom Maschinenbau bei digitalen Plattformen und Mehrwertdiensten wünschen. *VDMA e.V. / McKinsey & Company*.

[7] Schuh, G., Anderl, R., Gausemeier, J., ten Hompel, M., & Wahlster, W. (2017). *Industrie 4.0 Maturity Index. Die digitale Transformation von Unternehmen gestalten*. München: Herbert Utz.

7 Bildverweise

[Abb. 1] Computer-Icon von Freepik auf flaticon.com; Figuren erstellt auf figure.playmobil.com

[Abb. 2] Eisberg-Icon erstellt von Freepik auf flaticon.com

[Abb. 3] Warnzeichen-Icon von Gregor Cresna, restliche Icons von Freepik auf flaticon.com

[Abb. 4] wie Abb. 1, Server-Icon von srip auf flaticon.com

Maschinelles Lernen für die Situationserkennung einer aktiven Kniegelenkorthese

Machine Learning for Situation Detection of an Active Knee-Ankle-Foot-Orthosis

Philipp Mattfeldt, Bastian Latsch, Jürgen Hielscher, Sascha Desch, Niklas Schäfer, Romal Chadda, Roland Werthschützky, André Seyfarth, Mario Kupnik

Technische Universität Darmstadt, Fachgebiet Mess- und Sensortechnik, 64283 Darmstadt, Deutschland, bastian.latsch@tu-darmstadt.de

Kurzfassung

Um den Menschen bedarfsgerecht und ergonomisch unterstützen zu können, müssen Assistenzsysteme die Bewegungen und Bewegungsabsichten des Nutzers frühzeitig erkennen. Die Individualität der menschlichen Bewegung ist dabei eine besondere Herausforderung. Dieser Beitrag zeigt, dass maschinelles Lernen geeignet ist, um aus den Daten ausgewählter, am Körper getragener Sensoren sieben alltägliche Bewegungssituationen zu unterscheiden. Ein Convolutional Neural Network erkennt Merkmale in den Bewegungsdaten, welche nicht nur auf die aktuelle Situation schließen lassen, sondern auch als Frühindikatoren dienen, aus denen eine kommende Bewegung antizipiert werden kann. Dazu ermittelt das Netzwerk für jede vordefinierte Situation eine Plausibilität im laufenden Betrieb basierend auf den Sensorsignalen des Assistenzsystems. Das hier beschriebene Modell wurde mit Bewegungsdaten von sechs Personen trainiert, wodurch eine Erkennungsgenauigkeit von 99,1 % erreicht wurde. Die Robustheit der Erkennung ist mit der Einführung einer Konfidenz gegeben und somit ein Einsatz der Situationserkennung an einer aktiven Kniegelenkorthese möglich.

Abstract

In order to support humans ergonomically and in line with their needs, assistance systems have to recognize the user's movements and movement intentions at an early stage. Here, the individuality of human movement is a key challenge. This paper shows that machine learning is suitable to distinguish seven common movement situations from the data of selected sensors attached to the body. A Convolutional Neural Network detects features in the motion data that not only indicate the current situation but also provide indicators from which an upcoming motion can be anticipated. For this purpose, the network determines a plausibility for each predefined situation in the ongoing operation based on the measured values of the assistance systems. The model described was trained with the motion data of six participants resulting in a recognition accuracy of 99,1 %. The robustness of the recognition is given with the aid of a confidence, and thus an application of the situation recognition in an active knee orthosis is possible.

1 Einleitung

Der demografische Wandel und der damit einhergehende steigende Anteil älterer Menschen stellt die Gesellschaft vor eine große Herausforderung. Vor allem ältere oder körperlich eingeschränkte Menschen verfügen oft nur über eine eingeschränkte Mobilität [1]. Dabei können einfache Bewegungen wie Treppensteigen oder das Aufstehen von einem Stuhl zur Belastung werden. In einem solchen Fall helfen Assistenzsysteme [2].

In vorheriger Arbeit [3] wurde eine aktive Kniegelenkorthese entwickelt, die das Aufstehen und Treppensteigen unterstützt (Abb. 1a). Durch die Orthese wird gerade der aufgrund der Muskelschwäche fehlende Teil des Drehmoments im Knie kompensiert. Je nach Bewegungssituation und Bedürfnis des Anwenders sind allerdings unterschiedliche Unterstützungsfaktoren notwendig. Um das Zuschalten der Unterstützung im richtigen Zeitpunkt und mit angepasstem Drehmoment zu gewährleisten, ist die Erkennung der aktuellen Bewegungssituation im laufenden Betrieb essenziell.

Für die Erkennung und Analyse von Bewegungsabläufen werden in der Sportwissenschaft und Biomechanik mehrere bewegungsspezifische Messgrößen am menschlichen Körper zeitkontinuierlich erfasst. Hierbei wird zwischen kinemetrischen Messgrößen, wie Segmentbeschleunigung und Gelenkwinkel, physiologischen Messgrößen, wie der Muskelaktivität, und dynamometrischen Messgrößen, wie den Bodenreaktionskräften und dem Körperschwerpunkt (center of pressure, CoP), unterschieden. Mithilfe solcher Messgrößen ist eine Klassifizierung der aktuellen Bewegungssituation möglich [4].

Der Anspruch dabei ist eine individuell auf den Nutzer zugeschnittene und optimierte Verarbeitung. Dadurch wird eine über die Nutzungsdauer andauernde Flexibilität erreicht, die auf Änderungen im Bewegungsverhalten konstruktiv eingehen kann. Ein verbreiteter Ansatz zur Klassifizierung, der die genannten Punkte erfüllt, ist Machine Learning. Hierbei werden mittlerweile vermehrt Ansätze des Deep Learning für die Erkennung menschlicher Aktivitäten (human activity recognition) eingesetzt [5]. In diesem Zusammenhang kommt besonders häufig die Klasse der Convolutional Neural Networks (CNNs) zum Einsatz [5]. Beispiels-

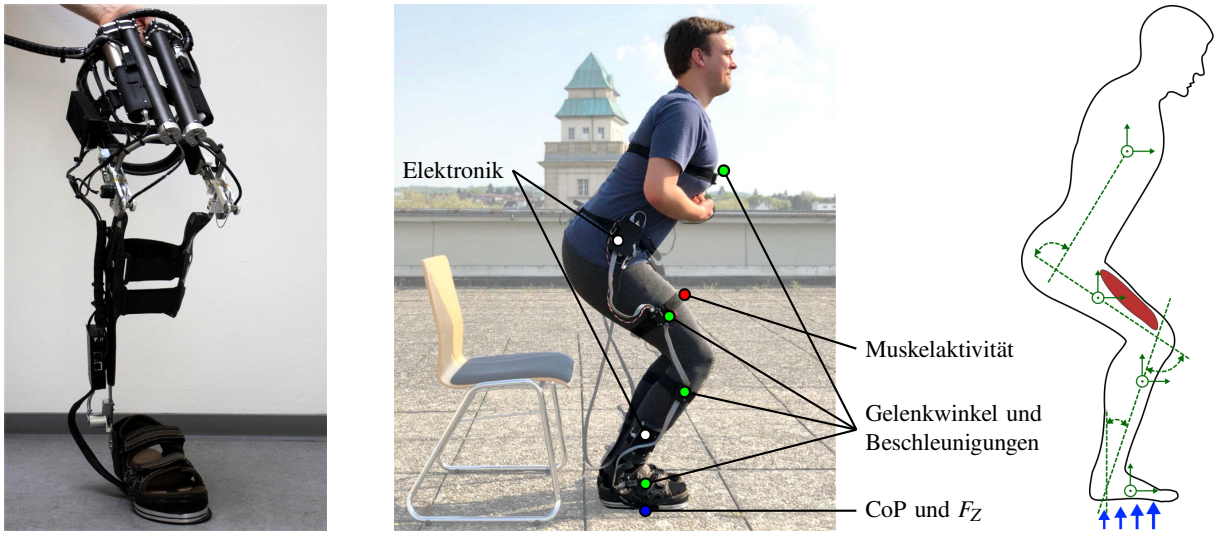


Abbildung 1 (a) Labormuster der aktiven Kniegelenkorthese nach [3] mit Kraftsensoren im Fußteil und Erfassung der Gelenkwinkel in Knie und Sprunggelenk. Die Aktorik ist am Oberschenkelteil befestigt. (b) Die Umsetzung der Sensorik-Orthese beinhaltet die zu erfassenden kinemetrischen Größen Gelenkwinkel sowie lineare Beschleunigungen und Rotationsgeschwindigkeiten (grün), die physiologische Messgröße Muskelaktivität der Muskelbündel am Oberschenkel (rot) sowie die dynamometrischen Messgrößen Körperschwerpunkt (CoP) und Bodenreaktionskräfte F_Z (blau). (c) Schematische Skizze der erfassten Messgrößen an der Sensorik-Orthese.

weise werden basierend auf den Gyroskop- und Beschleunigungsdaten eines Smartphones sechs verschiedene Aktivitäten (Gehen, Treppauf, Treppab, Sitzen, Stehen, Liegen) klassifiziert [6]. Viele dieser Ansätze berichten bereits hohe Erkennungsgenauigkeiten. Allerdings sind sie oftmals nicht für den Einsatz im laufenden Betrieb geeignet oder werden nicht für die Erkennung von Übergangssituationen, wie Aufstehen und Hinsetzen, ausgelegt. Beide Aspekte sind für die Anwendung bei einer aktiven Kniegelenkorthese jedoch essenziell. Ein weiterer wesentlicher Aspekt ist, dass unter realen Bedingungen auch unbekannte Situationen auftreten, welche dann oftmals trotzdem eindeutig einer bekannten Situation zugeordnet werden. Dabei gibt das Modell in der Regel keinen Aufschluss darüber, in welchem Maß der Erkennung vertraut werden kann.

In dieser Arbeit wird ein Modell zur Situationserkennung für die Anwendung an einer aktiven Kniegelenkorthese vorgestellt, das die zuvor genannten Herausforderungen adressiert. Hierzu wird für die Generierung von Trainings- und Testdaten zunächst eine passive Orthese mit Sensorik ausgestattet (Sensorik-Orthese). Das trainierte mehrschichtige CNN mit abschließendem Fully-Connected-Layer wird evaluiert. Hierbei wird untersucht, wie die Berücksichtigung personalisierter Trainingsdaten die Modellgüte beeinflusst. Darüber hinaus wird ein Gütemaß eingeführt, welches das Vertrauen quantifiziert, das einer Erkennung des Modells entgegengebracht werden kann.

2 Sensorik-Orthese

Das Ziel der Sensorik-Orthese ist, eine umfangreiche Erfassung verschiedener Messgrößen zu ermöglichen, ohne die Probanden in ihrer Bewegungsfreiheit einzuschränken. Hierfür wird bei der Entwicklung auf die Verwendung möglichst leichter und flexibler Materialien geachtet. So werden beispielsweise die Sprung-, Knie- und Hüftgelenkswinkel berührungslos aus den Orientierungen inertialer Messeinheiten (IMUs) ermittelt, welche jeweils mit elastischen Manschetten an Oberkörper, Oberschenkel, Unterschenkel und am Fuß appliziert sind (Abb. 1). Zusätzlich werden über die IMUs (BNO055, Bosch Sensortec, Reutlingen, Deutschland) die linearen Beschleunigungen sowie die Rotationsgeschwindigkeiten dieser Segmente hinsichtlich der drei Raumrichtungen erfasst. Die Detektion der Muskelaktivität des Kniestreckers erfolgt über Elektromyografie (EMG) mit Sensoren (MyoWare Muscle Sensor, SparkFun Electronics, Boulder, Colorado, USA) an den Oberschenkeln und Biegesensoren (Flex-Sensor 4.5", Spectra Symbol, Salt Lake City, Utah, USA), die in der Oberschenkel-Manschette integriert sind und so die Kontraktion der Oberschenkel-Muskulatur detektieren. Die Bodenreaktionskräfte sowie der Körperschwerpunkt werden mittels acht Wägezellen (FX1091, Measurement Specialities Inc., Hampton, Virginia, USA) in den Schuhen ermittelt. Die Sensorik-Orthese übermittelt die sich so ergebenden 64 Messgrößen mit einer Abtastrate von 1 kHz an einen Rechner zur weiteren Verarbeitung.

Mit der Sensorik-Orthese führten sechs Probanden unterschiedlichen Alters (39 $\pm$ 15 Jahre), Geschlechts (vier männlich, zwei weiblich), Größe (1,78 m $\pm$ 0,09 m) und Gewichts (76 kg $\pm$ 12 kg) jeweils 30 Minuten lang Bewegungen aus den sieben Alltagsbewegungen durch. Dabei wurde den Probanden keine Einschränkung hinsichtlich Reihenfolge oder auch Ausführungsart gegeben, um den Datensatz mit Diversität zu füllen und den Anteil der einzelnen Situationen (Tab. 1) nicht vorherzubestimmen.

Tabelle 1 Die Verteilung der repräsentierten Situationen im gesamten dreistündigen Datensatz berechnet sich aus ihrer empirischen Beobachtung und dem zeitlichen Anteil.

Situation	Anteil h_i	Umfang im Datensatz
Gehen	25 %	ca. 4500 Schritte
Stehen	15 %	ca. 40 Minuten
Sitzen	15 %	ca. 40 Minuten
Treppauf	12,5 %	jeweils ca. 2100 Schritte
Treppab	12,5 %	
Aufstehen	10 %	jeweils ca. 700 Vorgänge
Hinsetzen	10 %	

Für den hier verwendeten Ansatz mit überwachtem Lernen ist es essenziell, dass zu jedem Zeitpunkt im Datensatz das Label (sog. Ground Truth), also die korrekte Situation zu den erfassten Messgrößen, vorliegt. Das Labeln der Daten erfolgt in der Literatur meist manuell während der Erfassung oder im Nachhinein manuell unter Zuhilfenahme synchronisierter Videoaufnahmen [7] und stellt damit einen erheblichen Arbeitsaufwand dar. Zur Erleichterung dieses Prozesses wurde basierend auf einem Zustandsautomaten ein „Auto-Labeler“ entwickelt, der einen mehrere Sekunden umfassenden Zeitraum überblicken kann. Durch dieses zusätzlich verfügbare Wissen gegenüber einer Echtzeit-Verarbeitung kann robust mittels boolescher Logik zwischen den sieben Alltagssituationen unterschieden werden. Die Daten können so nach der Aufnahme automatisiert, optional mit geringfügig manueller Anpassung, für die spätere Nutzung aufbereitet werden. Abschließend werden 80 % der gelabelten Daten eines Probanden dem Trainingsdatensatz und die restlichen 20 % dem Testdatensatz zugeteilt.

3 Netzarchitektur der Situationserkennung

Für das Modell zur Situationserkennung wird ein künstliches neuronales Netz mit mehreren Schichten so genannter Convolutional-Layer und zwei finalen Fully-Connected-Layern gewählt. Die Inputs sind zeitliche Fenster T_F von zwei Sekunden Länge (Abb. 2). Die Convolutional-Layer extrahieren dabei zeitliche und signal-übergreifende Merkmale (Features) aus den Sensorsignalen. Die abschließenden Fully-Connected-Layer lernen schließlich aus der Kombination der jeweiligen Merkmale die Plausibilitäten für jede Situation. Entgegen vieler bisheriger Forschungen [6], [8] wird der Output-Layer nicht via Softmax-Funktion als One-Hot-Encoded gestaltet, sondern über eine lineare Aktivierungsfunktion als Regression. Dem aktuellen Trend der „explainable AI“ [9] folgend wird dieser Ansatz im Kontext der Orthese zur Steigerung der Nutzersicherheit gewählt. Er ermöglicht es, durch dem Modell nachgeschaltete Operationen wie die Bewertung einer Konfidenz, in die Entscheidung des Modells einzugreifen.
Das Modell lernt während des Trainings die Plausibilität einer jeden Situation als Funktion der Sensorsignale. Das Lernen dieser Funktion entspricht einer Optimierung der Gewichte, welche die freien Parameter der einzelnen Schichten sind. Im Falle der Convolutional-Layer entsprechen diese Parameter den Filtern, mit denen das Eingangssignal gefaltet wird, um das Ausgangssignal zu berechnen. Bei dem Fully-Connected-Layer sind die Gewichte die Einträge in einer Matrix, mit der das Eingangssignal multipliziert wird, um die Ausgänge zu erhalten. Um mit diesem Aufbau, der im Wesentlichen aus zwei linearen Operationen (Faltung und Matrix-Multiplikation) besteht, nichtlineare Funktionen wie die Situationserkennung der Orthese zu erlernen, wird zwischen den einzelnen Schichten eine Nichtlinearität eingebracht. Hierfür wird die ReLU-Funktion (Rectified Linear Unit) aufgrund ihrer schnellen Lernkonvergenz in Mehrschichtmodellen und der weiten Verbreitung in der Klassifizierung [10] genutzt.

4 F_1-Maß und Konfidenz zur Evaluierung der Modellgüte

Zur Evaluierung von Machine-Learning-Klassifizierern wird in der Regel eine Bewertung nach dem F_1-Maß mit *k-fold cross-validation* (hier $k = 3$) durchgeführt [11]. Das F_1-Maß ist ein Gütemaß, das die zwei Gütekriterien *Präzision* und *Trefferquote* gleichmäßig gewichtet über alle N Klassen (hier $N = 7$) kombiniert. Die Präzision einer Klasse gibt an, ob eine positive Entscheidung des Modells korrekt war, wenn eine Klasse als positiv gelabelt war. Die Präzision ist

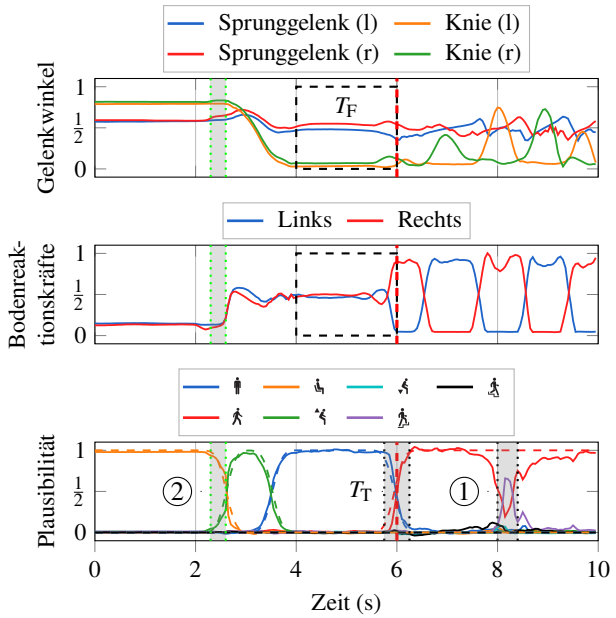


Abbildung 2 Die Sensordaten (Fensterbreite $T_\mathrm{F} = 2\,\mathrm{s}$) fließen zum aktuellsten Zeitpunkt (rot gestrichelt) in die Prognose des Modells ein. Aufgrund der natürlichen Unschärfe beim Situationsübergang wird ein Toleranzbereich $T_\mathrm{T} = \pm 250\,\mathrm{ms}$ definiert, in dem eine Transition des Modells als richtig bewertet wird. ① Dabei kann es auch zu einer Fehlentscheidung mit anschließender Korrektur zwischen Plausibilität (durchgezogen) und Ground Truth (gestrichelt) kommen. ② Das Modell erkennt Frühindikatoren für einen sich ankündigenden Situationswechsel.

somit das Verhältnis aus den Wahr-Positiv-Entscheidungen zur Summe aus Wahr-Positiv (WP) und Falsch-Positiv (FP). Die Trefferquote einer Klasse gibt an, ob das Modell eine Klasse als positiv erkannt hat, wenn sie dies in Wahrheit war. Die Trefferquote ist also das Verhältnis aus Wahr-Positiv-Entscheidungen zur Summe aus Wahr-Positiv und Falsch-Negativ (FN).

$$Präzision = \frac{WP}{WP+FP}, \quad Trefferquote = \frac{WP}{WP+FN} \tag{1}$$

$$F_1 = \sum_i^N 2 \cdot h_i \cdot \frac{Präzision_i \cdot Trefferquote_i}{Präzision_i + Trefferquote_i} \tag{2}$$

Zur Evaluierung der Modelle im laufenden Betrieb wird der Testdatensatz kontinuierlich (alle 10 ms) hinsichtlich der Situationserkennung ausgewertet. Für die Berechnung des F_1-Maßes wird an jedem dieser Zeitpunkte die dominante Situation als Argmax-Funktion der Plausibilitäten bestimmt. Eine Herausforderung stellt die Definition des exakten Übergangszeitpunkts zwischen zwei Situationen, sodass nicht immer eine scharfe Trennung möglich ist. Dies ist auch beim manuellen Labeln schwierig, da teilweise bei den Situationswechseln keine eindeutige Trennlinie existiert. Um einer Verfälschung der Modellgüte durch diese Unschärfe vorzubeugen, muss das Modell innerhalb eines Toleranzfensters $T_\mathrm{T} = \pm 250\,\mathrm{ms}$ den Situationswechsel erkennen, damit diese Entscheidung als richtig gewertet wird (Abb. 2).
Zur Steigerung der Nutzersicherheit ist das Überprüfen der getroffenen Entscheidung erforderlich. Hierzu wird das Modell um eine Konfidenz erweitert, die dem neuronalen Netz nachgeschaltet ist. Diese stellt ein Gütemaß dar, welche anhand eigens definierter Kriterien die Plausibilitäten des Modells in den vergangenen 200 ms bewertet. Zu diesen Kriterien zählen die relativen Differenzen aller Plausibilitäten und das Rauschen der Plausibilitäten. Ersteres bedeutet im Falle einer unsicheren Entscheidung des Modells, dass mehrere Situationen eine Plausibilität größer als Null besitzen. Letzteres bedeutet im Falle einer unsicheren Entscheidung des Modells, dass die Plausibilitäten stärker rauschen und die Transitionen zwischen zwei Situationen sehr steil sind. Beide Kriterien werden über Gewichtungsfaktoren miteinander kombiniert und stellen die Konfidenz dar. Unterschreitet die Konfidenz einen definierten Schwellenwert (z. B. 50 %), kann dies als Trigger für einen übergeordneten Sicherheitsmodus genutzt werden.

5 Experimente

Das vorgestellte Modell wird mit den Trainingsdaten trainiert und anhand von Testdaten, die das Modell zuvor nicht gesehen hat, evaluiert. Dabei ist für den Einsatz solch einer Situationserkennung vor allem die Robustheit des Modells relevant und wird daher in zwei Fällen untersucht. Im ersten Fall ist die universelle Einsetzbarkeit des Modells von Interesse. Dies wird durch eine allgemeine und breite Anwendung der Kniegelenkorthese motiviert und erfordert, dass das Modell auf die Daten eines unbekannten Anwenders adaptierbar ist. Hierzu werden in einer Versuchsreihe mehrere Modellvarianten mit gleichen Parametern, jedoch mit den Daten variierender Probanden, trainiert. Für die Evaluierung wird jeder Proband gegen alle Modelle getestet, sodass folgende Szenarien entstehen:

- **Personalisiert**: Test eines jeden Probanden mit dem Modell aus seinen individuellen Trainingsdaten;
- **Exklusiv**: Test eines jeden Probanden mit Modellen, die ausschließlich mit Daten anderer Probanden trainiert wurden, wobei die Anzahl der Probanden im Training erhöht wird;
- **Alle**: Test eines jeden Probanden mit einem universellen Modell, welches mit den Trainingsdaten aller Probanden trainiert wurde.

Im zweiten Fall zur Untersuchung der Robustheit wird das Verhalten des Modells bei unbekannten Situationen geprüft. Dies ist ein wesentlicher Sicherheitsaspekt für den Nutzer der Kniegelenkorthese. Im Alltag können neben den sieben hier behandelten Situationen weitere Situationen und besondere Bewegungen, wie beispielsweise das Stürzen des Nutzers, vorkommen. Auch in diesen Fällen ist die Robustheit der Situationserkennung für die Nutzersicherheit unabdingbar. Dafür wird das Modell Situationen ausgesetzt, die während des Trainings in dieser Form nie aufgetreten sind.

6 Ergebnisse & Diskussion

Das Experiment zum universellen Einsatz des Modells zeigt, dass die Genauigkeit mit steigender Anzahl an Probanden stetig ansteigt (Abb. 3). Probandenanzahl „1“ bis „5“ bilden die Gruppe *Exklusiv*. Eine Probandenanzahl von eins bedeutet, dass jeder Proband gegen die personalisierten Modelle jedes anderen Probanden getestet wurde. Bei sechs verfügbaren Probanden, wird also jeder Proband gegen die fünf personalisierten Modelle der anderen Probanden getestet. Bei einem Modell der Probandenanzahl „2“ enthalten die Modelle entsprechend alle möglichen Kombinationen von jeweils zwei Probanden, jedoch immer exklusive des zu testenden Probanden. Eine Probandenanzahl von sechs entspricht somit *Alle*. Neben der steigenden Genauigkeit sinkt gleichzeitig die Streuung zwischen den einzelnen Modellen einer Probandenanzahl. Dies wird vor allem beim Testen gegen die personalisierten Modelle der anderen Probanden „1“ deutlich. Hier weist das F_1-Maß teilweise eine starke Streuung von bis zu 17 % auf (Proband P3 und P6). Es wird allerdings auch deutlich, dass vereinzelte Probanden bereits gut durch personalisierte Modelle anderer Probanden repräsentierbar sind (Probanden P1 und P5). Dies weist darauf hin, dass sich die Bewegungsmuster dieser Probanden ähneln. Dennoch lassen sich personalisierte Modelle für einen universellen Einsatz optimieren. Wird die Anzahl der Probanden im Trainingsdatensatz erhöht, steigt die durchschnittliche Genauigkeit von 93,9 % bei einer Probandenanzahl von eins auf 98,2 % bei einem Modell mit fünf Probanden an. Werden zusätzlich Daten des zu testenden Probanden in den Trainingsdatensatz mit aufgenommen, steigt die Genauigkeit nochmals um 0,9 Prozentpunkte auf 99,1 %

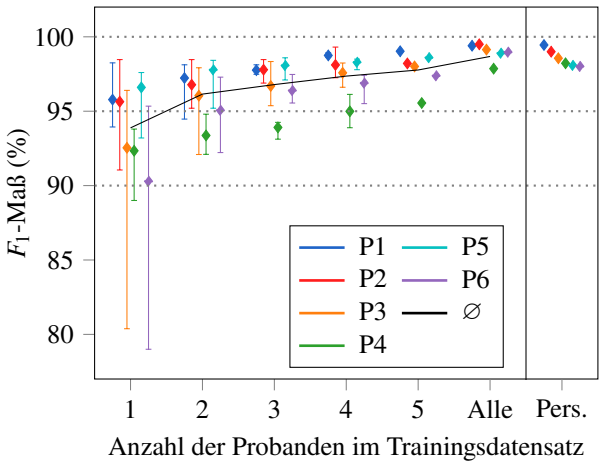


Abbildung 3 Beim Vergleich der Modellgenauigkeit mit unterschiedlicher Anzahl an Probanden im Trainingsdatensatz wird gegen mehrere Modelle bestehend aus den Daten von insgesamt sechs Probanden getestet. Ein farbiger Punkt markiert den Mittelwert eines Probanden (P1 bis P6) für diese Anzahl an Probanden, mit denen das Modell trainiert wurde. Der zugehörige Fehlerbalken beschreibt den Bereich der Modellgenauigkeit, der beim Testen gegen die Modelle dieser Probandenanzahl erreicht wird. Sowohl die Modellgenauigkeit als auch die Streuung der Genauigkeit verbessern sich mit steigender Probandenanzahl und übersteigen teilweise sogar die des personalisierten Modells (Pers.).

an (*Alle*). Dieses universelle Modell mit den Daten aller verfügbaren Probanden ist selbst im Vergleich zu den personalisierten Modellen eines jeden Probanden mit ihrer durchschnittlichen Genauigkeit von 98,6 % um 0,5 Prozentpunkte genauer. Dabei erfolgt die Erkennung mit einer durchschnittlichen Konfidenz von 99,4 % und einer Laufzeit von 5 ms pro Berechnung (GeForce GT440, NVidia, Santa Clara, USA). In der Konfusionsmatrix (Tab. 2) wird deutlich, dass nur in wenigen bestimmten Kombinationen die vom universellen Modell ermittelte dominante Plausibilität von der Ground Truth abweicht. Eine nähere Betrachtung der Abweichungen zeigt, dass in nahezu allen Fällen eine Abweichung aus der Unschärfe der Ground Truth resultiert. Diese existierende Unschärfe ist jedoch grundsätzlich nicht negativ, sondern bietet dem Modell die Möglichkeit, Frühindikatoren für einen kommenden Situationswechsel in Signalverläufen zu erkennen. So interpretiert beispielsweise das universelle Modell das Heranziehen der Füße und das Vorbeugen des Oberkörpers im 🧍 als Vorankündigung für 🚶 . Die Plausibilität steigt hier bereits auf ca. 30-40 % an (Abb. 2, ②). Durch diese Frühindikation des Modells ist die Orthese schneller bereit dem Anwender das unterstützende Drehmoment hinzuzuschalten, sobald dieser mit dem Belasten der Füße das ausschlaggebende Signal zur 🚶 Bewegung gibt.

Bei unbekannten Situationen wie beispielsweise dem *Einbeinstand mit Knie- und Fußwackeln* (Abb. 4, ①) erkennt das universelle Modelle dies als wechselndes 🧍 🚶 . Obwohl in diesem Fall keine eindeutige Ground Truth besteht, ist die Erkennung des universellen Modells nachvollziehbar und verständlich. Das initiale In-die-Knie-Gehen beim *In-der-Hocke-Wippen* ② wird souverän als 🚶 verbucht. Nachfolgend wechselt die Prognose wieder zurück zu 🧍 . Somit erkennt das universelle Modell nachvollziehbar diese Pose eher als Stehen statt Sitzen. Beim *Auf-Zehenspitzen-Stehen* ③ werden keine merklichen Impulse ausgelöst, von der Entscheidung 🧍 abzuweichen. Die letzte Testsituation, das *dreimalige Springen* ④, ist hingegen eine Herausforderung für das universelle Modell. Das initiale In-die-Knie-Gehen wird noch verständlicherweise als 🚶 erkannt. Die Landung resultiert jedoch in konfusen Plausibilitäten, aus denen keine dominante Situation abgeleitet werden kann. In diesem Fall greift die absichernde Konfidenz, sodass das Modell in den übergeordneten Sicherheitsmodus gelangt und die Sicherheit für den Nutzer weiterhin gegeben ist.

Tabelle 2 Der Konfusionmatrix des universellen Modells mit einem F_1-Maß von 99,1 % und einer Konfidenz von 99,4 % zufolge werden alle Situationen bis auf wenige Kombinationen richtig erkannt. Abweichungen sind in so gut wie allen Fällen auf die Unschärfe der Ground Truth zurückzuführen.

		Modellentscheidung						
		Stehen	Gehen	Sitzen	Aufstehen	Hinsetzen	Treppauf	Treppab
Ground Truth	Stehen	98,89	1,11	0	0	0	0	0
	Gehen	0,55	98,84	0	0	0,08	0,21	0,32
	Sitzen	0	0	99,81	0,19	0	0	0
	Aufstehen	0	0,27	0	99,73	0	0	0
	Hinsetzen	0	0	0	0	100	0	0
	Treppauf	0	0,88	0	0	0	99,12	0
	Treppab	0	0,6	0	0	0	0	99,4

7 Schlussfolgerung

Der vorgestellte Ansatz zeigt, dass die Situationserkennung mittels neuronaler Netze basierend auf der Messung von Bewegungsdaten möglich ist. Dabei hat sich eine größere Menge an Trainingsdaten als Vorteil gezeigt, um ein universelles Modell zu erstellen, welches durchschnittlich besser abschneidet als ein personalisiertes Modell. Ein Grund dafür sind Bewegungen, die ein Proband anders als üblich ausführt. Eine korrekte Erkennung wird in solchen Fällen durch die Diversität aller Probanden in einem universellen Modell möglich. Mit Einführung der Konfidenz sind unbekannte Situationen erkennbar und somit erhöht sich die Nutzersicherheit. In weiteren Untersuchungen wird ein größerer Trainingsdatensatz erhoben, um die Anwendung des universellen Modells bei einer erhöhten Probandenanzahl zu bestätigen. Damit lässt sich die aktive Kniegelenkorthese um ein universelles Modell zur Situationserkennung erweitern, um den Nutzer dem Bedarfsfall entsprechend zu unterstützen.

Danksagung

Diese Forschung wurde unterstützt von der Deutschen Forschungsgemeinschaft unter den Projektnummern 164569068 und 402740893.

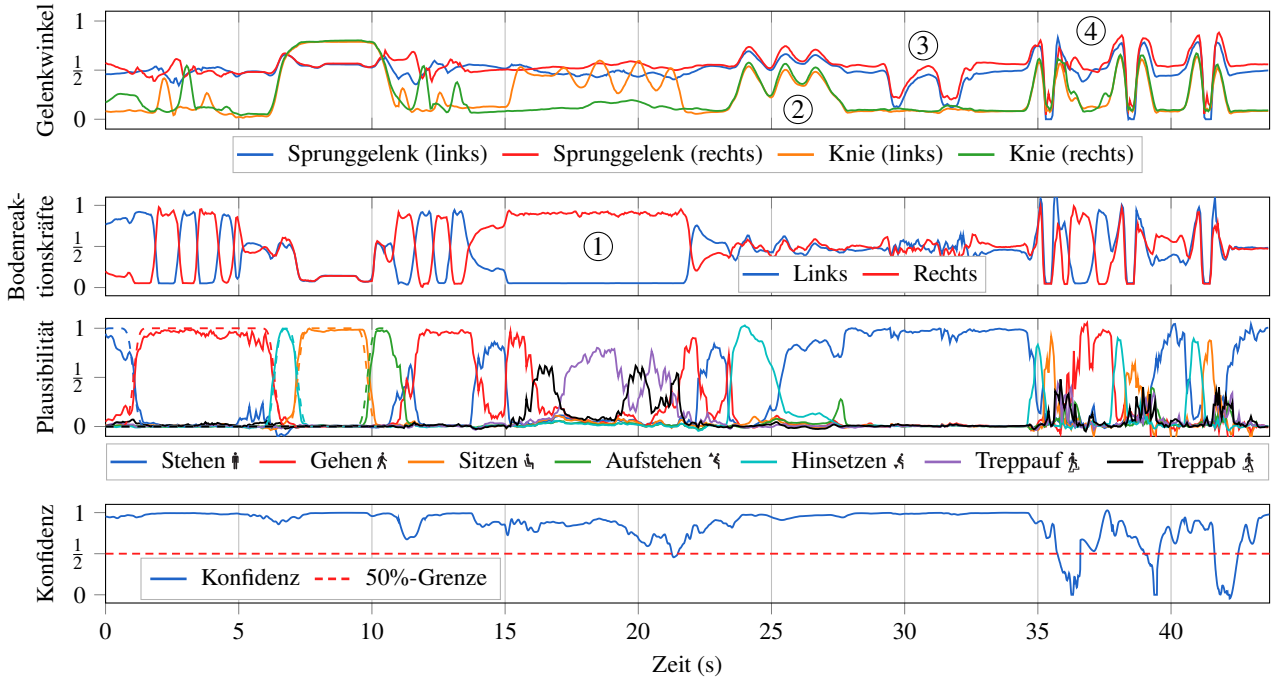


Abbildung 4 Aus den am Probanden gemessenen Sensorsignalen, wie Gelenkwinkel und Bodenreaktionskräfte, berechnet das Modell die Plausibilitäten für die aktuelle Bewegungssituation und gibt die Konfidenz der Berechnung an (normierte Verläufe). Bei angelernten, also bekannten, Situationen (bis ca. 15 s) sind die Plausibilitäten der einzelnen Situationen (durchgezogen) nahezu gleich mit der vorliegenden Ground Truth (gestrichelt). Anhand von dem Modell unbekannten Situationen lässt sich die Stärke unseres Ansatzes erkennen (ab 15 s), da in der Realität auch eine manuelle Einschätzung selten eindeutig ist. ① Beim *Einbeinstand mit Knie- und Fußwackeln* schwankt die Erkennung zwischen 🧍 und 🧍 , was nachvollziehbar der Situation am nächsten kommt. Da keine Plausibilität dominiert, sinkt die Konfidenz. ② Beim *Wippen in der Hocke* wird das In-die-Knie-Gehen korrekterweise als 🚶 erkannt. ③ Das *Stehen auf Zehenspitzen* besitzt wenig Einfluss auf die 🧍 Entscheidung. ④ Die Klassifizierung 🚶 beim Schwungholen zum *dreimaligen Springen* ist gefolgt von konfusen Plausibilitäten, woraufhin die Konfidenz absinkt.

Literatur

[1] S. R. Lord, R. D. Clark und I. W. Webster, „Postural Stability and Associated Physiological Factors in a Population of Aged Persons,“ en, *Journal of Gerontology*, Jg. 46, Nr. 3, S. M69–M76, Mai 1991.

[2] L. Zhang, G. Liu, B. Han u. a., „Assistive devices of human knee joint: A review,“ en, *Robotics and Autonomous Systems*, Jg. 125, S. 103 394, März 2020.

[3] P. P. Pott, S. I. Wolf, J. Block u. a., „Knee-ankle-foot orthosis with powered knee for support in the elderly,“ en, *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, Jg. 231, Nr. 8, S. 715–727, Aug. 2017.

[4] A. Gollhofer und E. Müller, Hrsg., *Handbuch Sportbiomechanik*, Ser. Beiträge zur Lehre und Forschung im Sport 171. Schorndorf: Hofmann-Verlag, 2009.

[5] J. Wang, Y. Chen, S. Hao u. a., „Deep learning for sensor-based activity recognition: A survey,“ en, *Pattern Recognition Letters*, Jg. 119, S. 3–11, März 2019.

[6] C. A. Ronao und S.-B. Cho, „Human activity recognition with smartphone sensors using deep learning neural networks,“ en, *Expert Systems with Applications*, Jg. 59, S. 235–244, Okt. 2016.

[7] A. Bulling, U. Blanke und B. Schiele, „A tutorial on human activity recognition using body-worn inertial sensors,“ en, *ACM Computing Surveys*, Jg. 46, Nr. 3, S. 1–33, Jan. 2014.

[8] B. Fang, Q. Zhou, F. Sun u. a., „Gait Neural Network for Human-Exoskeleton Interaction,“ *Frontiers in Neurorobotics*, Jg. 14, S. 58, Okt. 2020.

[9] A. Adadi und M. Berrada, „Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),“ *IEEE Access*, Jg. 6, S. 52 138–52 160, 2018.

[10] K. Hara, D. Saito und H. Shouno, „Analysis of function of rectified linear unit used in deep learning,“ in *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney, Ireland: IEEE, Juli 2015, S. 1–8.

[11] D. M. W. Powers, „Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,“ *International Journal of Machine Learning Technology*, 2011.

Formale semantische Repräsentation regelungstechnischer Methoden
Formal Semantic Representation of Methods in Automatic Control

Dipl.-Ing. Robert Heedt, Dr.-Ing. Carsten Knoll, Prof. Dr.-Ing. habil. Dipl.-Math. Klaus Röbenack
TU Dresden, Institut für Regelungs- und Steuerungstheorie, 01062 Dresden, Deutschland
E-Mail: {robert.heedt, carsten.knoll, klaus.roebenack}@tu-dresden.de

Kurzfassung

Regelungstechnisches Wissen in klassischer Textform ist schwer zugänglich, wenn man für die aktuelle Problemstellung nicht die exakten Begriffe kennt. Als Ergänzung zur klassischen Wissenrepräsentation schlagen wir daher ein „Methodennetz“ vor, bestehend aus Typen und Methoden angeordnet in einer Graphstruktur. Für eine konkrete Aufgabe kann dann mittels Tiefensuche ein prozeduraler Lösungsablauf generiert werden, dessen Nutzen für den Wissenstransfer wir am Beispiel der Trajektorienfolgeregelung für einen Dreifachpendelversuchsstand skizzieren.

Abstract

Written knowledge about automatic control theory is hard to access as it requires familiarity with the exact terminology. Therefore, we propose the “method net”, a supplement to classical knowledge representation, consisting of types and methods in a graph structure. From that, a schematic solution procedure can be generated for a specific problem. Trajectory tracking control for a triple pendulum is used to demonstrate how the proposed methode supports knowledge transfer.

1 Einleitung und Motivation

In vielen mechatronischen Produkten sind regelungstechnische Komponenten unverzichtbar. Die zugrundeliegende Regelungstheorie deckt inzwischen ein sehr breites Methodenspektrum ab und unterliegt, wie auch andere Wissensgebiete, einem beständigen Wachstum.
Als domänenunabhängige bzw. -übergreifende Querschnittsdisziplin steht sie jedoch vor einer besonderen Herausforderung bezüglich des Wissenstransfers, sowohl innerhalb der Disziplin als auch und besonders in potenzielle Anwendungsfelder. Gleichzeitig weisen die regelungstheoretischen Konzepte und Verfahren eine starke Heterogenität bezüglich expliziter und impliziter Verständnis- und Anwendungsvoraussetzungen auf – man vergleiche etwa einen einschleifigen P-Regelkreis mit einer nichtlinearen Mehrgrößen-Trajektorienfolgeregelung.
Als Konsequenz können einerseits fortgeschrittene regelungstechnische Methoden ihr Potential nicht voll entfalten, weil potentielle Anwender nicht von deren Existenz wissen, oder mangelndes Verständnis die Akzeptanz verhindert. Andererseits behindert suboptimaler Wissenstransfer auch die wissenschaftliche Neu- und Weiterentwicklung von Methoden.
Der vorliegende Beitrag schlägt deshalb einen formalen Ansatz als experimentelle Ergänzung zur bisher dominierenden Wissensrepräsentation vor.
Üblicherweise wird Wissen über geschriebene natürliche Sprache vermittelt, angereichert mit mathematischer Symbolik und Visualisierungen. Dieser textbasierte Informationsvorrat ist semantisch bisher bestenfalls mit meist willkürlich gewählten Schlagwörtern erschlossen.
Für ein gegebenes konkretes Problem setzt das Auffinden und Evaluieren einsetzbarer Lösungsmethoden mithin eine profunde Kenntnis der üblichen Fachterminologie voraus und ist selbst unter Zuhilfenahme syntaxbasierter Volltextsuchwerkzeuge aufwendig.
Inzwischen existieren aber fortgeschrittene Konzepte zur maschinenlesbaren Repräsentation von Wissen auf semantischer Ebene, wie etwa sog. Beschreibungslogiken oder darauf aufbauende sog. *Ontologien* [11, 3, 9]. Eine solche Repräsentationsform ermöglicht eine zielgerichtete Suche mit hoher Spezifität. Weiterhin erlaubt die semantische Repräsentation auch die Anwendung eines automatischen Schließers (engl. „reasoner“) zum Aufdecken innerer Widersprüche.
Allerdings ist festzustellen, dass bisherige Bestrebungen regelungstechnisches Wissen formal abzubilden aus Perspektive der Autoren nicht die erforderliche Funktionaliät und/oder Reichweite erreicht haben, um über interessante „Nischenlösungen“ wie z. B. [2] hinaus eine Wirkung zu erzielen und an der oben beschriebenen Problemlage signifikant etwas zu verändern.

Als potenzielle Alternative stellen wir deshalb mit dem *Methodennetz* und den daraus generierbaren *Aufgabenprozeduren* einen gleichzeitig sowohl anwendungsbezogenen als auch generalistischen Ansatz vor. Das Methodennetz formalisiert hierbei das Wissen darüber, wie typische regelungstechnische Aufgaben gelöst werden, d. h. mit welchen Teilschritten sich gegebene Informationen in gesuchte Informationen überführen lassen.

2 Das Methodennetz

2.1 Aufbau

Ziel des Methodennetzes ist es, Wissen über regelungstechnische Verfahren strukturiert menschen- und maschinenlesbar zu repräsentieren. Für eine konkrete regelungstechnische Aufgabe soll dann mithilfe des Methodennetzes auto-

matisch durch Verknüpfen hinterlegter Lösungsschritte eine Gesamtlösungsstrategie generiert werden. Diese Strategie wird nachfolgend als *Aufgabenprozedur* bezeichnet.
Da das Endergebnis auf eine prozedurale Beschreibung abgebildet wird, ist es naheliegend, zur Repräsentation des Wissens Konzepte der Programmierung zu übernehmen. Beim Entwurf von Software teilt man die Aufgabe typischerweise in Datentypen und Funktionen auf. Dementsprechend unterteilen sich auch die Bestandteile des Methodennetzes in zwei Kategorien: *Typen* und *Methoden*.

2.1.1 Typen

Ein *Typ* im Methodennetz entspricht einem Datentyp in einer konventionellen Programmiersprache, d. h. einer Beschreibung einer Teilmenge aller möglichen Daten, für die bestimmte Operationen möglich sind. Auf die Regelungstechnik bezogen, können dies grundlegende Rechentypen sein wie z. B. `Matrix` oder komplexere mathematische Objekte wie `DGLSystem`, `ZRLTI`[1], `Sprungantwort` oder `Trajektorienfolgeregler`.
Um später Methoden automatisiert miteinander zu verknüpfen, müssen diese ihre Ein- und Ausgänge in der Signatur hinreichend genau spezifizieren können. Somit werden unpraktikable Kombinationen möglichst ausgeschlossen – und zwar auf Typenebene, nicht erst zur Laufzeit. Für ein System in Zustandsraumdarstellung ist es beispielsweise entscheidend, ob dieses steuerbar und beobachtbar ist und wie viele Ein- und Ausgangsgrößen es besitzt. Man kann sich vorstellen, diese Informationen im Namen des Typs abzubilden (`ZRLTIControllableNotObservableSISO`); es wird aber offensichtlich, dass dieser Ansatz mit zunehmender Menge von Eigenschaften „kombinatorisch explodiert“.
Alternativ erlaubt das Methodennetz deshalb, *Parameter* als Teil des Typs zu spezifieren, auf deren Basis Methoden später Beschränkungen definieren können. Eine beispielhafte Typendefinition ist in Listing 1 dargestellt. Diese wird durch ein verschachteltes, assoziatives Array in YAML-Syntax[2] repräsentiert. Die Schlüssel sind durch Doppelpunkte von den Werten abgetrennt und die Verschachtelung durch Einrückungen realisiert. Der Parameter `Zustandsdimension` kann ganzzahlige Werte annehmen, `Normalform` und `Steuerbarkeit` hingegen nur Werte aus einer diskreten Menge von Optionen (sog. Enumeration)[3].
Diese parametrierten Typen sind mit sog. *dependent types* aus Sprachen wie Agda [12] vergleichbar, erlauben aber keine beliebigen Ausdrücke, sondern nur die angesprochenen Ganzzahlen und Enumerationen.

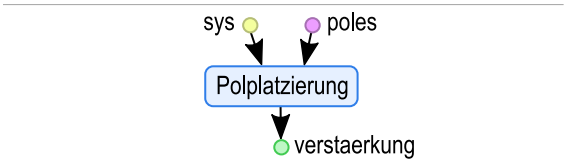
2.1.2 Methoden

Jede *Methode* beschreibt ein regelungstechnisches Verfahren oder einen Teilschritt eines solchen. Neben ihrem Namen spezifiziert eine Methode, welchen Typ die Objekte an den Eingängen besitzen dürfen, und welche Typen sich dafür an den Ausgängen ergeben.
Am Beispiel in Listing 2 wird das Prinzip deutlich. Die Methode besitzt zwei Eingänge mit den Namen `sys` und `poles`[4]. Für diese werden jeweils der erwartete Typ sowie Anforderungen an die Parameter festgelegt. Für die Zustandsdimension und die Anzahl der vorgegebenen Eigenwerte sind keine konkreten Werte vorgegeben, sondern es

```
types:
  ...
  ZRLTI:
    params:
      Zustandsdimension:
        type: Int
      Normalform:
        type: Normalform
      Steuerbarkeit:
        type: Steuerbarkeit
      ...
```

Listing 1 Ausschnitt aus der Definition des Typs `ZRLTI`.

```
methods:
  ...
  Polplatzierung:
    inputs:
      sys:
        type: ZRLTI
        params:
          Zustandsdimension: n
          Steuerbarkeit: <Steuerbar>
      poles:
        type: ListeEigenwerte
        params:
          Laenge: n
        tune: true
    outputs:
      erfolg:
        verstaerkung:
          type: Matrix
          params:
            Zeilen: 1
            Spalten: n
            Rolle: <Zustandsrueckfuehrung>
```

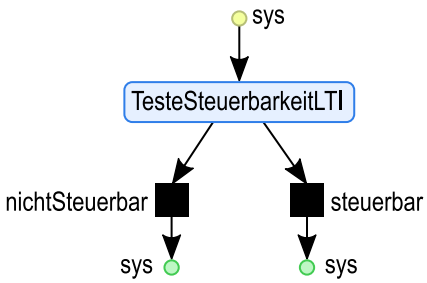


Listing 2 Definition der Methode `Polplatzierung` (YAML-Quelltext und schematische Visualisierung). Diese Methode hat nur eine Ausgangsoption (`erfolg`), vgl. Listing 3.

[1] Lineares zeitinvariantes System in Zustandsraumdarstellung
[2] YAML („Yet Another Markup Language“) ist eine Auszeichnungssprache, zur menschen- und maschinenlesbaren Datenserialisierung.
[3] Erlaubte Werte für diese Enumerationen sind separat definiert.
[4] Dieser Eingang ist als Designparameter (`tune`) markiert, es muss also kein Wert berechnet werden, da dieser vom Nutzer gegeben wird.

wird ein Platzhalter `n` definiert, auf den sich später bezogen werden kann. Dadurch, dass für beide Werte der gleiche Platzhalter verwendet wird, stellt die Methode außerdem sicher, dass Zustandsdimension und Anzahl der Eigenwerte gleich sind. Über den Typ des Ausgangs `verstaerkung` können Aussagen zur Dimension getroffen werden, wobei hier der Wert von `n` wieder verwendet wird.
Eine andere Besonderheit sind Fallunterscheidungen (nicht zu verwechseln mit mehreren simultanen Ausgängen). Wenn erst zur Laufzeit klar werden wird, welchen Typ die Ausgangsobjekte annehmen, kann das über *Ausgangsoptionen* kenntlich gemacht werden. Die Eigenschaft der Steuerbarkeit beispielsweise existiert – wie oben erläutert – auf Typenebene. Eine Methode, die eine Systemdarstellung auf diese Eigenschaft prüft, kann also nicht im Vorhinein wissen, wie das Ergebnis lautet. Listing 3 stellt dar, wie diese Fallunterscheidung abgebildet wird.

```
methods:
  ...
  TesteSteuerbarkeitLTI:
    inputs:
      sys:
        type: ZRLTI
        params:
          Steuerbarkeit: unset
    outputs:
      steuerbar:
        sys:
          type: ZRLTI
          params:
            Steuerbarkeit: <Steuerbar>
      nichtSteuerbar:
        sys:
          type: ZRLTI
          params:
            Steuerbarkeit: <NichtSteuerbar>
```



Listing 3 Definition der Methode `TesteSteuerbarkeitLTI` (YAML-Quelltext und Visualisierung).

Hier werden zwei Ausgangsoptionen definiert: `steuerbar` und `nichtSteuerbar`. Außerdem wird am Eingang mit dem Schlüsselwort `unset` gefordert, dass dieser Parameter noch nicht gesetzt ist. Dies verhindert später in den Lösungsprozeduren, dass die Prüfung unnötigerweise mehrmals durchgeführt wird, obwohl das System sich nicht verändert hat.

2.2 Inhaltlicher Rahmen

Das bis jetzt im Methodennetz dargestellte Wissen erhebt keinen Anspruch auf Vollständigkeit, sondern soll nur einen groben Rahmen vorgeben, an dem sich die zukünftige Erweiterung orientiert. Deshalb beschränkt sich der Inhalt thematisch hauptsächlich auf endlichdimensionale SISO-Systeme und deren Folgeregelung.
Ein Teil der Typen sind verschiedene Systemdarstellungen: `ZRLTI`, `ZRLTV`, `ZRNichtlinear` und `ÜTF`[5]. Eigenschaften wie Steuerbarkeit, mögliche Normalformdarstellungen oder die Zustandsdimension sind dabei schon im Typ als Parameter abgebildet.
Eine weitere Kategorie besteht aus mathematischen Grundtypen wie `Matrix` und `Zeitfunktion`. Um eine sinnvolle, automatische Verkettung von Methoden zu ermöglichen, erweist es sich als hilfreich, deren semantische Bedeutung auch mit im Typ abzubilden. Dazu kann der Parameter `ZeitfunktionRolle` beispielsweise die Werte `Eingangstrajektorie` oder `Zustandsrückführung` (für eine zeitabhängige Reglerverstärkung) annehmen.
Methoden teilen sich im Wesentlichen auf vier Kategorien auf:

1. **Konvertieren**: Methoden zum „verlustfreien“ Überführen zwischen verschiedenen Darstellungen wie `ÜTFzuZR` sowie Approximationen wie `Linearisieren`.
2. **Prüfen**: Tests auf Systemeigenschaften, die dann im Typ abgebildet werden, z. B. `TesteSteuerbarkeitLTI`, `TesteTeilerfremdheitÜTF`.
3. **Entwerfen**: Verfahren wie `Polplatzierung`, `ParametriereZieglerNichols` oder `EKFEntwurf`.
4. **Zusammensetzen**: Methoden, die Objekte zu einem funktionsfähigen Ganzen zusammenfügen, z. B. `BaueTrajektorienfolgeregler`, `BaueLuenbergerBeobachter`.

3 Aufgabenprozeduren

Die zweite Art von Graph – neben dem Methodennetz – ist die *Aufgabenprozedur*. Diese wird für eine konkrete Aufgabe aus dem Methodennetz generiert und ähnelt konzeptionell einem Programmablaufplan der herkömmlichen Programmierung. Sie beschreibt, welche Methoden in welcher Reihenfolge welche Objektinstanzen verarbeiten und welche Zwischenergebnisse damit entstehen.

3.1 Aufbau

Ausgangspunkt jeder Aufgabenprozedur sind

1. Eingangsspezifikation: Beschreibung von *Objekten*, die die gegebenen Informationen repräsentieren, sowie

[5] `ZRLTV` bezeichnet eine lineare zeitvariante Zustandsraumdarstellung. `ZRNichtlinear` meint hier „nicht zwangsläufig linear“; damit ist es die allgemeinste Systemdarstellung. `ÜTF` steht für Übertragungsfunktion.

2. Ausgangsspezifikation: eine Beschreibung der gesuchten Informationen in Form eines Zielobjektes.

Objekte besitzen jeweils einen eindeutigen Bezeichner und einen konkreten Typ, dessen Parameter konkrete Werte enthalten (siehe Abschnitt 2.1.1).
Eine gültige Aufgabenprozedur ist ein gerichteter Graph in dem sich Methodenaufrufe mit (Zwischenergebnis-)Objekten abwechseln. Den Abschluss bildet (mindestens) ein Objekt, was der Zielspezifikation entspricht.
Eine exemplarische Aufgabenprozedur, die die Methoden und Typen des letzten Abschnitts verknüpft, ist in Abbildung 1 dargestellt.

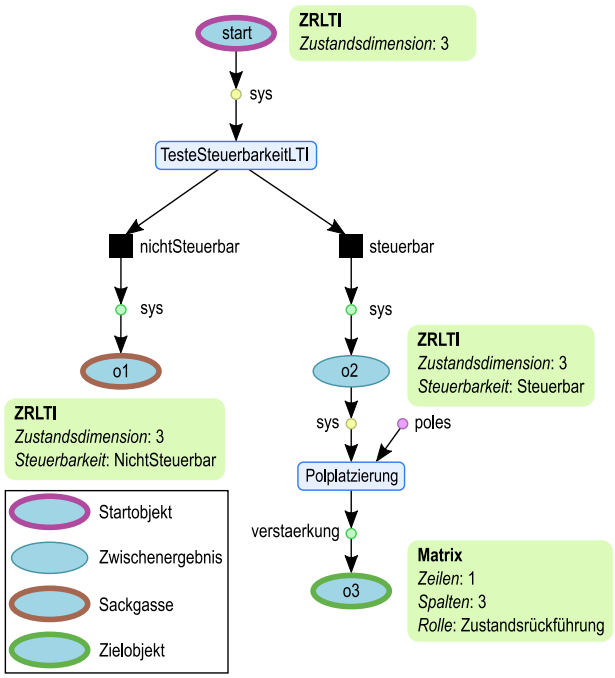


Abbildung 1 Eine beispielhafte Aufgabenprozedur, bestehend aus zwei Methodenaufrufen. Ein LTI-System in Zustandsraumdarstellung wird erst auf Steuerbarkeit überprüft, dann wird eine Rückführung entworfen, die die Pole wie gewünscht platziert. Ellipsen stehen für Objekte, abgerundete Rechtecke für Methodenaufrufe, Kreise für Ein- und Ausgänge, Quadrate für Ausgangsoptionen.

Hervorzuheben ist die zusätzliche Komplexität bei Methoden mit mehreren Ausgangsoptionen. Beim Abarbeiten der Aufgabenprozedur zur Laufzeit wird in diesen Methoden eine Entscheidung getroffen und nur der gewählte *Entscheidungsast*[6] weiter verfolgt. Deshalb dürfen bei einem Methodenaufruf immer nur Objekte verarbeitet werden, die auf dem gleichen Ast von Entscheidungen liegen; siehe Abb. 2.

3.2 Generieren der Aufgabenprozedur

In der aktuellen Implementierung [6] wird eine Aufgabenprozedur aus dem Methodennetz in zwei Schritten erzeugt:

1. Mittels Tiefensuche werden alle möglichen zu den Startobjekten passenden Methodenaufrufe angelegt, die neue Objekte erzeugen. Für neu geöffnete Entscheidungsäste wird diese Suche rekursiv ausgeführt. Die Suche bricht ab, wenn keine neuen Objekte mehr erzeugt werden können.
2. Alle Methodenaufrufe, die nicht zu einem Zielobjekt führen, werden entfernt.

[6] Dieser kann sich durchaus weiter verzweigen.

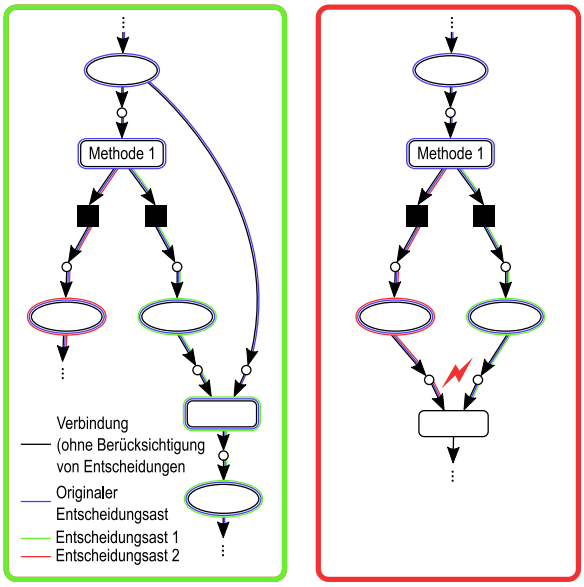


Abbildung 2 Erlaubte und nicht erlaubte Verbindungen. *Alle* Objekte und Methoden liegen auf dem ursprünglichen Entscheidungsast (blau). Methode 1 trifft Entscheidung zwischen zwei Ausgangsoptionen. Dadurch entstehen zwei neue Entscheidungsäste (rot und grün). Diese dürfen nachfolgend nicht kombiniert werden (rechts), Elemente des blauen Asts dürfen in beiden verwendet werden (links).

3.3 Beispiel: Trajektorienfolgeregelung für Dreifachpendel

In [7] wird ein Versuchsstand betrachtet, der aus einem Schlitten, der auf einer Schiene horizontal angetrieben wird und drei passiven Pendelgliedern, die über Lager am Schlitten befestigt sind, besteht, siehe Abb. 4. Weiterhin ist in [8] ein Video zu sehen, in dem das Dreifachpendel vom energieärmsten Zustand (alle Pendelglieder hängen nach unten) in die vollständig aufgerichtete Konfiguration (alle Pendelglieder stehen aufrecht) überführt wird.
Das Anwendungsszenario für das Methodennetz sei nun wie folgt: Ein grundsätzlich mit der Regelungstechnik vertrauter Nutzer möchte die gesehenen Ergebnisse reproduzieren und interessiert sich deshalb für die detaillierten Lösungsschritte. Die Bewegungsgleichungen des Systems werden als gegeben angenommen. Als Suchanfrage an das Methodennetz gibt er also folgende Start- und Zielspezifikation an:

```
start:
  type: DGLSystem
  params:
    Linear: <NichtLinear>
target:
  type: Trajektorienfolgeregler
```

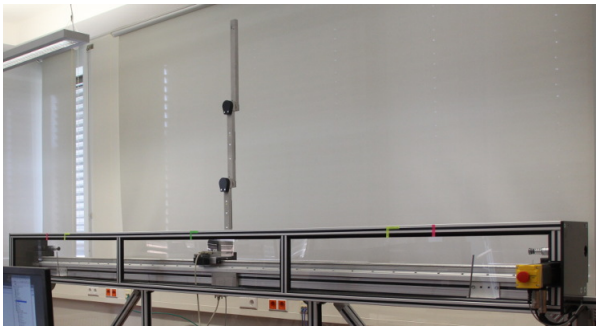


Abbildung 4 Dreifachpendelversuchsstand, bestehend aus aktuiertem Schlitten und drei passiven Pendelgliedern. Das Foto zeigt das System während die obere Ruhelage stabilisiert wird.

Damit wird automatisch eine Aufgabenprozedur generiert, die im rechten Teil von Abb. 3 dargestellt ist.
Es wurde also korrekt geschlossen, dass das System steuerbar[7] und beobachtbar sein muss, eine Trajektorie geplant werden muss, man das System entlang dieser Trajektorie linearisieren kann um einen Folgeregler zu entwerfen, und schlussendlich ein Beobachter benötigt wird. Dass man die Prüfung auf Steuerbarkeit und Beobachtbarkeit vertauschen kann, ist im Suchalgorithmus noch nicht abgebildet, weshalb dieser in der jetzigen Version zwei separate Pfade generiert.
Im Beispielszenario geht für den Nutzer aus dem Ergebnis hervor, dass die meisten Methoden prinzipiell bekannt sind, aber bezüglich der optimierungsbasierten Trajektorienplanung noch Recherchebedarf besteht.

[7] Die Verallgemeinerung von Steuerbarkeit auf Stabilisierbarkeit ist noch nicht im Methodennetz hinterlegt, siehe Abschnitt 2.2.

4 Softwareumsetzung

Zum Erstellen und Erkunden des Methodennetzes, sowie zum Generieren von Aufgabenprozeduren wurde eine prototypische Webanwendung implementiert [5, 6]. Diese basiert auf dem Python-Framework *Django*, sowie der Javascript-Bibliothek *vis.js*. Die Elemente des Methodennetzes sind in YAML spezifiziert. Der Nutzer kann diese im Webbrowser als Graph betrachten und Anfragen wie in Abschnitt 3.3 stellen, woraufhin die entsprechende Aufgabenprozedur generiert wird. Ein Screenshot der Webanwendung ist in Abb. 3 zu sehen.

5 Semantische Interpretation

Das beschriebene Methodennetz repräsentiert regelungstheoretisches Wissen in einer auf einen konkreten Anwendungsfall zugeschnittenen Weise: Für gegebene Ein- und Ausgangsspezifikationen kann eine zielführende Abfolge von Zwischenschritten (die Aufgabenprozedur) bestimmt werden. Um die in der Einleitung skizzierten grundsätzlichen Wissenstransferprobleme mit Hilfe (teil-)automatisierter Assistenzsysteme zu überwinden, ist dies aber noch nicht ausreichend. Beispielsweise ist im Methodennetz die *Bedeutung* der einzelnen Begriffe (z. B. „steuerbar“, „Zustand“) nicht abgebildet. Es ist daher erstrebenswert, das Methodennetz mit einer standardisierten semantischen Wissensrepräsentation zu verknüpfen.
Die gegenwärtig zur formalen Wissenrepräsentation verbreitetste Methode sind sog. semantische Tripel (Subjekt-Prädikat-Objekt) und daraus zusammengesetzte *Ontologi-*

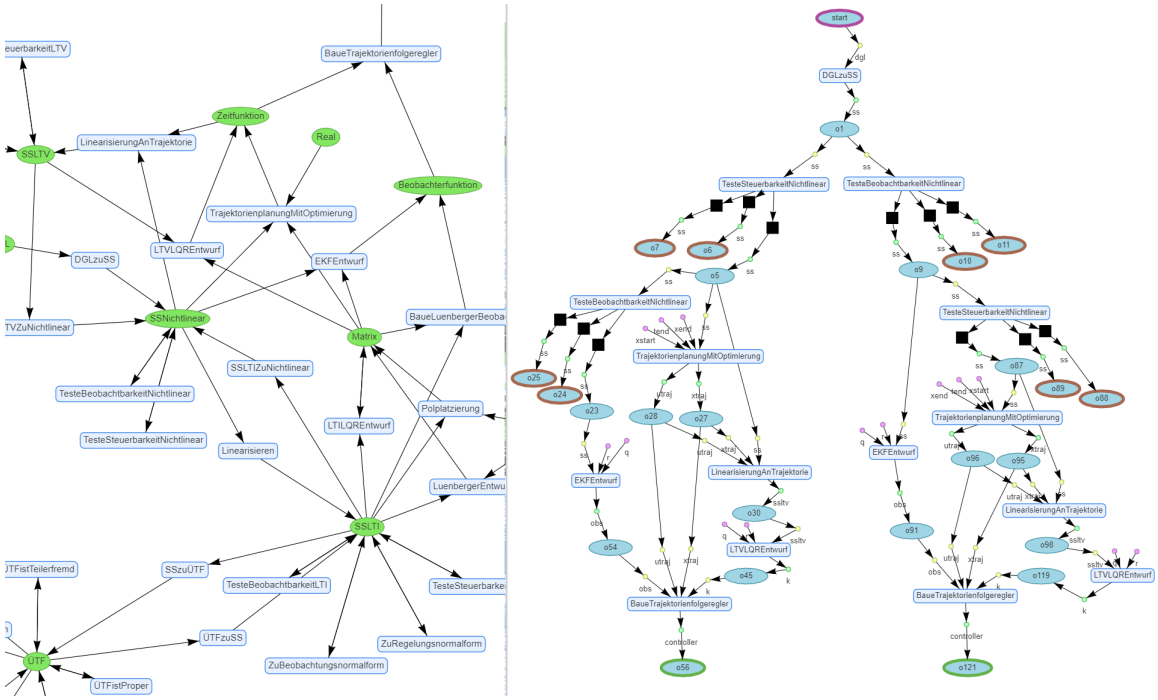


Abbildung 3 Screenshot der Webanwendung [5] (nur Ergebnisbereich). Links: Ausschnitt des Methodennetzes; Rechts: Aufgabenprozedur für Anwendungsbeispiel Dreifachpendelversuchsstand. Unter [6] ist die Abbildung in Originalauflösung verfügbar.

en. Darunter versteht man im informatischen Kontext eine *explizite formale* (d. h. maschinenlesbare) *Spezifikation einer geteilten* (d. h. konsensfähigen) *Konzeptualisierung* (begrifflichen Abdeckung) einer Wissensdomäne [4]. Im engeren Sinn wird damit ein in der *Web Ontology Language (OWL)* repräsentierter Datensatz bezeichnet.
Ontologische Wissenrepräsentation ist bisher insbesondere in der Medizin und Biologie weit verbreitet, siehe z. B. [3]. Für die Ingenieurwissenschaften und speziell für die Regelungstheorie sind den Autoren für den vorliegenden Anwendungsfall keine direkt nutzbaren Ontologien bekannt.

Als Ausgangspunkt für weitere Entwicklungen wird daher eine aus dem Methodennetz automatisch generierte und mit zusätzlichen Informationen angereicherte „Ontologie Regelungstechnischer Methoden“ (ORM) vorgeschlagen. Diese spezialisiert Konzepte aus der Basic Formal Ontology (BFO) [1], was eine zukünftige Interoperabilität mit anderen Ontologien erleichtert. Die ORM erlaubt den Einsatz der standardisierten Abfragesprache SPARQL um konkrete (abgeleitete) Fakten aus der Wissensbasis zu extrahieren, die nur implizit im Methodennetz repräsentiert sind. Als einfaches Demonstrations-Beispiel ist in [6] die Abfrage implementiert, welche der Methoden eine LTI-Zustandsraumdarstellung als Eingang verarbeiten. Auf Basis einer Volltextsuche wäre selbst diese simple Frage nur mit erheblichem manuellem Zusatzaufwand zu beantworten.

6 Zusammenfassung und Ausblick

Das unter [5, 6] verfügbare Methodennetz formalisiert regelungstechnische Verfahren sowohl menschen- als auch maschinenlesbar. Für eine konkrete Problemstellung kann mittels Tiefensuche eine sog. Aufgabenprozedur generiert werden, die dann algorithmisch abarbeitbar ist. Zudem dient das Methodennetz als Basis der Ontologie Regelungstechnischer Methoden (ORM), welche die repräsentierten Verfahren auf semantischer Ebene ergänzt.
Diese formale Repräsentation bietet durch zusätzliche Recherchemöglichkeiten das Potenzial, den inner- unter interdisziplinären Wissenstransfer zu verbessern. Die hier vorgeschlagenen Ansätze sind dabei nicht als fertige Lösungen sondern als Diskussionsbeitrag aufzufassen.

Das Fernziel ist dabei ein Assistenzssystem, das eine Aufgabensituation inhaltlich erfassen, Fragen beantworten und bei Entwurfsentscheidungen und Rechnungen unterstützen kann. Ein plausibler Zwischenschritt dazu ist die Integration der abstrakt repräsentierten Methoden mit den in [10] hinterlegten Beispielimplementierungen.

Danksagung

Die Autoren bedanken sich bei Johannes Bender für die Unterstützung bei der Literaturrecherche.

Literatur

[1] Robert Arp, Barry Smith und Andrew D. Spear. *Building ontologies with basic formal ontology*. MIT Press, 2015.

[2] Carmen Benavides u. a. „An ontology-based approach to knowledge representation for Computer-Aided Control System Design“. In: *Data & Knowledge Engineering* 118 (2018), S. 107–125.

[3] Christophe Dessimoz und Nives Škunca, Hrsg. *The Gene Ontology Handbook*. Springer New York, 2017.

[4] Nicola Guarino, Daniel Oberle und Steffen Staab. „What is an ontology?“ In: *Handbook on ontologies*. Springer Berlin, 2009, S. 1–17.

[5] Robert Heedt und Carsten Knoll. *Demoinstanz des Methodennetzes*. 2021. URL: https://methodnet.ackrep.org/.

[6] Robert Heedt und Carsten Knoll. *Methodennetz-Quellcode-Repositorium*. 2021. URL: https://github.com/TUD-RST/methodnet.

[7] Institut für Regelungs- und Steuerungstheorie, TU Dresden. *Versuchsstände am Institut – Technischer Bericht*. 2020. URL: https://github.com/TUD-RST/technical-reports.

[8] Institut für Regelungs- und Steuerungstheorie, TU Dresden. *Video des Aufschwingvorgangs für den Dreifachpendelversuchsstand*. 2021. URL: https://videocampus.sachsen.de/video/Swing-Up-and-stabilization-of-a-triple-pendulum-on-a-movable-cart/a4b68ff99aed87c9a095f322002ac7b7.

[9] Maria Keet. *An introduction to ontology engineering, v1.5*. College Publications, 2020.

[10] Carsten Knoll und Robert Heedt. „‚Automatic Control Knowledge Repository‘ – A Computational Approach for Simpler and More Robust Reproducibility of Results in Control Theory“. In: *24th International Conference on System Theory, Control and Computing (ICSTCC)*. 2020, S. 130–136.

[11] Jan Lunze. *Künstliche Intelligenz für Ingenieure: Methoden zur Lösung ingenieurtechnischer Probleme mit Hilfe von Regeln, logischen Formeln und Bayesnetzen*. 3. Auflage. Walter de Gruyter Berlin, 2016.

[12] Ulf Norell. „Dependently Typed Programming in Agda“. In: *Proceedings of the 6th International Conference on Advanced Functional Programming*. AFP'08. Springer Berlin, Heidelberg, 2008, S. 230–266.

Einsatz eines detaillierten Modells eines Lenksystems in hochdynamischen HiL-Tests von EPS-Motoren
Application of a detailed model of a steering system in highly dynamic HiL tests of EPS motors

Markus Hemmersbach, M.Sc.; Dipl.-Ing. Hermann Briese; Michael Haßenberg, M.Sc.;
DMecS GmbH & Co. KG, 51105 Köln, Deutschland, info@dmecs.de

Marcus Irmer, M.Sc.; Prof. Dr.-Ing. Hermann Henrichfreise;
Technische Hochschule Köln, Labor für Mechatronik, 50679 Köln, Deutschland, info@clm-online.de

Kurzfassung

Bei der Entwicklung von elektromechanischen Lenksystemen (Electric Power Steering, EPS) sind HiL-Tests ein gängiges Testverfahren. Zur Durchführung der HiL-Tests werden Powerpack-Prüfstände verwendet, mit denen die reale Betriebsumgebung eines EPS-Motors nachgebildet wird.
Der vorliegende Beitrag beschreibt den Einsatz des Powerpack-Prüfstandes aus [1] in hochdynamischen HiL-Tests eines EPS-Motors. Zur möglichst realistischen Nachbildung der Betriebsumgebung des EPS-Motors, wird ein neuartiges, detailliertes Lenkungsmodell [2] als Teilsystem eines Fahrzeugmodells verwendet. Zusammen mit dem Powerpack-Prüfstand und dessen Regelung bildet dieses Modell ein HiL-System, das die Durchführung der HiL-Tests unter Berücksichtigung physikalisch korrekter Parameter aller Teilmodelle ermöglicht. Mit dem Fishhook und dem doppelten Fahrspurwechsel werden zwei standardisierte Fahrversuche im HiL-Test absolviert. Anhand der Messergebnisse wird die Leistungsfähigkeit des HiL-Systems bei hochdynamischer Anregung nachgewiesen.

Abstract

Hardware-in-the-Loop (HiL) tests are common practice in the development process of Electric Power Steering (EPS) systems. The HiL tests are carried out on power pack test benches, which emulate the real environment, that an EPS motor is exposed to in a real vehicle.
This paper describes highly dynamic HiL tests of an EPS motor using the power pack test bench of [1]. To precisely emulate the real environment of the EPS motor, the detailed steering model of [2] is used and moreover embedded in a vehicle dynamics model. The resulting HiL system consists of the detailed steering model, the vehicle dynamics model, the power pack test bench, and a test bench actuator control. The HiL system is used to perform the HiL tests and allows realistic parameter sets for every involved model. During the HiL tests, the fishhook and the double lane-change are performed. Both are standardized road tests, that include highly dynamic excitations of the EPS system. Consequently, the measurement results are used to prove the performance of the power pack test bench in highly dynamic HiL tests.

1 Einleitung

Bei der Entwicklung von Regelungssystemen ist ein Hardware-in-the-Loop Test (HiL-Test) ein etabliertes Verfahren zur Validierung der korrekten Funktion eines geregelten Aktors in einem Labortest. Ziel des HiL-Tests ist es, die realen Betriebslasten des geregelten Aktors mit einem Prüfstand nachzubilden und so eine realistische Testumgebung für den Aktor und dessen Regelung zu schaffen.
Im Bereich elektromechanischer Lenksysteme (Electric Power Steering, EPS) von Kraftfahrzeugen werden seit Jahren elektrische EPS-Motoren als Aktoren eingesetzt. Hierbei hat sich die Bauweise sogenannter Powerpacks etabliert, bei denen der EPS-Motor zusammen mit der EPS-Regelung in einer Komponente (Powerpack) gefertigt wird.

Im Zuge von Entwicklungsarbeiten werden Powerpacks umfangreichen HiL-Tests an Powerpack-Prüfständen unterzogen. Dabei zählen, neben HiL-Tests zur Validierung der EPS-Regelung, zunehmend auch Tests zur Entwicklung von Fahrerassistenzsystemen und autonomem Fahren zum täglichen Einsatzspektrum von Powerpack-Prüfständen.
Die zur Durchführung der HiL-Tests erforderliche Testumgebung, bestehend aus dem Powerpack-Prüfstand und den Modellen zur Nachbildung der realen Betriebsumgebung eines Powerpacks, muss im Wesentlichen zwei Eigenschaften aufweisen:

1. Prüfstandsregelung mit ausreichender Robustheit/Regelgüte/Dynamik
2. Detailliertes Modell der Betriebsumgebung des Powerpacks im realen Fahrzeug

In der praktischen Anwendung besteht die Herausforderung darin, eine Testumgebung bereit zu stellen, die beiden Eigenschaften gleichermaßen gerecht wird.

Nur wenn dies zutrifft kann sichergestellt werden, dass sich das Regelungssystem in Betrieb nehmen lässt und gleichzeitig die Nachbildung der realen Betriebslasten im HiL-Test genau genug ist.
Im vorliegenden Beitrag wird am Beispiel des Powerpack-Prüfstands der Firma (Fa.) DMecS eine HiL-Anwendung gezeigt, die beide Anforderungen erfüllt. Der hierbei verwendete Powerpack-Prüfstand wird in [1] detailliert beschrieben und ist ein eigenständig entwickeltes Produkt der Fa. DMecS, das bei verschiedenen Kunden im Einsatz ist.
Zur Erstellung der HiL-Anwendung wird neben einem zielgerichteten Entwurf der Prüfstandsregelung, eine Einbindung des detaillierten Lenkungsmodells aus [2] in ein Fahrzeugmodell vorgenommen. Die Leistungsfähigkeit der HiL-Anwendung wird abschließend anhand von HiL-Tests zweier hochdynamischer Fahrversuche (Fishhook, doppelter Fahrspurwechsel) gezeigt.

2 Powerpack-Prüfstand

Bild 1 zeigt einen Powerpack-Prüfstand der Fa. DMecS, der zur Erzeugung der Ergebnisse in Kapitel 5 eingesetzt wird.



Bild 1 Powerpack-Prüfstand

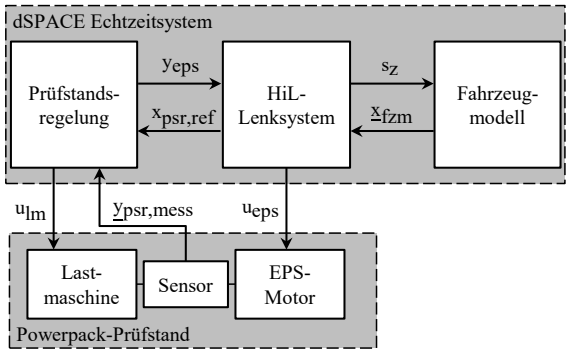
Das Konzept des Powerpack-Prüfstandes resultiert aus der langjährigen Erfahrung von DMecS bei der Entwicklung von EPS-Regelungen, sowie der Weiterentwicklung von Lenkungsprüfständen in diversen Industrieprojekten. Das Anwendungsspektrum des Prüfstandes erstreckt sich von der Systemidentifikation bis hin zu hochdynamischen HiL-Tests von elektrischen Antrieben. Konzeptionell ermöglicht der Prüfstand nicht nur HiL-Tests von Powerpacks aus EPS-Systemen, sondern kann auch zur Entwicklung von elektrischen Antrieben aus anderen Funktionseinheiten eines Fahrzeugs (elektrohydraulisches Bremssystem, Sitzverstellung, Radnabenmotoren, usw.) eingesetzt werden.
Die Funktionsweise des Prüfstandes im HiL-Test lässt sich anhand des Prüfstandsaufbaus erklären. Wie aus Bild 1 ersichtlich, sind die Lastmaschine sowie der EPS-Motor mit jeweils einer Kupplung über eine Sensoreinheit verbunden. Durch diese Verbindung ist es mit der Lastmaschine möglich, Schnittgrößen (Lage, Winkelgeschwindigkeit oder Moment) am EPS-Motor einzuprägen und dadurch die Nachbildung einer realen Betriebssituation im HiL-Test vorzunehmen. Aus Bild 1 wird ferner deutlich, dass anstelle eines Powerpacks aus einem Fahrzeug ein Elektromotor vergleichbarer Größe und Leistungsfähigkeit als EPS-Motor verbaut ist. Da dieser Elektromotor nicht über eine integrierte EPS-Regelung verfügt, wird diese in der HiL-Anwendung (vgl. Bild 2) als Teil des Echtzeitsystems realisiert. Der Einsatz des EPS-Motors anstelle eines Powerpacks ist eine Besonderheit des Prüfstandsaufbaus aus Bild 1, die den Einsatzbereich des Prüfstandes über die Entwicklung und den Test von herstellerspezifischen EPS Funktionen hinaus erweitert.

2.1 Prüfstandsregelung

Die Nachbildung der realen Betriebsumgebung des EPS-Motors wird durch eine hochdynamische Regelung der Lastmaschine des Powerpack-Prüfstands erreicht. Die hierzu verwendete Regelung wird als Prüfstandsregelung bezeichnet und stellt als Regelgröße eine Schnittgröße an der Welle des EPS-Motors ein. Als Schnittgröße wird, u.a. abhängig von der HiL-Anwendung, entweder die Lage, die Winkelgeschwindigkeit oder das Moment am EPS-Motor verwendet.
Neben einer hohen Regelungsbandbreite ist eine hohe Regelgenauigkeit erforderlich. Dies muss trotz Störungen wie Reibung, Drehmomentwelligkeit und Messrauschen realisiert werden. Etwaige Eigenschwingungen des Prüfstandssystems können durch die Prüfstandsregelung aktiv gedämpft werden.
Um diese Ziele zu erreichen, ist die Prüfstandsregelung als beobachter-basierter Zustandsregler [1] implementiert. Zudem wird die Regelgüte durch ein nichtlineares Modell zur Kompensation der Momentenwelligkeit der Lastmaschine verbessert.

3 HiL-Anwendung

In Bild 2 wird eine Übersicht über das HiL-System gegeben, mit dem die HiL-Tests in Kapitel 5 realisiert werden.



Legende:

s_z:	Auslenkung Zahnstange (z)
u_{eps}:	Stellgröße EPS-Motor (eps)
u_{lm} :	Stellgröße Lastmaschine (lm)
$\underline{x}_{fzm}$:	Vektor fahrdynamischer Größen Fahrzeugmodell (fzm)
$x_{psr,ref}$:	Führungsgröße Prüfstandsregelung (psr)
y_{eps}:	Bewegungsgröße EPS-Motor (eps)
$\underline{y}_{psr,mess}$:	Messgrößenvektor Prüfstandsregelung (psr)

Bild 2 HiL-System

Analog zur Grenze zwischen Soft- und Hardware wird in Bild 2 die Unterteilung in „dSPACE Echtzeitsystem“ und „Powerpack-Prüfstand“ vorgenommen. Das Echtzeitsystem umfasst die Subsysteme „Prüfstandsregelung“, „HiL-Lenksystem“ sowie „Fahrzeugmodell“, deren Wechselwirkungen nachfolgend erläutert werden.

Die Prüfstandsregelung (psr) erhält vom HiL-Lenksystem mit dem Signal $x_{psr,ref}$ eine Führungsgröße, die an der Welle des EPS-Motors eingestellt werden soll. Unter Berücksichtigung der Führungsgröße $x_{psr,ref}$ und der Messgrößen im Vektor $\underline{y}_{psr,mess}$ wird von der Prüfstandsregelung die zugehörige Stellgröße u_{lm} der Lastmaschine (lm) ermittelt und angesteuert.

Das HiL-Lenksystem, das in Kapitel 3.1 näher erläutert wird, beinhaltet ein detailliertes Lenkungsmodell mit EPS-Regelung und steuert über die Stellgröße u_{eps} den EPS-Motor an. Von der Prüfstandsregelung erhält das HiL-Lenksystem mit y_{eps} Größen des EPS-Motors zurück.

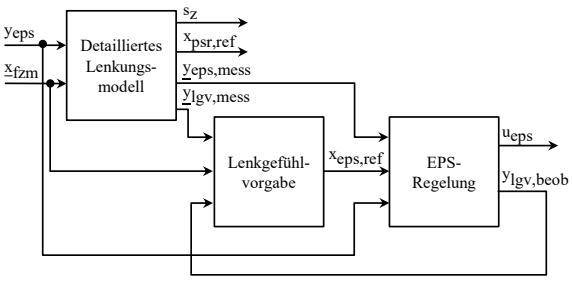
Wie bereits in Kapitel 2 erklärt, ist die Realisierung der EPS-Regelung als Teil des HiL-Lenksystems ein Sonderfall, der aus dem Einsatz eines Elektromotors (EPS-Motor) mit Leistungselektronik anstelle eines Powerpacks resultiert. Wird anstelle des EPS-Motors ein Powerpack am Prüfstand montiert, so erfolgt im HiL-Lenksystem der Ersatz der EPS-Regelung durch eine Sensoremulation für das Powerpack.

Neben der Verbindung mit der Prüfstandsregelung ist das HiL-Lenksystem mit einem Fahrzeugmodell (fzm) gekoppelt. Durch diese Kopplung wird das HiL-Lenksystem als Teil eines gesamten Fahrzeugs modelliert, sodass die Anregungen des HiL-Lenksystems im Vektor $\underline{x}_{fzm}$ jenen in einem realen Fahrzeug entsprechen. Als Fahrzeugmodell dient das „Automotive Simulation Models, ASM“ der Fa. dSPACE, dessen Parametrierung dem Kapitel 3.2 entnommen werden kann. Da das HiL-Lenksystem ein detailliertes Lenkungsmodell enthält, wurde das im ASM vorhandene Lenkungsmodell entfernt und mit der Zahnstangenauslenkung s_Z die vom ASM benötigte Eingangsgröße bereitgestellt.

3.1 HiL-Lenksystem

Das in Bild 3 dargestellte HiL-Lenksystem wird aus einem detaillierten Lenkungsmodell, einer EPS-Regelung sowie einem System zur Lenkgefühlvorgabe gebildet. Bei der EPS-Regelung handelt es sich um eine Handmomenten-Regelung, die in [2] beschrieben wurde. Zur korrekten Funktion der EPS-Regelung muss in Abhängigkeit fahrdynamischer Größen eine Führungsgröße (Soll-Handmoment) bereitgestellt werden, was durch das System zur Lenkgefühlvorgabe aus [3] gewährleistet wird.

Das detaillierte Lenkungsmodell wurde in [2] hergeleitet und ist ein nichtlineares Modell einer elektromechanischen Servolenkung, mit insgesamt acht Freiheitsgraden. Die Parametrierung des detaillierten Lenkungsmodells entspricht den physikalischen Parametern einer Lenkung eines Fahrzeugs aus der Mittelklasse.



Legende:

s_Z :	Auslenkung Zahnstange (z)
u_{eps} :	Stellgröße EPS-Motor (eps)
$x_{eps,ref}$:	Führungsgröße EPS-Regelung (eps)
$\underline{x}_{fzm}$:	Vektor fahrdynamischer Größen Fahrzeugmodell (fzm)
$x_{psr,ref}$:	Führungsgröße Prüfstandsregelung (psr)
y_{eps} :	Bewegungsgröße EPS-Motor (eps)
$\underline{y}_{eps,mess}$:	Messgrößenvektor EPS-Regelung (eps)
$y_{lgv,beob}$:	Beobachtete Eingangsgröße Lenkgefühlvorgabe (lgv)
$\underline{y}_{lgv,mess}$:	Messgrößenvektor Lenkgefühlvorgabe (lgv)

Bild 3 HiL-Lenksystem

Durch die Berücksichtigung aller relevanten Elastizitäten und nichtlinearen Eigenschaften einer realen Servolenkung weist das Modell einen hohen Detaillierungsgrad auf, der die Abbildung sogenannter residualer Eigenbewegungen ermöglicht. Diese Eigenbewegungen bleiben in konventionellen Lenkungsmodellen meist unberücksichtigt, obwohl sie signifikante Auswirkungen auf die Robustheit einer EPS-Regelung haben. [2] Die Verwendung des detaillierten Lenkungsmodells im HiL-Test verbessert deshalb die Nachbildung der realen Betriebsumgebung eines EPS-Motors und steigert dadurch die Genauigkeit der Ergebnisse der HiL-Tests.

3.2 Fahrzeugmodell

Das im HiL-System verwendete Fahrzeugmodell ist das „Vehicle Dynamics Model“ des ASM der Fa. dSPACE. Bei den HiL-Tests in Kapitel 5 wird für das Fahrzeugmodell der Parametersatz „SmallCar“ verwendet, der im Lieferumfang des ASM enthalten ist. Zur Modellierung der Reifenkräfte wird das „Magic Formula Tire Model“ eingesetzt. Neben der Implementierung des Fahrzeugmodells, werden mittels ASM die Experimentdefinitionen sowie die Experimentsteuerung der in Kapitel 4 beschriebenen Fahrversuche vorgenommen.

3.3 Stabilität der Prüfstandsregelung in der HiL-Anwendung

Im Allgemeinen ist zur Durchführung von HiL-Tests ein Regelungssystem in Betrieb zu nehmen, das sehr komplex ist. Die hohe Komplexität entsteht durch die Wechselwirkungen zwischen den beteiligten Teilsystemen, die bereits am HiL-System aus Bild 2 deutlich wurden.

Aus der Praxis ist bekannt, dass sich bei der Inbetriebnahme eines HiL-Systems meist nicht ohne weiteres ein stabiles Systemverhalten einstellt. Liegt ein instabiles System vor, werden häufig die Parameter der

eingesetzten Modelle so angepasst, dass sich ein stabiles Verhalten für das HiL-System ergibt. Dieses Vorgehen ist jedoch zu Lasten der Aussagekraft der HiL-Tests, da zur genauen Nachbildung der realen Betriebsumgebung des EPS-Motors physikalisch korrekte Modelle erforderlich sind.
Um die beschriebene Parameteranpassung zu vermeiden, werden im Vorfeld der Inbetriebnahme des HiL-Systems aus Bild 2 die beteiligten Modelle und Regelungen analysiert. Das Ziel der Analysen besteht darin, Anforderungen an das gewünschte Betriebsverhalten der Prüfstandsregelung abzuleiten. Nachdem diese Anforderungen bekannt sind, erfolgen die Auswahl und der zielgerichtete Entwurf der Prüfstandsregelung, wodurch der Betrieb des HiL-Systems mit physikalisch korrekten Modellparametern ermöglicht wird.

4 Standardisierte Fahrversuche

Mit dem Fishhook und dem doppelten Fahrspurwechsel werden im HiL-Test zwei Fahrversuche nach Industriestandard absolviert. Beide Experimente stellen eine Anregung mit großer Dynamik dar, bei denen sowohl die EPS- als auch die Prüfstandsregelung ihre Leistungsfähigkeit beweisen müssen.
Nachfolgend werden der Versuchsablauf und die Rahmenbedingungen der beiden Fahrversuche erläutert. Dabei liegt der Fokus auf einer kompakten Übersicht, die das Verständnis der Ergebnisse in Kapitel 5 ermöglicht.

4.1 Fishhook (NCAP)

Der Fishhook ist ein Fahrmanöver, das durch das „New Car Assessment Program, NCAP“ standardisiert ist und zur Untersuchung der Kipp-Neigung von Fahrzeugen bei schlagartigen Richtungswechseln dient. [4] Hierbei dient keine reale Fahrsituation als Vorbild, sondern es erfolgt die Anregung mit einem festgelegten Lenkwinkelprofil (vgl. Bild 4), das während der Versuchsdurchführung mit einem Lenkroboter eingeprägt wird.

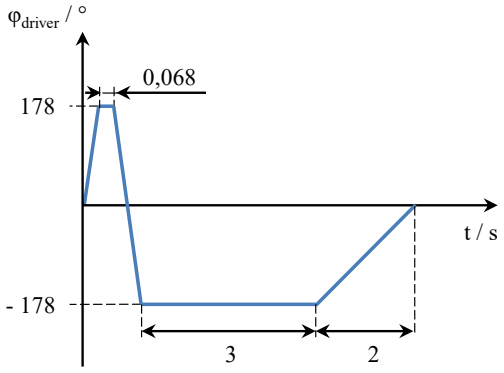


Bild 4 Lenkwinkelprofil beim Fishhook

Charakteristisch für das Lenkwinkelprofil sind zwei schnell aufeinander folgende Richtungswechsel gleicher Amplitude am Beginn des Profils, bei denen der Lenkwinkelgradient 720 °/s beträgt.

Im Vorfeld einer Versuchsdurchführung müssen die quantitativen Größen des Lenkwinkelprofils in einem Vorversuch für das jeweilige Testfahrzeug ermittelt werden. Für das gemäß Kapitel 3.2 parametrierte Fahrzeug, wurde der Vorversuch mit dem ASM simuliert und dadurch das in Bild 4 gezeigte Lenkwinkelprofil bestimmt. Dieses Lenkwinkelprofil wird während des in Kapitel 5.1 beschriebenen HiL-Tests mit einer Anfangsgeschwindigkeit von 70 km/h absolviert.

4.2 Doppelter Fahrspurwechsel (ISO 3888-1)

Der doppelte Fahrspurwechsel (gemäß Norm ISO 3888-1) ist ein Fahrversuch, der ein Ausweichmanöver nachbildet. [5] Der Versuchsablauf sieht hierzu das Durchfahren der in Bild 5 gezeigten Pylonengasse vor, die zwei parallel versetzte Fahrspuren vorgibt.

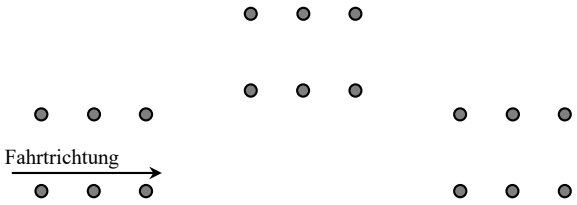


Bild 5 Pylonengasse beim doppelten Fahrspurwechsel

Während der Versuchsdurchführung, des in Kapitel 5.2 beschriebenen HiL-Tests, wird die Einfahrgeschwindigkeit in die Pylonengasse mit 100 km/h vorgegeben und das Fahrzeug vom Fahrermodell des ASM gesteuert.

5 Messergebnisse

Die Messergebnisse der HiL-Tests von Fishhook und doppeltem Fahrspurwechsel werden in diesem Kapitel diskutiert. Der Schwerpunkt liegt dabei auf der Bewertung der Leistungsfähigkeit des Powerpack-Prüfstands, inklusive der Prüfstandsregelung. Auf das Verhalten des Lenksystems, während der HiL-Tests, wird nicht weiter eingegangen.
Die Bewertung des Powerpack-Prüfstandes erfolgt durch einen Vergleich der Messergebnisse aus den HiL-Tests mit Referenzverläufen. Die Referenzverläufe stammen aus einer Simulation der Fahrversuche mit einem separaten Simulationsmodell. Dieses Simulationsmodell dient allein der Erzeugung der Referenzverläufe und beinhaltet das detaillierte Lenkungsmodell, die EPS-Regelung, das System zur Lenkgefühlvorgabe und das Vehicle Dynamics Model des ASM. Demnach kann mit diesem Modell das Verhalten des EPS-Systems im Fahrzeug, ohne das Vorhandensein des Powerpack-Prüfstands, simuliert werden.
Zeigen nun die Messergebnisse aus den HiL-Tests eine große Übereinstimmung mit den Simulationsergebnissen, so bildet der Powerpack-Prüfstand die reale Betriebsumgebung des EPS-Systems ohne Verfälschungen nach.

5.1 Messergebnis Fishhook

In Bild 6 sind die Ergebnisse des HiL-Tests für den Fishhook anhand des Lenkwinkels phi_{driver}, des Momentes T_{tb} im Torsionsstab der Lenkung und des Momentes $T_{eps,mot}$ des EPS-Motors dargestellt. Für die Momente T_{tb} und $T_{eps,mot}$ erfolgt jeweils der Vergleich zwischen der Simulation des EPS-Systems als Teil eines Fahrzeugs (Referenzverlauf ohne Powerpack-Prüfstand) und dem Messergebnis aus dem HiL-Test. Beim Lenkwinkel phi_{driver} wird nicht zwischen Simulation und Messung unterschieden, da in beiden Fällen das in Bild 4 festgelegte Lenkwinkelprofil eingeprägt wird.

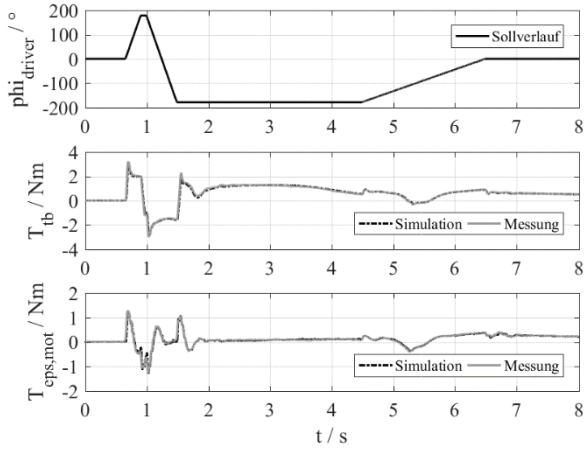


Bild 6 Vergleich von Simulation und Messung für den Fishhook

Der Vergleich von Simulations- und Messergebnissen für die Größen T_{tb} und $T_{eps,mot}$ zeigt über die gesamte Versuchsdauer eine sehr große Übereinstimmung. Daraus folgt, dass der Powerpack-Prüfstand bei HiL-Tests mit hochdynamischer Anregung die reale Betriebsumgebung des EPS-Motors sehr gut nachbildet. Eine Verfälschung der Ergebnisse durch den Powerpack-Prüfstand liegt nicht vor.
Folglich können mit der vorliegenden HiL-Anwendung belastbare Ergebnisse in HiL-Tests erzeugt werden, selbst wenn die Dynamik der Anregung im Grenzbereich der Leistungsfähigkeit eines EPS-Systems liegt.

5.2 Messergebnis doppelter Fahrspurwechsel

Das Messergebnis des HiL-Tests des doppelten Fahrspurwechsels wird in Bild 7 dargestellt. Die hierbei verwendeten Größen $phi_{driver,}$ T_{tb} und $T_{eps,mot}$ entsprechen jenen, die bereits bei der Darstellung der Ergebnisse des Fishhook eingeführt wurden. Im Gegensatz zum Fishhook wird beim doppelten Fahrspurwechsel kein festes Lenkwinkelprofil vorgegeben, sondern der Lenkwinkel vom Fahrermodell des ASM erzeugt. Aus diesem Grund wird in Bild 7 für den Lenkwinkel phi_{driver} ebenfalls zwischen Simulation und Messung unterschieden.

Anhand des Lenkwinkels phi_{driver} lässt sich der Versuchsablauf des doppelten Fahrspurwechsels wie folgt beschreiben. Der doppelte Fahrspurwechsel beginnt mit der Einfahrt in die Pylonengasse zum Zeitpunkt eine Sekunde und endet mit der Ausfahrt aus der Pylonengasse bei sechs Sekunden. Anschließend liegt eine Geradeausfahrt vor. Während der Geradeausfahrt (Zeitraum sechs bis zehn Sekunden) zeigt der Lenkwinkel phi_{driver} eine leichte Schwingung um Null.
Es ist anzunehmen, dass die Ursache dieser Schwingung in der Rückkopplung des detaillierten Lenkungsmodells mit dem Fahrermodell des ASM liegt. Da die Schwingung gleichermaßen in Simulation und Messung vorliegt und darüber hinaus zeitlich nach dem doppelten Fahrspurwechsel auftritt, wird sie in der weiteren Analyse nicht weiter betrachtet.

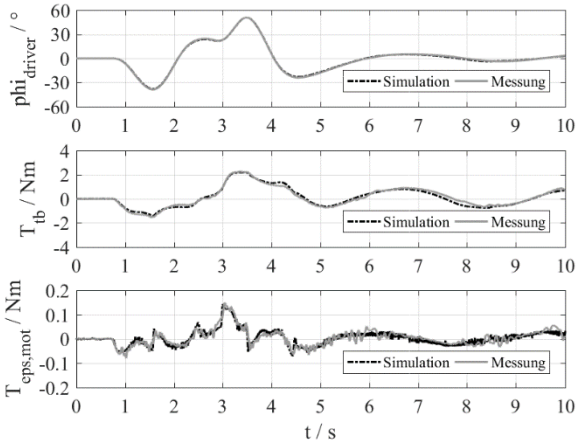


Bild 7 Vergleich von Simulation und Messung für den doppelten Fahrspurwechsel

Wie bereits beim Fishhook, zeigt der Vergleich von Simulations- und Messergebnis der Größen T_{tb} und $T_{eps,mot}$ beim doppelten Fahrspurwechsel eine sehr gute Übereinstimmung. Ein etwaiger Einfluss durch den Powerpack-Prüfstand ist erneut nicht vorhanden.
Im Vergleich zu den Ergebnissen des Fishhook in Bild 6, liegt das Niveau des EPS-Motormoments $T_{eps,mot}$ beim doppelten Fahrspurwechsel in Bild 7 deutlich niedriger. Ein Einfluss auf die Genauigkeit der Nachbildung der realen Betriebsumgebung durch den Powerpack-Prüfstand ist dennoch nicht vorhanden.

6 Zusammenfassung

In diesem Beitrag werden zwei hochdynamische HiL-Tests eines EPS-Motors beschrieben. Zur Durchführung der HiL-Tests wird ein Powerpack-Prüfstand [1] eingesetzt, der als Teil einer komplexen HiL-Anwendung die reale Betriebsumgebung des EPS-Motors detailliert nachbildet. Um den hohen Detaillierungsgrad der Nachbildung zu erreichen sind zwei Aspekte zentral.
Erstens wird ein präzises Modell der realen Betriebsumgebung eines EPS-Motors verwendet. Hierbei kommt ein detailliertes Lenkungsmodell [2] zum Einsatz, das als Teilsystem eines Fahrzeugmodells eingebunden wird.

Das detaillierte Lenkungsmodell umfasst acht Freiheitsgrade und berücksichtigt alle relevanten Elastizitäten und nicht linearen Eigenschaften einer Lenkmechanik. Diese werden in konventionellen Lenkungsmodellen meist vernachlässigt, sodass das detaillierte Lenkungsmodell die Genauigkeit der Nachbildung des realen EPS-Systems im HiL-Test verbessert.
Zweitens erfolgen die bedarfsgerechte Auswahl und der zielgerichtete Entwurf der Prüfstandsregelung. Dadurch wird eine Prüfstandsregelung mit hoher Regelgüte und Dynamik erreicht, die den robusten Betrieb der HiL-Anwendung unter Berücksichtigung realistischer physikalischer Parameter aller Teilmodelle ermöglicht.
Mit dem Fishhook und dem doppelten Fahrspurwechsel werden zwei standardisierte Fahrversuche im HiL-Test nachgestellt. Anhand der Messergebnisse der HiL-Tests wird nachgewiesen, dass durch die vorliegende HiL-Anwendung keine Beeinflussung des zu testenden EPS-Motors vorliegt. Somit können belastbare Ergebnisse in HiL-Tests erzielt werden, bei denen hochdynamische Anregungen für das EPS-System vorliegen.
Das Konzept des Powerpack-Prüfstands der Fa. DMecS ermöglicht neben den HiL-Tests von EPS-Motoren, auch HiL-Tests anderer elektrischer Antriebe, die beispielsweise in elektrohydraulischen Bremssystemen, elektrischen Sitzverstellungen oder Radnabenantrieben verbaut werden. Neben dem Einsatz im HiL-Test, besteht die Möglichkeit den gezeigten Powerpack-Prüfstand zur Systemidentifikation zu nutzen. Das vorgestellte Prüfstandskonzept wird bereits von verschiedenen Kunden im industriellen Umfeld eingesetzt und von der Fa. DMecS stetig weiterentwickelt.

7 Literatur

[1] Briese, H.; Farshizadeh, E.; Oedekoven, S.; Schubert, T.; Henrichfreise, H.: *Realistic Dynamical Testing of EPS Motors and ECUs by Means of a Hardware-in-the-Loop Test Bench*. In: 6th International Munich Chassis Symposium 2015. 16./17. Juni 2015 München. Springer Vieweg, Wiesbaden. ISBN 978-3-658-09711-0. S. 489-506.

[2] Irmer, M.; Haßenberg, M.; Briese, H.; Henrichfreise, H.: *Mechatronic system design for EPS systems with residual modes and variable, nonlinear plant behavior.* In: 20. Internationales Stuttgarter Symposium 2020. 17./18. März 2020 Stuttgart. Springer Vieweg, Wiesbaden. ISBN 978-3-658-29942-2. S. 469-482.

[3] Graßmann, O.; Henrichfreise, H.; Niessen, H.; Hammel, K.v.: *Variable Lenkunterstützung für eine elektromechanische Servolenkung*. In: 23. Tagung „Elektronik im Kraftfahrzeug", Haus der Technik, Stuttgart (2003)

[4] U.S. Department of Transportation, National Highway Traffic Safety Administration: *The Fishhook Maneuver Test Procedure*, Washington DC (2013)

[5] International Organization for Standardization: *ISO 3888-1: Passenger cars – Test track for a severe lane-change manoeuvre Part 1 Double lane-change.* Genf (1999)

Linear-Quadratisch-Gaußsche Positionsregelung des Hand Wheel Aktuators für ein Steer-by-Wire Lenksystem
Linear-Quadratic-Gaussian Position Control of the Hand Wheel Actuator for a Steer-by-Wire Steering System

Robert Gonschorek, M.Eng., ZF Friedrichshafen AG, Düsseldorf, Germany, e-mail: Robert.Gonschorek@ZF.com

Univ.-Prof. Dr.-Ing. Prof. h.c. Dr. h.c. Torsten Bertram, Technische Universität Dortmund, Institut für Roboterforschung, Dortmund, Germany, e-mail: torsten.bertram@tu-dortmund.de

Kurzfassung

Die Steer-by-Wire (SbW) Lenkung ist eine Schlüsseltechnologie für hochautomatisiertes Fahren. Für die automatisierte Fahrzeugquerführung ist neben einer präzisen Positionsregelung des SbW Front Axle Aktuators ebenso die präzise Positionsregelung des SbW Hand Wheel Aktuators wesentliche Voraussetzung. Dieser Beitrag beschreibt die Modellbildung, Systemanalyse, Synthese und Evaluation einer Positionsregelung für den Hand Wheel Aktuator des SbW Lenksystems. Ausgehend von einem nichtlinearen Modell der Regelstrecke wird ein vereinfachtes lineares Systemmodell entwickelt. Dieses bildet die Grundlage für die Analyse der Systemdynamik. Anschließend erfolgt der Entwurf einer Zwei-Freiheitsgrade-Linear Quadratisch Gaußschen Positionsregelung (2DOF LQG Regelung), welche es ermöglicht, das Führungs- und Störverhalten der Regelung unabhängig voneinander einzustellen und die gesamte Systemdynamik frei zu gestalten. Die Leistungsfähigkeit der Regelung wird anhand von Simulationen sowie Messungen am realen System bewertet. Das Regelungskonzept erfüllt die hohen Anforderungen für automatisiertes Fahren und ist neben SbW Lenksystemen ebenso auf elektromechanische Lenksysteme (EPS) applizierbar.

Abstract

The Steer-by-Wire (SbW) steering system is a key technology for highly automated driving. For automated lateral vehicle guidance, the precise position control of the SbW Front Axle Actuator and SbW Hand Wheel Actuator are essential prerequisites. This contribution presents the modeling, system analysis, synthesis, and evaluation of the position control for the Hand Wheel Actuator of a SbW steering system. Based on a nonlinear model of the plant a simplified linear system model is derived. This model yields the basis for the analysis of system dynamics. Thereafter, the synthesis of a Two-Degrees of Freedom-Linear Quadratic Gaussian Position Control (2DOF LQG control) is detailed, which allows the independent design of the closed-loop command and disturbance response as well as an arbitrary design of the entire system dynamics. The control performance is evaluated in simulation and by measurements at the real system. The presented control concept fulfills the high requirements for automated driving and is besides SbW Steering systems also applicable to conventional Electric Power Steering (EPS) systems.

1 Einleitung

Die Steer-by-Wire (SbW) Lenkung ist eine Schlüsseltechnologie für hochautomatisiertes Fahren. Für die automatisierte Fahrzeugquerführung ist neben einer präzisen Positionsregelung des SbW Front Axle Aktuators ebenso die präzise Positionsregelung des SbW Hand Wheel Aktuators wesentliche Voraussetzung. Hierbei sind hohe Anforderungen an die Dynamik und stationäre Genauigkeit der Regelung zu erfüllen, welche eine aktive Kompensation der nichtlinearen Reibung sowie die Berücksichtigung der elastischen Ankopplung des Lenkrads erfordern. Arbeiten von [1] legen nahe, dass der relevante Frequenzbereich für Fahrereingaben bis zu 3 Hz beträgt. Dieser Bereich kann deshalb als Mindestforderung an das Folgeverhalten der Positionsreglung betrachtet werden. Des Weiteren darf die Positionsregelung Fahrereingaben nicht als Störung interpretieren, sondern muss diese tolerieren. Dies schafft die Grundlage für eine kooperative Ausgestaltung des Regelungssystems, welche für zukünftige Fahrautomatisierungssysteme von zentraler Bedeutung ist. Für die Erfüllung der vorstehenden Anforderungen bietet sich der Einsatz einer Zwei-Freiheitsgrade-Linear Quadratisch Gaußschen Positionsregelung (2DOF LQG Regelung) an. Diese ermöglicht es, die Dynamik des Führungsverhaltens und des Störverhaltens unabhängig voneinander zu gestalten, indem eine dynamische Vorsteuerung mit einer beobachterbasierten optimalen Zustandsregelung kombiniert wird. Durch die Einbindung aller Systemzustände in die Regelungsaufgabe wird eine gezielte Gestaltung der gesamten Systemdynamik erreicht.

Dieser Beitrag beschreibt die Positionsregelung des Hand Wheel Aktuators eines SbW Lenksystems. Zunächst werden die Modellbildung und Systemanalyse der Regelstrecke dargelegt. Anschließend erfolgt der systematische Entwurf einer 2DOF LQG Regelung. Nach dem Entwurf der Regelung wird deren Leistungsfähigkeit anhand von Simulationen sowie durch Messungen am realen System evaluiert. Abschließend werden die Ergebnisse des Beitrags zusammengefasst und deren Bedeutung für zukünftige Forschungen und Entwicklungen herausgestellt.

2 Modellbildung und Systemanalyse

Bild 1 zeigt den Prüfstandsaufbau des Hand Wheel Aktuators eines SbW Lenksystems. Dieser besteht im Wesentlichen aus einem permanentmagneterregten Synchronmotor (1), welcher über ein Schneckengetriebe (2) mit der Lenksäule verbunden ist, einem elastischen Torsionsstab (3) und dem Lenkrad (4). Das Torsionsstabmoment sowie der Absolutwinkel der unteren Lenksäule sind Messgrößen des Systems und werden über entsprechende Sensoren erfasst.

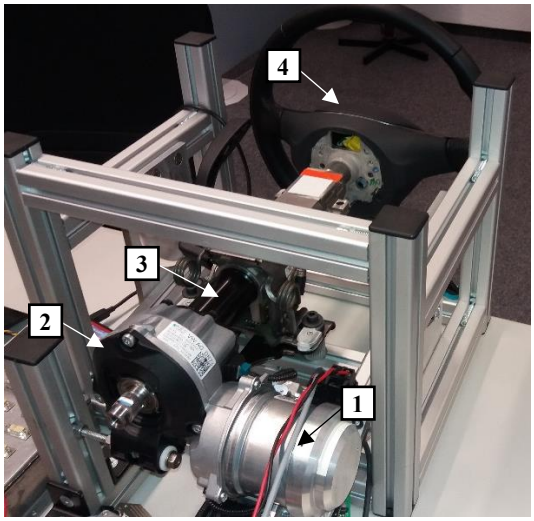


Bild 1 Hand Wheel Aktuator eines SbW Lenksystems

Zur Herleitung eines physikalischen Modells, welches das dynamische Verhalten des Systems hinreichend genau beschreibt, werden die mechanischen Systemkomponenten durch Feder-, Masse-, Dämpfer und Getriebe-Elemente modelliert. Daraus resultiert das folgende Ersatzsystem:

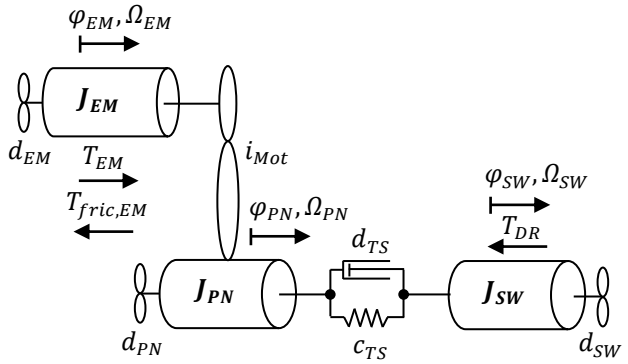


Bild 2 Physikalisches Ersatzsystem des Hand Wheel Aktuators

In diesem Kontext werden die Trägheit des Elektromotors und der Schnecke zu einer Ersatzträgheit (J_{EM}) zusammengefasst, welche über das Schneckenrad starr mit der Trägheit der Lenksäule (J_{PN}) verbunden ist. Die Trägheit der Lenksäule ist mit der Trägheit des Lenkrades (J_{SW}) über den elastischen Drehstab gekoppelt. Zur Modellierung der Elastizität des Drehstabs eignet sich ein Feder-Dämpfer Element mit linearer Steifigkeit (c_{TS}) und linearer Dämpfung (d_{TS}) zur Berücksichtigung der Materialdämpfung.
Die rotatorischen Bewegungen der Trägheiten werden durch die Zustandsgrößen Winkel (φ_i) und Winkelgeschwindigkeit (Ω_i) hinsichtlich inertialer Bezugssysteme beschrieben. Das mechanische Ersatzsystem wird durch das Motormoment (T_{EM}) angetrieben. Das vom Fahrer eingeprägte Handmoment (T_{DR}) sowie die nichtlineare Getriebereibung ($T_{fric,EM}$) wirken als Störgrößen. Zur Berücksichtigung von Lagerreibung werden den Drehmassen viskose Reibelemente zugeordnet (d_{EM}, d_{PN}, d_{SW}).
Durch den Freischnitt der einzelnen Drehmassen, die Erstellung von Freikörperbildern und die Ableitung der Bewegungsgleichungen (Newtonsches Gesetz) können die Differentialgleichungen des Systems bestimmt werden [2]:

$$J_{EM}\dot{\Omega}_{EM} = T_{EM} - d_{EM}\Omega_{EM} - T_{fric,EM} - T_{React} \quad (1)$$

$$J_{PN}\dot{\Omega}_{PN} = T_{React}i_{Mot} + T_{TB} - d_{PN}\Omega_{PN} \quad (2)$$

$$J_{SW}\dot{\Omega}_{SW} = -T_{DR} - T_{TB} - d_{SW}\Omega_{SW} \quad (3)$$

mit:

$$T_{TB} = c_{TS}(\varphi_{SW} - \varphi_{PN}) + d_{TS}(\Omega_{SW} - \Omega_{PN}) \quad (4)$$

Diese lassen sich durch Eliminierung der internen Reaktionsmomente (T_{React}) in Gl. (1) und (2) weiter vereinfachen. Ferner werden die Trägheiten von Motor und Lenksäule zu einer Ersatzträgheit ($J_{\overline{PN}}$) zusammengefasst:

$$J_{\overline{PN}}\dot{\Omega}_{PN} = T_{EM}i_{Mot} + T_{TB} - T_{fric,PN} - d_{\overline{PN}}\Omega_{PN} \quad (5)$$

$$J_{SW}\dot{\Omega}_{SW} = -T_{Dr} - T_{TB} - d_{SW}\Omega_{SW} \quad (6)$$

mit:

$$J_{\overline{PN}} = J_{PN} + i_{Mot}^2 J_{EM} \quad (7)$$

$$d_{\overline{PN}} = d_{PN} + i_{Mot}^2 d_{EM} \quad (8)$$

$$T_{fric,PN} = i_{Mot}T_{fric,EM} \quad (9)$$

Die Dynamik der Stromregelung des Elektromotors wird über ein Verzögerungsglied 1. Ordnung berücksichtigt, wobei ω_{bw} die Bandbreite der Regelung und T_{EM}^* das Motorsollmoment darstellen:

$$\dot{T}_{EM} = -\omega_{bw}T_{EM} + \omega_{bw}T_{EM}^* \quad (10)$$

Zur Herleitung eines linearen Systemmodells aus Gl. (5) bis (10) werden die nichtlinearen Terme ($T_{fric,PN}$, T_{DR}) als unbekannte Störeingänge behandelt. Ferner werden der Winkel φ_{PN}, die Winkelgeschwindigkeit Ω_{PN}, der Differenzwinkel $\Delta\varphi = \varphi_{SW} - \varphi_{PN}$, die Differenzwinkelgeschwindigkeit $\Delta\Omega = \Omega_{SW} - \Omega_{PN}$ und das Motormoment T_{EM} als Zustandsgrößen gewählt, wodurch sich die Gleichungen in die folgende Zustandsraumdarstellung überführen lassen:

$$\underline{\dot{\mathbf{x}}}_\mathbf{p} = \underline{\mathbf{A}}_\mathbf{p}\,\underline{\mathbf{x}}_\mathbf{p} + \underline{\mathbf{B}}_{\mathbf{c,p}}\,\underline{\mathbf{u}}_{\mathbf{c,p}} + \underline{\mathbf{B}}_{\mathbf{d,p}}\,\underline{\mathbf{u}}_{\mathbf{d,p}} \quad (11)$$

$$\underline{\mathbf{y}}_{\mathbf{o,p}} = \underline{\mathbf{C}}_{\mathbf{o,p}}\,\underline{\mathbf{x}}_\mathbf{p}$$

$$\underline{\mathbf{y}}_{\mathbf{m,p}} = \underline{\mathbf{C}}_{\mathbf{m,p}}\,\underline{\mathbf{x}}_\mathbf{p}$$

Mit

$$\underline{\mathbf{A}}_\mathbf{p} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & -\frac{d_{\overline{PN}}}{J_{\overline{PN}}} & \frac{c_{TS}}{J_{\overline{PN}}} & \frac{d_{TS}}{J_{\overline{PN}}} & \frac{i_{Mot}}{J_{\overline{PN}}} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -\frac{d_{SW}}{J_{SW}} + \frac{d_{\overline{PN}}}{J_{\overline{PN}}} & -\frac{c_{TS}}{J_{SW}} - \frac{c_{TS}}{J_{\overline{PN}}} & -\frac{d_{SW}+d_{TS}}{J_{SW}} - \frac{d_{TS}}{J_{\overline{PN}}} & -\frac{i_{Mot}}{J_{\overline{PN}}} \\ 0 & 0 & 0 & 0 & -\omega_{bw} \end{bmatrix}$$

$$\underline{\mathbf{B}}_{\mathbf{c,p}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \omega_{bw} \end{bmatrix}, \underline{\mathbf{B}}_{\mathbf{d,p}} = \begin{bmatrix} 0 & 0 \\ -\frac{1}{J_{\overline{PN}}} & 0 \\ 0 & 0 \\ \frac{1}{J_{\overline{PN}}} & \frac{1}{J_{SW}} \\ 0 & 0 \end{bmatrix}, \underline{\mathbf{x}}_{\mathbf{p}} = \begin{bmatrix} \varphi_{PN} \\ \Omega_{PN} \\ \Delta\varphi \\ \Delta\Omega \\ T_{EM} \end{bmatrix},$$

$$\underline{\mathbf{C}}_{\mathbf{m,p}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_{TS} & 0 & 0 \end{bmatrix}, \ \underline{\mathbf{u}}_{\mathbf{c,p}} = [T_{EM}^*]$$

$$\underline{\mathbf{C}}_{\mathbf{o,p}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \ \underline{\mathbf{u}}_{\mathbf{d,p}} = \begin{bmatrix} T_{fric,PN} \\ T_{DR} \end{bmatrix} \tag{12}$$

Die physikalischen Modellparameter wurden über eine Systemidentifikation bestimmt. Eine Validierung im Frequenzbereich zeigt gute Übereinstimmungen zwischen Modell und realem System für Frequenzen von 1 bis 30 Hz.

Zur Analyse des dynamischen Systemverhaltens werden die Pole und Nullstellen im Übertragungspfad von der Stellgröße (T_{EM}^*) zur Regelgröße (φ_{PN}) betrachtet. **Bild 3** zeigt die Pole und Nullstellen in der s-Ebene.

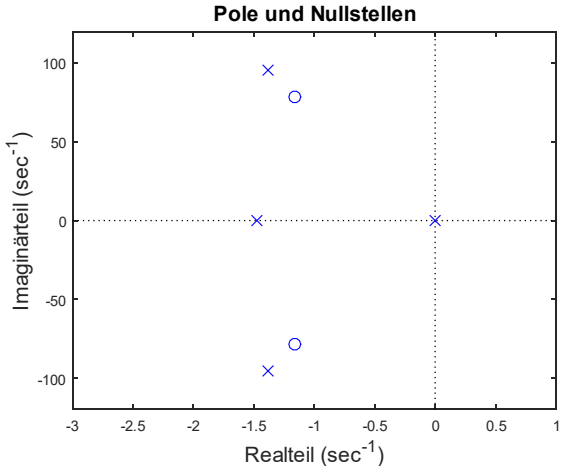


Bild 3 Pole und Nullstellen der Regelstrecke

Das System besitzt einen schnellen reellen Pol weit links, der die Dynamik der Motorregelung beschreibt (nicht dargestellt). Zudem verfügt es über einen Integrator-Pol durch die Integration der Winkelgeschwindigkeit zum Absolutwinkel sowie einen reellen Pol durch die Umsetzung des Motormoments in eine Winkelgeschwindigkeit. Aufgrund der Elastizität des Drehstabs tritt eine schwach gedämpfte, mechanische Resonanz bei ca. 15 Hz auf. Zudem entsteht durch die passive, elastische Kopplung des Lenkrads eine Tilgung bei ca. 12.5 Hz. Dieser Resonanz-Tilgungs-Effekt kann im ungünstigsten Fall durch die Regelung angeregt werden und führt zu unerwünschten Schwingungen in der geregelten Position. Hierdurch wird die erreichbare Bandbreite der Regelung begrenzt. Aus diesem Grund werden im Folgenden Methoden der Zustandsregelung angewendet, welche die mechanische Resonanz aktiv dämpfen und die Dynamik aller Zustände beeinflussen können.

3 Entwurf einer 2DOF LQG Regelung

Ausgehend von dem linearen Modell der Regelstrecke in Zustandsraumdarstellung (11) erfolgt in diesem Abschnitt der Entwurf einer 2DOF LQG Positionsregelung. Deren grundlegende Struktur ist in **Bild 4** dargestellt und resultiert aus der Kombination einer dynamischen Vorsteuerung mit einer LQG Regelung.

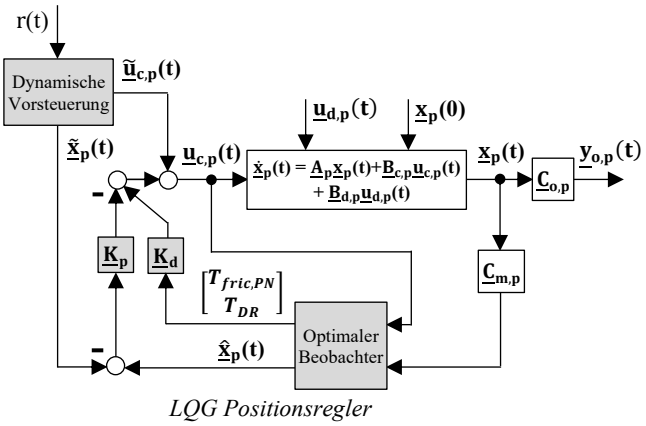


Bild 4 Zwei-Freiheitsgrade-LQG Positionsregelung

Dies hat den Vorteil, dass die Dynamik des Führungs- und des Störverhaltens unabhängig voneinander gestaltet werden können [3]. Zudem ist eine freie Auslegung der gesamten Systemdynamik möglich, da die Regelstrecke vollständig steuerbar ist und Methoden der Zustandsregelung zum Einsatz kommen [4].

Im folgenden Abschnitt wird zunächst der Entwurf einer LQG Regelung beschrieben, welche die Dynamik des Störverhaltens und die Stabilität definiert. Sie kombiniert eine optimale Zustandsregelung mit einem optimalen Zustandsbeobachter. Anschließend wird eine dynamische Vorsteuerung eingeführt, die eine gezielte Gestaltung der Dynamik des Führungsverhaltens ermöglicht. Die Kopplung von Regelung und dynamischer Vorsteuerung resultiert schließlich in der 2DOF LQG Regelung.

3.1 Optimale Zustandsregelung

Für die lineare Regelstrecke mit Zielausgangsgleichung:

$$\begin{aligned} \dot{\underline{\mathbf{x}}}_{\mathbf{p}} &= \underline{\mathbf{A}}_{\mathbf{p}}\,\underline{\mathbf{x}}_{\mathbf{p}} + \underline{\mathbf{B}}_{\mathbf{c,p}}\,\underline{\mathbf{u}}_{\mathbf{c,p}} \qquad \text{mit: } \underline{\mathbf{u}}_{\mathbf{d,p}} = \underline{\mathbf{0}} \\ \underline{\mathbf{y}}_{\mathbf{o,p}} &= \underline{\mathbf{C}}_{\mathbf{o,p}}\,\underline{\mathbf{x}}_{\mathbf{p}} \end{aligned} \tag{13}$$

wird eine optimale Regelung $\underline{\mathbf{u}}_{\mathbf{c,p}}$ gesucht, die den Zustandsvektor $\underline{\mathbf{x}}_{\mathbf{p}}$ von jedem beliebigen Anfangszustand in den Nullzustand, die Ruhelage, so überführt, dass der Wert eines Gütemaßes minimiert wird. Als Gütemaß wird eine quadratische Kostenfunktion J aus den gewichteten Ziel- und Stellgrößen verwendet:

$$J = \int_0^\infty \left[\underline{\mathbf{y}}_{\mathbf{o,p}}^{\mathbf{T}}\underline{\mathbf{Q}}\,\underline{\mathbf{y}}_{\mathbf{o,p}} + \underline{\mathbf{u}}_{\mathbf{c,p}}^{\mathbf{T}}\underline{\mathbf{R}}\,\underline{\mathbf{u}}_{\mathbf{c,p}}\right] \mathrm{d}t \tag{14}$$

Die Gewichtungsmatrizen $\underline{\mathbf{Q}} \geq 0$ und $\underline{\mathbf{R}} > 0$ sind Diagonalmatrizen, deren Elemente hohe Werte der zugehörigen Variablen in der Minimierung der Kostenfunktion bestrafen. Die Zustandsrückführung:

$$\underline{\mathbf{u}}_{\mathbf{c,p}} = -\underline{\mathbf{K}}_{\mathbf{p}}\,\underline{\mathbf{x}}_{\mathbf{p}} \tag{15}$$

ist die optimale Regelung, welche die quadratische Kostenfunktion über einen unendlichen Zeithorizont $[0, \infty]$

minimiert. Der Reglerverstärkungsvektor $\underline{\mathbf{K}}_\mathbf{p}$ folgt aus [5]:

$$\underline{\mathbf{K}}_\mathbf{p} = \underline{\mathbf{R}}^{-1}\underline{\mathbf{B}}_{\mathbf{c,p}}^{\mathbf{T}}\,\underline{\mathbf{S}} \tag{16}$$

mit $\underline{\mathbf{S}}$ der stationären Lösung der Matrix Riccati Gleichung:

$$\underline{\mathbf{A}}_\mathbf{p}^{\mathbf{T}}\,\underline{\mathbf{S}} + \underline{\mathbf{S}}\,\underline{\mathbf{A}}_\mathbf{p} - \underline{\mathbf{S}}\,\underline{\mathbf{B}}_{\mathbf{c,p}}\underline{\mathbf{R}}^{-1}\underline{\mathbf{B}}_{\mathbf{c,p}}^{\mathbf{T}}\,\underline{\mathbf{S}} + \underline{\mathbf{C}}_{\mathbf{o,p}}^{\mathbf{T}}\underline{\mathbf{Q}}\,\underline{\mathbf{C}}_{\mathbf{o,p}} = \underline{\mathbf{0}} \tag{17}$$

Die optimale Zustandsregelung bestimmt die Dynamik des Störverhaltens und die Stabilität.

Zur stationär genauen Störgrößenkompensation wird weiterhin eine Störgrößenaufschaltung hinzugefügt. Das erweiterte Regelungsgesetz ergibt sich damit zu:

$$\underline{\mathbf{u}}_{\mathbf{c,p}} = -\underline{\mathbf{K}}_\mathbf{p}\,\underline{\mathbf{x}}_\mathbf{p} + \underline{\mathbf{K}}_\mathbf{d}\,\underline{\mathbf{x}}_{\mathbf{dist}} \tag{18}$$

und setzt die Messbarkeit oder Schätzbarkeit der Störgrößen $\underline{\mathbf{x}}_{\mathbf{dist}}$ voraus. Die Verstärkungen $\underline{\mathbf{K}}_\mathbf{d}$ der Störgrößenaufschaltung können für die lineare Regelstrecke mit Zielgleichung und Störeingangsvektor $\underline{\mathbf{u}}_{\mathbf{d,p}} \neq \underline{\mathbf{0}}$ analytisch berechnet werden zu [6]:

$$\underline{\mathbf{K}}_\mathbf{d} = -\left[\underline{\boldsymbol{\phi}}\underline{\mathbf{B}}_{\mathbf{c,p}}\right]^{-1}\underline{\boldsymbol{\phi}}\underline{\mathbf{B}}_{\mathbf{d,p}} \tag{19}$$

$$\text{mit: } \underline{\boldsymbol{\phi}} = -\underline{\mathbf{C}}_{\mathbf{o,p}(1,1:5)}\left[\left(\underline{\mathbf{A}}_\mathbf{p} - \underline{\mathbf{B}}_{\mathbf{c,p}}\underline{\mathbf{K}}_\mathbf{p}\right)\right]^{-1}$$

3.2 Optimaler Zustandsbeobachter

Beim optimalen Zustandsreglerentwurf wurde die Messbarkeit der Zustands- und Störgrößen vorausgesetzt. Dies ist in der praktischen Implementierung nicht der Fall. Deshalb wird ein Zustandsbeobachter entworfen, der Schätzungen der Zustandsgrößen bereitstellt und unbekannte Störgrößen rekonstruiert.

Für den Entwurf eines Zustandsbeobachters wird die lineare Regelstrecke mit der Messausgangsgleichung:

$$\begin{aligned}\dot{\underline{\mathbf{x}}}_\mathbf{p} &= \underline{\mathbf{A}}_\mathbf{p}\,\underline{\mathbf{x}}_\mathbf{p} + \underline{\mathbf{B}}_{\mathbf{c,p}}\,\underline{\mathbf{u}}_{\mathbf{c,p}} + \underline{\mathbf{B}}_{\mathbf{d,p}}\,\underline{\mathbf{u}}_{\mathbf{d,p}}\\ \underline{\mathbf{y}}_{\mathbf{m,p}} &= \underline{\mathbf{C}}_{\mathbf{m,p}}\,\underline{\mathbf{x}}_\mathbf{p}\end{aligned} \tag{20}$$

um Integrator-Störmodelle mit unbekanntem Eingang:

$$\begin{aligned}\dot{\underline{\mathbf{x}}}_{\mathbf{dist}} &= \underline{\mathbf{A}}_{\mathbf{dist}}\,\underline{\mathbf{x}}_{\mathbf{dist}} + \underline{\mathbf{B}}_{\mathbf{dist}}\,\underline{\mathbf{u}}_{\mathbf{dist}}\\ \underline{\mathbf{y}}_{\mathbf{dist}} &= \underline{\mathbf{C}}_{\mathbf{dist}}\,\underline{\mathbf{x}}_{\mathbf{dist}}\end{aligned} \tag{21}$$

$$\text{mit: } \underline{\mathbf{A}}_{\mathbf{dist}} = \underline{\mathbf{0}},\ \underline{\mathbf{B}}_{\mathbf{dist}} = \underline{\mathbf{I}},\ \underline{\mathbf{C}}_{\mathbf{dist}} = \underline{\mathbf{I}}$$

erweitert [2]. Aus der Substitution von $\underline{\mathbf{u}}_{\mathbf{d,p}} = \underline{\mathbf{y}}_{\mathbf{dist}}$ folgt die erweiterte Regelstrecke:

$$\begin{bmatrix}\dot{\underline{\mathbf{x}}}_\mathbf{p}\\ \dot{\underline{\mathbf{x}}}_{\mathbf{dist}}\end{bmatrix} = \begin{bmatrix}\underline{\mathbf{A}}_\mathbf{p} & \underline{\mathbf{B}}_{\mathbf{d,p}}\underline{\mathbf{C}}_{\mathbf{dist}}\\ \underline{\mathbf{0}} & \underline{\mathbf{A}}_{\mathbf{dist}}\end{bmatrix}\begin{bmatrix}\underline{\mathbf{x}}_\mathbf{p}\\ \underline{\mathbf{x}}_{\mathbf{dist}}\end{bmatrix} + \begin{bmatrix}\underline{\mathbf{B}}_{\mathbf{c,p}} & \underline{\mathbf{0}}\\ \underline{\mathbf{0}} & \underline{\mathbf{B}}_{\mathbf{dist}}\end{bmatrix}\begin{bmatrix}\underline{\mathbf{u}}_{\mathbf{c,p}}\\ \underline{\mathbf{u}}_{\mathbf{dist}}\end{bmatrix}$$

$$\begin{bmatrix}\underline{\mathbf{y}}_{\mathbf{m,p}}\end{bmatrix} = \begin{bmatrix}\underline{\mathbf{C}}_{\mathbf{m,p}} & \underline{\mathbf{0}}\end{bmatrix}\begin{bmatrix}\underline{\mathbf{x}}_\mathbf{p}\\ \underline{\mathbf{x}}_{\mathbf{dist}}\end{bmatrix} \tag{22}$$

beziehungsweise in kompakter Matrix-/Vektornotation:

$$\begin{aligned}\dot{\underline{\mathbf{x}}}_\mathbf{a} &= \underline{\mathbf{A}}_\mathbf{a}\,\underline{\mathbf{x}}_\mathbf{a} + \underline{\mathbf{B}}_\mathbf{a}\,\underline{\mathbf{u}}_\mathbf{a}\\ \underline{\mathbf{y}}_\mathbf{a} &= \underline{\mathbf{C}}_\mathbf{a}\,\underline{\mathbf{x}}_\mathbf{a}\end{aligned} \tag{23}$$

Basierend auf dem erweiterten Modell der Regelstrecke wird ein optimaler Zustandsbeobachter (Kalman Filter) entworfen. Für das optimale Schätzproblem wird angenommen, dass die erweiterte Regelstrecke durch Prozessrauschen $\underline{\mathbf{w}}$ und Messrauschen $\underline{\mathbf{v}}$ gestört wird. Das Prozessrauschen wird dabei additiv zu den Eingängen angenommen, um die unbekannten Eingänge der Integrator-Störmodelle im Entwurf zu bestrafen sowie als Vorbereitung des Loop Transfer Recovery (LTR)-Verfahrens. Letzteres erreicht über ein Kalman Filter Einstellverfahren die asymptotische Rückgewinnung der Robustheitseigenschaften für vollständige Zustandsrückführung und trägt so zur robusten Auslegung der LQG Regelung bei [7,8].

Für die durch Rauschprozesse gestörte erweiterte Regelstrecke wird eine optimale Zustandsschätzung $\hat{\underline{\mathbf{x}}}_\mathbf{a}(\mathbf{t})$ gesucht, welche den stationären Wert der Summe der Schätzfehlerquadrate minimiert:

$$\mathrm{J} = \lim_{t\to\infty} \mathrm{sp}\left[\mathrm{E}\left\{\left(\underline{\mathbf{x}}_\mathbf{a} - \hat{\underline{\mathbf{x}}}_\mathbf{a}\right)\left(\underline{\mathbf{x}}_\mathbf{a} - \hat{\underline{\mathbf{x}}}_\mathbf{a}\right)^T\right\}\right] = \lim_{t\to\infty} \mathrm{sp}\left[\underline{\mathbf{P}}_\mathbf{e}\right] \tag{24}$$

Die optimale Lösung liefert der Zustandsbeobachter

$$\dot{\hat{\underline{\mathbf{x}}}}_\mathbf{a} = \left(\underline{\mathbf{A}}_\mathbf{a} - \underline{\mathbf{L}}\,\underline{\mathbf{C}}_\mathbf{a}\right)\hat{\underline{\mathbf{x}}}_\mathbf{a} + \underline{\mathbf{B}}_\mathbf{a}\,\underline{\mathbf{u}}_\mathbf{a} + \underline{\mathbf{L}}\,\underline{\mathbf{y}}_\mathbf{a} \tag{25}$$

der die quadratische Kostenfunktion über den unendlichen Zeithorizont $[0,\infty]$ minimiert. Die Beobachterrückführverstärkungen $\underline{\mathbf{L}}$ errechnen sich zu [5]:

$$\underline{\mathbf{L}} = \underline{\mathbf{P}}_\mathbf{e}^*\,\underline{\mathbf{C}}_\mathbf{a}^{\mathbf{T}}\,\underline{\mathbf{V}}^{-1} \tag{26}$$

mit $\underline{\mathbf{P}}_\mathbf{e}^*$ der stationären Lösung der Riccati Gleichung:

$$\underline{\mathbf{A}}_\mathbf{a}\underline{\mathbf{P}}_\mathbf{e}^* + \underline{\mathbf{P}}_\mathbf{e}^*\underline{\mathbf{A}}_\mathbf{a}^{\mathbf{T}} - \underline{\mathbf{P}}_\mathbf{e}^*\underline{\mathbf{C}}_\mathbf{a}^{\mathbf{T}}\underline{\mathbf{V}}^{-1}\underline{\mathbf{C}}_\mathbf{a}\underline{\mathbf{P}}_\mathbf{e}^* + \underline{\mathbf{B}}_\mathbf{a}\underline{\mathbf{W}}\,\underline{\mathbf{B}}_\mathbf{a}^{\mathbf{T}} = \underline{\mathbf{0}} \tag{27}$$

Die geschätzten Zustands- und Störgrößen werden dem Regelgesetz zugeführt. Durch die Einbindung der Integrator-Störmodelle in den Zustandsbeobachter besitzt der resultierende Regler einen integralen Anteil und ist robust gegenüber unbekannten Störungen und Modellfehlern.

3.3 Dynamische Vorsteuerung

Die LQG Regelung definiert die Dynamik des Störverhaltens und die Stabilität des geschlossenen Regelkreises. Um die Dynamik des Führungsverhaltens davon unabhängig zu gestalten, bietet sich der Einsatz einer dynamischen Vorsteuerung an (**Bild 5**).

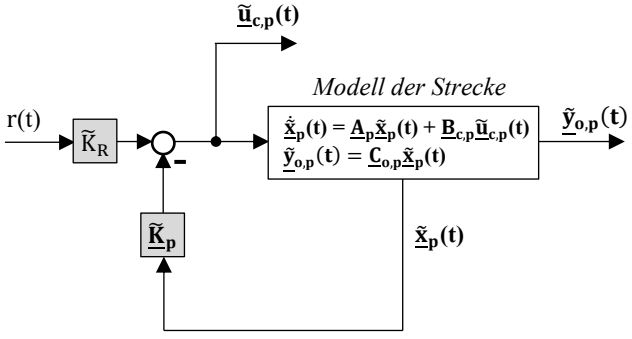


Bild 5 Dynamische Vorsteuerung für Führungsverhalten

Diese beinhaltet einen virtuellen Regelkreis, der online simuliert wird und ein Steuersignal generiert. Das erzeugte Steuersignal wird der realen Strecke zugeführt. Um die Wirkung des Steuersignals auf die Systemzustände vor dem Feedback-Regler zu verstecken, wird der virtuelle Zustandsvektor vom rückgeführten Zustandsvektor subtrahiert

(siehe **Bild 4**). Hierdurch toleriert die Feedbackregelung alle Eingriffe der dynamischen Vorsteuerung und bleibt inaktiv. Der virtuelle Regelkreis wird für ein gewünschtes Führungsverhalten ausgelegt. Hierfür bietet sich der Entwurf einer optimalen Zustandsregelung mit konstantem Vorfilter an, da die Zustandsgrößen des Streckenmodells direkt zugänglich sind [3]. Das Regelgesetz lautet:

$$\underline{\tilde{\mathbf{u}}}_{\mathbf{c,p}} = -\underline{\tilde{\mathbf{K}}}_{\mathbf{p}}\,\underline{\tilde{\mathbf{x}}}_{\mathbf{p}} + \tilde{\mathrm{K}}_{\mathrm{r}}\,\mathrm{r} \qquad (28)$$

Die optimale Rückführverstärkung $\underline{\tilde{\mathbf{K}}}_{\mathbf{p}}$ wird analog zu Kapitel 3.1 durch Minimierung einer quadratischen Gütefunktion bestimmt. Das konstante Vorfilter zur Sicherung der Führungsfolge berechnet sich analytisch zu [5]:

$$\tilde{\mathrm{K}}_{\mathrm{r}} = -\left[\underline{\mathbf{C}}_{\mathbf{o,p(1,1:5)}}\left[\left(\underline{\mathbf{A}}_{\mathbf{p}} - \underline{\mathbf{B}}_{\mathbf{c,p}}\underline{\tilde{\mathbf{K}}}_{\mathbf{p}}\right)\right]^{-1}\underline{\mathbf{B}}_{\mathbf{c,p}}\right]^{-1} \qquad (29)$$

Da es sich bei der dynamischen Vorsteuerung um eine Steuerung handelt, hat diese keinen Einfluss auf die Stabilität.

3.4 Zwei-Freiheitsgrade-LQG Regelung

Durch Kombination der dynamischen Vorsteuerung mit der LQG Regelung resultiert die in **Bild 4** gezeigte 2DOF LQG Positionsregelung.
Bei Abwesenheit von Modellfehlern, Störungen und unterschiedlichen Anfangswerten überträgt sich das Führungsverhalten des virtuellen Regelkreises auf den realen Ausgang. Der Feedbackregler bleibt in diesem Fall inaktiv. Treten jedoch verschiedene Anfangswerte, externe Störungen oder Modellfehler auf, so wird der LQG Regler sofort aktiv und bestimmt das Störverhalten. Somit lässt sich durch die 2DOF Struktur die Dynamik des Führungs- und des Störverhaltens unabhängig voneinander gestalten [6].

4 Evaluation der Regelungsgüte

4.1 Evaluation in der Simulation

Die Bewertung der Regelungsgüte erfolgt zunächst durch eine lineare Analyse des geschlossenen Regelkreises im Zeit- und Frequenzbereich. Hierbei werden Sprungantworten für eine Führungs- und Störanregung betrachtet sowie die Frequenzgänge der Führungsübertragungsfunktion und des offenen Regelkreises analysiert.

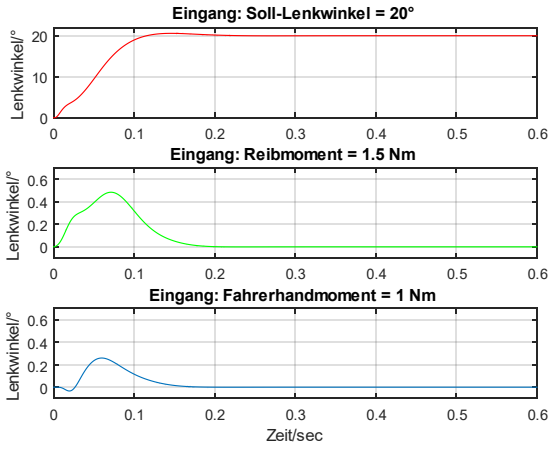


Bild 6 Sprungantworten für Führungs- und Störanregung

Bei einem Führungssprung auf einen Soll-Lenkwinkel von 20° erreicht die Regelung eine Anregelzeit von 0.116 sec, Überschwingweite von 3% und Einschwingzeit von 0.178 sec (siehe **Bild 6**). Die Aufschaltung eines Störsprungs für Reibwertstörungen von 1.5 Nm zeigt einen maximalen Lenkwinkelfehler von 0.5°, der nach ca. 0.17 sec stationär genau kompensiert ist. Sprungförmige Störungen des Fahrerhandmoments von 1 Nm bewirken einen Lenkwinkelfehler von 0.26°, welcher ebenfalls nach ca. 0.17 sec ausgeregelt ist. Die Regelung zeigt somit eine hohe Regelungsgüte und ist zudem robust gegenüber Störungen.
Zusätzlich zur Analyse der Zeitantworten werden die Frequenzgänge des offenen und geschlossenen Regelkreises analysiert (**Bild 7, 8**).

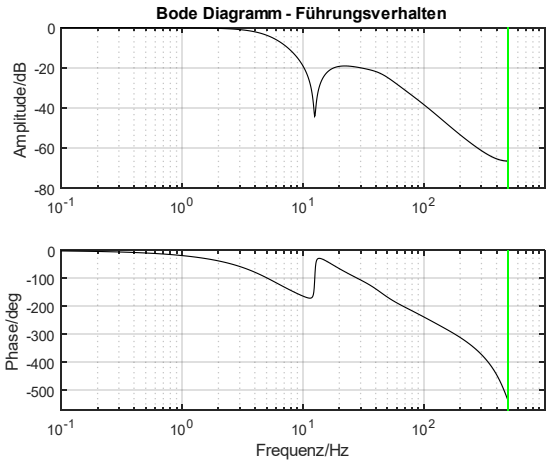


Bild 7 Frequenzgang des geschlossenen Regelkreises

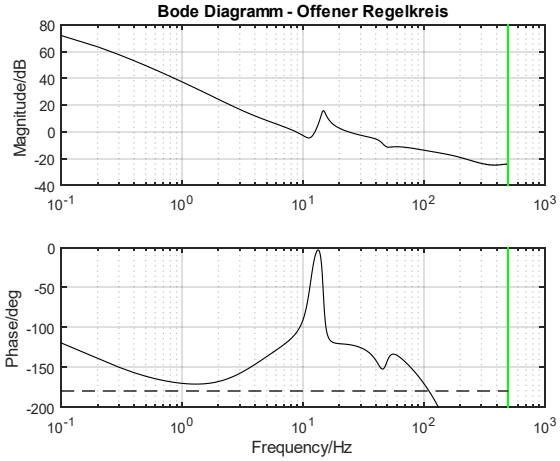


Bild 8 Frequenzgang des offenen Regelkreises

Hierbei zeigt die entworfene Regelung eine Bandbreite von 4.6 Hz und weist damit eine gute Dynamik des Führungsverhaltens auf. Weiterhin wird der Frequenzgang des offenen Regelkreises zur Beurteilung der Stabilität betrachtet. Durch die Anwendung des LTR Verfahrens werden gute Stabilitätsreserven erreicht. Der Regelkreis verfügt über eine Amplitudenreserve von 15 dB und eine Phasenreserve von 58°. Zur genaueren Untersuchung des Einflusses von Nichtlinearitäten der Regelstrecke auf die Regelungsgüte wird die Regelung in der numerischen Simulation bewertet. Hierbei werden Effekte wie Aktuatorbegrenzungen,

nichtlineare Reibung, Abtast- und Haltevorgänge, Signalquantisierung, und Streckenparameteränderungen detailliert analysiert. Die Regelung zeigt sich robust gegenüber den aufgeführten Einflüssen im gesamten Betriebsbereich.

4.2 Messungen am realen System

Die Ergebnisse aus der numerischen Simulation werden durch experimentelle Messungen am realen System bestätigt. **Bild 9** zeigt das Regelverhalten bei einer sinusförmigen Solllenkwinkelvorgabe mit einer Amplitude von 45° und einer Frequenz von 1 Hz. Die Regelung erreicht durch ihre 2DOF Struktur eine dynamische Sollwertfolge sowie eine stationär genaue Störunterdrückung. Sie folgt der Führungsgröße unabhängig von Störungen durch nichtlineare Reibung und unsichere Streckenparameter gut. Dabei werden auch bei dynamischen Solllenkwinkelvorgaben stets ruhige und schwingungsfreie Lenkradbewegungen erreicht.

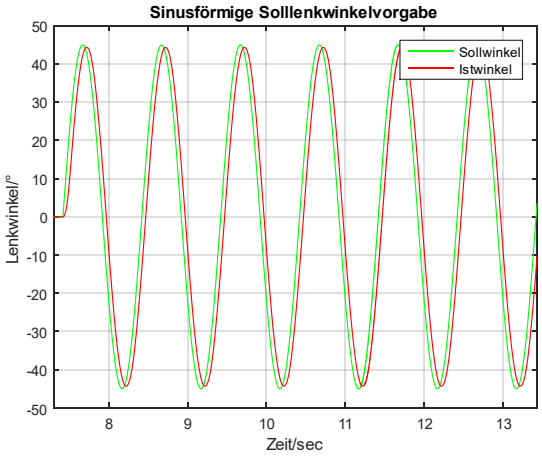


Bild 9 Regelverhalten bei sinusförmiger Solllenkwinkelvorgabe

Das Verhalten ist auf die direkte Modellierung der elastischen Kopplung des Lenkrads und deren gezielte Berücksichtigung im optimalen Reglerentwurf zurückzuführen. Die entworfene Regelung kann hierdurch eine aktive Dämpfung von Drehstabschwingungen aufgrund der mechanischen Resonanz des Systems realisieren.

Zudem bietet die Modellierung den Vorteil, separate Störmodelle für Lenkrad und Ritzel ansetzen zu können, wodurch eine getrennte Schätzung des Fahrerhandmoments und der nichtlinearen Reibung ermöglicht wird. Hierdurch kann eine selektive Störgrößenkompensation erreicht werden. Dies bedeutet, dass der Integralanteil der Regelung Fahrereingriffe akzeptiert, während Störungen durch nichtlineare Reibung stationär genau kompensiert werden. **Bild 10** zeigt hierzu eine Messung am realen System mit Fahrereingriff. Ohne selektive Störgrößenkompensation ($t < 15\text{sec}$) wird die Wirkung des Fahrerhandmoments als Störung interpretiert und durch den I-Anteil der Regelung kompensiert. Mit selektiver Störgrößenkompensation hingegen ($t \geq 15\text{sec}$) wird einzig die nichtlineare Reibung durch den I-Anteil der Regelung kompensiert während Fahrereingriffe zugelassen werden. Ein derartiges Verhalten schafft die Grundlage für eine kooperative Ausprägung des Regelungssystems, welche für die Interaktion zwischen Fahrer und Automatisierungssystem zukünftiger Fahrzeuge von zentraler Bedeutung ist.

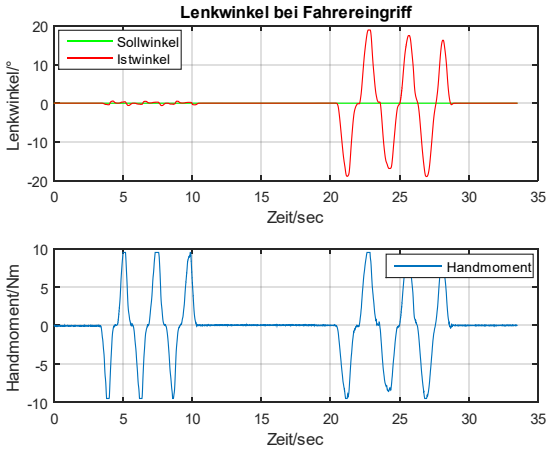


Bild 10 Lenkwinkel bei Fahrereingriff ohne ($t < 15\text{sec}$) und mit ($t \geq 15\text{sec}$) selektiver Störgrößenkompensation

5 Zusammenfassung

In diesem Beitrag wurde eine 2DOF LQG Positionsregelung für den Hand Wheel Aktuator eines SbW Lenksystems vorgestellt. Die Regelung erlaubt eine getrennte Gestaltung der Dynamik des Führungs- und des Störverhaltens. Durch die Einbindung von Störmodellen in den Beobachterentwurf wird darüber hinaus eine getrennte Schätzung des Fahrerhandmoments und der nichtlinearen Reibung ermöglicht. Hierdurch werden Reibstörungen durch den I-Anteil der Regelung robust kompensiert während Fahrerhandmomente toleriert werden. Dies bietet die Möglichkeit einer kooperativen Ausgestaltung der Regelung. Das Regelungskonzept ist neben Steer-by-Wire auch auf elektromechanische Lenksysteme applizierbar.

6 Literatur

[1] Pfeffer, P.; Harrer, M.: *Lenkungshandbuch*. Wiesbaden: Springer Vieweg Verlag, 2013.

[2] Franklin, G. F.; Powell, J. D.; Emami-Naeini, A.: *Feedback Control of Dynamic Systems*. New Jersey: Pearson Prentice Hall, 2018.

[3] Deutscher J.: *Zustandsregelung verteilt-parametrischer Systeme*, Heidelberg: Springer-Verlag, 2012.

[4] Lohmann, B.: *Entwurf im Zustandsraum – alle Systemgrößen einbeziehen*. In: Skolaut, W. (Hrsg.), Maschinenbau. Heidelberg: Springer-Verlag, 2014.

[5] Föllinger, O.: *Regelungstechnik – Einführung in die Methoden und ihre Anwendung*. Berlin: VDE, 2016

[6] Roppenecker, G.: *Zustandsregelung linearer Systeme*. Automatisierungstechnik, B. 57, H. 10, 2009.

[7] Skogestad, S.; Postlethwaite, I.: *Multivariable Feedback Control: Analysis and Design*. Chichester: John Wiley & Sons Inc., 2005.

[8] Kwakernaak, H.; Bosgra, O. H.; Meinsma, G.: *Design Methods for Control Systems*. Enschede: University of Twente, 2005.

Robuste Lenkgefühlvorgabe für elektromechanische Servolenkungen
Robust generation of a steering feel for electric power steering systems

Marcus Irmer, M.Sc.; Prof. Dr.-Ing. Hermann Henrichfreise,
Technische Hochschule Köln, Labor für Mechatronik, 50679 Köln, Deutschland, info@clm-online.de

Kurzfassung

Bei der Regelung elektromechanischer Servolenkungen besteht aufgrund von vernachlässigten Freiheitsgraden und nichtlinearen Eigenschaften der Regelstrecke innerhalb des Regelungsentwurfs weiterhin Handlungsbedarf für eine robuste Realisierung. Daher wird in diesem Beitrag eine Lenkungsregelung vorgestellt, welche bestehende Robustheitsprobleme löst und gleichzeitig eine nahezu beliebige Lenkgefühlvorgabe ermöglicht. Dazu wird auf Basis eines detaillierten nichtlinearen Modells einer elektromechanischen Servolenkung ein reduziertes linearisiertes Modell für den Entwurf eines optimalen Zustandsreglers und eines optimalen Zustandsbeobachters hergeleitet. Ferner wird ein Modell der Lenkgefühlvorgabe vorgestellt. Es ermittelt das Sollhandmoment, dass die fahrzeugführende Person am Lenkrad spüren soll. Dieses Modell der Lenkgefühlvorgabe ermöglicht es, dass der fahrzeugführenden Person ein weitgehend beliebiges Lenkgefühl vermittelt werden kann. Dabei zeigt das entstehende Regelungssystem sehr gute dynamische Eigenschaften und sehr gute Robustheitseigenschaften.

Abstract

The control of the driver's steering torque of electromechanical power steering systems is state of the art. However, due to nonlinear characteristics and degrees of freedom of the plant which are unconsidered in the control design, the challenge still is the robust implementation of this control approach. Therefore, this paper presents a control approach that solves the robustness problems and allows a nearly unconstrained design of the steering feel. Based on a detailed nonlinear model of an electromechanical power steering, a reduced linearized model for the design of an optimal state space controller and an optimal state space observer is derived. Furthermore, a feeling system is presented. It computes the requested steering torque that the driver should feel at the steering wheel. This feeling system allows that the driver can experience an almost freely configurable steering feel. In addition, the resulting control system shows very good dynamic characteristics and very good robustness characteristics.

1 Einleitung

Bisherige Lenkungsregelungen weisen wie in [1] gezeigt oft eine begrenzte Robustheit gegenüber im Regelungsentwurf nicht berücksichtigten Freiheitsgraden und nichtlinearen Eigenschaften der Regelstrecke auf. Daher wird in diesem Beitrag eine robuste Regelung des sogenannten Fahrerhandmoments einer elektromechanischen Servolenkung (Electric Power Steering, EPS) beschrieben. Ferner wird für diese Regelung ein Modell zur robusten Lenkgefühlvorgabe vorgestellt, wodurch der fahrzeugführenden Person ein weitgehend beliebiges Lenkgefühl vermittelt werden kann.
Für den Regelungsentwurf wird ein neuartiges Modell einer elektromechanischen Servolenkung verwendet. Dieses resultiert aus einem detaillierten nichtlinearen Modell einer elektromechanischen Servolenkung, dass alle relevanten Freiheitsgrade und nichtlinearen Eigenschaften berücksichtigt, die in einer realen Lenkung auftreten können. Dieses detaillierte Modell beschreibt die reale Regelstrecke ausreichend genau. Es wird in Kapitel 2 gezeigt. Der Entwurf eines dynamischen Kompensators für diese Regelstrecke wird daraufhin in Kapitel 3 beschrieben. Der Kompensator entsteht aus der Verkopplung eines optimalen Zustandsreglers mit einem optimalen Zustandsbeobachter. Die Ergebnisse der Systemanalyse des Regelungssystems bestehend aus detailliertem Modell der elektromechanischen Servolenkung und Kompensator sind anschließend in Kapitel 4 abgebildet. In Kapitel 5 wird darüber hinaus das Modell der Lenkgefühlvorgabe vorgestellt. Im folgenden Kapitel 6 und in Kapitel 7 werden dann Analysen des um das Modell der Lenkgefühlvorgabe erweiterten Regelungssystems durchgeführt. Das Strukturbild des so entstehenden Regelungssystems ist in Bild 1 dargestellt. In Kapitel 8 wird schließlich eine Zusammenfassung gegeben.

2 Modell der elektromechanischen Servolenkung

Das detaillierte nichtlineare Modell der elektromechanischen Servolenkung besitzt acht Freiheitsgrade und besteht aus einem Modell der Lenkmechanik sowie einem Modell eines stromgeregelten EPS-Motors. Das physikalische Ersatzmodell der Lenkmechanik ist in Bild 2 abgebildet. Dort sind die Freiheitsgrade des

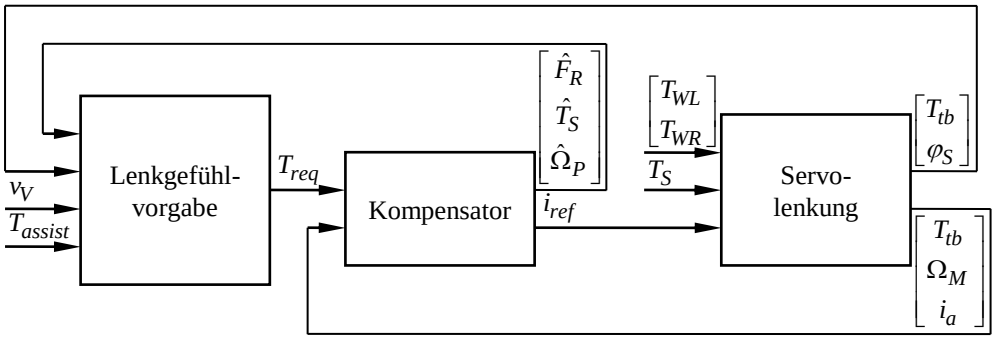


Bild 1 Strukturbild des Regelungssystems bestehend aus elektromechanischer Servolenkung, dynamischem Kompensator und Lenkgefühlvorgabe

Mehrkörpersystems mit Pfeilen dargestellt sowie dessen Körper mit den Bezeichnern S (steering, Lenkrad), P (pinion, Ritzel), R (rack, Zahnstange), N (nut, Spindelmutter), M (motor, EPS-Motor), C (casing, Lenkungsgehäuse), V (vehicle, Fahrzeug), WL (wheel left, linkes Vorderrad) und WR (wheel right, rechtes Vorderrad) versehen. Der Torsionsstab zwischen Lenkrad und Ritzel wird darüber hinaus durch die Steifigkeit c_{tb} und die Dämpfungskonstante b_{tb} modelliert.

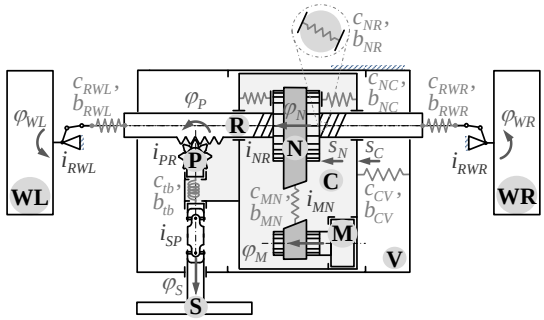


Bild 2 Physikalisches Ersatzmodell der Lenkmechanik

Die Körper sind durch weitere viskoelastischen Elemente und Getriebe miteinander gekoppelt. Diese werden mit den Bezeichnern c_{xy} sowie b_{xy} und i_{xy} dargestellt. Hierbei geben die Indizes x und y an, zwischen welchen beiden Körpern das viskoelastische Element oder das Getriebe platziert ist.

Die viskoelastischen Elemente und Getriebe besitzen zum Teil nichtlineare Eigenschaften. Zudem berücksichtigt das Modell der elektromechanischen Servolenkung weitere nichtlineare Eigenschaften, wie lastabhängige Reibung, mechanische Begrenzungen und Quantisierung, sowie Prozess- und Messrauschen. Damit geht dieses Modell wie in [1] beschrieben über die einfachen Modelle aus anderen Veröffentlichungen hinaus.

Die ausführliche Modellbildung und -analyse ist in [1-4] zu finden. Für dieses Modell der Regelstrecke wird im Folgenden ein dynamischer Kompensator entworfen.

3 Regelungsentwurf

Für die Regelung der elektromechanischen Servolenkung wird eine sogenannte Handmomentenregelung in Form einer LQG/LTR-Regelung realisiert. Dazu wird ein dynamischer Kompensator verwendet. Dieser besteht aus einem linearen statischen optimalen Zustandsregler (LQR) mit Führungs- und Störgrößenaufschaltung sowie einem linearen optimalen Zustandsbeobachter (LQE) mit Störgrößenschätzung. Der Kompensator ermöglicht dadurch eine aktive Schwingungsdämpfung und Störgrößenkompensation.

Grundlage für den Regler- und Beobachterentwurf stellt das nach [1,2] reduzierte linearisierte Modell der elektromechanischen Servolenkung dar. Es entspricht einem mechanischen Modell mit drei Freiheitsgraden, welches die dominantesten Eigenbewegungen der elektromechanischen Servolenkung abbildet. Dieses reduzierte Modell folgt durch modale Ordnungsreduktion aus dem detaillierten Modell der elektromechanischen Servolenkung mit acht Freiheitsgraden aus Kapitel 2. Dabei wurde das Reduktionsverfahren dahingehend modifiziert, dass das entstehende Regelungssystem bezüglich seiner Robustheit im Vergleich zu anderen Veröffentlichungen wesentlich verbessert wurde.

Das Regelungskonzept basiert auf [5-7]. Die Vorgehensweise des Entwurfs ist [1,5] zu entnehmen.

4 Systemanalyse

Das Verhalten der Regelung kann durch eine Analyse im Zeit- und Frequenzbereich mit linearen Modellen überprüft werden. Dazu wird zunächst das Regelungssystem bestehend aus dem linearisierten detaillierten Modell der elektromechanischen Servolenkung und dem dynamischen Kompensator, welcher mit dem zuvor erwähnten reduzierten Modell der elektromechanischen Servolenkung entworfen wurde, betrachtet.

Das Torsionsstabmoment T_{tb} repräsentiert die Regelgröße sowie das Sollhandmoment T_{req} die Führungsgröße des Regelungssystems. Der Sollstrom i_{ref} für den stromgeregelten EPS-Motor stellt die Steuereingangsgröße dar. Die Störgrößen der Regelstrecke werden durch das Handmoment T_S, welches die fahrzeugführende Person am Lenkrad einleitet, sowie durch die

Störmomente T_{WL} und T_{WR} um die Lenkachsen der Vorderräder infolge des Kontakts zwischen Reifen und Fahrbahn charakterisiert. Die Messgrößen der Regelstrecke sind das Torsionsstabmoment T_{tb}, die Winkelgeschwindigkeit Ω_M und der Strom i_a des EPS-Motors sowie der Lenkradwinkel φ_S (siehe Bild 1).
Bild 3 zeigt die Führungssprungantwort des geschlossenen Regelkreises für ein Sollhandmoment T_{req} von 1 Nm sowie die Störsprungantworten für ein tatsächliches Handmoment T_S von 1 Nm und ein Störmoment T_{WL} um die Lenkachse des linken Vorderrades von 50 Nm. Die Sprungantworten stellen dar, dass das Regelungssystem bei sprungförmiger Führungs- oder Störanregung stationäre Genauigkeit aufweist. Zudem sind die Überschwingweiten und Amplituden der Sprungantworten gering sowie die Ausregelzeiten hinreichend klein. Das Regelungssystem besitzt somit ein gutes dynamisches Verhalten.

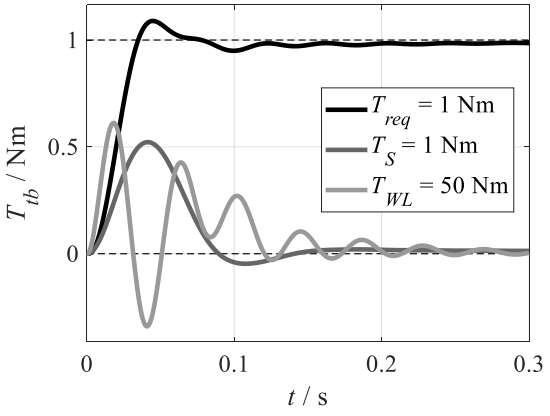


Bild 3 Sprungantworten des geschlossenen Regelkreises

In Bild 4 sind die Frequenzgangsortskurven des offenen Regelkreises abgebildet, wenn die Massenträgheitsmomente der Räder (J_{min}/J_{max}), alle Steifigkeiten (c_{min}/c_{max}) und die Übersetzungsverhältnisse von Spurstange und Spurhebel (i_{med}/i_{max}) innerhalb der Regelstrecke wie in [1,2] variiert werden. Der offene Regelkreis entsteht hierbei durch Schnitt des geschlossenen Regelkreises am Steuereingang i_{ref} der Regelstrecke.

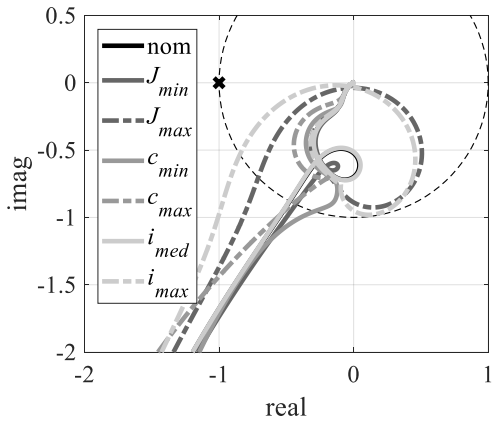


Bild 4 Frequenzgangsortskurven des offenen Regelkreises bei Parametervariationen innerhalb der Regelstrecke

Für das vorliegende System mit nur einer Steuereingangsgröße i_{ref} kann als Maß für die Robustheit die Amplituden- und Phasenreserve herangezogen werden. Den Frequenzgangsortskurven aus Bild 4 lässt sich entnehmen, dass das Regelungssystem stets eine nahezu unendliche Amplitudenreserve und eine Phasenreserve von mindestens 30° besitzt. Das Regelungssystem weist somit eine hohe Robustheit auf und bleibt auch bei großen Parametervariationen innerhalb der Regelstrecke robust stabil.
Eine detailliertere Beschreibung der Systemanalyse des Regelungssystems bestehend aus detailliertem Modell der elektromechanischen Servolenkung und Kompensator ist in [1-4] zu finden. Dieses Regelungssystem wird nun gemäß Bild 1 im nächsten Kapitel um das Modell der Lenkgefühlvorgabe erweitert.

5 Modell der Lenkgefühlvorgabe

Das Modell der Lenkgefühlvorgabe dient zur Erzeugung des Sollhandmoments T_{req}, mit dessen Hilfe der fahrzeugführenden Person eine Rückmeldung über den aktuellen Fahrzustand geben wird. Das Sollhandmoment T_{req} setzt sich dabei aus vier Anteilen zusammen, die in den folgenden Unterkapiteln beschrieben werden. Die entsprechenden Momente sind abhängig von der Schätzung $\hat{F}_R$ einer an der Zahnstange wirkenden Störkraft, der Schätzung $\hat{T}_S$ des tatsächlichen Handmoments, der Schätzung $\hat{\Omega}_P$ der Winkelgeschwindigkeit des Ritzels, dem Lenkradwinkel φ_S, der Fahrzeuggeschwindigkeit v_V sowie dem aktuellen Torsionsstabmoment T_{tb}. Die Störkraft F_R an der Zahnstange stellt dabei eine äquivalente Kraft infolge der Störmomente T_{WL} und T_{WR} um die Lenkachsen der Vorderräder dar. Diese Momente können mit den gegebenen Messgrößen nicht einzeln geschätzt werden.
Das Strukturbild des Modells der Lenkgefühlvorgabe ist in Bild 5 dargestellt.

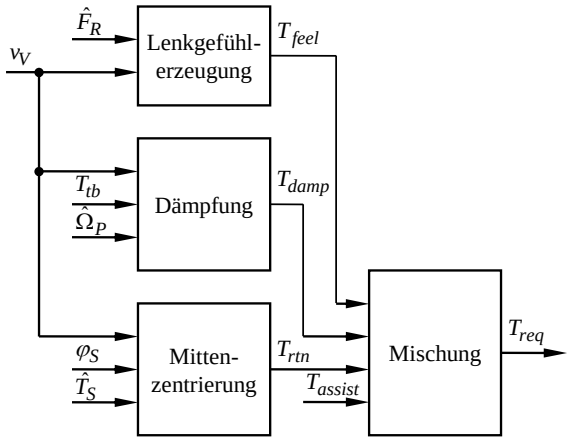


Bild 5 Strukturbild des Modells der Lenkgefühlvorgabe

5.1 Lenkgefühlerzeugung

Das Subsystem zur Lenkgefühlerzeugung ermittelt in Anlehnung an [6] abhängig von der Schätzung $\hat{F}_R$ der

an der Zahnstange wirkenden Störkraft mithilfe sogenannter Feeling Curves das Lenkmoment T_{feel}. Dabei ist bei niedrigen Fahrzeuggeschwindigkeiten ein komfortableres Lenkgefühl sowie bei hohen Fahrzeuggeschwindigkeiten ein direkteres Lenkgefühl gewünscht. Dies wird erreicht, indem jeweils eine Feeling Curve f_{pkg} für eine Fahrzeuggeschwindigkeit gleich Null und eine Feeling Curve f_{hwy} für betragsmäßig hohe Fahrzeuggeschwindigkeiten definiert wird, zwischen welchen in Abhängigkeit der aktuellen Fahrzeuggeschwindigkeit v_V mithilfe der Funktion f_v interpoliert wird. Hierbei gilt $f_v(v_V) \in [0,1]$. Das Lenkmoment T_{feel} berechnet sich dann zu

$$T_{feel} = f_{hwy}(\hat{F}_R)[1 - f_v(v_V)] + f_{pkg}(\hat{F}_R)\, f_v(v_V) \quad . \tag{1}$$

Bild 6 zeigt den qualitativen Verlauf der Feeling Curves in Abhängigkeit der geschätzten Störkraft $\hat{F}_R$ an der Zahnstange für niedrige und hohe Fahrzeuggeschwindigkeiten.

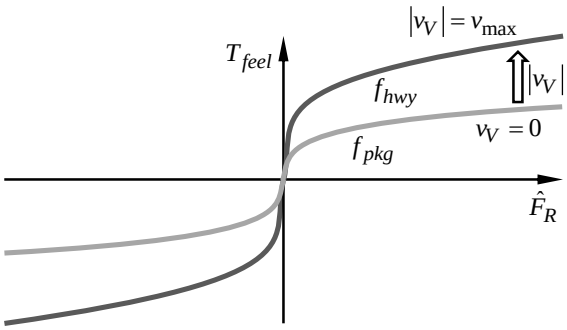


Bild 6 Vorgegebenes Lenkmoment in Abhängigkeit der geschätzten Störkraft an der Zahnstange und der Fahrzeuggeschwindigkeit

5.2 Dämpfung

Das Subsystem zur Dämpfung generiert angelehnt an [6] in Abhängigkeit der Schätzung $\hat{\Omega}_P$ der Winkelgeschwindigkeit des Ritzels ein Dämpfungsmoment T_{damp}. Dabei nimmt dieser Anteil zum Sollhandmoment T_{req} für eine steigende Fahrzeuggeschwindigkeit v_V zu sowie für ein größer werdendes Torsionsstabmoment T_{tb} ab und umgekehrt. Dies wird durch die Funktionen f_{b_v} und f_{b_T} charakterisiert. Das Dämpfungsmoment T_{damp} lässt sich dann durch

$$T_{damp} = f_{b_v}(v_V)\, f_{b_T}(T_{tb})\hat{\Omega}_P \tag{2}$$

ermitteln.

5.3 Mittenzentrierung

Im Subsystem zur Mittenzentrierung wird in Anlehnung an [8] abhängig vom Lenkradwinkel φ_S das Moment T_{rtn} zur Rückstellung des Lenkrads erzeugt. Dabei soll die Mittenzentrierung für ein großes Handmoment T_S harmonisch sowie für ein geringes Handmoment T_S, wenn zum Beispiel das Lenkrad losgelassen wird, stärker wirksam sein. Dies wird erreicht, indem jeweils eine Kennlinie f_{hrmc} für ein betragsmäßig großes Handmoment und eine Kennlinie f_{ctrg} für ein Handmoment gleich Null definiert wird, zwischen welchen in Abhängigkeit der Schätzung $\hat{T}_S$ des aktuellen Handmoments mithilfe der Funktion f_{rtn_T} interpoliert wird. Dabei gilt $f_{rtn_T}(\hat{T}_S) \in [0,1]$. Zudem ist das Rückstellmoment T_{rtn} für niedrige Fahrzeuggeschwindigkeiten kleiner als für hohe Fahrzeuggeschwindigkeiten. Dies wird durch die Funktion f_{rtn_v} mit $f_{rtn_v}(v_V) \in [0,1]$ berücksichtigt. Somit folgt für das Rückstellmoment

$$T_{rtn} = \{ f_{hrmc}(\varphi_S)\, f_{rtn_T}(\hat{T}_S) + f_{ctrg}(\varphi_S)[1 - f_{rtn_T}(\hat{T}_S)]\}\, f_{rtn_v}(v_V) \quad . \tag{3}$$

Bei Stillstand des Fahrzeugs ist das Rückstellmoment T_{rtn} gleich Null.
Der qualitative Verlauf der Kennlinien für eine harmonische oder stark wirksame Mittenzentrierung in Abhängigkeit des Lenkradwinkels φ_S ist in Bild 7 dargestellt.

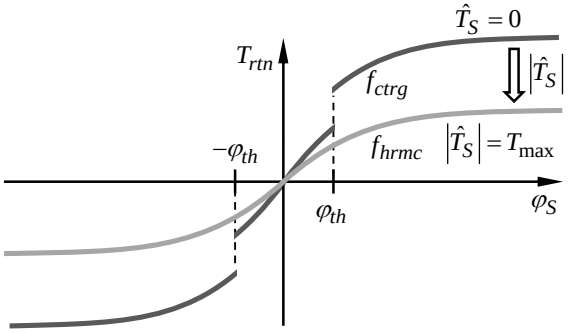


Bild 7 Rückstellmoment zur Mittenzentrierung des Lenkrads in Abhängigkeit des Lenkradwinkels und der Schätzung des Handmoments

5.4 Mischung

Das Subsystem zur Mischung bildet die gewichtete Summe der einzelnen Momente T_{feel}, T_{damp} und T_{rtn} aus den zuvor beschriebenen Subsystemen sowie des von Assistenzsystemen geforderten Handmoments T_{assist}. Diese Summe ist gleich dem Sollhandmoment T_{req}, das die Ausgangsgröße des Modells der Lenkgefühlvorgabe darstellt.
Durch den modularen Aufbau des Modells der Lenkgefühlvorgabe ist es möglich, einzelne Anteile zu vernachlässigen oder weitere Anteile zu berücksichtigen.

6 Robustheitsanalyse

Die Robustheit des Regelungssystems, welches um das zuvor beschriebene Modell der Lenkgefühlvorgabe erweitertet wird, lässt sich nun ebenfalls im Frequenzbereich anhand linearer Modelle analysieren. Bild 8 zeigt dafür exemplarisch die Frequenzgangsortskurven des offenen Regelkreises, welcher durch Schnitt des geschlossenen Regelkreises nach

Bild 1 am Steuereingang i_{ref} der Regelstrecke entsteht, wenn innerhalb der Regelstrecke wie zuvor die Massenträgheitsmomente (J_{min}/J_{max}), Steifigkeiten (c_{min}/c_{max}) und Übersetzungsverhältnisse (i_{med}/i_{max}) variiert werden. Das Regelungssystem besteht dabei aus dem linearisierten detaillierten Modell der elektromechanischen Servolenkung, dem mit dem reduzierten Modell der elektromechanischen Servolenkung entworfenen dynamischen Kompensator sowie dem linearisierten Modell der Lenkgefühlvorgabe. Die Linearisierung des Modells der Lenkgefühlvorgabe erfolgt hier für den Zustand, dass das Fahrzeug mit einer konstanten Geschwindigkeit v_V von 50 km/h ohne Einwirkung von Störungen geradeaus fährt und die fahrzeugführende Person kein Moment T_S am Lenkrad einprägt.

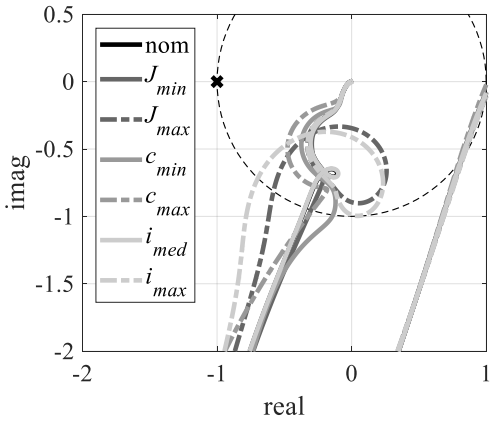


Bild 8 Frequenzgangsortskurven des offenen Regelkreises mit Lenkgefühlvorgabe bei Parametervariationen innerhalb der Regelstrecke (1)

Die Frequenzgangsortskurven aus Bild 8 mit weiterhin nahezu unendlichen Amplitudenreserven und großen Phasenreserven zeigen, dass das Modell der Lenkgefühlvorgabe die Robustheitseigenschaften des Regelungssystems aus Kapitel 4 nicht beeinträchtigt.
Eine weiterführende Robustheitsanalyse zeigt zudem, dass der Verlauf der Kennlinien innerhalb des Modells der Lenkgefühlvorgabe in einem sehr großen Bereich variiert werden kann, ohne dass Robustheitsprobleme auftreten. Dafür sind beispielhaft in Bild 9 die Frequenzgangsortskurven des offenen Regelkreises für die zuvor beschriebenen Parametervariationen dargestellt, wenn das Modell der Lenkgefühlvorgabe für eine erhöhte Fahrzeuggeschwindigkeit und eine starke Lenkaktivität der fahrzeugführenden Person linearisiert wird. Darüber hinaus wurde die Steigung der Feeling Curves maximal gewählt, sodass durch den EPS-Motor nahezu keine Unterdrückung der äquivalenten Störkraft F_R an der Zahnstange infolge des Kontakts zwischen Reifen und Fahrbahn stattfindet. Damit wird der fahrzeugführenden Person die gesamte Störkraft am Lenkrad vermittelt. Auch hier bleibt das Regelungssystem mit dem Modell der Lenkgefühlvorgabe robust stabil.

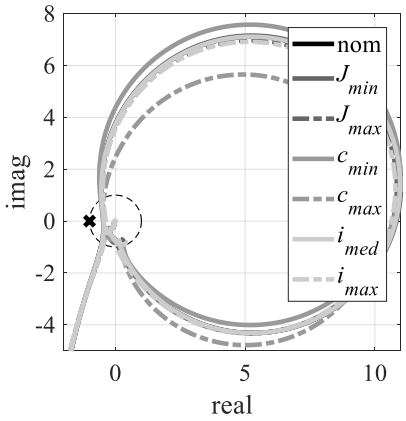


Bild 9 Frequenzgangsortskurven des offenen Regelkreises mit Lenkgefühlvorgabe bei Parametervariationen innerhalb der Regelstrecke (2)

Ähnliche Ergebnisse ergeben sich ebenfalls für andere analysierte Fahrsituationen, welche hier nicht dargestellt sind. Somit kann mit dem hier verwendeten Modell der Lenkgefühlvorgabe ein nahezu beliebiges Sollhandmoment eingestellt und gleichzeitig mit dem zuvor vorgestellten Kompensator ein ausreichend robustes Regelungssystem erzielt werden.

7 Simulationsergebnisse

In diesem Kapitel wird abschließend das Verhalten des nichtlinearen Regelungssystems analysiert. Das nichtlineare Simulationsmodell besteht aus dem detaillierten nichtlinearen Modell der elektromechanischen Servolenkung, dem mit dem reduzierten linearisierten Modell der elektromechanischen Servolenkung entworfenen dynamischen Kompensator sowie dem nichtlinearen Modell der Lenkgefühlvorgabe. Dabei werden lastabhängige Reibung im Kugelgewindetrieb, nichtlineare Federkennlinien und nichtlineare Übersetzungsverhältnisse sowie mechanische Begrenzungen berücksichtigt. Für glattere Verläufe der Zeitantworten werden Prozess- und Messrauschen sowie Quantisierung vernachlässigt. Zudem wird das nichtlineare Simulationsmodell um ein Fahrzeugmodell mit Reifenmodellen gemäß [9] zur realitätsnahen Berechnung der Störmomente T_{WL} und T_{WR} um die Lenkachsen der Vorderräder infolge des Kontakts zwischen Reifen und Fahrbahn erweitert.
Bild 10 zeigt die Simulationsergebnisse, wenn das nichtlineare Simulationsmodell ab 0,05 s sprungförmig mit einem Handmoment T_S von 1 Nm angeregt wird. Dies entspricht dem Fall, dass die fahrzeugführende Person stetig den Lenkradwinkel φ_S erhöht. Das Fahrzeug fährt dabei mit einer nahezu konstanten Geschwindigkeit v_V von 30 km/h.
Wie die Zeitantworten oben links in Bild 10 zeigen, folgt die Regelgröße T_{tb} gut der Sollgröße T_{req}. Die Regelung weist somit wie zuvor beschrieben ein gutes dynamisches Verhalten auf. Auch die Schätzung $\hat{T}_S$ des Handmoments folgt gut dem tatsächlichen Handmoment T_S. Dies ist den Zeitantworten unten rechts zu

entnehmen. In den Zeitantworten unten links erkennt man zudem, dass sich neben der Beobachterdynamik auch die lastabhängige Reibung im Kugelgewindetrieb auf die Schätzung $\hat{F}_R$ der an der Zahnstange wirkenden Störkraft auswirkt. Dazu ist unten links in Bild 10 ebenfalls die äquivalenten Störkraft F_R an der Zahnstange dargestellt. Diese kann mithilfe der Störmomente T_{WL} und T_{WR} um die Lenkachsen der Vorderräder sowie den nichtlinearen Übersetzungsverhältnissen i_{RWL} und i_{RWR} ermittelt werden.

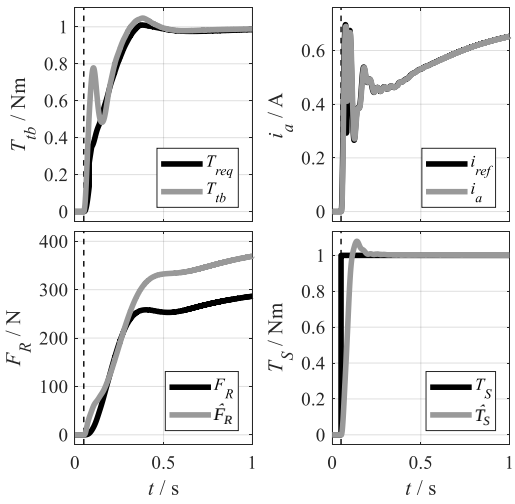


Bild 10 Zeitverläufe der Soll- und Regelgröße (oben links), der Stellgröße (oben rechts) sowie der Störgrößen (unten) des nichtlinearen Regelungssystem

Die Schätzung $\hat{F}_R$ der an der Zahnstange wirkenden Störkraft besitzt einen großen Einfluss auf die Bestimmung des Sollhandmoments T_{req}. Damit wirken sich die Beobachterdynamik und die lastabhängige Reibung im Kugelgewindetrieb auf das Sollhandmoment T_{req} und somit auch auf das Handmoment T_S, welches die fahrzeugführende Person spürt, aus. Dennoch wird mittels der Schätzung $\hat{F}_R$ der Störkraft die aktuelle Fahrsituation der fahrzeugführenden Person gut vermittelt.

8 Zusammenfassung

Der Beitrag skizziert eine robuste Lenkungsregelung in Form einer Handmomentenregelung mit Lenkgefühlvorgabe. Das entstehende Regelungssystem zeigt dabei ein gutes dynamisches Verhalten und eine gute Störunterdrückung. Es ist zudem robust gegenüber externen Störungen, nichtlinearem Systemverhalten und Freiheitsgraden der Regelstrecke, welche nicht im Regelungsentwurf berücksichtigt werden. Mit dieser Regelung und dem Modell der Lenkgefühlvorgabe kann daher ein nahezu beliebiges Sollhandmoment, welches die fahrzeugführende Person verspüren soll, ohne Beeinträchtigung der Robustheitseigenschaften eingestellt werden. Dies ermöglicht die Anpassung des Lenkgefühls an die jeweilige Fahrsituation. Das vorgestellte Regelungskonzept erfüllt somit die Anforderungen an Sicherheit und Fahrkomfort des Lenksystems.
Im nächsten Schritt soll das Regelungskonzept auf Steer-by-Wire Systeme übertragen werden.

9 Literatur

[1] Irmer, M.; Henrichfreise, H.: *Design of a robust LQG Compensator for an Electric Power Steering*. In: 21st IFAC World Congress. Elsevier, Berlin, 2020.

[2] Irmer, M.; Haßenberg, M.; Briese, H.; Henrichfreise, H.: *Mechatronic system design for EPS systems with residual modes and variable, nonlinear plant behavior*. In: 20. Internationales Stuttgarter Symposium. Springer Vieweg, Stuttgart, 2020. ISBN 978-3-658-29942-2. S. 469-482.

[3] Irmer, M.; Haßenberg, M.; Briese, H.; Henrichfreise, H.: *Robust Control of an Electric Power Steering with unconsidered Modes and Parameter Uncertainties*. In: 10th International Munich Chassis Symposium 2019. Springer Vieweg, München, 2019. ISBN 978-3-658-26434-5. S. 465-479.

[4] Wittler, G.; Haßenberg, M.; Briese, H.; Schubert, T.; Henrichfreise, H.: *Systematic model-based vibration analysis of a controlled electric power steering system*. In: 8th International Munich Chassis Symposium 2017. Springer Vieweg, München, 2017. ISBN 978-3-658-18458-2. S. 505-517.

[5] Henrichfreise, H.; Jusseit, J.; Niessen, N.: *Optimale Regelung einer elektromechanischen Servolenkung*. In: Fachtagung Mechatronik 2003. VDI, Fulda, 2003.

[6] Graßmann, O.; Henrichfreise, H.; Niessen, H.; Hammel, K. v.: *Variable Lenkunterstützung für eine elektromechanische Servolenkung.* In: 23. Tagung „Elektronik im Kraftfahrzeug“. Haus der Technik, Stuttgart, 2003.

[7] Niessen, H.; Henrichfreise, H.: *Vehicle steering system for controlling a steering or steering lock angle for at least one wheel of a vehicle*. Internationale Patentnummer WO 02/076806 A1, World Intellectual Property Organization, Genf, 2002.

[8] Niessen, H.; Starbeck, R. F.: *Steering mechanism of a vehicle, comprising a variable centering device*. Internationale Patentnummer W0 03/072417 A1. World Intellectual Property Organization, Genf, 2003.

[9] Irmer, M.; Henrichfreise, H.: *Design of a Simulation Environment for Testing the Control of Electric Power Steering Systems*. In: 21. Internationales Stuttgarter Symposium. Springer Vieweg, Stuttgart, 2021.

Nichtlineares Zweispurmodell eines LKW-Sattelaufliegers mit experimenteller Validierung der lateralen und vertikalen Reifenkräfte

Nonlinear Two-Track Model of a Semitrailer with Experimental Validation of Lateral and Vertical Tire Forces

Simon F. G. Ehlers,[1] Zygimantas Ziaukas,[1] Jan-Philipp Kobler,[2], Alexander Busch,[1] Tobias Ortmaier,[1] Mark Wielitzka[1]
[1]Leibniz Universität Hannover, Institut für Mechatronische Systeme, 30823 Garbsen, Deutschland, simon.ehlers@imes.uni-hannover.de
[2] BPW Bergische Achsen KG, 51674 Wiehl, Deutschland

Kurzfassung

Im Rahmen der Automatisierung von Nutzfahrzeugen halten verstärkt Assistenzsysteme Einzug in den LKW-Bereich. Dem Sattelauflieger kommt dabei eine wichtige Rolle zu. Dieser trägt aufgrund seiner unterschiedlich möglichen Beladungszustände und dem großen Anteil an der Gesamtmasse des Zuges, vor allem im beladenen Zustand, maßgeblich zur Fahrdynamik des gesamten Zuges bei. Um eine Grundlage für die weitere Entwicklung von Assistenzsystemen zu schaffen, wird in dieser Arbeit ein Zweispurmodell eines Sattelaufliegers entwickelt, welches die Quer- und Wankdynamik umfasst. Dies ermöglicht die nachfolgende Entwicklung von Beobachtern und Filtern zur Schätzung von nicht oder nur aufwendig messbaren Fahrzeugzuständen und -parametern. Eine Offline-Identifikation der unbekannten Modellparameter wird mittels einer Partikel-Schwarm Optimierung (PSO) durchgeführt. Die Validierung des Zweispurmodells wird auf Basis von Messfahrten eines Versuchsfahrzeugs vorgenommen, wobei der Fokus auf den lateralen und vertikalen Reifenkräften des Sattelaufliegers liegt, die durch Dehnungsmessstreifen (DMS) am Versuchsfahrzeug erfasst werden.

Abstract

As part of the automation of commercial vehicles, the number of assistance systems in this field is continuously increasing. The semitrailer plays an important role for the vehicles driving dynamics due to its highly varying loads and the large propotion to the total mass of the truck-semitrailer, especially when it is fully loaded. To create a basis for further development of assistance systems for the semitrailer, this paper presents a two-track model which includes the lateral and roll dynamics of the semitrailer. This enables the future development of observer and filter-based estimation of vehicle states and parameters, which are impossible or very difficult to measure. For offline identification of the unknown model parameters a Particle-Swarm-Optimization (PSO) algorithm will be used. The validaton of the model is based on measurements from a test vehicle. The focus is on the lateral and vertical tire forces of the semitrailer, which are measured at the test vehicle using strain gauges.

1 Einleitung

Automatisiert fahrende Nutzfahrzeuge bieten in Zukunft die Chance auf hohe Wirtschaftlichkeit für den Betreiber und auf erhöhte Sicherheit im Straßenverkehr. Bereits heutzutage werden verstärkt Advanced Driver Assistance Systems (ADAS) in Nutzfahrzeugen eingesetzt [1], z.B. das elektronische Stabilitäts-Programm (ESP) oder ein Spurhalteassistent. Vor allem der Sattelauflieger hat aufgrund seiner unterschiedlich möglichen Beladungszustände und dem großen Anteil an der Gesamtmasse des Zuges maßgeblichen Einfluss auf die Fahrdynamik. Die Kenntnis über die fahrdynamischen Größen des Sattelaufliegers, zum Beispiel die lateralen und vertikalen Reifenkräfte, ermöglicht die Entwicklung weiterer Assistenzsysteme zur Überwachung und Schätzung des Verschleißes von sicherheitsrelevanten Komponenten, wie z.B. der Reifen [2]. Des Weiteren kann auf Basis des fahrdynamischen Modells eine Aussage über die Fahrstabilität und -sicherheit von Manövern gegeben werden, was für die Pfadplanung von automatisierten Fahrzeugen von hoher Relevanz ist.

Für Personenkraftwagen gibt es bereits viele Ansätze zur fahrdynamischen Modellierung mit einem Zweispurmodell und anschließender Schätzung der lateralen und vertikalen Reifenkräfte mit Beobachterstrukturen oder Filtern [3]. Die Entwicklung von Modellen des Sattelaufliegers ist in der Literatur bisher jedoch weitaus weniger beschrieben, vor allem in Bezug auf die Reifenkräfte. Zwar gibt es Arbeiten, die ein Modell für die Zugmaschine und den Sattelauflieger erstellen, allerdings handelt es sich dabei meistens um Einspurmodelle, die zum Teil weder mit einem detaillierten Simulationsmodell noch in einer experimentellen Untersuchung validiert wurden [4]. Verschiedene lineare Einspurmodelle eines Sattelzuges werden in [4] miteinander verglichen und die Validierung durch ein detailliertes Mehrkörper-Simulationsmodell durchgeführt. Ein nichtlineares Einspurmodell wurde in [5] vorgestellt und die modellbasiert berechneten lateralen Reifenkräfte auf Basis von Kraftmesssensorik validiert. Da ein Einspurmodell die Reifen einer Achse zusammengefasst betrach-

tet, können die lateralen Reifenkräfte nicht für einzelne Reifen gesondert modelliert werden. Auch die Dynamik der vertikalen Reifenkräfte durch den Einfluss der Querbeschleunigung und Wankdynamik kann durch ein Einspurmodell nicht abgebildet werden. Unter dem Aspekt der Verschleißschätzung einzelner Komponenten (z.B. Reifen) oder der Fahrsicherheit (z.B. Umkippen des Sattelaufliegers) kann der Einsatz eines Zweispurmodells, trotz des erhöhten Modellierungs- und Rechenaufwands, sinnvoll sein. Sowohl Einspur- als auch Zweispurmodelle von mehreren Zugmaschine-Anhängerkombinationen unter Berücksichtigung eines minimalen Rechenaufwands mit einer anschließenden experimentellen Validierung des Modells werden in [6] präsentiert. In [7] wird ebenfalls ein detailliertes Zweispurmodell aufgebaut und abschließend experimentell validiert. Allerdings liegt bei beiden Arbeiten der Fokus der Validierung nicht auf den Reifenkräften sondern auf kinematischen Größen, wie z.B. Gierrate oder Querbeschleunigung.

Im Gegensatz dazu wird in dieser Arbeit die Validierung eines Zweispurmodells eines Sattelaufliegers auf Basis von experimentell bestimmten lateralen und vertikalen Reifenkräften durchgeführt. Dazu wird in Kapitel 2 das mathematische Modell des Sattelaufliegers als Zweispurmodell und des Zugfahrzeugs als Einspurmodell vorgestellt. Die Identifikation der Modellparameter mittels globalem Identifikationsalgorithmus wird in Kapitel 3 behandelt. In Kapitel 4 wird das Modellverhalten auf Basis der Messung der lateralen und vertikalen Reifenkräfte sowie der Gier- und Wankrate unter Betrachtung des Root-Mean-Square-Errors (RMSE) als Fehlermaß validiert. Eine Zusammenfassung der Erkenntnisse sowie ein Ausblick auf nachfolgende Forschungsarbeiten finden sich in Kapitel 5.

2 Modellierung des Sattelzuges

In dieser Arbeit wird die Zugmaschine als Einspurmodell modelliert, da der Fokus auf dem fahrdynamischen Verhalten des Sattelaufliegers liegt. Dieses Vorgehen verringert die Modellkomplexität und den Rechenaufwand. Der Sattelauflieger hingegen ist als Zweispurmodell umgesetzt, wobei der starre Fahrzeugaufbau über Feder- und Dämpferelemente mit den sechs Rädern verbunden ist. Da die Kommunikation zwischen Sattelauflieger und Zugmaschine im praktischen Einsatz keine Information über Antriebs- und Bremsmomente enthält, wird hier die Modellierung der Längsdynamik nicht betrachtet. Zur Modellierung nach Newton-Euler wird das Freikörperbild in **Abbildung 1** verwendet. Bei der Nummerierung der Achsen steht die erste Stelle für das i-te Glied im Zug und die zweite Stelle für die j-te Achse des i-ten Glieds. Für die Zugmaschine ($i = 1$) ergeben sich die Bewegungsgleichungen zu

$$(4m_{\mathrm{R}_1} + m_{\mathrm{A}_1})a_{y_1} = F_{y_{11}}\cos(\delta) + F_{y_{12}} - F_{\mathrm{K}y}\cos(\theta), \quad (1)$$

$$J_{z_1}\ddot{\psi}_1 = F_{y_{11}}\cos(\delta)l_{\mathrm{v}_1} - F_{y_{12}}l_{\mathrm{h}_1} + F_{\mathrm{K}y}\cos(\theta)l_{\mathrm{k}_1}, \quad (2)$$

wobei $a_{y_1} = \dot{v}_{y_1} + v_{x_1}\dot{\psi}_1$ die Querbeschleunigung der Zugmaschine, $F_{y_{1i}}$ die Reifenquerkräfte, δ den Lenkwinkel, J_{z_1} das Massenträgheitsmoment und θ den Knickwinkel zwischen der Zugmaschine und dem Sattelauflieger beschreibt. Die Größen l_{v_1},l_{h_1} und l_{k_1} stellen die Abstände der Vorder- bzw. Hinterachse sowie des Koppelpunkts zum Schwerpunkt dar. Die Achs- und Radmassen sind für jedes Rad vereinfacht in m_{R_i} zusammengefasst. Die Aufbaumasse des jeweiligen Gliedes ist m_{A_i}. Die Koppelkraft $F_{\mathrm{K}y}$ wirkt nur in y_2-Richtung (Querrichtung) am Königszapfen des Sattelaufliegers, da die Längsdynamik vernachlässigt wird. Weitere Effekte an der Sattelkupplung, wie z.B. eine Wankabstützung, werden nicht berücksichtigt.

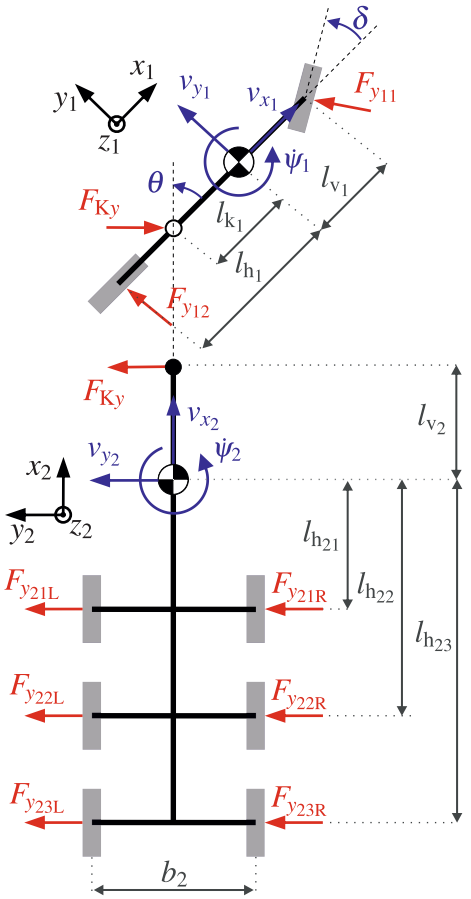


Abbildung 1 Freikörperbild der Zugmaschine und des Sattelaufliegers in der x-y-Ebene.

Die Bewegungsgleichungen des Sattelaufliegers ergeben sich zu

$$(6m_{\mathrm{R}_2} + m_{\mathrm{A}_2})a_{y_2} + m_{\mathrm{A}_2}h_{W_2}\ddot{\kappa}_2 = \sum F_{y_{2j\{\mathrm{L,R}\}}} + F_{\mathrm{K}y}, \quad (3)$$

$$\begin{aligned} J_{z_2}\ddot{\psi}_2 =& F_{\mathrm{K}y}l_{\mathrm{v}_2} - (F_{y_{21\mathrm{R}}} + F_{y_{21\mathrm{L}}})l_{\mathrm{h}_{21}} \\ &- (F_{y_{22\mathrm{R}}} + F_{y_{22\mathrm{L}}})l_{\mathrm{h}_{22}} - (F_{y_{23\mathrm{R}}} + F_{y_{23\mathrm{L}}})l_{\mathrm{h}_{23}}, \end{aligned} \quad (4)$$

mit der Querbeschleunigung $a_{y_2} = \dot{v}_{y_2} + v_{x_2}\dot{\psi}_2$, der Wankwinkelbeschleunigung $\ddot{\kappa}_2$, dem Massenträgheitsmoment J_{z_2} sowie dem Abstand des Aufbauschwerpunkts zur Wankachse h_{W_2}. Ein Freikörperbild der Wankdynamik ist in **Abbildung 2** gezeigt. Das Wankverhalten wird beschrie-

ben durch

$$\begin{aligned} J_{x_2}\ddot{\kappa}_2 =& (a_{y_2}\cos(\kappa_2)+g\sin(\kappa_2))m_{A_2}h_{W_2} \\ &+\frac{b_2}{2}(-F_{FD_{21R}}+F_{FD_{21L}}-F_{FD_{22R}} \\ &+F_{FD_{22L}}-F_{FD_{23R}}+F_{FD_{23L}}) \\ &-F_{Ky}h_{W_K}\cos(\kappa_2) \end{aligned} \tag{5}$$

mit der rechten bzw. linken Feder-Dämpfer Kraft $F_{FD_{2j\{L,R\}}}$ der j-ten Achse, der Spurweite b_2 und dem Abstand des Koppelpunkts zur Wankachse h_{W_K}. Das Heben des Aufbaus wird vernachlässigt. Die Kräfte in den rechten bzw.

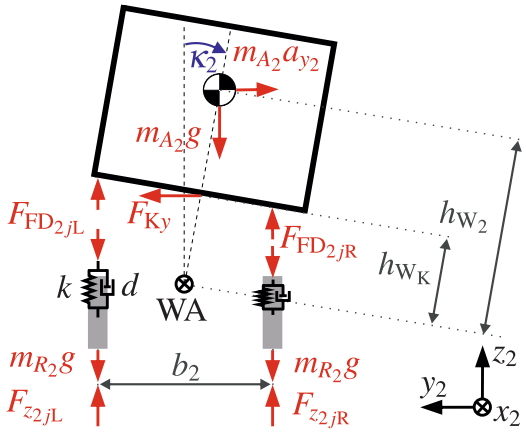


Abbildung 2 Freikörperbild in der y-z-Ebene zur Veranschaulichung der Wankdynamik.

linken Feder-Dämpfer Systemen (Steifigkeit k, Dämpfung d) der j-ten Achse setzen sich aus dem statischen und dynamischen Anteil der Federkraft sowie der Dämpferkraft zusammen zu

$$\begin{aligned} F_{FD_{2jR}} &= F_{stat_{2jR}}+\kappa_2\frac{b_2}{2}k+\dot{\kappa}_2\frac{b_2}{2}d \quad \text{und} \\ F_{FD_{2jL}} &= F_{stat_{2jL}}-\kappa_2\frac{b_2}{2}k-\dot{\kappa}_2\frac{b_2}{2}d. \end{aligned} \tag{6}$$

Zur Berechnung der statischen Aufstandskräfte an den Achsen 21, 22 und 23 wird aufgrund des Luftfedersystems die Annahme getroffen, dass sich die Last auf alle drei Achsen gleichmäßig verteilt. Somit ergibt sich

$$F_{stat_{11}} = (m_{A_1}l_{h_1}e+m_{A_2}l_{h_{22}}(l_{h_1}-l_{k_1}))\frac{g}{ef}, \tag{7}$$

$$F_{stat_{12}} = (m_{A_1}l_{v_1}e+m_{A_2}l_{h_{22}}(l_{v_1}+l_{k_1}))\frac{g}{ef}, \tag{8}$$

$$F_{stat_{2j\{L,R\}}} = \frac{1}{6}m_{A_2}g\frac{l_{v_2}}{e}, \tag{9}$$

mit $e = l_{h_{22}}+l_{v_2}$ und $f = l_{v_1}+l_{h_1}$. Weiterhin wird ein lineares Verhalten der Feder- und Dämpferelemente für alle Achsen angenommen. Eine weitere Annahme ist, dass sich die Reifenaufstandkräfte für die Zugmaschine aus den statischen Achskräften und den Gewichtskräften der Reifen und Achsen zusammensetzen. Für den Sattelauflieger gilt

$$F_{z_{2j\{L,R\}}} = F_{FD_{2j\{L,R\}}}+m_{R_2}g. \tag{10}$$

Die laterale kinematische Beziehung zwischen den beiden Gliedern des Zuges leitet sich aus der Bedingung der gleichen Geschwindigkeit beider Körper im Koppelpunkt ab

$$v_{y_1}-\dot{\psi}_1 l_{k_1} = v_{x_2}\sin(\theta)+(v_{y_2}+\dot{\psi}_2 l_{v_2})\cos(\theta). \tag{11}$$

Die Winkelgeschwindigkeit des Knickwinkels setzt sich aus der Differenz der Gierraten der beiden Glieder zusammen zu

$$\dot{\theta} = \dot{\psi}_2-\dot{\psi}_1. \tag{12}$$

Zur Modellierung der statischen Reifenquerkräfte $F_{y_{stat,ij\{L,R\}}}$ in Abhängigkeit des Schräglaufwinkels α bieten sich nichtlineare Modelle an. Vor allem bei hohen Schräglaufwinkeln, wie sie häufig beim Sattelauflieger in engen Kurven an der ersten und dritten Achse auftreten, ist ein lineares Modell nicht mehr zulässig, was in [7] anhand experimenteller Daten gezeigt wurde. In dieser Arbeit wird eine vereinfachte Version des Magic Tire Formula Model (MTFM) [8] genutzt

$$F_{y_{stat,ij\{L,R\}}} = \mu F_{z_{ij\{L,R\}}}\sin(C_{ij}\arctan(B_{ij}\alpha_{ij\{L,R\}})), \tag{13}$$

wie in [5] beschrieben. Der Parameter μ fungiert als Skalierungsfaktor der Funktion und beschreibt unter anderem den Reibwert im Reifen-Fahrbahn-Kontakt. Die Funktion enthält die Koeffizienten C_{ij} und

$$B_{ij} = \frac{1}{C_{ij}\mu F_{z_{ij\{L,R\}}}}c_{1_{ij}}\sin(2\arctan\left(\frac{F_{z_{ij\{L,R\}}}}{c_{2_{ij}}}\right)). \tag{14}$$

Daraus folgt, dass die zur Beschreibung des Querkraftverhaltens unbekannten Parameter μ, C_{ij}, $c_{1_{ij}}$ und $c_{2_{ij}}$ ermittelt werden müssen. Die Schräglaufwinkel ergeben sich für das Einspurmodell der Zugmaschine zu

$$\begin{aligned} \alpha_{11} &= \delta-\arctan\left(\frac{v_{y_1}+\dot{\psi}_1 l_{v_1}}{v_{x_1}}\right) \quad \text{und} \\ \alpha_{12} &= -\arctan\left(\frac{v_{y_1}-\dot{\psi}_1 l_{h_1}}{v_{x_1}}\right). \end{aligned} \tag{15}$$

Für das Zweispurmodell des Sattelaufliegers ergibt sich für die j-te Achse

$$\begin{aligned} \alpha_{2jR} &= -\arctan\left(\frac{v_{y_2}-\dot{\psi}_2 l_{h_{2j}}+\dot{\kappa}_2 h_{W_2}}{v_{x_2}+\dot{\psi}_2\frac{b_2}{2}}\right) \quad \text{und} \\ \alpha_{2jL} &= -\arctan\left(\frac{v_{y_2}-\dot{\psi}_2 l_{h_{2j}}+\dot{\kappa}_2 h_{W_2}}{v_{x_2}-\dot{\psi}_2\frac{b_2}{2}}\right). \end{aligned} \tag{16}$$

Die zeitliche Verzögerung des Aufbaus der lateralen Reifenkräfte bei dynamischen Manövern bzw. sich schnell ändernden Schräglaufwinkeln wird durch ein Verzögerungsglied erster Ordnung berücksichtigt

$$F_{y_{ij\{L,R\}}}+\frac{l_{ij}}{v_{x_i}}\dot{F}_{y_{ij}} = F_{y_{stat,ij\{L,R\}}}, \tag{17}$$

wobei l_{ij} die Einlauflängen sind [9]. Weitere Effekte am Reifen, wie z.B. ein möglicher Stick-Slip-Effekt bei hohen Schräglaufwinkeln, werden hier nicht berücksichtigt.
Für das Zustandsraummodell ergibt sich ein Zustandsvek-

tor mit $n_x = 15$ Zuständen

$$\boldsymbol{x} = (v_{y_1}, \dot{\psi}_1, \kappa_2, v_{y_2}, \dot{\kappa}_2, \dot{\psi}_2, \theta, F_{y_{11}}, F_{y_{12}}, F_{y_{21\mathrm{R}}}, F_{y_{21\mathrm{L}}}, F_{y_{22\mathrm{R}}}, F_{y_{22\mathrm{L}}}, F_{y_{23\mathrm{R}}}, F_{y_{23\mathrm{L}}})^{\mathrm{T}}. \tag{18}$$

Um aus den vorgestellten Gleichungen ein Zustandsraummodell zu erzeugen, wird mit Gleichung (1) die Koppelkraft $F_{\mathrm{K}y}$ eliminiert. Somit bleiben noch 4 Gleichungen (2)-(5) übrig. Unter Hinzunahme der zeitlichen Ableitung von (11), der Beziehung (12) und $\dot{\boldsymbol{x}}(3) = \boldsymbol{x}(5)$ sowie der 8 Gleichungen der Reifenkräfte (17), ergeben sich insgesamt 15 Gleichungen. Diese können in die Form

$$\boldsymbol{M}\dot{\boldsymbol{x}} = \tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{u}) \quad \text{mit} \quad \boldsymbol{u} = (\delta, v_{x_2}, \dot{v}_{x_2})^{\mathrm{T}} \tag{19}$$

als Eingangsvektor überführt werden, wobei $\dot{v}_{x_2}$ die Längsbeschleunigung des Sattelaufliegers ist. Anschließend kann Gleichung (19) zu

$$\dot{\boldsymbol{x}} = \boldsymbol{M}^{-1}\tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \tag{20}$$

aufgelöst werden. Die $n_y = 12$ Größen für die Ausgangsgleichung werden zu

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) = (\dot{\psi}_1, \dot{\psi}_2, \dot{\kappa}_2, \theta, F_{y_{21\mathrm{R}}}, F_{y_{21\mathrm{L}}}, F_{y_{23\mathrm{R}}}, F_{y_{23\mathrm{L}}}, F_{z_{21\mathrm{R}}}, F_{z_{21\mathrm{L}}}, F_{z_{23\mathrm{R}}}, F_{z_{23\mathrm{L}}})^{\mathrm{T}} \tag{21}$$

gewählt, welche im weiteren Verlauf für die Identifikation und Validierung (Kap. 3 und 4) genutzt werden.

3 Identifikation der Modellparameter

Einige der Parameter, die in dem Zustandsraummodell in (20) verwendet werden, sind direkt messbar. Dazu gehören geometrische Längen, wie die Achsabstände zu dem Königszapfen von der Zugmaschine und dem Sattelauflieger. Auch die Fahrzeugmassen $m_{\mathrm{A}_1} + 4m_{\mathrm{R}_1}$ bzw. $m_{\mathrm{A}_2} + 6m_{\mathrm{R}_2}$ sowie die Schwerpunktpositionen in Längsrichtung der Zugmaschine l_{v_1} und des Sattelaufliegers l_{v_2} können durch Auswiegen der Achsen experimentell bestimmt werden. Die Massenträgheitsmomente J_{z_1}, J_{z_2} und J_{x_2} werden vereinfacht aus der Annahme eines Rechtecks mit homogener Massenverteilung bestimmt. Die Ersatzmassen für die Räder und Achse m_{R_1} und m_{R_2} können Datenblättern entnommen werden. Die weiteren Parameter werden mit einer Partikel-Schwarm-Optimierung (PSO) [10] bestimmt. Dazu gehören die Reifenparameter μ, C_{ij}, $c_{1_{ij}}$, $c_{2_{ij}}$ und l_{ij}. Es werden hierbei vereinfachend die gleichen Parameter für die sechs Reifen sowie für die Feder-Dämpfer-Elemente k und d des Sattelaufliegers angenommen, da sich die Reifen und Achsen $j = 1,2,3$ konstruktiv nicht unterscheiden. Da die Position der Wankachse WA sowie die Schwerpunkthöhe h_{W_2} des Sattelaufliegers (siehe Abb. 2) experimentell schwierig zu bestimmen sind, werden diese Parameter ebenfalls offline identifiziert. Somit ergeben sich die 15 zu identifizierenden Parameter

$$\boldsymbol{p} = (\mu, C_{11}, C_{12}, C_{2j}, c_{1_{11}}, c_{1_{12}}, c_{1_{2j}}, c_{2_{11}}, c_{2_{12}}, c_{2_{2j}}, l_{11}, l_{12}, l_{2j}, k, d, h_{\mathrm{W_K}}, h_{\mathrm{W}_2})^{\mathrm{T}}. \tag{22}$$

Zur Identifikation von $\boldsymbol{p}$ wurden Messfahrten mit dem Versuchsfahrzeug (siehe Kapitel 4) durchgeführt, welche die Manöver Slalom, doppelter Spurwechsel und Kurvenfahrt bei jeweils unterschiedlichen Geschwindigkeiten enthalten. Es wird die Kostenfunktion

$$J = \sum_{l=1}^{n_y} \frac{\|\boldsymbol{y}_{\mathrm{m}}(l) - \boldsymbol{y}(l)\|^2}{\|\boldsymbol{y}_{\mathrm{m}}(l) - \bar{\boldsymbol{y}}_{\mathrm{m}}(l)\|^2} \tag{23}$$

genutzt, wobei $\boldsymbol{y}(l)$ der l-te Eintrag aus (21) ist, $\boldsymbol{y}_m(l)$ die korrespondierende gemessene Größe und $\bar{\boldsymbol{y}}_{\mathrm{m}}(l)$ der jeweilige Mittelwert ist. Ziel der PSO ist es, einen Parametersatz $\boldsymbol{p}_{\mathrm{opt}}$ zu finden, der die Kostenfunktion J minimiert. Die Optimierung wurde auf maximal 150 Iterationen beschränkt. Auf Basis des ermittelten Parametersatzes wird anschließend eine lokale Optimierung mit sequentieller quadratischer Programmierung [11] durchgeführt, was in [5] zu einem 3% besseren Resultat führt. Da das Ergebnis der PSO stochastischen Einflüssen unterliegt, wurde die gesamte Identifikation 60 Mal wiederholt. Auf Basis aller Durchläufe wurde der Parametersatz mit dem geringsten Wert der Kostenfunktion für die experimentelle Validierung ausgewählt.

4 Experimentelle Validierung

4.1 Messfahrzeug

Das Messfahrzeug besteht aus einer Zugmaschine und dem Sattelauflieger, welcher der europäischen Richtlinie 96/53/EC entspricht. Der Sattelauflieger ist mit einem luftgefederten Fahrwerk (BPW Eco Air, BPW Bergische Achsen KG, Wiehl, Deutschland) sowie einem Elektronischen-Bremssystem (EBS, ZF WABCO, Friedrichshafen, Deutschland) ausgestattet. Zur Messdatenaufnahme wird ein Data-Aquisition-System (DAQ) (Dewesoft d.o.o, Trbovlje, Slovenien) verwendet, welches sowohl analoge Eingänge als auch CAN-Schnittstellen aufweist. Die Kommunikation über CAN zwischen der Zugmaschine und dem Sattelauflieger (ISO 11992) wird vom DAQ-System erfasst und beinhaltet unter anderem die Längsgeschwindigkeit des Sattelaufliegers. Des Weiteren wird der CAN-Bus der Zugmaschine (nach SAE J 1939) mit einer Frequenz von 100 Hz aufgezeichnet und beinhaltet die Gierrate $\dot{\psi}_1$ und den Lenkradwinkel δ der Zugmaschine. Die lateralen und vertikalen Reifenkräfte $F_{y_{21\mathrm{R}}}, F_{y_{21\mathrm{L}}}, F_{y_{23\mathrm{R}}}, F_{y_{23\mathrm{L}}}, F_{z_{21\mathrm{R}}}, F_{z_{21\mathrm{L}}}, F_{z_{23\mathrm{R}}}$ und $F_{z_{23\mathrm{L}}}$ werden durch kalibrierte Dehnungsmessstreifen (DMS) an der ersten und dritten Achse des Sattelaufliegers gemessen



Abbildung 3 Im Versuch verwendete Zugmaschine-Sattelauflieger-Kombination.

und als analoge Eingänge mit 1000 Hz vom DAQ-System erfasst. Zur Messung der kinematischen Größen (Gierrate $\dot{\psi}_2$ und Wankwinkelgeschwindigkeit $\dot{\kappa}_2$) des Sattelaufliegers wird eine 6-DOF inertiale Messeinheit bei einer Abtastrate von 100 Hz verwendet. Zur Messung des Knickwinkels θ kommt ein in den Königszapfen integrierter Winkelsensor (Abtastrate 1000 Hz) zum Einsatz.
Zur Entwicklung des Fahrzeugmodells wurde die Software MATLAB/Simulink (The MathWorks, Inc., Natick, USA) und Maple (Waterloo Maple Inc., Waterloo, Kanada) genutzt. Die Identifikation der Parameter sowie die Validierung wurde nach der erfolgten Messdatenaufnahme offline in MATLAB durchgeführt.

4.2 Validierung

Um das Modell zu validieren, wird eine Sequenz von Manövern hinzugezogen, die nicht für die Identifikation der unbekannten Modellparameter (Kap. 3) genutzt wurde. Die Sequenzlänge beträgt 115 s und beinhaltet verschiedene querdynamische Manöver bei unterschiedlichen Geschwindigkeiten. Die Messdatenaufnahme erfolgte auf einem trockenen Asphaltuntergrund. In den ersten 40 s werden mehrere doppelte Spurwechsel bei etwa 15 km/h gefahren (Abschnitt I). Anschließend wird der Sattelzug auf 30 km/h beschleunigt und ab ca. 50 s eine Rechtskurve gefahren (Abschnitt II). Daraufhin wird auf 18 km/h verzögert, um eine 180° Kehrtwende in Form einer Linkskurve auszuführen (Abschnitt III). Nach anschließender Beschleunigung auf 40 km/h werden erneut mehrere doppelte Spurwechsel hintereinander durchgeführt (Abschnitt IV). Somit sind in dieser Sequenz viele für den Straßenverkehr relevante Fahrsituationen abgebildet.

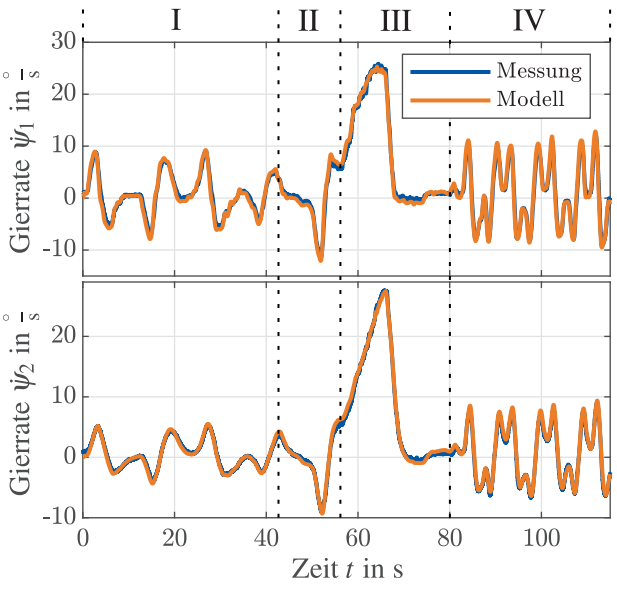


Abbildung 4 Gierrate von Zugmaschine und Sattelauflieger.

In **Abbildung 4** werden die modellbasiert berechneten Gierraten $\dot{\psi}_{1,2}$ von Zugmaschine und Sattelauflieger mit den Messungen verglichen. Es ist festzuhalten, dass in allen vier Abschnitten das Modell das Gierverhalten der Glieder gut abbilden kann. Der RMSE zwischen Modell und Messung beträgt für die Gierrate der Zugmaschine 0,76 $\frac{°}{s}$ und für den Sattelauflieger 0,39 $\frac{°}{s}$. Zudem ist in **Abbildung 5** die Wankwinkelgeschwindigkeit des Sattelaufliegers dargestellt, allerdings nur für Abschnitt IV. Da in den Abschnitten I-III nur eine geringe Anregung des Wankverhaltens vorhanden ist, lässt sich aufgrund des geringen Signal-Rausch-Verhältnisses der Messung keine Aussage über die Modellgüte treffen. Für den Abschnitt IV ist jedoch erkennbar, dass die Wankdynamik ausreichend angeregt wird und diese durch das Modell ebenfalls abgebildet werden kann (RMSE von 0,5 $\frac{°}{s}$).

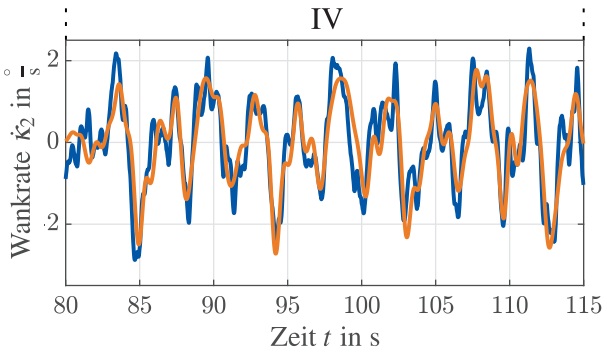


Abbildung 5 Wankwinkelgeschwindigkeit des Sattelaufliegers.

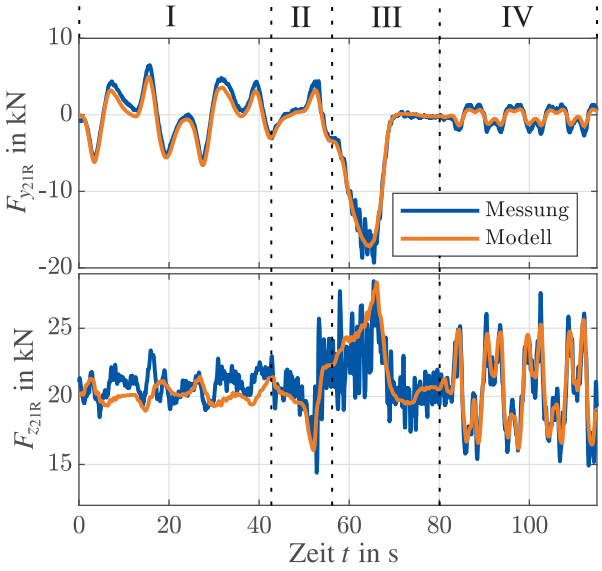


Abbildung 6 Laterale und vertikale Reifenkraft des rechten Reifens an der ersten Achse des Sattelaufliegers.

Die simulierten lateralen und vertikalen Reifenkräfte der ersten Achse des Sattelaufliegers werden mit den Messungen der DMS in **Abbildung 6** verglichen. Es ist zu erkennen, dass auch hier das Verhalten der lateralen Reifenkraft in allen Abschnitten I-IV gut abgebildet werden kann (RMSE von 0,74 kN). Das Verhalten der vertikalen Reifenkraft kann ebenfalls abgebildet werden (RMSE von 1,34 kN), wobei in Abschnitt III eine etwas schlechtere Ab-

bildungsgüte vorhanden ist. Für die Kräfte des linken Rei-

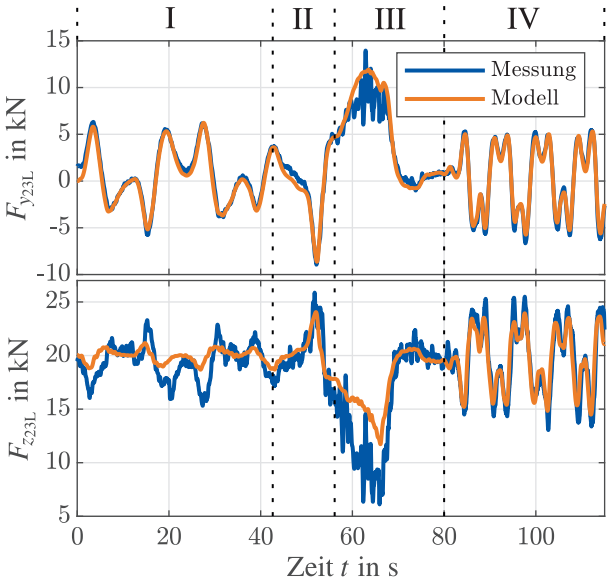


Abbildung 7 Laterale und vertikale Reifenkraft des linken Reifens an der dritten Achse des Sattelaufliegers.

fens der dritten Achse des Sattelaufliegers ist ähnliches zu beobachten (**Abbildung 7**). Das Modell kann die laterale Reifenkraft in Abschnitt I, II und IV gut abbilden (RMSE von 0,79 kN). Lediglich in Abschnitt III (enge Kurvenfahrt mit niedriger Geschwindigkeit) kommt es zu einer leicht schlechteren Abbildung des tatsächlichen Verhaltens. Dies ist ebenfalls für die simulierte vertikale Reifenkraft zu beobachten, die in Abschnitt III deutlich größer als die gemessene Kraft ist. Der RMSE für die vertikale Reifenkraft beträgt über alle Abschnitte 1,87 kN. Mögliche Erklärungsansätze beziehen sich auf das stark nichtlineare Systemverhalten in diesem Fahrzustand, da für dieses Modell einige Vereinfachungen angenommen wurden, wie lineare Feder- und Dämpfer-Elemente, ein vereinfachtes Reifenmodell (MTFM) und eine Vernachlässigung der Wankabstützung zwischen Zugmaschine und Sattelauflieger an der Sattelplatte. Einen weiteren Einfluss könnte eine leichte Querneigung der Fahrbahn haben. Weiterhin tritt bei den Messungen ein erhöhtes Rauschen im Kraftsignal, vor allem in Abschnitt III, auf. Dies könnte durch mögliche Stick-Slip-Effekte am Reifen oder durch den zusätzlichen Störeinfluss durch vertikale Straßenunebenheiten hervorgerufen werden.

5 Zusammenfassung

In dieser Arbeit wurde ein nichtlineares Zweispurmodell eines Sattelaufliegers vorgestellt, was die Quer- und Wankdynamik abbildet. Zur Modellierung der Reifenquerkräfte wurde eine vereinfachte Variante des Magic Tire Formula Model verwendet. Die unbekannten Modellparameter wurden durch eine Partikel-Schwarm-Optimierung bestimmt. Anhand von der Messung der Gier- und Wankrate sowie den lateralen und vertikalen Reifenkräften an Achse 1 und 3 des Sattelaufliegers wurde die Modellgüte validiert. Trotz einiger vereinfachenden Annahmen, wie lineare Feder-Dämpfer-Elemente oder die Vernachlässigung der Wankabstützung an der Sattelplatte, können die fahrdynamischen Zustände des Sattelaufliegers bei verhältnismäßig geringer Modellkomplexität gut abgebildet werden. Dies legt den Grundstein, um das Modell zur weiteren Entwicklung von Assistenzsystemen, wie der Zustands- und Parameterschätzung mit Beobachtern und Filtern, zu verwenden. Weitere Ansätze zur Steigerung der Modellgüte wären eine detailliertere Modellierung der Feder-Dämpfer-Systeme und der Kopplung zwischen Zugmaschine und Sattelauflieger.

Diese Arbeit ist im Rahmen des Projekts *IdenT* (19I19008A, 19I19008B) entstanden. Gefördert vom Bundesministerium für Wirtschaft und Energie aufgrund eines Beschlusses des Deutschen Bundestages.

6 Literatur

[1] Trigell, A. S. et al.: *Advanced vehicle dynamics of heavy trucks with the perspective of road safety*. In: Vehicle system dynamics, Vol. 55 Nr. 10, S. 1572-1617, Taylor & Francis, 2017

[2] Li, Y. et al.: *Analysis of impact factors of tire wear*. In: Journal of Vibration and Control., S. 833-840, 2011

[3] Jin, X.; Yin, G.; Chen, N.: *Advanced estimation techniques for vehicle system dynamic state: A survey*. In: Sensors, Vol. 19, Nr. 19, 2019

[4] Brock, Z.; Nelson, J., Hatton, R. L.: *A Comparison of Lateral Dynamic Models for Tractor-Trailer Systems*. In: 2019 IEEE Intelligent Vehicles Symposium (IV), S. 2052-2059, 2019

[5] Ziaukas, Z.; Kobler, J.-P.; Busch, A.; Wielitzka, M.; Ortmaier, T.: *Nonlinear Modeling of Truck-Semitrailer Lateral Dynamics Including Experimental Validation of Lateral Tire Forces*. In: The IAVSD International Symposium on Dynamics of Vehicles on Roads and Tracks, S. 1879-1888, 2019

[6] Ghandriz, T.; Jacobson, B.; Nilsson, P.; Laine, L.; Fröjd, N.: *Computationally Efficient Nonlinear One- and Two-Track Models for Multitrailer Road Vehicles*. In: IEEE Access, Vol. 8, S. 203854-203875, 2020

[7] Gäfvert, M.; Lindgärde, O.: *A 9-DOF tractor-semitrailer dynamic handling model for advanced chassis control studies*. In: Vehicle System Dynamics, Vol. 41, S. 51-82, 2004

[8] Pacejka, H. B: *Tire and vehicle dynamics*. 2. Aufl. Butterworth-Heinemann Ltd, 2006

[9] Schramm, D.; Hiller, M.; Bardini, R.: *Modellbildung und Simulation der Dynamik von Kraftfahrzeugen*. Vol. 2, Springer, 2013

[10] Kennedy, J.; Eberhart, R.: *Particle swarm optimization*. In: IEEE Proceedings of ICNN'95-International Conference on Neural Networks, Vol. 4, S. 1942-1948, 1995

[11] Nocedal, J.; Wright, S.: *Numerical optimization*. Springer Science & Business Media, 2006

Fein-Positionierung mit reflektierenden Markern
Fine-Positioning with Reflective Markers

Stefan-Octavian Bezrucav M.Sc., Tobias Haschke M.Sc., Markus Schmitz M.Sc., Simon Klamt B.Sc., Institut für Getriebetechnik, Maschinendynamik und Robotik, RWTH Aachen University, 52066 Aachen, Deutschland, {haschke, bezrucav, schmitz, klamt}@igmr.rwth-aachen.de
Dr.-Ing Jascha Paris, Dr.-Ing Stefan Kurtenbach, TRAPO AG, Industriestraße 1, 48712 Gescher, Deutschland, {jparis,skurtenbach}@trapo.de
Univ.-Prof. Dr.-Ing. Dr. h. c. Burkhard Corves, Institut für Getriebetechnik, Maschinendynamik und Robotik, RWTH Aachen University, 52066 Aachen, Deutschland, corves@igmr.rwth-aachen.de

Kurzfassung

Eine präzise Lokalisierung fahrerloser Transportsysteme (FTS) ist eine Grundvoraussetzungen für Logistikaufgaben in Produktionsstätten. Besonders in herausfordernden Situationen, wie dem Anfahren von Förderern, kann ein geringer Lokalisierungsfehler über den Erfolg des Be- und Entladevorgangs des FTS entscheiden. Aus diesem Grund werden im Rahmen der Arbeit zwei gesonderte Methoden zur präzisen Lokalisierung und anschließender Positionierung von FTS vorgestellt. Diese wurden für das *TRAPO Transport Shuttle* (TTS) der TRAPO AG entwickelt. Die Methoden basieren auf der Nutzung eines LiDAR-Sensors zur Detektion der Umgebung und reflektierenden Markern als Umgebungsmerkmal. Reflektierende Marker können das Licht des Sensors mit einer bestimmten Intensität reflektieren, sodass die Messpunkte auf den Markern aus den Messdaten herausgefiltert werden können. Im Anschluss kann die Pose, also die Position und Orientierung des Roboters mittels Laser-Triangulation berechnet werden.

Abstract

Precise localization of autonomous guided vehicles (AGVs) is a basic requirement for logistics tasks in production plants. Especially in challenging situations, such as the positioning in front of conveyors, a small localization error can decide on the success of the loading and unloading process of AGVs. For this reason, two separate methods for the precise localization and positioning of AGVs are presented in this thesis. These were developed for the *TRAPO Transport Shuttle* (TTS) of the TRAPO AG. The methods are based on the use of a LiDAR sensor for sensing the environment and reflective markers as environmental features. Reflective markers can reflect the light of the sensor with a specific intensity to filter out the measurement points on the markers from the measurement data. Afterwards the pose, i.e. the position and orientation of the robot can be calculated by laser triangulation.

1 Einleitung

Für die fehlerfreie Zusammenarbeit von hochautonomen Maschinen im Rahmen von Industrie 4.0 ist die Lokalisierung mobiler Systeme von großer Bedeutung. Besonders in kritischen Bereichen, wie zum Beispiel in der Nähe von Förderbändern, kann die aktuelle Pose, also die Position und Orientierung des Roboters über den Erfolg des Be- und Entladevorgangs entscheiden. Aus diesem Grund sind in den letzten Jahren einige Ansätze zur Lokalisierung und anschließender Positionierung von fahrerlosen Transportsystemen (FTS) entwickelt worden [1]. Zwei Ansätze, welche auf dem Einsatz eines LiDAR-Sensors und reflektierenden Markern als Umgebungsmerkmal basieren, werden in dieser Arbeit erläutert und miteinander auf Genauigkeit und Robustheit gegen Messungenauigkeiten verglichen. Im Anschluss werden diese Methoden mit dem *TRAPO Transport Shuttle* (TTS), dem FTS der TRAPO AG, validiert. Zuvor wird auf existierende Methoden zur Lokalisierung mobiler Roboter eingegangen.

2 Bestehende Methoden

Die Lokalisierung von Transportsystemen ist in Produktionsanlagen der Industrie 4.0 ausschlaggebend. Daher gibt es für die *Indoor-Lokalisierung* verschiedene Ansätze. Diese lassen sich in statische und flexible Lokalisierungsmethoden unterteilen. Die Unterschiede, sowie deren Vor- und Nachteile werden im Folgenden erläutert.

2.1 Statische Lokalisierungsmethoden

Unter statischen Lokalisierungsmethoden versteht man die Lokalisierung des FTS durch eine vordefinierten Route. Die Route wird dabei vor dem Start des eigentlichen Roboters definiert und während der Fahrt des Roboters nicht geändert. Eine verbreitete Methode im Bereich der Intralogistik ist das induktive Spurführungssystem, bei welchem die Pfade durch Leiterbänder im Boden der Produktionshalle festgelegt werden. Durch Magnete, welche an der Unterseite des Roboters montiert sind, wird ein Magnetfeld induziert, sobald dieser über die Leiterbänder fährt [2]. Dadurch ist eine Orientierung für den mobilen Roboter möglich. Neben dem induktiven Spurführungssystem ist die Definition des Pfades ebenfalls mit einem auf dem Boden befestig-

ten Magnetband möglich. Der Roboter kann sich wiederum mit einem Magneten an dem verlegten Magnetband orientieren [1].
In modernen FTS werden die statischen Lokalisierungs- und Navigationsmethoden jedoch selten eingesetzt. Obwohl die Implementierung der Methoden einfach ist, fehlt es den Methoden an Flexibilität [1].

2.2 Flexible Lokalisierungsmethoden

Die flexiblen Lokalisierungsmethoden sind im Gegensatz zu den statischen Verfahren vielseitig einsetzbar. Die Lokalisierung findet in diesem Fall ohne vordefinierten Weg statt. Der Roboter ist dementsprechend nicht an eine definierte Spur gebunden, sondern orientiert sich mit dem Einsatz verschiedener Sensoren in der Umgebung. Im Falle einer blockierten Spur ist die Umplanung des Pfades möglich. Daher können die Hindernisse ohne hohen Effizienzverlust umfahren werden. Aus diesem Grund werden in dynamisch wechselnden Umgebungen flexible Lokalisierungs- und Navigationsmethoden für Industrieroboter eingesetzt [2]. Allerdings ist zu beachten, dass die flexiblen Lokalisierungsmethoden im Vergleich zu den statischen Methoden einen Genauigkeitsverlust mit sich bringen können. Je nach Sensor (optisch, akustisch, etc.) können unterschiedliche Methoden zur Lokalisierung des FTS eingesetzt werden [3]. Eine mögliche Art der Lokalisierung ist die Lokalisierung mittels RFID-Chips. Der Roboter kann die RFID-Chips in der Umgebung registrieren und sich anhand der Chips orientieren [5]. Anstelle der RFID-Chips können Laser-Orientierungspunkte, sowie ein Laser-Sensor zur Lokalisierung des FTS eingesetzt werden. Anhand der Orientierungspunkte kann das FTS die aktuelle Pose relativ zu den Orientierungspunkte bestimmten. [1] Im Rahmen der Arbeit wird auf dieses Verfahren zurückgegriffen. Die Methoden, welche im Folgenden vorgestellt werden, nutzen einen LiDAR, *Light Detection and Ranging*-Sensor und reflektierende Marker in der Umgebung, um das FTS präzise zu lokalisieren.

3 Voraussetzungen

Für die Berechnung der aktuellen Pose des FTS werden Annahmen getroffen, um die Lokalisierung des Roboters zu vereinfachen. Da in dieser Arbeit Methoden zur präzisen Lokalisierung mit anschließender Positionierung vorgestellt werden, kann angenommen werden, dass sich der Roboter bereits in der Nähe des definierten Ziels befindet. Aus diesem Grund wird die Pfadplannung im Ganzen, also mit Vermeidung von Hindernissen etc. in diesem Paper als gegeben vorausgesetzt. Des Weiteren wird für die Lokalisierung eine genaue Platzierung der reflektierenden Markerpaare an den Rollerförderern als gegeben angenommen. Ein Markerpaar entspricht zwei reflektierenden Markern, welche beide zur Lokalisierung des FTS benötigt werden. Um eine hohe Genauigkeit der Methoden zu erzielen, müssen die Marker mit einem geringen Abstand am Rollenförderer angebracht werden. Auf die Positionierung der reflektierenden Markerpaare wird im weiteren Verlauf näher eingegangen.

4 FTS Lokalisierung

Im Folgenden werden zwei Methoden zur präzisen Lokalisierung fahrerloser Transportsysteme vorgestellt. Beide Lokalisierungsmethoden basieren auf der Nutzung eines LiDAR-Sensors zur Erforschung der Umgebung und reflektierenden Markern als Umgebungsmerkmal. Die Daten des LiDAR-Sensors werden im Anschluss an die Detektion der Umgebung für die Berechnung der aktuellen Pose $\chi = (x \;\; y \;\; \theta)^T$ des FTS in Referenz zu einem Koordinatensystem, definiert im Referenzpunkt M, verwendet. Die Methoden unterscheiden sich in der Berechnung der aktuellen Pose χ des Roboters. Zum besseren Verständnis werden in dieser Arbeit Koordinatensysteme in dem Sensor, dem Roboter, dem vorgegebenen Ziel, sowie dem Referenzpunkt M definiert. Mit Hilfe von homogenen Transformationen lassen sich die Translationen, sowie Rotationen der Koordinatensysteme beschreiben. Die gesuchte Pose χ, also die Pose zwischen dem Roboter und dem Referenzsystem M, lässt sich ebenfalls als homogene Transformation ${}^{R}T_{M}$ beschreiben. Diese Transformation lässt sich aus mehreren homogenen Transformationen zusammensetzen:

$$ {}^{R}T_{M} = {}^{R}T_{S}\,{}^{S}T_{M} \tag{1} $$

${}^{R}T_{S}$ entspricht der Transformation zwischen dem LiDAR-Sensor S und dem Roboter R und ${}^{S}T_{M}$ der Transformation zwischen dem Referenzpunkt M und dem Sensor. Dabei kann die Transformation ${}^{R}T_{S}$ als gegeben angenommen werden. Des Weiteren kann auch die Transformation ${}^{M}T_{Z}$, also die Transformation zwischen dem Ziel Z und dem Referenzpunkt M vorrausgesetzt werden. Somit wird lediglich die Transformation ${}^{S}T_{M}$ zur Berechnung benötigt. Diese ist durch die Lokalisierungsmethoden, welche im Folgenden vorgestellt werden, zu bestimmen. Die Koordinatensysteme und deren Transformationen ist Abbildung 1 zu entnehmen.

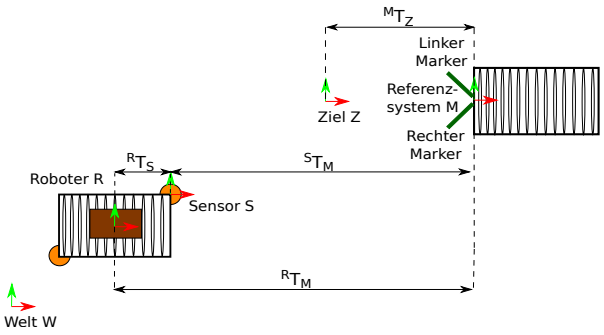


Abbildung 1 Definition der Koordinatensysteme und homogenen Transformationen

Für die Berechnung der Transformation ${}^{S}T_{M}$ nutzen die Lokalisierungsmethoden die Daten des LiDAR-Sensors. Aufgrund von Messungenauigkeiten des Sensors ist eine vorzeitige Aufbereitung der Daten nötig, um somit valide Lokalisierungsergebnisse zu erzielen. Die Aufbereitung wird durch Filter, welche in Reihe geschaltet sind, umgesetzt:

1. Intensitätsfilter: Da die reflektierenden Marker das

Licht mit einer bestimmten Intensität reflektieren, können die reflektierenden Marker innerhalb der Punktewolke des LiDAR-Sensors registriert werden. Der Intensitätsfilter nutzt diese Erkenntnis aus, um die Marker aus der Punktewolke herauszufiltern.

2. Ggf. Winkel- und Reichweitenfilter: In Produktionsstätten ist eine Identifizierung von mehreren Markerpaaren durchaus möglich. Somit ist die eindeutige Identifizierung des richtigen Markerpaares schwierig. Mit Hilfe des Winkel- und Reichweitenfilters wird der Bereich, in dem sich das Markerpaar befindet, eingeschränkt.

3. Ggf. Glättungsfilter: Aufgrund von Messungenauigkeiten des LiDAR-Sensor schwankt die Approximation der Pose. Aus diesem Grund werden Glättungsfilter, wie zum Beispiel der *Moving Average*-Filter zur Glättung der Messdaten verwendet.

4.1 Marker-Mittelpunkt Lokalisierung

Die Bestimmung der aktuellen Pose des Roboters in Relation zu den reflektierenden Markern ist mit Hilfe der Mittelpunkte der reflektierenden Marker möglich. Zur Berechnung der Mitte jedes Markers werden die Koordinaten der äußersten Punkte der Marker anhand trigonometrischen Gleichungen berechnen. Dies ist möglich, da wir durch die Messdaten des Sensors sowohl eine gemessene Weite R, sowie einen Winkel α für jeden Laserpunkt erhalten. Die äußersten Punkte auf der rechten und linken Seiten der zwei Marker P_{rr}, P_{rl}, P_{lr} und P_{ll} lassen sich, wie in Abbildung 2 beschrieben, berechnen:

$$P_i = \begin{pmatrix} R_i\cos(\alpha_i) \\ R_i\sin(\alpha_i) \\ 0 \end{pmatrix}, i \in \{rr, rl, lr, ll\} \tag{2}$$

Mit Hilfe der Eckpunkte der Marker ist die Berechnung der Mitte jeden Markers möglich. Die Position des Referenzsystems M kann im Anschluss aus den Markerzentren approximiert werden. Die Position wird in der Mitte zwischen den Mittelpunkten der Marker definiert. Hierbei wird angenommen, dass sich der Mittelpunkt der Marker genau im Zentrum zwischen den berechneten Eckpunkten P_{rr} und P_{rl} bzw. P_{lr} und P_{ll} liegt:

$$M = \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} = \begin{pmatrix} \frac{(P_{rl,x}+P_{rr,x})+(P_{ll,x}+P_{lr,x})}{4} \\ \frac{(P_{rl,y}+P_{rr,y})+(P_{ll,y}+P_{lr,y})}{4} \\ 0 \end{pmatrix} \tag{3}$$

Die Koordinaten des Systens M stimmen mit der Translation der Transformation ${}^{S}T_M$ überein, da der Ursprung des Sensor-Koordinatensystems exakt in der Laserquelle definiert ist. Neben der Translation wird für die Berechnung der homogenen Transformation ${}^{S}T_M$ eine Rotation der Koordinatensysteme zueinander gesucht. Für die genannte Problemstellung kann lediglich eine Rotation um die vertikale bzw. z-Achse vorliegen. Der damit verbundene Winkel, der sogenannte Gier-Winkel γ, lässt sich anhand

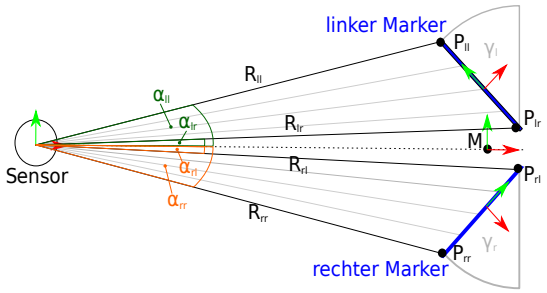


Abbildung 2 Funktionales Model der Marker-Mittelpunkt Lokalisierung

der Eckpunkte der Marker mittels Triangulation berechnen:

$$\Delta x_r = P_{rl,x} - P_{rr,x}, \qquad \Delta x_l = P_{ll,x} - P_{lr,x} \tag{4}$$

$$\Delta y_r = P_{rl,y} - P_{rr,y}, \qquad \Delta y_l = P_{ll,y} - P_{lr,y} \tag{5}$$

$$\gamma_r = \operatorname{atan}(|\frac{\Delta x_r}{\Delta y_r}|), \qquad \gamma_l = \operatorname{atan}(|\frac{\Delta x_l}{\Delta y_l}|) \tag{6}$$

$$\gamma_{total} = \begin{cases} \frac{\gamma_l - \gamma_r}{2}, & \Delta x_r > 0, \Delta x_l < 0 \\ \frac{\gamma_r - \gamma_l}{2} + \frac{\pi}{2}, & \Delta x_r < 0, \Delta x_l < 0 \\ \frac{\gamma_r - \gamma_l}{2} - \frac{\pi}{2}, & \Delta x_r > 0, \Delta x_l > 0 \\ \frac{\pi}{4}, & \Delta x_r = 0 \\ -\frac{\pi}{4} & \Delta x_l = 0 \end{cases} \tag{7}$$

Die Winkel γ_r bzw. γ_l repräsentieren die Winkel der approximierten Marker. Die Berechnung von γ_{total} hängt davon, wie der Sensor positioniert ist. Aus diesem Grund wird zwischen fünf möglichen Fällen unterschieden, welche in (7) definiert sind. Eine Visualisierung der Winkel lässt sich der Abbildung 2 entnehmen. Dabei ist anzumerken, dass der Winkel γ_{total} zwischen $-45°$ und $+45°$ gültig ist. Des Weiteren wird angenommen, dass die approximierten Winkel γ_r und γ_l kleiner als $90°$ sind. Diese Annahme ist möglich, da die Marker ansonsten nicht vom Sensor hätten detektiert werden können. Insgesamt lässt sich daher die Transformation zwischen dem Sensor und dem Referenzsystem M wie folgt beschreiben:

$${}^{S}T_M = \begin{pmatrix} \cos(\gamma_{ges}) & -\sin(\gamma_{ges}) & 0 & M_x \\ \sin(\gamma_{ges}) & \cos(\gamma_{ges}) & 0 & M_y \\ 0 & 0 & 1 & M_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{8}$$

Um aus der Transformation im Anschluss die gesuchte Pose χ zu erhalten, muss in einem ersten Schritt die Transformation ${}^{R}T_M$ berechnet werden. Anschließend kann die Pose χ aus der Transformation extrahiert werden. Eine Visualiserung der Marker-Mittelpunkt Lokalisierung lässt sich Abbildung 2 entnehmen.
Die beschriebene Methode bringt eine Ungenauigkeit mit sich, da die Annahme getroffen wurde, dass die Mittelpunkte der Marker exakt zwischen den äußersten Messpunkten zu finden sind. In der Realität ist das nicht zwingend der Fall. Aus diesem Grund lässt die Methode einen

Fehler zu, welcher Einfluss auf die Genauigkeit der Lokalisierung hat.

4.2 Marker-Schnittpunkt Lokalisierung

Zur Beseitigung des zuvor beschriebenen Lokalisierungsfehler wird ein weiteres Konzept zur Schätzung der aktuellen Pose zwischen dem Sensor und den Markern herangezogen. Die Marker-Schnittpunkt Lokalisierung benutzt im Gegensatz zur Marker-Mittelpunkt Lokalisierung alle Laserpunkte, welche auf die reflektierenden Marker treffen, um Pose in Relation zum Referenzsystem M zu berechnen. Dazu basiert die Methode auf der Approximation linearer Gleichungen, durch welche die reflektierenden Marker angenähert werden sollen. Die Geraden lassen sich durch folgende Gleichungen beschreiben:

$$y_r = a_r x_r + b_r \tag{9}$$

$$y_r = a_l x_r + b_l \tag{10}$$

Die Parameter a_r, a_l, b_r, b_l können zur Beschreibung der linearen Gleichungen mit Hilfe der Laserpunkte mittels linearer Regression berechnet werden:

$$\begin{aligned} b_r &= \frac{\sum_{i=1}^{n_r}(x_{r,i}-\overline{x_r})(y_{r,i}-\overline{y_r})}{\sum_{i=1}^{n_r}(x_{r,i}-\overline{x_r})^2}, \qquad a_r = \overline{y_r} - b_r\overline{x_r} \\ b_l &= \frac{\sum_{i=1}^{n_l}(x_{l,i}-\overline{x_l})(y_{l,i}-\overline{y_l})}{\sum_{i=1}^{n_l}(x_{l,i}-\overline{x_l})^2}, \qquad a_l = \overline{y_l} - b_l\overline{x_l} \end{aligned} \tag{11}$$

Hierbei entspricht $x_{r,i}$ bzw. $y_{r,i}$ der x-Koordinate bzw. y-Koordinate des i-ten Laserpunktes auf dem rechten Marker und $x_{l,i}$ bzw. $y_{l,i}$ der x-Koordinate bzw. y-Koordinate des i-ten Laserpunktes des linken Markers. Des Weiteren entspricht $\overline{x_r}$ und $\overline{y_r}$ den Durchschnittswerten der x- und y-Koordinaten des rechten Markers. Folglich entsprechen $\overline{x_l}$ und $\overline{y_l}$ den Durchschnittswerten des linken Markers. Für die Bestimmung der Pose zwischen dem Sensor und dem Referenzsystem M wird bei der Marker-Schnittpunkt Lokalisierung der Schnittpunkt zwischen den zuvor bestimmten Geraden berechnet:

$$M = \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} = \begin{pmatrix} \frac{b_l - b_r}{a_r - a_l} \\ a_r(\frac{b_l - b_r}{a_r - a_l}) + b_r \\ 0 \end{pmatrix} \tag{12}$$

Hierbei ist darauf zu achten, dass die Pose des Referenzsystems M nicht mit dem Referenzsystem der Marker-Mittelpunkt Lokalisierung übereinstimmt. Die Position des Referenzsystems ist für die anschließende Positionierung ohne Relevanz, solange die Kenntnis über die verwendete Methode an die Positionierung weitergegeben wird. Die Koordinaten des Punktes M entsprechen, wie bei der Marker-Mittelpunkt Lokalisierung, der Translation der Transformation ${}^{S}T_M$. Neben der Translation sind Kenntnisse über die Orientierung bzw. der Rotation für die Bestimmung der Transformation ${}^{S}T_M$ notwendig. Die Rotation wird, genauso wie bei der Marker-Mittelpunkt Lokalisierung, durch einen Rotationswinkel, dem sogenannten Gierwinkel γ beschrieben. Dieser Gierwinkel kann aus den Schnittwinkeln der zuvor bestimmten Geraden berechnet werden:

$$\gamma_r = \text{atan}(|a_r|), \qquad \gamma_l = \text{atan}(|a_l|) \tag{13}$$

$$\gamma_{total} = \begin{cases} \frac{\gamma_r - \gamma_l}{2}, & b_r < 0, b_l > 0 \\ \frac{\gamma_l - \gamma_r}{2} + \frac{\pi}{2}, & b_r > 0, b_l > 0 \\ \frac{\gamma_l - \gamma_r}{2} - \frac{\pi}{2}, & b_r < 0, b_l < 0 \\ \frac{\pi}{4}, & \sum_{i=1}^{n_r}(x_{r,i}-\overline{x_r})^2 = 0 \\ -\frac{\pi}{4}, & \sum_{i=1}^{n_l}(x_{l,i}-\overline{x_l})^2 = 0 \end{cases} \tag{14}$$

Bei der Berechnung des Gier-Winkels γ_{total} muss ebenfalls, wie bei der Marker-Mittelpunkt Lokalisierung, zwischen verschiedenen Fällen unterschieden werden. Dies ist davon abhängig, wie der Sensor zu den Markern orientiert ist. Die unterschiedlichen Fälle sind in (14) definiert.
Die gesuchte Pose χ lässt sich anschließend wiederum aus der Transformation ${}^{R}T_M$ extrahieren. Ein funktionales Modell der Marker-Schnittpunkt Lokalisierung lässt sich der Abbildung 3 entnehmen.

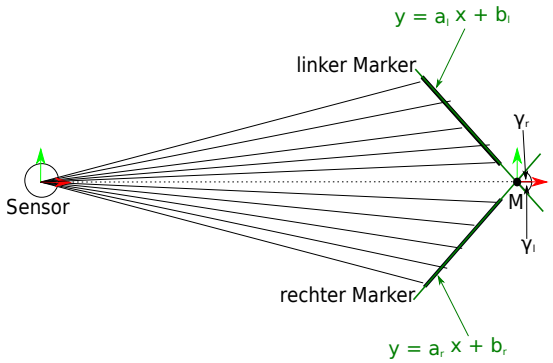


Abbildung 3 Funktionales Modell der Marker-Schnittpunkts Lokalisierung

5 FTS Positionierung

Neben der Bestimmung der aktuellen Pose des Roboters in Relation zum Referenzpunkt M, muss das FTS präzise an dem vorgegebenen Ziel positioniert werden. Dazu wird das Positionierungsmodul verwendet, welches mit Hilfe eines P-Reglers umgesetzt wurde. Dieser Regler berechnet zuerst die Differenz zwischen der aktuellen Pose χ und der Zielpose χ_{Ziel}, welche sich aus der Transformation ${}^{M}T_G$ extrahieren lässt. Der P-Regler nutzt die berechnete Differenz, um die Geschwindigkeitswerte für die x- und y-Richtung, sowie die Rotationsgeschwindigkeit $\dot{\theta}$ zu regeln:

$$\dot{\chi} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = f(\Delta\chi) = K_p \cdot \begin{pmatrix} x_{Ziel} - x \\ y_{Ziel} - y \\ \theta_{Ziel} - \theta \end{pmatrix} \tag{15}$$

Für die Regelung der Geschwindigkeitswerte $\dot{\chi}$ werden Regelparameter, welche in der Diagonalmatrix K_p definiert sind, als gegeben angenommen. Im Anschluss werden die berechneten Geschwindigkeitswerte $\dot{\chi}$ an die Robotersteuerung weitergegeben.

6 Validierung

Die Methoden zur Lokalisierung und Positionierung wurden mit dem *Robot Operating System*, ROS, für das fahrerlose Transportsystem TTS der TRAPO AG implementiert [7]. Beide Module lassen sich innerhalb von ROS eigenständig aufrufen, wobei die FTS-Lokalisierung als eigenständiger Knoten und die FTS-Positionierung als Plugin des ROS-Paketes *Move Base* implementiert ist [8].

6.1 Das TTS der TRAPO AG

Anwendung finden die Module zur präzisen Lokalisierung mit anschließender Positionierung in autonomen Systemen der TRAPO AG. Das TTS entspricht einem flexiblen, modularen FTS, welches mit Mecanum-Rädern, also einem omnidirektionalen Antrieb, ausgestattet ist [9]. Wird dem TTS ein neues Ziel vorgegeben, wird der Roboter zuerst mittels Pfadplannung in die Nähe des Ziels navigiert. Sobald sich das FTS in der Nähe des Ziels befindet, werden die zuvor beschriebenen Module zur präzisen Lokalisierung und Positionierung verwendet. Um die Korrektheit der Methoden zu Überprüfen, werden die Methoden im Folgenden mit Hilfe von simulativen, sowie realen LiDAR-Sensormesswerten validiert. Für die Validierung mit realen Sensormesswerten wurde ein FTS der TRAPO AG verwendet. Dieses FTS ist in der Abbildung 4 dargestellt.



Abbildung 4 Das TTS mit Rollenförderer-Modul

6.2 Ergebnisse

Die Validierung erfolgt mit simulierten Messwerten, wie auch mit realen Messwerten eines LiDAR-Sensors. Die Ergebnisse der Validierung werden im Folgenden erläutert.

6.2.1 Simulierte Bedingungen

Für die Validierung mittels simulierten LiDAR-Sensordaten wurde die Simulationsumgebung *Gazebo* für die Generierung der Messdaten des LiDAR-Sensors verwendet [6]. Um die Messwerte möglichst realistisch darzustellen, werden Messungenauigkeiten in die Simulation integriert. Dazu wurde auf die simulierten Messwerte eine Normalverteilung $N(\mu, \sigma^2)$ gelegt. Zur Validierung der Lokalisierungsmethoden wurde der Sensor aus unterschiedlichen Distanzen vor dem Markerpaar platziert. Die Entfernung reicht von circa $0,25$ Meter bis $2,5$ Meter. Die Markerpaare waren dabei in einem $90°$ Winkel zueinander platziert und der Winkel zwischen der Referenzebene und den Marker betrug jeweils $45°$. Des Weiteren wurde für die Validierung eine Markerbreite von 15 Zentimetern angenommen. Der Approximationsfehler der Position, welche für beide Methoden durch die Messungenauigkeiten des Sensors entsteht, ist für circa 500 Iterationen pro Methode und pro Distanz in Abbildung 5 dargestellt.

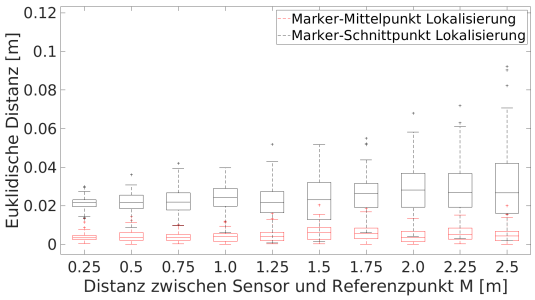


Abbildung 5 Ergebnisse der simulierten Positionsschätzung

Anhand der Boxplots in Abbildung 5 lässt sich erkennen, dass die Marker-Mittelpunkt Lokalisierung für die Approximation der Position robuster als die Marker-Schnittpunkt Lokalisierung gegen Messungenauigkeiten ist. Der zuvor beschriebene Fehler, welcher durch die Definition der Marker-Mittelpunkt Lokalisierung entsteht, hat demnach einen geringen Einfluss auf die Genauigkeit der Lokalisierung.
Für die Approximation des Gier-Winkels ist die Marker-Schnittpunkt Lokalisierung genauer und robuster gegen Messungenauigkeiten. Hierbei muss allerdings gesagt werden, dass die Ergebnisse der Marker-Schnittpunkt Lokalisierung von externen Faktoren, wie der Positionierung der reflektierenden Marker oder der Markerbreite abhängig sind. Des Weiteren ist auch der Winkel zwischen den Markern für die Genauigkeit der Marker-Schnittpunkt Lokalisierung ausschlaggebend. Die Ergebnisse der simulativen Validierung lassen sich Abbildung 6 entnehmen.

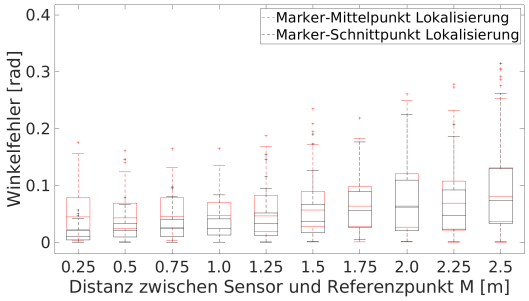


Abbildung 6 Ergebnisse der Schätzung des Gier-Winkels

6.2.2 Reale Bedingungen

Um das Verhalten der Methoden unter realen Bedingungen zu validieren, wurden des Weiteren reale LiDAR-Sensordaten zur Lokalisierung eines FTS verwendet. Da in diesem Fall die exakte Position des Roboters nicht bekannt ist, ist die Berechnung eines Fehlers nicht möglich. Dennoch können die zwei Methoden miteinander auf

Robustheit gegen Messunsicherheiten verglichen werden. Für den Vergleich wurde das Markerpaar in einer Entfernung von 1 Meter vor dem LiDAR-Sensor platziert. Die reflektierenden Marker, mit einer Markerbreite von circa 10 Zentimeter, wurden in einem 90° Winkel zueinander angeordnet. Anhand der Abbildung 7 lassen sich die zu-

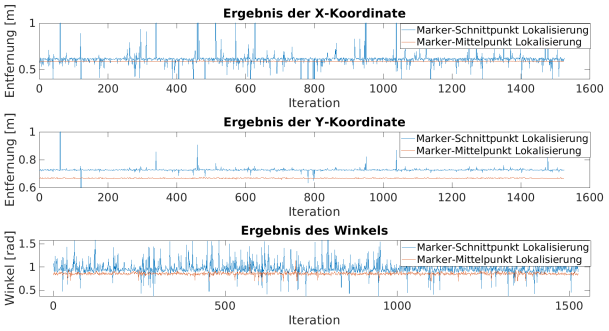


Abbildung 7 Ergebnisse der Lokalisierung mittels realen Messwerten

vor getroffenen Aussagen größtenteils bestätigen. Die Bestimmung der Position ist mit der Marker-Mittelpunkt Lokalisierung im Vergleich zur Marker-Schnittpunkt Lokalisierung robuster gegen Messungenauigkeiten. Mit realen Messwerten besitzt die Marker-Schnittpunkt Lokalisierung sowohl für die Positions-, sowie die Orientierung, größere Schwankungen als die Marker-Mittelpunkt Methode. Das kann daran liegen, dass externe Faktoren, wie zum Beispiel die Breite des Markers, einen Einfluss auf die Berechnung des Winkels mittels Marker-Schnittpunkt Lokalisierung haben. Aus diesem Grund ist die Marker-Mittelpunkt Methode der Marker-Schnittpunkt Methode zur Lokalisierung eines FTS für die zuvor beschriebenen Bedingungen zu bevorzugen. Die Ergebnisse der Validierung unter realen Bedingungen lassen sich der Abbildung 7 entnehmen.

7 Zusammenfassung

Eine präzise Indoor-Lokalisierung und -Positionierung von fahrerlosen Transportsystemem in kritischen Bereichen, wie zum Beispiel in der Nähe von Förderbändern, ist eine große Herausforderung der Industrie 4.0. Aufgrund hoher Flexibilitätsanspruche werden heutzutage flexible Lokalisierungsmethoden bevorzugt. Die Methoden bringen im Vergleich zu den statischen Methoden einen Genauigkeitsverlust mit sich, welcher in kritischen Situationen, wie zum Beispiel beim Be- und Entladen der Transportsysteme, zu Problemen führen kann. Aus diesem Grund werden gesonderte Methode in kritischen Bereiche, sogenannte Feinlokalisierer und -positionierer, eingesetzt. In dieser Arbeit wurden zwei Methoden zur präzisen Lokalisierung und Positionierung von FTS vorgestellt und auf Genauigkeit und Robustheit gegen Messungenauigkeiten miteinander verglichen. Beide Methoden benutzen zur Lokalisierung einen LiDAR-Sensor und reflektierende Marker als Umgebungsmerkmal. Anhand von Validierungen konnte gezeigt werden, dass die Marker-Mittelpunkt Lokalisierung der Marker-Schnittpunkt Lokalisierung zur präzisen Bestimmung der Pose eines FTS zu bevorzugen ist. Dabei ist anzumerken, dass die Genauigkeit der Methoden von externen Faktoren, wie zum Beispiel der Markerposition, der Markerbreite oder der Intensitäten der Umgebung, beeinflusst werden, sodass die Ergebnisse von diesen Faktoren abhängig sind.

8 Ausblick

Die beschriebenen Methoden zur präzisen Lokalisierung und Positionierung bieten Potenzial zur Verbesserung. Hierzu lässt sich zum Beispiel eine robustere Berechnung der Pose durch genauere Filter erreichen. Neben der Robustheit der Methoden ist ebenfalls die Anwendbarkeit von großer Bedeutung. Die Methoden lassen sich lediglich mit den Sensordaten eines Sensors ausführen. Eine Erweiterung auf mehrere Sensoren wäre wünschenswert, um die gesamte Umgebung des FTS in die Lokalisierung und Positionierung mit einzubeziehen.

9 Literatur

[1] Sheiner, M. & Bostelmann, R. (Hrsg.): *Literature Review of Mobile Robots for Manufacturing*. US Department of Commerce, National Institute of Standards and Technology, 2015, S. 8ff.

[2] Ullrich, G. (Hrsg.): *Fahrerlose Transportsysteme, Eine Fibel*. Wiesbaden: Springer-Verlag, 2014 ISBN 978-3-8348-2591-9, S. 4ff.

[3] Hesse, S. & Schnell, G. (Hrsg.): *Sensoren für die Prozess- und Fabrikautomation*. Plauen: Springer-Verlag, 2014 ISBN 978-3-658-05866-1, S. 25ff.

[4] Biswas, J. & Veloso, M. (Hrsg.): *Depth Camera Based Indoor Mobile Robot Localization and Navigation* 2012 IEEE International Conference on Robotics and Automation, RiverCentre, Saint Paul, Minnesota, USA, May 14-18, 2012

[5] Luimula, M. & Sääskilahti, K. & Partala, T. & Pieskä, S. & Alaspää J. *Remote navigation of a mobile robot in an RFID-augmented environment* Springer-Verlag, 06.06.2009, DOI 10.1007/s00779-009-0238-3

[6] Open Source Robotics Foundation: *Gazebo, Robot simulation made easy* http://gazebosim.org, aufgerufen am: 23.12.2020

[7] Quigley, M. & Conley, K. & Gerkey, B. & Faust, j. & Foote, T. & Leibs, J. & Wheeler, R. & Ng, A. Y. *ROS: an open-source Robot Operating System* 2009, ICRA Workshop on open source software, Band 3, Ausgabe 3.2, Seiten 5 http://robotics.stanford.edu/~ang/papers/icraoss09-ROS.pdf

[8] ROS Move Base: *Move Base* http://wiki.ros.org/move_base, aufgerufen am: 23.12.2020

[9] TRAPO AG: *TRAPO AG - We move your success* https://www.trapo.de/, aufgerufen am: 23.12.2020

Fahrbahnzustandskarten aus der cloudbasierten Informationsfusion von Kamera- und Fahrdynamikinformationen

Road Condition Maps from Cloud-based Information Fusion of Camera and Vehicle Dynamics

Alexander Busch[1], Tim Ahrenhold[2], apl. Prof. Dr.-Ing. Roman Henze[2], Prof. Dr.-Ing. Tobias Ortmaier[1], Dr.-Ing. Mark Wielitzka[1]

[1]Gottfried Wilhelm Leibniz Universität, Institut für Mechatronische Systeme, 30823 Garbsen, Deutschland, alexander.busch@imes.uni-hannover.de

[2]Technische Universität Braunschweig, Institut für Fahrzeugtechnik, 38106 Braunschweig, Deutschland, t.ahrenhold@tu-braunschweig.de

Kurzfassung

Der Reibwert zwischen Reifen und Fahrbahn bestimmt maßgeblich die fahrdynamischen Grenzen und eine vertrauenswürdige Schätzung trägt daher zur höheren Verkehrssicherheit bei. Ein vielversprechender Ansatz zur Bereitstellung prädiktiver Informationen in Form von Fahrbahnzustandskarten folgt aus der Fusion von Informationen ganzer Fahrzeugflotten mit Informationen der Fahrumgebung. Um eine Grundlage für entsprechende Cloudanwendungen zu schaffen, werden in dieser Arbeit Methoden und experimentelle Ergebnisse zur Erstellung von statischen Fahrbahnzustandskarten vorgestellt. Im durchgeführten Experiment wird eine Testbahn mehrfach mit einem Versuchsfahrzeug befahren und dabei gewonnene Informationen aus der Kamera und der Fahrdynamik mittels bayessche Statistik fusioniert.

Abstract

The coefficient of friction between tire and road significantly determines the driving dynamic limits and therefore a reliable estimation contributes to higher road safety. A promising approach to provide predictive information in form of road condition maps leads to the fusion of Information of vehicle fleets and environmental information. In order to provide a basis for corresponding cloud applications methods and experimental results for the development of static road condition maps are presented. In the realized experiment a test track is driven several times with a test vehicle and thereby information from the camera and the vehicle dynamics are fused using Bayesian statistics.

1 Einleitung

Der Reibwert zwischen Reifen und Fahrbahn bestimmt maßgeblich die fahrdynamischen Grenzen eines Fahrzeugs und eine vertrauenswürdige Schätzung kann bei einer Vielzahl heutiger Assistenzsystemen wie dem Notbremsassistent oder der Antriebsschlupfregelung zur höheren Verkehrssicherheit beitragen [1,2]. In der Zukunft ist insbesondere der den Reibwert beeinflussende Fahrbahnzustand als Teil des Umfeldmodells eine wesentliche Voraussetzung für autonomes Fahren [3].

Es gibt eine Vielzahl an Sensoren und Ansätze mit den jeweiligen individuellen Vor- und Nachteilen, um auf den Reibwert zu schließen [4]. Beispielsweise kann aus Fahrdynamiksensoren mit einem Beobachter eine Schätzung des maximalen Reibwerts realisiert werden, wobei die Verfügbarkeit einer vertrauenswürdigen Schätzung in hohem Maße von der vorliegenden Fahrdynamik abhängig ist [5]. Im Gegensatz dazu lassen sich mit einer Kamera zwar häufiger Informationen ableiten, jedoch zumeist mit geringerem Informationsgehalt zum Fahrbahnzustand. Beispielsweise können künstliche neuronale Netze oder Verfahren des maschinellen Lernens verwendet werden, um anhand von Bildern der Fahrzeugfrontkamera eine Klassifikation des Fahrbahnzustandes in Klassen wie trocken, nass oder verschneit vorzunehmen [6].

Es gibt bereits verschiedene Foschungsaktivitäten, die darauf aufbauend eine Fusion von heterogenen Reibwertinformationen auf Ebene des Fahrzeugs oder übergeordnet in einer Cloudanwendung anstreben. Eine multisensorbasierte Reibwerterkennung auf Fahrzeugebene mit einem Fuzzy-Expertensystem wird beispielsweise in [7] vorgestellt. Ein Ansatz unter Ausnutzung serienmäßiger Informationsquellen wie der FAS-Kamera, der Fahrdynamik und Scheibenwischerstatus und der Hinzunahme von Wetterinformationen wird in [4] vorgestellt. In einer Cloudanwendung lassen sich darüber hinaus örtlich- und zeitlich veränderliche Fahrbahnzustandskarten über eine Fusion heterogener Informationsquellen wie globaler Informationen zum Verkehr oder Wetter und Informationen aus der Fahrzeugsensorik einzelner Verkehrsteilnehmer generieren [4, 8]. Über die Vernetzung von Verkehrsteilnehmern mit dem Fahrzeugumfeld können prädiktive Informationen zum Fahrbahnzustand des künftigen Fahrbahnabschnittes bereitgestellt werden.

Das Projekt Road Condition Cloud (RoCCl) hat ebenfalls

zum Ziel, Methoden zur Erstellung von örtlich- und zeitlich veränderlichen Fahrbahnzustandskarten aus der Fusion von Informationen verschiedener Verkehrsteilnehmer zu entwickeln und diese simulativ und experimentell zu untersuchen. In dieser Arbeit werden experimentelle Ergebnisse zur Erstellung von statischen Fahrbahnzustandskarten aus der Fusion von Kamera- und Fahrdynamikinformationen präsentiert. Dabei wird eine Testbahn mit bekannten und konstanten Fahrbahnzuständen mit einem Versuchsfahrzeug mehrfach befahren und als Fahrbahnzustandskarte in der RoCCl abgebildet. Dazu wird in Kapitel 2 zunächst die Anwendung zur Auswertung von Kamerabildern basierend auf künstlichen neuronalen Netzen vorgestellt. In Kapitel 3 wird anschließend die verwendete Methode und der Beobachter zur Schätzung des Reibwerts aus der Fahrdynamik erläutert. In Kapitel 4 wird die entwickelte Simulationsumgebung zur RoCCl schematisch vorgestellt und die experimentellen Ergebnisse zur Fusion und zur Realisierung von statischen Fahrbahnzustandskarten präsentiert. Das Kapitel 5 schließt mit einer Zusammenfassung sowie einen Ausblick auf nachfolgende Forschungsarbeiten.

2 Klassifikation des Fahrbahnzustandes aus Kamerabildern

Dieser Abschnitt stellt zusammenfassend eine Methode zur Klassifikation des Fahrbahnzustandes aus Bilddaten mit tiefen faltenden künstlichen neuronalen Netzen dar, die auf Weiterentwicklungen eigener Vorarbeiten in [6, 9] beruhen. In den Vorarbeiten wurden verschiedene Bildbereiche und Netztopologien hinsichtlich der erzielbaren Genauigkeit sowie ein Pruning-Ansatz zur Reduktion der Modellkomplexität untersucht. Im Folgenden wird der zugrundeliegende und kontinuierlich weiterentwickelte Bilddatensatz vorgestellt, in Grundrissen die Methodik erläutert und Klassifikationsergebnisse für die im Rahmen dieser Arbeit verwendete Netztopologie präsentiert.

2.1 Datensatz

Der Bilddatensatz besteht aus einer individuellen Auswahl an Bildern aus öffentlich zugänglichen Bilddatensätzen wie beispielsweise KITTI [10], Cityscapes [11], Oxford Robotcar [12] und BDD100K [13], welcher um eigenes Bildmaterial sowie lizenzfreiem Bildmaterial von Videoplattformen ergänzt wurde. Der zusammengestellte Bilddatensatz umfasst insgesamt 35.424 Bilder und fünf Klassen: *trockener Asphalt* (AT), *nasser Asphalt* (AN), *trockenes Kopfsteinpflaster* (KT), *nasses Kopfsteinpflaster* (KN) und *Schnee* (S). Die **Tabelle 1** zeigt die Zusammensetzung des entwickelten Bilddatensatzes und die Zuordnung zu den einzelnen Klassen. Einschränkend bezüglich der Vollständigkeit hinsichtlich relevanter Klassen ist noch hinzuzufügen, dass Bilder von eisbedeckten Fahrbahnen nicht extrahiert werden konnten, womit eine mögliche Klasse Eis zum aktuellen Stand unberücksichtigt bleibt.
Weiterhin zeigt das **Bild 1** eine exemplarische Auswahl an Bildern verschiedener Bildquellen und Klassen sowie den trapezförmigen Bildbereich, welcher den Ausschnitt der Fahrbahn darstellt der rektifiziert als Eingangsbild zur Klassifikation verwendet wird.

Tabelle 1 Zusammensetzung des Bilddatensatzes

Bildquelle	AT	AN	AT	KN	S
KITTI	1.450	-	37	-	-
Robotcar	1.005	1.103	-	-	-
Cityscapes	5.492	29	331	-	-
BDD100K	4.546	1.713	51	-	127
Lizenzfreie Bilder	1.850	9	20	-	6.254
Eigene Bilder	1.502	2229	3.004	4.457	215
Gesamt	15.845	5.083	3.443	4.457	6.596

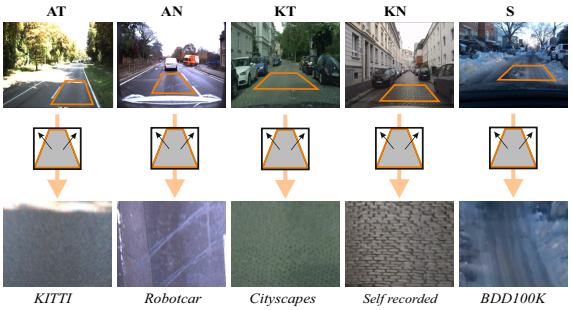


Abbildung 1 Exemplarische Bildauswahl für die Klassen: Asphalt trocken (AT), Asphalt nass (AN), Kopfsteinpflaster trocken (KT), Kopfsteinpflaster nass (KN) und Schnee (S) [9]

2.2 Methode und Ergebnisse

Zur Klassifikation des Fahrbahnzustandes wird ein SqueezeNet1.1 verwendet und anhand des Bilddatensatzes in Tabelle 1 trainiert. Der Datensatz in Tabelle 1 wird hierfür in einen Trainings-, Validierungs- und einen Testdatensatz eingeteilt. Dabei werden Bilder aufgenommen an verschiedenen Tagen, Monaten oder gar Jahren in verschiedene Datensätzen verteilt. Wenn möglich wird zusätzlich eine örtliche Trennung nach verschiedenen Städten oder auch nach verschiedenen Ortschaften vorgenommen. Während die lernbaren Parameter im Trainingsprozess anhand der Trainingsdaten optimiert werden, erfolgt die Auswahl eines Parametersatzes anhand der Validierungsdaten. Zusätzlich wird im Lernvorgang eine künstliche Datenvergrößerung des Trainingsdatensatzes vorgenommen, wobei jedes Bild in einem definierten Intervall zufällig gedreht, gespiegelt und verzerrt wird. Da ein Transferlernansatz verwendet wird, werden unterschiedliche Lernraten für die Merkmalsextraktion- und Klassifikationschichten gewählt. Durch die Wahl einer geringeren Lernrate für die Schichten der Merkmalsextraktion werden gelernte Merkmale grundsätzlich beibehalten und gegebenenfalls geringfügig weiterentwickelt.
Der vom Trainings- und Validierungsprozess unabhängige Testdatensatz dient abschließend zur Bewertung der im Lernvorgang bestimmten Parameter. Die am Testdatensatz mit dem SqueezeNet1.1 ausgewerteten Wahrheitsmatrix ist in **Bild 2** dargestellt. Es ist zu erkennen, dass mit einem F_1-Maßen von 94,32 % eine hohe Klassifikationsgüte auf

dem Testdatensatz erreicht wird. Die niedrigste Trefferquote mit 85,8 % wird bei der Klasse AN erzielt. Dabei wird die Klasse AN häufig als AT oder als S klassifiziert. Die höchste Trefferquote mit 97,1 % wird bei der Klasse AT erreicht. Zusätzlich zur prädizierten Klasse lassen sich dem neuronalen Netz ebenfalls die prädizierten Klassenwahrscheinlichkeiten für alle Klassen entnehmen und für eine übergeordnete Fusion nutzen. Es sei jedoch abschließend angemerkt, dass nach Meinung des Autors der zusammengestellte Datensatz noch nicht ausreichend Bilder umfasst, um alle im Straßenverkehr möglichen Umgebungsbedingungen und Fahrsituationen abzudecken.

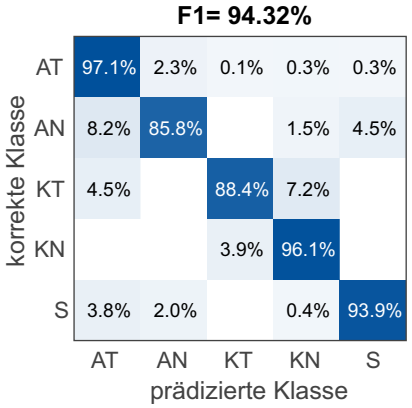


Abbildung 2 Wahrheitsmatrix des Testdatensatzes für das SqueezNet1.1

3 Reibwertschätzung aus der Fahrdynamik mittels UKF

In diesem Abschnitt wird der Beobachter und das zugrundeliegende Fahrdynamikmodell vorgestellt, mit dem die Schätzung der Wahrscheinlichkeitsdichte für den maximalen Reibwert auf Basis der Fahrdynamik erfolgt.

3.1 Methode

Zur Schätzung des maximalen Reibwerts aus der Fahrdynamik wird als Beobachter ein Unscented KALMAN-Filter (UKF) eingesetzt. Bei diesem rekursiven Filter-Algorithmus werden zunächst in einem Prädiktionsschritt die Systemzustände mit vorhandenem Modellwissen prädiziert und im anschließenden Korrekturschritt mit vorliegender Messung eine korrigierte Schätzung realisiert. Das zugrundeliegende Modell der Fahrzeugdynamik, in dem der maximale Reibwert $\mu_{\max}$ einen Modellparameter repräsentiert, ist ein nichtlineares Zweispurmodell in Zustandsraumdarstellung und ist in **Bild 3** dargestellt. Eine Schätzung von $\mu_{\max}$ wird über eine joint-Beobachterstruktur ermöglicht, in welcher der Zustandsvektor um $\mu_{\max}$ erweitert und die zugehörige Dynamik zu null angenommen wird. Der resultierende Zustandsvektor ergibt sich zu

$$x = (\dot{\psi}, \beta, \kappa, \dot{\kappa}, \omega_i, v_S, \mu_{\max})^T, \tag{1}$$

wobei $\dot{\psi}$ die Gierrate, β den Schwimmwinkel, κ den Wankwinkel, $\dot{\kappa}$ die Wankrate, ω_i die Winkelgeschwindigkeit des i-te Rades, v_S die Fahrzeugschwerpunktgeschwindigkeit beschreibt. Die Modelleingänge sind der Lenkradwinkel δ_L, das Motormoment M_M, die Getriebeübersetzung i_G und der Bremsdruck p_B. Die im Korrekturschritt verwendeten Messungen sind die Längs- und die Querbeschleunigung im Einbauort der Sensorik $a_{x,E}$ und $a_{y,E}$, die Winkelgeschwindigkeit ω_i der vier Räder und die Gierrate $\dot{\psi}$ und entsprechen den zur Verfügung stehenden Messungen im gängigen Serienstandard. Über das UKF lässt sich der Reibwert sowohl bei längs- als auch querdynamischen Fahrmanövern schätzen. Zusätzlich ist eine lokale Sensitivitätsanalyse zur Anregungsüberwachung implementiert, woraus sich ein sensitivitätsbasiertes-UKF (s-UKF) ableitet. Dabei wird in jedem Korrekturschritt die lokale Sensitivität des Reibwerts bezogen auf die Systemausgänge berechnet und bei Unterschreitung eines vom Anwender gewählten Schwellenwertes die Reibwertschätzung ausgesetzt. Hierdurch soll ein Driften der Schätzung bei schlechter Beobachtbarkeit verhindert werden.

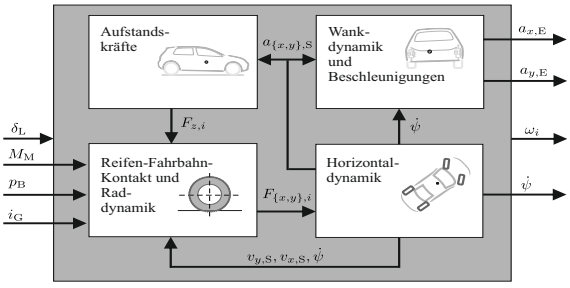


Abbildung 3 Modell der Fahrzeugdynamik

Innerhalb der Beobachterstruktur wird eine Schätzung der aktuellen Wahrscheinlichkeitsdichte $\mathcal{N}(\hat{x}_k, \hat{P}_{\hat{x}_k})$ unter der Annahme einer Normalverteilung mit dem Erwartungswert $\hat{x}_k$ und der zugehörigen Kovarianz $\hat{P}_{\hat{x}}$ definiert. Neben der Schätzung des Erwartungswertes kann die Kovarianz herangezogen werden, um eine Schätzung des Kredibilitätsintervalls $\left[\theta^-, \theta^+\right]$ zu realisieren. Dieses gibt an, dass der unbekannte wahre Wert x mit einer gewissen Wahrscheinlichkeit $p(\theta^- \leq x \leq \theta^+)$ im entsprechenden Intervall liegt. Die symmetrischen Intervallgrenzen ergeben sich zu

$$\theta^{\pm}_{k,j} = \hat{x}_{k,j} \pm \nu \sqrt{\hat{P}_{k,jj}}, \qquad j = 1, ..., n_x \tag{2}$$

mit einem Skalierungsfaktor ν, wobei für $\nu = 2$ die 2σ-Umgebung folgt.

3.2 Ergebnisse

Das **Bild 4** zeigt beispielhaft die Schätzung des Kredibilitätsintervalls für den maximalen Reibwert anhand eines Vergleichs vom UKF mit dem s-UKF bei einem querdynamischen Fahrmanöver. Es sei angemerkt, dass das Fahrdynamikmodell im dargestellten Fahrmanöver im Vergleich zum Modell aus Abschnitt 3.1 zusätzlich die Nickdynamik und instationären Querkraftaufbau berücksichtigt. Beim dargestellten Fahrmanöver sind mehrere Fahrbahnwechsel über die Fahrbahnuntergründe Asphalt, Beton, Kopfsteinpflaster und Stahl durchgeführt worden, wobei alle Fahr-

bahnen mit Ausnahme der asphaltierten Fahrbahn bewässert sind. Es ist die Schätzung des Erwatungswertes $\hat{\mu}_{\max}$ und die 2σ-Umgebungen beider Beobachter dargestellt. Es ist zu erkennen, dass mit beiden Beobachtern für die beiden Niedrigreibwert-Fahrbahnen Kopfsteinpflaster und Stahl hinsichtlich Literaturwerte plausible maximale Reibwerte geschätzt werden. Der Unterschied zwischen dem UKF und dem s-UKF wird dabei zu Beginn des Fahrmanövers deutlich. Während der verhältnismäßig geringen Fahrdynamik, wird im Vergleich zum UKF die Schätzung von $\hat{\mu}_{\max}$ beim s-UKF ausgesetzt. Zusätzlich ist die in Abhängigkeit der Fahrdynamik variierende 2σ-Umgebungen erkennbar. Insbesondere ist die Aufweitung auf den Fahrbahnuntergründen Asphalt und Beton ersichtlich, welche auf eine unsicherere Schätzung des Erwartungswerts hindeutet. Das dargestellte Fahrmanöver verdeutlicht, inwiefern die Schätzung der aktuellen Wahrscheinlichkeitsdichte mit einem s-UKF als Informationsquelle zur cloudbasierten Fusion dienen kann.

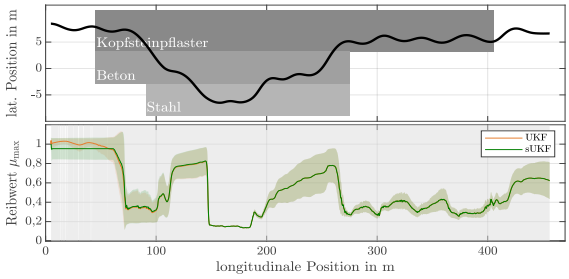


Abbildung 4 Reibwertschätzung und 2σ-Umgebung [5]

4 Fahrbahnzustandskarten aus der cloudbasierten Informationsfusion

Im Projekt Road Condition Cloud (RoCCl) ist eine Simulation für eine Cloudanwendung entwickelt worden, die die Erstellung und Bewertung von örtlich- und zeitlich veränderlichen Fahrbahnzustandskarten aus der Fusion von Informationen verschiedener Verkehrsteilnehmer und Umgebungsinformationen ermöglicht. Dabei können sowohl simulierte Fahrzeugflotten als auch experimentelle Messdaten berücksichtigt werden. Nachfolgend wird zunächst der schematische Aufbau der Simulationsumgebung erläutert und anschließend experimentelle Ergebnisse zur Erstellung von Fahrbahnzustandskarten aus der Fusion von Kamera- und Fahrdynamikinformationen präsentiert.

4.1 Road Condition Cloud

Der schematisch Aufbau Der Road Condition Cloud (RoCCl) ist in **Bild 5** dargestellt. Die RoCCl besteht aus vier modularen Blöcken: Datenaufnahme, Datenverarbeitung, Fusion und Entscheider. In der Datenaufnahme können eine beliebige Anzahl an Informationsquellen eines Fahrzeugs sowie mehrerer Fahrzeuge abgebildet werden. Dabei lassen sich sowohl experimentelle Messdaten aus Versuchsfahrten einbinden als auch mit hinterlegten Fahrdynamikmodellen eine Flottensimulation simulativ abbilden. In der Datenverarbeitung werden die entsprechenden Informationen individuell verarbeitet und in Form von Klassen und Klassenwahrscheinlichkeiten mit den zugehörigen Positionsinformationen an den Fusions-Block weitergeleitet. Im Fall experimenteller Messdaten kann die Datenverarbeitung auch auf entsprechender Hardware im Fahrzeug erfolgen. Im Fusionsblock sind verschiedene Fusionsalgorithmen implementiert, welche die Fusion der eingehenden Informationen mit denen in der RoCCl hinterlegten Informationen ermöglichen. Im Entscheider erfolgt die Aktualisierung, eine zeitliche Prädiktion sowie die Visualisierung der Fahrbahnzustandskarten. In der entwickelten Simulationsumgebung lassen sich verschiedene Untersuchungen unter Variation der Flottenteilnehmer, der Fusionsalgorithmen, der Reibwertklassen und der örtlichen sowie zeitlichen Auflösung der Fahrbahnzustandskarte durchführen.

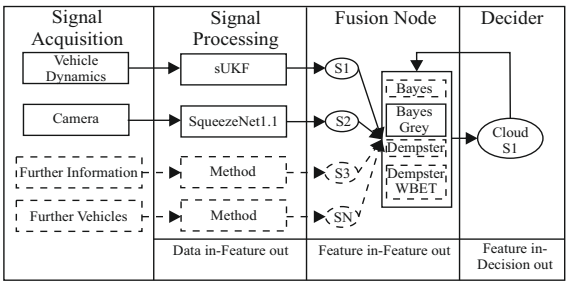


Abbildung 5 Schematische Darstellung der Cloudsimulation

4.2 Ergebnisse

In diesem Abschnitt wird eine Teilfunktionalität der RoCCl anhand von Fahrversuchen präsentiert. Dabei werden die im Versuchsfahrzeug mit dem Beobachter und dem SqueezeNet1.1 gewonnenen Informationen zum Fahrbahnzustand mittels Bayessche Wahrscheinlichkeitstheorie durch lokale Fusion in statische Fahrbahnzustandskarten überführt. Hierzu wird eine Testbahn mehrfach mit dem Versuchsfahrzeug befahren. Ein Teilbereich der Fahrbahn ist mit Kopfstein gepflastert und kontinuierlich bewässert. Der restliche befahrene Teil der Testbahn ist asphaltiert und trocken. Hieraus ergeben sich für den Zeitraum der Versuchsfahrten die definierten Fahrbahnzustände AT und KN. Eine zeitliche Veränderung bei fehlender Anregung wird in dieser Arbeit nicht vorgenommen.
Das **Bild 6** zeigt die resultierende Fahrbahnzustandskarte nach 6 Umrundungen mit dem Versuchsfahrzeug. Die Fahrbahnzustandskarte ist ausgenommen der Kurven in 10m Fahrbahnabschnitte unterteilt und beinhaltet 4 Reibwertklassen, wobei die zugehörigen Reibwertintervalle der Legende zu entnehmen sind. Die Klasse unbekannt wird angezeigt, wenn keine Informationen zum Fahrbahnzustand im jeweiligen Fahrbahnabschnitt vorliegen oder keine Reibwertklasse eine höhere Wahrscheinlichkeit als 50% aufweist. Die Transparenz gibt die Wahrscheinlichkeit einer Klasse wieder, wobei ein höherer Transparenzgrad stellvertretend für eine geringere Klassenwahrscheinlich-

keit ist. Es ist zu erkennen, dass alle Fahrbahnabschnitt der Klasse KN ($\mu_{\max} = 0,2 - 0,5$) nach 6 Runden in der RoCCl abgebildet werden. Ebenfalls werden größtenteils die Fahrbahnabschnitte der Klasse AT mit hoher Wahrscheinlichkeit richtig abgebildet. Lediglich ein Fahrbahnabschnitt auf der Geraden und wenige Fahrbahnabschnitte im Bereich der Kurven werden fälschlicherweise im Bereich $\mu_{\max} = 0,5 - 0,9$ geschätzt. An der Transparenz der blauen Farbe ist jedoch zu erkennen, dass für die Mehrheit dieser Bereiche eine hohe Unsicherheit vorliegt. Zusätzlich ist bei einigen Fahrbahnabschnitten auf der Geraden der Reibwertbereich unbekannt.

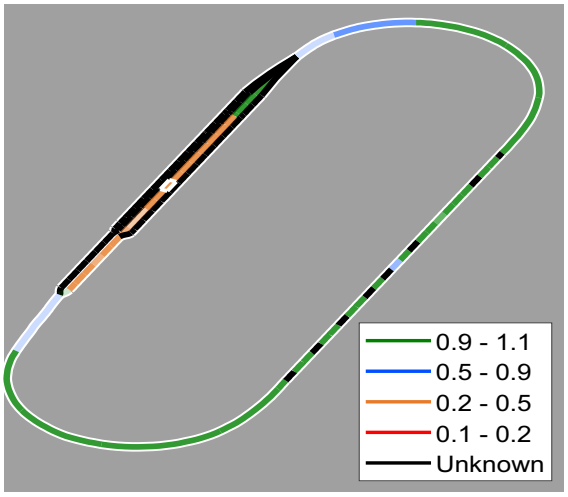


Abbildung 6 Fahrbahnzustandskarte nach 6 Umrundungen

Um einen Einblick in die Entstehung der Fahrbahnzustandskarte aus Bild 6 zu geben, zeigt **Bild 7** für den weiß markierten Fahrbahnabschnitt (KN) die Klassenwahrscheinlichkeiten für die RoCCl, die Kamera- und die Fahrdynamikinformationen sowie die zeitliche Abfolge der Fusionen. Die Klassenwahrscheinlichkeiten für die Kamerainformationen werden denen vom SqueezeNet1.1 prädizierten Klassenswahrscheinlichkeiten entnommen und den in Bild 6 vorgegeben Reibwertklassen zugeordnet. Die Klassenwahrscheinlichkeiten der Fahrdynamikinformationen werden abhängig von der Anregung bzw. der lokalen Sensitivität unterschiedlich bestimmt. Im Falle ausreichend hoher Anregung (lokale Sensitivität überschreitet Schwellenwert) ergeben sich die Klassenwahrscheinlichkeiten aus der vom s-UKF prädizierten Wahrscheinlichkeitsdichte $\mathcal{N}(\hat{\mu}_{\max}, \hat{P}_{\hat{\mu}_{\max}})$, welche entsprechend den Reibwertklassen in der RoCCl diskretisiert wird. Im Falle zu geringer Anregung (lokale Sensitivität unterschreitet Schwellenwert) werden die Klassenwahrscheinlichkeiten aus der Beschleunigung des Fahrzeugs nach [4] bestimmt. Dabei werden aus der vorliegenden kombinierten Beschleunigung Wahrscheinlichkeiten für die betrachteten Reibwertklassen abgeleitet. Die Informationen in der Cloud ergeben sich aus der Fusion der eingehenden Kamera- und Fahrdynamikinformationen sowie der zuvor gespeicherten Informationen in der Cloud mittels bayessche Statistik nach [14]:

$$(x|Y_1^1 Y_1^2 ... Y_1^n) = \frac{(x|Y_1^1)(x|Y_1^2)...(x|Y_1^n)(x|Y_0^1 Y_0^2 ... Y_0^n)}{(x|Y_0^1)(x|Y_0^2)...p(x|Y_0^n)}. \tag{3}$$

Das “p“ wird aus Darstellungsgründen weglassen, sodass sich beispielsweise $p(x|Y_1^1)$ zu $(x|Y_1^1)$ ergibt. Durch zusätzliche Normalisierung der zuvor genannten Gleichung lässt sich die Klassenwahrscheinlichkeit einer Klasse x für einen Fahrbahnabschnitt bestimmen. Dabei beschreibt Y_0^n eine vorangegangene und Y_1^n eine aktuelle Information einer Informationsquelle n. Es wird zusätzlich sowohl bei den Kamera-, den Fahrdynamikinformationen als auch in der Cloud eine maximale Klassenwahrscheinlichkeit von 90% festgelegt. Hierdurch soll eine eine Konvergenz gegen 100% ausgeschlossen werden und eine ausreichende Veränderlichkeit der Cloudinformationen sichergestellt werden. Der dargestellte Zeitraum umfasst etwa 700s, innerhalb dessen 6 Umrundungen und Fusionen durchgeführt worden sind. Der Vorteil verschiedener Informationsquellen wird bereits innerhalb der ersten Fusion deutlich. Während anhand der Kamera ähnliche Wahrscheinlichkeiten für die Bereiche $\mu_{\max} = 0,2 - 0,5$ und $\mu_{\max} = 0,5 - 0,9$ hervorgehen, wird aufgrund eines vorliegenden Assistenzsystemeingriffes über den Beobachter eine hohe Wahrscheinlichkeit für den Bereich $\mu_{\max} = 0,2 - 0,5$ prädiziert. Bei der anschließenden zweiten Durchfahrt und Fusion kann aufgrund einer Konstantfahrt keine Aussage über den Reibwert auf Basis der Fahrdynamik gemacht werden. Wohingegen aus den Kamerabildern über das SqueezeNet1.1 zu 90% der Bereich $\mu_{\max} = 0,2 - 0,5$ prädiziert wird. Dieser Umstand führt zur Erhöhung der entsprechenden Klassenwahrscheinlichkeit in der Cloud von etwa 68% auf 90%. Aufgrund der gespeicherten Informationen in der Cloud wird selbst bei widersprüchlichen Aussagen von Kamera und Fahrdynamik bei der 3. und 4. Fusion der wahre Bereich $\mu_{\max} = 0,2 - 0,5$ prädiziert. Es sei jedoch kritisch angemerkt, dass sowohl für die Kamera- als auch Fahrdynamikinformationen die durch die jeweiligen Methoden prädizierten Wahrscheinlichkeiten verwendet wurden. Diese werden häufig durch ein Neuronales Netz oder einen Beobachter überschätzt, sodass eine vermeintliche Sicherheit der entsprechenden Informationen suggeriert wird. Es lassen sich jedoch für jede Informationsquelle anhand von Testdatensätzen validierte Klassenwahrscheinlichkeiten für den Reibwert bestimmen.

5 Zusammenfassung

In dieser Arbeit wurden Ergebnisse des Projektes Road Condition Cloud (RoCCl) zur Erstellung von Fahrbahnzustandskarten aus der cloudbasierten Informationsfusion gezeigt. Diese umfassten experimentelle Ergebnisse zur lokalen Fusion von Kamera- und Fahrdynamikinformationen mittels Bayes’sche Wahrscheinlichkeitstheorie. Zur Auswertung der Kamera wurde ein SqueezeNet1.1 verwendet, welches den Fahrbahnzustand in 5 Klassen klassifiziert. Zur Auswertung der Fahrdynamik wurde ein s-UKF

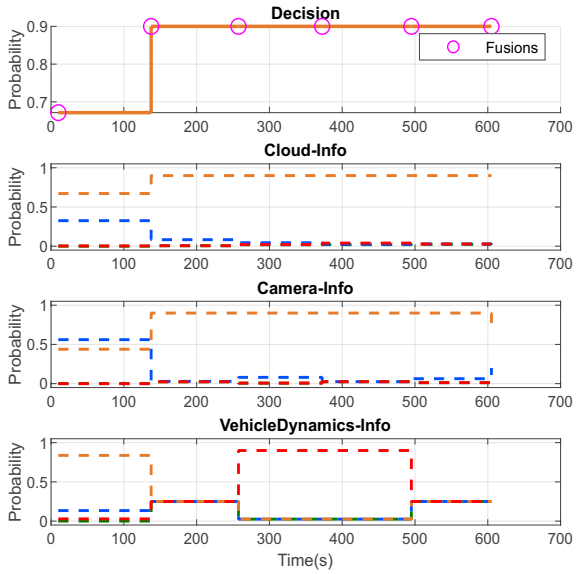


Abbildung 7 Ergebnisse der Fusion für den weiß markierten Fahrbahnabschnitt in Bild 6

mit zugrundeliegendem nichtlinearen Zweispurmodell verwendet. Es wurde mit einem Versuchsfahrzeug eine Testbahn mehrfach umrundet und die dabei gewonnene Fahrbahnzustandskarte präsentiert. Hierbei konnte gezeigt werden, dass mit den gewählten Informationen und Methoden nahezu alle Fahrbahnabschnitte dem richtigen Reibwertbereich für nassen Kopfsteinpflaster und trockenen Asphalt zugeordnet werden konnten. Lediglich vereinzelnd sind trockene Fahrbahnabschnitte als nasse Bereiche aus der Fusion hervorgegangen. Zusätzlich wurde für einen Fahrbahnabschnitt die zeitliche Abfolge der Fusionen dargestellt, wobei der Vorteil verschiedener Informationenquellen sowie mehrerer Umrundungen aufgezeigt werden konnte.

Zukünftig sind eine Vielzahl verschiedener experimenteller und simulativer Untersuchgen angedacht. In experimenteller Hinsicht erscheint die Hinzunahme weiterer Informationsquellen mit geringerem Informationsgehalt wie z.B. die Scheibenwischeraktivität und die Temperatur als sinnvoll. In der Simulation hingegen sind Auswertungen hinsichtlich der Anzahl an Verkehrsteilnehmer, der Reibwertklassen und der Fahrbahnabschnitte sinnvoll, um Aussagen über eine optimale örtliche Diskretisierung treffen zu können. Weiterer Forschungsbedarf besteht ebenfalls in der Realisierung zeitlich und örtlich prädiktiver Fahrbahnzustandskarten, welche eine fehlende Anregung durch eine Fahrzeugflotte und wechselnde Umweltbedingungen berücksichtigen.

Die Autoren danken der Deutschen Forschungsgemeinschaft (DFG) für die Förderung dieses Projekts.

6 Literatur

[1] Jarisa, W.; Henze, R.; Kücükay F.; Hartmann; B.: *Sensitivitätsanalyse von Fahrerassistenzsystemen in Bezug auf Fahrbahnzustände*. AAET 2016.

[2] Busch, A.; Kleyman, V.; Wielitzka, M.; Ortmaier, T.: *Adaptive model predictive traction control for electric vehicles*. In: 2019 18th European Control Conference (ECC). IEEE, 2019. S. 1239-1244.

[3] Hartmann, B.; Amthor, M.; Jarisa, W.: *Fahrbahnzustandserkennung als grundlegender Baustein für das Umfeldmodell*. Reifen-Fahrwerk-Fahrbahn (Internationale VDI-Tagung) 2015, ISSN 0083-5560, S. 339-356.

[4] Jarisa, W.: *FUTURE TECHNOLOGY–Road condition classification using information fusion.*. 7th International Munich Chassis Symposium 2016. Springer Vieweg, Wiesbaden, 2017.

[5] Wielitzka, M.: *Zustands- und Reibwertschätzung in der Fahrzeugdynamik*. Berichte des TEWISS Verlags. IMES 01/2020. ISBN: 978-3-95900-426-8.

[6] Busch, A.; Fink, D.; Laves; MH.; Ziaukas, Z.; Wielitzka, M.; Ortmaier T.: *Classification of Road Surface and Weather-Related Condition Using Deep Convolutional Neural Networks*. In: The IAVSD International Symposium on Dynamics of Vehicles on Roads and Tracks. Springer, Cham, 2019. S. 1042-1051.

[7] Trabelsi, A.: *Automotive Reibwertprognose zwischen Reifen und Fahrbahn*. FORTSCHRITT BERICHTE-VDI REIHE 12 VERKEHRSTECHNIK FAHRZEUGTECHNIK, 608.

[8] Panahandeh, G.; Ek, E.; Mohammadiha, N.: *Road friction estimation for connected vehicles using supervised machine learning*. In 2017 IEEE intelligent vehicles symposium (IV) (pp. 1262-1267). IEEE.

[9] Fink, D.; Busch, A.; Wielitzka, M.; Ortmaier T.: *Resource Efficient Classiffication of Road Conditions through CNN Pruning*. In: The IAVSD International Symposium on Dynamics of Vehicles on Roads and Tracks. Springer, Cham, 2019. S. 1042-1051.

[10] Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R.: *Vision meets robotics: The kitti dataset*. The International Journal of Robotics Research, 32(11), 1231-1237.

[11] Cordts, M.; et al.: *The cityscapes dataset for semantic urban scene understanding*. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, S. 3213-3223.

[12] Maddern, W.; Pascoe, G.; Linegar, C.; Newman, P.: *1 year, 1000 km: The Oxford RobotCar dataset*. The International Journal of Robotics Research, 2017, 36. Jg., Nr. 1, S. 3-15.

[13] Yu, F. et al.: *Bdd100k: A diverse driving video database with scalable annotation tooling*. arXiv preprint arXiv:1805.04687 Titel anhand dieser ArXiv-ID in Citavi-Projekt übernehmen, 2018, 2. Jg., Nr. 5, S. 6.

[14] Challa, S.; Koks, D.: *Bayesian and dempster-shafer fusion.*. Sadhana, 2004, 29(2), 145-174.

Systemidentifikation und modellbasierte prädiktive Pfadfolgeregelung eines automatisierten knickgelenkten Muldenkippers

System Identification and Model Predictive Path Following Control for an Automated Articulated Dump Truck (ADT)

Arash Shahirpour, M.Sc., Christian-Eike Framing, M.Sc., Dr.-Ing. René Zweigel, Univ.-Prof. Dr.-Ing. Dirk Abel
RWTH Aachen University, Institut für Regelungstechnik, Aachen, Deutschland, (a.shahirpour, c.framing, r.zweigel, d.abel)@irt.rwth-aachen.de

Kurzfassung

Knickgelenkte Muldenkipper werden aufgrund ihrer hohen Transportkapazität bei gleichzeitig guter Manövrierbarkeit im Bergbau vielfach eingesetzt. Aus Effizienz- und Sicherheitsgründen wird zukünftig ein automatisierter Betrieb dieser Fahrzeuge angestrebt. Aufgrund der besonderen Lenkkinematik und Totzeiten in der Aktorik sind übliche Ansätze zur Fahrzeugautomatisierung nicht ohne weiteres auf solche Fahrzeuge übertragbar. In diesem Artikel wird zunächst die spezielle Lenkkinematik des Fahrzeugs modelliert und die Systemverzögerungen mit experimentellen Daten identifiziert. Darauf aufbauend wird eine modellprädiktive Regelung (MPR) vorgestellt, die zwei Referenzpunkte auf der Vorder- und Hinterachse auf einem gegebenen Sollpfad stabilisiert. Der Vergleich zwischen der MPR und einem signalbasierten Regler zeigt, dass die MPR die Ergebnisse verbessert, indem sie keine Dauerschwingungen bei höheren Geschwindigkeiten und deutlich kleinere Abstandsfehler (bis zu 94% in manchen Kurven) aufweist. Aus Sicht der Autoren existiert im Stand der Technik kein knickgelenktes Fahrzeug, das mit der Berücksichtigung der Dynamik und Verzögerungen geregelt wird.

Abstract

Articulated Dump Trucks are one of the main means of transporting heavy goods in mining industries due to their capability to carry heavy loads while maintaining high maneuverability. In order to increase efficiency and security, the automated operation of these vehicles is desired. A big challenge on this way is that the conventional control algorithms, which are generally used for automated driving, cannot be directly used here. This is due to the special movement and the great transport delays (dead times) typical to this type of steering. In the course of this work, the steering kinematics are first modelled, and then the system delays are identified with the help of experimental data. Based on the model, a model predictive control (MPC) concept is introduced, which stabilizes two reference points on the front and rear axle on a given path, while considering the delays. A comparison between the MPC and a signal based controller shows how the MPC does not have remaining oscillations in higher speeds and achieves better distance errors (up to 94% in some cases). From the authors' point of view, there are no articulated vehicle control concepts in the literature that consider the delays.

1 Einleitung

Im BMBF-Verbundprojekt ARTUS wird eine autonome Flotte von knickgelenkten Bergbaufahrzeugen für hybride, d.h. über- und untertägige, umweltschonende und nachhaltige Gewinnung mineralischer Rohstoffe aufgebaut. Das Projekt trägt damit zu besseren Arbeitszeiten, erhöhter Effizienz und einer längeren Lebensdauer von Bergbaumaschinen und -prozessen und damit zur Ressourcenschonung bei. Eines der zu automatisierenden Fahrzeuge ist in **Abbildung 1** abgebildet.

Die Automatisierung knickgelenkter Fahrzeuge ist kein neues Forschungsthema und wurde z.B. schon im Jahr 1999 von Altafini [1] untersucht. In [1] modelliert der Autor das Fahrzeug in Frenet-Koordinaten, sodass die laterale und longitudinale Bewegung voneinander entkoppelt werden und separat steuerbar sind. Durch diese Art von Modellierung werden die beiden Abstände zwischen der Vorder- und Hinterachse und den jeweiligen Soll-Pfaden im Modell betrachtet. Nach der Linearisierung wird eine Zustandsrückführung als Regler verwendet. Während die Regelung gute Ergebnisse liefert, kann sie, wie alle nicht prädiktiven Regelungen, auf größere Knicke im Pfad erst nach deren Auftreten reagieren. Regelung mit Zustandsrückführung wird ebenfalls in den Arbeiten [2–4] für unterschiedliche Modellierungskonzepte angewandt.

Eine nichtlineare modellbasierte prädiktive Regelung (NMPR) wurde in [5] vorgeschlagen. Obwohl diese Regelung gute Ergebnisse liefert, kann sie wegen der hohen Rechenzeit nicht die Echtzeitanforderungen für einen Real-



Abbildung 1 Knickgelenkter Muldenkipper im Tagebau, Quelle: Framing / RWTH Aachen

betrieb erfüllen. In der hier vorgestellten Anwendung ist beispielsweise eine Stellgrößenberechnung mit 100 hz erforderlich.
Eine lineare modellbasierte prädiktive Regelung (MPR) wird in Beiträgen wie z.B. [6] und [7] vorgestellt. In diesen Artikeln werden allerdings die Totzeiten und Verzögerungen des Fahrzeugs nicht berücksichtigt.
Das in diesem Beitrag betrachtete knickgelenkte Bergbaufahrzeug weist zur Bewältigung der hohen Lasten eine robuste Bauweise auf. Zur Identifikation der Verzögerungen in der Knickgelenk- und Antriebsdynamik des Fahrzeugs wurden Testfahrten in einem Bergwerk durchgeführt. In der vorliegenden Arbeit werden die Dynamiken durch lineare Verzögerungselemente modelliert und anhand der Messdaten identifiziert (Kapitel 2). Die identifizierten Modelle werden einerseits genutzt um eine realitätsnahe Simulationsumgebung aufzubauen (Kapitel 3) und um andererseits den modellbasierten Regler zu entwerfen (Kapitel 4). Die resultierende MPR benutzt das linearisierte Modell und deshalb einen einfacheren Solver, der die Echtzeitanforderungen erfüllt. Die MPR soll in der Lage sein, eine Stabilisierung der Fahrzeugsreferenzpunkte, die auf beiden Achsen liegen, auf dem Soll-Pfad, ohne Dauerschwingungen und mit Berücksichtigung der Totzeit und Verzögerungen zu ermöglichen. In Kapitel 5 werden die Ergebnisse eines signalbasierten Reglers (Stanley-Regler) und der MPR verglichen.
In diesem Projekt wird der Soll-Pfad von einem vorgelagerten Modul geplant, das von dem Projektpartner indurad umgesetzt wird. Die dazu notwendige Umgebungserfassung und Befahrbarkeitsanalyse basiert im Gegensatz zu vielen anderen Automatisierungsansätzen ausschließlich auf Radarscannern in 2D und 3D, um einen Automatikbetrieb unter Schnee und Staub zu ermöglichen.

2 Systemidentifikation

Das hier betrachtete Fahrzeug ist über eine Schnittstelle mit zwei verschiedenen Stellgrößen ansteuerbar: die longitudinale Beschleunigung und die Knickwinkelgeschwindigkeit. Eine Vorgabe dieser Stellgrößen wird am Fahrzeug durch zwei unterlagerte Regelkreise umgesetzt, deren Parametrierung den Autoren nicht vorliegt. Um die entstehende Verzögerung im Antriebs- und Lenkverhalten zu modellieren, werden Verzögerungelemente 1. Ordnung (PT1) mit nachgeschalteter Totzeit verwendet. Für die Identifikation der Parameter wurden dem Fahrzeug Sprungfunktionen mit unterschiedlichen Amplituden vorgegeben. Die Messfahrten wurden mit sprunghaft ansteigenden und fallenden Stellgrößen, vom Stillstand und während der Fahrt mit unterschiedlichen Geschwindigkeiten und mit nacheinander einzeln angeregten Stellgrößen durchgeführt.
Um die Beschleunigungswerte bzw. die Knickwinkelrate zu finden, wurde die Geschwindigkeit bzw. der Knickwinkel zuerst mit dem smoothdata Befehl von MATLAB mit einem Gaussian-Weighted Moving Average Filter mit einem Fenster von 35 bzw. 20 Datenpunkte gefiltert und dann numerisch abgeleitet, weil die Werte nicht als Rohdaten vorhanden waren.

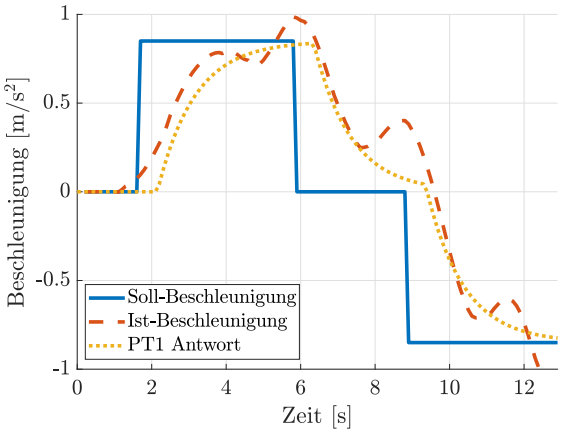


Abbildung 2 Identifikationsergebnisse - Beschleunigung

Anschließend wurde SystemIdentifikation Toolbox von MATLAB benutzt, um die Parameter der PT1-Elemente mit den Messdaten zu optimieren. Weil das Fahrzeug über einen internen Regler verfügt, der bei negativer Beschleunigung den Bremsdruck anpasst, wird das Bremsverhalten nicht separat identifiziert.

2.1 Beschleunigung

Das Fahrzeug zeigt bei der Beschleunigung kein einheitliches Verhalten und seine Verzögerung kann je nach Situation (bergauf, bergab, mit und ohne Beladung) variieren. Grund dafür ist, dass bei unterschiedlichen Fahrsituationen und Fahrbahneigenschaften unterschiedlich große Drehmomente bei den Rädern gebraucht werden, um eine bestimmte Beschleunigung zu erreichen. Die Getriebedynamik, die durch die Schaltung stark nichtlinear ist, verstärkt ebenfalls die Variation in Verzögerung bei unterschiedlichen Messfahrten.
Im Rahmen der Messfahrten wurde das Fahrzeug nur auf ebener Fahrbahn ohne Beladung mehrmals getestet und das durchschnittliche Verhalten wurde für die Identifikation genommen. Das identifizierte Verhalten kann, falls nötig, nach weiteren Tests um die Abhängigkeit von Beladung und Bodensteigung erweitert werden.
Das Beschleunigungsverhalten zeigt ebenfalls keine einheitliche Totzeit, d.h. die Totzeit variiert zwischen 0,45 und 0,55 s. Als Durchschnittswert wird daher 0,50 s als Totzeit angenommen. Die Tests zeigen weiterhin, dass auf ebener Fahrbahn die maximale Beschleunigung $a_{max} = 0{,}78\,\mathrm{m/s^2}$ beträgt. Aus der Identifikation ergibt sich folgende Beschleunigungsübertragungsfunktion:

$$\frac{a_{ist}}{a_{soll}} = \frac{1}{s+T_a} e^{-T_{t_a} s}, \tag{1}$$

bei der a die Beschleunigung ist. Für die Zeitkonstante gilt $T_a = 1\,\mathrm{s}$ und für die Totzeit $T_{t_a} = 0{,}5\,\mathrm{s}$. **Abbildung 2** zeigt die Soll-Beschleunigung und vergleicht den Ist-Wert mit der Antwort der Übertragungsfunktion.

2.2 Knickwinkelgeschwindigkeit

Verglichen mit dem Beschleunigungsverhalten ist für die Vorgabe der Knickwinkelgeschwindigkeit eine konsistentere Dynamik bei verschiedenen Steigungen und Beladungen zu beobachten. Obwohl das Lenkverhalten bei manueller Lenkung von Beladung und der Bodensteigung abhängig ist, ergibt sich bei Verwendung der unterlagerten autonomen Lenkregelung ein konsistentes Verhalten. Grund dafür ist, dass die Knickwinkelgeschwindigkeit beim autonomen Betrieb auf 12° 1/s begrenzt ist und immer erreicht werden kann, ohne dass der Öldruck zuerst stark aufgebaut werden muss, während sie beim manuellen Betrieb bis 20° 1/s zunehmen kann allerdings unterschiedliche Verzögerungen und Totzeiten abhängig von Motordrehzahl und Beladung zeigt. Deshalb kann dieses Verhalten auch anhand einer einzelnen Übertragungsfunktion modelliert werden.
Analog zum Beschleunigungsverhalten ist auch für die Knickwinkelgeschwindigkeit eine variierende Totzeit zu beobachten. Der Öldruck für das hydraulische Knickgelenk hängt direkt von der Motordrehzahl ab. Er wird erst dann aufgebaut, wenn eine Lenkbewegung oder eine andere Fahrzeugfunktion mit hydraulischer Hilfsenergie angefragt wird. Abhängig von der bereits anliegenden Motordrehzahl variiert die Totzeit bis zur Umsetzung einer angefragten Lenkbewegung. Hier wurde nach mehreren Tests 0,6 s als Durchschnittswert für die Totzeit identifiziert. Die Übertragungsfunktion resultiert zu:

$$\frac{\omega_{ist}}{\omega_{soll}} = \frac{K_\omega}{s+T_\omega} e^{-T_{t_\omega} s}. \tag{2}$$

Hier gilt für die Zeitkonstante $T_\omega = 2,\mathrm{s}$, für die Verstärkungsfaktor $K_\omega = 2$ und für die Totzeit $T_{t_a} = 0{,}6\,\mathrm{s}$. Dabei bezeichnet ω_{soll} die angefragte und ω_{ist} die am Gelenk vorliegende Knickwinkelgeschwindigkeit. In **Abbildung 3** ist ein Vergleich der gemessenen und der identifizierten Dynamik für eine Sprungvorgabe dargestellt.

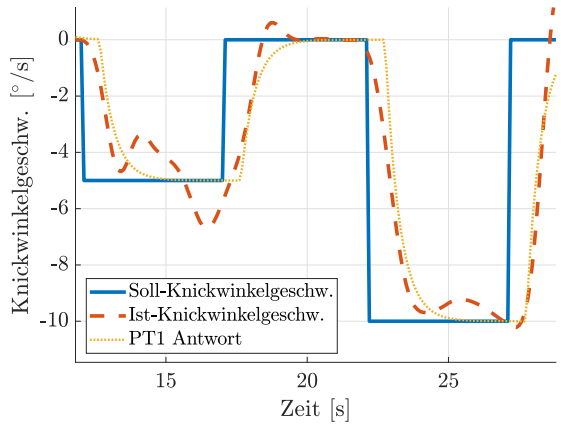


Abbildung 3 Identifikationsergebnisse - Knickwinkelgeschwindigkeit

3 Modellierung eines knickgelenkten Fahrzeugs

3.1 Kinematisches Modell im lokalen Bezugssystem

Zur Modellierung des Fahrzeugverhaltens in der Simulation wurde ein kinematisches Modell basierend auf [9] gewählt (siehe **Abbildung 4**):

$$\dot{x} = v_V \cos\psi_V, \tag{3a}$$
$$\dot{y} = v_V \sin\psi_V, \tag{3b}$$
$$\dot{\psi}_V = \frac{\sin\phi}{L_H + L_V\cos\phi} v_V + \frac{L_H}{L_H + L_V\cos\phi}\omega, \tag{3c}$$
$$\dot{\phi} = \omega. \tag{3d}$$

Dabei sind x und y die Koordinaten des Mittelpunktes der Vorderachse im lokalen Bezugssystem. ψ_V und v_V sind der Gierwinkel und die Geschwindigkeit des Vorderkörpers. L_V bzw. L_H sind der Abstand zwischen der Vorder- bzw. Hinterachse und dem Knickpunkt. ϕ und ω beschreiben den Knickwinkel und die Knickwinkelgeschwindigkeit. Alle Winkel sind in Richtung der Pfeile in Abbildung 4 positiv.

3.2 Kinematisches Modell in Frenet-Koordinaten

Durch die Modellierung in Frenet-Koordinaten werden die laterale und die longitudinale Bewegung voneinander entkoppelt und separat steuerbar. Deshalb wird dieses Modell als Reglermodell benutzt. Für die Regelung wird allerdings nur die laterale Bewegung betrachtet.
Morin et al. modellieren in [10] einen mobilen Roboter mit drei Rädern in Frenet-Koordinaten. Das gleiche Prinzip wird hier für das knickgelenkte Fahrzeug angewandt.
Ein Pfad wird als bekannt vorausgesetzt. Der Punkt auf dem Pfad, der den kleinsten Abstand zur Vorderachse des Fahrzeugs hat, wird als V bezeichnet. V existiert und ist eindeutig, solange sich das Fahrzeug in hinreichender Nähe zum Pfad befindet (vgl. Abbildung 4). Um das kinematische Verhalten zu beschreiben, müssen noch zwei Variablen definiert werden. Die resultierende Gleichung für die zweite Variable wird direkt im Modell benutzt.

- s ist das Kurvenstück im Punkt V und wird mit der folgenden Gleichung berechnet (siehe [10]):

$$\dot{s} = v_V \frac{\cos\Delta\psi_V}{1 - d_V\kappa_V}. \tag{4}$$

- $\Delta\psi_V = \psi_V - \psi_{V_{Pfad}}$ ist die Orientierung des Vorderkörpers relativ zu der Tangente des Pfades im Punkt V.

Dabei ist v_V die Geschwindigkeit der Vorderachse und d_V der Abstand zwischen der Vorderachse und dem Punkt V. Die Krümmung, der Kehrwert des Kurvenradius, im Punkt V auf dem Sollpfad (κ_V) kann laut Definition mit:

$$\kappa_V(s) = \frac{\partial\psi_{V_{Pfad}}}{\partial s} \tag{5}$$

oder mit:

$$\dot{s}\kappa_V = \dot{\psi}_{V_{Pfad}} \tag{6}$$

beschrieben werden. Der Vorteil der Herleitung mit der Krümmung ist, dass ihr Wert für Geraden 0 ist, während der Radius für Geraden gegen ∞ geht. Nun kann mithilfe der Gleichungen (3c) und (4) die Gleichung für die Orientierung $\Delta\psi_V$ neu formuliert werden:

$$\begin{aligned}\Delta\dot{\psi}_V &= \dot{\psi}_V - \dot{\psi}_{V_{Pfad}} \\ &= \dot{\psi}_V - \dot{s}\kappa_V(s) \\ &= \frac{v_V \sin\phi + L_H\omega}{L_H + L_V\cos\phi} - \dot{s}\kappa_V(s) \\ &= \frac{v_V \sin\phi + L_H\omega}{L_H + L_V\cos\phi} - v_V\kappa_V\frac{\cos\Delta\psi_V}{1-d_V\kappa_V}.\end{aligned} \tag{7}$$

Gleichung (7) beschreibt zusammen mit der folgenden Gleichung aus [10] das kinematische Verhalten des Vorderkörpers:

$$\dot{d}_V = v_V \sin\Delta\psi_V. \tag{8}$$

Analog können Gleichungen (7) und (8) für den Hinterkörper aufgestellt werden, so dass sich das folgende nichtlineare System für die Beschreibung des kinematischen Verhaltens des Gesamtfahrzeugs ergibt:

$$\begin{bmatrix} \dot{d}_V \\ \Delta\dot{\psi}_V \\ \dot{d}_H \\ \Delta\dot{\psi}_H \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} v_V \sin\Delta\psi_V \\ \frac{v_V\sin\phi + L_H\omega}{L_H + L_V\cos\phi} - v_V\kappa_V\frac{\cos\Delta\psi_V}{1-d_V\kappa_V} \\ v_H\sin\Delta\psi_H \\ \frac{v_V\sin\phi - L_V\cos\phi\,\omega}{L_H + L_V\cos\phi} - v_H\kappa_H\frac{\cos\Delta\psi_H}{1-d_H\kappa_H} \\ \omega \end{bmatrix} =: \boldsymbol{f}(\cdot). \tag{9}$$

4 Pfadfolgeregelung

4.1 Stanley-Regler

Stanley ist der Name eines autonomen Fahrzeugs der Stanford Universität, das an der DARPA Grand Challenge im Jahr 2005 teilgenommen und gewonnen hat [8]. Dieser Regler berücksichtigt weder die Kinematik, noch die Totzeit und Verzögerungen und kann deshalb mit wenig Aufwand auf dem Fahrzeug implementiert und getestet werden. Hier wird der Regler zu Vergleichszwecken benutzt.

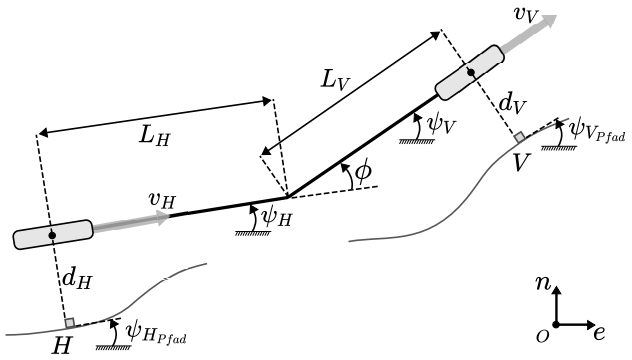


Abbildung 4 Kinematik eines knickgelenkten Fahrzeugs

Die eingesetzte Pfadfolgeregelung basiert auf einer nichtlinearen Rückführung des Abstands d_V zwischen der Vorderachse und dem Soll-Pfad (siehe Abbildung 4). Gleichung (10) zeigt das geschwindigkeitsabhängige Stellgesetz der Lenkung.

$$\phi(t) = (\psi_V - \psi_{V_{Pfad)}}) + \arctan\frac{k d_V(t)}{v_V(t)}, \tag{10}$$

k ist als Gain-Parameter fahrzeugabhängig einzustellen. $\psi_{V_{Pfad}}$ entspricht dem Winkel der Pfadtangente an dem der Vorderachse nächstliegenden Punkt V.

4.2 Modellbasierte prädiktive Regelung

Die MPR setzt zur Prädiktion des Systemverhaltens ein linearisiertes Einspurmodell ein, das um die identifizierte Dynamik erweitert ist und auf dem Modell im Kapitel 3.2 basiert. Die MPR berücksichtigt die Krümmung des Soll-Pfades als zukünftige bekannte Störgröße und optimiert die Knickwinkelgeschwindigkeit, um die Abstände zwischen der Vorder- und Hinterachse und dem Soll-Pfad zu minimieren.
Ein weiterer Vorteil bei der Modellierung in Frenet-Koordinaten ist, dass das angestrebte Regelungsziel einer guten Pfadfolge direkt der Ruhelage des Systems $\boldsymbol{f}(\cdot) = \boldsymbol{0}$ entspricht. Deshalb können die Matrizen des linearisierten Systems für $\boldsymbol{x}_R = \boldsymbol{0}$, $u_R = 0$ und $\boldsymbol{z}_R = \boldsymbol{0}$ berechnet werden. Die Zustandsraumdarstellung für das lineare zeitvariante System lautet (vgl. Gleichung (9)):

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}u + \boldsymbol{E}\boldsymbol{z}, \tag{11}$$

$$\boldsymbol{x} = \begin{bmatrix} d_V \\ \Delta\psi_V \\ d_H \\ \Delta\psi_H \\ \phi \end{bmatrix},\ u = \omega,\ \boldsymbol{z} = \begin{bmatrix} \kappa_V \\ \kappa_H \end{bmatrix}. \tag{12}$$

Dabei ist $\boldsymbol{x}$ der Zustandsvektor, u der Eingang und $\boldsymbol{z}$ der Vektor der Eingangsstörungen. Die Systemmatrizen ergeben sich aus der Linearisierung:

$$\begin{aligned}\boldsymbol{A} &= \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}_R,u_R,\boldsymbol{z}_R},\ \boldsymbol{b} = \frac{\partial \boldsymbol{f}}{\partial u}\Big|_{\boldsymbol{x}_R,u_R,\boldsymbol{z}_R}, \\ \boldsymbol{E} &= \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}}\Big|_{\boldsymbol{x}_R,u_R,\boldsymbol{z}_R}.\end{aligned} \tag{13}$$

Nachdem das Modell zeitlich diskretisiert wird, kann die Kostenfunktion der MPR im zeitdiskreten Fall wie folgt definiert werden:

$$J = \sum_{k=1}^{N} \boldsymbol{x}_k^T \boldsymbol{Q}\boldsymbol{x}_k + \sum_{k=0}^{N-1} r u_k^2. \tag{14}$$

$\boldsymbol{Q}$ und r bei der Gleichung (14) bezeichnen die Gewichtsmatrix für die Systemzustände bzw. den Gewichtsparameter für die Stellgröße und N ist die Länge des diskreten Prädiktionshorizonts.
Die MPR fasst die Pfadkrümmung über den Prädiktionshorizont als bekannte Störung auf und bekommt in jeder Iteration die Krümmung der nächsten N Punkte auf dem

Pfad, die Tv voneinander entfernt sind, wobei T die Abtastzeit der MPR und v die Geschwindigkeit des entsprechenden Körpers ist. Das heißt, dieser Wert beträgt Tv_V für den Vorder- und Tv_H für den Hinterkörper.
Die Systemgleichungen können zur Berücksichtigung der Stelldynamik um die Gleichung (2) erweitert werden.
Um auch die Totzeit in den Stellgrößen zu kompensieren, wird in der MPR der geschätzte Fahrzeugzustand in 0,6 Sekunden (vgl. Gleichung (2)) eingesetzt. Dafür wird der aktuelle Zustand mithilfe von Gleichung (3) numerisch integriert. Für die Integralrechnung wird die Geschwindigkeit konstant angenommen und für die Knickwinkelgeschwindigkeit die über die letzten 0,6 Sekunden gerechneten Werten genommen.

5 Versuchsaufbau und Ergebnisse

5.1 Versuchsaufbau

In der Simulation kann gezeigt werden, ob der entwickelte Regler in der Lage ist, das Fahrzeug mit hinreichender Genauigkeit einem vorgegebenen Soll-Pfad mit konstanter Geschwindigkeit folgen zu lassen. Dafür werden für das Fahrzeugmodell die Systemgleichungen (3) zusammen mit den Identifikationsergebnissen (Gleichungen (1) und (2)) in MATLAB/Simulink implementiert.
Für die Länge des diskreten Prädiktionshorizonts wird $N = 20$ genommen und das Modell der MPR hat die Abtastzeit $T = 0.2$ Sekunden.
Als Teststrecke werden 3 Kurven benutzt. Die erste Strecke ist ein Kreis mit einem Radius von 16 m, die zweite Strecke ist ein gleichseitiges Rechteck mit einer Seitenlänge von 60 m und die dritte Strecke ist eine Sinus-Kurve mit einer Amplitude von 15 m und einer Periode von 66,6 Sekunden. Jede Strecke wird mit zwei Geschwindigkeiten $v_1 = 7{,}2$ km/h und $v_2 = 18$ km/h gefahren.

5.2 Ergebnisse

Abbildungen 5 und **6** zeigen die Simulationsergebnisse für den Stanley-Regler und die MPR. Hier werden die Abstandsfehler für Vorder- und Hinterachse bei Kurvenverläufen über der Zeit darstellt. Das Fahrzeug hat in allen Fahrten die konstante Geschwindigkeit von 7,2 km/h auf Abbildung 5 und 18 km/h auf Abbildung 6.
Das erste Bild in Abbildung 5 zeigt den Abstandsfehler während einer Kreisfahrt. Die Oszillationen im Abstandsfehler treten auf, da der Soll-Pfad nur eine Näherung der Kreisform aus Geradenstücken darstellt. Aus den Verläufen mit MPR ist ersichtlich, dass die Mitte der Hinterachse auf der kurvenäußeren und die Mitte der Vorderachse auf der kurveninneren Seite des Soll-Pfads geführt werden. So kann die Quadratsumme des Abstandsfehlers über den Prädiktionshorizont minimiert werden. Die Verläufe mit Stanley-Regler zeigen, dass dort beide Achsmittelpunkte auf der kurvenäußeren Seite und in einem größeren Abstand zum Pfad liegen.
Das zweite Bild in Abbildung 5 zeigt eine Fahrt mit einem rechteckigen Soll-Pfad. Da die vorausliegende Pfadkrümmung der MPR als Störgröße über den Prädiktionshorizont bekannt ist, wird der maximale Abstandsfehler verglichen zum Stanley-Regler stark reduziert.
Bei dem dritten Bild in Abbildung 5 wird eine Sinus-Kurve als Soll-Pfad vorgegeben. Hier ist nochmals ein geringerer Abstandsfehler bei der MPR verglichen zum Stanley-Regler erkennbar.
Während der Stanley-Regler bei kleinerer Geschwindigkeit noch vergleichbare Ergebnisse mit der MPR liefert, wird der größere Abstandsfehler bei einer Geschwindigkeit von 18 km/h deutlicher, was auf der Abbildung 6 zu erkennen ist.
Abbildung 7 zeigt den simulierten Fahrzeugpfad bei Vorgabe eines sinusförmigen Soll-Pfades und bei einer Geschwindigkeit von 18 km/h. Hier ist genauer zu sehen , dass das Verhalten des Stanley Reglers für echte Anwendung nicht akzeptabel ist, weil er unter anderem große und unvorhersagbare Abweichungen bewirkt.

6 Zusammenfassung und Ausblick

In diesem Beitrag wurde das dynamische Verhalten eines realen knickgelenkten Fahrzeugs bei Vorgabe von Lenkbewegungen und Beschleunigungen mithilfe von PT1-Elementen modelliert. Deren Parameter wurden aus gesammelten Messdaten identifiziert. Das so bestimmte dynamische Übertragungsverhalten wurde danach in ein kinematisches Modell in Frenet-Koordinaten integriert. An-

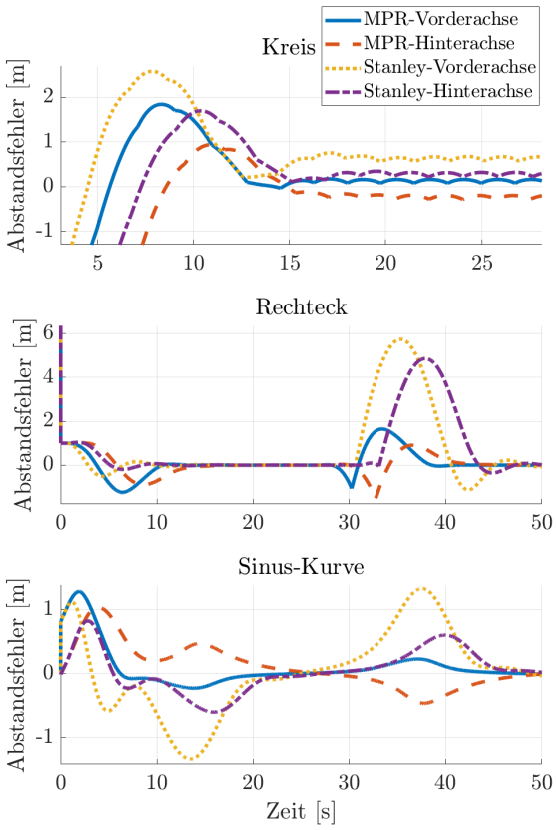


Abbildung 5 Simulationsergebnisse für i) Kreis ii) Rechteck iii) Sinus-Kurve, mit 7,2 km/h

schließend wurde zur hochgenauen Fahrzeugregelung eines knickgelenkten Fahrzeugs eine MPR entwickelt und implementiert, die unter Berücksichtigung der Stelldynamik die Abstände der Achsen zum Soll-Pfad minimiert.
Die Ergebnisse stellen dar, dass die MPR das Regelungsziel eines kleinen Abstandsfehlers und minimales Überschwingens bei sprunghaften Krümmungsänderungen erfüllen kann. Weil die Teststrecken so entworfen wurden, dass sie möglichst reelle Fahrsituationen widerspiegeln, kann jetzt festgestellt werden, dass die MPR in der Lage ist, mit dem echten Fahrzeug getestet zu werden.
Um dynamische Effekte besser untersuchen zu können, wird im nächsten Schritt eine erweiterte Simulationsumgebung entwickelt. Diese basiert auf einem dynamischen Zweispurmodell, bei dem auch Schlupfeffekte modelliert werden. Dynamisches Fahrverhalten soll auch in der MPR berücksichtigt werden. Dazu wird ein dynamisches Einspurmodell integriert. Das Ziel ist auch bei höheren Geschwindigkeiten, bei denen Schlupfeffekte nicht mehr vernachlässigt werden können, eine hohe Regelgüte und Stabilität zu haben.

7 Danksagung

Dieser Beitrag entstand im Rahmen des durch das Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekts ARTUS (FKZ 033R126DN).

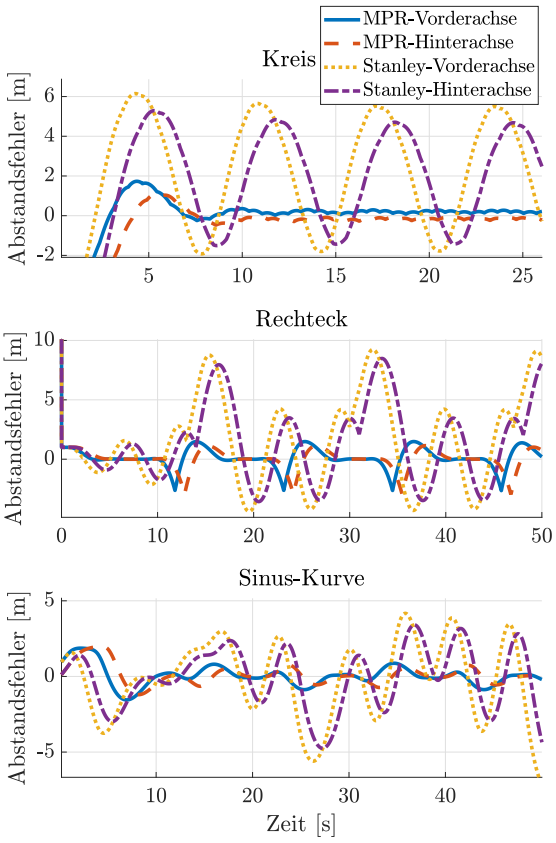


Abbildung 6 Simulationsergebnisse für i) Kreis ii) Rechteck iii) Sinus-Kurve, mit 18 km/h

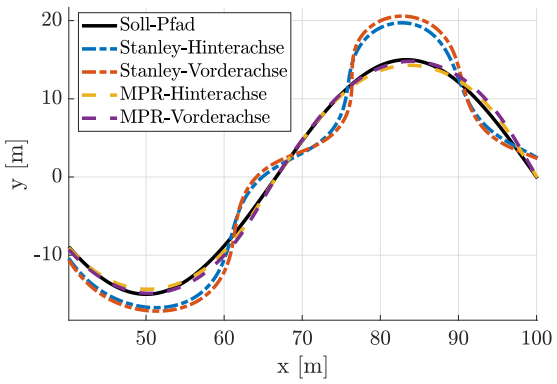


Abbildung 7 Sinus-Kurvenfahrt mit 18 km/h

8 Literatur

[1] Altafini, C.: *A Path-Tracking Criterion for an LHD Articulated Vehicle*. The International Journal of Robotics Research 18, no. 5 (1999): S. 435-441.

[2] Ridley, P.; Corke, P.: *Load Haul Dump Vehicle Kinematics and Control*. Journal of Dynamic Systems, Measurement, and Control 125, no. 1. (2003), S. 54-59.

[3] Hemami, A.; Polotski, V.: *Path Tracking Control Problem Formulation of an LHD Loader*. The International Journal of Robotics Research 17, no. 2. (1998), S. 193-199.

[4] DeSantis, R. M.: *Modeling and Path-Tracking for a Load-Haul-Dump Mining Vehicle*. Journal of Dynamic Systems, Measurement, and Control. (1997), S. 40-47.

[5] Bai, G.; Liu, L.; Meng, Y.; Weidong, L.; Gu, Q.; Ma, B.: *Path Tracking of Mining Vehicles Based on Nonlinear Model Predictive Control*. Applied Sciences 9, no. 7. (2019), S. 1372.

[6] Nayl, T; Nikolakopoulos, G; Gustfsson, T.: *Switching Model Predictive Control for an Articulated Vehicle under Varying Slip Angle*. In: 20th Mediterranean Conference on Control and Automation 2012 (MED), pp. 890-895. IEEE, 2012.

[7] Bai, G.; Meng, Y.; Liu, L.; Luo, W.; Gu, Q.; Li, K.: *A New Path Tracking Method Based on Multilayer Model Predictive Control*. Applied Sciences 9, no. 13. (2019), S. 2649.

[8] Thrun, S.; Montemerlo, M.; Dahlkamp, K,; Stavens, D.; Aron, A.; Diebel, J.; Fong, P.; et al.: *Stanley: The Robot That Won the DARPA Grand Challenge*. Journal of field Robotics 23, no. 9. (2006), S. 661-692.

[9] Delrobaei, M.; Kenneth, A.: *Design and Steering Control of a Center-Articulated Mobile Robot Module*. Journal of Robotics. (2011).

[10] Morin, P.; Samson, C.: *Motion Control of Wheeled Mobile Robots*. Springer handbook of robotics 1. (2008), S. 799-826.

Mechatronisches System für die modellbasierte Datenfusion zur Ermittlung des psychologischen Fahrerzustands

Mechatronic System for Model-Based Data Fusion to Determine the Cognitive Driver State

Alexander Hugenroth, HELLA GmbH & Co. KGaA, E-LAB, 59552 Lippstadt, Deutschland, alexander.hugenroth@tu-dortmund.de

Dr.-Ing. Alexander Warnecke, HELLA GmbH & Co. KGaA, E-LAB, 59552 Lippstadt, Deutschland, alexander.warnecke@hella.com

Univ.-Prof Dr.-Ing. Prof. h.c. Dr. h.c. Torsten Bertram, TU Dortmund, Lehrstuhl für Regelungssystemtechnik, 44221 Dortmund, Deutschland, torsten.bertram@tu-dortmund.de

Kurzfassung

Die Anzahl der Fahrerassistenzsysteme in modernen Fahrzeugen steigt stetig. Hierbei gibt es viele Systeme, die den Fahrer warnen können, wie zum Beispiel vor einem Fahrzeug im toten Winkel oder Müdigkeit. Jedoch sind Systeme, die dahingegen die Aufmerksamkeit des Fahrers ermitteln, noch nicht verbreitet. Dieser Beitrag zeigt auf, wie durch einen Sensoraufbau und dazugehörige Algorithmen der psychologische Fahrerzustand der Langzeitaufmerksamkeit (Vigilanz) bestimmt werden kann. Dazu werden in dieser ersten Machbarkeitsanalyse die Daten des Sensoraufbaus durch ein künstliches neuronales Netzwerk fusioniert. Durch das im Netzwerk abgebildete Modell, wird auf den psychologischen Zustand des Fahrers geschlossen. Der ermittelte Fahrerzustand kann zukünftig für einen sicheren Umgang mit den Automatisierungsstufen, wie SAE Level 3, genutzt werden. Zudem kann eine Rückmeldung an den Fahrer erfolgen, sodass dieser sich seiner Situation bewusst wird oder der Fahrerzustand könnte durch eine Regelung beeinflusst werden.

Abstract

The number of driver assistance systems in modern vehicles is increasing. There are many systems that can warn the driver, for example of a vehicle in the blind spot or fatigue. However, systems that detect the driver's attention are not yet widespread. In this contribution we will investigate how a sensor setup and the corresponding programming can be used to determine the psychological driver state of long-term attention, i.e. vigilance. For this purpose, in this initial feasibility analysis, the data from the sensor setup will be fused by an artificial neural network. The psychological state of the driver is then inferred by the model mapped in the network. The determined driver state can be used in the future for a safe handling of automation levels, such as SAE Level 3. In this way, feedback to the driver is also possible, so that the driver becomes aware of his situation or the state could be influenced by a control loop.

1 Einleitung

Mit modernen Fahrzeugen geht meistens ein Gewinn in Sicherheit und Komfort einher. Dabei haben sich die Sicherheits- und Komfortstandards immer weiter durch eine Automatisierung und Assistenzsysteme erhöht. Hierbei spielt neben der sensorgestützten Umfelderfassung auch die Fahrerbeobachtung eine wesentliche Rolle. Bisher liegt bei der Fahrerbeobachtung der Fokus auf einer Müdigkeitserkennung. Diese wird überwiegend durch die Analyse von Fahrereingaben durchgeführt, wie von [5] gezeigt. Jedoch fällt diese Vorgehensweise durch eine erhöhte Automatisierung der Fahraufgabe vermutlich weg. Dazu kommen neue Aufgaben auf die Fahrerbeobachtung zu, da sich die Automatisierung nicht nur negativ auf die Müdigkeit sondern auch auf die Aufmerksamkeit auswirkt. So kann davon ausgegangen werden, dass die Reize für den Fahrer immer geringer und schwächer werden. Bei einer längeren Aufmerksamkeitsaufgabe (mehr als 30 min) treten schwache und seltene Reize auf, wie [7] definiert hat. So kann psychologisch zwischen dauerhaft aufmerksam (vigilant) und unaufmerksam (hypovigilant) bei einer solchen Aufgabe unterschieden werden. Diese Unterscheidung wird in dem vorliegenden Fall durch eine Datenfusion in einem nichtlinearen Modell klassifiziert. Da der Ansatz auf künstliche neuronale Netzwerke als Modell für die Klassifizierung setzt, werden Daten aus einem Versuch mit neun Probanden herangezogen. Eine genauere Beschreibung des Versuchs sowie der zugehörigen Algorithmen wird in einem Beitrag in der Zeitschrift Forschung im Ingenieurwesen mit dem Titel *Untersuchung und Klassifikation der Fahreraufmerksamkeit bei längerer partiell automatisierter Fahrt* im Nachgang der vorliegenden Veröffentlichung erfolgen. In Zukunft soll durch ein solches System die Sicherheit im Fahrzeug erhöht werden. Zum einen wird es möglich Warnungen an den Fahrer abzugeben, wenn dieser unaufmerksam wird, zum anderen könnte das Fahrzeug direkt Einfluss auf die ermittelte Aufmerksamkeit des Fahrers nehmen, indem die Aufgabenverteilung zwischen Fahrzeug und Fahrer anders gewichtet verteilt oder der Fahrer durch Nebentätigkeiten grundsätzlich aufmerksam gehalten wird.



Abbildung 1 Sensoraufbau

2 Sensoraufbau

Um für die Bestimmung der Vigilanz des Fahrers verschiedene Körperfunktionen erfassen zu können, wird ein Sensorsystem benötigt. Ein Überblick über mögliche Sensoren in einem Fahrzeug wird in [6] gegeben. Bei [8] werden verschiedene Messmethoden für die Müdigkeit zusammengetragen, sodass sich auch in dem vorliegenden Beitrag der Sensoraufbau an diesen Beispielen orientiert.
Bei dem hier vorliegenden Beitrag wird eine Kombination aus verschiedenen Sensoren verwendet. Dabei stehen sowohl kontaktlose als auch kontaktbehaftete Sensoren zur Messung der physiologischen Merkmale zur Verfügung. In Abbildung 1 werden die Positionen der Sensoren angegeben. Es werden zwei Kameras auf dem Armaturenbrett, ein Sensor am Brustkorb und ein Sensor an der Stirn der Versuchsperson in dem Sensoraufbau verwendet. In der Abbildung 1 sind die kontaktlosen Kameras mit roter und die kontaktbehafteten Sensoren mit grüner Markierung farblich hervorgehoben.

2.1 Elektroenzephalogramm

Von der Gehirnaktivität ausgehend lassen sich Aktivitäten in der Hirnrinde (Kortex) durch an der Schädeloberfläche auftretende Spannungsunterschiede (Gehirnwellen) bestimmen. Für die Messung der Gehirnwellen wird ein Elektroenzephalogramm (EEG) benutzt. In [1] wird beschrieben, dass diese Messung die besten Ergebnisse für eine Fahrerzustandseinschätzung gibt. Dabei wird darauf hingewiesen, dass der Sensor kontaktbehaftet ist und zusätzlich ein Kontaktgel verwendet wird, sodass die Versuchsperson abgelenkt bzw. beeinträchtigt werden könnte. Um bei der vorliegenden Studie die Messung möglichst nichtinvasiv zu halten, wird ein kabelloses Stirnband mit trockenen Elektroden benutzt. Das verwendete Stirnband ist mit vier Elektroden und einer Referenzelektrode ausgestattet. Dies ist ausreichend, da keine qualitative Aussage über die örtliche Herkunft von Gehirnwellen gegeben werden soll. Wenn eine starke Änderung der Gehirnwellen vorliegt, dann kann von verschiedenen Zuständen des Gehirns ausgegangen werden, wie [12] schon für wenige Elektroden zur Müdigkeitserkennung von Zugfahrern zeigt und verifiziert. Im vorliegenden Fall wird das Frequenzspektrum der Spannungswerte in fünf überlappende Frequenzbänder [3] aufgeteilt:
Delta (δ) [1-4 Hz], Theta (θ) [4-8 Hz], Alpha (α) [7.5-13 Hz], Beta (β) [13-30 Hz] und Gamma (γ) [30-44 Hz]. Ebenso werden auch Kombinationen der Frequenzbänder als Merkmal für die Erkennung von Müdigkeit [4] herangezogen. So werden in dem vorliegenden Beitrag ebenso neben den oben beschriebenen Frequenzbändern auch folgende Kombinationen von Frequenzbändern verwendet: $\frac{\delta}{\alpha}$, $\frac{\alpha}{\beta}$, $\frac{\delta+\alpha}{\alpha+\beta}$ und $\frac{\delta}{\beta}$

2.2 Elektrokardiogramm

Da ein Elektrokardiogramm (EKG) über elektrische Signale vom Herzen gemessen wird, muss dieses mit einem kontaktbehafteten Sensor durchgeführt werden. Hier wird ebenso wie bei dem EEG ein kabelloser Sensor verwendet, um keine Störungen durch Verkabelungen zu erhalten. Dieser Sensor misst sowohl die Herzfrequenz (HF) in Herzschläge pro Minute, als auch die Zeit zwischen den einzelnen Herzschlägen (R-R-Intervallen) r_i, die für die Berechnung der Herzratenvariabilität H über die Zeitspanne $T = 60$ s verwendet wird:

$$H = \sqrt{\sum_{i=0}^{T} (r_{i+1} - r_i)^2} \tag{1}$$

Die Messwerte (HF, r_i) werden mit 1 Hz ausgegeben, wobei die Zeit zwischen einzelnen Herzschlägen r_i gegebenenfalls unter eine Sekunde betragen kann, sodass mehrere Messwerte gleichzeitig ausgegeben werden können.

2.3 Kamera

Um eine kontaktlose Beobachtung des Fahrers zu erhalten, werden, zwei Kameras verwendet. Eine ist, wie in Abbildung 1 zu sehen, hinter dem Lenkrad und die zweite an der A-Säule auf der Beifahrerseite auf dem Armaturenbrett befestigt. Die Kameras nehmen Farbbilder mit einer Auflösung von 1920 mal 1080 Pixel mit 30 Bildern pro Sekunde auf. Um die Informationen über das Auge des Fahrers zu extrahieren, wird im ersten Schritt der in [11] beschriebene Algorithmus verwendet. Dieser liefert Schlüsselpunkte der Augen in Pixelkoordinaten, mit denen dann die Breite des linken Auges b_l und des rechten Auges b_r sowie die Höhe des linken Auges h_l und des rechten Auges h_r jeweils als Pixelabstand bestimmt werden können. Aus diesen Werten wird dann wie folgt das Höhen- zu Breitenverhältnis r eines Quadrupels bestimmt:

$$r = \frac{\frac{h_l}{b_l} + \frac{h_r}{b_r}}{2} \tag{2}$$

Beispielhaft ist in Abbildung 2 das Höhen- zu Breitenverhältnis r einer Versuchsperson abgebildet, wobei in diesem Fall der Wert bei geöffnetem Auge etwa 0.3 beträgt und bei geschlossenem Auge bei etwa 0.07 liegt. Um Messausreißer aus dem Signal r zu entfernen, werden alle Messwerte, die mehr als 0.05 nach oben, da nur Augenschließungen in diesem Fall interessant sind, von dem Mittelwert (in dem gezeigten Fall $\varnothing r = 0.27$) abweichen, auf den Mittelwert gesetzt. Dieses Signal ist jedoch immer noch verrauscht, sodass ein Tiefpassfilter mit einer Grenzfrequenz von 15 Hz (halbe Abtastfrequenz) angewendet wird, wel-

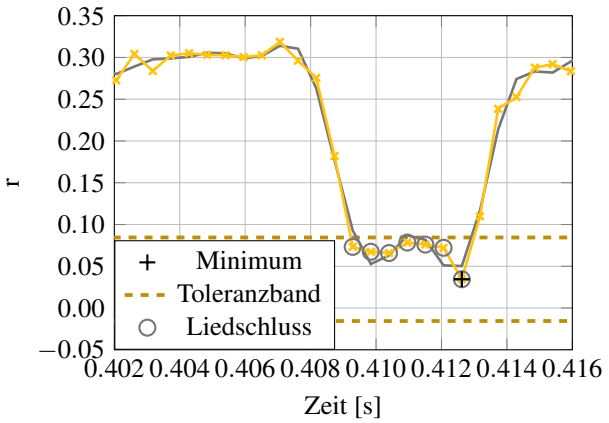


Abbildung 2 Höhen- zu Breitenverhältnis für eine Augenschließung und Bestimmung der Schließzeit

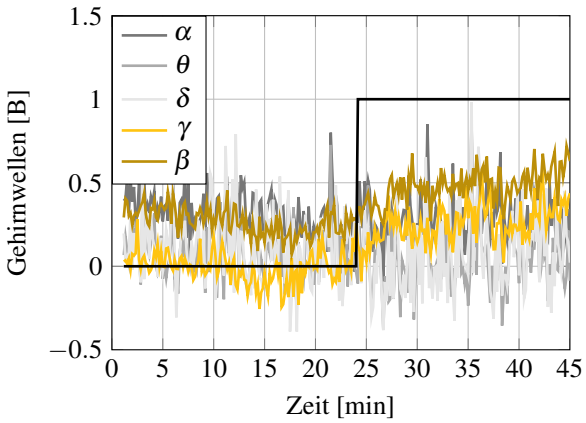


Abbildung 3 Gehirnwellen einer Versuchsperson mit eingezeichneter Beschriftung (Ground Truth)

ches durch die graue Linie in der Abbildung 2 visualisiert wird. Das Filter dient dazu, besser die zeitlichen Minima zu finden, um eine Augenschließung von Messrauschen zu unterscheiden. Dazu werden nur Minima mit einer Schartenhöhe von mindestens 0.07 als Augenschließung gezählt, sodass keine Nebenminima durch das Rauschen hinzukommen. So wird ein Minimum nur gezählt, wenn der Wert r vor und nach dem Minimum um mehr als 0.07 steigt.
Um die Schließdauer zu bestimmen, werden von dem Minimum nach links und rechts alle Messpunkte markiert, die in einem Toleranzband von 0.05, wie in der Abbildung 2 gezeigt, um den Wert des Minimums liegen. Zuletzt wird noch die Anzahl der markierten Messpunkte durch die Anzahl der Messungen in dem Intervall von 60 s bestimmt, um die prozentuale Schließdauer, also den PERCLOS-Wert, wie in [9] zu erhalten.

3 Datenauswertung

Da die Größe der Stichprobe in der vorliegenden Studie sehr gering ist, wird an dieser Stelle das methodische Vorgehen zur Datenverarbeitung und eine grundsätzliche Aussagekraft der Vorgehensweise analysiert. So wird trotz der geringen Größe des Datensatzes ein künstliches neuronales Netzwerk vorgestellt und trainiert, um eine Datenverarbeitung für einen größeren Datensatz vorzubereiten. Die Datenauswertung wird offline umgesetzt und dient zur Analyse, ob und mit welchen Sensoren eine Erkennung des psychologischen Zustands möglich ist. In den folgenden Abschnitten wird beschrieben, wie die Daten der verschiedenen Sensoren zusammengeführt werden, um sie dann mit einem künstlichen neuronalen Netzwerk zu klassifizieren. Hierbei soll im Fokus stehen, welches Klassifikationsergebnis des psychologischen Zustands (Vigilanz) bei verschiedenen Kombinationen von Sensoren, also verschiedenen Merkmalen, auftritt. Dazu werden beim Training die Daten einzelner Sensoren aus der Menge der Trainingsdaten genommen und die Ergebnisse auf ungesehenen Testdaten miteinander verglichen.

3.1 Datenvorbereitung

Da die Sensoren unterschiedliche Abtastraten aufweisen, muss der Datensatz auf eine gemeinsame Zeitskala mit definierten Abständen synchronisiert werden. Dabei wird eine Abtastzeit von 10 s gewählt. Diese wird so gewählt, um eine praktikable Balance zwischen den Merkmalen des EEGs mit einer Abtastzeit von 100 ms und den Merkmalen PERCLOS sowie H mit einer Berechnung über 60 s, also einer Abtastzeit von 60 s, zu schaffen. Um die Messwerte für diese festen Zeitschritte zu bestimmen, wird eine lineare Interpolation der Messdaten zwischen benachbarten Zeitpunkten durchgeführt. Das bedeutet, dass die Messung einer Versuchsperson über 45 Minuten (2700 s) 270 Datenpunkte erzeugt. Da in der ersten Minute zu Versuchsbeginn noch starke Bewegungsartefakte das EEG stören, werden diese Messwerte gelöscht, sodass 264 Datenpunkte weiter verarbeitet werden können. Für das verwendete überwachte Lernverfahren ist es notwendig, die Daten in die verschiedenen Klassen (Vigilant und Hypovigilant) einzuordnen.
In diesem Fall werden die Gehirnwellen für eine Einordnung benutzt, um eine erste Variante der Datenverarbeitung zu erproben. Dabei wird bei einer großen Veränderung der β- und γ-Wellen auf einen Zustandswechsel geschlossen. So kann in Abbildung 3 das Vorgehen an einem Beispiel aus den Trainingsdaten betrachtet werden, bei dem die Gehirnwellen ab etwa Minute 24 eine größere Veränderung aufweisen, da die β- und γ-Wellen im Vergleich zu den anderen Gehirnwellen ansteigen. So wird in diesem Fall auch die Kennzeichnung (Ground Truth) des kognitiven Fahrerzustands in schwarz dargestellt. In diesem Zeitpunkt ändert sich der Zustand von aufmerksam (Vigilant) zu unaufmerksam (Hypovigilant) und zeigt so, dass hier ein Zustandswechsel vorliegt. Die weiteren Daten werden nach dem gleichen Prinzip manuell beschriftet, was aufgrund der individuellen Unterschiede nicht immer präzise durchgeführt werden kann. So kann bei einigen Probanden der Zeitpunkt des Zustandswechsels zeitlich nur grob bestimmt werden, da die Änderungen der Gehirnwellen bei jeder Versuchsperson unterschiedlich stark ausgeprägt sind.

Tabelle 1 Merkmale für das Training der Datenverarbeitung

	Körperteil	Merkmal
1	Gehirn	δ
2		θ
3		α
4		β
5		γ
6		$\frac{\delta}{\alpha}$
7		$\frac{\alpha}{\beta}$
8		$\frac{\delta+\alpha}{\alpha+\beta}$
9		$\frac{\delta}{\beta}$
10	Herz	HF
11		H
12	Augen	PERCLOS

Nach dem Erkennen von starken Änderungen und dem Beschriften der Daten muss noch dafür gesorgt werden, dass die Werte jedes Sensors gleich von dem Netzwerk verarbeitet werden können. Dazu werden alle Daten von dem Mittelwert befreit und über die Varianz normiert. So wird sichergestellt, dass kein Merkmal vom Netzwerk bevorzugt wird und die Gewichte sich in der gleichen Größenordnung befinden. Die Daten der neun Versuchspersonen werden in Trainings- und Testdaten aufgeteilt. Dabei werden die Daten von acht Versuchspersonen für die Trainingsdaten verwendet und die Daten der zufällig verbliebenen Versuchsperson als Testdaten herausgestellt. In den folgenden Abschnitten bleibt diese Aufteilung gleich.

3.2 Struktur des künstlichen neuronalen Netzwerks

Um die Klassifizierung gleichzeitig mit einer Fusion der Merkmale auf hoher Abstraktionsebene zu realisieren, werden in dem vorliegenden Beitrag künstliche neuronale Netzwerke verwendet. Da davon ausgegangen werden kann, dass sich Müdigkeit kumuliert, ist es wichtig, den zeitlichen Verlauf der Merkmale zu berücksichtigen, sodass bei der Struktur die Wahl auf Long Short-Term Memory (LSTM) Netzwerke [2] gefallen ist. In Abbildung 4 ist die Struktur des vollständigen Netzwerks abgebildet, die aus einer LSTM-Schicht, einer vollständig verknüpften Schicht (FC) und einer SoftMax-Schicht besteht. Der Eingang wird mit Sequenzen mit einer Länge von 264 Messpunkten, wie in Abschnitt 3.1 beschrieben, belegt. Die Anzahl der Sequenzen bestimmt sich aus der Anzahl der Merkmale, die für die Klassifizierung benutzt werden. In diesem Fall ist das Netzwerk mit den zwölf Merkmalen aus der Tabelle 1 dargestellt, die sich zu neun Merkmalen aus dem Gehirn, zwei Merkmalen aus dem Herzen und einem Merkmal aus den Augen aufteilen. Die LSTM-Schicht besteht aus 300 verstecken Einheiten, die parallel mit den Sequenzen gespeist werden. Heraus kommen 300 Sequenzen, die durch die LSTM-Schicht verarbeitet werden, die nun als Eingang für das vollständig verknüpfte Netzwerk dienen. Dieses fungiert als Klassifikator, sodass es eine Sequenz von Tupeln ausgibt, welche durch die SoftMax-Schicht zu einer Sequenz von Klassifikationsergebnissen verarbeitet wird.

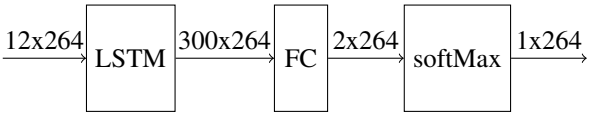


Abbildung 4 Struktur des künstlichen neuronalen Netzwerks

Das Training wird über eine bestimmte Anzahl von 150 Epochen durchgeführt. Diese Anzahl wird so groß gewählt, sodass der Fehler nur noch einen kleinen Gradienten besitzt und sich kaum mehr verbessert. Die initiale Lernrate wird auf 0.02 festgesetzt, welche nach der Hälfte der Epochen noch einmal auf 0.002 verkleinert wird.

3.3 Ergebnisse der Fahrerzustandserkennung

In diesem Abschnitt werden die Ergebnisse der Datenverarbeitung vorgestellt. Es werden verschiedene künstliche neuronale Netzwerke verglichen, die unterschiedlich trainiert sind, aber immer die in Abschnitt 3.2 vorgestellte Struktur aufweisen. Der Unterschied liegt darin, dass eine verschiedene Anzahl von Merkmalen aus den zwölf Merkmalen bei dem Training verwendet wird. Bei jedem Ergebnis wird das Netzwerk auf Daten einer Versuchsperson ausgewertet, die nicht beim Training verwendet wird. Das Netzwerk ermittelt den Fahrerzustand über die Versuchszeit und anschließend wird das Ergebnis mit dem vorher manuell beschrifteten Wert für den Fahrerzustand (Ground Truth) verglichen. Die Genauigkeit wird in diesem Fall durch die Anzahl der richtig positiven Klassifikationen geteilt durch die Gesamtanzahl der Klassifikationen bestimmt, welche im Optimalfall 100 % betragen kann. Hierbei sind hohe Werte der Genauigkeit zu erwarten, denn über 80 % der Zeit bleibt der Zustand in diesem speziellen Fall gleich, sodass ein Netzwerk ohne die Prädiktion eines Zustandswechsel bereits 80 % Genauigkeit erreicht. Als weiteres Maß wird der Zeitunterschied zwischen dem prädizierten Zustandswechsel und dem aus der Beschriftung hervorgehenden Zustandswechsel verwendet, sodass ein absoluter Vergleichswert vorliegt. Dieser sollte im Optimalfall möglichst gering sein.

Als Referenz der Genauigkeit ist in Abbildung 5 ein Ergebnis von einem vollständig verknüpften Netzwerk zu sehen, welches nicht die zeitliche Abfolge der Daten berücksichtigt, da jeder Messpunkt zufällig und von anderen unabhängig für das Training benutzt wird. Es wird deutlich, dass bei einer Genauigkeit von etwa 65 % viele Messwerte der Sequenz falsch zugeordnet werden und das Ergebnis stark zwischen den Zuständen hin und her springt.

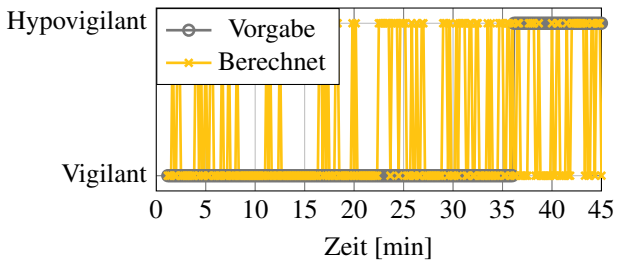


Abbildung 5 Klassifikation auf den Testdaten mit allen verfügbaren Merkmalen und einem einfachen Netzwerk

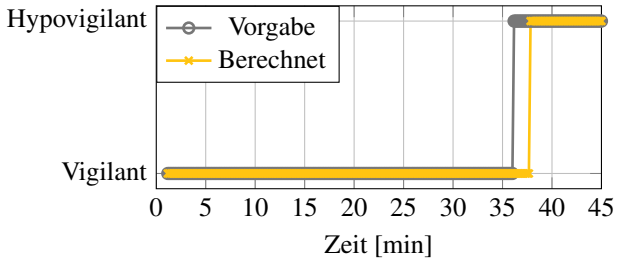


Abbildung 6 Klassifikation auf einer Sequenz der Testdaten mit allen verfügbaren Merkmalen

Nachfolgend werden nun die Ergebnisse der Netzwerke mit der Struktur aus Abbildung 4 gezeigt und erläutert. In Abbildung 6 kann das Ergebnis des Netzwerks gesehen werden, welches mit allen zwölf Merkmalen auf der Basis von acht Versuchspersonen trainiert worden ist. Bei den Testdaten kann eine Genauigkeit von 96 % festgestellt werden. Der zeitliche Verzug zwischen dem Zustandswechsel in der Beschriftung und dem vom Netzwerk ermittelten Zustandswechsel beträgt 100 s.
In Abbildung 7 ist das Ergebnis des Netzwerks zu sehen, bei dem die Gehirnwellen von dem EEG Stirnband nicht mehr als Merkmale zum Training benutzt werden, es werden nur die drei verbleibenden Merkmale (Herz und Auge) ausgewertet. Hier wird deutlich, dass im Ergebnis kaum ein Unterschied zu dem Netzwerk besteht, welches mit allen Merkmalen trainiert worden ist. Die Genauigkeit ist mit etwa 96 % identisch zu den Ergebnissen in Abbildung 6. Ebenso beträgt der Zeitverzug zwischen dem Zustandswechsel 100 s. Dies ist erstaunlich, da neun der zwölf Merkmale nicht mehr benutzt werden. Besonders die

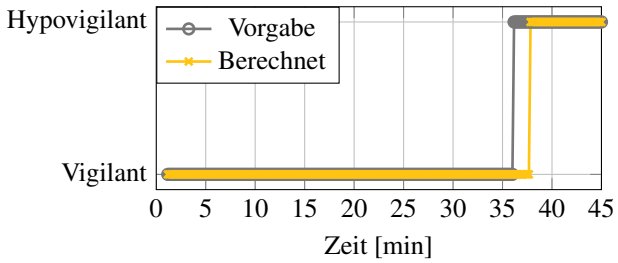


Abbildung 7 Klassifikation auf einer Sequenz der Testdaten mit allen verfügbaren Merkmalen außer EEG

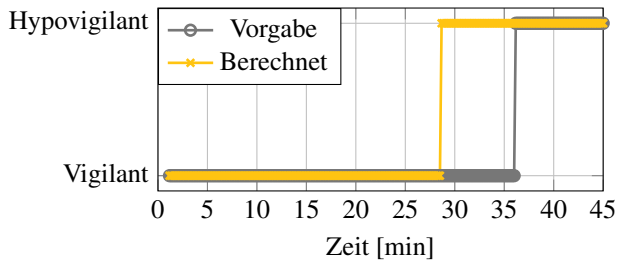


Abbildung 8 Klassifikation auf einer Sequenz der Testdaten nur mit PERCLOS als Merkmal

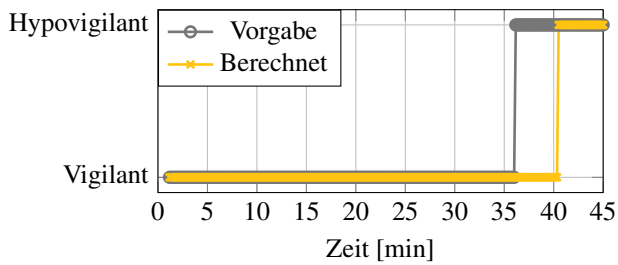


Abbildung 9 Klassifikation auf einer Sequenz der Testdaten nur mit Herzdaten (*HF*, *H*) als Merkmal

Gehirnwellen, die für die manuelle Bestimmung des Zustandswechsels herangezogen werden, werden hier nicht mehr verwendet. Um die Robustheit des Verfahrens zu testen, wird eine Kreuz-Validierung durchgeführt. Dazu werden Test- und Trainingsdaten mehrfach vertauscht, indem immer eine andere Versuchsperson als Testdatensatz fungiert und die übrigen Acht als Trainingsdatensatz. Dies wird für alle neun Versuchspersonen wiederholt, so dass jede Versuchsperson einmal als Testdatensatz herangezogen wird. Hierbei erreicht dieses Netzwerk eine Genauigkeit von ∅ 86,2 %, die sich um etwa 10 % absolut von der Auswertung des einzelnen Testdatensatzes verringert.
Wird das Netzwerk nur mit dem Merkmal der Kameraauswertung (PERCLOS) trainiert, dann ist das Ergebnis in Abbildung 8 zu sehen. Hier kann erkannt werden, dass sich das Ergebnis verschlechtert hat und sich mit etwa 83 % also 13 % absolut schlechter darstellt. Es wird deutlich, dass das Netzwerk den Wechsel des Zustands deutlich zu früh vorhersagt, sodass der Fehler zwischen der Minute 28 und 36 entsteht. Somit lässt sich hier die Zeitdifferenz zwischen den Zustandswechseln mit 450 s angeben. Diese ist im Vergleich zu den beiden Netzwerken, die mit mehr Merkmalen trainiert werden, um 90 s größer.
Ebenso kann die Auswertung für das Netzwerk durchgeführt werden, welches nur mit den Merkmalen des EKG (also *HF* und dem Wert *H*) trainiert wird. Das Ergebnis ist in Abbildung 9 zu sehen. In diesem Fall wird der Wechsel des Fahrerzustands erst in Minute 40 von dem Netzwerk bestimmt, sodass hier eine Genauigkeit von etwa 90 % resultiert. Dieses Ergebnis ist also 6 % absolut schlechter als das mit allen Merkmalen trainierte, jedoch 7 % absolut besser als das Ergebnis, welches nur mit den Merkmalen der Kamera trainiert wird. Hier ist der zeitliche Versatz 270 s

und somit 70 s länger als bei den Netzwerken, die mit mehr Merkmalen trainiert werden, jedoch 180 s weniger als bei dem Netzwerk, welches nur mit den Merkmalen aus dem Kamerasensor trainiert wird.

4 Zusammenfassung

In diesem Beitrag wird ein mechatronisches System analysiert, welches den Zustand des Fahrers wahrnehmen kann. Dazu werden drei verschiedene Sensoren verwendet (EEG, EKG, Kamera), deren Signale mit einem künstlichen neuronalen Netzwerk verarbeitet werden. Das Netzwerk bildet ein Modell des Fahrers ab und kann mit den Signalen der Sensoren einen Fahrerzustand berechnen. Der Fahrerzustand wird in diesem Fall visuell aus den Informationen des EEG, also den Gehirnwellen, herausgearbeitet, um eine Beschriftung (Ground Truth) für das Training zu bekommen. Daraufhin werden verschiedene Kombinationen von Merkmalen für das Training benutzt, um die Notwendigkeit dieser Merkmale einschätzen zu können. Zunächst wird das Training ohne die Merkmale aus den Hirnwellen durchgeführt (nur mit EKG und Kamera). Danach wird das Training nur auf den einzelnen Sensoren durchgeführt (EKG oder Kamera). Hierbei zeigt sich, dass das Modell auch ohne die Gehirnwellen als Merkmal ein gleichwertiges Ergebnis erzielen kann, jedoch mit den Merkmalen aus einzelnen Sensoren schlechter funktioniert. Die Verzugszeit ist mit minimal 100 s eher gering im Gegensatz zum Zeitraum von mindestens 30 min, in dem die Effekte der Vigilanz auftreten können. Somit wird hier eine Fusion aus mehreren Sensoren mit einem LSTM-Netzwerk für die Klassifizierung des Fahrerzustands vorgeschlagen.
In Zukunft sollen alle Merkmale kontaktlos erfasst werden. So stellt [10] ein Verfahren vor, mit dem einzelne Herzschläge aus Farbbildern von dem Gesicht erkannt werden können, es existiert jedoch noch keine Analyse, ob auch die Herzratenvariabilität hierüber bestimmt werden kann. Dazu können sich auch noch weitere Merkmale, wie zum Beispiel die Sitzposition oder Bewegungen im Fahrersitz, positiv auf die Erfassung des Fahrerzustands auswirken. Hierzu wurde bereits ein Sitzdrucksensorsystem in den Sensoraufbau implementiert, sodass weitere Analysen dazu folgen werden.

5 Literatur

[1] Hecht, T., Feldhütter, A., Radlmayr, J., Nakano, Y., Miki, Y., Henle, C., Bengler, K.: A review of driver state monitoring systems in the context of automated driving. In: Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), *Advances in intelligent systems and computing*, vol. 823, pp. 398–408. Springer International Publishing, Cham (2019). DOI 10.1007/978-3-319-96074-6{\textunderscore}43

[2] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)

[3] Interaxon: MuseIO | Available Data (2015, geprüft 03.11.2020). URL `https://web.archive.org/web/20181105231756/http://developer.choosemuse.com/tools/available-data#Absolute_Band_Powers`

[4] Jap, B.T., Lal, S., Fischer, P., Bekiaris, E.: Using eeg spectral components to assess algorithms for detecting fatigue. Expert Systems with Applications **36**(2), 2352–2359 (2009)

[5] Krajewski, J., Sommer, D., Trutschel, U., Edwards, D., Golz, M.: Steering wheel behavior based estimation of fatigue. Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment pp. 118–124 (2009). DOI 10.17077/drivingassessment.1311

[6] Leonhardt, S., Leicht, L., Teichmann, D.: Unobtrusive vital sign monitoring in automotive environments—a review. Sensors **18**(9) (2018). DOI 10.3390/s18093080

[7] Mackworth, N.H.: The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology pp. 6–21 (1948). DOI 10.1080/17470214808416738

[8] Sahayadhas, A., Sundaraj, K., Murugappan, M.: Detecting driver drowsiness based on sensors: a review. Sensors **12**(12), 16937–16953 (2012)

[9] Sommer, D., Golz, M.: Evaluation of perclos based current fatigue monitoring technologies. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 4456–4459. IEEE (2010)

[10] Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W.: Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics **31**(5), 1–8 (2012)

[11] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, Y. A. Sheikh: Openpose: Realtime multi-person 2d pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019)

[12] Zhang, X., Li, J., Liu, Y., Zhang, Z., Wang, Z., Luo, D., Zhou, X., Zhu, M., Salman, W., Hu, G., et al.: Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable eeg. Sensors **17**(3), 486 (2017). DOI 10.3390/s17030486. URL `http://dx.doi.org/10.3390/s17030486`

Untersuchung charakteristischer Parameter des Spurfolgeverhaltens zur Generierung einer Referenz für hochautomatisiertes Fahren

Investigation of characteristic parameters of lane following behavior to generate a reference for highly automated driving

Laurin Vasile, Barbara Seitz, Christoph Däsch, Mercedes-Benz AG, 70372 Stuttgart, Deutschland, Vorname.Nachname@Daimler.com

Prof. Dr. Dieter Schramm, Universität Duisburg-Essen, Lehrstuhl für Mechatronik, 47057 Duisburg, Deutschland, dieter.schramm@uni-due.de

Kurzfassung

Mit steigendem Automatisierungsgrad der Fahrzeugführung reduziert sich im gleichen Maße auch die Einflussmöglichkeit der Fahrer auf die Fahrweise. Folglich bedarf es zunächst einer Referenz aus natürlichem Fahrverhalten, um zu bewerten, wie Menschen in Zukunft hochautomatisiert gefahren werden möchten. In diesem Beitrag wird am Beispiel des Spurfolgeverhaltens in Autobahnkurven ein neuartiger Ansatz präsentiert, der aus natürlichem Fahrverhalten eine Referenz zur Eingrenzung der Systemauslegung für hochautomatisiertes Fahren generiert. Das Fahrverhalten wird durch die Abweichung von der Spurmitte entlang der Kurve und die Pendelfrequenz quantifiziert. Durch Analyse eines vorhandenen Fahrsimulator-Datensatzes konnten Verhaltenstendenzen bei Autobahnkurvenfahrt anhand des quantifizierten Fahrverhaltens abgeleitet werden. Die Ergebnisse können zur Evaluierung des Komfort- und Sicherheitsempfindens beim hochautomatisierten Fahren in Autobahnkurven herangezogen werden. Weiterhin liefert die vorliegende Arbeit einen Beitrag zur Klärung der Frage, wie natürliches Fahrverhalten beim hochautomatisierten Fahren abzubilden ist.

Abstract

As the degree of automation in vehicle guidance increases, the driver's ability to influence the driving style is reduced to the same extent. Consequently, a reference of natural driving behavior is needed first to evaluate, how people would like to be driven in a highly automated way in the future. Based on lane following behavior on freeway curves, this article presents a new approach, which generates a natural driving behavior reference to reduce the system design domain within the area of highly automated driving. The driving behavior is quantified by the deviation of the lane centerline along the curve and the pendulum frequency. By analyzing an existing driving simulator data set, behavioral tendencies in freeway curve driving were derived with the quantified behavior. The results can be used to evaluate the perception of comfort and safety during highly automated driving on freeway curves. Furthermore, this work contributes to clarifying the question of how to represent natural driving behavior in highly automated driving.

1 Einleitung

Die Einführung aktiver Fahrerassistenzsysteme in den vergangenen Jahrzehnten hat die Sicherheit der Passagiere in kritischen Fahrsituationen kontinuierlich erhöht [1]. Darüber hinaus wurde der Fahrer mit Hilfe teilautomatisierter Fahrerassistenten bei der Aufgabe der Fahrzeugführung weiter entlastet. Für die Zukunft der Mobilität wird eine vollständige Entlastung des Fahrers durch vollautomatisiertes Fahren für mehr Komfort und Sicherheit angestrebt. Die entscheidende Frage, die sich jedoch bei dieser Entwicklung stellt, ist, welche Fahrweise als komfortabel und sicher empfunden wird.

Trotz einiger Untersuchungen, die sich mit dem Komfort- und Sicherheitsempfinden der Fahrzeuginsassen beim hochautomatisierten Fahren befassen, besteht bis heute keine Einigkeit darüber, welche Fahrstilausprägung bevorzugt wird [4–7]. Die verschiedenen Ansätze unterscheiden sich vor allem im Grad der Abweichung zu menschlichem Fahrverhalten. Gründe hierfür sind die Vielzahl an Szenarien, die sich im Straßenverkehr ergeben und die daran gekoppelte subjektive Gewichtung äußerer Einflussgrößen. Als Beispiel seien hier die Beschleunigungen, relativen Abstände zu anderen Verkehrsteilnehmern oder die Positionierung in der eigenen Spur genannt. Die Festlegung allgemeingültiger fahrdynamischer Kenngrößen oder Trajektorien, die den präferierenden Fahrstil definieren, wird dadurch erschwert.

Um weitere Erkenntnisse darüber zu gewinnen, wie der Mensch gefahren werden möchte, wird in diesem Beitrag zunächst am Beispiel des Spurfolgeverhaltens in Autobahnkurven das natürliche Verhalten des Fahrers innerhalb der eigenen Spur untersucht. Für die Analyse wird ein vorhandener Fahrsimulator-Datensatz verwendet. Dieser enthält aufgezeichnete Fahrdaten von 102 Probanden. Während des Versuchs folgen die Probanden einem Kurvenverlauf auf der Autobahn mit unterschiedlichen Radien und natürlichem Verkehr. Zur Quantifizierung des Fahrverhaltens in Kurven werden in der Literatur vorwiegend die von den Fahrern gewählte Beschleunigung und Längs-

geschwindigkeit herangezogen [8–11]. Der Abstand des Fahrzeugs zur Spurmitte und das Pendelverhalten stehen hingegen nicht im Fokus der meisten Betrachtungen. Die Positionierung innerhalb der eigenen Spur ist jedoch ein wichtiger Indikator für geplante Fahrmanöver [18]. Durch eine entsprechende Positionierung in der Spur kann damit die Vorhersehbarkeit für andere Verkehrsteilnehmer erhöht und die Auftrittswahrscheinlichkeit von kritischen Fahrsituationen reduziert werden [5]. Für die Akzeptanz hochautomatisierter Fahrfunktionen spielt neben dem Sicherheitsempfinden auch der erlebte Komfort eine wichtige Rolle. In Ergänzung zur weitreichenden Literatur wird daher das laterale Schwingungsverhalten als komfortrelevanter Parameter zur Quantifizierung des Fahrverhaltens in Autobahnkurven hinzugezogen. Die geschaffene Referenz ermöglicht eine Evaluierung des Komfort- und Sicherheitsempfindens einer hochautomatisierten Fahrt, die anhand der ermittelten Parameter natürliches Fahrverhalten abbildet.

2 Bisherige Arbeiten

Gerade vor dem Hintergrund des hochautomatisierten Fahrens hat das natürliche Fahrverhalten besonders in den vergangenen Jahren an Bedeutung gewonnen und wird im Kontext unterschiedlichster Anwendungsbereiche untersucht. Xie et al. [11] verwenden an einem statischen Fahrsimulator aufgenommene Fahrdaten, um sie in die Längsregelung eines Kurvenfahrt-Modells zu integrieren. Betrachtet werden die Beschleunigungswerte bei Kurveneinund -ausgang sowie die Längsgeschwindigkeit während der Kreisbogenfahrt. Vor allem im Bereich des maschinellen Lernens werden aufgenommene Fahrdaten zur Planung von Trajektorien genutzt, um ein möglichst menschenähnliches Fahrverhalten zu erzeugen. Guo et al. [13] definieren den Fahrkorridor durch verhaltensinduzierte künstliche Potentialfelder, die in unterschiedlichen urbanen Verkehrssituationen generiert wurden. Ein Ansatz zur automatischen Umwandlung von naturalistischen Fahrdaten in übergeordnete Fahrverhaltens-Cluster präsentieren Bender et al. [12].

Darüber hinaus wird das Fahrverhalten auch im Zusammenhang mit der Straßenplanung untersucht. In der Studie von Abel und Møller [14] führt das Vorhandensein von Fahrbahnrandstreifen zu einer Erhöhung der Kurvengeschwindigkeit und der Tendenz, bei Fahrt auf der inneren Spur näher am Rand zu fahren. Cerni und Bassani [15] entwickeln auf Basis naturalistischer Fahrdaten zwei Modelle, die die Kurvenkrümmung der gefahrenen Trajektorie in Beziehung zur Kurvenkrümmung der Straße setzen. Havránek et al. [16] untersuchen die Positionierung in der Spur für sechs verschiedene Kurven mit Radien unterhalb 200 m, vor und nach der Platzierung von Spurrand- und Trennungsmarkierungen auf einspurigen Landstraßen. Weiterhin ist das natürliche Fahrverhalten Gegenstand einiger Untersuchungen zur Objektivierung des Komfort- und Sicherheitsempfindens beim hochautomatisierten Fahren. Griesche et al. untersuchen in [4] die Bewertung unterschiedlicher Fahrstile für ein Spurwechselmanöver beim hochautomatisierten Fahren auf der Autobahn. Abhängig von der Situation bevorzugten die meisten Probanden ihren eigenen oder einen ähnlichen Fahrstil. Zu einem etwas anderen Ergebnis kommen Basu et al. in [6]. Bei der Bewertung des Fahrstils beim hochautomatisiertern Fahren auf der Autobahn und in urbaner Umgebung bevorzugten die Probanden einen deutlich defensiveren Fahrstil als den eigenen. Ein zusätzlich interessantes Ergebnis aus dieser Studie: Über 80 % der Probanden bevorzugten einen Fahrstil, den sie für den eigenen hielten, der aber bezüglich der Abstände zu anderen Verkehrsteilnehmern und Längsgeschwindigkeit einem defensiveren Fahrstil entsprach.

Die oben genannten Studien liefern wichtige Erkenntnisse bei der Klärung der Frage, inwieweit und in welchen Situationen natürliches Fahrverhalten beim hochautomatisierten Fahren präferiert wird. Es besteht jedoch weiterer Bedarf zur Quantifizierung natürlicher Fahrverhaltensparameter, um deren Einfluss auf das Komfort- und Sicherheitsempfinden und damit die Akzeptanz hochautomatisierter Fahrten zu untersuchen. Die vorliegende Arbeit liefert hierzu einen neuartigen Ansatz zur Quantifizierung des natürlichen Fahrverhaltens anhand des Spurfolgeverhaltens in Autobahnkurven durch die Abweichung von der Spurmitte und das Pendelverhalten.

3 Versuchsaufbau und Methodik

Für eine Einordnung der Ergebnisse werden im folgenden Abschnitt die Rahmenbedingungen des Versuchs und die Auswahl der betrachteten Daten beschrieben.

3.1 Fahrsimulator

Der Versuch wurde in einem hochdynamischen Fahrsimulator mit hoher Wiedergabetreue eines realen Fahrgefühls durchgeführt. Die Bewegungsplattform bilden ein luftgelagertes 12,5 m langes einachsiges Schienensystem und ein elektromechanischer Hexapod. Auf diesem befindet sich ein CFK-Dom, der mit 4,5 m Breite und einem Durchmesser von 7,5 m Platz für ein vollständiges Fahrzeug bietet. **Abbildung 1** zeigt die Möglichkeit je nach Anwendungsfall den Dom um 90° zum Schienensystem zu drehen, um Längs- oder Querdynamiken mit bis zu $10\,\mathrm{m/s^2}$ zu simulieren. Die Visualisierung der Umgebung für die Probanden wird über acht Projektoren, die eine 360° Rundumsicht erzeugen, innerhalb des Doms realisiert. Gegenüber vielen anderen Fahrsimulatoren kann dadurch sowohl eine realitätsnahe vestibuläre Wahrnehmung der Beschleunigungen im Fahrzeug als auch eine visuelle Wahrnehmung der Umgebung erzeugt werden. Ein weiterer Vorteil im Zusammenhang mit der vorliegenden Untersuchung ist die präzisere Bestimmung der Fahrzeugposition in der Spur gegenüber sensorbasierender Fahrzeuglokalisierung. Für eine detailliertere Beschreibung des Fahrsimulators wird auf [2, 3] verwiesen.

3.2 Simulationsumgebung

Die Versuchsfahrt wird auf einer zweispurigen Autobahn mit einem Standstreifen durchgeführt. Die Gegenfahrbahn ist durch eine Leitplanke und den dazwischen liegenden Grünstreifen getrennt. Das Spurfolgeverhalten wird

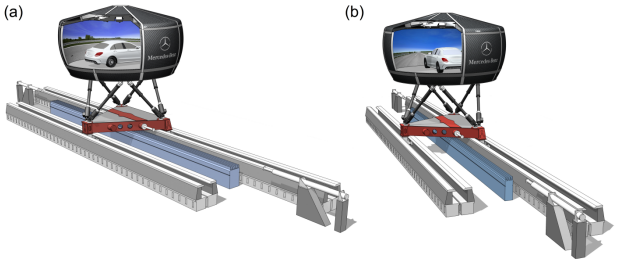


Abbildung 1 Fahrsimulatoraufbau mit a) Quer- und b) Längsanordnung zum Schlitten

in verschiedenen, aufeinander folgenden Kurven mit einer Gesamtlänge von 7,36 km betrachtet. **Abbildung 2** zeigt den Verlauf der Kurvenkrümmung des untersuchten Abschnitts.

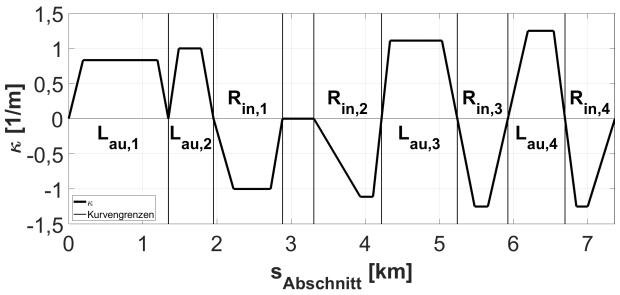


Abbildung 2 Kurvenkrümmung κ entlang der Wegstrecke $s_{Abschnitt}$ des untersuchten Abschnitts. Die einzelnen Kurven sind durch vertikale Linien getrennt. Linkskurven $L_{au,1-4}$ sind durch eine positive und Rechtskurven $R_{in,1-4}$ durch eine negative Krümmung definiert

Insgesamt werden acht Kurven betrachtet (s. **Abbildung 2**). Jede Kurve teilt sich in drei Abschnitte auf: Eingangsklothoide, konstanter Kreisbogen und Ausgangsklothoide. Die Radien der Strecke sind grundsätzlich so ausgelegt, dass bei vorgegebener Geschwindigkeit zur reinen Durchfahrt der Kurve keine Querbeschleunigungen oberhalb von $1{,}4\,\mathrm{m/s^2}$ auftreten. Durch Pendelbewegungen innerhalb der eigenen Spur kann dieser Wert abweichen, wobei sich die Querbeschleunigungen über den gesamten Abschnitt zu 99 % zwischen $\pm 2\,\mathrm{m/s^2}$ bewegen. Die von den Probanden gefahrene Durchschnittsgeschwindigkeit beträgt $M = 122\,\mathrm{km/h}$ ($SD = \pm 6{,}31\,\mathrm{km/h}$). In der vorliegenden Arbeit wird zwischen Kurvenfahrten auf der Innen- und Außenspur über die Richtung der Kurve unterschieden. Da sich die Probanden im untersuchten Abschnitt durchgängig auf der rechten Fahrspur befinden, entsprechen Linkskurven einer Kurvenfahrt auf der äußeren (L_{au}) und Rechtskurven einer Kurvenfahrt auf der inneren Spur (R_{in}). Die Kurvenradien betragen 800, 900, 1000 und 1200 m.

Der Abstand d_{lat} zwischen Fahrzeug und Spurmitte wird im Koordinatensystem der Straße gemessen, mit der Kurventangente als x-Achse und der dazu orthogonal liegenden y-Achse (s. **Abbildung 3**). Die verlängerte y-Achse verbindet den Bezugspunkte der Straße P_S mit dem des Fahrzeugs P_F. Der Bezugspunkt des Fahrzeugs P_F ist durch den Kreuzungspunkt der diagonalen Verbindungslinien definiert, die durch die Mittelpunkte der Radaufstandsflächen verlaufen.

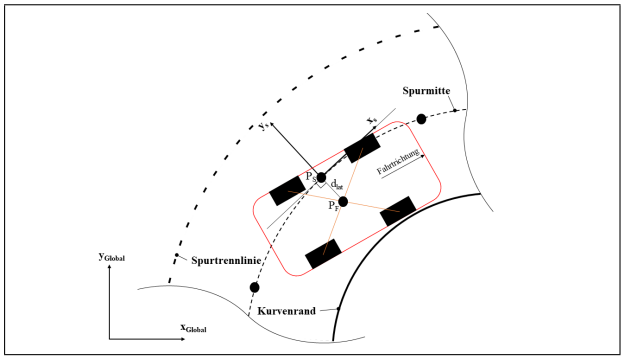


Abbildung 3 Definition laterale Abweichung zur Spurmitte

3.3 Beschreibung der ausgewählten Daten

Die gesamte Strecke wird manuell von dem jeweiligen Proband durchfahren. Die Probanden sind dazu angehalten, die Straßenverkehrsordnung zu befolgen, Spurwechsel nur auf Anweisung durchzuführen und das Geschwindigkeitslimit von 120 km/h zu beachten. Des Weiteren wird keinen fahrfremden Tätigkeiten während der Fahrt nachgegangen. Der simulierte Verkehr erzeugt keine kritischen Situationen. Die Abschnitte zwischen den Spurwechseln erfolgen ohne Anweisungen, wodurch von einem natürlichen Spurfolgeverhalten ausgegangen wird.

3.4 Probandenkollektiv

An der Studie haben insgesamt 102 Probanden teilgenommen. Durch Motion Sickness oder eine starke Abweichung von den Fahranweisungen reduzierte sich die Anzahl auf 89 (44 Frauen, 45 Männer). Der Altersdurchschnitt beträgt 46 Jahre mit Probanden zwischen 20 und 75 Jahren. Die von den Probanden durchschnittlich zurückgelegte Strecke beträgt 18 782 km/Jahr, wodurch von überwiegend erfahrenen Fahrern ausgegangen werden kann. Während für die Kurven $L_{au,2-4}$ und $R_{in,1-4}$ die vollständige Anzahl an Probanden zur Verfügung steht, ergibt sich für Kurve $L_{au,1}$ eine reduzierte Anzahl von 80. Grund hierfür sind nicht abgeschlossene Spurwechselvorgänge der Probanden vor Beginn der Kurve.

4 Analyse Fahrsimulatordaten

Im folgenden Abschnitt werden zur Quantifzierung des natürlichen Fahrverhaltens der Fahrkorridor und das Pendelverhalten analysiert.

4.1 Fahrkorridor

Das Spurfolgeverhalten der Probanden in der Kurve wird nicht durch eine einzige gemittelte Trajektorie, sondern einen Fahrkorridor beschreiben, der abhängig von der Verteilung aller Probanden gebildet wird. Hierzu werden die

Querabweichungen zur Spurmitte an jedem Ortspunkt in einem ersten Schritt über eine Dichtefunktion dargestellt. Das Ergebnis ist ein dreidimensionales Feld, das die Probandendichte bezüglich der Querposition in der Spur über den gesamten Verlauf der Kurve zeigt. Die Draufsicht ist in **Abbildung 4** dargestellt. Die x-Achse beschreibt die Wegstrecke der Kurve und die y-Achse die normierte Querabweichung zur Spurmitte nach links und rechts. Die z-Achse gibt die Höhe der Probandendichte an, farblich aufsteigend in ihrem Maß von blau nach rot. Die normierte Querabweichung bezieht sich auf die maximale Querposition, die das Fahrzeug auf einer Seite der Spurmitte abweichen kann, bevor es den Spurrand berührt. Weiterhin wird die maximal mögliche Querabweichung von der Spurmitte in beiden Richtungen im Folgenden als verfügbarer Fahrraum bezeichnet. Die in Fahrtrichtung linke Hälfte der Spur ist durch einen positiven und die rechte Hälfte durch einen negativen y-Wert definiert. Um den Zusammenhang zwischen Querabweichung und Kurvenabschnitt besser zu visualisieren, ist die Kurve durch zwei vertikale Linien in ihre drei Bereiche unterteilt: Eingangsklothoide, konstanter Kreisbogen und Ausgangsklothoide. Die Spurmitte ist durch eine gestrichelte Linie dargestellt. Der Fahrkorridor ist durch zwei durchgängige Linien, die in der Horizontalen verlaufen, gekennzeichnet.

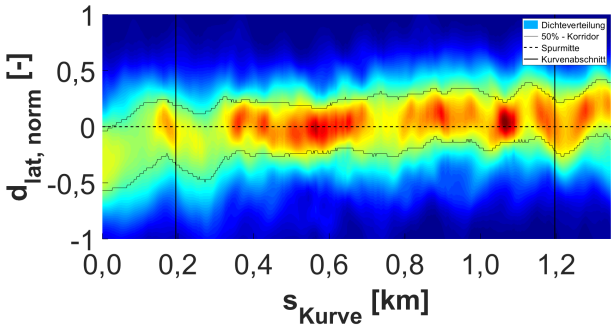


Abbildung 4 Linkskurve $L_{au,1}$: Dichteverteilung der normierten Abweichung zur Spurmitte $d_{lat,norm}$ entlang der Kurvenwegstrecke s_{Kurve}. Die Spurmittenabweichung ist über die Hälfte des verfügbaren Fahrraums jeweils nach links und rechts normiert

Die Integration der Dichtefunktion an einem Ortspunkt über die gesamte Spur ergibt immer 1, wodurch sich die Verteilung der Probanden im jeweiligen Abschnitt berechnen lässt. Der Fahrkorridor kann entlang eines Wegstreckenpunkts an verschiedenen Positionen mit unterschiedlicher Intervallbreite für dieselbe Anzahl an Probanden definiert werden. Ziel ist es daher, ein sinnvolles Kriterium für die Position und die Breite des Fahrkorridors festzulegen. Als besonders hilfreich hat sich hier das Verhältnis von Intervallbreite zu Anzahl der sich darin befindenden Probanden herausgestellt, durch das gleichzeitig die Position des Fahrkorridors bestimmt wird. Das beste Ergebnis wurde mit einem Anteil von 50 Prozent der Probanden erzielt. Der daraus resultierende Fahrkorridor konzentriert sich über allen Kurven auf durchschnittlich $M = 24{,}32\,\%$ $(SD = \pm 1{,}65\,\%)$ des verfügbaren Fahrraums innerhalb der Spur. Für die Untersuchung des Fahrkorridors wird zunächst die Streuung der durchschnittlichen Geschwindigkeitsprofile und deren möglicher Einfluss auf die Querabweichung zur Spurmitte betrachtet. Die beobachtete Korrelation $(r = 0{,}06,\ p < 0{,}05)$ ist als statistisch nicht relevanter Effekt zu interpretieren und erlaubt daher eine Analyse unabhängig von der Geschwindigkeitsstreuung.

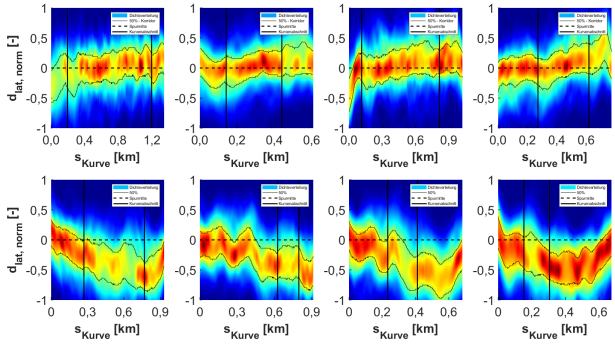


Abbildung 5 Dichteverteilung der normierten Abweichung zur Spurmitte $d_{lat,norm}$ entlang der Kurvenwegstrecke s_{Kurve}. Die oberen Graphen zeigen die Linkskurven $L_{au,1-4}$ (v. l. n. r.), die unteren Graphen zeigen die Rechtskurven $R_{in,1-4}$ (v. l. n. r.)

Für ein besseres Verständnis des folgenden Abschnitts sind in **Abbildung 5** alle Kurven mit der jeweiligen Dichteverteilung dargestellt. Ein deutlicher Unterschied zeigt sich im Grad der Spurmittenabweichung und deren Verlauf entlang der Kurve. So tendieren die Probanden auf der inneren Fahrspur noch weiter am Rand zu fahren als auf der äußeren Fahrspur und die Kurve im Kurvenausgang zu schneiden. In den Ausgangsklothoiden weichen die Probanden im Höchstwert sogar mit durchschnittlich 85,25 % von der Spurmitte ab. Des Weiteren zeigt der Verlauf, dass der Fahrkorridor auf der inneren und äußeren Fahrspur jeweils einem klaren Trend folgt. Die Probanden auf der äußeren Spur halten sich überwiegend im Bereich um die Mitte auf mit leichter Tendenz zur Kurveninnenseite. Ein starker Anstieg der Spurmittenabweichung zum Kurvenausgang hin ist nicht zu erkennen. Vor allem in den Klothoidenausgängen beträgt die maximale Abweichung zur Spurmitte nur knapp die Hälfte des Höchstwertes, der auf den inneren Fahrspuren gemessen wird. Weiterhin nimmt die Spurmittenabweichung auf der inneren Fahrspur von Beginn bis zum Ende der Kurve zu, bevor sie ihr Maximum im Klothoidenausgang erreicht und dann wieder Richtung Spurmitte abnimmt. Das hat zur Folge, dass bei einem Richtungswechsel der Kurve, die folgende mit einer starken Tendenz von außen angefahren wird. Eine Ausnahme bildet Kurve $L_{in,1}$, deren Eintrittskorridor durch kurz zuvor abgeschlossene Spurwechselmanöver beeinflusst ist. Dies führt trotz der davor liegenden Geraden zu einem Anfahren der Linkskurve von außen.

4.2 Schwingungsanalyse

Für die Generierung einer Referenz zur Bewertung des hochautomatisierten Fahrens in Kurven mit natürlichem Fahrverhalten ist eine reine Betrachtung des Fahrkorridors

nicht ausreichend. Es ist daher notwendig auch das Pendelverhalten der Probanden zu analysieren, das über den Fahrkorridor nicht abgebildet werden kann. Zur Beurteilung des Pendelverhaltens werden die Frequenzanteile der überlagerten Schwingung, der laterale Abstand zwischen den Umkehrpunkten der Pendelbewegung (im Weiteren Verlauf Pendeldistanz genannt) sowie die Verteilungen der jeweiligen Daten herangezogen.

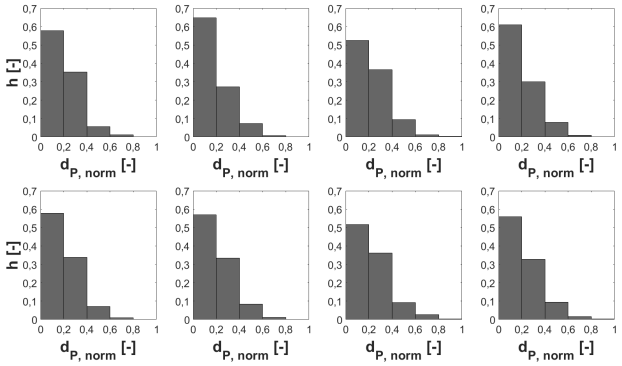


Abbildung 6 Häufigkeitsverteilung h der normierten Pendeldistanz $d_{P,norm}$ für die Linkskurven $L_{au,1-4}$ (oberere Graphen, v. l. n. r.) und Rechtskurven $R_{in,1-4}$ (untere Graphen, v. l. n. r.)

Die Pendeldistanz wird über den verfügbaren Fahrraum in der Spur normiert und beschreibt die Distanz, die zwischen zwei Richtungswechseln quer zur Spur zurückgelegt wird. Die Normierung ermöglicht einen Vergleich unterschiedlicher Spurbreiten bei weiteren Untersuchungen. **Abbildung 6** zeigt die Häufigkeitsverteilung der normierten Pendeldistanz abhängig von ihrer Größe für die äußere (obere Graphen) und innere (untere Graphen) Fahrspur. Für beide Spuren ergibt sich ein ähnliches Bild. Die überwiegende Anzahl der Pendeldistanzen liegt mit $M = 57{,}29\,\%$ $(SD = 4{,}01\,\%)$ im Größenbereich von $0-20\,\%$ des verfügbaren Fahrraums. Während auch im Größenbereich von $20-40\,\%$ immer noch ungefähr ein Drittel der Werte liegen, fallen nur weniger als $10\,\%$ in den Bereich von $60-80\,\%$ des verfügbaren Fahrraums.

Neben der Pendeldistanz ist vor allem der Frequenzbereich durch seinen direkten Einfluss auf das Komfortempfinden von großem Interesse. Für die Analyse wird die Pendelfrequenz zunächst durch die Anzahl der Umkehrpunkte und die Dauer der Kurvendurchfahrt berechnet. Die durchschnittliche Frequenz für alle Kurven beträgt $M = 0{,}36\,\mathrm{Hz}$ $(SD = \pm 0{,}01)$ und zeigt eine geringe Varianz. Ein etwas anderes Bild ergibt sich bei der Betrachtung einzelner Frequenzanteile. Diese werden innerhalb einer Kurve mit Hilfe einer diskreten Fourier-Transformation (DFT) identifiziert [17]. Anschließend werden nur Frequenzen der Amplitudenmaxima betrachtet und als signifikant gewertet. Maxima unterhalb von 2 cm haben keinen entscheidenden Einfluss auf das Gesamtbild der lateralen Schwingung und werden deshalb nicht betrachtet. Aufgrund der unterschiedlichen Kurven- und damit Datenlänge, ergeben sich unterschiedliche Frequenzauflösungen für jede Kurve. Für das resultierende Frequenzintervall sind die Unterschiede jedoch vernachlässigbar klein und stellen kein Problem für den Vergleich der Kurven dar. Die Häufigkeitsverteilung der Frequenzen für die innere (untere Graphen) und äußere (obere Graphen) Fahrspur ist in **Abbildung 7** dargestellt.

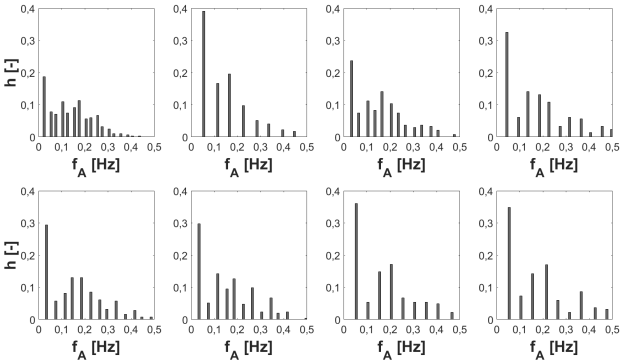


Abbildung 7 Häufigkeitsverteilung h der Frequenzanteile f_A bei Kurvenfahrt für die Linkskurven $L_{au,1-4}$ (obere Graphen, v. l. n. r.) und Rechtskurven $R_{in,1-4}$ (untere Graphen, v. l. n. r.)

Insgesamt bewegen sich die Frequenzanteile zwischen $0-0{,}5\,\mathrm{Hz}$. Es zeigt sich, dass vor allem der Bereich um $0{,}05\,\mathrm{Hz}$ den Hauptfrequenzanteil aller Probanden abbildet. Darüber hinaus ergeben sich für alle Kurven zwischen $0{,}1-0{,}23\,\mathrm{Hz}$ ein bis zwei weitere Häufigkeitsmaxima. Durch die unterschiedlichen Betrachtungsweisen der Pendelfrequenz ergeben sich zwei Referenzen, die zur Evaluierung des Komforts- und Sicherheitsempfindens beim hochautomatisierten Fahren in Kurven genutzt werden können.

5 Zusammenfassung und Ausblick

In diesem Beitrag wird am Beispiel des Spurfolgeverhaltens in Autobahnkurven ein neuartiger Ansatz präsentiert, der eine Referenz für das natürliche Fahrverhalten generiert, um den Systemauslegungsraum für hochautomatisiertes Fahren einzugrenzen. Neu ist der Fokus auf die Quantifizierung der Spurmittenabweichung und Pendelfrequenz zur Ableitung von Verhaltenstendenzen bei Autobahnkurvenfahrt. Für die Analyse wurde ein vorhandener Fahrsimulator-Datensatz mit aufgezeichneten Fahrdaten von 102 Probanden verwendet. Die Ergebnisse zeigen, dass sich die Spurmittenabweichung und deren Verlauf entlang der Kurve grundsätzlich zwischen Innen- und Außenspur unterscheiden. Die Tendenz der Probanden, auf der inneren Fahrspur noch weiter am Rand zu fahren als auf der äußeren Fahrspur, wurden durch die Korridorgrenzen visualisiert und durch die normierte Abweichung von der Spurmitte quantifiziert. Die höchsten Werte der Spurmittenabweichung zur Kurveninnenseite wurden bei Fahrten auf der Innenspur in den Ausgangsklothoiden erreicht und zeigen zusammen mit dem Verlauf, dass Menschen längere Kurven zum Ausgang hin schneiden. Für die Pendelfrequenz ergaben sich abhängig von der Berechnungsmethode unterschiedliche Frequenzbereiche. Die durch eine DFT ermittelten und als signifikant bewerteten Frequenzanteile liegen in einem Bereich zwischen $0-0{,}23\,\mathrm{Hz}$. Die durch die Anzahl der Pendelbewegungen und der Dauer der Kur-

venfahrt ermittelten Frequenzen liegen jedoch bei durchschnittlich 0,36 Hz.
Fahrverhaltensparameter, die mit dem vorgestellten Ansatz ermittelt werden, können als Basis zur Evaluierung des Komfort- und Sicherheitsempfindens beim hochautomatisierten Fahren in Autobahnkurven verwendet werden. Anhand der Parameter lassen sich weitere Erkenntnisse darüber gewinnen, mit welcher Fahrverhaltensausprägung beim hochautomatisierten Fahren ein Fahrgefühl generiert werden kann, dass dem bei einem selbstgesteuerten Fahrzeug entspricht. Für zukünftige Arbeiten liegt es daher nahe, das Spurfolgeverhalten auch in Kurven mit engeren Radien zu quantifizieren und den Einfluss von Kurvenlänge, Anzahl und Breite der Spuren sowie das Kurvenfahrtverhalten auf der linken (schnelleren) Spur miteinzubeziehen, um einen breiten Bereich an Umgebungsparametern abzudecken.

6 Literatur

[1] Schramm, D.; Schweig, S.: Fahrerassistenzsysteme – ein Überblick. In: Proff H., Brand M., Schramm D. (eds) Altersgerechte Fahrerassistenzsysteme. Springer Gabler, Wiesbaden, S. 39-53, 2017.

[2] Klüver, M.: *Can We Trust Driving Simulator Studies? The Behavioral Validity of the Daimler AG Driving Simulators.* Dissertation, Fachbereich 02 – Sozialwissenschaften, Medien und Sport, Johannes Gutenberg-Universität Mainz, 2016.

[3] Bellem, H., Klüver M, Schrauf M, Schöner H-P, Hecht H, Krems JF.: *Can We Study Autonomous Driving Comfort in Moving-Base Driving Simulators? A Validation Study*. Human Factors, 2017.

[4] Griesche, S.; Nicolay, E.; Assmann, D.; Dotzauer, M.; Käthner, D.: *Should my car drive as I do? What kind of driving style do drivers prefer for the design of automated driving functions?* 7. Braunschweiger Symposium Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel (AAET), ITS automotive nord e.V., pp. 185-204, 2016.

[5] Syed, A.: *Toward human-like automated driving: Learning spacing profiles from human driving data.* Dissertation, Dept. Comput. Eng., Wayne State Univ., Detroit, MI, USA, 2017.

[6] Basu, C.; Yang, Q.; Hungerman, D.; Singhal, M.; Dragan, A.: *Do You Want Your Autonomous Car To Drive Like You?* HRI '17: ACM/IEEE International Conference on Human-Robot Interaction Vienna Austria March, pp. 417-425, 2017.

[7] Festner, M.: *Objektivierte Bewertung des Fahrstils auf Basis der Komfortwahrnehmung bei hochautomatisiertem Fahren in Abhängigkeit fahrfremder Tätigkeiten.* Dissertation, Fakultät für Ingenieurwissenschaften, Abteilung Maschinenbau und Verfahrenstechnik, Universität Duisburg-Essen, 2019.

[8] Lichtenstein, A.; Stößel, C.; Clemens, C.: *Der Mensch im Mittelpunkt technischer Systeme* Tagungsband der 8. Berliner Werkstatt Mensch-Maschine-Systeme, Düsseldorf: VDI Verlag, S. 202-207, 2009.

[9] Mumcuoglu, M. E.; Alcan, G.; Unel, M.; Cicek, O.; Mutluergil, M.; Yilmaz, M.; Koprubasi, K.: *Driving Behavior Classification Using Long Short Term Memory Networks.* 2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), Torino, Italy, pp. 1-6, 2019.

[10] Wang, B.; Hallmark, S.; Savolainen, P.; Dong, J.: *Examining vehicle operating speeds on rural two-lane curves using naturalistic driving data.* Accident Analysis & Prevention,Volume 118, pp. 236-243, 2018.

[11] Xie, J.; Gong, J.; Wu, S.; Xiong, G.; Lu, C.: *A personalized curve driving model for intelligent vehicle.* 2017 IEEE International Conference on Unmanned Systems (ICUS), Beijing, pp. 301-306, 2017.

[12] Bender, A.; Agamennoni, G.; Ward, J. R.; Worrall, S.; Nebot, E. M.: *An Unsupervised Approach for Inferring Driver Behavior From Naturalistic Driving Data.* IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3325-3336, 2015.

[13] Guo, C.; Owaki, T.; Kidono, K.; Machida, T.; Terashima, R.; Kojima, Y.: *Toward human-like lane following behavior in urban environment with a learning-based behavior-induction potential map.* 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, pp. 1409-1416, 2017.

[14] Abele, L.; Møller, M.: *The Relationship between Road Design and Driving Behavior.* 3rd International Conference on Road Safety and Simulation, Indianapolis, Indiana, 2011.

[15] Cerni, G.; Bassani, M. *Naturalistic driving data collection to investigate into the effects of road geometrics on track behaviour.* Transportation Research Part C: Emerging Technologies, Volume 77, pp. 1-15, 2017.

[16] Havránek, P.; Zůvala, R.; Špaňhel, J.; Herout, A.; Valentová, V.; Ambros, J.: *How does road marking in horizontal curves influence driving behaviour?* In: Eur. Transp. Res. Rev. 12, Artikel 33, 2020.

[17] Heinzel, G.; Rüdiger, A.; Schilling, R.: *Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows.* Max Plank Inst. 12., 2002.

[18] Wissing, C.: *Trajektorienprädiktion für das automatisierte Fahren.* Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität Dortmund, 2020.

Fahrstilanalyse zur Erhöhung der Nutzerakzeptanz automatisiert energieeffizienter Fahrzeuglängsführung
Driving Style Analysis for an Increasing Driver Acceptance of Automated Energy-Efficient Longitudinal Vehicle Guidance

Daniel Fink, M.Sc.; Jan-Domenik Czarnecki, B.Sc; Dr.-Ing. Mark Wielitzka, Leibniz University Hannover, Institut für Mechatronische Systeme, 30823 Garbsen, Deutschland, daniel.fink@imes.uni-hannover.de; Dr.-Ing. Christoph Schweers; Dr.-Ing. Ahmed Trabelsi, IAV GmbH, 10587 Berlin

Kurzfassung

Die Nutzungshäufigkeit von Fahrerassistenzsystemen zur teilautomatisierten Fahrzeuglängsführung steigt in der Regel mit der Möglichkeit, das Fahrverhalten des Systems den eigenen Ansprüchen anzupassen. Da sich diese Ansprüche fahrerindividuell mitunter stark unterscheiden, bietet es sich an, das Fahrverhalten während der manuellen Fahrzeugführung zu analysieren und diese Informationen zur Parametrierung der Fahrerassistenzsysteme zu nutzen. Dazu wird im Rahmen dieser Arbeit ein regelbasiertes Verfahren zur Fahrstilanalyse vorgestellt, bei dem, basierend auf seriensensorischen Messgrößen und Routeninformationen, neben einer dynamikbezogenen Fahrstilbewertung, eine fahrerspezifische Präferenz im Zielkonflikt einer energie- und zeitsparenden Fahrweise abgeschätzt wird. Zur Sensitivierung der Fahrstilanalyse umfasst das Verfahren eine Methode zur Erkennung relevanter Fahrmanöver.

Abstract

For most drivers, the use of an assistance system for semi-automated longitudinal vehicle control depends on how the system's driving behavior meets their individual preferences. In order to generate information that can be utilized to adapt assistance systems through re-parametrizing, this paper presents a rule-based method to analyze the driver's behavior during manual driving. On the one hand, driving style-relevant situations are detected and analyzed regarding the occurring driving dynamics. On the other hand, the presented method enables to evaluate the driver's preferences concerning the trade-off between a minimal driving time and reducing the energy consumption.

1 Einleitung

Auf dem Weg zum automatisierten Fahren werden Teilaufgaben der Fahrzeugführung zunehmend von Fahrerassistenzsystemen (FAS) übernommen. Viele dieser Systeme, wie beispielsweise ein Abstandsregeltempomat (ACC), tragen neben einer Komfortsteigung dazu bei, die allgemeine Verkehrs- und Fahrsicherheit zu erhöhen [1]. Darüber hinaus existieren weitere FAS, die darauf abzielen den Energiebedarf des Fahrzeugs zu senken [2]. Ob diese FAS tatsächlich verwendet und deren Vorteile genutzt werden, obliegt letztlich jedoch allein der fahrzeugführenden Person. Diese Verwendungsentscheidung ist häufig daran geknüpft, inwieweit das assistenzsystemspezifische Fahrverhalten der fahrzeugführenden Person zusagt. So kann eine Unterrepräsentation fahrerindividueller Fahrzeugführungspräferenzen dazu führen, dass Assistenzsysteme ungenutzt bleiben und ihr Potential zur Steigerung von Fahrsicherheit und Energieeffizienz verfällt.

Um die Fahrerakzeptanz assistierter Fahrzeugführung zu erhöhen, gilt es das Fahrverhalten von FAS so zu adaptieren, dass dieses den fahrereigenen Ansprüchen angepasst wird. Um Informationen über diese Ansprüche automatisiert zu generieren, bietet es sich an, die fahrereigene Art der manuellen Fahrzeugführung zu analysieren. Da sich der eigene Fahrstil vor allem in der Fahrdynamik widerspiegelt [3], basieren etablierte Verfahren zur Analyse des Fahrerverhaltens auf der Auswertung von Beschleunigungs- und Geschwindigkeitswerten. So wird in [4] ein Fuzzylogik-basierter Ansatz vorgestellt, mit dem der Fahrstil basierend auf Beschleunigungs- und Geschwindigkeitswerten einer von vier Aggressivitätsstufen zugeordnet werden kann. Dabei erfolgt stets die Betrachtung der gesamten Fahrt, ohne zwischen verschiedenen Fahrmanövern zu differenzieren oder verkehrsbedingte Beeinflussungen zu berücksichtigen. In [5] wird eine Umsetzung einer Fahrstilanalyse mit Kurvenmanövererkennung vorgestellt. Dabei wird zunächst anhand eines Schwellwertes für die Gierrate erkannt, ob eine Kurvenfahrt vorherrscht und um welche Art es sich handelt (Straßenkrümmung, U-Turn, Spurwechsel). Anschließend wird die erkannte Kurvenart, über einen dynamischen Zeitnormierungsansatz [6], mit bekannten Manövern definierten Fahrstils verglichen, um zwischen „aggressiven" und „nicht aggressiven" Kurvenfahrverhalten zu unterscheiden. Als Messsystem dient dabei ein Smartphone. Ähnlich wird in [7] vorgegangen, um abhängig von Längs- und Querbeschleunigungsschwellwerten Kurvenfahrten, Beschleunigungs- oder Verzögerungsmanöver zu erkennen und diese als „sicher" oder „unsicher" einzustufen. Weitere Verfahren und Möglichkeiten der Fahrstilanalyse werden beispielsweise in [8] beschrieben.

Derzeitige Verfahren zur Fahrstilanalyse zeigen, dass sich allgemeine Fahrerklassen wie beispielsweise „sportlich“, oder „komfortabel“ auf Basis von Fahrdynamikdaten unterscheiden lassen. Häufig wird jedoch bei der Bewertung der Fahrstile nicht differenziert, ob eine Fahrsituation entsprechend eigener Vorlieben ausgeführt, oder diese durch andere Verkehrsteilnehmer bedingt bzw. hervorgerufen wurde. Somit werden auch Fahrerreaktionen miteinbezogen, die letztlich nicht immer repräsentativ für den tatsächlichen Fahrstil des Fahrers sind. So sind beispielsweise auch sehr aggressive Fahrertypen, bei nichtvorhandener Überholmöglichkeit, dazu gezwungen ihre Geschwindigkeit an vorausfahrende Fahrzeuge anzupassen.

Darüber hinaus adressieren derzeitige Analyseverfahren zumeist ausschließlich das dynamische Fahrerverhalten, eine Aussage über eine beabsichtigt energieeffiziente oder zeitsparende Fahrweise wird hingegen nicht getroffen. Daher reichen die so gewonnen Fahrerpräferenzinformation nicht immer aus, um auch energiebedarfsbezogene FAS zu personalisieren [2]. Zudem existieren bereits heute Forschungs- und Entwicklungsansätze einer teilautomatisierten Fahrzeugführung, bei der eine fahrstreckenbezogene Energiebedarfsreduktion erfolgt. Als Grundlage dient dabei häufig die Planung und Optimierung einer Geschwindigkeitstrajektorie für den bevorstehenden Streckenabschnitt. Hierbei erfordert der Optimierungsprozess die Gewichtung der Kostenterme Energiebedarf und Fahrtdauer [9], auf die fahrerabhängig unterschiedlich viel Wert gelegt werden kann. Es bedarf daher, sowohl für die assistierte als auch für die teilautomatisierte Fahrzeugführung, zusätzlicher Information über entsprechende Fahrervorlieben, um diese im Zielkonflikt eines geringen Energiebedarfs und einer kurzen Fahrtdauer zu berücksichtigen.

Gegenstand dieser Arbeit ist daher der Aufbau eines Verfahrens zur umfangreichen Analyse des Fahrstils, sodass einerseits fahrdynamische Präferenzen und andererseits fahrerindividuelle Abwägungsvorlieben zwischen kurzer Fahrtdauer und geringem Energiebedarf ermittelt werden können.

2 Methodischer Aufbau des Fahrstilanalyseverfahrens

Das Gesamtkonzept des entwickelten Fahrstilanalyseverfahrens ist in **Bild 1** dargestellt. Als Grundlage dienen seriensensorische Messgrößen, die an einem modernen Versuchsfahrzeug aufgenommen werden. Neben Beschleunigungs- und Geschwindigkeitswerten werden zusätzlich Navigationsdaten wie beispielsweise Tempolimits mitberücksichtigt. Darüber hinaus erfolgt die Integration einer Manövererkennung, mit der Fahrsituationen erkannt werden, bei denen sich Fahrzeugführungspräferenzen besonders ausprägen. Unter Auswertung der Distanz zum Vorderfahrzeug, wird zudem automatisiert unterschieden, ob eine Fahrhandlung tatsächlich gewollt ist und somit den Fahrstil der fahrzeugführenden Person widerspiegelt. Auf diese Weise wird gewährleistet, dass Fahrsituationen, die durch andere Verkehrsteilnehmer beeinflusst oder hervorgerufen werden, nicht zur Fahrstilbewertung herangezogen werden.

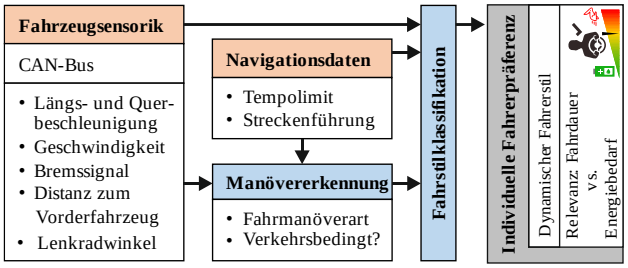


Bild 1 Konzeptionierug: Fahrerstilanlyse

Zur methodischen Beschreibung der Fahrstilanalyse wird zunächst das Verfahren zur Erkennung unterschiedlicher Fahrmanöver vorgestellt. Anschließend wird das Vorgehen zur Ermittlung, ob eine Fahrsituation durch andere Verkehrsteilnehmer bedingt wurde, skizziert. Abschließend erfolgt die Vorstellung einer manöverseparaten Fahrstilklassifikation sowie einer globalen Fahrtbewertung hinsichtlich Energiebedarf und Fahrdauerpräferenz.

2.1 Fahrmanövererkennung

Der Fahrstil einer fahrzeugführenden Person prägt sich verstärkt bei bestimmten Fahrmanövern und weniger bei Konstantfahrten aus [5, 7]. Daher bietet es sich an, den Fahrstil vorwiegend während dieser Manöver zu analysieren. Dafür wird zunächst zwischen *Beschleunigen*, *Verzögern* und *Kurvenfahrt* unterschieden. Bei Verzögerungs- und Beschleunigungsfahrten wird zusätzlich *Anhalten* bzw. *Anfahren* abgegrenzt. Um Start und Endpunkte für Fahrmanöver zu erkennen, werden schwellwertgebundene Bedingungen für bestimmte seriensensorische Messgrößen am Fahrzeug gemäß **Tabelle 1** definiert.

Tabelle 1 Schwellwerte zur Fahrmanövererkennung

Fahrmanöver	*Messgröße*	*Bedingung*
Beschleunigen	Längsbeschleunigung	$> 0{,}3\ \mathrm{m/s^2}$
└ Anfahren	Fahrzeugstartgeschwindigkeit	$< 5\ \mathrm{km/h}$
Verzögern	Längsbeschleunigung	$< -0{,}4\ \mathrm{m/s^2}$
│	Bremspedal betätigt	$= true$
└ Anhalten	Fahrzeugendgeschwindigkeit	$< 5\ \mathrm{km/h}$
Kurvenfahrt	Lenkradwinkel	$> 30\ °$

Die Eignung der spezifischen Beschleunigungsschwellwerte für die Manövererkennung wird beispielsweise in [10] gezeigt. Zur Plausibilisierung werden zudem Bremspedal- und Lenkradwinkelsignale hinzugezogen. Die Messsignale werden über einen gleitenden Mittelwert gefiltert. Werden die Bedingungen eines Manövers für mehr als 2 Sekunden erfüllt, gilt dieses als erkannt. Manöver gleicher Art, die innerhalb von 2 Sekunden aufeinander folgen, werden zusammengefasst und als ein Manöver betrachtet, um beispielsweise ungleichmäßig ausgeführte Verzögerungs- und Beschleunigungsvorgänge nicht zu trennen. Für Beschleunigungs- und Verzögerungsmanöver wird zudem festgelegt, dass mindestens eine Geschwindigkeitsänderung von 15 km/h erfolgen muss. Ansonsten werden die Manöver verworfen. Um Überholmanöver und

Spurwechsel nicht fälschlicherweise als Kurvenfahrt zu klassifizieren, gilt die zusätzliche Bedingung, dass sich die Fahrzeugorientierung nach erkannter Kurvenfahrt um mindestens 20° verändert.

2.2 Vorderfahrzeugerkennung

Um sicherzustellen, dass zur Bewertung des Fahrstils lediglich Fahrmanöver herangezogen werden, die nach eigenen Vorlieben von der fahrzeugführenden Person ausgeführt wurden, wird eine Vorderfahrzeugerkennung integriert. Ziel ist es, fremdverschuldete Fahrsituationen auszusortieren, um die Repräsentanz tatsächlicher Fahrzeugführungspräferenzen im Ergebnis der Fahrstilanalyse zu erhöhen. Dazu wird das Frontradarsignal des Versuchsfahrzeuges ausgewertet. In Anlehnung an [11], wird r_{VF} als einheitslose Beziehung zwischen Fahrzeuggeschwindigkeit v_{x} in [km/h] und Abstand zum Vorderfahrzeug x_{VF} in [m] als

$$r_{\mathrm{VF}} = \frac{2 \cdot x_{\mathrm{VF}} \cdot 1\,\mathrm{km}}{v_{\mathrm{x}} \cdot 1\,\mathrm{m} \cdot 1\,\mathrm{h}} \qquad (1)$$

eingeführt. Unterschreitet r_{VF} während eines Manövers den Wert 1, gilt dieses als verkehrsbedingt beeinflusst, wird verworfen und somit nicht zur Bewertung des Fahrstils herangezogen.

2.3 Manöverbewertung

Der Fahrstil in erkannten und nicht als fremdverschuldet eingestuften Manövern wird anschließend, bezogen auf die Fahrdynamik, bewertet. Dafür bietet sich eine regelbasierte Bewertungsmethode an, da in der Literatur bereits zahlreiche Kriterien zur Fahrdynamikbewertung definiert wurden [5, 10]. Letztlich basiert eine solche Bewertung allerdings auf subjektivem Empfinden [12]. Um dieser Unsicherheit, unter Berücksichtigung etablierter Kriterien, gerecht zu werden, wird methodisch in dieser Arbeit ein Fuzzylogik-Ansatz (siehe [13]) zur Manöverbewertung herangezogen. In Anlehnung an [4] wird zunächst ein Fahrstilbewertungswert eingeführt der zwischen 0 und 1 liegt. Dazu wird das Intervall $[-0{,}33;\ 1{,}33]$ in vier Bereiche eingeteilt, die entweder eine eher geringe „G“, durchschnittliche „D“, hohe „H“, oder sehr hohe „S“ Fahrdynamik repräsentieren. Anstatt diesen Bereichen feste Fahrstilbewertungswertgrenzen zuzuordnen, werden sie über die Funktion eines Zugehörigkeitsgrades gemäß **Bild 2** definiert.

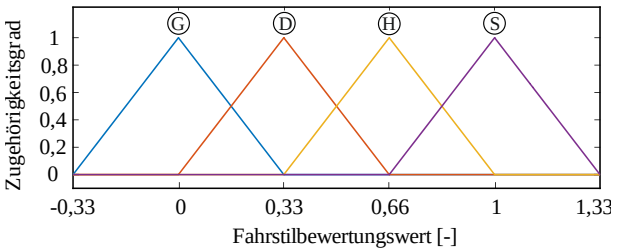


Bild 2 Zugehörigkeitsverteilung der Fahrstilbewertungsklassen

Um für erkannte Beschleunigungs- und Anfahrtsmanöver einen dynamikbezogenen Fahrstilbewertungswert zu ermitteln, wird das Maximum der, während des Manövers aufgetretenen, Fahrzeuglängsbeschleunigung betrachtet. Dazu wird zunächst das Intervall $[0; 5\,\mathrm{m/s^2}]$ über Zugehörigkeitsgradfunktionen in vier Bereiche unterteilt (siehe **Bild 3**).

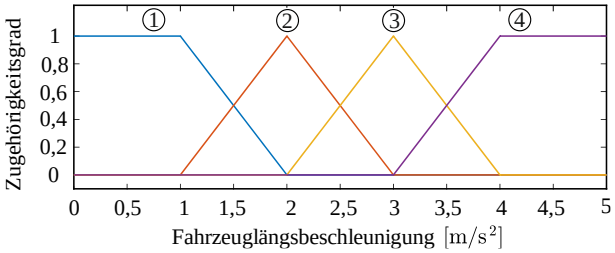


Bild 3 Zugehörigkeitsverteilung Beschleunigungsbewertung

Physikalisch bedingt, ändert sich das maximale Beschleunigungsvermögen aufgrund steigender Fahrwiderstände abhängig von der Fahrzeuggeschwindigkeit. Daher gilt es die Leistungsgrenzen des Versuchsfahrzeuges zu ermitteln. Dazu werden mehrere Identifikationsmanöver mit maximal durchgetretenem Gaspedal ausgeführt. **Bild 4** zeigt die dabei auftretenden Längsbeschleunigung in Abhängigkeit der Fahrzeuggeschwindigkeit.

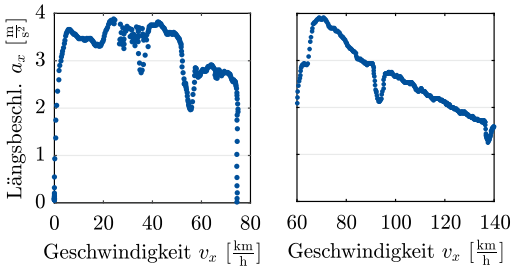


Bild 4 Maximale Fahrzeuglängsbeschleuningung in Abhängigkeit der Geschwindigkeit

Es wird deutlich, dass zwischen 0 und 70 km/h Längsbeschleunigungen von bis zu 4 $\mathrm{m/s^2}$ erreicht werden, während die maximal mögliche Beschleunigung bei höheren Geschwindigkeiten signifikant abfällt. Um diese Geschwindigkeitsabhängigkeit bei der Beschleunigungsmanöverbewertung zu berücksichtigen, wird die Fahrzeuggeschwindigkeit in fünf Bereiche mit Zugehörigkeitsgradfunktion unterteilt (siehe **Bild 5**).

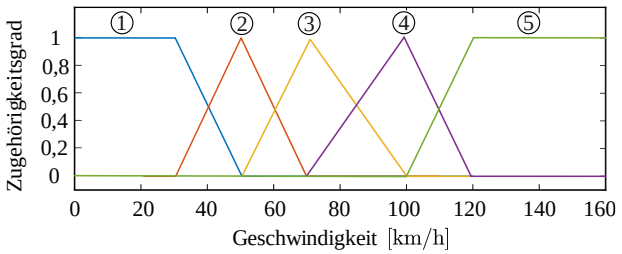


Bild 5 Zugehörigkeitsverteilung der Fahrzeuggeschwindigkeit

Zur geschwindigkeitsbezogenen Anfahrt- und Beschleunigungsmanöverbewertung auf Basis von Zugehörigkeitsgraden, werden in Anlehnung an [4] und durch Auswertungen eigener Fahrversuche, die Regeln in **Tabelle 2 (links)** definiert. Dadurch wird gewährleistet, dass Leistungsgrenzen berücksichtigt werden. So führt beispielsweise eine Längsbeschleunigung von 3 $\mathrm{m/s^2}$ bei 100 km/h dazu, dass die Fahrstilbewertung für das entsprechenden Beschleunigungsmanöver im Bereich sehr hoher Dynamik

liegt. Für Eingangswerte, deren Zugehörigkeitsgrad sich auf mehrere Bereiche aufteilt, erfolgt, durch Anwendung aller Regeln mit jeweiliger Inferenz, die Ermittlung von Zugehörigkeitswerten zu mitunter mehreren Fahrstilbewertungsbereichen (vgl. **Bild 1**). Über die Zugehörigkeitswerte wird jeder Bereich nach oben begrenzt, sodass der Flächeninhalt des resultierenden Trapezes berechnet werden kann. Die Abszissenkoordinate des Schwerpunkts der vereinigten Gesamtflächen aller Trapeze führt schließlich auf den Fahrstilbewertungswert. Eine detailliertere Beschreibung der Fuzzylogik-Methode sowie deren Regelanwendung kann beispielsweise [13] entnommen werden. Veranschaulichend zeigt **Bild 6** die Fahrstilbewertung von Anfahrts- und Beschleunigungsmanövern in Abhängigkeit maximal aufgetretener Längsbeschleunigung und der Manöveranfangsgeschwindigkeit.

Tabelle 2 Fahrstilbewertungsregeln

Regelnr.	*Beschl.-bereich*	*Operator*	*Geschw.-bereich*	*Fahrstilbereich*	*Regelnr.*	*Geschw.-bereich*	*Operator*	*Rel. Geschw. zu Tempolimit*	*Fahrstilbereich*
Dynamikbewertung					Fahrtdauerrelevanzbewertung				
1	1	and	1	G	1	1	and	1	G
2	2	and	1	G	2	2	and	1	G
3	3	and	1	D	3	3	and	1	D
4	4	and	1	S	4	4	and	1	H
5	1	and	2	G	5	5	and	1	S
6	2	and	2	G	6	1	and	2	G
7	3	and	2	H	7	2	and	2	G
8	4	and	2	S	8	3	and	2	D
9	1	and	3	G	9	4	and	2	D
10	2	and	3	D	10	5	and	2	S
11	3	and	3	H	11	1	and	3	D
12	4	and	3	S	12	2	and	3	D
13	1	and	4	G	13	3	and	3	S
14	2	and	4	D	14	4	and	3	S
15	3	and	4	S	15	5	and	3	S
16	4	and	4	S	16	1	and	4	H
17	1	and	5	D	17	2	and	4	H
18	2	and	5	S	18	3	and	4	H
19	3	and	5	S	19	4	and	4	S
20	4	and	5	S	20	5	and	4	S
					21	1	and	5	S
					22	2	and	5	S
					23	3	and	5	S
					24	4	and	5	S
					25	5	and	5	S

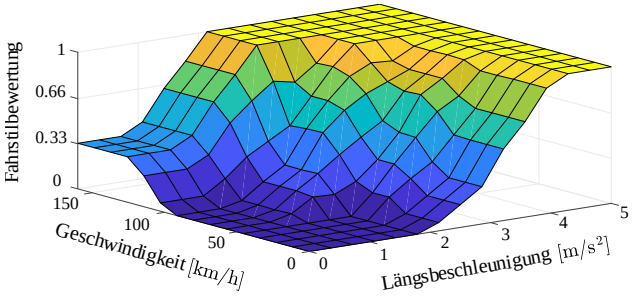


Bild 6 Zusammenfassung: Beschleunigungsmavöverbewertung

Die Analyse und Bewertung von Anhalte- und Verzögerungsmanövern erfolgt geschwindigkeitsunabhängig über die Ermittlung der maximal aufgetretenen negativen Längsbeschleunigung bzw. sogenannter Bremsverzögerung. Die in Fahrversuchen mit Bremsmanövern ermittelte, maximal aufgetretene Bremsverzögerung beträgt gerundet $7\ \mathrm{m/s^2}$. Die zugehörigkeitsgradabhängige Bereichsunterteilung der Fahrzeuglängsbeschleunigung, zur Bewertung von Anhalte- und Verzögerungsmanövern erfolgt daher gemäß **Bild 7**. Die Bewertungsregeln werden dabei so gewählt, dass die bereichsspezifischen Zugehörigkeitsgrade direkt für die Bereiche zur Fahrstilbewertung übernommen werden.

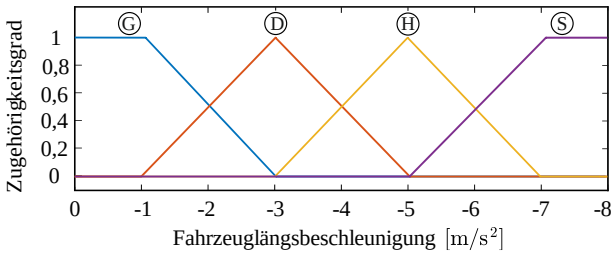


Bild 7 Zugehörigkeitsverteilung: Verzögerungsbewertung

Zur fahrdynamischen Bewertung detektierter Kurvenfahrten wird die gemessene, maximal auftretende Fahrzeugquerbeschleunigung herangezogen. Um Anhaltswerte für die Querbeschleunigung der dynamikbezogenen Fahrstilklassen zu generieren, wurden mehrere, von verschiedenen Personen durchgeführte, Testfahrzeug-Kurvenfahrten ausgewertet (vgl. **Bild 8**).

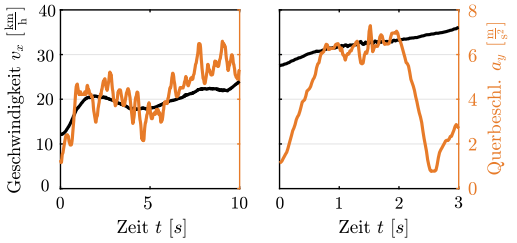


Bild 8 Fahrsituationen zur Parametrierung der Kurvenbewertung

Im Straßenverkehr lag die Querbeschleunigung bei normaler Fahrweise im gerundeten Mittel bei $\mathbf{4\ m/s^2}$. Bei geforderter, bewusst aggressiven Fahrweise konnte ein gemittelter Wert von etwa $\mathbf{7\ m/s^2}$ bestimmt werden, sodass ab diesem Querbeschleunigungswert eine Fahrzeugführung mit sehr hoher Dynamik unterstellt wird. Selbst bei bedacht geringdynamischer Fahrweise wurden im Mittel Querbeschleunigungswerte von bis zu $\mathbf{3\ m/s^2}$ gemessen. Anhand dieser Untersuchungen erfolgt die Bereichseinteilung der Querbeschleunigung gemäß **Bild 9**. Kurvenfahrten werden ebenfalls geschwindigkeitsunabhängig bewertet. Daher erfolgt die Definition der Zugehörigkeitsgradfunktionen so, dass auch hierbei eine direkte Zuordnung zu Fahrstilbewertungsbereichen erfolgen kann.

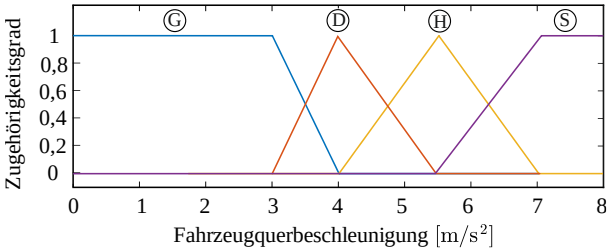


Bild 9 Zugehörigkeitsgradverteilung der Querbeschleunigung zur Bewertung von Kurvenfahrten

2.4 Fahrtdauerbezogene Fahrstilbewertung

Neben der Manövererkennung und dessen dynamikbezogener Bewertung, gilt es mit dem vorgestellten Fahrstilanalyseverfahren eine Aussage darüber zu treffen, wie viel Wert die fahrzeugführende Person auf einen geringen Energiebedarf, im Kontext einer kurzen Fahrtdauer, legt. Unter der Annahme, dass eine fahrzeugführende Person grundsätzlich ein Minimum an eingesetzter Energie favorisiert, wird komplementär dazu ermittelt, in welchem Maß diese Person eine möglichst geringe Fahrtdauer durch ihre Fahrweise forciert.

Eine bewusste Reduktion der Fahrtzeit für eine bestimmte Route kann letztlich vor allem durch eine Erhöhung der Fahrzeuggeschwindigkeit erreicht werden. Zwar ist die maximal zulässige Geschwindigkeit in der Regel durch die Straßenverkehrsordnung vorgeschrieben, jedoch neigen Personen, mit besonderem Fokus auf eine kurze Fahrtdauer, dazu, diese Tempolimits zu überschreiten. Zudem liegt nahe, dass vorwiegend Personen, die der Fahrtzeit wenig Relevanz beimessen, das Geschwindigkeitslimit nicht immer ausreizen, während durchschnittliche Fahrertypen ihre Fahrtgeschwindigkeit nahe dieses Limits wählen [14]. Ausgehend davon erfolgt die Bewertung der fahrerindividuellen Fahrtzeitrelevanz auf Basis der relativen Abweichung der Fahrzeuggeschwindigkeit zum jeweils vorherrschenden Tempolimit. Zur Ermittlung des geltenden Tempolimits, werden Navigationsdaten herangezogen und ausgewertet. Analog zu Abschnitt 2.3 wird eine Bereichseinteilung für das relative Verhältnis der Fahrzeuggeschwindigkeit zum aktuellen Tempolimit gemäß **Bild 11** definiert. Die Zuordnung erfolgt abhängig von den Fahrzeuggeschwindigkeitsbereichen aus **Bild 5** über **Tabelle 2 (rechts).** So kann beispielsweise gewährleistet werden, dass die Fahrtdauerrelevanz durch zehnprozentige Tempolimitüberschreitung bei erlaubten 100 km/h höher bewertet wird als bei erlaubten 30 km/h. Zur Ermittlung eines fahrtdauerbezogenen Fahrstilbewertungswertes werden die Bereiche aus **Bild 2** verwendet, sodass während unbeeinflusster Fahrt bewertet wird, ob die fahrzeugführenden Person einen „geringen“, „durchschnittlichen“, „hohen“ oder „sehr hohen“ Wert auf eine kurze Fahrtdauer legt. Diese Bewertung wird sowohl bei erkanntem Vorderfahrzeug (vgl. Abschnitt 2.2) als auch während erkannten Fahrmanövern verworfen, da eine Tempolimitabweichung dabei zumeist aus einer manöverbedingten Geschwindigkeitsanpassung resultiert. Komplementär zum Bewertungswert der Fahrtdauerrelevanz, wird durch Subtraktion dieses Wertes von 1 ein Maß dafür definiert, wie viel Wert auf eine energiesparende Fahrweise gelegt wird.

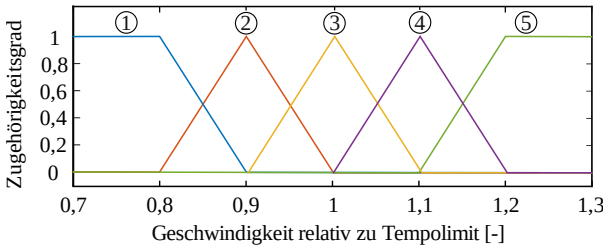


Bild 11 Zugehörigkeitsgradverteilung der Geschwindigkeit relativ zum Tempolimit

3 Evaluation der Fahrstilbewertung

Zur Evaluation des vorgestellten Fahrstilanalyseverfahrens zeigt **Bild 10** exemplarisch die dynamikbezogene Bewertung diverser Fahrsituationen für zwei unterschiedliche Fahrertypen. Dabei wird links ein Anhaltemanöver durch überschreiten des Längsbeschleunigungsschwellwertes von 0,4 m/s^2 erkannt. Während des Manövers tritt eine maximale Bremsverzögerung von $-5{,}07\ \mathrm{m}/s^2$ auf. Dieser Wert liegt mit einem Zugehörigkeitsgrad von 0,97 im Bereich hoher und mit 0,03 im Bereich sehr hoher Dynamik. Gemäß **Bild 2** ergibt sich damit, über eine Flächeninhalts- und Schwerpunktsberechnung, ein Fahrstilbewertungswert von 0,68 für das Manöver. Aufgrund einer sehr hohen Längsbeschleunigung von 4,13 m/s^2 im anschließend erkannten Anfahrtsmanöver, wird dieses durch Anwendung der Regeln aus **Tabelle 2 (links)** ausschließlich dem Bereich „S“ und somit einem Bewertungswert von 1 zugeordnet. Darüber hinaus werden zwei Kurvenfahrten nach Überschreiten des Lenkradwinkelschwellwertes erkannt. Während die erste aufgrund einer Dauer von unter zwei Sekunden verworfen wird, erfolgt die Bewertung der zweiten über eine maximalen Querbeschleunigung von 5,73 m/s^2, zu 0,7. Unter Betrachtung von r_{VF} ist ersichtlich, dass es links zu keiner signifikanten Beeinflussung durch andere Verkehrsteilnehmer gekommen ist. In der rechts gezeigten Fahrsituation wird hingegen deutlich, wie sowohl ein Kurven- als auch ein Verzögerungsmanöver aufgrund eines verhältnismäßig geringen Abstandes zum Vorderfahrzeug ($r_{VF} < 1$) verworfen wird. Bei der arithmetischen Mittelung der Manöverbewertungen zu einem Gesamtwert für den dynamikbezogenen Fahrstil, bleiben diese Manöver letztlich unberücksichtigt. Zum Ende der Fahrsituation ist zudem erkennbar, wie zwei direkt aufeinanderfolgende Beschleunigungsmanöver zu einem zusammengefasst werden, da zwischen ihnen der Längsbeschleunigungsschwellwert lediglich kurzzeitig unterschritten wird.

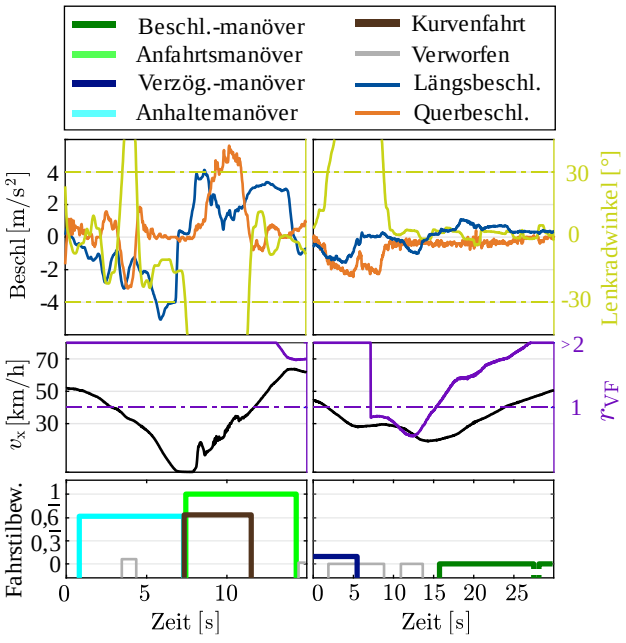


Bild 10 Dynamikbezogene Fahrstilbewertung

Die fahrtzeitbezogene Fahrstilbewertung ist in **Bild 12** veranschaulicht. Es wird deutlich, dass die Bewertung der Fahrtzeitpräferenzen lediglich in Bereichen erfolgt, in denen weder ein Vorderfahrzeug das Fahrerverhalten beeinflusst, noch eine manöverbedingte Geschwindigkeitsanpassung vorliegt. So wird die fahrzeitbezogene Fahrstilbewertung selbst bei verworfenen Manövern weiterhin ausgesetzt. Für alle Fahrabschnitte in denen die Bewertung nicht ausgesetzt wird, kann schließlich ein durchschnittlicher Bewertungswert individueller Fahrdauerpräferenzen berechnet werden. Über die Differenz zu 1 lässt sich dieser in den Kontext eines Bestrebens Energie einzusparen setzen.

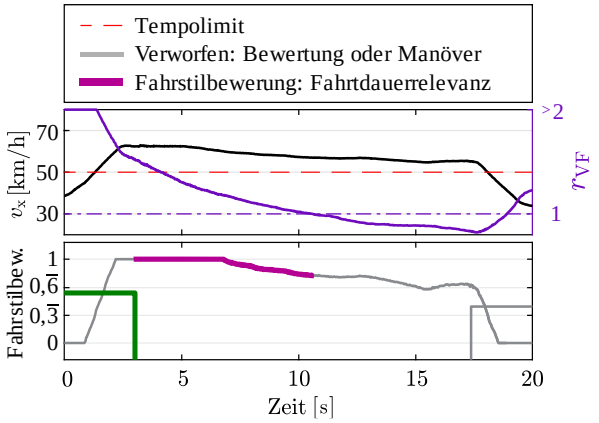


Bild 12 Fahrzeitbezogene Fahrstilbewertung

4 Zusammenfassung

In dieser Arbeit wird ein Verfahren zur Fahrstilbewertung vorgestellt, das auf einem regelbasierten Fuzzylogik-Ansatz basiert. Einerseits lassen sich mit diesem Verfahren, über eine Detektion und Bewertung relevanter Fahrmanöver, dynamisch Fahrervorlieben ermitteln. Andererseits wird ermöglicht individuelle Präferenzen der fahrzeugführenden Person im Zielkonflikt einer kurzen Fahrtdauer und einer energiesparenden Fahrweise abzuschätzen.
Diese Abschätzung beruht derzeit allerdings allein auf der Bewertung der gewählten Fahrzeuggeschwindigkeit in Relation zum vorherrschenden Tempolimit. Weiterführend gilt es daher bei der energiebedarfsbezogenen Fahrstilbewertung differenziert zu analysieren, inwieweit bei der Fahrzeugführung energiebewusst gehandelt wird. Dazu ließe sich das vorgestellte Verfahren erweitert. Um zu ermitteln, inwieweit sich die fahrzeugführende Person vorrausschauenden verhält, ließen sich beispielsweise Abstände zum Vorderfahrzeug und Tempolimitänderungen in Zusammenhang mit Gaspedalstellungen und Bremsdrücken bringen. Zudem bleibt einschränkend festzuhalten, dass das vorgestellte Verfahren auf Regeln aus Literaturwerten und fahrzeugspezifischen Leistungsgrenzen basiert. Um generalisierbare Analyseergebnisse zu erhalten, sind Fahrstilbewertungswerte, Regeln und Schwellwerte im Rahmen empirischer Nutzerstudien zu verifizieren. Anhand der dabei gewonnen Erkenntnisse lassen sich diese dann weiter anpassen. Dazu bietet unter anderem das Anlernen eines adaptiven Neuro-Fuzzy-Inferenzsystems eine vielversprechende Möglichkeit die Generalisierbarkeit des Analyseverfahrens signifikant zu steigern.

5 Literatur

[1] Stiller, C.: Fahrerassistenzsysteme - Von realisierten Funktionen zum vernetzt wahrnehmenden, selbstorganisierenden Verkehr. In: Maurer, M. u. Stiller, C. (Hrsg.): Fahrerassistenzsysteme mit maschineller Wahrnehmung. Berlin/Heidelberg: Springer-Verlag 2005, S. 1–20

[2] Hülsebusch, D.: Fahrerassistenzsysteme zur energieeffizienten Längsregelung - Analyse und Optimierung der Fahrsicherheit, Karlsruher Institut für Technologie 2018

[3] Hartwich, F., Beggiato, M. u. Krems, J. F.: Driving comfort, enjoyment and acceptance of automated driving - effects of drivers' age and driving style familiarity. Ergonomics 61 (2018) 8, S. 1017–1032

[4] Aljaafreh, A., Alshabatat, N. u. Najim Al-Din, M. S.: Driving style recognition using fuzzy logic. In: Staff, I. (Hrsg.): 2012 IEEE International Conference on Vehicular Electronics and Safety. [Place of publication not identified]: IEEE 2012, S. 460–463

[5] Johnson, D. A. u. Trivedi, M. M.: Driving style recognition using a smartphone as a sensor platform. Driving style recognition using a smartphone as a sensor platform - 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE 2011 - 2011, S. 1609–1615

[6] Dynamic Time Warping. In: Müller, M. (Hrsg.): Information retrieval for music and motion. With 26 tables. Berlin: Springer 2007, S. 69–84

[7] Eren, H., Makinist, S., Akin, E. u. Yilmaz, A.: Estimating driving behavior by a smartphone. IEEE Intelligent Vehicles Symposium (IV), 2012. 3-7 June 2012, Alcalá de Henares, Madrid, Spain. Piscataway, NJ: IEEE 2012, S. 234–239

[8] Marina Martinez, C., Heucke, M., Wang, F.-Y., Gao, B. u. Cao, D.: Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey. IEEE Transactions on Intelligent Transportation Systems 19 (2018) 3, S. 666–676

[9] Freuer, A.: Ein Assistenzsystem für die energetisch optimierte Längsführung eines Elektrofahrzeugs. Wiesbaden: Springer Fachmedien Wiesbaden 2016

[10] Dorr, D., Grabengiesser, D. u. Gauterin, F.: Online driving style recognition using fuzzy logic. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE 2014 - 2014, S. 1021–1026

[11] bussgeldkatalog.org: Sicherheitsabstand zum Vordermann, 2020. https://www.bussgeldkatalog.org/sicherheitsabstand-landstrasse/#sicherheitsabstand_zum_vordermann, abgerufen am: 10.01.2021

[12] Broberg, T. u. Dukic Willstrand, T.: Safe mobility for elderly drivers--considerations based on expert and self-assessment. Accident; analysis and prevention 66 (2014), S. 104–113

[13] Zadeh, L. A., Klir, G. J. u. Yuan, B.: Fuzzy sets, fuzzy logic, and fuzzy systems. Selected papers. Advances in fuzzy systems, Bd. 6. Singapore: WORLD SCIENTIFIC 1996

[14] Hautzinger, H.: Regelverstöße im Straßenverkehr. Berlin: GDV 2011

Robotergestützte Handhabung im kryogenen Arbeitsumfeld
Robot-supported handling in a cryogenic working environment

Dr.-Ing. Frank Ihmig, Fraunhofer-Institut für Biomedizinische Technik IBMT, 66280 Sulzbach/Saar, Deutschland, frank.ihmig@ibmt.fraunhofer.de
Dipl.-Ing. Philipp Jahn, Leibniz Universität Hannover, Institut für Montagetechnik, 30823 Garbsen, Deutschland, jahn@match.uni-hannover.de
Dipl.-Ing. Timo Koch, Fraunhofer-Institut für Biomedizinische Technik IBMT, 66280 Sulzbach/Saar, Deutschland, timo.koch@ibmt.fraunhofer.de
Dipl.-Phys. Margit Biehl, Fraunhofer-Institut für Biomedizinische Technik IBMT, 66280 Sulzbach/Saar, Deutschland margit.biehl@ibmt.fraunhofer.de
Prof. Dr.-Ing. Annika Raatz, Leibniz Universität Hannover, Institut für Montagetechnik, 30823 Garbsen, Deutschland raatz@match.uni-hannover.de

Kurzfassung

In der heutigen Industrie ist Automatisierung ein allgegenwärtiger Faktor, selbst in Nischenanwendungen wie der Kryokonservierung. Aber die manuelle Handhabung von biologischen Proben ist in den meisten Kryobanken immer noch die Regel. Diese konservierten Proben werden bei tiefen Temperaturen von bis zu -196 °C unter Verwendung sperriger Schutzkleidung und Sicherheitsausrüstung eingelagert, umgeordnet und ausgelagert. Dabei besteht einerseits ein erhebliches Verletzungsrisiko für die Anwender durch Kaltverbrennungen, andererseits kann die Probenintegrität während des Handhabungsprozesses durch jede Art von Erwärmung beeinträchtigt werden. Die Entwicklung eines Parallelroboters ist ein aussichtsreicher Ansatz, um ein robotergestütztes Handhabungssystem im kryogenen Arbeitsumfeld zu realisieren. In diesem Beitrag beschreiben wir die Konstruktion eines monolithischen Greifersystems und die Entwicklung der integrierten Steuerelektronik. Ein Elektromagnet kommt als Aktor zum Einsatz. Wir berichten über die Charakterisierung der Komponenten, die Notwendigkeit einer Regelung sowie erste Ergebnisse zur Funktionstüchtigkeit.

Abstract

In today's industry, automation is an omnipresent factor, even in niche applications such as cryopreservation. But manual handling of biological samples is still the standard in most cryobanks. These preserved samples are stored, sorted and retrieved at low temperatures down to -196 °C using bulky protective clothing and safety equipment. There is a significant risk of injury to the user from cold burns, and the integrity of the samples can be compromised by any kind of warming during the handling process. The development of a parallel robot is a promising approach to realize a robot-supported handling system in a cryogenic working environment. In this paper, we describe the design of a monolithic gripper system and the development of the integrated control electronics. An electromagnet is used as actuator. We report on the characterization of the components, the necessity of a closed-loop control and first results on the functionality.

1 Einleitung

Die Kryokonservierung ermöglicht eine unbefristete Lagerung von biologischem Zellmaterial, wie beispielsweise Stammzellen oder Gewebe. Hierfür werden minimale Mengen des Zellmaterials in Kryoröhrchen portioniert, die meist in der Gasphase eines Kryolagerbehälters aufbewahrt werden. Im Allgemeinen bezeichnet Kryokonservierung das Einfrieren und Lagern von Zellmaterialien bei Temperaturen unterhalb von -130 °C; als Kühlmittel wird flüssiger Stickstoff benutzt. Die Gasphase bildet einen Temperaturgradienten, der von oben nach unten etwa zwischen -130 °C und -190 °C liegt.

Da es für die klassische Probenhandhabung (hierzu zählt das Ein- und Auslagern, sowie das Sortieren und Umlagern der Kryoröhrchen) bisher kaum Automatisierungslösungen gibt, um die Probenhandhabung bei diesen tiefkalten Temperaturen zu ermöglichen, die für die Probenqualität unkritisch sind, wird dies nach wie vor unter aufwendigen Schutzmaßnahmen manuell durchgeführt. Hierbei werden die Lagertürme von Hand aus dem Kryobehälter gezogen, wobei sie durch die Luftfeuchte vereisen. Zudem werden durch dieses Vorgehen alle Proben der Raumtemperatur ausgesetzt, wodurch ein Risiko für das Lagergut durch An- oder Auftauen entsteht. Bei der Handhabung stellen die manuellen Tätigkeiten ein Verletzungsrisiko für den Anwender durch Kaltverbrennungen dar. Außerdem besteht ein erhöhtes Risiko der Vertauschung oder gar des Verlustes von Kryoröhrchen.

Eine weitere Herausforderung besteht in der Reduzierung der notwendigen Handhabungszeiten. Die Arbeitsabläufe wie Einlagern, Auslagern und Sortieren von einzelnen Proben sind sehr zeitaufwendig, insbesondere wenn es sich um größere Mengen handelt, z.B. beim Zusammenstellen von ausgesuchten Proben für klinische Studien. Ein hoher Pro-

bendurchsatz ist so nicht möglich. Ein grundlegendes Hindernis sind hier vor allem die notwendigen Schutzmaßnahmen für das Personal, insbesondere die materialbedingt bei diesen Temperaturen steifen Isolierhandschuhe. Um die Gefahren der manuellen Handhabung von Kryoröhrchen zu minimieren und gleichzeitig die technischen Herausforderungen zu meistern, die eine kryogene Arbeitsumgebung mit sich bringt, ist eine vollständige Automatisierung wünschenswert. Die logische Konsequenz ist die Entwicklung eines robotergestützten Handhabungssystems.

Im Rahmen des von der DFG geförderten Projektes "Methoden zur Automatisierung von Handhabungsprozessen unter kryogenen Umgebungsbedingungen" wird am Institut für Montagetechnik der Leibniz Universität Hannover in Kollaboration mit dem Fraunhofer-Institut für Biomedizinische Technik IBMT mittelfristig ein Parallelroboter entwickelt, um eine solche Vollautomatisierung zu realisieren. Die Handhabung der tiefkalten Proben in einem Lagerbehälter soll mithilfe eines monolithisch gefertigten Greifersystems realisiert werden. Monolithische Konstruktionen bestehen aus nur einem kohärenten Bauteil, es gibt also keine gegeneinander bewegten Komponenten. Die Greifbewegung selbst muss somit durch eine elastische Verformung des Greifers erfolgen. Als Aktor wird ein Elektromagnet eingesetzt.

Die für die Steuerung notwendigen Komponenten wie Mikrocontroller und Messverstärker sollen am Träger des Greifersystems montiert werden und unterliegen denselben extremen Umgebungsbedingungen. Aufgrund der zu erwartenden temperaturabhängigen Änderung der elektrischen Eigenschaften des Elektromagneten ist eine Regelung des Aktorstroms erforderlich. Die Energie zur Versorgung des Greifersystems soll in der geplanten Anwendung kabellos über induktiv gekoppelte Schwingkreise übertragen werden. In diesem Beitrag beschreiben wir die Konstruktion des Greifersystems und den Aufbau der integrierten Steuerelektronik. Wir berichten über die Charakterisierung der elektronischen Komponenten sowie über erste Ergebnisse zur Funktionstüchtigkeit des Greifersystems bei kryogenen Temperaturen.

2 Grundlagen

In diesem Kapitel erläutern wir kurz einige Grundlagen zu nachgiebigen Mechanismen und zum Betrieb elektronischer Komponenten bei tiefen Temperaturen.

2.1 Nachgiebige Mechanismen

Im Gegensatz zu Starrkörpermechanismen, die aus starren, gegeneinander bewegten Teile bestehen, basieren nachgiebige Mechanismen auf der lokal begrenzten, elastischen Verformung bestimmter Bauteilbereiche. In ihrer einfachsten Form sind diese Bereiche als Dünnstellen mit einem stark reduzierten Querschnitt ausgeführt. In komplexeren Strukturen wie dem in diesem Projekt angestrebten Parallelroboter werden kompliziertere Strukturen in Form von Festkörpergelenken eingesetzt, hier in **Bild 1** gezeigt. Durch den Einsatz moderner Werkstoffe können diese Gelenke vergleichbare Belastungen ertragen wie konventionelle Drehgelenke ähnlicher Baugröße. Monolithische Konstruktionen, wie der im Folgenden vorgestellte Greifer, haben eine Reihe von Vorteilen, die sie für den Einsatz in einer kryogenen Arbeitsumgebung prädestinieren: Sie sind reibungs-, spiel- und verschleißfrei, können ohne Schmierung betrieben werden und sind unempfindlich gegenüber der Einwirkung von Schmutz oder Fremdkörpern. Die Ausnutzung der elastischen Verformung als Bewegungskonzept präsentiert aber auch einige Herausforderungen: Die erreichbaren Biegewinkel monolithischer Strukturen sind verhältnismäßig klein und die mechanischen Eigenschaften der meisten verwendbaren Materialien verschlechtern sich unter kryogenen Bedingungen noch einmal erheblich. Um diese Probleme zu lösen, sind komplexe geometrische Konstruktionen notwendig.



Bild 1 Komplexer nachgiebiger Mechanismus in Form von kaskadierten Festkörpergelenken

2.2 Tieftemperaturelektronik

Bei den Begriffen Tieftemperaturelektronik oder Kryoelektronik denkt man häufig an Supraleitung. Kryoelektronik basiert in praktischen Anwendungen aber überwiegend auf Halbleitern oder auf einer Kombination aus beiden. Typische Anwendungen sind hochempfindliche Strahlungsdetektoren, Mikrowellenempfänger, Raumfahrzeuge und seit neuestem auch Quantencomputer. Aus der Literatur ist bekannt, dass einige Halbleiterbauelemente wie z.B. MOSFETs und auf der CMOS-Technologie basierende ICs eine verbesserte Performanz bei tiefen Temperaturen haben können, wie z. B. niedrigerer ON-Widerstand und schnellere Schaltgeschwindigkeit [1].

Passive Bauelemente, wie Kondensatoren, können unterschiedliches Verhalten in Kapazität und Verlustfaktor in Abhängigkeit des Dielektrikums aufweisen [2]. Während Elektrolytkondensatoren ungeeignet sind, zeigen Kondensatoren mit Dielektrikum aus NP0/C0G, PPS oder PP recht stabiles Verhalten bei tiefen Temperaturen. Unter den Widerständen sind solche basierend auf Dünnfilm, Metallfilm und Drahtwicklung gute Kandidaten [3].

Für die Aufbau- und Verbindungstechnik sind vor allem Anwendungen mit häufigen Temperaturzyklen über einen weiten Temperaturbereich kritisch. Dies ist bei der hier beschriebenen Anwendung nicht der Fall, da die Kühlkette bei kryokonservierten Proben nicht unterbrochen werden darf.

Bei der Verwendung von Weichlotlegierungen ist zu beachten, dass mit fallender Temperatur die Festigkeit steigt während die Duktilität sinkt [4]. Eine PbSn-Legierung als typisches Weichlot kann bei kryogenen Temperaturen verwendet werden und behält ihre Duktilität bei, wenn sie einen hohen Pb-Anteil von mindestens 60% hat. Typische Leiterplatten aus glasfaserverstärktem Epoxid sind ebenfalls für tiefe Temperaturen geeignet [5].

3 Systemdesign und Messmethoden

Das Systemdesign beinhaltet die Konstruktion des monolithischen Greifers und die Entwicklung der integrierten Steuerelektronik. Außerdem beschreiben wir in diesem Kapitel den experimentellen Teil zur Charakterisierung der elektronischen Komponenten.

3.1 Mechanische Konstruktion

Die Greiferkonstruktion wurde auf Basis eines Elektromagneten erarbeitet (siehe **Bild 2**). Elektromagnete sind robuste, langlebige und kostengünstige Aktoren, die in großer Auswahl kommerziell verfügbar sind. Allerdings muss die Temperaturabhängigkeit der elektrischen Eigenschaften eines Elektromagneten berücksichtigt werden, um eine konstante Magnetkraft bei tiefen Temperaturen zu gewährleisten. Wir verwenden einen Haftmagneten (Red Magnetics ITS-MS-3025) mit einer Anzugskraft von 100 N und einer Betriebsspannung von 12 VDC, der eine Anzugsplatte aus VACOFLUX® 50 (Eisen-Cobalt-Legierung mit besonders hoher Sättigungsmagnetisierung und geringer Remanenz), an seine Stirnfläche zieht, und dadurch eine Zusammenführung der Greifzange bewirkt.

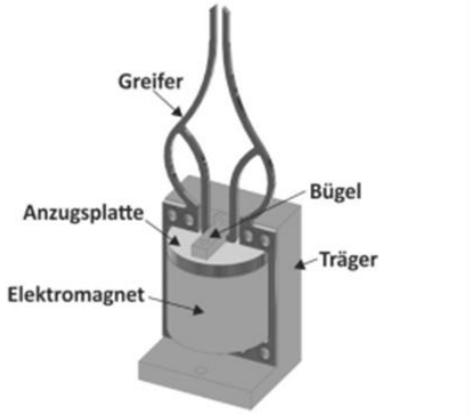


Bild 2 Monolithisches Greifersystem mit Elektromagnet

Ähnlich wie bei den geplanten Festkörpergelenken des Parallelroboters wurde für den Greifer ein monolithisches Greifprinzip umgesetzt. Dabei wird der innen hohle Deckel (8,6 mm Außendurchmesser) der ausgesuchten Kryoröhrchen (1,1 ml Kryoröhrchen mit Füllgewicht 1,8 g für 96er RoboRack von Micronic, Niederlande) mit von innen auseinanderspreizenden Zangen gegriffen.
Um eine energieeffiziente Handhabung zu gewährleisten, muss der Greifer lediglich zum Einführen der Greifzange in die Deckelvertiefung und, nach erfolgtem Transfer, zum Freisetzen des Röhrchens kurzzeitig bestromt werden. Die Haltekraft während des Transfervorgangs wird stromlos durch die verformungsbedingte Rückstellkraft des Werkstoffs (Edelstahl 1.4301) zur Verfügung gestellt. Mittels Simulation der dabei auftretenden Spannungen und Verschiebungen im CAD-Programm SolidWorks wurde das Design optimiert.

3.2 Schaltungsentwicklung und Regelung

Für den Greifer bestand die Anforderung eine Steuerelektronik mit Mikrocontroller zu entwickeln, die auch im Tieftemperaturbereich bis -196 °C funktionstüchtig bleibt. Es kamen konventionelle elektronische Bauelemente zum Einsatz. **Bild 3** zeigt den Schaltungsteil für die Aktorsteuerung.

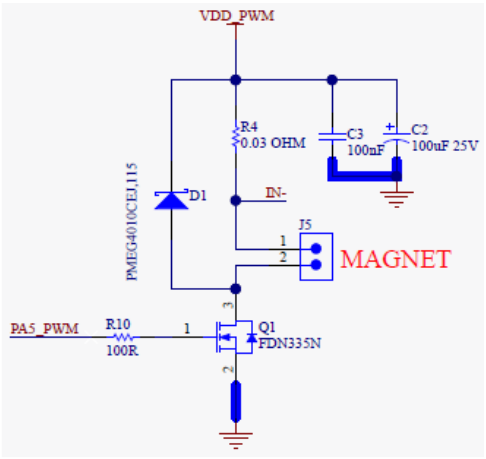


Bild 3 Schaltungsteil für die PWM-Aktorsteuerung

Die Greiferkonstruktion umfasst zwei Zustände: Öffnen und Schließen. Somit ist die Ansteuerung des Elektromagneten eine An-Aus-Anwendung. Dies erfolgt mit einem pulsweitenmodulierten (PWM) Steuersignal zum schnellen Schalten eines MOSFETs (Q1, On Semiconductor FDN337N), der in Reihe zum Aktor geschaltet ist. Parallel zum Elektromagnet ist eine Schottky-Diode (D1, Nexperia PMEG4010BEA) geschaltet, um im ausgeschalteten Zustand des MOSFETs die noch vorhandene Energie der Induktivität in die Schaltung zurückzuführen, ohne den MOSFET zu beschädigen. Das PWM-Signal wird mit einem Mikrocontroller (STMicroelectronics STM32L0) generiert, der aus einem ARM Cortex-M0+ 32-Bit-RISC-Kern besteht. Für die Stromüberwachung des Aktors wird ein Strommessverstärker (U2, Texas Instruments INA190A4) eingesetzt (**Bild 4**).

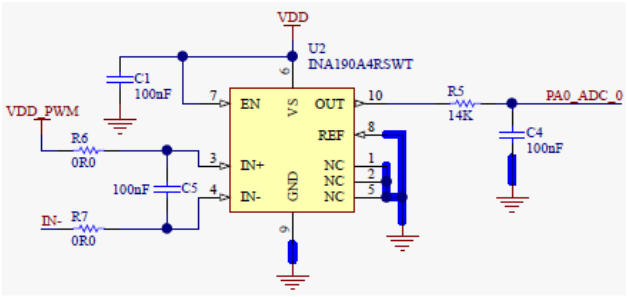


Bild 4 Schaltungsteil für die Stromüberwachung des Aktors

Messungen am Aktor zeigten, dass für das Schließen des Greifers bei Raumtemperatur ein Tastverhältnis von etwa 50 % ausreicht. Für die Simulation mit LTspice wurde der Elektromagnet als Induktivität mit Reihenwiderstand dargestellt und ein PWM-Signal mit einer Frequenz von 25 kHz und einem Tastverhältnis von 50 % generiert.
Ein weiterer wichtiger Schaltungsteil der Steuerelektronik ist die Spannungskonvertierung und -regelung. Damit werden aus der induktiv übertragenen Energie die Betriebsspannungen für den Elektromagneten und die Elektronik generiert. Um die 12 VDC für den Elektromagneten zu generieren, wurde ein Step-Down Regler vorgesehen. Für die Erzeugung der 3,3 VDC der Elektronik, wurde ein LDO-Regler eingesetzt.
Aufgrund der zu erwartenden starken temperaturabhängigen Abnahme des Widerstandes des Elektromagneten wurde eine Regelung entworfen, um das Tastverhältnis des PWM-Signals zur Bestromung des Elektromagneten in Abhängigkeit der Umgebungstemperatur zu regeln. Das Diskretisieren des Reglers ist abhängig von der Abtastfrequenz des AD-Wandlers. Die Abtastfrequenz wurde mit 100 Hz zum einen so gewählt, dass der Mikrocontroller genug Zeit für das Durchführen des Regelalgorithmus zwischen zwei Messwerten hat, und zum andern so dimensioniert, dass der resultierende diskrete Regler eine schnelle Einschwingzeit hat. Der Regelkreis wurde so konzipiert, dass der gemessene Strom in Abhängigkeit der temperaturabhängigen Änderung des Aktorwiderstands auf den Regelsollwert gebracht wird, um eine konstante Haltekraft über den betrachteten Temperaturbereich von +20 °C bis -196 °C zu erzeugen.
Der Regelkreis hat das Verhalten eines Verzögerungsgliedes 1. Ordnung (PT1-Glied). Die Anwendung erfordert einerseits eine schnelle Reaktion der Aktorik, d.h. eine hohe Stellgeschwindigkeit, andererseits kann sich aufgrund der Temperaturabhängigkeit des Aktors die Zeitkonstante des Regelkreises verstellen. Das bedeutet, dass das Regelverhalten des Systems vom Arbeitspunkt abhängt. Zudem sollte der Regler möglichst keine verbleibende Regelabweichung vorweisen.
Aus diesen Gründen wurde ein PID-Regler ausgewählt, der keine Arbeitspunkteinstellung benötigt und trotzdem ein stabiles Regelverhalten über den gesamten Arbeitsbereich der Aktorik liefern kann. Die Einstellung der Regelparameter und die Simulation des Regelkreises erfolgte mit Hilfe von MATLAB Simulink. Die Simulation zeigte, dass bei abnehmendem elektrischen Widerstand des Aktors mit einer abnehmenden Anstiegszeit und bei den niedrigsten Temperaturen mit einem kurzen Überschwingen zu rechnen ist.

3.3 Charakterisierung der elektronischen Komponenten

Der ausgewählte Elektromagnet und die Halbleiterkomponenten der Schaltung (Mikrocontroller, Strommessverstärker, Spannungsregler, MOSFET und Schottky-Diode) wurden bei tiefen Temperaturen zunächst im flüssigen Stickstoff bei -196 °C getestet und nach der Gewährleistung ihrer Funktionsfähigkeit wurden sie in einem computergesteuerten Einfrierautomaten (Sylab Icecube 14S) mit definierter Kühlrate im Temperaturbereich von +20 °C bis -180 °C charakterisiert. Der Temperaturbereich wurde in Schritten von -20 °C mit einer Haltezeit von 5 Minuten bei jedem Temperaturschritt und einer Kühlrate von 10 °C/min durchfahren.
Als Messequipment wurden der Präzisions-Impedanz-Analyzer Agilent 4294A, das Digitalmultimeter Keithley 2000 und das Oszilloskop Agilent DSO1012A verwendet. Für den Elektromagneten wurden die Induktivität und der elektrische Widerstand gemessen. Als wesentlicher temperaturabhängiger Parameter des MOSFETs wurde der Durchlasswiderstand (ON-Widerstand) ermittelt. Für die Charakterisierung des Strommessverstärkers war es wichtig herauszufinden, wie sich der Verstärkungsfaktor und die Stromaufnahme mit fallender Temperatur ändern. Für den Mikrocontroller wurden die Taktfrequenz und die AD-Wandlung untersucht. Bei der Schottky-Diode wurde die Durchlassspannung in Abhängigkeit der Temperatur gemessen. Für die Spannungsregler wurde der Wirkungsgrad für verschiedene Lastfälle bei -196 °C im Vergleich zu Raumtemperatur ermittelt.
Der funktionale Test der realisierten Steuerelektronik erfolgte eingetaucht im flüssigen Stickstoff sowie im Einfrierautomaten mit etwas feinerer Abstufung im Temperaturbereich von -130 °C bis -180 °C. Dieser Temperaturbereich wurde in Schritten von -10 °C mit einer Haltezeit von 4 Minuten bei jedem Temperaturschritt und einer Kühlrate von 10 °C/min durchfahren. Außerdem erfolgte ein funktionaler Test des Greifersystems im Dauerbetrieb, eingetaucht im flüssigen Stickstoff, um eine erste Einschätzung über die Belastbarkeit des Aktors geben zu können.

4 Ergebnisse

Im Folgenden beschreiben wir die Messergebnisse zur Charakterisierung der elektronischen Komponenten sowie erste Resultate zur Funktionstüchtigkeit des Greifersystems bei kryogenen Temperaturen.

4.1 Komponentenverhalten bei kryogenen Temperaturen

Der elektrische Widerstand und die Induktivität des Aktors zeigten eine lineare Abnahme beim Abkühlen von Raumtemperatur auf -196 °C. Für die Induktivität ergab sich insgesamt eine leichte Verringerung von 147,70 mH auf 131,82 mH um etwa 11 %. Demgegenüber nahm der elektrische Widerstand des Elektromagneten beim Abkühlen um das 7,8-fache ab und zwar von 37,5 Ω auf 4,8 Ω. Die Messung der Schottky-Diode ergab ebenfalls ein lineares Verhalten mit einer Zunahme der Durchlassspannung um das 3,2-fache. Dioden zeigen in der Regel einen linearen Anstieg der Durchlassspannung bei tiefen Temperaturen, weswegen sie auch als kryogene Temperatursensoren eingesetzt werden können.
Das Ergebnis für die Messung des ON-Widerstandes des MOSFET-Schalters ist in **Bild 5** dargestellt. Das Diagramm zeigt die Messpunkte und einen polynomiellen Fit dritter

Ordnung. Der ON-Widerstand hat einen nicht linearen Verlauf im Temperaturbereich von +20 °C bis -196 °C. Mit fallender Temperatur bis -140 °C nimmt er um etwa 58 % ab, gefolgt von einem Anstieg um das 1,6-fache. Dieser Verlauf folgt dem Trend der temperaturabhängigen Mobilität freier Ladungsträger im Halbleiter. Zunächst nimmt die Ladungsträgermobilität aufgrund der abnehmenden Gitterstreuung kontinuierlich mit fallender Temperatur zu [6], was mit der Abnahme des ON-Widerstandes korrespondiert. Sobald allerdings die Streuung an Fremdatomen dominierend wird, steigt der ON-Widerstand wieder an [7].

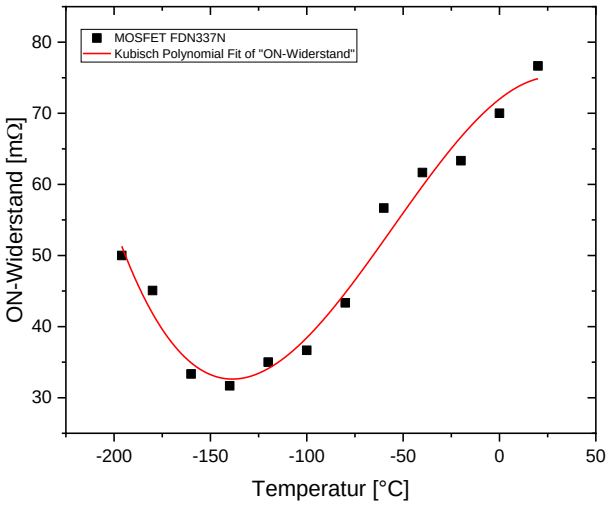


Bild 5 Verlauf des ON-Widerstandes des MOSFETs als Funktion der Temperatur

Die Messung der Leistungsaufnahme des Strommessverstärkers ergab bei -196 °C insgesamt eine Abnahme von etwa 54 % im Vergleich zu Raumtemperatur. Der Verstärkungsfaktor blieb hingegen sehr stabil mit einer geringfügigen Abnahme von etwa 4 % bei -196 °C.
Die Messergebnisse der Taktfrequenz des Mikrocontrollers zeigten eine kontinuierliche Abnahme mit fallender Temperatur von knapp 16 MHz bei Raumtemperatur auf etwa 13,5 MHz im flüssigen Stickstoff. Dies entspricht einer Abnahme von insgesamt etwa 15 % über den betrachteten Temperaturbereich. Diese Änderung führte zu keiner Beeinträchtigung der Timer-Funktionalität und der AD-Wandlung.
Fünf verschiedene Step-Down Regler wurden evaluiert, um einen geeigneten Typ zu identifizieren. Als geeignet erwies sich der LM5164 von Texas Instruments, der in MOSFET-Technologie aufgebaut ist. Dieser Buck-Konverter kann Eingangsspannungen bis 100 VDC verarbeiten und liefert Ausgangsströme bis 1 A. Darüber hinaus ist die Taktrate des Reglers einstellbar. **Bild 6** zeigt die Messergebnisse für den Wirkungsgrad des Reglers bei Raumtemperatur und im flüssigen Stickstoff. Der Regler weist, außer bei hoher Last, einen niedrigeren Wirkungsgrad bei -196 °C auf als bei Raumtemperatur. Ist die Eingangsspannung nahe der Ausgangspannung (Vin = 15 VDC und Vout = 12 VDC, schwarz gestrichelte Kurve) sinkt der Wirkungsgrad bei geringer Last noch stärker, da der höhere Ruhestrom bei tiefen Temperaturen immer stärker zum Tragen kommt. Insgesamt neun LDO-Regler wurden evaluiert, um einen geeigneten Typ zu identifizieren. Hier erwies sich der TPS7B6933 von Texas Instruments als geeignet.

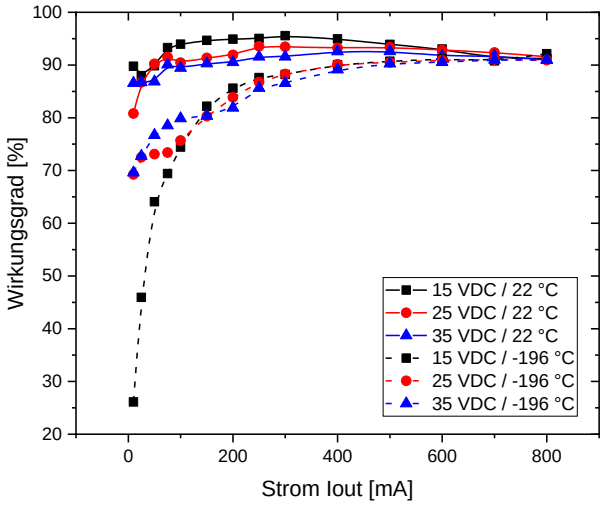


Bild 6 Wirkungsgrad des Step-Down-Reglers bei Raumtemperatur und im flüssigen Stickstoff

4.2 Funktionsmuster

Aus dem Komponentenverhalten bei kryogenen Temperaturen resultierten keine Einschränkungen für die geplante Anwendung. Die Steuerelektronik wurde in Form einer 2-lagigen FR4-Platine (90 mm x 43,5 mm) realisiert, die am Greifer montiert werden kann (siehe **Bild 7**).

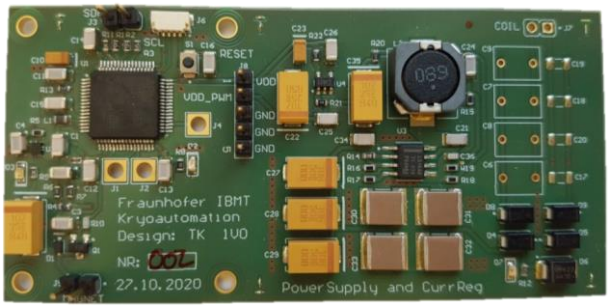


Bild 7 Foto der Platine für die Steuerelektronik

Die Abtastfrequenz des AD-Wandlers blieb über den gesamten Temperaturbereich stabil bei 98 Hz und änderte sich nur geringfügig auf 97 Hz bei -196 °C. Die Ausgangsspannung des Strommessverstärkers blieb als Regelgröße ebenfalls stabil. Die Regelung des Tastverhältnisses ergab eine Abnahme von 46 % bei Raumtemperatur auf 10 % im flüssigen Stickstoff. Dementsprechend ergab sich bei Betrieb des Greifers auch eine Reduzierung der Leistungsaufnahme von etwa 1,4 W bei Raumtemperatur auf etwa 0,3 W bei -196 °C. Das simulierte Verhalten des Regelkreises konnte bestätigt werden. Die Regelung erreichte bei tiefen Temperaturen aufgrund der kürzeren Anstiegs- und Einschwingzeit den Sollwert mit einem kurzen Überschwingen schneller. Außerdem erfolgte ein funktionaler Test im Dauerbetrieb bei -196 °C (siehe **Bild 8**). Der Dauerbetrieb wurde

nach knapp 180 Stunden beendet, in denen die Aktorik 461.785 Schaltzyklen erfolgreich ausführte.
Nach der Bestückung mehrerer Platinen zeigte sich, dass nicht alle Mikrocontroller des evaluierten Typs bis -196 °C funktionsfähig bleiben. Auch wenn ein Mikrocontroller bis -130 °C oder bis -180 °C funktionsfähig bleibt, bedeutet dies nicht, dass er auch im flüssigen Stickstoff funktionieren wird. Diese Ausfälle waren allerdings stets in reversibler Form. Der Grund dafür könnte eine Streuung der Prozessparameter in der Fertigung des Herstellers sein, die sich mit fallender Temperatur immer stärker bemerkbar macht. Für die Realisierung solcher Systeme bedeutet dies einen erhöhten Aufwand, da jeder Mikrocontroller vor der Bestückung getestet werden muss.

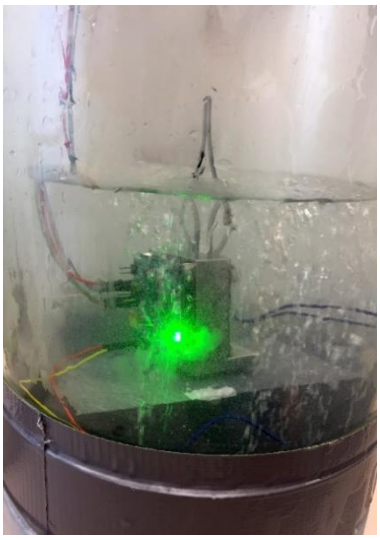


Bild 8 Funktionaler Test des Greifersystems im flüssigen Stickstoff

5 Zusammenfassung

Mit der in diesem Beitrag beschriebenen Entwicklung eines monolithischen Tieftemperatur-Greifersystems sind wir unserem mittelfristigen Ziel, ein robotergestütztes Handhabungssystem für Kryobanken in Form eines Parallelroboters zu realisieren, nähergekommen.
Das realisierte Funktionsmuster arbeitet energieeffizient bei tiefen Temperaturen. Der geöffnete Zustand ermöglicht das Festhalten von Kryoröhrchen ohne Bestromung des Aktors. Das Freisetzen erfordert eine kurzzeitige Bestromung. Die digitale PWM-Regelung hält die Bestromung des Aktors konstant, indem das Tastverhältnis in Abhängigkeit der Umgebungstemperatur angepasst wird. Das Tastverhältnis konnte bei -196 °C insgesamt um den Faktor 4,6 im Vergleich zu Raumtemperatur reduziert werden, d.h. der Energiebedarf des Aktors verringert sich kontinuierlich mit fallender Temperatur, wodurch sich sogar ein Vorteil für den Betrieb bei kryogenen Temperaturen ergibt. Ein initialer funktionaler Test des Greifersystems im Dauerbetrieb, eingetaucht im flüssigen Stickstoff, wurde erfolgreich durchgeführt. Weitere Dauertests sind geplant, um mögliche Langzeiteffekte, wie Parameterdrift oder Materialermüdung, zu untersuchen. Die Energieversorgung erfolgte bisher mit einem Labornetzteil. Im nächsten Schritt planen wir die kabellose Energieversorgung über eine induktive Übertragungsstrecke.
Die Ergebnisse dieses Beitrags liefern darüber hinaus wichtige Informationen über die Auswahl und das Verhalten von elektronischen Komponenten für die Realisierung von Schaltungen bei kryogenen Umgebungstemperaturen.

6 Danksagung

Die in diesem Beitrag vorgestellten Ergebnisse wurden im Rahmen des Projekts "Methoden zur Automatisierung von Handhabungsprozessen unter kryogenen Umgebungsbedingungen" (Projektnummer: 349906175) erzielt. Die Autoren danken der Deutschen Forschungsgemeinschaft (DFG) für die finanzielle Unterstützung dieses Projekts.

7 Literatur

[1] Singh R.; Baliga B.J.: *Cryogenic Operation of Silicon Power Devices*. Boston, MA, USA: Springer US; 1998.

[2] Pan M.-J.: *Performance of capacitors under DC bias at liquid nitrogen temperature*. Cryogenics 2005; 45(6):463–7.

[3] Bourne J.; Schupbach R.; Hollosi B.; Di J.; Lostetter A.; Mantooth H.A.: *Ultra-Wide Temperature (-230°C to 130°C) DC-Motor Drive with SiGe Asynchronous Controller*. In: IEEE Aerospace Conference 2008. 1./8. März 2008 Big Sky, MT, USA. S. 1–15.

[4] Ekin J.W.: *Experimental techniques for low-temperature measurements: Cryostat design, material properties, and superconductor critical-current testing*. Reprinted (with corr.). Oxford: Oxford Univ. Press; 2007.

[5] Takeda T.; Takano S.; Shindo Y.; Narita F.: *Deformation and progressive failure behavior of woven-fabric-reinforced glass/epoxy composite laminates under tensile loading at cryogenic temperatures*. Composites Science and Technology 2005; 65(11-12):1691–702.

[6] Clark W.F.; El-Kareh B.; Pires R.G.; Titcomb S.L.; Anderson R.L.: *Low temperature CMOS-a brief review*. IEEE Trans. Comp., Hybrids, Manufact. Technol. 1992; 15(3):397–404.

[7] Balestra F.; Ghibaudo G.: *Brief review of the MOS device physics for low temperature electronics*. Solid-State Electronics 1994; 37(12):1967–75.

Konstruktion eines Festkörpergelenkgreifers mit bistabiler Formgedächtnisaktorik

Dominik Scholtes [1], Stefan Seelecke [2], Paul Motzki [1,2]

[1]Zentrum für Mechatronik und Automatisierungstechnik (ZeMA gGmbH), Saarbrücken
[2]Fachbereich Systems Engineering, Fachbereich Materialwissenschaften und Werkstofftechnik, Universität des Saarlandes, Saarbrücken

Kurzfassung

In der industriellen Fertigung sind die meisten Bearbeitungsschritte mit dem Transportieren und Positionieren von Werkstücken verbunden. Die aktiven Schnittstellen zwischen Handhabungssystem und Werkstück sind industrielle Greifer, die vor allem im kleinteiligen Bereich oft pneumatisch angetrieben werden. Auf dem Weg zu höherer Energieeffizienz und digitalen Fabriken sind Unternehmen auf der Suche nach neuen Antriebstechnologien mit mehr Sensorintegration und besseren Wirkungsgraden.
Gängige Aktoren wie Magnete und Elektromotoren sind in vielen Fällen zu schwer und groß für eine direkte Integration in das Greifsystem. Aufgrund ihrer hohen Energiedichte sind Formgedächtnislegierungen (FGL) geeignet, diese Nachteile herkömmlicher Aktuatoren zu überwinden. Zusätzlich verfügen sie über sogenannte Self-Sensing Fähigkeiten, die zu einer sensorlosen Überwachung und Steuerung des Antriebssystems führen. Ein weiterer Nachteil konventioneller Greifer ist ihr Aufbau, der auf beweglichen Teilen, die besonders in abrasiven Umgebungen schnell verschleißen. Dies kann durch Festkörpergelenke, die konventionelle Gelenke ersetzen, vermieden werden. In der vorliegenden Arbeit wird der Entwicklungsprozess eines Funktionsprototyps für einen elastischen Greifer, der von einer bistabilen FGL-Antriebseinheit angetrieben wird, für industrielle Anwendungen umrissen. Der Schwerpunkt liegt auf der Entwicklung des FGL-Antriebs, während ein erster Designansatz für den nachgiebigen Greifmechanismus mit Festkörpergelenken vorgestellt wird. Das Ergebnis ist ein funktionierender Greifer-Prototyp, der hauptsächlich aus 3D-gedruckten Teilen besteht. Erste Ergebnisse von Validierungsversuchen werden diskutiert.

Abstract

Within industrial manufacturing most processing steps are accompanied by transporting and positioning of workpieces. The active interfaces between handling system and workpiece are industrial grippers, which often are driven by pneumatics, especially in small scale areas. On the way to higher energy efficiency and digital factories, companies are looking for new actuation technologies with more sensor integration and better efficiencies.
Commonly used actuators like solenoids and electric engines are in many cases too heavy and large for direct integration into the gripping system. Due to their high energy density shape memory alloys (SMA) are suited to overcome those drawbacks of conventional actuators. Additionally, they feature self-sensing abilities that lead to sensor-less monitoring and control of the actuation system. Another drawback of conventional grippers is their design, which is based on moving parts, are prone to wear, especially in abrasive environments. This can be overcome by flexure hinges that dispense with bearings and guides.
In the presented work, the development process of a functional prototype for a compliant gripper driven by a bistable SMA actuation unit for industrial applications is outlined. The focus lies on the development of the SMA actuator, while the first design approach for the compliant gripper mechanism with solid state joints is proposed. The result is a working gripper-prototype which is mainly made of 3D-printed parts. First results of validation experiments are discussed.

1 Motivation und Einleitung

Die globale Bewegung zur Senkung des Energieverbrauchs und Steigerung der Effizienz ist auch im industriellen Umfeld ein großes Thema. Parallel dazu hält mit der Entwicklung von Industrie 4.0 mehr Digitalisierung Einzug in alle großen Unternehmen. Dazu gehört der Bedarf an intelligenteren Maschinen mit mehr Sensorintegration und intelligenter Zustandsüberwachung.
Neben Magnetaktoren und Elektromotoren ist die Pneumatik in den meisten Unternehmen immer noch die verbreitetste Antriebstechnologie. Obwohl sie viele Vorteile hat, ist die Pneumatik dafür bekannt, dass sie vor allem aufgrund von Verlusten in den Rohrleitungen, Ventilen und Anschlüssen ineffizient ist. Besonders in sauberen Produktionen bringt die Druckluft zusätzlich Partikel und Öl in den Prozess. Die Überwachung und Steuerung von pneumatischen Systemen basiert auf externen Sensoren, die nur grundlegende Daten liefern. Besonders in Montagelinien wird eine große Anzahl dieser pneumatischen Greifer zum Positionieren, Handhaben und Transportieren von Werkstücken eingesetzt. Ziel der vorgestellten Arbeit ist es, einen funktionsfähigen Prototyp zu entwickeln, der das Potenzial aufweist, diese Greifer in Zukunft zu ersetzen.

Der entwickelte Greifer soll eine energieeffiziente, intelligente und leichte Alternative sein, die komplett auf pneumatische Systeme und Komponenten verzichtet, aber eine ähnliche Funktionalität wie heutige Standardgreifer aufweist.
Auf dem Weg zu neuen Lösungen, die höhere Effizienz als auch bessere Digitalisierung ermöglichen, können intelligente Materialien eine große Rolle spielen. Neben ihren aktorischen Eigenschaften können sie gleichzeitig als Sensor genutzt werden. Dies führt zu sensorlosen Aktuator Systemen, die leicht, kompakt und energieeffizient sind. Während Dielektrische Elastomere (DEs) für ihre hohe Energieeffizienz und schnelle Betätigungsfrequenzen bekannt sind [1], weisen Formgedächtnislegierungen (FGL) die höchste bekannte Energiedichte auf [2]. Da die Ansteuerung auf einem thermischen Effekt beruht, sind die Frequenz und der Wirkungsgrad einer FGL ans sich begrenzt. Daher wurden intelligente Steuerungs- und Designstrategien entwickelt, die diese Nachteile ausgleichen. Neuere Forschungen zeigen verschiedene Ansätze für energieeffiziente FGL-basierte Antriebssysteme [3], [4].
Darüber hinaus basiert das Design traditioneller industrieller Greifer auf einem Zusammenbau von beweglichen Teilen mit Linearführungen, Lagern und Scharnieren [5]. Der Bedarf an Präzisionsteilen in Verbindung mit dem hohen Montageaufwand macht herkömmliche Greifer teuer und komplex. Alle Führungen und Lager sind anfällig für Verschleiß, insbesondere in Prozessumgebungen mit abrasiven Medien. Bei elastischen Mechanismen (compliant mechanisms) hingegen, die bereits von Mikrogreifern bekannt sind, werden alle Gelenke und Führungen durch flexible Teile ersetzt. Die gesamte Greiferkinematik kann als ein einziges Teil gefertigt werden, was heutzutage mit einer Vielzahl von 3D-Druckverfahren realisiert werden kann. Durch das Fehlen von Toleranzen, die mechanischen Führungen und Lagern aufweisen, wird die Präzision und Wiederholgenauigkeit durch den Einsatz von Festkörpergelenken erhöht [6].

2 Grundlagen

FGL-Aktuatoren werden in den meisten Anwendungen in Form dünner Drähte verwendet. Sie lassen sich bei niedrigen Temperaturen leicht verformen und kehren bei Erwärmung über ihre Übergangstemperatur in ihre ursprüngliche Form zurück. Dieses Verhalten wird als Zwei-Weg-Formgedächtnis-Effekt (ZWE) bezeichnet und basiert auf einer Phasenumwandlung von Martensit (niedrige Temperatur) zu Austenit (hohe Temperatur) [7]. Die heute am besten untersuchte und am häufigsten verwendete Legierung ist Nickel-Titan (NiTi) mit einer Übergangstemperatur von ca. 90 °C und einem maximalen Hub von ca. 5 %. Es gibt bereits eine Reihe von Anwendungen für FGL-basierte Aktuatoren in der Automobilindustrie und in Elektronikartikeln. Beispiele sind Pneumatikventile und Verriegelungsmechanismen [8], [9]. In neueren Forschungskonzepten wurden FGL-getriebene Pick-and-Place Systeme, Roboteranwendungen und medizinische Geräte entwickelt [10]-[13]. In diesen Konzepten werden die Vorteile von FGL-Aktoren wie hohe Energiedichte, geringes Gewicht, geräuschlose Betätigung, hohe Kräfte und kurze Aktivierungszeiten gezeigt. Die Dynamik eines FGL-Aktors hängt von der Abkühlzeit des Drahtes ab, die direkt mit dem Verhältnis von Oberfläche zu Querschnittsfläche des Drahtes zusammenhängt. Aus diesem Grund führt die Bündelung vieler dünner Drähte zu einer ähnlichen Kraftabgabe wie ein äquivalenter Draht mit großem Durchmesser, kann aber Frequenzen von bis zu 35 Hz erreichen [14], [15]. Um energieeffiziente Systeme zu schaffen, die von FGL angetrieben werden, muss der Wärmeübergang vom FGL-Draht auf seine Umgebung minimiert werden. Dies kann z. B. auf zwei verschiedene Arten erreicht werden: mit einer schnellen Aktivierung und einem energiefreien Positionshalten [3], [4], [11]. Bei einer schnellen Aktivierung ist der Strom zum Aufheizen des Drahtes ein kurzer Impuls mit großer Amplitude, was zu einem adiabatischen Aufheizverhalten des Drahtes führt. Verschiedene Regelstrategien können dann angewendet werden, um eine Überhitzung des FGL zu verhindern. Ein energieloses Halten der Position kann mit einem bistabilen Mechanismus erreicht werden. Der Mechanismus führt des Weiteren zu einem Übersetzungsverhältnis des FGL-Hubs und hoher Haltekräfte am Abtrieb. Die Kombination eines bistabilen FGL-Aktors mit einer Hochgeschwindigkeitsansteuerung reduziert den Energieverbrauch des Systems auf ein Minimum.
In der industriellen Fertigung sind Greifer die aktive Schnittstelle zwischen Handhabungssystem und Werkstück. Ihre Funktionen können sein, eine bestimmte Position und Orientierung des Werkstücks temporär beizubehalten, aber auch zu verändern, Prozesskräfte und -momente zu halten und bestimmte technische Abläufe zu erfüllen [16]. Die für diese Arbeit relevantesten Greifertypen sind die so genannten Backengreifer. Das Greifen eines Objekts oder Werkstücks mit diesen Greifern kann in drei Gruppen eingeteilt werden: reiner Formschluss, partieller Formschluss kombiniert mit Kraftschluss und reiner Kraftschluss [16]. Die Backenbewegung selbst lässt sich im Wesentlichen in zwei Formen einteilen, die in Abbildung 1 dargestellt sind: winklige Backenbewegung a) und parallele Backenbewegung b). Beide können innen- und außengreifend sein.

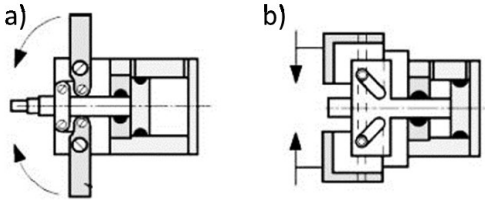


Abbildung 1: Schematische Darstellung der Greifbackenbeweg ung bei herkömmlichen Industriegreifern [16]

Das Antriebssystem, das die Greifbacken bewegt, ist in den meisten Fällen pneumatisch angetrieben. Auch elektrische Antriebe und Elektromagnete werden als Aktoren eingesetzt. Bislang konnten sich intelligente Materialien als Aktoren für Makrogreifer noch nicht

etablieren, obwohl auf diesem Gebiet geforscht wird. Im Bereich der Mikrogreifer gibt es mehrere Beispiele, die die Vorteile von piezoelektrischen Elementen und Formgedächtnislegierungen zeigen [17], [18].
Das Gleiche gilt für Kinematiken, die auf Festkörpergelenken basieren. Elastische Mechanismen sind in Mikrogreifern gut etabliert, da die vorhandenen Lager und Führungen zu groß sind zu großes Spiel aufweisen. Es gibt aber bereits Ansätze, Festkörpergelenk basierte Kinematiken in Industriegreifern zu etablieren [19]. Vorteile wie kleinere Toleranzen, bessere Präzision, weniger Einzelteile stehen Nachteilen wie elastischen Gegenkräften und höherem Entwicklungsaufwand gegenüber.
Die Kombination aus einer Antriebseinheit auf Basis von FGL-Drähten und einer nachgiebigen Greiferkinematik ist ein vielversprechender neuer Ansatz für intelligente und innovative Greifsysteme.

3 Entwicklung der elastischen Kinematik

Als erster Schritt in der Entwicklung des Funktionsprototyps wird ein Konzept der nachgiebigen Kinematik erstellt. Ziel ist es, eine erste funktionstüchtige Version zu erhalten, zu diesem Zeitpunkt ohne Bauraumbedarf. Der Grundaufbau des nachgiebigen Mechanismus ist in Abbildung 2 dargestellt. Sie besteht aus zwei trapezförmigen Gelenken (3) und zwei Blattfedergelenken (2), die im Wesentlichen als Schubstangen arbeiten und mit der Antriebseinheit verbunden sind. Der gesamte Prototyp wird mittels FDM-Druck hergestellt. Als Material wurde PETG gewählt, da es eine hohe Flexibilität und Zähigkeit bei gleichzeitig guten Druckergebnissen aufweist. Das beispielhafte Werkstück (1) hat einen Durchmesser von 15,9 mm, eine Dicke von 2,9 mm und eine Masse von 0,005 kg. Das Konzept zeigt eine winklige Backenbewegung und hat kein großes Übersetzungsverhältnis von Antrieb zu Abtrieb. Nachdem der Prototyp manuell getestet wurde und das gewünschte Hubverhalten zeigt, folgen mehrere Iterationsschleifen. Der Entwurf wird modifiziert, um die Federkräfte der Biegescharniere zu reduzieren, eine flache Betätigungseinheit und austauschbare Greifbacken zu ermöglichen. Das Ergebnis wird in Kapitel 5 beschrieben und dargestellt.

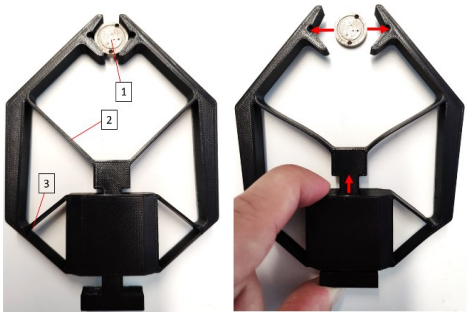


Abbildung 2: Erstest 3D gedrucktes Konzept der Greiferkinematik auf Basis von Festkörpergelenken

4 Entwicklung des bistabilen FGL-Aktuators

Das Patent von *Motzki und Seelecke*, auf dem der entwickelte FGL-Aktuator basiert, enthält eine Vielzahl von Grundkonstruktionen für einen bistabilen FGL Aktuator [20]. Um relativ kompakte Außenmaße für den Greiferantrieb zu erreichen, werden die FGL-Drähte parallel zur Blattfeder gesetzt, wie in Abbildung 3 dargestellt.

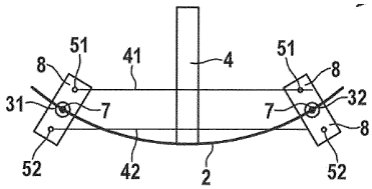


Abbildung 3: Skizze eines FGL Aktors mit einer bistabilen, drehbar gelagerten Blattfeder in parallal dazu angeordneten FGL Drähten [20]

Nummer 41 und 42 in der Abbildung sind die FGL-Drähte, 2 die Blattfeder und 8 die Hebelarme, an denen die FGL-Drähte in den Punkten 51 und 52 befestigt sind. Wesentliche Merkmale der Antriebsmechanik sind die Übertragung des FGL-Hubs auf den Antriebshub in 4 über die Länge der Hebelarme und vor allem die beiden stabilen energielosen Positionen des Systems. Die Ausgangskraft des Aktors wird durch die Parameter der Blattfeder bestimmt. Hier ist vor allem deren Dicke relevant. Obwohl die genaue Kraft für diesen ersten Funktionsprototyp nicht wichtig ist, muss sie stark genug sein, um die Gegenkraft der nachgiebigen Mechanik zu überwinden und das Werkstück in den Greifbacken zu sichern. Ein Greifzyklus (Backen schließen und wieder öffnen) sollte etwa 1 Sekunde dauern. Da diese Frequenz nur vom Drahtdurchmesser abhängig ist, liegt der maximal mögliche Drahtdurchmesser bei 100 µm, wenn keine zusätzlichen Kühlmedien berücksichtigt werden [21]. Um eine ausreichende Ausgangskraft zu erreichen, müssen mehrere NiTi-Drähte mechanisch parallel gebündelt werden.
Sechs 100 µm NiTi-Drähte werden mit 5 mm breiten Edelstahlblechen verbunden. Dies wird mit einem Widerstandsschweißprozess und einem kontrollierten Vorspannverfahren für jeden einzelnen Draht realisiert. Die hohe Verbindungsfestigkeit beim Widerstandsschweißen von dünnen NiTi-Drähten mit Edelstahl wurde von Scholtes et al. gezeigt [22]. Die Länge dieser Bündel hängt vom Hub ab, der benötigt wird, um die Blattfeder von einer Position in die andere schnappen zu lassen. Mit dem gewünschten Hub in der Mitte des bistabilen Elements und dessen Länge kann der Winkel in den Drehpunkten mit Hilfe einer Spline-Funktion im CAD modelliert werden. Der vorgestellte Aktor hat einen Hub der bistabilen Feder von 4 mm. Der Abstand zwischen den Drehpunktaufnahmen beträgt 39 mm. Bei einem Hebel von 1,5 mm und einem FGL-Hub von 3,5 % ergibt sich eine Drahtlänge von 35 mm. In Abbildung 4 ist die endgültige CAD-Baugruppe des Aktors dargestellt.

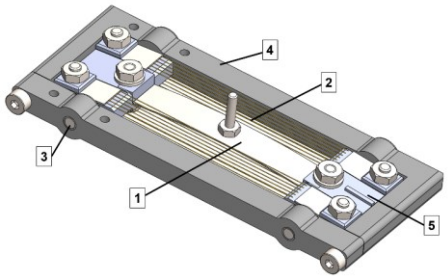


Abbildung 4: CAD-Modell des bistabilen FGL-aktors: 1: Blattfeder mit Gewindestange als Abtrieb; 2: FGL-Drahtbündel; 3. Passstifte als Welle; 4. Rahmen mit Buchsen; 5: Federklemme mit Befestigungspunkten für FGL-Aktoren

Alle Teile, mit Ausnahme der Blattfeder und der Drahtbündel, werden aufgrund der besseren Auflösung und Präzision mit einem SLA-Drucker hergestellt. Die Klemmen (5) sind auf Passstiften (3) montiert, die in Buchsen im Rahmen (4) gelagert sind. Die elektrische Verbindung der FGL-Drähte wird mit selbstklebender Kupferfolie realisiert, die auf den Klemmen angebracht ist. Die Bündel auf jeder Seite der Feder sind elektrisch in Reihe geschaltet, während sie vom gegenüberliegenden Bündelpaar isoliert sind.
Zur Versorgung und Ansteuerung des Aktors wird eine mikrocontrollerbasierte Elektronik entwickelt. Sie verarbeitet auch das Widerstandssignal. Als Eingangssignal dient ein 24-Volt-Signal, wie es z. B. von einer SPS-Steuerung erzeugt wird. Ein Hochspannungsimpuls versorgt die FGL-Drähte mit einer bestimmten Energiemenge. Die Einschaltdauer beträgt ca. 100 ms. Je nach Umgebungstemperatur werden 1400 - 1800 mJ an Energie benötigt, damit das bistabile Element in die entgegengesetzte Position schnappt. Die Widerstandsänderung der NiTi-Drähte wird gemessen und als Positionsrückmeldung des Aktors an die SPS-Steuerung zurückgesendet.

5 Zusammenbau des Greifers und erste Validierung

Zur Befestigung des Aktors an der Greiferkinematik ist ein Klemmsystem vorgesehen. Es ermöglicht die einfache Demontage des Greifers und erleichtert die Wartung und Feinabstimmung des Systems. In Abbildung 5 ist eine Explosionszeichnung der Greiferbaugruppe in CAD dargestellt. Die Befestigungsteile sind mit dem Greiferkörper mit integrierter Kinematik verschraubt und ficxieren die Aktorik.

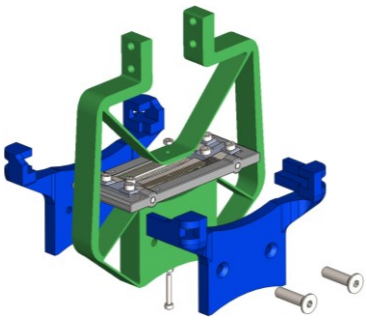


Abbildung 5: CAD-Modell des Greifer Zusammenbaus als Explosionsansicht: In grün die elastische Kinematik, blau die Befestigungsteile für die Aktorik

In Abbildung 6 ist der fertig montierte Greifer mit elektrischen Anschlüssen dargestellt. Die Greifbacken sind austauschbar, um zu verschiedene Werkstücken zu passen und auch den Abstand einzustellen. Der gesamte Greifer wiegt etwa 150 g.

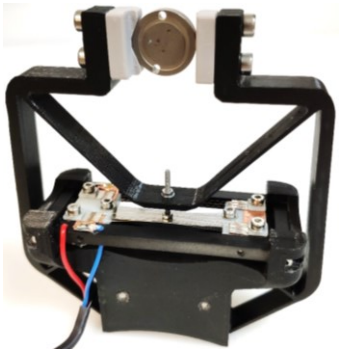


Abbildung 6: Foto des komplett zusammengebauten Festkörper-Greifers mit bistabilem FGL-Antrieb, der ein Beispiel-Werkstück greift

Für erste Validierungstests des Greifers wird ein Versuchsaufbau konzipiert. Ziel ist es, die Bewegung einer Greifbacke über die Zeit im Vergleich zum Eingangssignal zu messen. Die Positionsmessung erfolgt mit einem *Keyence LK-G37* Laser-Triangulationssensor, während der Greifer auf einem Breadboard befestigt ist. Die Datenerfassung erfolgt mit Hilfe von *NI LabVIEW*. Abbildung 7 zeigt ein beispielhaftes Ergebnis der Messungen. Das 24-V-Eingangssignal von der SPS-Steuerung, das einer logischen 1 entspricht, ist in rot und sichtbar.

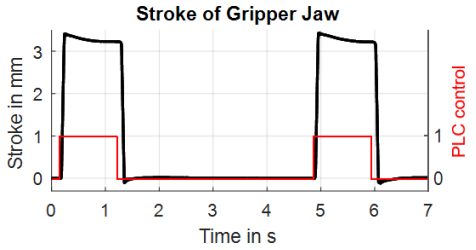


Abbildung 7: Grafische Darstellung der Messung von zwei Aktivierungszyklen des FGL-Greifers

Das Eingangssignal schaltet für 1,2 s auf 1 und dann wieder auf 0. Dieser Zyklus wird dann nach ca. 4 s manuell wiederholt. Die Bewegung der Greifbacke folgt dem Eingangssignal. Der maximale Hub beträgt 3,4 mm, stabilisiert sich aber bei 3,2 mm. Da nur eine Backe beobachtet wird und da sie sich parallel bewegen, wird ein Gesamthub der Backen von 6,4 mm erreicht. Das Überschwingen, das beim Öffnen und Schließen zu beobachten ist, ist auf einen FGL-Hub zurückzuführen, der größer als nötig ist und die bistabile Feder weiter als ihre eigene stabile Position biegt. Mit reduziertem Energieeintrag oder einer optimierten Steuerelektronik kann dieses Verhalten reduziert werden. Die Messungen zeigen, dass das Umschalten zwischen geöffnetem und geschlossenem Zustand weniger als 100 ms dauert. Eine Abkühlzeit für die Drahtbündel von 1,2 s ist ausreichend und kann durch eine weiter optimierte Ansteuerung des Ansteuerstroms sogar noch reduziert werden.

6 Zusammenfassung und Ausblick

Mittels relativ neuer Entwicklungen und Forschungsergebnisse auf dem Gebiet der FGL Aktorik wird ein fortgeschrittenes neues Greifkonzept für das industrielle Umfeld vorgestellt. Die Machbarkeit und Funktionalität eines FGL-angetriebenen bistabilen Greifers auf Basis einer elastischen Kinematik wird in der vorgestellten Arbeit aufgezeigt. Der Greifer erreicht eine Greiffrequenz von 0,8 Hz bei einem Backenöffnungshub von 6,4 mm. Die Schaltzeit selbst liegt unter 100 ms. Er kann ein kleines Werkstück sicher greifen und transportieren. Das System ist leicht, hat einen geringen Energieverbrauch, ist geräuschlos und arbeitet ohne zusätzliche Medien außer Elektrizität. Die Self-Sensing Eigenschaften der NiTi-Drähte ermöglichen eine Positionserkennung des Greifers sowie eine Zustandsüberwachung ohne zusätzliche externe Sensoren. Das Widerstandssignal der Aktordrähte wird in Zukunft auch für eine Regelung des Aktivierungsimpulses genutzt. Dazu wird der Schnapppunkt der Feder als charakteristisches Verhalten des Drahtwiderstandes erkannt. Im Moment der Erkennung wird die Stromzufuhr abgeschaltet. Dies führt zu einem noch weiter reduzierten Energieverbrauch, schnelleren Betätigungs- und Zykluszeiten sowie der Fähigkeit, sich an wechselnde Umgebungsbedingungen anzupassen. Ein erster Vergleich zu einem pneumatischen Standardgreifer verspricht eine Energieeinsparung von bis zu 80 %.

Danksagung

Die Autoren danken der Robert Bosch GmbH für die Finanzierung des Forschungsprojekts.

Literatur

[1] P. Linnebach, F. Simone, G. Rizzello, and S. Seelecke, “Development, manufacturing, and validation of a dielectric elastomer membrane actuator–driven contactor,” *J. Intell. Mater. Syst. Struct.*, vol. 30, no. 4, pp. 636–648, 2019.

[2] H. Janocha, T. Bonertz, and G. Pappert, *Unkonventionelle Aktoren : eine Einführung*. München: Oldenbourg Wissenschaftsverlag, 2013.

[3] P. Motzki, T. Gorges, M. Kappel, M. Schmidt, G. Rizzello, and S. Seelecke, “High-speed and high-efficiency shape memory alloy actuation,” *Smart Mater. Struct.*, vol. 27, no. 7, p. 075047, Jul. 2018.

[4] P. Motzki and S. Seelecke, “Bi-stable SMA Actuator,” *Actuator 16 - 15th Int. Conf. New Actuators*, pp. 317–320, 2016.

[5] “SCHUNK Grippers.” [Online]. Available: https://schunk.com/de_en/gripping-systems/category/gripping-systems/schunk-grippers/. [Accessed: 18-Apr-2020].

[6] M. N. Mohd Zubir and B. Shirinzadeh, “Development of a high precision flexure-based microgripper,” *Precis. Eng.*, vol. 33, no. 4, pp. 362–370, Oct. 2009.

[7] D. C. Lagoudas, *Shape Memory Alloys*, vol. 1. Boston, MA: Springer US, 2008.

[8] J. Mohd Jani, M. Leary, A. Subic, and M. A. Gibson, “A review of shape memory alloy research, applications and opportunities,” *Mater. Des.*, vol. 56, pp. 1078–1113, 2014.

[9] “FG-Aktoren (Kleinantriebe) - Kunststoffverarbeitung Hoffmann GmbH.” [Online]. Available: https://www.hoffmann-kunststoffe.de/aktoren/. [Accessed: 18-Apr-2020].

[10] P. Motzki, F. Khelfa, L. Zimmer, M. Schmidt, and S. Seelecke, “Design and validation of a reconfigurable robotic end-effector based on shape memory alloys,” *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 1, pp. 293–303, 2019.

[11] F. Welsch, M. Schmidt, S. M. Kirsch, S. Seelecke, and P. Motzki, “Vacuum gripper system based on bistable SMA actuation,” *ASME 2018 Conf. Smart Mater. Adapt. Struct. Intell. Syst. SMASIS 2018*, vol. 1, no. September 2019, 2018.

[12] Y. Goergen *et al.*, “Shape memory alloys in continuum and soft robotic applications,” *ASME 2019 Conf. Smart Mater. Adapt. Struct. Intell. Syst. SMASIS 2019*, no. December, 2019.

[13] P. Motzki and S. Seelecke, “Industrial Applications for Shape Memory Alloys,” in *Reference Module in Materials Science and Materials Engineering*, Elsevier, 2019.

[14] R. Britz *et al.*, “SMA Wire Bundles - Mechanical and Electrical Concepts,” *ACTUATOR 2018 16th Int. Conf. New Actuators*, no. June, 2018.

[15] S. H. Song, J. Y. Lee, H. Rodrigue, I. S. Choi, Y. J. Kang, and S. H. Ahn, “35 Hz shape memory alloy actuator with bending-twisting mode,” *Sci. Rep.*, vol. 6, no. February, pp. 1–13, 2016.

[16] G. J. Monkman, S. Hesse, R. Steinmann, and H. Schunk, “Robot Grippers,” *Robot Grippers*, pp. 1–452, 2007.

[17] M. Kohl, B. Krevet, and E. Just, “SMA microgripper system,” *Sensors Actuators A Phys.*, vol. 97–98, pp. 646–652, Apr. 2002.

[18] F. Glashuette, “FFZ Glashuette - Greiftechnik,” 2019. [Online]. Available: http://ffz-glashuette.com/greiftechnik/mikrogreifer-montage/. [Accessed: 22-Apr-2020].

[19] “Greifer | MONOLITIX.” [Online]. Available: http://www.monolitix.com/115-0-Greifer.html.

[20] P. Motzki and S. Seelecke, “Bistable Actuator Device Having a Shape Memory Element,” US2019/0203701, 2019.

[21] SAES Group, “SmartFlex Springs and Wires.” [Online]. Available: https://www.saesgetters.com/sites/default/files/SMARTFLEX SPRINGS AND WIRES.pdf. [Accessed: 06-Jul-2018].

[22] D. Scholtes, R.-K. Zäh, M. Schmidt, P. Motzki, B. Faupel, and S. Seelecke, “Resistance Welding of NiTi Actuator Wires,” in *ACTUATOR 2018: 16th International Conference on New Actuators*, 2018, no. June, pp. 500–504.

Aktor- und Sensorelemente basierend auf Dielektrischen Elastomeren zur Verwendung in tragbaren Assistenztools

S. Gratz-Kelly [1], B. Fasolt[2], S. Nalbach [1,2], S. Seelecke[1, 2], P. Motzki [1,2] und G. Rizzello [1]

1 Fachrichtung Systems Engineering, Universität des Saarlandes, Saarbrücken, Germany
2 Zentrum für Mechatronik und Automatisierungstechnik (ZeMA) gGmbH, Saarbrücken, Germany

Email: sebastian.gratz-kelly@imsl.uni-saarand.de, b.fasolt@zema.de, sophie.nalbach@imsl.uni-saarland.de, gianluca.rizzello@imsl.uni-saaland.de, p.motzki@zema.de, stefan.seelecke@imsl.uni-saarland.de

Kurzfassung

Aufgrund der stark fortschreitenden Mensch-Maschine-Interaktion in einer Industrie 4.0 Umgebung wird die flexible und sichere Handhabung von Werkstücken sowie die Rückmeldung des Handhabungsgerätes an den Werker, beispielsweise durch haptisches Feedback, immer wichtiger. Dielektrische Elastomere (DEs) eignen sich aufgrund ihrer hohen Integrationsdichte und der gleichzeitigen Nutzbarkeit als Aktor und Sensor besonders um die beschriebenen Aufgaben zu erfüllen. Insbesondere für die Verwendung in mobilen Anwendungen und für die Integration in Textilien, beispielsweise in Werkerkleidung, sind DEs aufgrund ihrer Flexibilität und Energieeffizienz prädestiniert.
Der vorliegende Beitrag präsentiert erste Ergebnisse der Entwicklung eines intelligenten Handschuhs mit integrierten DEs, zur Detektion der Handposition und Gesten des Nutzers sowie der haptischen Rückmeldung an diesen.

Abstract

Due to the high demand for human-machine-interaction, especially in the industry 4.0 environment, a flexible and safe interaction between the handling device and the worker, for example by haptic feedback, is very important. Dielectric Elastomers (DEs) are known for their use as actuator and sensor elements at the same time. These properties together with a high energy efficiency as well as flexibility make them perfectly suitable for mobile application, in particular for integration into textiles. This paper shows first results of the development of an intelligent glove prototype. DE elements are used for gesture recognition of and haptic feedback to the worker.

1 Einleitung

Durch die immer stärkere Vernetzung von Mensch und Maschine in der Produktion wird die Interaktion und der Informationsaustausch des Menschen mit der digitalen Komponente immer wichtiger. Um eine möglichst einfache und intuitive Interaktion zu gewährleisten wird die flexible und sichere Handhabung von Werkstücken durch den Werker sowie die intuitive Befehlseingabe an das Handhabungsgerät und die Rückmeldung des Handhabungsgerätes an den Werker, beispielsweise durch haptisches Feedback, immer wichtiger. In dem vorliegenden Beitrag wird ein Konzept eines intelligenten Handschuhs, mit integrierten Sensoren und Aktuatoren auf Basis von intelligenten Materialien, dargestellt, mit dem genau diese Interaktion umsetzt wird.
Dielektrische Elastomere (DEs) eignen sich aufgrund der gleichzeitigen Nutzbarkeit als Aktor und Sensor, sowie der flexiblen Anpassung der Geometrie, besonders zur Verwendung als integrierte Aktor-Sensor-Elemente. Wegen des geringen Gewichts und der hohen Energieeffizienz sowie der variablen Ansteuerbarkeit, sind DEs zudem für die Verwendung in Textilien, beispielsweise in Handschuhen, prädestiniert.
Durch die Einbettung von DE-Elementen an verschiedenen Gelenkstellen sowie an den Handflächen und der Handrückseite eines Handschuhs soll zum einen die genaue Handposition ermittelt werden können, wodurch beispielsweise Gesten erkannt werden können, und zum anderen soll ein entsprechend differenziertes haptisches Feedback gegeben werden. Als haptisches Feedback sind unterschiedlichste Reize, wie Vibrationen mit verschiedenen Frequenzen oder Kraftpulse, möglich. Zusätzlich kann durch die entsprechende Anpassung der DEs weitere physikalische Größen wie insbesondere Kraft (bspw. Greifkraft) aber auch Temperatur und Feuchtigkeit gemessen werden. Diese breite Einsetzbarkeit der DE-Elemente als vielseitiger Sensor und Aktor erlauben die zweckmäßige Verwendung von DEs in tragbaren Assistenztools. Durch die Entwicklung von zusätzlichen perzeptiven Elementen, wie textilintegrierten Knöpfen oder eines Mini-Joysticks, ergeben sich für den Benutzer eine Vielzahl an Interaktionsmöglichkeiten mit der zu bedienenden Maschine.
Zudem kann durch die Einbettung in die Industrie 4.0 Umgebung ein lernfähiges Assistenztool entwickelt werden, welches beispielsweise die gewünschte Bewegung des Werkers erkennt und eine entsprechende Krafteinwirkung durch das Handhabungsgerät auf das Werkstück impliziert.
In dem präsentierten Paper wird sich näher mit einzelnen speziellen Anwendungen der DE-Sensoren und -Aktoren im Kontext von intelligenten Assistenztools beschäftig, um das Assistenztool später durch weitere Elemente zu erweitern. Somit ist es möglich ein modulares System zu

entwickeln, welches auf anwendungsspezifische Anforderungen angepasst werden kann.

2 Grundlagen DEs

Im folgenden Abschnitt werden kurz die Grundlagen und die Funktionsweise von dielektrischen Elastomeren sowie die Konzepte zur Entwicklung eines intelligenten Assistenztools erläutert.

DEs sind eine Untergruppe der elektroaktiven Polymere, welchen gemein ist, dass sie aufgrund einer elektrischen Spannung eine Gestaltänderung erfahren oder eine mechanische Verformung in eine elektrisch messbares Signal wandeln [1].

DEs basieren auf einem Elastomerfilm (typischerweise 10-100 µm Dicke), auf welchem beidseitig eine sehr dünne Elektrode (wenige µm) aufgebracht ist [2]. Hierdurch entsteht ein flexibler Kondensator, der sowohl als Aktor als auch als Sensor genutzt werden kann [3]. Beim Aktoreffekt wird die Gestaltänderung durch elektrostatische Kräfte hervorgerufen. Dies geschieht durch das Anlegen einer elektrischen Spannung (kV-Bereich) an den beiden nachgiebigen Elektroden, woraufhin sich diese, aufgrund der aufgebrachten gegenpaarigen Ladungen, anziehen (vgl. Bild 1) und so einen elektrostatischen Druck (Maxwelldruck) auf den dielektrische Film (Dielektrikum) ausüben. Dieser wird hierdurch in Dickenrichtung komprimiert, was aufgrund der Inkompressibilität des verwendeten Elastomers zusätzlich zu einer Flächenausdehnung führt .

Umgekehrt ändern sich, infolge einer wirkenden äußeren Kraft, die Geometrie des DEs und folglich auch dessen Kapazität sowie dessen Elektrodenwiderstand. Durch eine Kapazitätsmessung an dem DE-Element kann, je nach mechanischem Aufbau, die Kraft, welche senkrecht auf das DE-Element wirkt, oder die Dehnung des Elastomers in Längsrichtung gemessen werden. Beides führt zu einer messbaren und mit der Verformung korrelierbaren Kapazitätsänderung (siehe Bild 1).

DEs lassen somit die energieeffiziente Umwandlung von elektrischer in mechanische Energie und umgekehrt zu [4].

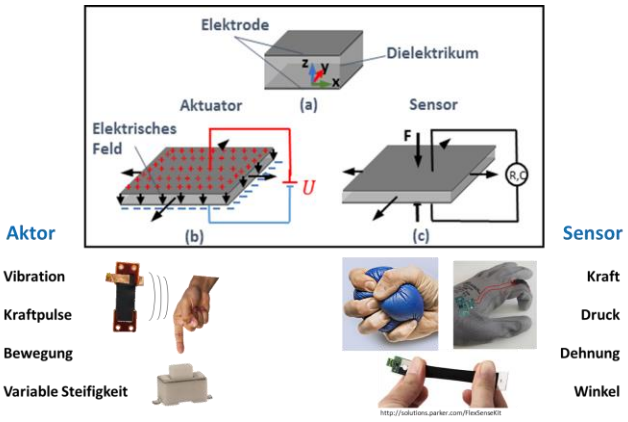


Bild 1 *Funktionsprinzip* von DEs: *(a) DE in unbelasteter Konfiguration; Geometrieänderung (b) nach Anlegen einer elektrischen Spannung (c) nach Verformung* durch externe Kräfte und Anwendungsmöglichkeiten

DEs können auf Grund ihrer flexiblen Geometrie weitreichend eingesetzt werden. So ist der Einsatz in Industrieventilen sowohl aktorisch als auch sensorisch möglich, die Verwendung als Linearaktor und beispielsweiße als Pumpaktor sowie die Textilintegrierbarkeit und die mechanische Flexibilität ergeben ein weites Eisatzgebiet für DEs. Speziell die gute Integrierbarkeit in Textilien, die einfach anpassbaren Geometrie, ihrer inhärenten Flexibilität / Dehnbarkeit (>100%) und des einfachen Aufbaus sowie die variable Ansteuerbarkeit der DEs sind für den Aufbau integrierte Aktor-Sensor-Elemente ausschlaggebend. Insbesondere die somit gleichzeitige Verwendbarkeit der DEs als Aktor und Sensor und der sehr schmale Aufbau prädestinieren diese zur Verwendung für den vorliegenden Anwendungsfall.

2.1 Sensorik

Für die Hand- bzw. Fingerhaltungsmessung und Greifkraftmessungen wird die sensorische Eigenschaft von DEs ausgenutzt. Dieses basiert wie bereits erwähnt auf der Kapazitätsänderung der DEs bei Dehnung. Die Kapazität eines DEs lässt sich mit der Formel für einen Plattenkondensator beschreiben:

$$C = \varepsilon_0 \varepsilon_r \cdot \frac{A}{d} = \varepsilon_0 \varepsilon_r \cdot \frac{lb}{d} \tag{1}$$

Hierbei ist ε_0 die elektrische Feldkonstante, ε_r die relative Permittivität des Dielektrikums, A die Elektrodenfläche, l die Länge und b die Breite des DEs sowie d die Dicke der Elastomerschicht. In Gleichung (1) ist zu erkennen, dass die Kapazität des DEs linear mit der Elektrodenfläche ansteigt und sich inversproportional zur Dicke des Dielektrikums verhält. Aufgrund der Inkompressibilität des Dielektrikums ergibt sich ein linearer Zusammenhang zwischen Ausgangsgeometrie und verformter Geometrie. Damit sind Flächenänderung und Dickenänderung direkt über die Ausgangsgeometrie miteinander verknüpft:

$$A(z) = \frac{A_0 \cdot d_0}{d(z)} \tag{2}$$

Mit dem Einsetzen von Gleichung (2) in Gleichung (1) folgt die dickenabhängige Kapazität zu

$$C(z) = \varepsilon_0 \varepsilon_r \cdot \frac{A_0 \cdot d_0}{d(z)^2}$$
$$\Rightarrow d(z) = \sqrt{\varepsilon_0 \varepsilon_r \cdot \frac{A_0 \cdot d_0}{C(z)}} \tag{3}$$

Mit gleichem Vorgehen ergibt sich die längenabhängige Kapazität des DE-Elementes zu

$$C(x) = \varepsilon_0 \varepsilon_r \frac{A_0 \cdot l(x)^2}{{l_0}^2 \cdot d_0}$$
$$\Rightarrow l(x) = \sqrt{\frac{C(x) \cdot {l_0}^2 \cdot d_0}{A_0 \varepsilon_0 \varepsilon_r}} \tag{4}$$

Aus Gleichung (3) kann somit eine Funktion für eine Kraftmessung (bspw. Greifkraft) aus der Dickenänderung und mit Gleichung (4) eine Funktion für eine Längenmessung (bspw. Handstellung) hergeleitet werden. Wobei zu

beachten ist, dass sich beide Effekte auch überlagern können und daher eine angepasste Auswerteelektronik nötig ist.
Zur Messung der Handstellung und Handposition ist es notwendig die Sensorelemente an geeigneten Stellen an der Hand zu platzieren. Insgesamt besitz die Hand 29 Freiheitsgrade, wobei hierbei zwei Freiheitsgrade des Handgelenkes mit einberechnet und die Räumliche Ausrichtung der Hand mit 6 Freiheitsgraden als beliebig angenommen sind.
Je nach angestrebter Komplexität des Handschuhs werden folglich nur die für eine Anwendung wichtigen Freiheitsgrade erfasst.

2.2 Aktorik

Neben der sensorischen Verwendung der DE-Elemente können bei exakt gleichem Aufbau dieselben Elemente auch als Aktor verwendet werden. Grundlage des Aktorprinzips ist der bereits erwähnte Maxwelldruck, der beim Anlegen einer elektrischen Spannung entsteht. Die anzulegende Spannung liegt, um erkennbare mechanische Auslenkungen zu erzeugen, typischerweise im Bereich zwischen 50 und 80 V/µm [5]. Die Erzeugung und Abschirmung der hohen Spannungen stellt spezielle Anforderungen an die Ansteuerelektronik und das Anschlusskonzept des DE-Elements, sowie insbesondere an die Isolation zur Hand. Aufgrund der sehr geringen Ströme, welche sich im Normalbetrieb im µA-Bereich befinden, ist die Leistungsaufnahme und entsprechend die Energiemenge (<30 mJ) der DEs jedoch sehr gering. Für die Platzierung der Aktorelemente gibt es, entsprechend zu der Platzierung der Sensorelemente, bevorzugte Bereiche auf der Hand. Hierzu sind in Bild 2 die rezeptiven Felder der Handinnenfläche dargestellt. Insbesondere für Vibrationen im Frequenzbereich von 50 – 300 Hz, sind die Parcini-Körperchen besonders empfindlich, weshalb an den entsprechenden Stellen der Haut eine Platzierung von Aktorelementen sinnvoll ist.

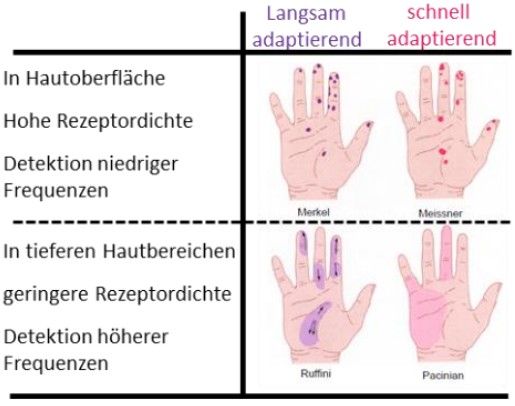


Bild 2 Rezeptive Felder der Handinnenfläche

Neben der speziellen Platzierung von Aktorelementen in der Handinnenfläche, können an der Handoberseite die DE-Elemente, welche zur Handpositionsmessung eingesetzt werden, ebenfalls als Aktor Verwendung finden und ein haptisches Feedback erzeugen. Der Einsatz von Aktorelementen an der Handrückseite ist zudem von Vorteil um dem Benutzer die Arbeitsfähigkeit an der Handinnenfläche zu gewährleisten. Neben der aktuatorischen Verwendung der DEs um ein haptisches Feedback zu erzeugen können auch andere Reize, beispielsweise akustischer Art, erzeugt werden. Aufgrund der flexiblen Ansteuerbarkeit der DE-Elemente ist es zudem möglich eine Superposition von niederfrequenten (haptisch) und höherfrequenten (akustisch) Signalen zu generieren.

2.3 Auslegung

Die Auslegung der Sensorelemente ist prinzipiell nach Gleichung (3) einfach möglich. Die Grundkapazität lässt sich über die geometrischen Größen des DEs bestimmen und je nach Art der Verformung kann die Kapazitätsänderung über die Geometrieänderung berechnet werden. In Bild 3 (a) ist beispielhaft die Kapazitätsänderung eines Streifensensors über den Gelenkwinkel dargestellt. Um DE-Elemente zudem zur Kraftmessung nutzen zu können sind weitere Anpassungen der Elemente notwendig (siehe [7]). In Bild 3 (b) sind für 2 mögliche Variationen der DE Geometrie zu Kraftmessung dargestellt. Zudem sind in den Graphen die jeweiligen Kapazitätsverläufe über der Dehnung der DE Elemente und über der bei der Auslenkung gemessenen Kraft aufgeführt.
Die Auslegung der Aktorik bedarf zunächst einer Kraftmessung der DE-Elemente mit angelegtem elektrischem Feld und ohne angelegtes Feld. Um eine Bewegung des DE Elements erzeugen zu können muss ein angepasster Vorspannmechanismus entwickelt werden. Der Vorspannmechanismus wird so ausgelegt, dass mit dem Kraftdifferential zwischen geringstem und maximalem Feld, ein möglichst hoher Hub des Aktors bzw. eine möglichst hohe Aktorkraft genutzt werden kann. Als Vorspannmechanismus können verschiedene Systeme mit einem entsprechenden Kraft-Weg-Verhalten genutzt werden. Mögliche Vorspannmechanismen sind beispielsweise lineare Federn, Magnete, nichtlineare Federn, oder eine zweite DE-Membran. Die speziellen Vorgehensweisen bei der entsprechenden Wahl des Vorspannmechanismus sind in ([8], [9]) dargestellt.

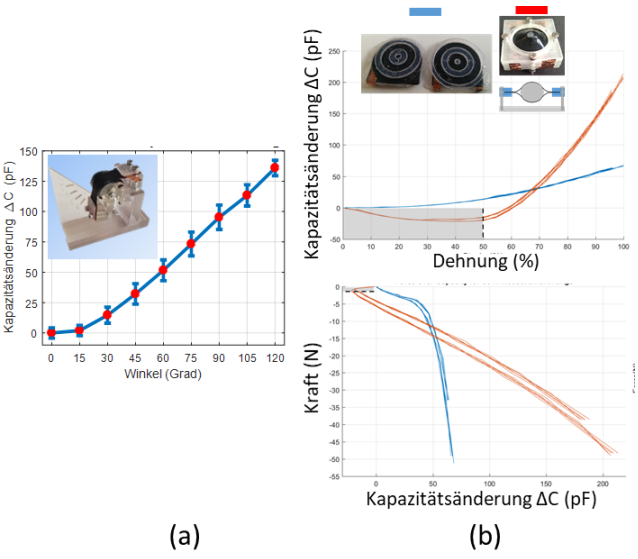


Bild 3 *Kapazitätsverlauf eines DE-Elements (a) zur Winkelmessung und* (b) zur Kraftmessung [7]

3 Prototyp Design

Der Aufbau des Prototypen soll zunächst die einzelnen Unterfunktionen des Gesamtsystems in einer etwas vereinfachten Form abbilden. Der Aufbau des späteren Handschuhs wird im Folgenden kurz aufgezeigt. Ein Teil der Unterfunktionen, die mit einem ersten Prototyp realisiert werden, sind im nächsten Abschnitt aufgeführt.

3.1 Position und Geometrie der DE-Elemente

Die Unterfunktionen bedürfen gesonderter Anpassungen der DE-Elemente. Zum einen werden zur Gesten- und Positionserkennung Sensorelemente in Streifenform verwendet, die direkt über den jeweilig zu sensierenden Gelenken positioniert werden. Zur Greif- und Drückkraftmessung werden zudem gesonderte runde Elemente mit angepasstem Vorspannmechanismus entwickelt und an den Fingerspitzen und der Handinnenfläche integriert. Haptisches Feedback wird ebenfalls über speziell angepasst Elemente eingebracht. Hierzu werden verschiedene Konzepte aufgezeigt und verglichen. In Bild 4 ist entsprechend der konzeptuelle Aufbau eines Prototypen dargestellt. Hierbei sind pro Finger jeweils ein DE-Sensorelement angedacht, sowie ein Sensorik-Haptik-Element an den Fingerspitzen zur Kraftmessung und um haptisches Feedback zu ermöglichen. Zusätzlich wird am Handrücken und am Handgelenk jeweils ein haptisch-akustisches Element eingebunden.

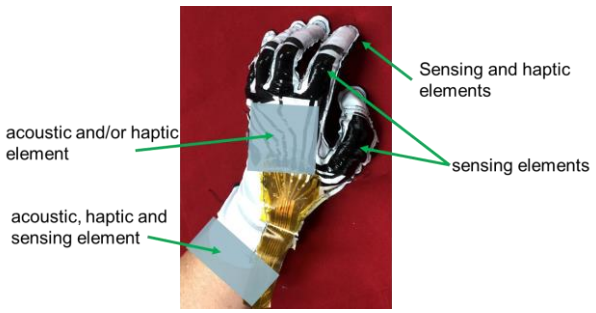


Bild 4 Konzept eines Protoyps mit Kombination verschiedener Einzelelemente

Zudem werden weitere Elemente, wie beispielsweise Knopf- und Eingabeelemente oder akustisches und visuelles Feedback, in einzelnen gesonderten Mustern aufgebaut und getestet.

3.2 Elektronik

Zur Ansteuerung und zum Auslesen der DE-Elemente sind jeweils gesonderte Elektronikschaltungen notwendig. Folgend werden kurz verschiedene Topologien für Kapazitätsmess- und Hochspannungsschaltungen aufgeführt und eine Auswahl für die ersten Prototypen getroffen.
Um die Dehnung des DEs zu bestimmen wird wie aus Gleichung (3) ersichtlich dessen Kapazität gemessen. Zur Verwendung in der vorliegenden Anwendung ist es notwendig eine kompakte und günstige Elektronik zu verwenden. Als Grundprinzipien der Kapazitätsmessung lassen sich hauptsächlich drei Topologiearten verwenden. Zum einen die Lade bzw. Entladezeitmessung der Kapazität, die Phasenverschiebung und Amplitudenänderung bei Anlegen einer Wechselspannung und die Messung der Schwingfrequenz eines mit der Kapazität erzeugten Schwingkreises. Die drei Methoden haben jeweils Unterschiede in Komplexität, Bauteilaufwand, Kosten, Genauigkeit und Auflösung. Die zunächst am einfachsten zu realisierende Topologie mit kleinstem Bauteilaufwand ist die Messung der Schwingkreis Frequenz. Diese Lösung wird im vorliegenden Fall verwendet und mit einer monostabilen Kippstufe NE555, welche in vielen Industrieanwendungen bereits zum Einsatz kommt, realisiert.
Für die aktorische Ansteuerung der DEs sind hohe Feldstärken bzw. hohe Spannungen notwendig. Folglich ist es auch hier notwendig eine kompakte und preiswerte Lösung zu realisieren. Zur Hochspannungserzeugung stehen verschiedene Topologien mit unterschiedlichen Eigenschaften, wie beispielsweiße Greinacher Schaltung (Spannungsverdoppler), StepUp-Wandler, Flyback Converter oder Resonanz Wandler, zur Verfügung. Auch hierbei wird die Verwendung der Topologie mit geringstem und günstigstem Bauteilaufwand, ein StepUp-Wandler mit nachgeschalteter Greinacher Schaltung, gewählt.

3.3 Textilintegration

Eines der wichtigsten Eigenschaften sowohl für die Elektronik als auch für die DE-Elemente selbst ist die Integrierbarkeit in Textilien. Hierbei sind insbesondere die mechanische Flexibilität und eine geringe Dicke wichtig. Die DEs sind mit einer Dicke von 20 – 50 µm prädestiniert für die Textilintegration. Ein wichtiger Aspekt um die nutzerfreundliche Tragbarkeit zu gewährleisten ist die Kontaktierungstechnik. Da die DE Elemente selbst relativ einfach in Textilien eigebracht werden können liegt in der Integration der Verbindungselemente, zwischen DE und Elektronik, ein spezielles Augenmerk bei dem Aufbau der Prototypen. Insbesondere die Flexibilität der Verbindungselemente, ein möglichst geringer Übergangswiderstand, ein schmaler Aufbau und die Notwendigkeit auch Hochspanungssignale übertragen zu können, stellen die wichtigsten zu erfüllenden Randbedingungen dar.

3.4 Integration in Industrie 4.0

Da das spätere Assistenztool insbesondere in digitalisierten Industrieumgebungen eingesetzt wird, ist eine Einbettung in die Kommunikationsstruktur der Umgebung notwendig. Zum einen muss, je nach Anwendungsfall, eine Verbindung zu einem übergeordneten Datenverarbeitungssystem und des Weiteren zu Maschinen, Robotern und weiteren Assistenztools aufgebaut werden. Zudem können dem Werker situationsspezifisch Hilfestellungen gegeben werden und seine Produktionsschritte überprüft und dokumentiert werden. In Bild 5 ist hierzu ein prinzipielles Schema der Kommunikation in der Produktionsumgebung skizziert. Durch die Verwendung des intelligenten Handschuhes werden dem Bediener zudem ergonomische Vorteile beim Arbeiten mit schweren Lasten, durch Kraft und Gestenerkennung sowie durch die Verwendung weiterer integrierter Bedienelemente, geboten und eine flexible und intuitive Bedienbarkeit erzielt. Weitere Ergonomieverbesserungen sind beispielsweise durch die Verwendung der Handpositions- und Kraftdaten zur Impedanz-/Admittanzregelung eines Handsteuergerätes möglich.



Bild 5 Einbettung des Assistenztools in dieIndustrie 4.0 Umgebung

4 Prototypaufbau und Messungen

Im Folgenden wird zum einen das Sensorprinzip für die Gestenerkennung an einem vereinfachten Aufbau dargestellt und zum anderen verschiedene Aktorprinzipien verglichen. Ergebnisse für realisierte Kraftsensoren sind in [7] veröffentlicht, es wurden ebenfalls verschiedene Konzepte aufgeführt und eine Auswahl verglichen.

4.1 Sensorik

Um das Prinzip der Gestenerkennung zu validieren wurde ein Prototyp mit 5 DE-Elementen hergestellt und vermessen. Die Verbindungstechnik wurde mit auf Kapton gedruckten Leiterbahnen, welche auf die DE-Kontaktstellen aufgeklebt werden, realisiert. In Bild 6 (a) ist der Aufbau eines Prototypenmessprüfstands dargestellt. Die 5 DE-Sensoren (einer pro Finger) können jeweils einzeln ausgelenkt werden. Mit einer Multiplexerschaltung werden die 5 Sensoren nacheinander mithilfe der Messelektronik (LCR Meter) gemessen (Bild 6 (b)).

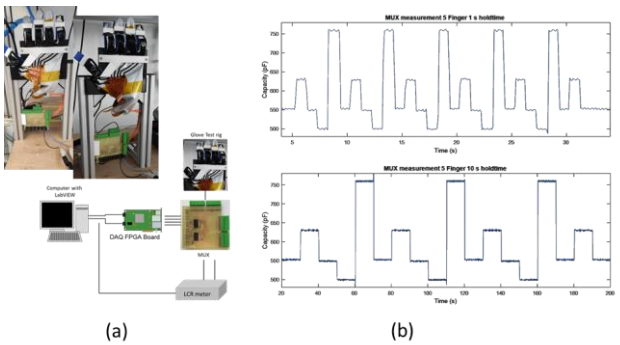


Bild 6 Sensorauslesung: (a) Prototypen Prüfstand und schematischer Messaufbau sowie (b) Ausgangssignale des gemuxten Kapazitätssignals

Die Stufen der Messung zeigen die Grundkapazitäten der einzelnen Sensoren. In Bild 7 sind die Änderungen der einzelnen Sensoren jeweils dargestellt. Von links oben nach rechts unten wurde hier der Daumen, der Zeigefinger, der Mittelfinger, der Ringfinger und der kleine Finger bewegt. In den Messungen sind teilweise leichte Übersprecher auf die anderen Sensorwerte zu erkennen, eine Zuordnung des gedehnten Sensors ist aber eindeutig möglich. Die Kapazitätsänderung bei einer maximalen Dehnung der DE-Elemente liegt je nach Finger zwischen 15 und 50 pF. Mit einer entsprechend ausgelegten Messelektronik ist dies Kapazitätsdifferenz in einer hohen Auflösung möglich.

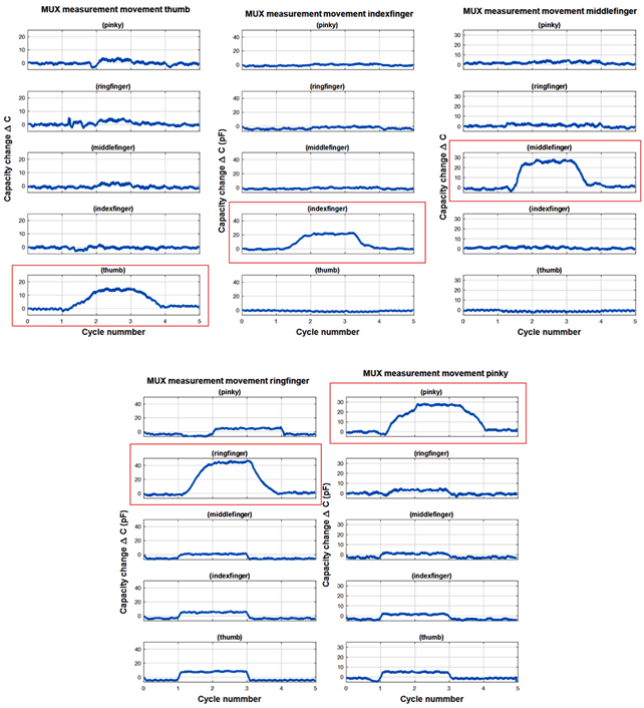


Bild 7 Kapazitätsverlauf eines DE-Elements (a) zur Winkelmessung (b) zur Kraftmessung

4.1 Aktorik

Neben der Sensorik ist das haptische Feedback des Assistenztools ein essentielles Feature. Durch die Verwendung von dielektrischen Elastomeren ist ein schmaler Aufbau der Aktorik mit einem hohen Signalumfang (Ansteuerung von niedrigen bis hohen Frequenzen, unterschiedliche Intensitäten, beliebige Signalformen) möglich. Zudem ist eine gute Integrierbarkeit in Textilien gegeben. Um die Krafterzeugung der DEs in einem haptischen Element nutzen zu können sind verschiedene Realisierungen möglich. Hierzu sind in Bild 8 verschieden Konzepte zur Erzeugung der haptischen Stimuli verglichen.

Geometrie / Messprinzip	Nr.	Beschreibung	Intensität	Komplexität	Flachbau
Haut	1	DE Vorgespannt (orth. Kraftwirkung)	●	●	●
NBS Haut	2	DE mit NBS (orth. Kraftwirkung)	●	●	●
PET / Silicone Dome Haut	3	PET / Silikon Dom als Vorspannmech. (orth. Kraftwirkung)	●	●	●
NBS Haut Haut	4	NBS und DE in Balkenrichtung (tang. Kraftwirkung)	●	●	●
Haut	5	Siehe 4 aber NBS in Ebene (tang. Kraftwirkung)	●	●	●
Haut	6	2 NBs in Ebene (tang. Kraftwirkung)	●	●	●
Haut	7	Strukturierter DE (orth. Kraftwirkung)	●	●	●
Haut Form	8	DE mit vorgeformter Silikonschicht (orth. Kraftwirkung)	●	●	●

Bild 8 Vergleich einer Auswahl verschiedener Konzepte zur Erzeugung haptischer Stimuli

Aus den vorgestellten Konzepten wurde eine Auswahl getroffen und für diese eine Blockierkraftmessung durchgeführt. Die drei zunächst ausgewählten Konzepte mit den entsprechenden Messungen sind in Bild 9 dargestellt. Wie zu erkennen liegen die resultierenden Kräfte in der gleichen Größenordnung. Für alle Aufbauten wurde ein Sinusförmiges Eingangssignal mit 2.8 kV und unterschiedlichen Frequenzen (1-180 Hz) angelegt.
Da die im späteren Aktorelement je nach Aufbau aufgebrachte Kraft jedoch in unterschiedliche Richtung wirken kann (senkrecht oder lateral), ist eine Aussage über die Stärke der Stimuli anhand der Messungen nur grob möglich. Dies wird anhand von Messungen mit Versuchspersonen noch genauer Untersucht, um eine qualitative Einschätzung der Intensität der unterschiedlichen Stimuli vergleichen zu können.

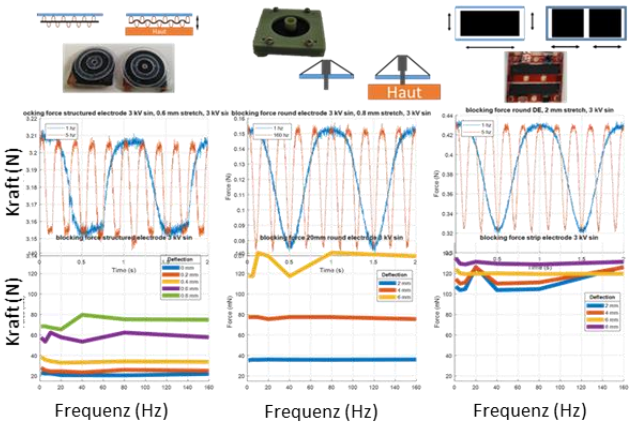


Bild 9 Blockierkraftmessung für verschiedene Aktorkonzepte

5 Zusammenfassung

In dem vorliegenden Beitrag wurde ein Prototyp für ein Assistenztool für eine Industrie 4.0 Umgebung vorgestellt. Hierbei wurden die Eigenschaften intelligenter Materialien (dielektrische Elastomere), welche sowohl als Sensor als auch als Aktor betrieben werden können, genutzt. Insbesondere die schlanke Integration der DE-Elemente in Textilien wurde gezeigt. Die Vermessung der Prototypen zeigt, dass eine Funktionsfähigkeit des Gesamtsystems gegeben ist. Zur Weiterentwicklung des Gesamtsystems bedarf es allerdings noch weiterer Verbesserungen, insbesondere in Elektronik (Auslese- und Ansteuerelektronik), Integration aller Features, Geometrieanpassung der Elemente und der Verbindungstechnik.

6 Danksagung

Ein besonderer Dank gilt der W. L. Gore & Associates GmbH für die Unterstützung im Entwicklungsprozess, sowie der Ausstattung mit Material.

7 Literatur

[1] R. Shankar, T. K. Ghosh und R. J. Spontak, „Dielectric elastomers as next-generation polymeric actuators.," *Soft Matter B. 3, The RoyalSociety of Chemistry,* pp. Nr. 9, S. 1116, 2007.

[2] A. York, J. Dunn und S. Seelecke, „Experimental characterization of the hysteretic and rate-dependent electromechanical behavior of dielectric electro-active polymer actuators.," *Smart Materials and Structures Bd. 19,* pp. Nr. 9, S. 094014, IOP Publishing (2010).

[3] G. Rizzello, D. Naso, A. York und S. Seelecke, „Closed loop control of dielectric elastomer actuators based on self-sensig displacement feedback.," *Smart Materials and Structures Bd. 25,* 2016.

[4] J. Biggs, K. Danielmeier, J. Hitzbleck und J. Krause, „Elektroaktive Polymere: Entwicklungen und Perspektiven dielektrischer Elastomere.," *Angewndte Chemie Bd. 125,* 2013.

[5] F. Carpi, D. Rossi, R. Kornbluh und R. Pelrine, „Dielectric elastomers as electromechanical transducers," in *Fundamentals, materials, devices, models and applications of an emerging electroactive polymer*, 2011.

[6] R. Pelrine, R. Kornbluh und P. J. Joseph, „Electrostriction of polymer dielectrics withcompliant electrodes as a means of actuation," *Sensors and Actuators A: Physical64,* 1998.

[7] S. Gratz-Kelly, A. Meyer, P. Motzki, S. Nalbach, G. Rizzello und S. Seelecke, „Force measurement based on dielectric elastomers for an intelligent glove providing worker assessment in the digital production," in *Electroactive Polymer Actuators and Devices (EAPAD) XXII*, 2020.

[8] D. Bruch, P. Loew, S. Hau, G. Rizzello und S. Seelecke, „Fast Model-based Design Of Large Stroke Dielectric Elastomer Membrane Actuators Biased With Pre-stressed Buckled Beams," in *Electroactive Polymer Actuators and Devices (EAPAD) XX*, 2018.

[9] S. Hau, D. Bruch, G. Rizello, P. Motzki und S. Seelecke, „Silicone based dielectric elastomer strip actuators coupled with non-linear biasing elements for large actuation strains," *Smart Materials and Structures,* 2018.

In-situ-Auswuchten rotierender Massen auf Basis bewegungsgesteuerter magnetischer Fluide
In-Situ-Balancing of Rotating Systems using Movement Controlled Magnetic Fluids

Valentin Schreiner, Technische Universität Berlin, Institut für Maschinenkonstruktion und Systemtechnik, 10623 Berlin, Deutschland, valentin.schreiner@emk.tu-berlin.de
Prof. Dr. Ing. Jürgen Maas, Technische Universität Berlin, Institut für Maschinenkonstruktion und Systemtechnik, 10623 Berlin, Deutschland, juergen.maas@tu-berlin.de

Kurzfassung

Durch Unwucht erzeugte Vibrationen in rotierenden Systemen können, je nach Anwendungsfall, entweder vorteilhaft genutzt werden oder, wie in den meisten Fällen, den sicheren Betrieb beeinträchtigen. Während das Prinzip des Auswuchtens bereits Stand der Technik ist, sind Systeme, die die Massenverteilung während des Betriebs variieren können, kaum zu finden. Ein vielversprechender Ansatz ist die gezielte Verschiebung von magnetischen Flüssigkeiten durch gesteuerte Magnetfelder, was Gegenstand dieses Beitrags ist. In dem hier vorgestellten Beitrag wird dieser Ansatz anhand eines realisierten Prototyps demonstriert. Das vorgeschlagene System erfordert nur eine aktive Spule, um die magnetische Flüssigkeit gezielt in situ zu bewegen und damit einen Massenausgleich zu erreichen. Mithilfe von Multiphysik-Simulationen wird das Systemverhalten analysiert und abschließend in experimentellen Untersuchungen validiert.

Abstract

Depending on the application, vibrations generated by unbalance in rotating systems can, either be used to advantage or, impair safe operation, as in most cases. While the principle of balancing is already state of the art, systems varying the mass distribution during operation are hard to find. A promising approach is the targeted displacement of magnetic fluids controlled by magnetic fields, which is the subject of this work. In this contribution, the proposed approach is demonstrated by means of a realized prototype. The proposed system requires only one active coil to selectively move the magnetic fluid in situ and achieve mass balance. With the support of multiphysics simulations, the system behavior is analyzed, optimized and finally validated in experimental investigations.

1 Einleitung

Zur Vermeidung und Kontrolle von Schwingungen in einem rotierenden System können zwei grundsätzlich unterschiedliche Ansätze Verwendung finden. Zum einen kann die Anregung von Schwingungen durch Veränderung der Massenverteilung in einen ausgeglichenen Zustand reduziert werden, zum anderen kann die auftretende Anregung durch eine geeignete Lagerung so weit gedämpft werden, dass die verbleibenden Schwingungen auf einen tolerierbaren Betrag reduziert werden [1]. In der industriellen Praxis muss immer ein Kompromiss zwischen den beiden Ansätzen gefunden werden.

Ein gängiger Ablauf für das Auswuchtverfahren basiert dabei auf mehreren Schritten. Zunächst wird das System auf eine definierte Geschwindigkeit beschleunigt und die daraus resultierenden Vibrationen werden gemessen. Anhand der aufgezeichneten Daten kann der Betrag und die Winkellage der Unwucht ermittelt werden, um die erforderliche Ausgleichsmasse zu berechnen. Je nach Verfahren wird entweder Material abgetragen (negativer Massenausgleich), zusätzliches Material aufgetragen (positiver Massenausgleich) oder Material entlang des Umfangs verlagert. Mit diesem konventionellen Auswuchtverfahren [1,2] kann die Verteilung der Masse nur im Stillstand beeinflusst und fest eigestellt werden.

In manchen technischen Anlagen kann sich die Massenverteilung jedoch im Betrieb oder während des Hochlaufvorgangs ändern. Für solche Systeme kann nicht das klassische Auswuchtverfahren verwendet werden, sondern es müssen sogenannte Echtzeitverfahren eigesetzt werden. Dabei wird die Unwucht kontinuierlich während des Betriebes ausgeglichen. Je nach Wirkmechanismus lassen sich die Echtzeit-Auswuchtverfahren in zwei weitere Untergruppen einteilen. Passive Verfahren nutzen die aus der Unwucht resultierenden Kräfte bzw. Auslenkungen, um ohne äußeren Eingriff oder Energiezufuhr einen ausgeglichenen Zustand zu erreichen.

Dies kann z. B. durch ein auf beweglich gelagerten Kugeln basierendes System [3-5] oder ein hydraulisches Auswuchtsystem [6] erreicht werden. Im Gegensatz dazu verändern aktive Systeme die Massenverteilung, indem zusätzliche Energie eingesetzt [1,7] oder dem rotierenden System Masse hinzugefügt wird [8,9]. Eine diesbezügliche Möglichkeit, die bisher nur wenig untersucht wurde, ist das Auswuchten mit magnetorheologischen Flüssigkeiten (MRF), wie in [10] und [11] vorgestellt.

MRF sind Funktionswerkstoffe, deren scheinbare Viskosität und Aggregatzustand von einem äußeren Magnetfeld abhängen. Die Grundbestandteile sind eine Trägerflüssigkeit (meist synthetisches Öl) und feine Carbonyleisenpartikel.

Ohne den Einfluss äußerer Felder sind die Partikel homogen in der Trägerflüssigkeit verteilt. Werden sie einem Magnetfeld ausgesetzt, ordnen sie sich entlang der Feldlinien zu Ketten an und erhöhen die Fließgrenze τ_0 der MRF um mehrere Größenordnungen, was als MR-Effekt bezeichnet wird [12]. MRF wird typischerweise in verschiedenen Aktoren wie Bremsen, Kupplungen [13] oder Dämpfern [14] eingesetzt. Zusätzlich kann die MRF auch durch wohldefinierte Volumenkräfte bewegt werden, wie es in [15] und [16] als MRF-Bewegungssteuerung vorgestellt und erfolgreich angewendet wird.

In den bereits erwähnten Publikationen [10] und [11] erfolgt die Änderung der Massenverteilung durch die magnetisch induzierte MRF-Bewegungssteuerung. Die magnetorheologische Flüssigkeit ist in einem zylindrischen Behälter auf dem rotierenden Teil des Systems eingeschlossen. Während in beiden Veröffentlichungen die Unwucht erfolgreich ausgeglichen wurde, sind die verwendeten Systeme relativ komplex und erfordern eine permanente Zufuhr von elektrischer Energie, um den ausgeglichenen Zustand aufrecht zu erhalten. In [10] wird zudem eine große Anzahl von Elektromagneten und eine entsprechende Leistungselektronik benötigt, um ein rotierendes Magnetfeld zu erzeugen. Das in [11] vorgestellte System benötigt zwar eine geringere Anzahl von Elektromagneten, diese müssen aber aufgrund ihrer Integration auf dem Rotor über Schleifringe mit Energie versorgt werden.

Das im Rahmen dieses Beitrages entwickelte System benötigt nur einen stationären Elektromagnet und hält den Gleichgewichtszustand im Betrieb auch ohne elektrische Energie aufrecht. Somit können die Nachteile der zuvor genannten Systeme überwunden werden.

Um die Funktion und den Aufbau des vorgeschlagenen Systems zu erläutern, wird in Abschnitt 2 das grundlegende Funktionsprinzip sowie der mögliche Betriebsbereich des Systems erklärt. In Abschnitt 3 werden Simulationsergebnisse für verschiedene Betriebsfälle vorgestellt, welche in Abschnitt 4 der experimentellen Validierung gegenübergestellt werden. Abschließend werden Messungen zum Funktionsnachweis des vorgeschlagenen magnetischen Auswuchtsystems gezeigt.

2 Funktionsprinzip des In-situ-Auswuchtsystems

Das angewandte Funktionsprinzip basiert auf der Verschiebung von magnetisierbarer Masse in radialer Richtung und soll anhand **Bild 1** erläutert werden. Als bewegte Masse wird MRF verwendet, wobei das Gesamtvolumen gleichmäßig auf drei auf dem Umfang angeordnete Kammern verteilt ist. Die einzelnen Kammern sind nur bis zu einem bestimmten Grad gefüllt, sodass innerhalb der Kammern die MRF bewegt werden kann und sich die Massenverteilung und der Zustand des Gesamtsystems einstellen lässt.

Mittels eines Permanentmagneten im Inneren jeder Kammer wird die MRF in einem Bereich möglichst nahe an der Drehachse, also auf einem kleinen Wirkradius, als Festkörper fixiert. Durch gezieltes Verflüssigen der MRF an einem der Unwucht gegenüberliegenden Punkt kann der Auswuchtprozess durchgeführt werden. Bewegt sich ein Teil der MRF aus dem Wirkbereich des Permanentmagneten heraus, wird er durch radiale Beschleunigung weiter in den Außenbereich der Kammer bewegt. Die zur Kompensation verwendete Unwucht ist dann das Produkt aus bewegter MRF und Änderung deren mittleren Durchmessers:

$$U_{\mathrm{MRF}} = m_{\mathrm{MRF}} \cdot r_{\mathrm{MRF}} \tag{1}$$

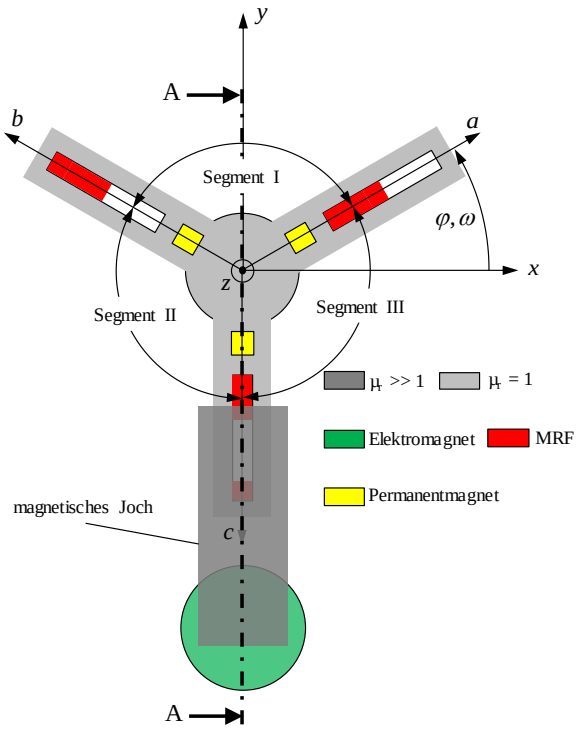


Bild 1 Schematische Darstellung des Konzepts. Kammer a: gesamte MRF in innerer Position, Kammer b: gesamte MRF in äußerer Position, Kammer c: teilweise bewegte MRF.

Um die ursprüngliche Massenverteilung wiederherzustellen, muss die nach außen bewegte MRF wieder in den inneren Bereich der Kammer bewegt werden. Bei einer horizontalen Drehachse, die im Folgenden angenommen wird, kann die Schwerkraft für diese Rückwärtsbewegung genutzt werden. Durch das Anhalten des Rotors bei einer nach oben gerichteten Kammer fließt die MRF zurück. Diese Bewegung kann jedoch durch den MR-Effekt behindert werden. Neben der Schwerkraft kann, bei entsprechender Positionierung, auch der Elektromagnet genutzt werden, um das Magnetfeld gezielt zu beeinflussen und damit den Rückfluss der MRF verbessern.

Das Funktionsprinzip der angewandten MRF-Bewegungssteuerung wird anhand von **Bild 2** erläutert. Die MRF ist in der Kammer verschiedenen Kräften ausgesetzt, für den Fall, dass es zu keiner Bewegung der MRF kommt, befinden sich diese Kräfte im Gleichgewicht. Aufgrund ihrer Masse wirken Schwerkraft und Rotationsbeschleunigungen auf die MRF. Ebenso wirken Druckkräfte über die Kammerwände und die freie Oberfläche auf die MRF. Die magnetischen Eigenschaften der

MRF führen zu einer magnetisch induzierten Kraft, der Kelvinkraft $\mathbf{f}_k$, die von der Größe und dem Gradienten der Magnetfeldstärke abhängt. Schließlich ermöglicht der MR-Effekt, auch in Ruhe, die Übertragung einer Scherspannung τ.

Um die Bewegung der MRF innerhalb der Kammer zu verhindern, muss die resultierende Schubspannung τ aufgrund der zuvor beschriebenen Kräfte unterhalb der magnetisch induzierten Fließgrenze τ_0 liegen. Um einen Teil der MRF nach außen zu bewegen, muss entweder die Schubspannung τ erhöht oder die magnetische Flussdichte B in der MRF geschwächt werden. Infolgedessen kann die MRF durch die Volumenkraft der Radialbeschleunigung $\mathbf{f}_a$ nach außen bewegt werden.

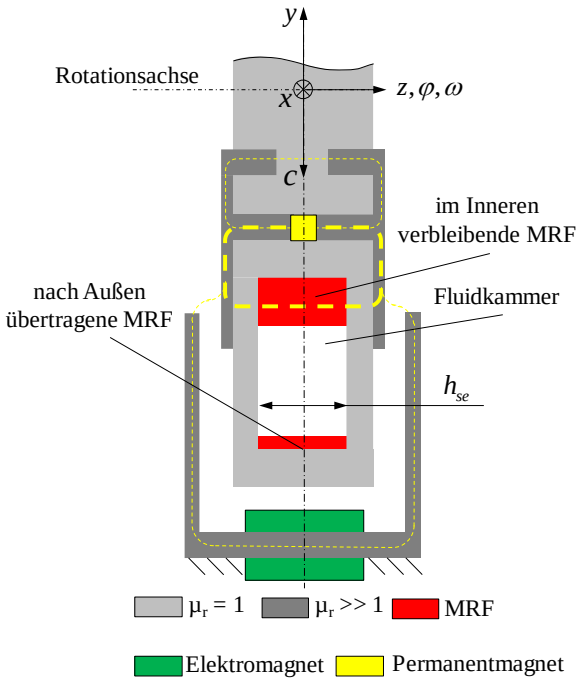


Bild 2 Schnittansicht nach dem in Bild 1 dargestellten Verlauf A-A zur Darstellung des magnetischen Kreises.

Aus dem beschriebenen Funktionsprinzip lässt sich eine grundlegende Grenze des nutzbaren Drehzahlbereichs ableiten. Dieser Bereich ist in **Bild 3** dargestellt. Darin ist der quadratische Verlauf der Schubspannung τ_a durch die Radialbeschleunigung a gezeigt, der entsprechend [17] nach Gleichung 2 berechnet werden kann.

$$\tau_a = \frac{\omega^2 \cdot h_{se} \cdot y \cdot \rho}{2} \tag{2}$$

Darüber hinaus ist in Bild 3 der Bereich dargestellt, in dem die entgegenwirkende Fließgrenze τ_0 eingestellt werden kann. Die Fließgrenze τ_0 wird durch die magnetische Flussdichte B in der MRF und damit durch den Elektromagneten gesteuert. Der Maximalwert der Fließgrenze $\tau_0(B_{max})$ wird bei deaktiviertem Elektromagnet erreicht. Durch Erhöhung des Stroms im Elektromagneten I_{EM} wird der magnetische Fluss innerhalb der MRF abgesenkt und die Fließgrenze τ_0 sinkt, bis bei einer entsprechenden Kompensation des magnetischen Flusses in der MRF $\tau_0(B_{min})$ erreicht wird.

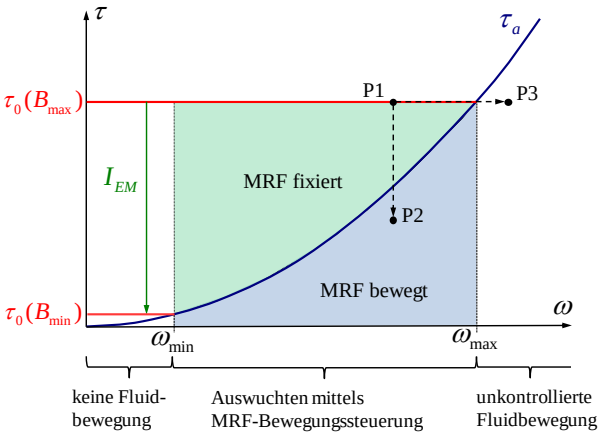


Bild 3 Betriebsdiagramm, eingeteilt in drei verschiedene Drehzahlbereiche.

Vernachlässigt man den Einfluss der Schwerkraft und der Kelvinkraft, so ergeben sich die Schnittpunkte zwischen der Schubspannung τ_a und den Bereichsgrenzen der Fließgrenze τ_0 für die drei Bereiche. Zwischen Stillstand und der Winkelgeschwindigkeit ω_{min} kann der magnetische Fluss nicht ausreichend kompensiert werden und die Fließgrenze liegt oberhalb der, durch die Radialbeschleunigung a verursachten, Schubspannung τ_a:

$$0 \le \omega < \omega_{min} \Rightarrow \tau_a < \tau_0(B_{min}) \tag{3}$$

In diesem Betriebsbereich ist keine Flüssigkeitsbewegung und entsprechend auch kein Auswuchten möglich. Oberhalb von ω_{max}, wie in Gleichung 4 angegeben, übersteigt die Schubspannung τ_a die maximale Fließgrenze $\tau_0(B_{max})$, die Fluidbewegung kann nicht gesteuert werden und ein spezifisches Auswuchten ist nicht möglich.

$$\omega_{max} < \omega \Rightarrow \tau_a > \tau_0(B_{max}) \tag{4}$$

Im Drehzahlbereich zwischen ω_{min} und ω_{max} ist die MRF einerseits durch die Fließgrenze $\tau_0(B_{max})$ fixiert, wenn der Elektromagnet nicht bestromt wird. Andererseits kann die MRF nach außen verschoben werden, indem der Elektromagnet bestromt und dadurch die Fließgrenze $\tau_0(B)$ gesenkt wird. Nur in diesem Bereich ist ein kontrollierter Auswuchtvorgang möglich.

3 Modellbasierte Simulation der Fluidbewegungssteuerung

Die Bewegung von Fluiden wird üblicherweise über die Navier-Stokes-Gleichung beschrieben. Da die MRF ferromagnetische Eigenschaften aufweist, muss zur Berechnung zusätzlich eine magnetisch induzierte Volumenkraft, die sogenannte Kelvinkraft $\mathbf{f}_k$, berücksichtigt werden [15]. Die Navier-Stokes-Gleichung wird daher um einen zusätzlichen Term $\mathbf{f}_k$ erweitert:

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\nabla)\,\mathbf{v}\right) = -\nabla p + \eta \cdot \Delta \mathbf{v} + \rho\left(\mathbf{g} + \omega^2 r\right) + \mathbf{f}_k . \tag{5}$$

Darin beschreibt ρ die Dichte der MRF, v das Geschwindigkeitsfeld, p den Druck, g die Beschleunigung aufgrund des Erdschwerefelds, ω die Winkelgeschwindigkeit und r die radiale Position der MRF. Die Kelvinkraft $\mathbf{f}_k$ ergibt sich zu:

$$\mathbf{f}_k = \mu_0 (\mathbf{M} \cdot \nabla) \mathbf{H}, \tag{6}$$

mit μ_0 als Vakuumpermeabilität, der Magnetisierung **M** und der magnetischen Feldstärke **H**.

Die Simulation der Fluidbewegung wird in *Comsol Multiphysics* als Zweiphasen-Strömung (MRF und Luft) durchgeführt, wobei die Level-Set-Methode angewendet wird. Der Level-Set-Parameter ϕ unterscheidet dabei die verschiedenen Phasen MRF ($\phi = 1$) und Luft ($\phi = 0$). Um die magnetisch induzierte Fließgrenze $\tau_0(B)$ der MRF zu berücksichtigen, wird das Bingham-Papanastasiou-Modell [18], zur Bestimmung der scheinbaren Viskosität angewendet.

Das Modell wird nun verwendet, um das Systemverhalten zu untersuchen und kritische Werte zu identifizieren. Zunächst wird die kritische Drehzahl ermittelt, bei der die MRF ohne Einwirkung des Elektromagneten bewegt wird. Die Rotation des Aktors wird in der Simulation durch die als Beschleunigung wirkende Zentrifugalbeschleunigung berücksichtigt. Während der Simulation wird die Rotationsgeschwindigkeit stetig erhöht.

Zur Auswertung der Ergebnisse wird der Betrag der Unwucht U_{MRF}, der sich aus der Verschiebung der MRF ergibt, herangezogen. Die inkrementelle Unwucht für jedes Element der Simulation dU_{MRF} kann aus der Masse der MRF multipliziert mit ihrem Abstand zur Drehachse bestimmt werden. Durch Integration über das Volumen der Kammer wird die Unwucht, die sich aus der Flüssigkeitsverteilung ergibt berechnet:

$$U_{MRF} = \iiint (\rho_{MRF} \cdot \phi + \rho_{air} \cdot (1 - \phi)) \cdot y \, dx \, dy \, dz \tag{7}$$

Das Ergebnis der Auswertung ist in **Bild 8** dargestellt. Das Diagramm zeigt, dass unterhalb von $\omega = 25$ rad/s eine unterkritische Drehzahl angenommen werden kann.

Zur Untersuchung des Auswuchtvorgangs, wird in einer weiteren Simulation das Verhalten unter dem Einfluss des bestromten Elektromagneten betrachtet. Dazu wird der Elektromagnet mit einem zeitlich veränderlichen Strom beaufschlagt. Um die Bedingung zu erreichen, dass die Kammer den aktivierten Luftspalt des Elektromagneten während fünf aufeinander folgender Umdrehungen durchläuft, wird der in **Bild 4** dargestellte zeitlich variable Strom I_{EM} eingesetzt.

Der Anfangszustand der Flüssigkeitsverteilung zum Zeitpunkt $t = 0$ s ist in **Bild 5 a)** dargestellt. Wie in Bild 5 zu erkennen ist, ändert sich die Unwucht während der ersten drei Umdrehungen trotz Aktivierung des Elektromagneten nicht. In der Flüssigkeitsverteilung ist jedoch bei $t = 1$ s in Bild 5 b) eine erste Inhomogenität zu beobachten. Am Ende der vierten Umdrehung tritt erstmals eine deutlich sichtbare Veränderung der Unwucht auf. Bevor der Elektromagnet das nächste Mal bei $t = 1{,}4$ s angesteuert wird, ist ein stationärer Zustand erreicht, der in Bild 5 c) dargestellt ist. Die letzten beiden Stromimpulse führen zu einer Änderung der Unwucht um je ca. 100 g mm, was schließlich zu dem in Bild 5 d) und e) dargestellten Zustand führt.

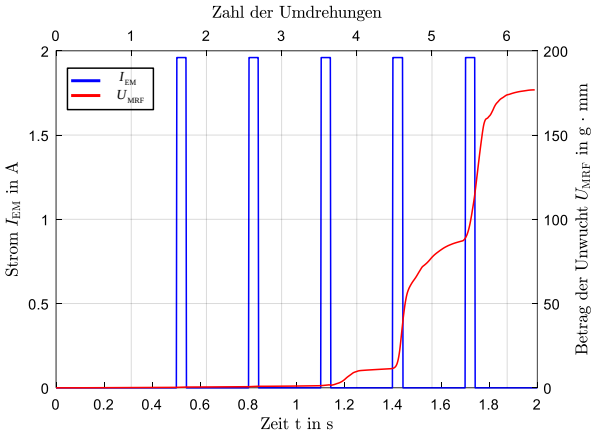


Bild 4 Der zur Simulation vorgegebene Stromverlauf I_{EM} sowie die resultierende Unwucht U_{MRF}.

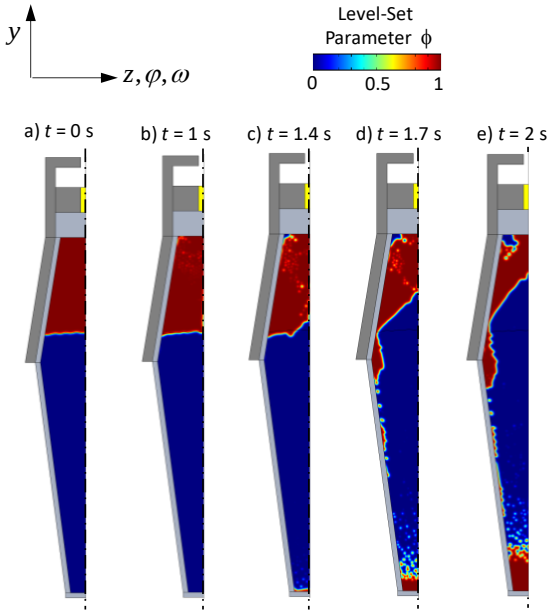


Bild 5 Simulationsergebnisse der MRF-Bewegung durch den in Bild 4 vorgegebenen Strom im Elektromagneten I_{EM}.

4 Experimentelle Validierung des vorgeschlagenen Systems

Der für die experimentelle Validierung verwendete Aufbau ist in **Bild 6** dargestellt. Er besteht aus einem drehzahlgeregelten elektrischen Antrieb, dem Rotor mit der MRF, der Apparatur zur Unwuchtmessung und einem Elektromagneten zur Beeinflussung der Massenverteilung der MRF. Die Antriebseinheit verfügt über einen optischen-Geber, mit dem die Winkellage und die Drehzahl ermittelt werden können. Der Rotor ist starr mit der Antriebswelle verbunden. Beide befinden sich auf der schwingenden Grundplatte, deren Auslenkung durch einen Lasertriangulationssensor erfasst wird.

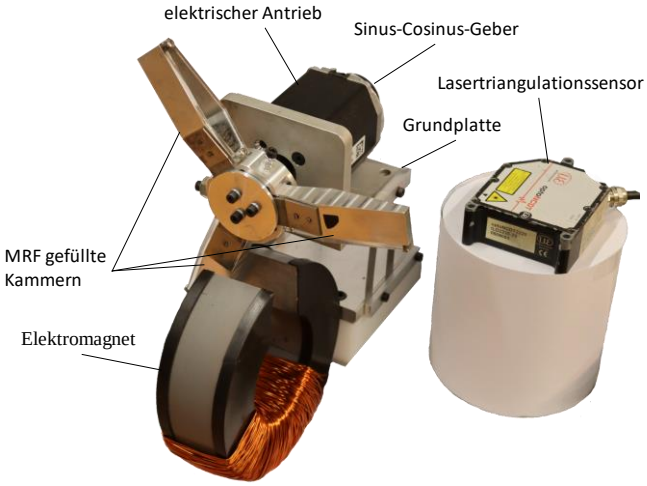


Bild 6 Versuchsaufbau zur experimentellen Validierung des Funktionsprinzips, bestehend aus Antrieb, Rotor, stationärem Elektromagnet und Unwuchtmessung.

Das schwingende System wird, wie in **Bild 7** dargestellt, als schwach gedämpfter Einmassenschwinger modelliert. Da die Unwucht eine periodische, cosinusförmige Anregung zur Folge hat, kann die ursächliche Unwucht durch Auswertung von Amplitude und Phasenlage der Beschleunigung ermittelt werden. Durch Demodulation des Beschleunigungssignals und Berücksichtigung der Systemmasse m_s können Amplitude und Phasenlage der Anregung ermittelt werden.

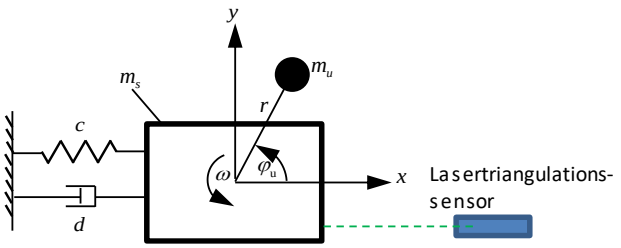


Bild 7 Mechanisches Ersatzmodell zur Bestimmung der unbekannten Unwucht $u = m_u \cdot r$.

Um die in der Simulation ermittelte kritische Drehzahl zu validieren, wird eine Kammer des Aktors, wie in **Bild 8 a)** zu sehen, mit 5 ml Basonetic 5030 befüllt. Die verbleibende Unwucht beträgt 3,9 g mm.

Durch schrittweises Erhöhen der Drehzahl kann die kritische Drehzahl durch gleichzeitige Aufzeichnung des Unwuchtzustandes ermittelt werden. Die aufgezeichnete Unwucht in Bild 8 ist in Abhängigkeit der Drehzahl dem aus der Simulation ermittelten Wert gegenübergestellt.

Die Ergebnisse von Simulation und Experiment stimmen gut überein, wobei die Simulation bei höheren Drehzahlen eine geringere Verschiebung prognostiziert.

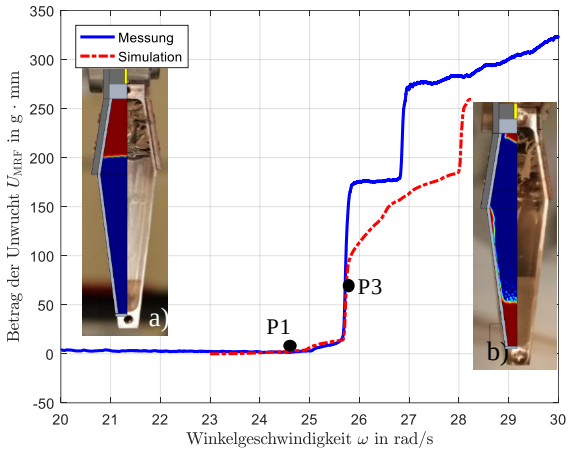


Bild 8 Bestimmung der kritischen Winkelgeschwindigkeit ω_{crit}, Vergleich zwischen Experiment und Simulation. a) Beginn des Experiments, b) Ende des Experiments.

Zum Nachweis der Funktionsfähigkeit, wird ein Auswuchtlauf durchgeführt, indem alle drei Kammern mit einer Menge von 16,5 g MRF gefüllt und zusätzliche Gewichte hinzugefügt werden, die eine anfängliche Unwucht verursachen. Bei einer Winkelgeschwindigkeit von $\omega = 20$ rad/s wird der automatische Auswuchtvorgang gestartet. Amplitude und Winkel der Unwucht werden dabei aufgezeichnet, das Ergebnis ist in **Bild 9** dargestellt und in **Tabelle 1** aufgeführt. Die anfängliche Unwucht erfordert das Bewegen der MRF in Kammer a und Kammer b. Durch den verwendeten Ablauf wird zunächst Kammer b als die effizienteste Möglichkeit identifiziert und eine entsprechende Masseverschiebung durchgeführt. Diese ist durch den schwarzen Pfeil gekennzeichnet. Nachdem eine gewisse Masse verschoben wurde, ergibt sich die Lage der resultierenden Unwucht (Phasenlage etwa 190°), sodass nun mit Kammer a effizienter korrigiert werden kann. Zuletzt wird eine Phasenlage der Unwucht von etwa 300° erreicht und es muss wiederum Kammer b aktiviert werden. Am Ende des dargestellten Verlaufs wird der festgelegte Grenzwert (5 g mm) unterschritten und damit ein während der Rotation stabiler, ausgewuchteter Zustand erreicht, mit dem der Vorgang beendet wird. Wie in Tabelle 1 vermerkt, wurde die Unwucht um mehr als 99 % vermindert.

Tabelle 1 Ergebnisse des automatisierten Auswuchtlaufs.

Wert Unwucht	**Start**	**Ende**
Phasenlage	$\varphi_{u,start} = 262\,°$	$\varphi_{u,end} = 169\,°$
Amplitude	$\hat{u}_{start} = 255\,g$ mm	$\hat{u}_{end} = 2,34\,g$ mm

Die weitere Entwicklung sieht vor, mit einem weiterentwickelten Algorithmus zeitgleich MRF in zwei Kammern zu verschieben. Dadurch ist es möglich, sich auf direktem Weg dem ausgewuchteten Zustand anzunähern. Dieser bisher nicht experimentell bestätigte Ablauf ist in Bild 9 als grüner Pfeil dargestellt.

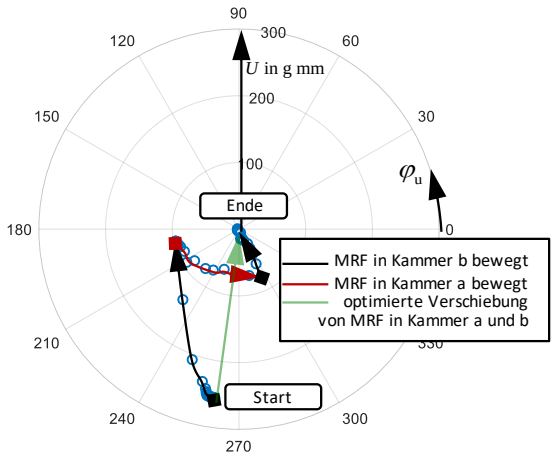


Bild 9 Ergebnisse des automatisierten Auswuchtlaufs. Schwarze Pfeile: Kammer a aktiviert, roter Pfeil: Kammer b aktiviert.

5 Zusammenfassung

In diesem Beitrag wurde ein neuartiges System zum automatischen In-situ-Auswuchten vorgestellt. Das Funktionsprinzip des Systems basiert auf einer gesteuerten MRF-Fluidbewegungssteuerung unter dem Einfluss von Rotationsbeschleunigung, Kelvinkraft, Schwerkraft und magnetisch induzierter Fließgrenze. Zur Untersuchung der Systemcharakteristik wurde das Fluidverhalten der MRF mit dem Bingham-Papanastasiou-Modell abgebildet. Durch die Berücksichtigung der Fließgrenze konnte eine gute Übereinstimmung zwischen numerischer Simulation und experimenteller Untersuchung erreicht werden. Die durch die Simulation ermittelte kritische Drehzahl stimmt sehr gut mit dem experimentell ermittelten Wert überein. Der Einfluss des steuernden Elektromagneten auf die Fluidbewegung in der Simulation konnte ebenfalls nachgewiesen werden. Es wurde ein Versuchsaufbau des Systems gefertigt und mit einer schwingungsbasierten Unwuchtmessung ausgestattet. Mit Hilfe des Versuchsaufbaus konnte eine aufgebrachte Unwucht mit zufälliger Verteilung um über 99 % reduziert und damit das Potenzial des vorgestellten Systems zum In-situ-Auswuchten nachgewiesen werden.

6 Literatur

[1] Fomi Wamba, F.: *Automatische Auswuchtstrategie für einen magnetgelagerten elastischen Rotor mit Auswuchtaktoren*. Dissertation. Technische Universität Darmstadt, 2009.

[2] Parkinson, A. and Bishop, R.: *Residual vibration in modal balancing.* Journal of Mechanical Engineering Science Vol. 7, pp. 33-39. 1965.

[3] Huang, W., Chao, C., Kang, J., and Sung, C.: *The application of ball-type balancers for radial vibration reduction of high-speed optic disk drives.* Journal of Sound and Vibration Vol. 250. 2002.

[4] Majewski, T., Szwedowicz, D. and Melo, M.: *Self-balancing system of the disk on an elastic shaft*. Journal of Sound and Vibration, Volume 359, 2015.

[5] Rodrigues, D. et al.: *Experimental investigation of a single-plane automatic balancing mechanism for a rigid rotor.* Journal of Sound and Vibration. Vol. 330. 2011.

[6] Bae, S., et al.: *Dynamic analysis of an automatic washing machine with a hydraulic balancer.* Journal of Sound and Vibration Vol. 257, pp. 3–18, 2002.

[7] Li-Fang, C., Xi, C. and Jin-Ji, G.: *A study on electromagnetic driven bi-disc compensator for rotor autobalancing and its movement control*. WSEAS Transactions on Systems and Control Vol. 5, 2010.

[8] Xin, P., Haiqi, W., Jinji, G. and Weimin, W.: *New Liquid Transfer Active Balancing System Using Compressed Air for Grinding Machine*, Journal of Vibration and Acoustics, 2015.

[9] Zhang, X., Liu, X. and Zhao, H.: *New Active Online Balancing Method for Grinding Wheel Using Liquid Injection and Free Dripping*. Journal of Vibration and Acoustics, 2018.

[10] Nakamoto, K., Adachi, K. and Shirase K.: *Proposal of Real-Time Balancing mechanism Using Magnetic Fluid for Machine Tool Spindle*. Manufacturing Systems and Technologies for the New Frontier: The 41st CIRP, 2008.

[11] Zhang, X et al.: *An Online Active Balancing Method Using Magnetorheological Effect of Magnetic Fluid*. Journal of Vibration and Acoustics Vol. 141, 2019.

[12] Carlson, J.D, Janocha, H. (Hrsg.): *Magnetorheological fluid actuators.* Adaptronics and Smart Structures - Basics, Materials, Design and Applications. 2 ed., Springer, Berlin Heidelberg, p. 184-204, 2007.

[13] Maas, J.; Güth, D.; Wiehe, A.: *MRF-Actuator Concepts for HMI and Industrial Applications*. SPIE Smart Structures/NDE, Vol. 7977, S. 797714, San Diego, Kalifornien, USA, März 2011.

[14] A.S. Tan, M. Dahlmann and T. Sattel.: *Limitation of an adjustable two degree-of-freedom damper system for directional damping*. Special Issue: 89th Annual Meeting of the International Association of Applied Mathematics and Mechanics. Vol. 18, Issue 1, 2018.

[15] Güth, D., Schamoni, M. and Maas, J.: *Magnetic fluid control for viscous loss reduction of high-speed MRF brakes and clutches with well-defined fail-safe behavior*. Smart Materials and Structures, Vol. 22, No. 9, 094010, 2013.

[16] Krüger, T., Hegger, C., Schreiner, V. and Maas, J.: *Mass adaptation for moving parts using a magnetically induced placement of a certain amount of magnetorheological fluid.* Proceedings of the SPIE, Volume 10967, id. 109671X 8 pp. (2019).

[17] Hegger, C. and Maas, J.: *Smart Sealing for MR-Fluid Actuators.* Journal of Intelligent Material Systems and Structures, Volume 30 Issue 5, 2018.

[18] Papanastasiou, T.: *Flows of Materials with Yield*. Journal of Rheology, Vol. 31, pp. 385-404., 1987.

Impedanzregelung eines Luftmanipulationssytems bestehend aus einem Multirotor und einem Sechs-Gelenk-Manipulator

Impedance Control of an Aerial Manipulator Composed of a Multirotor and a 6-DOF Manipulator

Chao Yao, TU Dresden, Institut für Automatisierungstechnik, 01069 Dresden, Deutschland, chao.yao@tu-dresden.de

Stephan Schwarz, TU Dresden, Institut für Automatisierungstechnik, 01069 Dresden, Deutschland, stephan_-andreas.schwarz@tu-dresden.de

Klaus Janschek, TU Dresden, Institut für Automatisierungstechnik, 01069 Dresden, Deutschland, klaus.janschek@tu-dresden.de

Kurzfassung

In dieser Arbeit wird eine Impedanzregelung eines freifliegenden, sechsachsigen Manipulators vorgestellt. Der Manipulator ist dabei an einem vollaktuierten Multirotor befestigt, dessen Pose durch eine geometrische Lageregelung geregelt wird. Der vorgestellte Impedanzregler passt die Endeffektor-Zielpose anhand der auf den Endeffektor wirksamen Kräfte und Momente an. Die angepasste Referenzpose wird durch Lösen der inversen Kinematik in den Gelenkraum überführt und die Winkel an die Gelenkregelung übergeben. Die Validierung erfolgt simulativ im Gazebo-Framework, wo die Mechanik und Sensormessungen simuliert werden. Außerdem wurde die Regelgüte der Impedanzregelung durch Untersuchungen der Eigenfrequenzen bestimmt. Die SIL-Tests zeigen, dass das Manipulationssystem der Kontaktkraft wie ein Feder-Masse-Dämpfer-System ausweicht. Aufgrund des Messrauschens entstehen erwartete Abweichungen zur modellierten Impedanz.

Abstract

This paper presents an impedance control for a free-flying 6-DOF manipulator, whereby the effective contact forces can be reduced to avoid possible damages. The manipulator is attached to a fully-actuated multirotor, the pose of which is controlled by a geometric controller. For the manipulator, the presented impedance controller adjusts the desired pose of the end-effector based on the forces and torques acting on the end effector. The adjusted reference-pose is transferred to the joint space by solving the inverse kinematics and the angles are provided to the joint control. The validation was carried out with Gazebo-simulator, which simulates the mechanics and sensor measurements. In addition, the performance of the impedance control was evaluated by investigating the eigen frequency. The SIL-tests show that the system reacts as expected to the force like a spring-mass-damper system. Due to measurement noises, expected deviations from the modeled impedance arise.

1 Einleitung

Abseits der bereits ausgiebig erforschten Manipulatoren mit fester Basis, gewann die luftgestützte Manipulation in den letzten Jahren zunehmend an Bedeutung. Dabei wird an einem Luftfahrzeug ein Manipulator befestigt, der mit seinem Umfeld interagieren kann [6, 10]. In den letzten Jahren wurden einige Untersuchungen zu verschiedenen Kombinationen einer Drohne mit einem Manipulator durchgeführt und neue Ansätze entwickelt. So wurden unter anderem in [1, 10] Regler für unteraktuierte Drohnen mit angebauten, komplexen Manipulatoren entworfen. In beiden Fällen wurde die Redundanz auf Kosten der Nutzlast und deutlich erhöhter Reglerkomplexität erreicht. Gioioso et al. [3] erweiterten einen Quadrocopter um einen starren Greifer, welcher in der Nähe des Schwerpunktes der Drohne befestigt wurde und mit der Umgebung interagieren kann. Dieser Reglerentwurf basiert dabei auf dem für kontaktlosen Betrieb konzipierten Regler. Diese Ansätze betrachten das Luftfahrzeug und den Manipulator als einzelne Komponenten, sodass der Einfluss des Arms für die Drohne als externe Größe betrachtet wird. In [4] wurde ein Hexacopter mit einem dreiachsigen Manipulator zu einem Gesamtsystem zusammengeführt und vorgestellt. Durch das redundante System konnten zusätzliche Anforderungen eingeführt und erfüllt werden. Damit der Luftmanipulator seine Zielposition sicher erreichen kann, ist es notwendig, dass das Luftfahrzeug stabil eine bestimmte Pose anfahren und halten kann. Dazu wird das Prinzip der vollaktuierten Drohne angewendet, bei dem die Rotoren gegenläufig verkippt werden [5, 7, 8, 9]. Sobald der Manipulator nun mit der Umgebung interagiert, wirken Kontaktkräfte auf den Endeffektor. Befindet sich die Zielpose des Manipulators nun innerhalb des Objektes, drückt der Arm immer stärker gegen das Objekt. Dadurch steigt die Kontaktkraft an, was zur Beschädigung des Objektes oder zum Absturz der Drohne führen kann. Um dies zu vermeiden, ist es ratsam die Zielpose des Endeffektors bei wirkenden Kräften anzupassen [2]. Eine Impedanzregelung dient im Allgemeinen dazu, die Pose des Endeffektors in Abhängigkeit der wirkenden Kräfte (bzw. Momente) so zu

korrigieren, dass die Störkräfte bzw. -momente nachgiebig eingeleitet werden und somit kein Schaden durch den Manipulator entsteht. Dabei wird eine gewünschte Impedanz festgelegt, nach der sich das System verhalten soll.
In dieser Arbeit wird die Drohnenpose durch eine eigenständige Lageregelung ähnlich wie in [7] geregelt. Anhand der Zielpose und der auf den Endeffektor wirksamen Kräfte und Momente soll durch den Impedanzregler eine neue Referenzpose gebildet werden, die an die Lageregelung übergeben wird. Ziel der Regelung ist es, trotz wirksamer Stör- oder Kontaktkräfte ein stabiles Führungsverhalten des Endeffektors und der Drohne zu gewährleisten. Als Impedanz wurde ein Feder-Masse-Dämpfer-System modelliert, wodurch die Verläufe gedämpften Schwingungen entsprechen. Die angepasste Referenzpose wird durch numerische Integration mithilfe des expliziten Euler-Verfahrens generiert. Diese wird anschließend durch Lösen der inversen Kinematik in den Gelenkraum überführt und die Winkel an die Gelenkregelung übergeben. Die entwickelten Module wurden im Robot Operating System (ROS) implementiert. Zusätzlich wurde eine grafische Nutzeroberfläche erstellt, mit der benötigte Eingaben und Parameteränderungen getätigt werden können. Die Validierung erfolgt mit dem Gazebo-Simulator, welcher die Mechanik und Sensormessungen simuliert. Die Regelgüte der Impedanzregelung wurde durch Untersuchungen der Eigenfrequenzen bestimmt.
Der Rest dieses Beitrags ist wie folgt organisiert. Das System wird in Abschnitt 2 modelliert. Anschließend wird die Regelung in Abschnitt 3 entworfen. Abschnitt 4 stellt die Simulationskonfiguration in Gazebo vor und bewertet anhand der Ergebnisse die Regelgüte des Impedanzreglers. Der Abschnitt 5 schließt diesen Beitrag mit einer Zusammenfassung und einem Ausblick ab.

2 Modellierung

2.1 Nomenklatur und Referenzsysteme

Für eine eindeutige Darstellung der Zusammenhänge wird im Folgenden ein skalarer Wert durch einen kursiven Kleinbuchstaben, beispielsweise a geschrieben. Ein Vektor wird durch einen fettgedruckten, kursiven Kleinbuchstaben mit einem Index links oben beschrieben. Dieser Index definiert das Referenzsystem, in dem der Vektor dargestellt wird. Somit wird beispielsweise der Vektor ${}^{A}\boldsymbol{v}$ relativ zum {A}-System dargestellt. Um die Anzahl der Indizes gering zu halten, wird bei einem Vektor, welcher im Inertialsystem {I} dargestellt ist, dieser Index weggelassen. Alle Matrizen werden in dieser Arbeit durch einen fettgedruckten, kursiven Großbuchstaben geschrieben. Die folgenden Formeln beschreiben die euklidische Transformation von {B} nach {A}:

$$ {}^{A}\boldsymbol{v} = {}^{A}_{B}\boldsymbol{R}\,{}^{B}\boldsymbol{v} + {}^{A}\boldsymbol{p}_{B}, \quad {}^{A}_{B}\boldsymbol{R} \in SO(3) \ . \tag{1} $$

Die Rotationsmatrix ${}^{A}_{B}\boldsymbol{R}$ beschreibt die Orientierung des {B}-Systems relativ zum {A}-System. ${}^{A}\boldsymbol{p}_{B} \in \mathbb{R}^3$ ist die Position der Ursprung von {B} beschrieben in {A}. Die Transformation kann kompakt mit 4×4 homogenen Transformationsmatrizen beschrieben werden:

$$ {}^{A}_{B}\boldsymbol{T} = \begin{bmatrix} {}^{A}_{B}\boldsymbol{R} & {}^{A}\boldsymbol{p}_{B} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix} \ . \tag{2} $$

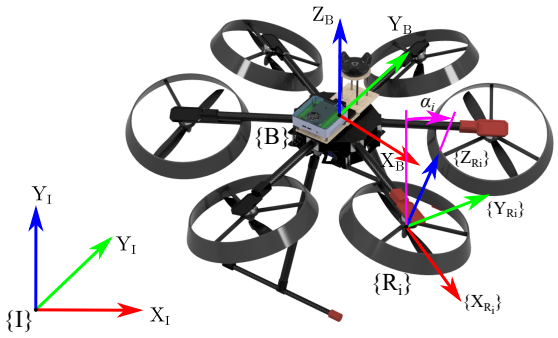


Abbildung 1 Definition der Referenzsysteme des Hexacopters sowie des Inertialsystems und der Kippwinkel

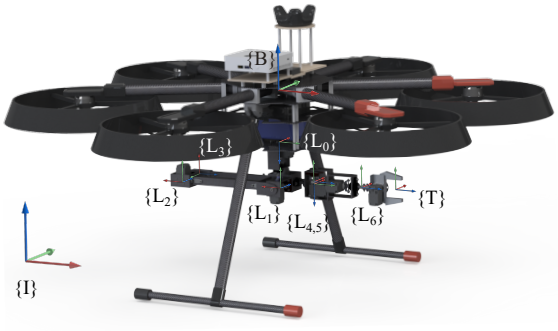


Abbildung 2 Referenzsysteme des sechsachsigen Manipulators

Im **Bild 1** sind die für die Betrachtungen des Hexacopters benötigten Referenzsysteme dargestellt. Diese beinhalten das Inertialsystem {I} und das Körper-System {B}. Das Inertialsystem ist dabei raumfest, während das körperfeste System fest mit der Drohne verbunden ist und sich in deren Schwerpunkt befindet. Das Rotor-System des i-ten Rotors R_i ist fest mit dem Rotor verbunden und definiert durch die Rotation von {B} um die Z_B-Achse mit dem Winkel $\gamma_i = \frac{(2i-3)\pi}{6}$, $i = 1\ldots6$. Die Z_{R_i}-Achse ist mit der Schubrichtung ausgerichtet. Die Winkel α_i werden dementsprechend Kippwinkel genannt.
Für den Manipulator werden zusätzliche Referenzsysteme definiert. Im **Bild 2** ist der sechsachsige Arm mit den definierten Systemen abgebildet. $\{\mathrm{L}_0\}$ bezeichnet das Arm-Basis-System. Das Endeffektor-System {T} liegt im Tool-Center-Point (TCP) und ist gemäß der Abbildung ausgerichtet.

2.2 Drohnen-Dynamik

Es wird angenommen, dass die Schubkraft f_T und das Reaktionsmoment τ_D, generiert von den einzelnen Propellern, proportional zum Quadrat der Propellerdrehzahl sind

$$ f_T = k\omega^2 \tag{3} $$

$$ \tau_D = b\omega^2 \ , \tag{4} $$

wobei k und b die entsprechenden Koeffizienten sind.
Die Gleichung für die translatorische Bewegung der Drohne in {I} ist

$$m\ddot{\boldsymbol{p}} = \boldsymbol{f}_G + \boldsymbol{f}_T + \boldsymbol{f}_{\text{ext}} \quad , \tag{5}$$

wobei $\boldsymbol{f}_G = \begin{bmatrix} 0 & 0 & -mg \end{bmatrix}^\top$ die Schwerkraft, $\boldsymbol{f}_T$ die Gesamtschubkraft und $\boldsymbol{f}_{\text{ext}}$ die externen Kräfte darstellen. Die Gesamtschubkraft $\boldsymbol{f}_T$ ist die Summe der Schubkräfte der Rotoren, die nach {I} transformiert werden:

$$\boldsymbol{f}_T = \sum_{i=1}^{6} {}^I_B\boldsymbol{R}\, {}^B_{R_i}\boldsymbol{R}\, f_{T,i}\, \boldsymbol{e}_3 = \sum_{i=1}^{6} {}^I_B\boldsymbol{R}\, {}^B\boldsymbol{e}_i\, k\omega_i^2 \quad , \tag{6}$$

wobei ${}^B\boldsymbol{e}_i$ ein Einheitsvektor ist, der die Schubrichtung des i-ten Rotors zeigt.
Die Rotation wird im körperfesten Koordinatensystem {B} beschrieben. Mit der Annahme, dass das Trägheitsmoment der Drohne $\boldsymbol{I}_B$ konstant ist, ergibt sich die Eulersche Kreiselgleichung als

$${}^B\boldsymbol{\tau}_T + {}^B\boldsymbol{\tau}_D + {}^B\boldsymbol{\tau}_{\text{ext}} = \boldsymbol{I}_B\, {}^B\dot{\boldsymbol{\omega}} + {}^B\boldsymbol{\omega} \times \boldsymbol{I}_B\, {}^B\boldsymbol{\omega} \quad , \tag{7}$$

wobei ${}^B\boldsymbol{\omega}$ die Winkelgeschwindigkeit der Drohne in {B} darstellt. Das Drehmoment resultiert aus der Schubkraft ${}^B\boldsymbol{\tau}_T$, dem Reaktionsmoment ${}^B\boldsymbol{\tau}_D$ und dem externen Moment ${}^B\boldsymbol{\tau}_{\text{ext}}$. Das MomentŚ ${}^B\boldsymbol{\tau}_T$ kann wie folgt bestimmen werden:

$${}^B\boldsymbol{\tau}_T = \sum_{i=1}^{6} {}^B\boldsymbol{\tau}_{T_i} = \sum_{i=1}^{6} {}^B\boldsymbol{r}_i \times {}^B\boldsymbol{f}_{T,i} \tag{8}$$

mit

$${}^B\boldsymbol{r}_i = \begin{bmatrix} l\cos\gamma_i & l\sin\gamma_i & \Delta h \end{bmatrix}^\top \quad , \tag{9}$$

wobei Δh die Höhendifferenz zwischen dem Ursprung von $\{\mathrm{R}_i\}$ und {B} bezeichnet.
Das in (4) vorgestellte Reaktionsmoment wird mit

$${}^B\boldsymbol{\tau}_D = \sum_{i=1}^{6} (-1)^i\, b\, \omega_i^2\, {}^B\boldsymbol{e}_{R_i} \tag{10}$$

bestimmt. Der Term $(-1)^i$ bezeichnet die Richtung des Moments.

2.3 Vorwärtskinematik des Manipulators

Aufgabe der Vorwärtskinematik ist es, anhand der Gelenkwinkel $\boldsymbol{\theta}^* \in \mathbb{R}^6$ die Pose des Endeffektor-System {T} relativ zu $\{\mathrm{L}_0\}$ zu bestimmen. Da die Referenzsysteme nach der Denavit-Hartenberg Konvention definiert wurden, können die benötigten Transformationsmatrizen über die D-H Parameter bestimmt werden. Dabei ist die Transformation zwischen zwei konsekutiven Koordinatensystemen wie folgt definiert:

$${}^{i-1}_{i}\boldsymbol{T} = \begin{bmatrix} c\theta_i & -s\theta_i c\alpha_i & s\theta_i s\alpha_i & a_i c\theta_i \\ s\theta_i & c\theta_i c\alpha_i & -c\theta_i s\alpha_i & a_i s\theta_i \\ 0 & s\alpha_i & c\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

mit $sx \coloneqq \sin x$, $cx \coloneqq \cos x$.
Die D-H Parameter für den Manipulator sind in **Tabelle 1** aufgeführt.

Tabelle 1 D-H Parameter für den Manipulator

Link	θ_i [rad]	d_i [m]	α_i [rad]	a_i [m]
1	$\theta_1^* - \frac{\pi}{2}$	0	$-\frac{\pi}{2}$	0
2	θ_2^*	0	0	0.245
3	$\theta_3^* + \frac{\pi}{2}$	0.0642	$-\frac{\pi}{2}$	0
4	$\theta_4^* - \frac{\pi}{2}$	0.3007	$\frac{\pi}{2}$	0
5	θ_5^*	0	$-\frac{\pi}{2}$	0
6	θ_6^*	0.1232	0	0

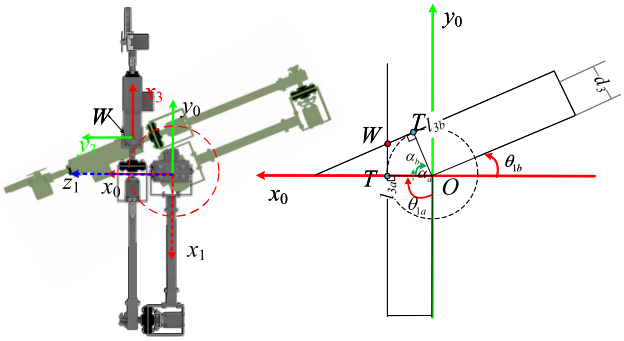


Abbildung 3 Draufsicht des Manipulators in zwei Konfigurationen zur Berechnung von Gelenk 1

Um die gewünschte Pose des Endeffektors, dargestellt im Arm-Basis-System $\{\mathrm{L}_0\}$, zu erhalten, nutzt man die folgende Kettenregel:

$${}^0_T\boldsymbol{T} = {}^0_1\boldsymbol{T}\, {}^1_2\boldsymbol{T} \cdots {}^5_6\boldsymbol{T}\, {}^6_T\boldsymbol{T} \quad . \tag{12}$$

2.4 Inverse Kinematik des Manipulators

Das Ziel der inversen Kinematik ist es, aus einer gegebenen Pose des Endeffektor-Systems die dazugehörigen Gelenkwinkel $\boldsymbol{\theta}^*$ zu ermitteln.
Die Handgelenkposition kann durch

$${}^0\boldsymbol{p}_{0W} = {}^0\boldsymbol{p}_{06} - {}^0\boldsymbol{p}_{36} \tag{13}$$

berechnet werden, wobei ${}^0\boldsymbol{p}_{jk}$ den Vektor vom Ursprung des $\{\mathrm{L}_j\}$-Systems zu dem des $\{\mathrm{L}_j\}$-Systems darstellt. ${}^0\boldsymbol{p}_{36}$ kann durch folgenden Zusammenhang bestimmt werden.

$${}^0\boldsymbol{p}_{36} = d_6 - {}^0\boldsymbol{z}_6, \quad \text{mit} \quad {}^0\boldsymbol{z}_6 = {}^0_6\boldsymbol{R}\, {}^6\boldsymbol{z}_6 \quad . \tag{14}$$

${}^0\boldsymbol{p}_{06}$ und ${}^0_6\boldsymbol{R}$ können aus der Transformationsmatrix ${}^0_6\boldsymbol{T}$ entnommen werden.
Für eine gegebene Handgelenkposition ${}^0\boldsymbol{p}_{0W}$ existieren zwei Lösungen für θ_1 (siehe **Bild 3** links):

$$\theta_{1,a}^* = \operatorname{atan2}(y_{Ta}, x_{Ta}) \tag{15}$$

$$\theta_{1,b}^* = \operatorname{atan2}(y_{Tb}, x_{Tb}) \quad . \tag{16}$$

Dabei liegen (y_{Ta}, x_{Ta}) und (y_{Tb}, x_{Tb}) in der x_0-y_0-Ebene und beschreiben die Koordinaten der beiden Tangentialpunkte T, an denen der Unterarm des Manipulators den Kreis um den Ursprung von $\{\mathrm{L}_0\}$ mit dem Radius d_3 berührt. Dies ist in der Abbildung rechts veranschaulicht. Die

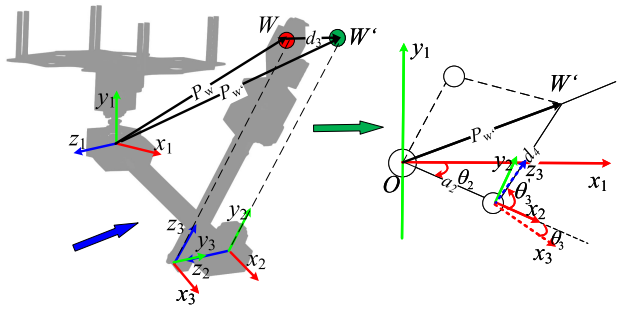


Abbildung 4 Seitenansicht des Manipulators zur Berechnung der Gelenke 2 und 3

beiden Punkte werden mit den Gleichungen

$$\begin{bmatrix} x_T & y_T \end{bmatrix} \cdot \begin{bmatrix} x_T - x_W & y_T - y_W \end{bmatrix}^\top = 0 \tag{17}$$

$$x_T^2 + y_T^2 = d_3^2 \tag{18}$$

bestimmt. Eine zusätzliche Bedingung für die Existenz einer Lösung ist $\sqrt{x_W^2 + y_W^2} \geq d_3$.
Für Gelenke 2 und 3 existieren wiederum 2 mögliche Konfigurationen (siehe **Bild 4**): Einmal den Ellenbogen nach oben und einmal nach unten. Es gibt somit 4 mögliche Gelenkwinkelkombinationen der ersten drei Gelenke, um eine bestimmte Handgelenkposition anzufahren. Als Ergebnis erhält man für das zweite und dritte Gelenk des Manipulators folgende Werte:

$$\theta_3^* = \arccos\left(\frac{x_W^2 + y_w^2 + z_W^2 - a_2^2 - a_3^2}{2a_2 d_4}\right) - \pi \tag{19}$$

$$\theta_2^* = \operatorname{atan2}(s_2, c_2) \tag{20}$$

mit

$$s_2 = \frac{(a_2 + d_4 c_3)\, z_W - d_4 s_3 \sqrt{x_W^2 + y_W^2}}{x_W^2 + y_W^2 + z_W^2} \tag{21}$$

$$c_2 = \frac{(a_2 + d_4 c_3) \sqrt{x_W^2 + y_W^2} + d_4 s_3 z_W}{x_W^2 + y_W^2 + z_W^2} \quad . \tag{22}$$

Bildet man die Rotationsmatrix ${}^3_6\boldsymbol{R}$ einerseits entsprechend des bereits gegebenen Zusammenhangs, andererseits aus den D-H Parametern der letzten drei Gelenke, so erhält man die folgende Gleichung zur Bestimmung von $\theta^*_{4,5,6}$:

$${}^0_3\boldsymbol{R}^\top(\theta_1^*, \theta_2^*, \theta_3^*) \cdot {}^0_6\boldsymbol{R} = {}^3_6\boldsymbol{R}(\theta_4^*, \theta_5^*, \theta_6^*) \quad . \tag{23}$$

3 Reglerentwurf

3.1 Lageregelung der Drohne

Da die unbekannten externen Kräfte f_{ext} und τ_{ext} als Störgrößen betrachtet werden, können diese für den Reglerentwurf ignoriert werden. Die Bewegungsgleichungen (5) und (7) werden zu

$$\begin{bmatrix} \ddot{\boldsymbol{p}} \\ {}^B\dot{\boldsymbol{\omega}} \end{bmatrix} = \underbrace{\begin{bmatrix} -g\boldsymbol{e}_3 \\ -\boldsymbol{I}_B^{-1}\left({}^B\boldsymbol{\omega} \times \boldsymbol{I}_B\, {}^B\boldsymbol{\omega}\right) \end{bmatrix}}_{\boldsymbol{h}} + \underbrace{\begin{bmatrix} \frac{1}{m}\,{}^I_B\boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_B^{-1} \end{bmatrix}}_{\boldsymbol{J}} \boldsymbol{u} \tag{24}$$

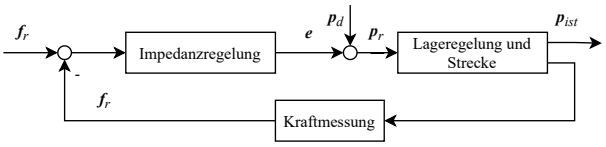


Abbildung 5 Regelkreis der Impedanzregelung auf Grundlage der Lageregelung

umgeschrieben. $\boldsymbol{J}$ ist dabei eine diagonale, invertierbare Matrix. Mit der Einführung des virtuellen Eingangs $\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_T^\top & \boldsymbol{v}_R^\top \end{bmatrix}^\top$, wird (24) mit dem Regelgesetz

$$\boldsymbol{u} = \boldsymbol{J}^{-1}\left[-\boldsymbol{h} + \boldsymbol{v}\right] \tag{25}$$

linearisiert, sodass $\begin{bmatrix} \ddot{\boldsymbol{p}}^\top & {}^B\dot{\boldsymbol{\omega}}^\top \end{bmatrix}^\top = \boldsymbol{v}$ gilt.
Um die gewünschte Trajektorie $\{\boldsymbol{p}_d, \dot{\boldsymbol{p}}_d, \ddot{\boldsymbol{p}}_d, \boldsymbol{R}_d, {}^B\boldsymbol{\omega}_d, {}^B\dot{\boldsymbol{\omega}}_d\}$ verfolgen zu können, werden folgende Fehlergrößen eingeführt:

$$\tilde{\boldsymbol{e}}_P = \boldsymbol{p}_d - \boldsymbol{p} \tag{26}$$

$$\tilde{\boldsymbol{e}}_V = \dot{\boldsymbol{p}}_d - \dot{\boldsymbol{p}} \tag{27}$$

$$\tilde{\boldsymbol{e}}_R = \frac{1}{2}\left[{}^I_B\boldsymbol{R}^\top \boldsymbol{R}_d - \boldsymbol{R}_d^\top\, {}^I_B\boldsymbol{R}\right]_\vee \tag{28}$$

$$\tilde{\boldsymbol{e}}_\Omega = {}^I_B\boldsymbol{R}^\top \boldsymbol{R}_d\, {}^B\boldsymbol{\omega}_d - {}^B\boldsymbol{\omega} \tag{29}$$

Hier bezeichnet $[\cdot]_\vee$ die inverse Operation des Operators $[\cdot]_\times$ der schiefsymmetrischen Matrix. Anhand dieser Definitionen kann der virtuelle Eingang $\boldsymbol{v}$ gemäß einem PID-Regler mit

$$\boldsymbol{v}_T = \ddot{\boldsymbol{p}}_d + \boldsymbol{K}_{T,D}\dot{\tilde{\boldsymbol{e}}}_P + \boldsymbol{K}_{T,P}\tilde{\boldsymbol{e}}_P + \boldsymbol{K}_{T,I}\int_0^t \tilde{\boldsymbol{e}}_P \,\mathrm{d}t \tag{30}$$

$$\boldsymbol{v}_R = {}^B\dot{\boldsymbol{\omega}}_d + \boldsymbol{K}_{R,D}\tilde{\boldsymbol{e}}_\Omega + \boldsymbol{K}_{R,P}\tilde{\boldsymbol{e}}_R + \boldsymbol{K}_{R,I}\int_0^t \tilde{\boldsymbol{e}}_R \,\mathrm{d}t \tag{31}$$

aufgestellt werden. Die Verstärkungsmatrizen $\boldsymbol{K}_{*,*}$ sind positiv definite Diagonalmatrizen.

3.2 Impedanzregelung des Manipulators

Der Impedanzregler soll zu der bestehenden Lageregelung hinzugefügt werden. Die Führungsgröße des Systems ist dabei die gewünschte, auf den Endeffektor wirksamen Kraft. Anhand der tatsächlich wirksamen Kraft wird daraus, über die Kraftabweichung, eine Posenabweichung bestimmt, durch den die gewünschte Endeffektor-Pose angepasst werden kann. Der schematische Regelkreis für diese Impedanzregelung kann dabei durch die allgemeine Darstellung gemäß **Bild 5** veranschaulicht werden.
Die Kräftebilanz und daraus die mechanische Impedanz eines Feder-Masse-Dämpfer-Systems können mit

$$f_E(t) = m\ddot{x}(t) + D\dot{x} + cx(t) \tag{32}$$

$$\frac{F_E}{X(s)} = ms^2 + Ds + c \tag{33}$$

aufgestellt werden. Die gewünschte Trajektorie des Endeffektors liegt dabei in der Form $\begin{bmatrix} {}^0\boldsymbol{p}_d^\top & {}^0_T\boldsymbol{R}_d^\top & {}^0\boldsymbol{v}_d^\top & {}^0\boldsymbol{\omega}_d^\top & {}^0\dot{\boldsymbol{v}}_d^\top & {}^0\dot{\boldsymbol{\omega}}_d^\top \end{bmatrix}^\top$ vor. Die resultierende Trajektorie wird durch

$\begin{bmatrix} {}^0\boldsymbol{p}_r^\top & {}^0_T\boldsymbol{R}_r^\top & {}^0\boldsymbol{v}_r^\top & {}^0\boldsymbol{\omega}_r^\top & {}^0\dot{\boldsymbol{v}}_r^\top & {}^0\dot{\boldsymbol{\omega}}_r^\top \end{bmatrix}^\top$ definiert.
Die Fehlerdynamik kann mit

$$\boldsymbol{M}_T\Delta\dot{\boldsymbol{v}} + \boldsymbol{D}_T\Delta\boldsymbol{v} + \boldsymbol{K}_T\Delta\boldsymbol{p} = \Delta\boldsymbol{f} \tag{34}$$

$$\boldsymbol{I}_R\Delta\dot{\boldsymbol{\omega}} + \boldsymbol{D}_R\Delta\boldsymbol{\omega} + \boldsymbol{K}_R\Delta\boldsymbol{\phi} = \Delta\boldsymbol{\tau} \tag{35}$$

für den Manipulator bestimmt werden. Die Trägheitsmatrix $\mathrm{diag}\{\boldsymbol{M}_T, \boldsymbol{I}_R\} \in \mathbb{R}^{6\times6}$, die Dämpfung $\mathrm{diag}\{\boldsymbol{D}_T, \boldsymbol{D}_R\} \in \mathbb{R}^{6\times6}$ sowie die Steifigkeit $\mathrm{diag}\{\boldsymbol{K}_T, \boldsymbol{K}_R\} \in \mathbb{R}^{6\times6}$ sind positiv definite Diagonalmatrizen, deren Verstärkungsfaktoren entsprechend der gewünschten Anforderungen an das Systemverhalten gewählt werden. $\Delta\boldsymbol{\phi}$ ist der Orientierungsfehler, parametriert durch die Euler-Winkel.
Mit den Vektoren $\boldsymbol{x} = \begin{bmatrix}\Delta\boldsymbol{p}^\top & \Delta\boldsymbol{\phi}^\top & \Delta\boldsymbol{v}^\top & \Delta\boldsymbol{\omega}^\top\end{bmatrix}^\top$ und $\boldsymbol{\xi} = \begin{bmatrix}\Delta\boldsymbol{f}^\top & \Delta\boldsymbol{\tau}^\top\end{bmatrix}^\top$ werden Differenzialgleichungen (34) und (35) in Gleichungen erster Ordnung umgewandelt:

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{\xi} \tag{36}$$

mit

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T}^{-1} \\ -\boldsymbol{K}_T\boldsymbol{M}_T^{-1} & \boldsymbol{0} & -\boldsymbol{D}_T\boldsymbol{M}_T^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{K}_R\boldsymbol{I}_R^{-1} & \boldsymbol{0} & -\boldsymbol{D}_R\boldsymbol{I}_R^{-1} \end{bmatrix} \tag{37}$$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{M}_T^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_R^{-1} \end{bmatrix} . \tag{38}$$

Die Matrix $\boldsymbol{T}$ beschreibt die Beziehung zwischen der Zeitableitung der Euler-Winkel und der Winkelgeschwindigkeit:

$$\boldsymbol{\omega} = \begin{bmatrix} 1 & 0 & -\sin\vartheta \\ 0 & \cos\varphi & \cos\vartheta\sin\varphi \\ 0 & -\sin\varphi & \cos\vartheta\cos\varphi \end{bmatrix} \dot{\boldsymbol{\phi}} = \boldsymbol{T}(\boldsymbol{\phi})\,\dot{\boldsymbol{\phi}} \ . \tag{39}$$

Mit einer frei wählbaren Schrittweite h kann nun das System (36) numerisch über beispielsweise das explizite Euler-Verfahren integriert werden.
Durch Bestimmen der entsprechenden Abweichung $\boldsymbol{x}$ lässt sich abschließend anhand der gewünschten Zieltrajektorie die neue Referenztrajektorie für den Endeffektor bestimmen. Diese kann an die bestehende Lageregelung des Manipulators übergeben werden. Als Lageregelung wird die Gelenkregelung verwendet, weshalb die Pose mit der inversen Kinematik vorverarbeitet werden muss.

4 Simulation

Die vom Regler realisierte virtuelle Impedanz entspricht einem Feder-Masse-Dämpfer-System, wobei die Regelgüte in diesem Fall durch Vergleich der erwarteten und der realen Systemantwort bewertet werden kann. Die Startpose wird mit ${}^0\boldsymbol{p}_d = \begin{bmatrix}0\,\mathrm{m} & 0.3\,\mathrm{m} & 0.3\,\mathrm{m} & 0° & 0° & 0°\end{bmatrix}$ festgelegt. In **Tabelle 2** sind die Messbereiche sowie die Messungenauigkeiten von den simulierten F/T-Messungen dargestellt. Wegen Platzmangels werden die Ergebnisse bzgl. der translatorischen Bewegungen in z-Richtung dargestellt. Damit der Reglerausgang einer gedämpften Schwingung entspricht, die jedoch ein signifikantes Schwingen aufweist, werden die Regelparameter auf $\boldsymbol{M}_T = 10\boldsymbol{I}_3$, $\boldsymbol{D}_T = 5\boldsymbol{I}_3$ und $\boldsymbol{K}_T = 80\boldsymbol{I}_3$ gesetzt.

Tabelle 2 Messbereiche und Ungenauigkeiten der F/T-Messungen

	f_x	f_y	f_z	$\tau_{x,z}$	τ_y
Messbereich (±)	145 N	145 N	290 N	5 Nm	5 Nm
Ungenauigkeit	1.25 %	1.0 %	0.75 %	1.25 %	1.5 %

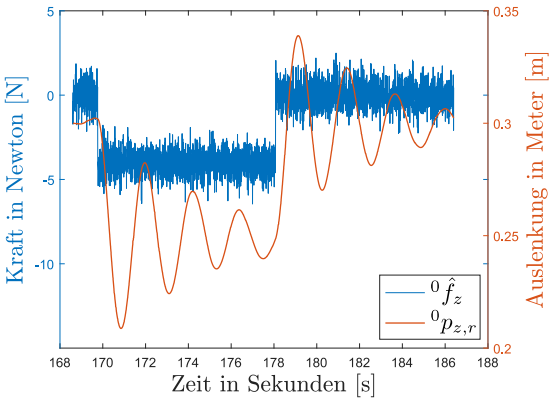


Abbildung 6 Verläufe von z-Komponente der resultierenden Endeffektor-Position bei Kraftsprung

Bei dem Versuch wird erst eine Kraft von $f_{z,\mathrm{stör}} = 4\,\mathrm{N}$ entlang der z-Achse des Inertialsystems aufgebracht und diese nach kurzer Zeit wieder entfernt, um eine Schwingung auszulösen. Die z-Komponente der resultierenden Endeffektor-Position wird im **Bild 6** aufgezeichnet.
Durch Umformen der Differenzialgleichung (34) und Nullsetzen der Störgrößen, kann das System in die allgemeine Form eines linearen gedämpften Oszillators überführt werden. Das erzeugte Feder-Masse-Dämpfer-System kann dabei durch die Eigenfrequenz ω_0 und den Dämpfungsgrad d_0 charakterisiert werden. Die Untersuchungen folgen dabei den Ausführungen zur Sensitivitätsanalyse in [11]. Die gewünschte Eigenfrequenz und der Dämpfungsgrad können den eingestellten Regelparametern entsprechend für die Regelung der Position durch

$$\omega_{0,d} = \sqrt{\frac{k_T}{m_T}} = 2.8284\,\mathrm{Hz} \tag{40}$$

$$d_{0,d} = \frac{d_T\,\omega_{0,d}}{2k_T} = \frac{d_T}{2m_T\,\omega_{0,d}} = 0.0884 \tag{41}$$

und analog dazu für die Orientierung bestimmt werden. Anhand des vom Regler bestimmten resultierenden Posenverlaufs kann eine nichtlineare Fit-Funktion der Form

$$x(t) = \hat{x}\cdot e^{-\delta\cdot t}\cdot\sin(\omega\cdot t + \varphi_0) + x_\infty \ . \tag{42}$$

bestimmt werden, die einer gedämpften Schwingung entspricht (siehe **Bild 7**). Die daraus resultierende, geschätzte Eigenfrequenz und der Dämpfungsgrad lassen sich durch

$$\omega_{0,e} = \sqrt{\omega^2 + \delta^2} = 2.7975\,\mathrm{Hz} \tag{43}$$

$$d_{0,e} = \delta/\omega_{0,e} = 0.082 \ . \tag{44}$$

bestimmen. Daraus kann der relative Fehler zwischen den

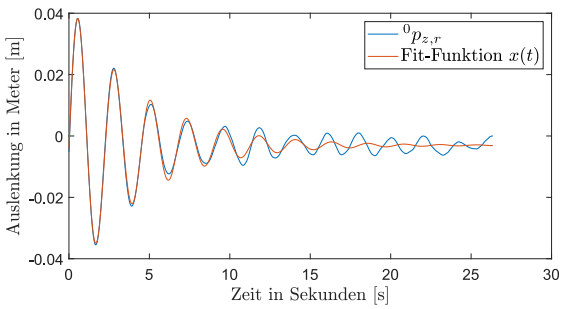


Abbildung 7 Verläufe von z-Komponente der resultierenden Endeffektor-Position bei Kraftsprung

gewünschten und resultierenden Werten durch folgende Formeln bestimmt werden:

$$e_{\omega_0} = \left(\frac{\omega_{0,d} - \omega_{0,e}}{\omega_{0,d}} \right) \times 100\% = 1.093\,\% \tag{45}$$

$$e_{d_0} = \left(\frac{d_{0,d} - d_{0,e}}{d_{0,d}} \right) \times 100\% = 7.228\,\% \ . \tag{46}$$

Das Simulationsexperiment bestätigt das gewünschte Antwortverhalten einer gedämpften Schwingung. Diese ist aufgrund von Messunsicherheiten jedoch merkbar verrauscht, wodurch relative Fehler der Eigenfrequenz und des Dämpfungsgrads entstehen.

5 Zusammenfassung

Im Rahmen dieser Arbeit wurden die Impedanzregelung und Lageregelung für ein Luftmanipulationssystem entwickelt. Der Manipulator ist dabei an einem vollaktuierten Hexacopter befestigt. Die Regelung der Drohne erfolgt durch exakte Linearisierung des Systems. Der Impedanzregler passt die erzeugte Trajektorie für den Endeffektor in Abhängigkeit der wirksamen Kräfte und Momente an. Als Impedanz wurde ein Feder-Masse-Dämpfer-System modelliert, wodurch die Verläufe gedämpften Schwingungen entsprechen. Die Regelgüte der Impedanzregelung wurde durch Untersuchungen der Eigenfrequenzen bestimmt. Aufgrund des Messrauschens entstehen erwartete Abweichungen zur modellierten Impedanz.
In Ausblick auf künftige Arbeiten sollte die Entkoppelung für den Arm und die Drohne implementiert werden. Die Impedanzregelung des Gesamtsystems soll als alternatives Konzept untersucht werden.

Literatur

[1] K. Baizid, G. Giglio, F. Pierri, M. A. Trujillo, G. Antonelli, F. Caccavale, A. Viguria, S. Chiaverini, and A. Ollero. Behavioral control of unmanned aerial vehicle manipulator systems. *Autonomous Robots*, 41 (5):1203–1220, 2017.

[2] E. Cataldi, G. Muscio, M. A. Trujillo, Y. Rodriguez, F. Pierri, G. Antonelli, F. Caccavale, A. Viguria, S. Chiaverini, and A. Ollero. Impedance control of an aerial-manipulator: Preliminary results. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3848–3853, 2016.

[3] G. Gioioso, M. Ryll, D. Prattichizzo, H. H. Bülthoff, and A. Franchi. Turning a near-hovering controlled quadrotor into a 3d force effector. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6278–6284, 2014. doi: 10.1109/ICRA.2014.6907785.

[4] M. Kamel, S. Comari, and R. Siegwart. Full-body multi-objective controller for aerial manipulation. In *2016 24th Mediterranean Conference on Control and Automation (MED)*, pages 659–664, 2016. doi: 10.1109/MED.2016.7536005.

[5] E. Kaufman, K. Caldwell, D. Lee, and T. Lee. Design and development of a free-floating hexrotor uav for 6-dof maneuvers. In *2014 IEEE Aerospace Conference*, pages 1–10. IEEE, 2014.

[6] A. Ollero, G. Heredia, A. Franchi, G. Antonelli, K. Kondak, A. Sanfeliu, A. Viguria, J. R. Martinez-de Dios, F. Pierri, J. Cortes, A. Santamaria-Navarro, M. A. Trujillo Soto, R. Balachandran, J. Andrade-Cetto, and A. Rodriguez. The aeroarms project: Aerial robots with advanced manipulation capabilities for inspection and maintenance. *IEEE Robotics Automation Magazine*, 25(4):12–23, 2018. doi: 10.1109/MRA.2018.2852789.

[7] S. Rajappa, M. Ryll, H. H. Bülthoff, and A. Franchi. Modeling, control and design optimization for a fully-actuated hexarotor aerial vehicle with tilted propellers. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4006–4013. IEEE, 2015.

[8] R. Rashad, P. Kuipers, J. Engelen, and S. Stramigioli. Design, modeling, and geometric control on se (3) of a fully-actuated hexarotor for aerial interaction. *ArXiv e-prints*, 2017.

[9] M. Ryll, H. H. Bülthoff, and P. R. Giordano. A novel overactuated quadrotor uav. *IEEE Transactions on Control Systems Technology*, 2015.

[10] M. Tognon, B. Yüksel, G. Buondonno, and A. Franchi. Dynamic decentralized control for protocentric aerial manipulators. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6375–6380, 2017.

[11] C. Yao, M. Schuster, Z. Jiang, K. Janschek, and M. Beitelschmidt. Sensitivity analysis of model-based impedance control for physically interactive hexarotor. *IFAC-PapersOnLine*, 52(15):597 – 602, 2019. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2019.11.741. 8th IFAC Symposium on Mechatronic Systems MECHATRONICS 2019.

Teach-In für das kraftsensitive automatische Wickeln eines Seiles im dreidimensionalen Raum

Teach-In for Force Sensitive Automatic Rope Winding in Three Dimensions

Michael Schwegel*,[1], Hanspeter Augschöll[1], Christian Hartl-Nesic[1], Andreas Kugi[1,2],
1 TU Wien, Institut für Automatisierungs- und Regelungstechnik, 1040 Wien, Österreich, michael.schwegel@tuwien.ac.at,
2 Center for Vision, Automation & Control, Austrian Institute of Technology, 1210 Wien, Österreich

Kurzfassung

Seile und Kabel stellen durch ihre formlabile Struktur für die automatisierte industrielle Handhabung eine große Herausforderung dar. In dieser Arbeit wird ein Demonstrations- und ein Roboterwerkzeug für das Teach-In eines Wickelprozesses im dreidimensionalen Raum vorgestellt. Eine angepasste Pfadfolgeregelung ermöglicht hierbei eine optimale Nutzung des erreichbaren Arbeitsraums eines Roboters zur automatischen Wiedergabe der eingelernten Prozesse. Ein praktischer Versuch zeigt, wie dadurch eine komplexe Seilwickelaufgabe mit einer definierten Zugspannung im Seil auf intuitive Weise eingelernt und automatisiert werden kann.

Abstract

Ropes and cables pose a great challenge for the automated industrial handling due to their flexible nature. In this work, a demonstration tool and a robot tool are presented for the teach-in of a rope winding process in the three-dimensional space. A tailored path-following controller allows for an optimal use of the reachable working space of a robot to automatically reproduce the demonstrated processes. A practical experiment shows how a complex rope winding task with a defined tension in the rope can be taught and automated in an intuitive way.

1 Einleitung

Die automatisierte Handhabung von Seilen, Kabeln und Klebebändern in der Industrie basiert weitestgehend auf spezialisierten maschinellen Lösungen. Die flexible Handhabung von Verkabelungen und Klebeprozessen weist für die Automatisierung eine Reihe von Herausforderungen auf. Neben der Vorgabe eines geeigneten Pfades ist auch die formlabile Natur des Materials und die Zugspannung im Material entscheidend. Applikationen von Stoffen und Bändern auf 3D-Oberflächen wurden bereits untersucht, siehe z.B. [1–3]. In dieser Arbeit wird mit Hilfe eines 7-Achs-Knickarm-Roboters ein Seil im dreidimensionalen Raum auf- bzw. abgewickelt und über unterschiedlich konfigurierte Abrollpunkte gespannt.

Die Offline-Programmierung einer solchen Aufgabe stellt durch die möglichen Kollisionen des Roboters mit dem elastischen Seil eine Herausforderung dar. Die Programmierung durch Teach-In ist für die vorliegende Aufgabe jedoch besonders geeignet, da die Beschränkungen durch das Seil direkt während der Demonstration sichtbar werden und sofort berücksichtigt werden können, siehe z.B. [4]. Des Weiteren kann auch die nötige Zugspannung im Seil im jeweiligen Bereich des Pfades direkt bei der Demonstration vorgegeben werden.

Da das Seil beim kinästhetischen Lernen eine große Einschränkung darstellen würde, soll die Demonstration mit einem geeigneten mechatronischen Werkzeug erfolgen, vergleiche auch [5]. Dieses Werkzeug wird nicht nur zur Bestimmung der Position des Pfades zur Abwicklung des Seils benötigt, sondern es müssen auch die Orientierung der Spule und die Zugspannung im Seil gemessen werden, um den Prozess hinreichend genau abbilden zu können.

In diesem Beitrag wird ein Teach-In Werkzeug präsentiert, das neben der Positions- und Orientierungsinformation auch die Kraftinformation erfasst. Mit dem entworfenen Werkzeug wird das Wickeln eines Seils um im dreidimensionalen Raum angeordnete Stangen demonstriert, siehe **Abbildung 1**. Hierbei muss eine definierte Zugspannung im Seil realisiert und aufgezeichnet werden. Eine am Flansch des Roboters montierte motorisierte Seilspule ermöglicht die Wiedergabe des eingelernten Prozesses. Experimente zeigen die Funktionsweise des vorgestellten Konzepts.

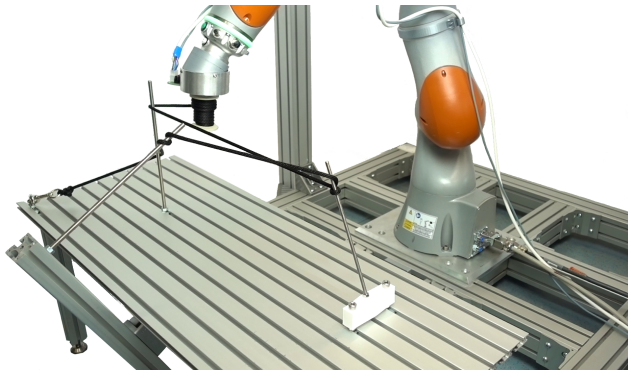


Abbildung 1 Experimenteller Aufbau.

Diese Arbeit ist wie folgt gegliedert: in Abschnitt 2 wird

die mechatronische Ausführung des Programmier- und des Roboterwerkzeugs zum automatisierten Seilwickeln beschrieben. Darauf folgend wird in Abschnitt 3 die Regelung des Knickarmroboters beschrieben. Abschnitt 4 enhält die experimentellen Ergebnisse eines beispielhaften Seilwickelprozesses. Die Erkenntnisse dieser Arbeit werden in Abschnitt 5 zusammengefasst.

2 Mechatronische Ausführung

Die Demonstration der Aufgabe basierend auf einem instrumentierten Werkzeug bietet im vorliegenden Szenario mehrere Vorteile gegenüber dem kinästhetischen Lernen. Wenn das Demonstrationswerkzeug zum Teach-In vom Prinzip her gleich wie das Roboterwerkzeug aufgebaut ist, dann vereinfacht sich der Transfer vom Teach-In zur automatisierten Ausführung, unter anderem dadurch, dass gleiche Signale (zum Beispiel Pose und Kraft in einer bestimmten Richtung) aufgezeichnet werden können.
In den folgenden beiden Abschnitten wird die Kontruktion des Demonstrations- und des Roboterwerkzeugs näher erläutert.

2.1 Demonstrationswerkzeug

Um die Demonstration des Seilwickelprozesses intuitiv zu gestalten, wurde ein Demonstrationswerkzeug entwickelt, das eine einhändige Handhabung des Seils ermöglicht. Eine schematische Darstellung des Werkzeugs in Explosionsansicht ist in **Abbildung 2** dargestellt. Eine Bremsvorrichtung zum Einprägen der gewünschten Zugspannung im Seil wurde mit Hilfe einer Bremsscheibe realisiert. Diese kann durch Kraftaufbringung auf den Bremshebel durch die Messwelle des Drehmomentsensors eine Zugkraft auf das Seil einprägen.

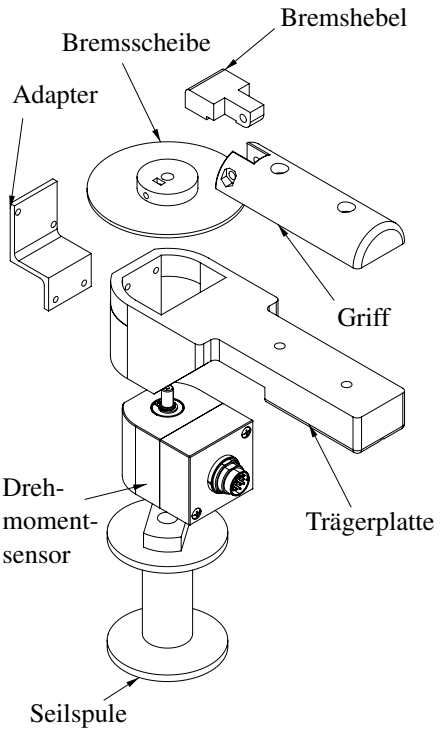


Abbildung 2 Explosionsdarstellung des Demonstrationstools.

Durch die Verwendung eines einachsigen Drehmomentsensors werden die Anteile der Zugkraft des Seils entlang der Seilspulenachse nicht erfasst. Zur Messung aller Interaktionskräfte könnte das Werkzeug mit einem 6-Achsen Kraft-Drehmomentsensor ausgestattet werden. In der praktischen Anwendung hat sich jedoch gezeigt, dass durch die Winkelfehler bei der Messung der Seilzugspannung nur kleine Anteile der Kraft vernachlässigt werden. Des Weiteren kann durch eine orthogonale Führung der Spulenachse zum Seil der Messfehler minimiert werden. Diese Orientierung des Werkzeugs verhindert zusätzlich ein Losrollen des Seils von der Spule.
Um einen Seilwickelvorgang aufzunehmen, muss neben der Seilzugspannung auch die Pose des Demonstrationswerkzeugs gemessen werden. Zur Bestimmung der Pose (Position und Orientierung) der Seilspule wurde das optische Messsystem *OptiTrack* verwendet. Hierzu werden Infrarotmarker auf einem Adapter des Demonstrationswerkzeugs angebracht, durch die die Pose der Spule im dreidimensionalen Raum bestimmt werden kann.

2.2 Roboterwerkzeug

Aufgabe des Roboterwerkzeugs ist es, einen Seilwickelvorgang automatisiert reproduzieren zu können. Um die Seilzugspannung anhand der Motorströme einstellen zu können, wurde für die Spule ein Direktantriebskonzept eingesetzt. Dieses minimiert Reibungsverluste im Antrieb und ermöglicht eine niedrige Bauhöhe des Werkzeugs. Dadurch wird die Belastung der Motorwelle und der konstruktive Aufwand am Endeffektor zur Lagerung und Führung der Seilspule minimiert.
Bringt man die Seilspulenachse nun parallel zur letzten Drehachse des Roboters an, so erhält man eine endlos drehbare Achse, die jedoch redundant zur letzten Drehachse des seriellen Roboters ist. Wird jedoch die Spulenachse gegen die letzte Drehachse des Roboters gekippt, so ergibt sich ein zusätzlicher aktuierter Freiheitsgrad. Das am Roboterflansch montierte Roboterwerkzeug ist in **Abbildung 3** dargestellt.
Im folgenden Abschnitt wird gezeigt, wie dieser zusätzliche Freiheitsgrad genützt werden kann, um auch schwierige Posen im Arbeitsraum erreichen zu können und bietet damit eine deutliche Erleichterung der Teach-In Aufgabe. Zusätzlich kann durch die geeignete Wahl des Kippwinkels eine bessere Erreichbarkeit der zu erwartenden Konfigurationen sichergestellt werden. Dies verhindert das Durchfahren von singulären Roboterkonfigurationen. Durch den Einsatz schneller Fertigungsverfahren für Prototypen, wie dem 3D-Druck, kann die Neigung des Endeffektors flexibel für die jeweilige Anwendung angepasst werden.

3 Roboterregelung

Um den Roboter entlang des beim Teach-In aufgezeichneten Pfades zu führen, wird die Zeitinformation verworfen und nur der geometrische Pfad herangezogen. Dieser Pfad der Position und Orientierung der Seilspule wird um die

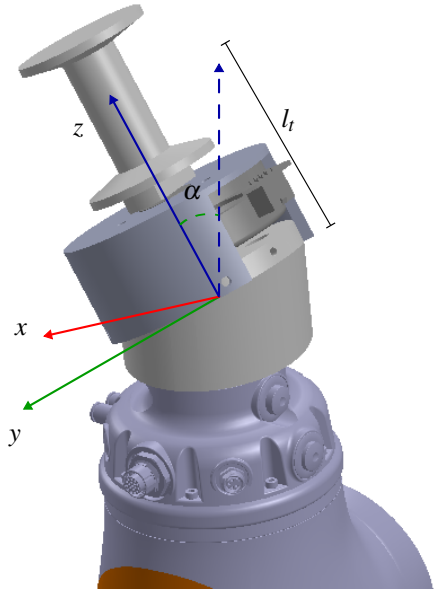


Abbildung 3 Roboterwerkzeug zur automatischen Seilwicklung mit Seitneigung zur Nutzung der redundanten Freiheitsgrade.

zurgehörige Seilzugspannung ergänzt. Da sich die Seilzugspannung präzise und schnell durch den Spulenmotor regeln lässt, kann diese Aufgabe als unterlagerte Regelung angesehen und getrennt von der Robterregelung behandelt werden. Diese unterlagerte Regelung der Seilzugspannung wurde direkt durch eine Drehmomentenregelung des Spulenmotors realisiert.
In diesem Abschnitt wird eine Pfadfolgeregelung hergeleitet, mit deren Hilfe der durch einen Teach-In Prozess gewonnene Seilpfad automatisiert gewickelt werden kann. Durch die Reduktion des Ausgangsraums kann der bearbeitbare Raum des Roboters erweitert werden. Dazu wird zuerst das verwendete Robotermodell vorgestellt und dann die nötigen Terme für die Regelung hergeleitet. Zuletzt wird die Reduktion des Arbeitsraums durchgeführt und die Nullraumregelung erläutert.

3.1 Robotermodell und Pfaddefinition

Im Folgenden wird der kollaborative 7-Achs-Roboter *Kuka LBR iiwa 14 R820* mit dem Starrkörpermodell

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{n}(\mathbf{q},\dot{\mathbf{q}}) = \tau + \tau_{ext}, \tag{1}$$

betrachtet, wobei $\mathbf{q}$ die generalisierten Koordinaten des Roboters, τ die Motordrehmomente, τ_{ext} externe Drehmomente und

$$\mathbf{n}(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{C}(\mathbf{q},\dot{\mathbf{q}})\,\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) \tag{2}$$

beschreiben. Die Position $\mathbf{y}_t$ und die Orientierung $\mathbf{y}_r$ des Werkzeugs

$$\mathbf{y} = \begin{bmatrix}\mathbf{y}_t \\ \mathbf{y}_r\end{bmatrix} = \begin{bmatrix}\mathbf{h}_t(\mathbf{q}) \\ \mathbf{h}_r(\mathbf{q})\end{bmatrix} = \mathbf{h}(\mathbf{q}) \tag{3}$$

sollen nun einem Pfad Γ, parametriert durch $\sigma(\vartheta)$, folgen. Der Pfadparameter ϑ liegt im Intervall $\mathscr{T}$. Der Pfad besteht aus einem translatorischen und einem rotatorischen Anteil $\sigma^{\mathrm{T}}(\vartheta) = \begin{bmatrix}\sigma_t^{\mathrm{T}}(\vartheta) & \sigma_r^{\mathrm{T}}(\vartheta)\end{bmatrix}$. Der gewünschte Pfad muss dreifach stetig differenzierbar und regulär sein, es muss also gelten

$$\sigma_t'(\bar{\vartheta}) = \left(\frac{\partial \sigma_t}{\partial \vartheta}\right)(\bar{\vartheta}) \neq \mathbf{0} \tag{4}$$

für alle $\bar{\vartheta} \in \mathscr{T}$. Der aktuelle Fortschritt des Ausgangs $\mathbf{y}_t$ am Pfad wird nun durch den Punkt minimalen Abstands durch

$$\vartheta^* = \arg\min_{\vartheta\in\mathscr{T}} \|\mathbf{y}_t - \sigma_t(\vartheta)\|_2^2, \tag{5}$$

bestimmt.

3.2 Koordinatentransformation

Im Folgenden werden Koordinaten des Ausgangsraums festgelegt. Dazu wird nach [6] ein Paralleltransport-Koordinatensystem an den Pfad gelegt, das sich mit dem aktuellen Referenzpunkt ϑ^* am Pfad mitbewegt. Für die Translation entlang des Pfades wird der Tangentialvektor des Pfades

$$\mathbf{e}_1(\vartheta) = \frac{\sigma_t'(\vartheta)}{\|\sigma_t'(\vartheta)\|_2}. \tag{6}$$

herangezogen. Die zweite Koordinatenrichtung $\mathbf{e}_2(\vartheta)$ folgt aus den Bedingungen für den Paralleltransport des Koordinatensystems entlang des Pfades

$$\mathbf{e}_2'(\vartheta) = \gamma_2(\vartheta)\mathbf{e}_1(\vartheta), \quad \mathbf{e}_2(\vartheta_0) = \mathbf{e}_{2,0} \tag{7a}$$
$$0 = 1 - \mathbf{e}_2^{\mathrm{T}}(\vartheta)\mathbf{e}_2(\vartheta) \tag{7b}$$
$$0 = \mathbf{e}_1^{\mathrm{T}}(\vartheta)\mathbf{e}_2(\vartheta), \tag{7c}$$

wobei sich der Skalar $\gamma_2(\vartheta)$ aus (7c) ergibt, siehe [6] zur Implementierung dieser Gleichungen. Der verbleibende Vektor $\mathbf{e}_3(\vartheta)$ ist durch

$$\mathbf{e}_3(\vartheta) = \mathbf{e}_1(\vartheta) \times \mathbf{e}_2(\vartheta) \tag{8}$$

gegeben.
Die Koordinaten des Fortschritts entlang des Pfades beziehungsweise der Abweichung vom Pfad können nun im zuvor eingeführten Koordinatensystem in der Form

$$\hat{\mathbf{y}}_t = \begin{bmatrix}\eta & \xi_1 & \xi_2\end{bmatrix}^{\mathrm{T}}, \tag{9}$$

mit der Bogenlänge (gemessen von einem Referenzpunkt ϑ_0) entlang des Pfads

$$\eta = \int_{\vartheta_0}^{\vartheta^*} \|\sigma_t'(\tau)\|_2 d\tau \tag{10}$$

und der transversalen Abweichung vom Pfad

$$\xi_1 = \mathbf{e}_2^{\mathrm{T}}(\vartheta^*)(\mathbf{y}_t - \sigma_t(\vartheta^*)) \tag{11}$$
$$\xi_2 = \mathbf{e}_3^{\mathrm{T}}(\vartheta^*)(\mathbf{y}_t - \sigma_t(\vartheta^*)) \tag{12}$$

angegeben werden.
Das rotatorische Teilsystem wird durch die Orientierung des Werkzeugkoordinatensystems bestimmt, das heißt

$$\zeta_r = \mathbf{y}_r = \mathbf{h}_r(\mathbf{q}). \tag{13}$$

Deren Ableitung wird jedoch mit der Drehmatrix $\mathbf{R}_0^s(\vartheta^*)$

in das beim Teach-In Vorgang aufgezeichnete gewünschte Werkzeugkoordinatensystem

$$\dot{\zeta}_r^{\Delta} = (\mathbf{R}_0^s)^{\mathrm{T}}(\vartheta^*)\mathbf{J}_r(\mathbf{q})\dot{\mathbf{q}} \tag{14}$$

transformiert, sodass die Orientierung der z-Achse des Werkzeugs explizit auftritt. Die zugehörigen Winkel werden formal mit ζ_r^{Δ} bezeichnet, sie müssen jedoch nicht explizit berechnet werden.
Damit lautet der virtuelle Ausgang der Pfadfolgeregelung

$$\hat{\mathbf{y}} = \begin{bmatrix}\hat{\mathbf{y}}_t & \hat{\mathbf{y}}_r\end{bmatrix}^{\mathrm{T}} = \begin{bmatrix}\eta & \xi_1 & \xi_2 & \zeta_r^{\Delta}\end{bmatrix}^{\mathrm{T}}. \tag{15}$$

Die zugehörige Zeitableitung ergibt sich zu

$$\dot{\hat{\mathbf{y}}} = \underbrace{\begin{bmatrix}\beta(\mathbf{y}_t)\mathbf{e}_1^{\mathrm{T}}(\vartheta^*) & \mathbf{0} \\ \mathbf{e}_2^{\mathrm{T}}(\vartheta^*) & \mathbf{0} \\ \mathbf{e}_3^{\mathrm{T}}(\vartheta^*) & \mathbf{0} \\ \mathbf{0} & (\mathbf{R}_0^s)^{\mathrm{T}}(\vartheta^*)\end{bmatrix}}_{\mathbf{L}(\mathbf{q})} \begin{bmatrix}\dot{\mathbf{y}}_t \\ \dot{\mathbf{y}}_r\end{bmatrix}, \tag{16}$$

wobei

$$\beta(\vartheta^*) = \frac{1}{1-\alpha(\vartheta^*)}, \quad \alpha = \frac{(\mathbf{y}_t - \sigma_t(\vartheta^*))^{\mathrm{T}}\sigma_t''(\vartheta^*)}{\|\sigma_t'(\vartheta^*)\|_2^2}. \tag{17}$$

Mit der Jacobi-Matrix

$$\mathbf{J}(\mathbf{q}) = \frac{\partial \mathbf{h}(\mathbf{q})}{\partial \mathbf{q}} = \begin{bmatrix}\mathbf{J}_t(\mathbf{q}) \\ \mathbf{J}_r(\mathbf{q})\end{bmatrix} \tag{18}$$

kann damit die auf den Pfad bezogene Jacobi-Matrix

$$\hat{\mathbf{J}}(\mathbf{q}) = \mathbf{L}(\mathbf{q})\mathbf{J}(\mathbf{q}) \tag{19}$$

gefunden werden.

3.3 Reduktion des Arbeitsraums

Der verwendete Roboter hat 7 Freiheitsgrade $\mathbf{q}$. Die Wahl eines geeigneten Arbeitsraums (auch Ausgangsraum) führt zu einem 5-dimensionalem Raum mit den Koordinaten $\tilde{\mathbf{y}}$, wobei das Tilde-Symbol für den reduzierten Arbeitsraum steht, siehe auch [7]. Hierbei wird die Rotation um die z-Achse des Werkzeugkoordinatensystems nicht dem Arbeitsraum zugeschrieben, da diese durch die Rotation der Spule frei drehbar ist, siehe Abbildung 3. Die Differenz der Freiheitsgrade des Konfigurationsraums und des Arbeitsraums entspricht dem Nullraum. Dieser Nullraum wird in weiterer Folge genutzt, um ein sekundäres Regelungsziel zu erreichen.
Berücksichtigt man die Rotation um die z-Achse nicht im Arbeitsraum, so vereinfacht sich die Jacobi-Matrix (19) zu

$$\tilde{\mathbf{J}}(\mathbf{q}) = \begin{bmatrix}\hat{\mathbf{J}}_t \\ \hat{\mathbf{J}}_{r,e_x} \\ \hat{\mathbf{J}}_{r,e_y}\end{bmatrix} \tag{20}$$

und es gilt $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{J}}(\mathbf{q})\dot{\mathbf{q}}$. Leitet man letztere Beziehung nochmals nach der Zeit ab, so folgt

$$\ddot{\tilde{\mathbf{y}}} = \dot{\tilde{\mathbf{J}}}(\mathbf{q})\dot{\mathbf{q}} + \tilde{\mathbf{J}}(\mathbf{q})\ddot{\mathbf{q}} \tag{21}$$

und mit der Starrkörperdynamik (1) gilt

$$\ddot{\tilde{\mathbf{y}}} = \dot{\tilde{\mathbf{J}}}\dot{\mathbf{q}} + \tilde{\mathbf{J}}\left[\mathbf{M}^{-1}(\tau + \tau_{ext} - \mathbf{n})\right]. \tag{22}$$

3.4 PD-Regelung im Arbeitsraum

Die statische Zustandsrückführung

$$\tau = \mathbf{n} - \tau_{ext} + \mathbf{M}\tilde{\mathbf{J}}^{\dagger}\left(\mathbf{v} - \dot{\tilde{\mathbf{J}}}\,\dot{\mathbf{q}}\right) + \tau_N \tag{23}$$

mit der Pseudoinversen

$$\tilde{\mathbf{J}}^{\dagger} = \tilde{\mathbf{J}}^{\mathrm{T}}\left(\tilde{\mathbf{J}}\tilde{\mathbf{J}}^{\mathrm{T}}\right)^{-1} \tag{24}$$

bringt das System in die Form $\ddot{\tilde{\mathbf{y}}} = \left[\mathbf{v}_t^{\mathrm{T}}, \mathbf{v}_r^{\mathrm{T}}\right]^{\mathrm{T}} = \mathbf{v}$ mit dem virtuellen Stelleingang $\mathbf{v}$. Das generalisierte Drehmoment τ_N in (23) wird zur Regelung des Nullraums verwendet und wird im nachfolgenden Abschnitt beschrieben.
Bezeichnet man mit

$$\tilde{\mathbf{e}}_t = \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{y}}_t^d\right) = \begin{bmatrix}\eta - \eta^d \\ \xi_1 - \xi_1^d \\ \xi_2 - \xi_2^d\end{bmatrix}. \tag{25}$$

den Regelfehler des translatorischen Teilsystems (Superskript d gibt die jeweilige Referenzgröße an), dann führt das PD-Regelgesetz mit Vorsteuerung

$$\mathbf{v}_t = \ddot{\tilde{\mathbf{y}}}_t^d - \mathbf{K}_{1,t}\dot{\tilde{\mathbf{e}}}_t - \mathbf{K}_{0,t}\tilde{\mathbf{e}}_t, \tag{26}$$

zu der Fehlerdynamik

$$\ddot{\tilde{\mathbf{e}}}_t + \mathbf{K}_{1,t}\dot{\tilde{\mathbf{e}}}_t + \mathbf{K}_{0,t}\tilde{\mathbf{e}}_t = \mathbf{0}. \tag{27}$$

Wählt man $\mathbf{K}_{1,t}$ und $\mathbf{K}_{0,t}$ als Diagonalmatrizen mit positiven Einträgen, so sind die Fehlerdynamiken entkoppelt und klingen exponentiell ab.
Der Orientierungsfehler $\tilde{\mathbf{e}}_r$ wird anhand des Fehlerquaternions $\mathbf{e}_{\mathscr{Q}}$ berechnet. Dieses beschreibt die relative Drehung der gewünschten Orientierung der Spule $\mathscr{Q}^d$ zur aktuellen Orientierung des Werkzeugkoordinatensystems $\mathscr{Q}_e$. Durch das Quaternionenprodukt

$$\mathbf{e}_{\mathscr{Q}} = \begin{Bmatrix}\mu & \varepsilon\end{Bmatrix} = \mathscr{Q}^d \otimes \mathscr{Q}_e^{-1} \tag{28}$$

ergibt sich der Orientierungsfehler als Quaternion mit dem Skalarteil μ und dem Vektorteil ε. Das resultierende Fehlerquaternion wird genau dann zu $\mathbf{e}_{\mathscr{Q}} = \{\mu \quad \mathbf{0}\}$, wenn die Orientierung des Werkzeugs $\mathscr{Q}_e$ der gewünschten Orientierung $\mathscr{Q}^d$ entspricht, siehe dazu [8] und [9]. Da der dritte Eintrag des Vektorteils $\varepsilon = \begin{bmatrix}\varepsilon_1 & \varepsilon_2 & \varepsilon_3\end{bmatrix}^{\mathrm{T}}$ der Verdrehung um die z-Achse entspricht, folgt der reduzierte Orientierungsfehler zu $\tilde{\mathbf{e}}_r = \begin{bmatrix}\varepsilon_1 & \varepsilon_2\end{bmatrix}^{\mathrm{T}}$. Der virtuelle Eingang des rotatorischen Teilsystems kann nun zu

$$\mathbf{v}_r = \dot{\omega}^d - \mathbf{K}_{1,r}\left(\omega - \omega^d\right) - \mathbf{K}_{0,r}\tilde{\mathbf{e}}_r \tag{29}$$

mit den positiv definiten (Diagonal-)Matrizen $\mathbf{K}_{1,r}$ und $\mathbf{K}_{0,r}$ gewählt werden, wobei ω den aktuellen und ω^d den gewünschten Vektor der Winkelgeschwindigkeiten des Werkzeugs angeben. Die gewünschte Winkelgeschwindigkeit wird mit Hilfe der schiefsymmetrischen Matrix

$$\mathbf{S}(\omega_{0,0}^s) = \dot{\mathbf{R}}_0^s(\vartheta^*)\dot{\vartheta}^*\mathbf{R}_0^s(\vartheta^*)^{\mathrm{T}} \tag{30}$$

bestimmt, wobei $\omega_{0,0}^s$ der Vektor der Winkelgeschwindigkeit der Spule relativ zum Basiskoordinatensystem beschreibt und im Basiskoordinatensystem ausgedrückt ist. Um die z-Achse des Werkzeugkoordinatensystems, also

die Drehrichtung der Spule, explizit auszudrücken, wird $\omega^s_{0,0}$ ins Werkzeugkoordinatensystem transformiert

$$\omega^s_{0,s} = \left(\mathbf{R}^s_0\right)^{\mathrm{T}}(\vartheta^*)\,\omega^s_{0,0}. \tag{31}$$

Dadurch ergibt sich der gewünschte Vektor der Winkelgeschwindigkeiten des Werkzeugs zu

$$\omega^d = \begin{bmatrix}\omega^s_{0,s_x} & \omega^s_{0,s_y}\end{bmatrix}^{\mathrm{T}}, \tag{32}$$

wobei die z-Komponente weggelassen wird. Einen Stabilitätsbeweis ohne Reduktion des Arbeitsraums gemäß Abschnitt 3.3 kann in [8, 9] gefunden werden. Experimente legen jedoch nahe, dass auch die nur 2-achsige Regelung zu einem stabilen Regelkreis führt.

3.5 Nullraumregelung

Der erweiterte Nullraum kann nun genutzt werden, um ein sekundäres Regelungsziel umzusetzen. Dabei werden im präsentierten Szenario die Achsen des Roboters auf ihre Mittelstellung geregelt, also $\mathbf{q}_0 = \mathbf{0}$. Dadurch kann die Erreichbarkeit eines größeren Arbeitsraums ermöglicht werden und es können Seilpfade gewickelt werden, die ohne Nutzung des erweiterten Nullraums den Roboter in kinematische Beschränkungen führen. Die hierarchische Gliederung von Regelzielen kann zum Beispiel durch eine Projektion erfolgen, siehe dazu auch [7]. Unter Verwendung der reduzierten Jacobi-Matrix (20) wird die *dynamisch konsistente* Projektion

$$\mathbf{N}_J(\mathbf{q}) = \mathbf{I} - \tilde{\mathbf{J}}^{\mathrm{T}}\tilde{\mathbf{J}}^{\mathbf{M}}, \tag{33}$$

berechnet. Hierin sind $\mathbf{I}$ die Einheitsmatrix und $\tilde{\mathbf{J}}^{\mathbf{M}}$ die mit der Massenmatrix gewichtete Pseudoinverse

$$\tilde{\mathbf{J}}^{\mathbf{M}} = \mathbf{M}^{-1}\tilde{\mathbf{J}}^{\mathrm{T}}\left(\tilde{\mathbf{J}}\mathbf{M}^{-1}\tilde{\mathbf{J}}^{\mathrm{T}}\right)^{-1}. \tag{34}$$

Durch diese Nullraumprojektion können Stelleingänge in den Nullraum projiziert werden, wobei garantiert ist, dass Kräfte und Beschleunigungen unabhängig vom primären Regelungsziel eingebracht werden. Es sei erwähnt, dass umgekehrt die sekundäre Regelungsaufgabe nicht unabhängig von der Primäraufgabe gemacht wird.
Mit Hilfe der Nullraumprojektionsmatrix (33) wird der Nullraum mit dem PD-Reglgesetz

$$\tau_{\mathrm{N}} = \mathbf{N}_J\left(-\mathbf{K}_{\mathrm{N1}}\dot{\mathbf{q}} - \mathbf{K}_{\mathrm{N0}}(\mathbf{q} - \mathbf{q_0})\right) \tag{35}$$

stabilisiert. Hierin stellen $\mathbf{K}_{\mathrm{N1}}$ und $\mathbf{K}_{\mathrm{N0}}$ diagonale Verstärkungsmatrizen mit positiven Einträgen dar.
Durch dieses Regelungskonzept bestehend aus Pfadfolge- und Nullraumregelung wird die Orientierung bezüglich der x- und y-Achse des Werkzeugkoordinatensystems durch (29) geregelt, während die Orientierung der z-Achse des Endeffektorkoordinatensystems durch (35) eingestellt wird, um den Roboter in kinematisch günstige Konfigurationen zu bringen.

4 Experimentelle Ergebnisse

Im folgenden Abschnitt werden die Ergebnisse eines praktischen Experiments dargestellt. Zur Demonstration der vorgestellten Methodik wurden drei Stangen im Arbeitsraum des Roboters platziert, siehe Abbildung 1. Dabei werden die ersten beiden Stangen mindestens einmal vollständig umschlungen, während die dritte Stange zur Umlenkung des Seils dient.

4.1 Demonstration eines Seilwickelvorgangs

Der durch eine Demonstration mit dem vorgestellten Werkzeug aufgenomme Pfad ist in **Abbildung 4** gezeigt. Dabei wurde die Kurve durch eine Spline-Kurve dritter Ordnung angenähert, um die nötige Differenzierbarkeit zu gewährleisten.

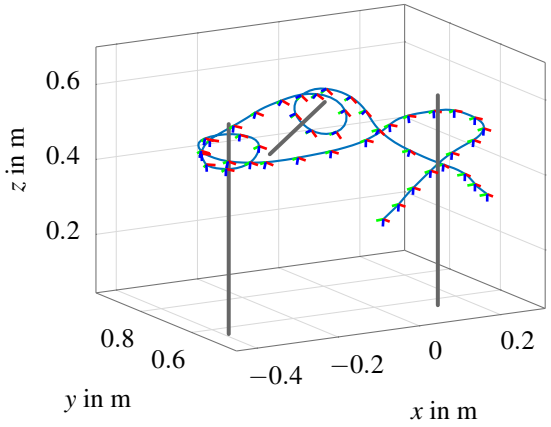


Abbildung 4 Aufgenommener parametrierter Pfad $\sigma(\vartheta)$ mit Koordinatenachsen zur Illustration der gewünschten Orientierung und umwickelte Stangen in ursprünglicher Konfiguration.

Die zweite Stange ist elastisch gelagert, sodass diese durch den eingebrachten Seilzug geneigt werden kann, siehe Abbildung 1. Nachdem die zweite Stange umwickelt wurde, wird die Seilzugspannung erhöht, sodass die Stange in einer geneigten Stellung verbleibt. Die hierzu nötige Solltrajektorie des skalierten gewünschten Motordrehmoments ist in **Abbildung 5** dargestellt.

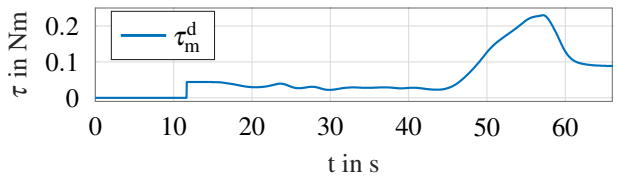


Abbildung 5 Drehmoment τ^d_m an der Spule zur Einprägung der gewünschten Seilzugspannung.

4.2 Wiedergabe durch den Roboter

Bei der Wiedergabe des Experiments durch den Roboter wurde eine genaue Reproduktion des Seilwickelvorgangs erreicht. Durch den am Roboterwerkzeug angebrachten Motor konnte die gewünschte Seilzugspannung hinreichend genau vorgegeben werden, sodass die demonstrierte Neigung der elastisch gelagerten zweiten Stange durch das Seil mit hoher Genauigkeit durch den Roboter reproduziert wurde.

Die während der Roboterbewegung aufgetretenen Fehler sind in **Abbildung 6** dargestellt. Hierbei wird deutlich, dass

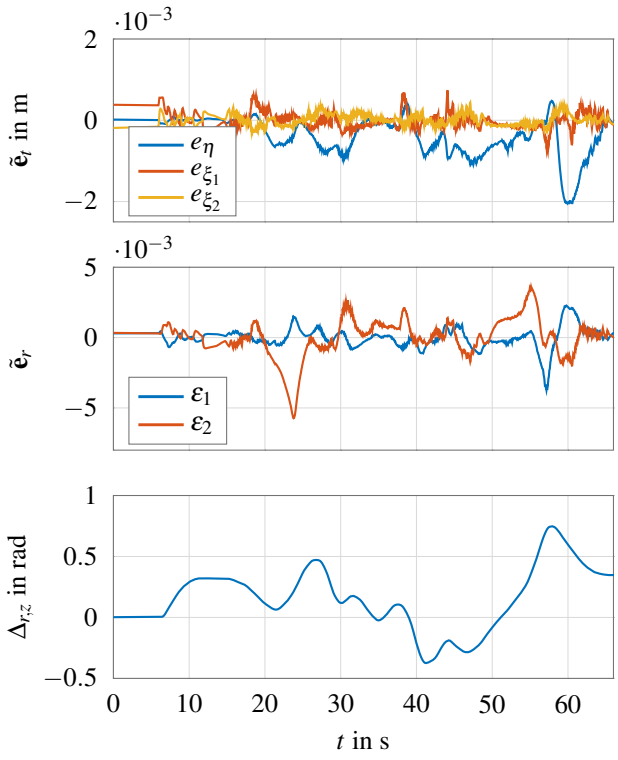


Abbildung 6 Regelfehler während des Experiments.

sowohl der Pfadfortschritt η als auch die Positionierung der Rolle mit Genauigkeiten im Millimeterbereich eingehalten werden konnten. Der etwas größere Fehler des Pfadfortschritts ist auf die erhöhte Seilzugspannung zurückzuführen, die am Ende der Trajektorie benötigt wird.
Die Quaternionenfehlerkoordinaten ε_1 und ε_2 bestätigen die Konvergenz des Algorithmus zur gewünschten Orientierung in den Rotationen um die x- und y-Achse.
Der dritte Graph in Abbildung 6 gibt die Abweichung $\Delta_{r,z}$ der Rotation der z-Achse des Werkzeugs von seiner Nullstellung an. Durch die signifikante Verdrehung der z-Achse wird der Nutzen der eingeführten Reduktion des Arbeitsraums und der Nullraumregelung deutlich. Diese Verdrehung erlaubt es dem Roboter in einer kinematisch günstigeren Konfiguration das Werkzeug in die gewünschte Pose zu bringen.

5 Zusammenfassung

In dieser Arbeit wurde das Teach-In und die darauf basierende automatische Abarbeitung eines Seilwickelprozesses mit Hilfe spezialisierter Werkzeuge vorgestellt. Mit dem Demonstrationswerkzeug kann einhändig ein Seil um frei wählbare Wickelpunkte gespannt werden. Durch ein optisches Trackingsystem und einen Drehmomentsensor können Position, Orientierung und Seilzugspannung während des Wickelprozesses aufgezeichnet werden. Durch die Schiefstellung des Roboterwerkzeugs konnte der nutzbare Arbeitsraum des 7-Achs-Roboters merklich vergrößert werden. Im Weiteren wurde eine speziell formulierte Pfadfolgeregelung verwendet, bei der die Rotation der Seilspule als Nullraumbewegung betrachtet wird. Dies sorgt für kinematisch günstigere Konfigurationen des Roboters während des Seilwickelprozesses.
Der Seilwickelprozess mit einem vom Pfadfortschritt abhängigen Seilzug konnte mit hoher Präzision wiedergegeben werden. Durch das einfach gestaltete Demonstrationswerkzeug konnte der gewünschte Prozess schnell eingelernt werden. Speziell die Reichweite des Roboters stellte sich bei den Demonstrationen als größtes Hindernis heraus. Diese Probleme konnten durch die vorgestellte Regelung stark vermindert werden. In zukünftigen Arbeiten könnte die Trajektorie optisch aufgezeichnet und die Absolutgenauigkeit bei der Wiedergabe durch den Roboter weiter verbessert werden.

6 Literatur

[1] J. Scholliers and H. Van Brussel: Computer-integrated filament winding: computer-integrated design, robotic filament winding and robotic quality control, Composites Manufacturing, vol. 5, 1994, S.15-23.

[2] S. Flixeder, T. Glück, and A. Kugi, Force-based cooperative handling and lay-up of deformable materials: Mechatronic design, modeling, and control of a demonstrator, Mechatronics, vol 47, 2017, S.246 - 261.

[3] C. Hartl-Nesic, T. Glück and A. Kugi: Surface-Based Path Following Control: Application of Curved Tapes on 3-D Objects, IEEE Transactions on Robotics, 2020.

[4] A. Billard, S. Calinon, R. Dillmann, and S. Schaal: Robot Programming by Demonstration, in Springer Handbook of Robotics, B. Siciliano and O. Khatib, Eds. Berlin, Heidelberg: Springer, 2008, S. 1371–1394.

[5] S. Wrede, C. Emmerich, R. Grünberg, A. Nordmann, A. Swadzba, and J. Steil: A user study on kinesthetic teaching of redundant robots in task and configuration space, Journal of Human-Robot Interaction, vol 2, S. 56-81, 2013.

[6] B. Bischof, T. Glück, and A. Kugi: Combined Path Following and Compliance Control for Fully Actuated Rigid Body Systems in 3-D Space, IEEE Transactions on Control Systems Technology, vol. 25, no. 5, S. 1750–1760, 2017.

[7] A. Dietrich, C. Ott, and A. Albu-Schäffer: An Overview of Null Space Projections for Redundant, Torque-Controlled Robots, The International Journal of Robotics Research, vol. 34, no. 11, S. 1385–1400, 2015.

[8] J. S. Yuan: Closed-Loop Manipulator Control using Quaternion Feedback, IEEE Journal on Robotics and Automation, vol. 4, no. 4, S. 434–440, 1988.

[9] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo: Robotics: Modelling, Planning and Control, London, Springer, 2009.

Ein Lösungsansatz zur Verbesserung der Genauigkeit von kraftgeregelten Applikationen basierend auf einem seriellen Roboter mit integrierter Sensorik

An approach to improve the accuracy of force-controlled applications based on a serial robot with integrated sensor technology

Prof. Dr.-Ing. Rainer Müller, ZeMA – Zentrum für Mechatronik und Automatisierungstechnik, 66121 Saarbrücken, Deutschland, rainer.mueller@zema.de

Ali Kanso M. Sc., ZeMA – Zentrum für Mechatronik und Automatisierungstechnik, 66121 Saarbrücken, Deutschland, a.kanso@zema.de

Marco Schneider M. Sc., ZeMA – Zentrum für Mechatronik und Automatisierungstechnik, 66121 Saarbrücken, Deutschland, m.schneider@zema.de

Kurzfassung

Spezielle Sensorik sowie ausgewählte Regelungsstrategien befähigen sensitive Robotersysteme mit ihrer Umgebung zu interagieren. Diese Interaktion mit der Umgebung kann in Form von kraftgeregelten Applikationen umgesetzt werden, welche insbesondere im industriellen Umfeld hinsichtlich der Automatisierung von Montage- und Fertigungsprozessen ein hohes Anwendungspotential bieten. Oft werden die benötigten externen Kräfte und Momente rechnerisch auf Basis von 1D Sensoren, montiert in jeder Achse des Roboters, bestimmt. Diese Ermittlung weist allerdings häufig eine Ungenauigkeit auf. Die Ursache dabei ist die Ungenauigkeit des kinematischen und dynamischen Modells des Roboters [1, 2]. Deshalb wird im Rahmen dieses Beitrags ein Ansatz zur Verbesserung der Genauigkeit von kraftgeregelten Applikationen basierend auf einem seriellen Roboter mit integrierter Sensorik vorgestellt. Darüber hinaus werden auf die Grundlagen der Kräfteberechnung eingegangen, sowie letztlich der Nutzen des Optimierungsprozesses anhand eines Validierungsszenarios in Form des Kuka LBR iiwa im Zusammenspiel mit einem dreidimensionalen Bauteil veranschaulicht.

Abstract

Special sensor technology and selected control strategies enable sensitive robot systems to interact with their environment. This interaction with the environment can be implemented in the form of force-controlled applications, which offer a high application potential especially in the industrial environment with regard to the automation of assembly and manufacturing processes. Often, the required external forces and torques are determined computationally based on 1D sensors mounted in each axis of the robot. However, this determination often has an inaccuracy. The main cause is the inaccuracy of the kinematic and dynamic model of the robot [1, 2]. Therefore, this paper presents an approach to improve the accuracy of force controlled applications based on a serial robot with integrated sensors. Furthermore, the basics of force calculation are discussed, and finally the benefits of the optimization process are illustrated by means of a validation scenario in the form of the Kuka LBR iiwa in interaction with a three-dimensional component.

1 Einleitung

Mechatronische Robotersysteme und deren effektiver Einsatz eröffnen zahlreiche Anwendungsfelder in Forschung und Industrie. Allerdings werden sie auch heute primär für repetitive, gefährliche oder schwere Aufgaben eingesetzt [1]. Um das Potential solcher mechatronischer Systeme auszuschöpfen, existieren seit nunmehr einigen Jahren sog. sensitive Robotersysteme auf dem Markt, deren Einsatzszenarien sich weit über die von konventionellen seriellen Robotern erstrecken [2]. Vor allem die spezielle Sensorik sowie ausgewählte Regelungsstrategien befähigen sensitive Robotersysteme mit ihrer Umgebung zu interagieren. Insbesondere im industriellen Umfeld im Kontext der Automatisierung zeigt sich ein hohes Anwendungspotential durch sensitive Roboterapplikationen. Dabei ist es möglich, zum einen das Roboterverhalten auf Basis der Sensorik zu adaptieren und zum anderen die Umgebung zu identifizieren.

2 Stand der Technik

2.1 Kraftgeregelte Applikationen

Kraftgeregelte Applikationen weisen die Gemeinsamkeit einer adäquaten Sensorik auf, die allerdings auf unterschiedliche Art und Weise realisiert werden kann. So nutzen unterschiedliche Roboterhersteller Sensortechnik, welche sich zum einen im physikalischen Prinzip und zum anderen an der Stelle der Kraft-/ Momentenaufnahme unterscheiden. Infolgedessen existiert heutzutage eine Vielzahl an Robotern, die eine solche sensitive Funktionalität besitzen. Diese technologische Erweiterung ebnet den Weg für kollaborative Roboteranwendungen, die neben der Sicherstellung der physisch sicheren Zu-

sammenarbeit zwischen Mensch und Roboter ebenfalls feinfühlige Bewegungen bzw. Aufgaben ermöglichen. Die Entwicklung von sensitiven seriellen Robotern beflügelt Weiterentwicklungen innerhalb der Robotik und zahlreicher Begleittechnologien. Allerdings stellt sie diese auch aufgrund hoher Erwartungshaltungen seitens der Industrie vor großen Herausforderungen. [5]

Um die angesprochenen Kontrast zwischen den eingesetzten Technologien der Hersteller deutlich zu machen, kann zunächst die Stelle bzw. der Ort der Kraft-/ Momentenaufnahme voneinander abgegrenzt werden. Dabei wird nämlich differenziert zwischen der Aufnahme

- am Flansch
- in den Gelenken
- in der Basis.

Daraus ergeben sich unterschiedliche Spezifikationen hinsichtlich der Sensitivität der einzelnen Roboter, die zur Verdeutlichung in Tabelle 1 aufgeführt sind.

Eig.: / *Roboter:*	*Kraft-/ Momentenaufnahme*
KUKA LBR iiwa	Integrierte Drehmomentsensoren in jeder Achse
Universal Robots UR 10e	Stromüberwachung (Zweikanalig), Integrierter Kraft-Drehmomentsensor im Flansch
Yaskawa MOTOMAN HC10	Integrierte Kraft-Momentensensoren in allen 6 Achsen (optische Sensoren)
Franka Emika Panda	Drehmomentsensoren in allen 7 Achsen
Fanuc CR-35iA	Kraft- und Drehmomentsensor in der Basis
Denso Robotics Cobotta	Geschwindigkeits- und Drehmomentsensoren in allen 6 Achsen
Doosan Robotics M1013	Drehmomentsensoren in allen 6 Gelenken

Tabelle 1 Übersicht der sensitiven Eigenschaften ausgewählter industrietauglicher sensitiver Roboter [6] [7] [8] [9] [10] [11] [12].

Ohne auf die Funktionsweise selbst einzugehen sei an dieser Stelle erwähnt, dass die Mehrzahl der genutzten Sensorik auf sog. Dehnungsmessstreifen (DMS) beruht, weshalb diesbezüglich lediglich auf die einschlägige Literatur [13] hingewiesen wird.

Zur Unterstreichung des hohen Potentials in industrieller Umgebung sind in Abb. 1 zwei Fallstudien aus dem Gebiet der spanenden Fertigung aufgezeigt. Dabei werden Aluminiumgussteile mit Hilfe eines 6D Kraft-Momentensensor am Flansch geschliffen, poliert und gefräst. Darüber hinaus existieren ebenfalls praktische Anwendungen diesbezüglich im Rahmen von Montageoperationen.

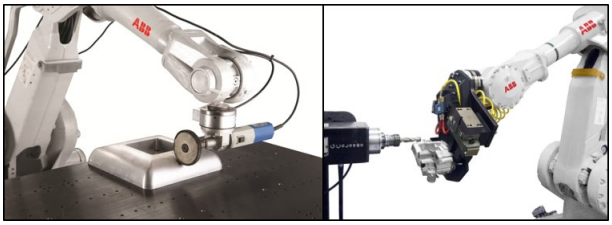


Abbildung 1 Beispiele kraftgeregelter und sensitiver Bearbeitungsprozesse [14]

2.2 Modellierung eines Roboters

Zur Beschreibung der Kinematik und Dynamik eines Handhabungsgerätes existieren grundsätzlich jeweils geeignete Ersatzmodelle. Diese sollen zum einen die Eigenschaften des realen Gerätes möglichst genau wiedergeben, zum anderen jedoch eine einfache mathematische Beschreibung zulassen. [1]

Um das direkte und inverse kinematische Problem eines Roboters zu berechnen und somit Informationen über dessen Bewegungszustand zu erlangen, existieren verschiedene Modelle. Die dabei am weitesten verbreitete Konvention zur Modellierung eines Handhabungsgerätes ist die sog. DH-Konvention. Diese Denavit-Hartenberg-Konvention modelliert einen Roboter mit Hilfe von Koordinatensystemen (KS) sowie spezifischen Parametern, wodurch u.a. das Lösen des direkten und inversen kinematischen Problems ermöglicht wird. [15, 16]

Da im Rahmen dieses Beitrags der Roboter mit seiner Umgebung interagiert und somit u.a. Kontaktkräfte auftreten, ist eine dynamische Betrachtung des Roboters sinnvoll. Insbesondere der Zusammenhang zwischen den Antriebsdrehmomenten und der vorgegebenen Trajektorie mitsamt den Geschwindigkeits- und Beschleunigungsprofilen sowie den externen Kräften und Momenten wird durch die direkte und inverse Roboterdynamik beschrieben. Diese ist ebenso wie die zuvor beschriebene kinematische Modellierung in der Robotersteuerung implementiert.

Wenn die Newton-Euler-Gleichungen symbolisch für einen beliebigen Manipulator ausgewertet werden, ergibt sie eine Bewegungsgleichung, die in der folgenden Form geschrieben werden kann [16]:

$$\underline{\tau} = \mathbf{M}(\underline{q})\ddot{\underline{q}} + \mathbf{C}(\underline{q},\dot{\underline{q}})\dot{\underline{q}} + \mathbf{G}(\underline{q}) + \mathbf{J}_G{}^T\underline{F} \tag{1}$$

Hier ist $\underline{\tau}$ der Vektor der Antriebsdrehmomente, die am Gelenk durch den Antrieb hervorgerufen werden, $\mathbf{M}$ ist die Massenträgheitsmatrix, $\mathbf{C}$ ist die Matrix der Coriolis- und Zentrifugalmomente, $\mathbf{G}$ bildet die Matrix der Gewichtsmomente, also die durch Gravitation hervorgerufene Drehmomente. $\mathbf{J}_G{}^T$ stellt die transponierte geometrische Jacobimatrix dar und $\underline{F}$ den Vektor der kartesischen Kräfte und Momente, die auf den Endeffektor wirken. Diese Bewegungsgleichung dient der dynamischen Modellierung eines Roboters und besteht aus nichtlinearen, gekoppelten Differentialgleichungen zweiter Ordnung. Dadurch ist es möglich

den Ort $\underline{q}$, die Geschwindigkeit $\dot{\underline{q}}$ und die Beschleunigung $\ddot{\underline{q}}$ mit den Antriebsdrehmomenten zu verknüpfen.

2.3 Regelungsstrategien für Roboter mit offener Hauptkette

In der Steuerung von Handhabungsgeräten werden unterschiedliche Arten von Reglern verwendet. Dazu zählen i.d.R. Positions- und Kraftregler. Erstere kommen beim Abfahren einer definierten Trajektorie im Raum zum Einsatz, während letztere bevorzugt bei einem Kontakt des Endeffektors mit der Umgebung des Roboters Verwendung finden. [17]

Mit Bezug auf die Thematik des vorliegenden Beitrags und dem verwendeten Roboter KUKA LBR iiwa stellen die

- Positionsregelung
- Impedanzregelung
- Kraft- und Momentenregelung

die wichtigsten Regler dar und werden infolgedessen in den nächsten Abschnitten näher beschrieben.

2.3.1 Positionsregelung

Grundsätzlich können zwei unterschiedliche Regelungsstrategien verfolgt werden - nämlich modellfreie und modellbasierte Regler. Wie die Bezeichnung der Regler bereits impliziert, unterscheiden sie sich in der Kenntnis über das dynamische Modell des Roboters (vgl. Kapitel 2.2). [18]

2.3.2 Impedanzregelung

Die Impedanzreglung beruht ebenfalls auf dem dynamischen Modell des Roboters und setzt folglich detaillierte Kenntnisse des Modells in Form von speziellen, expliziten Gleichungen (vgl. Gl. 1) voraus. Im Allgemeinen sind modellbasierte Regelungen weniger trivial als modellfreie und somit aufwendiger zu implementieren. [19] Ist jedoch die Interaktion des Roboters mit seiner Umgebung gewünscht, so ist die Impedanzregelung eine häufig genutzte Regelungsart. Im Grunde fußt die Impendanzregelung auf einem virtuellen Masse-Feder-Dämpfer-System. Dabei kann durch Adaption der Parameter Steifigkeit und Dämpfung das Verhalten des Roboters bei physikalischem Kontakt maßgeblich beeinflusst werden.

2.3.3 Kraft-/ Momentenregelung

Bei direkter physikalischer Interaktion des Endeffektors mit der Umgebung ist wie bereits erklärt eine reine Positionsregelung unzureichend, eine reine Kraftregelung allerdings auch. Typische Aufgaben wie Entgraten, Polieren oder Montieren erfordern oftmals eine Kraftregelung in einer vordefinierten kartesischen Raumrichtung. Deshalb ist es sinnvoll einen hybriden Kraft- Positionsregler zu verwenden, dessen Ziel es ist, gleichzeitige die Regelung sowohl der Endeffektor-Bewegung als auch der Kontaktkräfte zu gewährleisten. Dies erfolgt durch eine Unterteilung in zwei Bereiche, also einer Entkopplung der Teilprobleme. [19]

Abhängig von der auszuführenden Operation muss zunächst ermittelt werden, welche der translatorischen und rotatorischen Freiheitsgrade entweder positions- oder kraftgeregelt werden. Dies geschieht unter Zuhilfenahme der sog. Selektionsmatrizen $\mathbf{S}$ und $\tilde{\mathbf{S}}$. Diese haben die Dimension 6×6. Die Matrix $\mathbf{S}$ repräsentiert die kraft-/ momentgeregelten Freiheitsgrade und $\tilde{\mathbf{S}}$ die positionsgeregelten Freiheitsgrade. Weist nun das entsprechende Element auf der Hauptdiagonalen der Matrix eine 1 auf, so gilt diese Richtung bzw. rotatorische Komponente als kraft- bzw. als momentengeregelt. Alle anderen Elemente der Matrix sind 0. Da die Hauptdiagonlaen von $\mathbf{S}$ und $\tilde{\mathbf{S}}$ zueinander komplementär sind, ergibt sich mit $\mathbf{I}$ als Einheitsmatrix folgender Zusammenhang [20]:

$$\tilde{\mathbf{S}} = \mathbf{I} - \mathbf{S} \tag{2}$$

Abbildung **??** zeigt das Blockschaltbild der hybriden Kraft-/ Lageregelung.

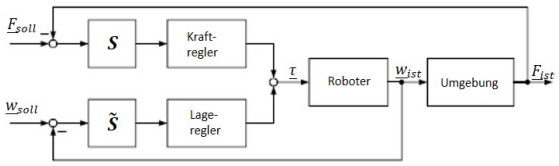


Abbildung 2 Blockschaltbild der hybriden Kraft-/ Lageregelung, in Anlehnung an [20]

3 Entwicklungsprozess

3.1 Grundlagen der Kräfteberechnung

Die Kräfteberechnung stellt im Kontext der sensitiven Robotik und kraftgeregelten Applikationen einen zentralen Punkt dar. Ohne die Kräfteberechnung sind Rückschlüsse auf die Interaktion mit der Umgebung nicht möglich und folglich insbesondere bei Mensch-Roboter-Applikationen und kraft- und momentengeregelten Aufgaben ein unabdingbarer Bestandteil. In Abbildung 3 ist das Prinzip der Kräfteberechnung abstrahiert dargestellt.

Wirken nun Kräfte und Momente, die durch den Vektor

$${}^{TCP}\underline{K}_{TCP} = \begin{pmatrix} F_x & F_y & F_z & M_x & M_y & M_z \end{pmatrix}^T \tag{3}$$

repräsentiert werden auf den Schwerpunkt S des montierten Werkzeuges, so muss zunächst der Einfluss der Gravitation eliminiert werden, welcher unmittelbar mit den kinematischen und dynamischen Parametern wie Masse, Schwerpunktlage und DH-Parametern zusammenhängt. Dazu existiert ein Vektor ähnlich dem Vektor ${}^{TCP}\underline{K}_{TCP}$, welcher die durch die Gravitation verursachten kartesischen Kräfte und Momente berücksichtigt. Dieser wird wie

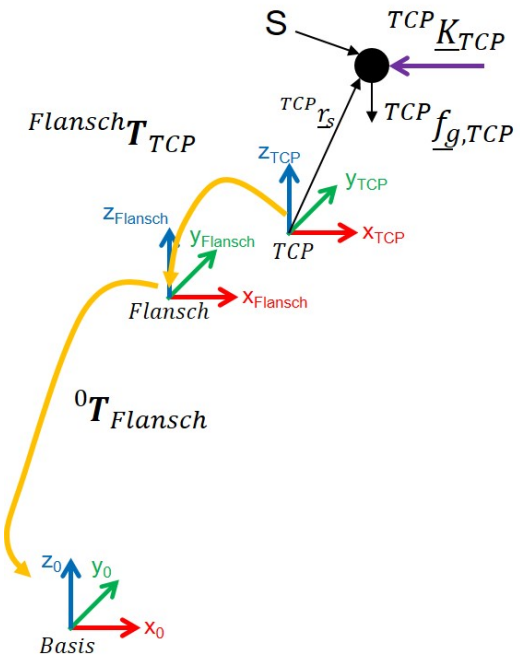


Abbildung 3 Grundlage der Kräfteberechnung

folgt berechnet:

$$ {}^{TCP}\underline{K}_{g,TCP} = \begin{pmatrix} {}^{TCP}\underline{F}_{g,TCP} \\ {}^{TCP}\underline{M}_{g,TCP} \end{pmatrix} = \begin{pmatrix} {}^{TCP}\mathbf{D}_0 \cdot \begin{pmatrix} 0 \\ 0 \\ {}^{0}\underline{f}_{g,TCP} \end{pmatrix} \\ {}^{TCP}\underline{r}_s \times {}^{TCP}\underline{F}_{g,TCP} \end{pmatrix} \tag{4} $$

Dabei stellt die Matrix ${}^{TCP}\mathbf{D}_0$ die Drehmatrix des Basis-KS bezogen auf das TCP-KS, der Vektor ${}^{TCP}\underline{r}_s$ die Schwerpunktlage des Werkzeuges und der Vektor ${}^{0}\underline{f}_{g,TCP}$ die Gravitationskraft am TCP dar. Somit ergibt sich der Vektor der kartesischen externen Kräfte und Momente ${}^{TCP}\underline{K}_{e,TCP}$ zu.

$$ {}^{TCP}\underline{K}_{e,TCP} = {}^{TCP}\underline{K}_{TCP} - {}^{TCP}\underline{K}_{g,TCP} \tag{5} $$

Soll nun die Belastung bzw. die Kräfte und Momente in einem beliebigen Koordinatensystem dargestellt werden, so muss eine Umrechnung erfolgen. Im Folgenden ist ein Beispiel der Umrechnung von einer Belastung am TCP im TCP-KS zu einer Belastung am Flansch im Flansch-KS angeführt. Zunächst erfolgt die Umrechnung in das Ziel Koordinatensystem mit Hilfe der entsprechenden Drehmatrix.

$$ {}^{Flansch}\underline{K}_{e,TCP} = \begin{pmatrix} {}^{Flansch}\mathbf{D}_{TCP} \cdot {}^{TCP}\underline{F}_{e,TCP} \\ {}^{Flansch}\mathbf{D}_{TCP} \cdot {}^{TCP}\underline{M}_{e,TCP} \end{pmatrix} \tag{6} $$

Außerdem muss das durch die Abstände der Koordinatensysteme hervorgerufene Moment Berücksichtigung finden. Somit ergibt sich:

$$ {}^{Flansch}\underline{K}_{e,Flansch} = {}^{Flansch}\underline{K}_{e,TCP} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ {}^{Flansch}\underline{r}_{TCP} \times {}^{Flansch}\underline{F}_{e,TCP} \end{pmatrix} \tag{7} $$

3.2 Problemstellung

Der hier verwendete Roboter - Kuka LBR iiwa - verfügt, wie bereits beschrieben, über eindimensionale Drehmomentsensoren in jeder Achse [6]. Um allerdings die im Kontext der sensitiven Robotik essentiell wichtigen externen Kräfte und Momente an beliebigen Orten und im Bezug auf beliebige Koordinatensysteme (vgl. Kapitel 3.1) zu bestimmen ist eine spezifische Berechnung notwendig. Der Zusammenhang hier lautet:

$$ \underline{K} = (\mathbf{J}_G{}^T)^{-1} \cdot \underline{\tau} \tag{8} $$

Das bedeutet die gemessenen Drehmomente in den einzelnen Achsen, repräsentiert durch den Vektor $\underline{\tau}$, werden mithilfe der geometrischen Jacobimatrix in die auf den Roboter wirkenden Kräfte und Momente gebildet. An dieser Stelle wird bereits deutlich, dass die Genauigkeit der Berechnung zum einen von dem kinematischen Modell des Roboters abhängt, da die Jacobimatrix unmittelbar auf diesem beruht. Die Ungenauigkeiten in der kinematischen Modellierung in Form der Denavit-Hartenberg-Notation, werden unmittelbar auf die Berechnung der externen Kräfte und Momente übertragen. Darüber hinaus ist wie zuvor beschrieben die Dynamik ein weiterer wichtiger Einflussfaktor auf die Genauigkeit der Bestimmung. Durch die dynamische Modellierung des Roboters in Form von Massen, Trägheitsmomenten sowie Schwerpunktlagen der einzelnen Glieder, werden etwaige Ungenauigkeiten unmittelbar auf die externen Kräfte und Momente weitergegeben. Diese Einflüsse sorgen konkret dafür, dass bspw. bei einer Bewegung des Roboters im freien Raum, bei der keinerlei äußere Belastung auf den Roboter einwirkt, externe Kräfte und Momente aufgezeigt werden.

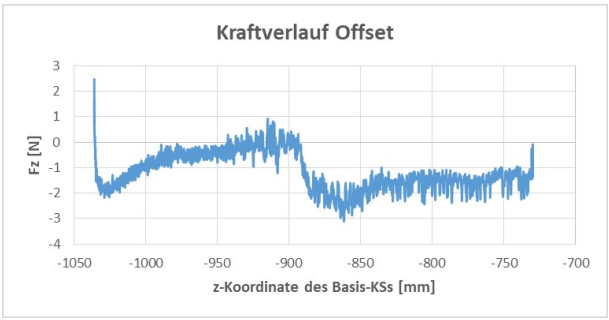


Abbildung 4 Offset (${}^{TCP}\underline{F}_z$) entlang der Trajektorie

Abbildung 4 zeigt die beschriebenen Ungenauigkeiten in Form eines Offset entlang einer definierten Trajektorie, aufgetragen in Abhängigkeit der z-Komponente des Basis-Koordinatensystems. Im Rahmen dieser Applikation wird eine maximale Abweichung von ca. 5,5 N erreicht, welche unter gewissen Umständen, nämlich bei hoch genauem applizieren von Kräften, nicht hinnehmbar ist. Folglich ist es wichtig, diese Ungenauigkeiten zu eliminieren bzw. möglichst stark zu kompensieren.

Des weiteren sei an dieser Stelle erwähnt, dass die Berech-

nung der externen Kräfte und Momente nach Gl. 8 ebenfalls von der Konfigurationen der kinematischen Struktur des Roboters abhängt. Allerdings existieren hier Ansätze in Form von Optimierungsalgorithmen, welche die Konfiguration der Kinematik unter Ausnutzung der Redundanz des KUKA LBR iiwas so anpassen, dass etwaige Ungenauigkeiten im Vorfeld ausgeschlossen werden [22].

3.3 Lösungsansatz - Parameteridentifikation

Ein Lösungsansatz die beschriebenen Ungenauigkeiten und Abweichungen der externen Kräfte und Drehmomente zu kompensieren liegt offensichtlich darin deren Ursache, nämlich die Modellierung, zu verbessern. Dazu existieren sog. Parameteridentifikations-Verfahren. Ziel dieser Verfahren ist es die realen kinematischen und dynamischen Parameter des Handhabungsgerätes zu identifizieren. Das heißt die Parameter werden durch mathematische Optimierungsalgorithmen bestimmt und schließlich im zugrundeliegenden Modell korrigiert. [1] Diese Methode ist allerdings zum einen mathematisch sehr aufwendig und zum anderen können einige Bauteile wie bspw. Kabel im Inneren der kinematischen Struktur des Roboter praktisch überhaupt nicht modelliert werden, da während der Bewegung des Roboter deren Position und Orientierung nicht gänzlich definiert ist. Außerdem ist i.d.R. externe Sensorik wie bspw. Lasertracker erforderlich, um die Modellabweichungen zu quantifizieren. Infolgedessen bietet sich ein mehr praktischer Ansatz in Form eines Offsets bzw. Look-up-tables an, welches die Diskrepanzen ausgleichen soll.

3.4 Lösungsansatz - statisches Offset

Eine aus der Praxis gängige Methode zur Erhöhung der Genauigkeit bspw. eines Messverfahrens liegt darin den messwertunabhängigen systematischen Fehler (Offset) durch einen Korrekturwert zu kompensieren [21]. Dies kann bei einer konstanten Soll-Ist-Abweichung relativ einfach umgesetzt werden.

Übertragen auf die Thematik des vorliegenden Beitrags, bedeutet dies, dass im ersten Schritt das Offset aufgenommen werden muss. Dies geschieht durch eine Referenzfahrt. Die Trajektorie der kraftgeregelten Roboterapplikation wird abgefahren und in deren Verlauf die berechneten externen Kräfte und Momente aufgenommen, allerdings interagiert der Roboter hier nicht mit dem Bauteil und befindet sich folglich im unbelasteten Zustand. Handelt es sich bei dem Offset schließlich um eine konstante Verschiebung zum Sollwert ($Offset_{soll} = \begin{pmatrix}0N & 0N & 0N & 0Nm & 0Nm & 0Nm\end{pmatrix}^T$), so kann die sensitive Roboterapplikation durch diesen Korrekturwert im Rahmen der Kraft-/ Momentenreglung berücksichtigt werden. Ein Beispiel diesbezüglich ist in [22] gegeben. Bei der dort vorgestellten Wickelapplikation handelt es sich um eine lineare Bewegung parallel zu einer Raumachse, weshalb eine nahezu konstante Abweichung auftritt, die wie zuvor beschrieben kompensiert werden kann.

4 Entwurf eines Offsets in einer dynamischen Umgebung

Handelt es sich allerdings bspw. um eine komplexe Freiformfläche, die durch eine kompliziertere mathematische Beschreibung der Bahn in Form eines Splines abgefahren werden muss, so ist ein statischer Korrekturwert nicht mehr ausreichend. Hier kommt ein Offset in einer dynamischen Umgebung zum Tragen, wodurch das Offset präzise für die gegebene Pose entlang der Bahn angepasst werden kann.

Das bedeutet wie zuvor in Kapitel 3.4 beschrieben dienen die im unbelasteten Zustand aufgezeichneten Messwerte entlang einer Trajektorie als Look-up-table und werden mit den für den Roboter gesetzten und als nominal geltenden Werten verrechnet, um somit die Ungenauigkeit der berechneten externen Kräfte und Drehmomente zu kompensieren. Im Einzelnen heißt dies beispielsweise, dass auf den Soll-Wert der Kraft das Offset aus dem Look-up-table addiert wird, um letztlich den gewünschten Betrag der Kraft zu erreichen.

Um nun die Kraftkomponenten des hybriden Kraft-/ Positionsregler während der Bewegung des Roboters zu variieren kommt die SmartServo-Umgebung des KUKA LBR iiwa [23] zum Einsatz, welche Änderungen dieser zur Laufzeit, ohne Abbruch der bestehenden Bewegung und folglich ohne Stoppen, erlaubt. Dabei wird unmittelbar auf den Feininterpolator des Roboters zugegriffen, was das Feld der Anwendungsmöglichkeiten, insbesondere sensitiver Roboteraufgaben, immens erweitert.

Zur Verdeutlichung des Prozesses ist dieser in Abb. 5 schematisch dargestellt.

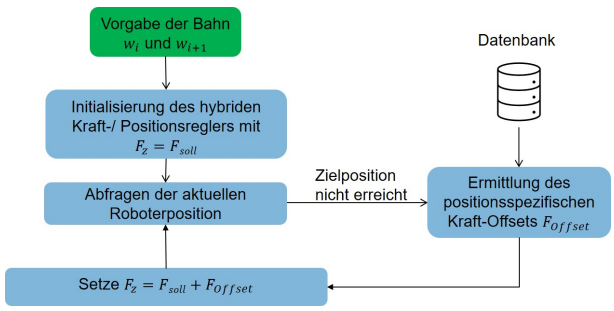


Abbildung 5 Schematische Darstellung der Implementierung eines Offsets in dynamischer Umgebung

5 Umsetzung und Validierung

Der in Kapitel 4 angeführten Entwurf eines Offsets in einer dynamischen Umgebung wird anhand eines praktischen Anwendungsfalls validiert. Für die Validierung des vorgestellten dynamischen Offsets in Kapitel 4 wurde ein Schleifprozess in Form eines Entgratens eines dreidimensionalen Bauteils gewählt. Bei der relevanten Gratbahn handelt es sich um eine Freiformfläche, welche mittels

einer geeigneten Spline-Interpolation abgefahren werden muss. Das korrespondierende Offset (vgl. Abb. 4) wurde wie mehrfach beschrieben aufgenommen und nach dem Schema aus Abb. 5 angewandt. Für F_{soll} wurde hier 15 N gewählt.

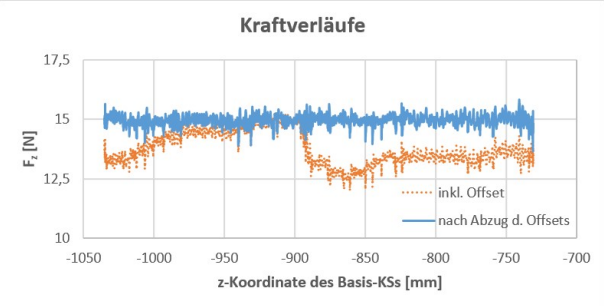


Abbildung 6 Validierung des Offsets in dynamischer Umgebung

Wie aus Abbildung 6 deutlich hervorgeht, ist es nach dem erläuterten Prozess möglich die Genauigkeit der auf ein Bauteil wirkenden Kraft signifikant zu erhöhen. Die orangene, gestrichelte Kurve zeigt den Kraftverlauf inklusive des Offsets und die blaue Kurve die tatsächlich applizierte Kraft nach Abzug des Offsets. Dies ist in Tabelle 2 statistisch belegt.

	Offset	*Nach Optimierung*
Spannweite	5,6 N	3,4 N
Standardabweichung	0,7 N	0,2 N

Tabelle 2 Statistischer Vergleich

6 Zusammenfassung und Ausblick

Durch den Einsatz sensitiver Robotersysteme eröffnen sich neue Anwendungsfelder im Kontext der Montage- und Fertigungstechnik. Allerdings können den hohen Erwartungen an Robotern insbesondere hinsichtlich ihrer Genauigkeit und Einsatzfähigkeit oft nicht Rechnung getragen werden. So treten bei kraftgeregelten Applikationen Diskrepanzen zwischen der tatsächlich beaufschlagten und der vorgegeben Kraft auf, welche nur schwerlich toleriert werden können, weshalb Maßnahmen der Verbesserung der kraftgeregelten Applikationen unabdingbar sind.

Durch praktische Versuche konnte die inhärente Ungenauigkeit des Roboters aufgrund seines kinematischen und dynamischen Modells aufgezeigt und durch eine Implementierung eines Offsets in einer dynamischen Umgebung im Rahmen einer in der Praxis typische Montageaufgabe kompensiert werden, was letztlich die Genauigkeit kraftgeregelter Applikationen erhöht.

Allerdings sollte an dieser Stelle abschließend kritisch angemerkt werden, dass jeder Offsetverlauf bzw. jeder Look-up-table für jede Trajektorie initial bestimmt werden muss, was mit einem gewissen Aufwand verbunden ist.

7 Literatur

[1] Müller, Rainer: *Verbesserung des kinematischen und dynamischen Bewegungsverhaltens von Handhabungsgeräten mit geschlossenen kinematischen Teilketten*. Dissertation, 1996.

[2] Wiest, Ulrich: *Kinematische Kalibrierung von Industrierobotern*. Dissertation 2001.

[3] Berthold Heinrich, Petra Linke, und Michael Glöckler: *Grundlagen Automatisierung: Erfassen - Steuern - Regeln*. Wiesbaden, Springer Vieweg, 2020.

[4] Birgit Vogel-Heuser, Thomas Bauernhansl, und Michael ten Hompel: *Handbuch Industrie 4.0: Bd. 1: Produktion*. Berlin, Springer Vieweg, 2017.

[5] Mathias Brandstötter, Damir Mirkovic, und Michael Hofbaur: *Mobile Manipulation - Eine altbekannte Technologie findet durch sensitive Robotertechnologie Einzug in die Industrie*. In C-AR2017 - Conference on Automation and Robotics 2017.

[6] KUKA AG, 2020.

[7] Universal Robots, 2020.

[8] Yaskawa, 2020.

[9] FRANKA EMIKA GmbH, 2020.

[10] FANUC Deutschland GmbH, 2020.

[11] DENSO Robotics, 2020.

[12] Doosan Robotics, 2020.

[13] Stefan Keil: *Dehnungsmessstreifen*. Wiesbaden, Springer Vieweg, 2017.

[14] Konradin-Verlag Robert Kohlhammer GmbH: *Roboter mit Tastsinn*. Internetdokument, 2016.

[15] Richard S. Hartenberg, Jacques Denavit: *A kinematic notation for lower pair mechanisms based on matrices*. Journal of applied mechanics, 1955.

[16] John J. Craig: *Introduction to robotics: Mechanics and control*. 2005.

[17] R. V. Patel, F. Shadpey: *Control of Redundant Robot Manipulators: Theory and Experiments*. Berlin Heidelberg, Springer-Verlag GmbH Vieweg, 2005.

[18] Wolfgang Weber: *Industrieroboter: Methoden der Steuerung und Regelung*. 2019.

[19] Bruno Siciliano, Oussama Khatib: *Springer Handbook of Robotics*. 2016.

[20] Winkler, Alexander: *Sensorgeführte Bewegungen stationärer Roboter*. Dissertation, 2015.

[21] A. Blum, R. Müller, M. Scholer, und A. Kanso *P3.6 Untersuchung eines Prozessroboters als Messroboter in der Montage*. AMA Service GmbH, Von-Münchhausen-Str. 49, 31515 Wunstorf, 2019.

[22] Kanso, Ali: *Methodik zur Realisierung einer sensitiven Montageaufgabe basierend auf dem Handhabungsgerät und intelligenter Sensorik*. Dissertation, 2021.

[23] KUKA Roboter GmbH: *KUKA Sunrise.Connectivity Servoing 1.7*. 2015.

Kraftsensitive Kalibriermethode für Industrieroboter
Force-sensitive calibration method for industrial robots

M. Sc. Stefan Schütz, B. Eng. Nikolaus Elsner, Dr.-Ing. Christian Henke, Prof. Dr.-Ing. Ansgar Trächtler
Fraunhofer-Institut für Entwurfstechnik Mechatronik IEM, 33102 Paderborn, Deutschland
{stefan.schuetz, nikolaus.elsner, christian.henke, ansgar.traechtler}@iem.fraunhofer.de

Kurzfassung

Moderne industrielle Fertigungsprozesse sind durch den Einsatz von Industrierobotern geprägt. Diese verlieren im Laufe ihrer Nutzungsdauer allerdings an Genauigkeit, etwa durch Nulllagenfehler, die aus Fertigungstoleranzen, Verschleiß und Crashfahrten resultieren. Zur Kompensation solcher Ungenauigkeiten und der damit einhergehenden Steigerung der Prozessqualität wird eine regelmäßige Roboterkalibration empfohlen. Der Beitrag stellt ein neuartiges kraftsensitives Verfahren im Bereich der Closed-Loop-Methoden vor, das die vollautomatische Kalibrierung von Industrierobotern zur Kompensation von Achsnulllagenfehlern ermöglicht.

Abstract

Modern industrial manufacturing processes are characterized by the use of industrial robots. Over the course of their service life, however, these robots lose accuracy, for example due to joint zero-position errors resulting from manufacturing tolerances, wear and tear and crash drives. In order to reduce such inaccuracies and to increase process quality, regular robot calibration is recommended. This paper presents a novel force-sensitive method in the field of closed-loop methods which allows the fully automatic calibration of industrial robots for the compensation of joint zero-position errors.

1 Einleitung

Die Roboterkalibrierung spielt heutzutage eine immer wichtiger werdende Rolle bei der Integration und dem Betrieb von Robotern in computerintegrierten Fertigungs- und Montagesystemen, da die Ergebnisse der Roboterkalibrierung zu erheblichen Genauigkeitsverbesserungen und damit zur Qualitätssteigerung der Produktionsprozesse bzw. zur Einsparung von Kosten führen können [1]. Die Roboterkalibrierung ermöglicht die Implementierung von zuvor offline geplanten und simulierten Roboteraufgaben. Dafür müssen die Abmessungen der realen und der virtuellen Roboter, die in den Offline-Programmierumgebungen zur Generierung der Roboterprogramme eingesetzt werden, nahezu identisch sein. Andernfalls führen die Unterschiede in den Abmessungen zu Fehlern bei der Roboterpositionierung während der Ausführung des Roboterprogramms im Produktionsbetrieb. Darüber hinaus ist die Kalibrierung für den generellen Betrieb von Industrierobotern wesentlich, da der Roboter-Controller ein kalibriertes Steuerungsmodell der Kinematik zur Berechnung der Vorwärts- und Rückwärtstransformationen bei der genauen Positionierung des Endeffektors voraussetzt.

2 Stand der Technik

Im Folgenden soll ein kurzer Überblick über die Grundlagen zur Modellbildung von Industrierobotern mit serieller Kinematik sowie über bestehende Methoden der Kalibrierung gegeben werden, um die darauf vorgestellte Kalibriermethode einordnen zu können.

2.1 Modellierung offener kinematischer Ketten

Es sei $\boldsymbol{q} = [q_1, \ldots, q_n]^T \in \mathbb{R}^n$ der Vektor der Gelenkkoordinaten eines Roboters mit n Freiheitsgraden. Die Pose des Roboter-Endeffektors wird durch den Vektor $\boldsymbol{r}_\mathrm{E} = [x, y, z, \alpha, \beta, \gamma]^T$ dargestellt, wobei x, y, z die Positionen im Basiskoordinatensystem und α, β, γ die Orientierung in Minimaldarstellung beschreiben. Der Ursprung des kartesischen Endeffektorkoordinatensystems wird auch als *Tool Center Point* (TCP) bezeichnet [2]. Zur Ermittlung der TCP-Pose $\boldsymbol{r}_\mathrm{E}$ auf Basis der bekannten achsspezifischen Gelenkkoordinaten $\boldsymbol{q}$ wird die kinematische Struktur des Roboters durch eine Folge von Transformationen vom Basis- zum TCP-Koordinatensystem modelliert. Die relative Stellung des Koordinatensystems $\boldsymbol{K}_i$ zum benachbarten Koordinatensystem $\boldsymbol{K}_{i-1}$ wird durch eine homogene Transformationsmatrix $\boldsymbol{T}_i$ beschrieben:

$$\boldsymbol{K}_i = \boldsymbol{T}_i \cdot \boldsymbol{K}_{i-1} \text{ mit} \tag{1}$$

$$\boldsymbol{T}_i = \begin{bmatrix} \boldsymbol{R}_i & \boldsymbol{t}_i \\ \boldsymbol{0} & 1 \end{bmatrix} \in \mathbb{R}^{4\times4}, \tag{2}$$

wobei $\boldsymbol{R}_i \in \mathbb{R}^{3\times3}$ die Rotationsmatrix und $\boldsymbol{t}_i \in \mathbb{R}^{3\times1}$ den Verbindungsvektor zwischen den beiden Koordinatensystemen darstellt. Die Multiplikation der n Transformationsmatrizen sowie der Transformationsmatrix $\boldsymbol{T}_\mathrm{Tool}$ vom Handflansch- zum Endeffektorkoordinatensystem führt schließlich zur Beschreibung der Endeffektorpose in Form der homogenen Transformationsmatrix

$$\boldsymbol{T}_\mathrm{E} = \boldsymbol{T}_1 \cdot \boldsymbol{T}_2 \cdot \ldots \cdot \boldsymbol{T}_n \cdot \boldsymbol{T}_\mathrm{Tool}. \tag{3}$$

Mit der Konvention nach DENAVIT-HARTENBERG (DH) wird eine Einzeltransformation $\boldsymbol{T}_i$ durch

$$\boldsymbol{T}_i(\theta_i, d_i, a_i, \alpha_i) = \boldsymbol{R}_z(\theta_i)\boldsymbol{T}_z(d_i)\boldsymbol{T}_x(a_i)\boldsymbol{R}_x(\alpha_i) \tag{4}$$

mit den vier Parametern θ_i, d_i, a_i und α_i beschrieben, wobei $\boldsymbol{T}_x$ und $\boldsymbol{T}_z$ reine Translationen entlang der x- und z-Achse sowie $\boldsymbol{R}_x$ und $\boldsymbol{R}_z$ reine Rotationen um die x- und die z-Achse darstellen. Bei Drehgelenken stellt θ_i, bei Schubgelenken d_i die zeitveränderliche Gelenkkoordinate q_i dar, die restlichen drei Parameter sind konstant [3]. Nach einer Überführung in die Minimaldarstellung folgt aus (3) die auch als Vorwärtstransformation bezeichnete Gleichung der direkten Kinematik

$$\boldsymbol{r}_E = \boldsymbol{f}(\boldsymbol{q}). \tag{5}$$

Bei der Lösung des inversen kinematischen Problems

$$\boldsymbol{q} = \boldsymbol{f}^{-1}(\boldsymbol{r}_E) \tag{6}$$

werden die Gelenkkoordinaten $\boldsymbol{q}$ bestimmt, die zu einer gegebenen Endeffektorpose $\boldsymbol{r}_E$ führen. Diese nichtlineare Rückwärtstransformation ist i. A. nicht eindeutig lösbar [2].

Die Differentialkinematik erster Ordnung mit der analytischen Jacobi-Matrix $\boldsymbol{J}(\boldsymbol{q}) = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{q}} \in \mathbb{R}^{m \times n}$ bildet die Gelenkgeschwindigkeiten $\dot{\boldsymbol{q}}$ auf den Vektor $\dot{\boldsymbol{r}}_E$ ab, der die Lineargeschwindigkeiten sowie die zeitlichen Änderungen der Orientierung des TCP umfasst:

$$\dot{\boldsymbol{r}}_E = \boldsymbol{J}(\boldsymbol{q})\dot{\boldsymbol{q}}. \tag{7}$$

Zur Bestimmung der erforderlichen Gelenkgeschwindigkeiten $\dot{\boldsymbol{q}}$, die der Roboter ausführen muss, um einer in Weltkoordinaten gegebenen Endeffektor-Trajektorie $\dot{\boldsymbol{r}}_E$ zu folgen, kann die Rückwärtstransformation auf Basis der inversen Differentialkinematik erfolgen:

$$\dot{\boldsymbol{q}} = \boldsymbol{J}^{-1}(\boldsymbol{q})\dot{\boldsymbol{r}}_E. \tag{8}$$

Dafür sei vorausgesetzt, dass die Jacobi-Matrix $\boldsymbol{J}(\boldsymbol{q}) \in \mathbb{R}^{m \times n}$ mit $m = n$ quadratisch und regulär ist und die inverse Matrix $\boldsymbol{J}^{-1}(\boldsymbol{q})$ somit existiert. In singulären Stellungen des Roboterarms verliert $\boldsymbol{J}(\boldsymbol{q})$ an Rang und Gleichung (8) führt zu unendlich hohen Gelenkgeschwindigkeiten. Darüber hinaus tritt bereits in der Nähe zu kinematischen Singularitäten eine Bewegungseinschränkung des Endeffektors in einer oder mehreren Richtungen auf. Daher gilt die Vermeidung kinematischer Singularitäten als eine der Hauptaufgaben bei der Steuerung von Robotern [4].

2.2 Kalibrierung von Industrierobotern

Generell kann die industrielle Kalibrierung nach SCHRÖER [5] und GEUENS ET AL. [6] in die Roboter- und Zellenkalibrierung unterteilt werden. Letztere umfasst dabei die Vermessung von Bauteilen und Werkzeug sowie der Umgebung des Robotersystems, um die im Rahmen der Offline-Programmierung benötigte virtuelle Roboterzelle an die Realität anzupassen. Die Roboterkalibrierung, die zur softwareseitigen Verbesserung der Positioniergenauigkeit dient, unabhängig vom verwendeten Werkzeug oder sonstiger Peripherie, wird nach ROTH ET AL. [7] in drei Stufen eingeteilt: Die *Level 1-Kalibrierung* umfasst die Bestimmung des korrekten Zusammenhangs zwischen dem Signal der Gelenkwegaufnehmer sowie der tatsächlichen Gelenkverschiebung. Dies beinhaltet i. d. R. die Kalibrierung der Kinematik des Antriebs und der Gelenksensorik. Die *Level 2-Kalibrierung* umfasst die Kalibrierung des gesamten kinematischen Modells des Roboters, also die Bestimmung der kinematischen Grundgeometrie sowie der korrekten Gelenkwinkelbeziehungen. Die *Level 3-Kalibrierung* wird als nicht-kinematische Kalibrierung bezeichnet, wobei nicht-geometrische Effekte wie Nachgiebigkeiten, Spiel und Reibung in den Gelenken sowie Parameter im dynamischen Modell des Roboters betrachtet werden.

Die Mehrheit aller Veröffentlichungen betrachtet die kinematische Kalibrierung von Robotern, also die Level 2-Kalibrierung [1]. In diesem Bereich wird nach [8] weiter zwischen *Open-Loop*- und *Closed-Loop*-Methoden unterschieden. Die Open-Loop-Methoden erfordern ein externes Messsystem, um die Pose des Roboterhandflanschs bzw. des daran befestigten Endeffektors zu vermessen. Dabei kann die Anzahl der gemessenen Posekomponenten von sechs (volle Pose mit drei Positionen und drei Orientierungen) bis hin zu nur einer einzigen Posekomponente variieren. Durch die Bewegung aller Gelenke wird der Roboter in einer Reihe von Stellungen frei positioniert („open loop“) und die resultierenden Posen vermessen. Anschließend werden die Parameter des kinematischen Modells im Rahmen einer nichtlinearen Optimierung der Gesamtmenge aller Posen identifiziert [9]. Bei den Closed-Loop-Methoden wird der Endeffektor des Roboters so mit der Umgebung fixiert, dass sich eine *bewegliche* kinematische Kette bildet. Die Kalibrierung erfolgt anschließend durch die Vermessung der Gelenkpositionen mit dem internen Messsystem des Roboters ohne ein externes Messsystem [10]. Die Zwangsbeschränkung des Endeffektors kann dabei von voller Steifigkeit (sechs Beschränkungen, bei kinematisch redundanten Systemen) bis hin zu nur einer Bewegungsbeschränkung variieren. Die Schätzung der Parameter erfolgt wie bei den Open-Loop-Methoden.

Weiterhin wird zwischen *lokalen* und *globalen* Ansätzen der Roboterkalibrierung unterschieden. Bei den lokalen Verfahren werden diskrete Messungen der TCP-Position durch sequentielles Abfahren eines 3D-Gitters im relevanten Arbeitsbereich des Roboters durchgeführt [11]. Resultierende Abweichungen in Form von Positionierfehlern werden als Korrekturwerte in der Steuerung des Roboters hinterlegt. Das Verfahren wird überwiegend bei Werkzeugmaschinen eingesetzt, da diese i. d. R. über genau definierte, am kartesischen Koordinatensystem ausgerichtete Achsen verfügen, und sich Betrag und Richtung der berechneten Korrekturwerte nicht ändern. Bei Industrierobotern mit Knickarm-Kinematik und meist sechsdimensionalem Arbeitsraum gestaltet sich dies deutlich schwieriger. Die globalen Verfahren wiederum nutzen ein Modell der nominalen Kinematik des Roboters, das schrittweise um weitere Einflussparameter wie Achsabstände und -schiefstände sowie Elastizitäten u. s. w. erweitert wird und das Verhalten des Roboters im gesamten Arbeitsbereich beschreiben

soll. Nach der Modellierung erfolgt ein Abgleich des nominalen Modells mit dem realen Roboter, wobei die Positionierabweichungen an vorgegebenen Kalibrierpositionen gemessen und durch die Parameteridentifikation minimiert werden. Zur Steigerung der Absolutgenauigkeit können die identifizierten Modellparameter daraufhin unmittelbar in der Steuerung des Roboters hinterlegt werden oder zur einmaligen Korrektur der anzufahrenden Roboterposen mittels einer Kalibriersoftware genutzt werden [12].

2.3 Anforderungen an die Methode

BEYER gibt in [11] einen Überblick über die Fehlereinflüsse auf die Genauigkeit von Industrierobotern und fasst verschiedene Untersuchungen zur Relevanz von Fehlerquellen zusammen. Die Vermessung der Lage des Werkstücks im Arbeitsraum sowie des TCP gegenüber dem Handflansch des Roboters stellen demnach die größten Fehlerquellen für die Posegenauigkeit bei der Offline-Programmierung dar. Bezüglich des Roboters selbst sind die größten systematischen Fehler auf die Nullstellungen der Gelenke zurückzuführen. Diese Nulllagenfehler bilden einen Anteil von 80 bis 90 % der geometrischen Fehlereinflüsse, die aus Justagefehlern, Fertigungs- und Montagetoleranzen sowie aus Beschädigungen durch Crash-Fahrten resultieren und zu Abweichungen zwischen dem kinematisch-geometrischen Modell und dem realen Roboter führen.
Der Großteil der bekannten Ansätze zur kinematisch-geometrischen Kalibrierung (Level 1 bis 2) weist nach [1] die folgenden Nachteile auf:

- Für den ordnungsgemäßen Betrieb der Messgeräte ist geschultes Personal erforderlich.
- Die Datenerhebung ist ermüdend, zeitaufwendig und schwierig zu automatisieren.
- Die Messtechniken sind hauptsächlich für die Roboterkalibrierung in Laborumgebung auslegt.
- Der Messaufbau und die -verfahren erfordern viel menschliches Eingreifen, was sie für die Vor-Ort-Kalibrierung von Robotern in einer industriellen Umgebung ungeeignet macht.

Die im Folgenden vorgestellte Kalibriermethode wird also gänzlich auf die Kompensation der Nulllagenfehler zugeschnitten, was bereits eine deutliche Reduzierung der Posefehler des Roboters verspricht. Darüber hinaus ermöglicht das Verfahren den vollautomatischen Einsatz im industriellen Umfeld ohne menschliches Eingreifen, um die genannten Nachteile der bekannten Verfahren zu vermeiden.

3 Methode zur kraftsensitiven Roboterkalibrierung

Die auf kinematischen Modellen basierende Kalibrierung wird i. A. als globale Kalibrierungsmethode betrachtet, die die Genauigkeit des Roboters über das gesamte Volumen des Arbeitsraums verbessert, und besteht nach ELATTA ET AL. [1] aus den vier aufeinander folgenden Schritten

- Modellierung,
- Messung,
- Identifikation und
- Kompensation oder Korrektur.

Zur Modellierung der nominalen Kinematik wird die DH-Konvention (siehe Kap. 2.1) verwendet. Dabei werden die vier Parameter θ_i, d_i, a_i und α_i um einen Einflussparameter $q_{\mathrm{Off},i}$ zur Beschreibung der Nulllagenfehler in Form von Gelenkoffsets erweitert, die im Rahmen der Kalibrierung ermittelt werden sollen. In allgemeiner Form kann die Berücksichtigung von Gelenkoffsets $\boldsymbol{q}_{\mathrm{Off}}$ bei der direkten Kinematik als

$$\boldsymbol{r}_{\mathrm{E}} = \boldsymbol{f}(\boldsymbol{q}, \boldsymbol{q}_{\mathrm{Off}}) = \boldsymbol{f}(\boldsymbol{q} + \boldsymbol{q}_{\mathrm{Off}}) \tag{9}$$

und bei der inversen Kinematik als

$$\boldsymbol{q} = \boldsymbol{f}^{-1}(\boldsymbol{r}_{\mathrm{E}}) + \boldsymbol{q}_{\mathrm{Off}} \tag{10}$$

formuliert werden [13].

3.1 Ansatz und Messaufbau

Der grundlegende Ansatz der kraftsensitiven Kalibriermethode zur Kompensation der Nulllagenfehler beruht darauf, den Roboter ohne Kenntnis der korrekten Gelenkpositionen q_i reproduzierbar in eine bekannte, eindeutig definierte Kalibrierpose $\boldsymbol{r}_{\mathrm{E,K}}$ zu bringen. Dafür wird ein Pyramidenstumpf-förmiger Kalibrierdorn am Roboterhandflansch, zusätzlich zu dem vom Kalibrierprozess unabhängigen Endeffektor, angebracht. Ein sogenanntes Kalibriertarget, das die Negativform zum Kalibrierdorn bildet, wird auf einem 6D-Kraft-Momenten-Sensor befestigt und im Arbeitsraum des Roboters positioniert. Das taktile Zusammenführen und der resultierende Formschluss zwischen Kalibrierdorn und -target sorgen somit für eine eindeutige Kalibrierpose des Endeffektors aufgrund der konstruktiven Gestaltung der Kalibrierpartner. Der Messaufbau ist in **Abbildung 1** dargestellt und zeigt den Roboter UR10 der Fa. Universal Robots, der ebenfalls für die in Kap. 4 beschriebene Validierung genutzt wird, in der Kalibrierpose $\boldsymbol{r}_{\mathrm{E,K}}$.

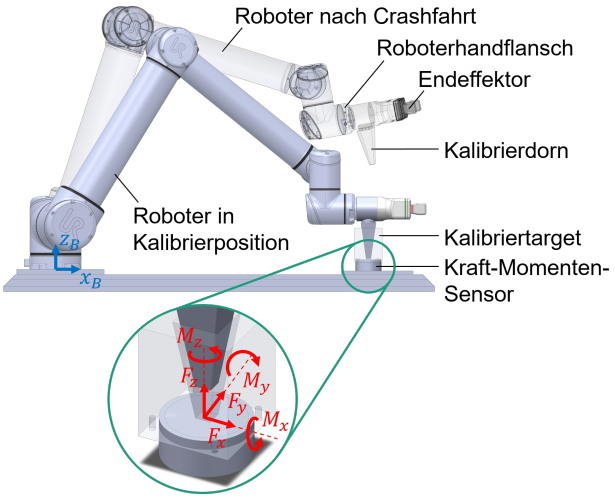


Abbildung 1 Messaufbau zur Roboterkalibrierung einschließlich der Kräfte und Momente am 6D-Kraft-Momenten-Sensor

Die initiale Konfiguration an korrekten Gelenkwinkeln (nach Werkskalibrierung), die zu dieser Kalibrierpose führen, wird einmalig im Vektor $\boldsymbol{q}_{\mathrm{K}}$ abgespeichert. Verliert der Roboter während des Betriebs, etwa aufgrund einer Crashfahrt, seine Gelenkpositionen, muss der Roboter wieder in die Kalibrierpose gebracht werden. Dort führt der entstandene Nulllagenfehler zu einer Differenz

$$\boldsymbol{q}_{\Delta} = \boldsymbol{q}_{\mathrm{K}} - \boldsymbol{q} \tag{11}$$

zwischen der initial abgespeicherten Konfiguration $\boldsymbol{q}_{\mathrm{K}}$ und der mit den internen Winkelaufnehmern gemessenen Gelenkposition $\boldsymbol{q}$. Die anschließende Korrektur der somit identifizierten Fehler erfolgt durch eine Anpassung des eingeführten Gelenkoffsets $\boldsymbol{q}_{\mathrm{Off}}$ auf dem Roboter-Controller.

3.2 Automatische Vermessung und Identifikation der Nulllagenfehler

Für den praktischen Einsatz der Kalibriermethode im industriellen Umfeld soll der Vorgang des taktilen Zusammenführens von Kalibrierdorn und -target automatisch erfolgen. Dafür wird der in **Abbildung 2** dargestellte Gesamtablauf der Kalibriermethode um die initiale Phase der Kontaktfindung und die darauffolgende Phase der Kontaktregelung erweitert.

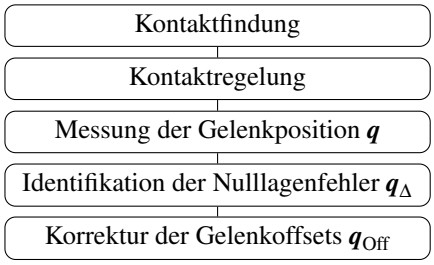


Abbildung 2 Ablauf des automatischen Kalibriervorgangs

Zur Kontaktfindung startet der Roboter aus einer Initialkonfiguration $\boldsymbol{q}_{\mathrm{Init}}$ mit den an dieser Stelle noch unbekannten Nulllagenfehlern $\boldsymbol{q}_{\Delta}$. Die resultierende Initialpose des Kalibrierdorns $\boldsymbol{r}_{\mathrm{E,Init}}$ muss sich dabei in einem gewissen „Einzugsbereich“, der durch den Öffnungswinkel der Negativform des Pyramidenstumpfs gebildet wird, über dem Kalibriertarget befinden, sodass der Kontaktfall gewährleistet werden kann. Außerdem darf die finale Kalibrierpose $\boldsymbol{r}_{\mathrm{E,K}}$ nicht durch eine zu $\boldsymbol{q}_{\mathrm{K}}$ alternative Gelenkstellung der Mehrfachlösungen der inversen Kinematik des Roboters erreicht werden (vgl. z. B. Konfiguration „Ellbogen unten“ statt „Ellbogen oben“ in [2]). Dies würde zu Fehlern bei der finalen Berechnung der Gelenkoffsets führen und ist daher unbedingt zu vermeiden. Der Kalibrierdorn wird anschließend durch eine lineare Endeffektorbewegung mit konstanter Geschwindigkeit in Richtung des Kalibriertargets bewegt. Überschreitet die gemessene Kraft F_z einen gewissen Schwellwert, gilt die Phase der Kontaktfindung als abgeschlossen.

Während der nächsten Phase sollen die im Kraftwinder $\boldsymbol{F} = [F_x, F_y, F_z, M_x, M_y, M_z]^T$ zusammengefassten Kontaktkräfte und -momente zwischen Kalibrierdorn und -target geregelt werden. Die Rückführung der sechs Messgrößen des Kraft-Momenten-Sensors erlaubt den Entwurf einer expliziten Kraft-Momentenregelung mit unterlagerter Geschwindigkeitsregelung des Endeffektors. Aufgrund der vorliegenden Kontaktgeometrie liegt ein verkoppeltes Mehrgrößensystem vor, da z. B. eine Kontaktkraft in x-Richtung über den anliegenden Hebelarm zwischen der Kontaktstelle und dem Kraft-Momenten-Sensor zu einem Drehmoment M_y führt (vgl. Abb. 1). Beim Entwurf der dezentralen Regelungen werden diese Verkopplungen der einzelnen Teilsysteme allerdings nicht berücksichtigt und der Entwurf der Teilregler erfolgt an den isolierten Teilsystemen. Dieses Verfahren erlaubt den Einsatz der bekannten PI-Standardregler und eignet sich daher für den industriellen Einsatz [14]. Die Auslegung der $i = 1,\ldots,6$ Einfachregelkreise mit den Übertragungsbeiwerten $k_{\mathrm{R},i}$ und den Nachstellzeiten $T_{\mathrm{n},i}$ der Übertragungsfunktionen

$$G_{\mathrm{K},i}(s) = k_{\mathrm{R},i}\left[1 + \frac{1}{T_{\mathrm{n},i} \cdot s}\right] \tag{12}$$

der PI-Regler erfolgt mit den Einstellregeln nach ZIEGLER und NICHOLS [15]. Die Sollwerte der Kraft-Momentenregelung werden mit Ausnahme der Vorgaben $F_{z,Soll} = -10\,\mathrm{N}$ und $M_{z,Soll} = 0{,}1\,\mathrm{Nm}$, die den Kontakt zwischen Kalibrierdorn und -target gewährleisten sollen, gleich Null gesetzt. Das Blockschaltbild der internen und externen Roboterregelung wird in **Abbildung 3** dargestellt. Aufgrund nichtlinearer Effekte, wie etwa der Haftreibung, erreicht der Kalibrierdorn in ersten praktischen Versuchen nicht immer eindeutig die gewünschte Endlage. Daher wird ein Toleranzband um die jeweiligen Sollwerte eingeführt, dessen Erreichen für einen erfolgreichen Abschluss der Kontaktregelphase sorgt.

3.3 Optimale Positionierung des Kalibriertargets

Durch die Beschränkung des Kalibrierverfahrens auf die Kompensation von Nulllagenfehlern bei 6-Achs-Knickarmrobotern reduziert sich die Anzahl der notwendigen Messposen auf 1 und die nichtlineare Optimierung zur Identifikation weiterer Parameter entfällt. Diese treibt die Anzahl der notwendigen Messposen üblicherweise in die Höhe, um den Einfluss von Messfehlern und stochastischen Störungen auf das Identifikationsergebnis zu reduzieren. Zusätzlich reicht es bei den globalen Verfahren der Roboterkalibrierung aus, lediglich einen Teilbereich des Arbeitsraums zu betrachten, da sich die Ergebnisse i. A. auf die restlichen Bereiche übertragen lassen.

Bei der Verwendung einer einzigen Messpose ist allerdings deren Lage entscheidend. Daher sollte die Messpose nicht in der Nähe zu Singularitäten des Roboters liegen [11]. Eine Vielzahl von Veröffentlichungen wie [16, 17] hat sich bereits mit der optimalen Positionierung von Messposen beschäftigt. Im Rahmen des hier vorgestellten Kalibrierverfahrens soll das Maß der Manipulierbarkeit nach YOSHIKAWA [18] herangezogen werden, das die Distanz einer Gelenkstellung $\boldsymbol{q}$ zu kinematischen Singularitäten charakterisiert:

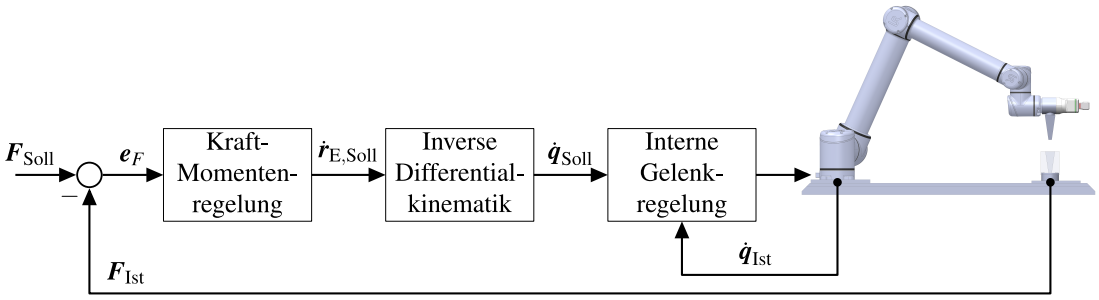


Abbildung 3 Blockschaltbild der internen und externen Regelung des Roboters

$$H_{\text{man}}(\boldsymbol{q}) = \sqrt{\det[\boldsymbol{J}(\boldsymbol{q})\boldsymbol{J}(\boldsymbol{q})^T]}. \tag{13}$$

Nimmt das Manipulierbarkeitsmaß den Wert $H_{\text{man}}(\boldsymbol{q}) = 0$ an, so befindet sich der Roboter mit seiner Gelenkkonfiguration in einer Singularität. Zur Positionierung des Kalibriertargets optimal sind daher Posen im Arbeitsraum, die der Roboter mit maximaler Manipulierbarkeit erreichen kann. **Abbildung 4** zeigt die Werte des kinematischen Manipulierbarkeitsmaßes in der $x_B y_B$-Ebene des UR10 auf der Höhe $z_B = 0\,\text{m}$ (vgl. Montagehöhe des Kraft-Momenten-Sensors in Abb. 1). Aufgrund der kinematischen Struktur des UR10, dessen erstes Gelenk q_1 mit einem Offset in z_B-Richtung unmittelbar über dem Basiskoordinatensystem bei $x_B = 0\,\text{m}, y_B = 0\,\text{m}$ liegt, verlaufen die Werte der Manipulierbarkeit konzentrisch um den Ursprung. So zeichnet sich der Rand des Arbeitsbereichs mit einem Radius von $1,3\,\text{m}$ ab, auf dem die äußeren kinematischen Singularitäten liegen und die Manipulierbarkeit sichtlich abfällt. Die Optimalstellen zur Positionierung des Kalibriertargets liegen auf einem etwas geringeren Radius.

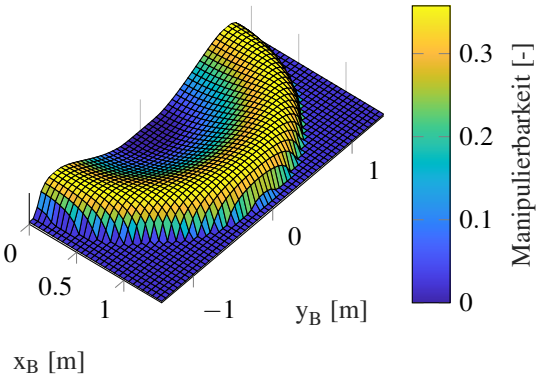


Abbildung 4 Manipulierbarkeit des UR10 in $x_B y_B$-Ebene bei der Höhe $z_B = 0\,\text{m}$

4 Validierung und quantitative Bewertung

Zur Eignung des Verfahrens ist es notwendig, dass die finale Kalibrierpose $\boldsymbol{r}_{\text{E,K}}$ aus verschiedenen unbekannten Startposen $\boldsymbol{r}_{\text{E,Init}}$ zuverlässig und reproduzierbar erreicht wird. Daher soll die Wiederholgenauigkeit zur Validierung und quantitativen Bewertung des Verfahrens untersucht werden. Die Steuerungs- und Regelungsfunktionen werden hierzu auf einer SPS der Fa. Beckhoff implementiert, die mit dem Roboter-Controller des UR10 sowie der Auswerteeinheit des verwendeten 6D-Kraft-Momenten-Sensors der Fa. ATI über eine auf TCP/IP-basierende echtzeitfähige Datenschnittstelle kommuniziert. Je nach Roboter und entsprechendem Controller kann die inverse Differentialkinematik (siehe Abb. 3) ebenfalls auf der SPS implementiert werden, um dem Roboter-Controller die Gelenkgeschwindigkeiten $\dot{\boldsymbol{q}}_{\text{Soll}}$ vorzugeben. Alternativ können die am Reglerausgang vorliegenden Sollwerte der Endeffektorgeschwindigkeiten $\dot{\boldsymbol{r}}_{\text{E,Soll}}$ direkt an den Roboter-Controller gesendet werden und die inverse Differentialkinematik wird intern berechnet.

Entsprechend der Richtlinie VDI 2861 zur Prüfung der kinematischen Kenngrößen von Industrierobotern wird die Standardabweichung s mit

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{14}$$

und dem arithmetischen Mittelwert $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ der Messwerte x_i in $n = 30$ Messungen betrachtet. Dabei werden die Startposen des Kalibrierdorns über dem Kalibriertarget mit max. Positions- und Winkelabweichungen von $20\,\text{mm}$ und $30\,°$ zufällig um die Initialpose $\boldsymbol{r}_{\text{E,Init}}$ vorgegeben. Die Ergebnisse der Messreihe sind in **Tabelle 1** dargestellt.

Entsprechend der technischen Spezifikation des Herstellers beläuft sich die Wiederholgenauigkeit des UR10 auf einen Wert von $\pm 0,1\,\text{mm}$. Damit liegt die maximale Standardabweichung des Kalibrierverfahrens unter der Wiederholgenauigkeit des Roboters und die vorgestellte kraftsensitive Kalibriermethode zur Kompensation von Nulllagenfehlern kann als validiert betrachtet werden.

Tabelle 1 Standardabweichungen der Gelenkwinkel sowie der Posekomponenten des Endeffektors

Gelenk	*Standardabw.*	*Posekomp.*	*Standardabw.*
q_1	$0,0103\,°$	x	$0,0898\,\text{mm}$
q_2	$0,0000\,°$	y	$0,0572\,\text{mm}$
q_3	$0,0143\,°$	z	$0,0224\,\text{mm}$
q_4	$0,0634\,°$	α	$0,0606\,°$
q_5	$0,0699\,°$	β	$0,0510\,°$
q_6	$0,0295\,°$	γ	$0,0176\,°$

5 Zusammenfassung und Ausblick

Der Beitrag stellt ein neuartiges kraftsensitives Kalibrierverfahren vor, das in den Bereich der Closed-Loop-Methoden eingeordnet werden kann. Entgegen der bekannten Verfahren in diesem Bereich kann auf eine feste Fixierung des Endeffektors mit der Umgebung verzichtet werden, indem die geschlossene kinematische Schleife durch eine Kraft-Momentenregelung zur automatischen Kontaktfindung zwischen einem Kalibrierdorn und einem Kalibriertarget hergestellt wird. Aufgrund der konstruktiven Gestaltung der Kalibrierkörper erreicht der Roboter eine eindeutige Kalibrierpose, in der die Nulllagenfehler identifiziert und als Korrekturwerte in Form von Gelenkoffsets in das Steuerungsmodell des Roboters integriert werden können. Der Ansatz wurde bereits erfolgreich in einem Industrieprojekt umgesetzt. Dabei traten regelmäßig Nulllagenfehler in den Hauptachsen der werkzeugführenden Industrieroboter nach prozessbedingten Kontakten mit dem Werkstück auf. Die Vollautomatisierung des Messvorgangs erlaubte die Durchführung regelmäßiger Wartungsintervalle zur Kompensation der Nulllagenfehler.
Die Methode zeichnet sich darüber hinaus durch den Einsatz kostengünstiger Messmittel aus. **Abbildung 5** stellt die Positionierung des Kalibriertargets im gemeinsamen Arbeitsbereich mehrerer Industrieroboter dar, was eine weitere Kostenersparnis ermöglicht.

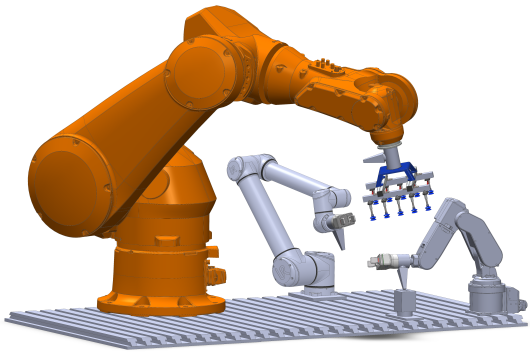


Abbildung 5 Positionierung des Kalibriertargets im gemeinsamen Arbeitsbereich mehrerer Industrieroboter

Weiteres Optimierungspotential wird in der zentralen Regelung der Kräfte und Momente mit der Berücksichtigung von deren Verkopplung während der Kontaktregelung anstelle der isolierten Festwertregelungen erwartet.

6 Literatur

[1] Elatta, A. Y.; Gen, L. P.; Zhi, F. L.; Daoyuan, Y.; Fei, L.: *An Overview of Robot Calibration*. In: Information Technology Journal 3.1 (2004), S. 74-78.

[2] Weber, W.: *Industrieroboter: Methoden der Steuerung und Regelung*. 2019, ISBN 9783446460607.

[3] Denavit, J.; Hartenberg, R. S.: *A kinematic notation for lower-pair mechanisms based on matrices*. In: Journal of Applied Mechanics 22 (1955), S. 215-221.

[4] Wenz, M.: *Automatische Konfiguration der Bewegungssteuerung von Industrierobotern*. Diss., Univ. Karlsruhe, 2008, ISBN 9783832520571.

[5] Schröer, K.: *Improvement of Robot Industrial Standardisation: IRIS*. In: 8th International Symposium on Measurement and Control in Robotics, Czech Technical University, 1998.

[6] Geuens, F.; van den Bossche, J.; Blanckaert, P.; van den Bossche, A.: *Standardization of Robotic Cell Calibration: The Key to Successful Integration of OLP.* In: Proceedings of the International Symposium on Industrial Robots, Detroit, 1997.

[7] Roth, Z.; Mooring, B.; Ravani, B.: *An overview of robot calibration.* In: IEEE Journal on Robotics and Automation 3.7 (1987), ISSN 08824967, S. 377-385.

[8] Hollerbach, J. M.; Wampler, C. W.: *The Calibration Index and Taxonomy for Robot Kinematic Calibration Methods*. In: The International Journal of Robotics Research 15.6 (1996), ISSN 0278-3649, S. 573-591.

[9] Mooring, B. W.; Roth, Z. S.; Driels, M. R.: *Fundamentals of Manipulator Calibration*. New York: Wiley, 1991.

[10] Bennett, D. J.; Hollerbach, J. M.: *Autonomous calibration of single-loop closed kinematic chains formed by manipulators with passive endpoint constraints*. In: IEEE Transactions on Robotics and Automation 7.5 (1991), ISSN 1042-296X, S. 597-606.

[11] Beyer, L.: *Genauigkeitssteigerung von Industrierobotern*. Diss., Aachen: Shaker, 2005.

[12] Bongardt, T.: *Methode zur Kompensation betriebsabhängiger Einflüsse auf die Absolutgenauigkeit von Industrierobotern*. Diss., Techn. Univ. München, 2003.

[13] Corke, P.: *Robotics, vision and control: Fundamental Algorithms In MATLAB*. Cham: Springer, ISBN 978-3-319-54412-0, 2017.

[14] Angermann, A.: *Entkopplung von Mehrgrößensystemen durch Vorsteuerung am Beispiel von kontinuierlichen Fertigungsanlagen*. Diss., Techn. Univ. München, 2004.

[15] Föllinger, O.; Konigorski, U.; Lohmann, B.; Roppenecker, G.; Trächtler, A.: *Regelungstechnik: Einführung in die Methoden und ihre Anwendung*. Berlin, Offenbach: VDE Verlag GmbH, ISBN 978-3-8007-4202-3 8, 2016.

[16] Borm, J.-H.; Meng, C.-H.: *Determination of Optimal Measurement Configurations for Robot Calibration Based on Observability Measure*. In: The International Journal of Robotics Research 10.1 (1991), ISSN 0278-3649, S. 51-63.

[17] Nahvi, A.; Hollerbach, J. M.: *The noise amplification index for optimal pose selection in robot calibration*. In: IEEE International Conference on Robotics and Automation (1996), ISBN 0-7803-2988-0, S. 647-654.

[18] Yoshikawa, T.: *Manipulability of Robotic Mechanisms*. In: The International Journal of Robotics Research 4.2 (1985), ISSN 0278-3649, S. 3-9.

Entwicklung einfacher modularer Seilroboter ohne Verwendung von Kraftsensorik

Development of simple modular cable-driven robots without the use of force sensors

Dipl.-Ing. Jonas Bieber, Dipl.-Ing. David Bernstein, Dipl.-Ing. Micha Schuster, Prof. Dr.-Ing. Michael Beitelschmidt, Technische Universität Dresden, Professur für Dynamik und Mechanismentechnik, 01307 Dresden, Deutschland, jonas.bieber@tu-dresden.de

Kurzfassung

Die meisten im Einsatz befindlichen Seilroboter sind auf die Erfüllung hoher Anforderungen bezüglich Last, Genauigkeit und Zuverlässigkeit ausgelegt, woraus eine hohe Systemkomplexität resultiert. Um die daraus folgenden Einstiegshürden in die Seilrobotertechnologie für Anwendungen mit verringerten Anforderungen zu reduzieren, wird ein Ansatz zur Vereinfachung des Seilroboteraufbaus vorgestellt. Dabei wird unter anderem auf komplexe Seilführungen, Getriebe und Kraftsensorik verzichtet. Für seilroboterspezifische Komponenten kommen additive Fertigungsverfahren zum Einsatz. Im Rahmen einer zentralen Regelung von Posen werden Seilkräfte über die Motorströme eingestellt. Für eine genaue Vorgabe der Kräfte werden Möglichkeiten zur Reduktion der Auswirkungen der Drehmomentwelligkeit diskutiert. Anhand eines planaren, redundanten Seilroboterprototyps wird das Folgen einer Trajektorie gezeigt.

Abstract

Most cable-driven parallel robots fulfill high requirements regarding load, accuracy and reliability. This results in a high system complexity. To facilitate the entry into cable-driven robotics for applications with reduced requirements, an approach to simplify the robot design is presented. Complex cable guidance, gears and force sensors are ommited. Additive manufacturing is used for cable robot specific components. The cable forces are set via motor currents as part of a centralized closed-loop pose control system. Possibilities for reducing the effects of torque ripple in order to set accurate forces are discussed. By using a planar redundant cable robot prototype, a trajectory is demonstrated.

1 Einleitung

Seilroboter sind parallelkinematische Systeme, bei denen eine Plattform mithilfe von Seilen definiert bewegt, bzw. dieser eine spezifische Last aufgeprägt werden kann. Da Seile im Vergleich zu starren Verbindungselementen leicht und flexibel sind, können mithilfe von Seilrobotern hohe Beschleunigungen und Geschwindigkeiten erreicht werden und große Arbeitsräume abgedeckt werden.

Die wohl prominenteste kommerzielle Anwendung von Seilrobotern besteht aktuell in der Positionierung und Bewegung von Kamerasystemen bei Konzerten oder Sportveranstaltungen. Abseits davon sind Seilroboter bisher eine Nischentechnologie und finden überwiegend prototypische Anwendung u.a. in der Lagerlogistik [1], Bauindustrie [2, 3] und Forschung [4, 5]. Resultierend aus hohen Anforderungen an Last, Genauigkeit und Zuverlässigkeit handelt es sich bei den aktuellen Entwicklungen typischerweise um komplexe mechatronische Systeme. Diese Komplexität führt zu einem hohen Aufwand und beträchtlichen Kosten bei Anschaffung, Implementierung und Wartung. Dies macht den Einsatz von Seilrobotern für viele Anwendungen kaum attraktiv und ist ein Hindernis bei der Erschließung neuer Anwendungsfelder.

Es gibt aber auch Projekte, die zeigen, dass auch Seilroboter mit einem einfachen Hardwareaufbau zielführend sein können. So wird beispielsweise daran gearbeitet, mit Seilrobotern automatisiert Bilder an Hausfassaden zu sprühen [6] oder große Objekte additiv zu fertigen [7, 8]. Darüber hinaus sind auch Anwendungen für einfache Seilroboter im Consumer-Bereich denkbar. So könnten Seilroboter beispielsweise als Kameraträger auch abseits von Großveranstaltungen oder als Eingabegeräte für Computerspiele mit haptischem Feedback genutzt werden.

Das Projekt RopeBot an der Technischen Universität Dresden ist aus einer studentischen Initiative hervorgegangen. Vorläufiges Ziel ist die Erweiterung der Einsatzbreite von Seilrobotern durch Vereinfachung und Modularisierung des Systemaufbaus. Das entwickelte System lässt sich mit geringem Aufwand in bestehende Infrastruktur integrieren. Es wird evaluiert, welche Anforderungen bezüglich Last, Genauigkeit und Zuverlässigkeit mithilfe vergleichsweise günstiger Komponenten und geringer mechanischer Systemkomplexität erfüllt werden können. Die Ansteuerung wird unter Nutzung des aktuellen Stands der Forschung realisiert. Für Anwendungen mit reduzierten Anforderungen resultieren damit aus dem RopeBot-Projekt Empfehlungen für einen hinreichenden Seilroboter-Aufbau.

Dieser Beitrag behandelt grundsätzliche Aspekte vereinfachter Seilroboter und demonstriert die Machbarkeit anhand eines zweidimensionalen, redundanten Prototyps.

2 Systemaufbau

Das zentrale Element eines Seilroboters ist eine in n Freiheitsgraden bewegbare Plattform, an der m Seile befestigt

sind. Jedes Seil wird auf eine angetriebene Seilwinde geführt, mithilfe welcher die Seillänge bzw. Seilkraft vorgeben werden kann (siehe **Abbildung 1**). Die Redundanz r ergibt sich für Seilroboter aus

$$r = m - n. \tag{1}$$

Häufig werden verspannte Seilroboter mit einer Redundanz $r \geq 1$ eingesetzt, um beliebig gerichtete Lasten bei geringer Plattformmasse aufbringen zu können.

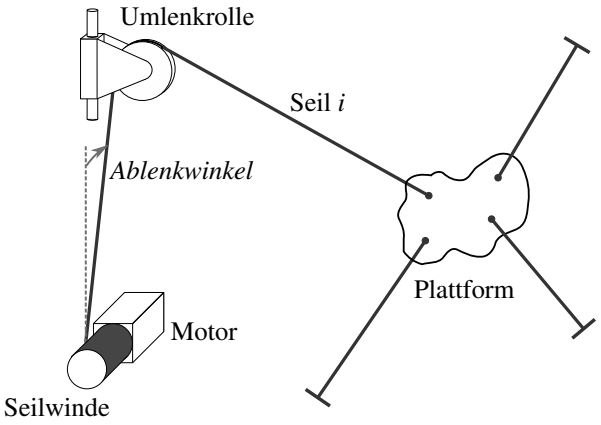


Abbildung 1 Grundlegender Aufbau des Seilroboters

Das vorgestellte System ist grundsätzlich als verspannter und nichtverspannter Seilroboter mit beliebiger Seilanzahl m umsetzbar. Im Folgenden wird der Aufbau mit den umgesetzten Vereinfachungen vorgestellt.

2.1 Komponenten

Seilroboter verwenden üblicherweise Seilwinden, auf denen das Seil einlagig aufgewickelt wird. So besteht durch einen konstanten Windenumfang eine deterministische Beziehung zwischen dem Drehwinkel der Winde und der aufgewickelten Seillänge. Um ein überlappungsfreies Aufwickeln des Seils zu gewährleisten, werden häufig rollenbasierte Seilführungen verwendet. Dabei wird das Seil zwischen der Plattform und der Winde über eine oder mehrere Umlenkrollen geführt. Eine besonders kontrollierte Wicklung kann mit axial zur Winde mitgeführten Umlenkrollen erreicht werden [9, 10]. Auf eine derartige Mechanik wird im Projekt Ropebot verzichtet. Stattdessen wird das Seil lediglich über eine einzelne schwenkbare Rolle zu der Seilwinde umgelenkt. Über den Wickelbereich der Winde verläuft eine schraubenförmige Keilrille, die für eine einlagige Aufwicklung sorgt. Dafür müssen die Seilwinden so zur darauffolgenden Umlenkrolle montiert werden, dass das Seil unter kleinem Ablenkwinkel (siehe Abbildung 1) in die Rille geführt wird. Um einerseits diesen veränderlichen Ablenkwinkel bei den kinematischen Berechnungen vernachlässigen zu können und andererseits ein Herausspringen des Seils aus der Keilrille zu verhindern, ist ein hinreichend großer Abstand zwischen der Seilwinde und der Umlenkkrolle erforderlich.
Als Antrieb der Seilwinden kommen wartungsarme, bürstenlose Elektromotoren zum Einsatz, die sich durch eine hohe Effizienz auszeichnen. Statt vergleichsweise teure Industrieservomotoren mit integriertem Motorcontroller zu verwenden, werden leistungsstarke Motoren aus dem Consumer-Bereich verwendet. Die Ansteuerung der Motoren erfolgt über externe Motorcontroller. Für die erforderliche Winkelmessung werden inkrementelle Encoder eingesetzt.
Bei vielen Seilrobotern treibt der Motor die Seilwinde über ein Getriebe an. So können mit kleineren Motoren hohe Drehmomente auf die Seilwinden realisiert werden. Eine Vereinfachung im Projekt RopeBot besteht darin, dass die Seilwinde direkt mit der Motorwelle verbunden ist und kein Getriebe eingesetzt wird. Das Übersetzungsverhältnis zwischen Motormoment und Seilkraft wird somit durch den Windenradius bestimmt. Aus dem Windenradius und der erforderlichen Gesamtseillänge ergibt sich die notwendige Windenbreite. Um die Windenbreite und somit den maximalen Ablenkwinkel klein zu halten, wird ein großer Windenradius angestrebt. Um das daraus resultierende hohe Drehmoment aufbringen zu können, eignen sich als Motoren besonders langsamlaufende Außenläufermotoren.

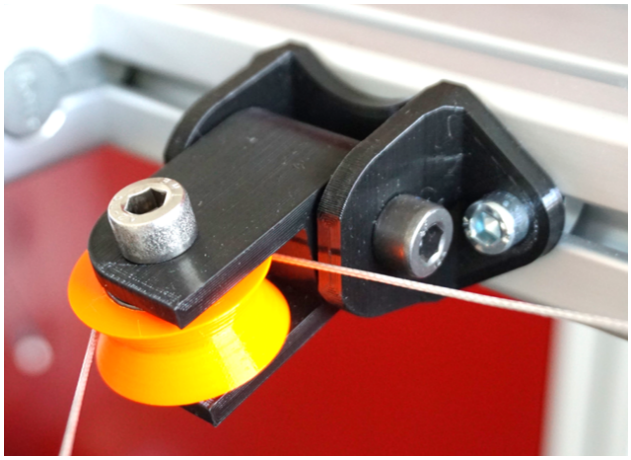


Abbildung 2 Umlenkrolle bestehend aus additiv gefertigten Teilen, Kugellagern und Verbindungsschrauben

Seilroboterspezifische Komponenten wie Seilwinden und Umlenkrollen sind überlicherweise Spezialanfertigungen. Aufgrund hoher Lastanforderungen werden dafür bei vielen Seilrobotern aufwändig herzustellende Bauteile aus Metall verwendet. Eine im Projekt RopeBot umgesetzte Vereinfachung besteht in der additiven Fertigung dieser Bauteile aus Kunststoff (PETG). Darüber hinaus wird bei mechanischen Komponenten auf Normbauteile zurückgegriffen. Bezogen auf die schwenkbaren Umlenkrollen (siehe **Abbildung 2**) bedeutet dies einen Aufbau aus gedruckten Kunststoffteilen, Kugellagern und Befestigungsschrauben. Die gedruckten Winden sind auf drehbar gelagerten Stahlwellen fixiert, die über Kupplungen mit den Motoren verbunden sind. Jeweils zwei Antriebseinheiten sind gemeinsam an einem Motorblockmodul befestigt (siehe **Abbildung 3**). Diese Modularisierung erlaubt den flexiblen Einsatz in vielen unterschiedlichen Konfigurationen.

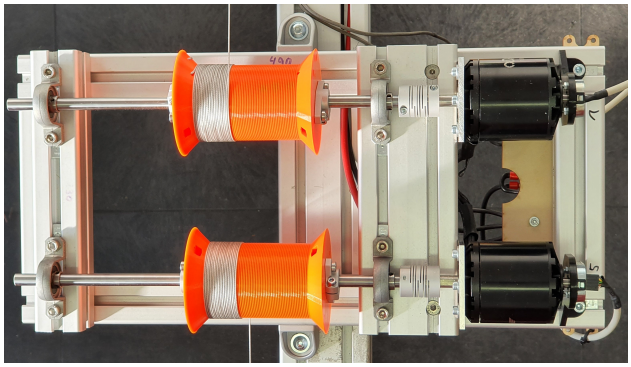


Abbildung 3 Motorblock mit zwei Antriebseinheiten

2.2 Regelung

Wie in der Robotik üblich, wird die Pose $\mathbf{p}$ im kartesischen Arbeitsraum angegeben, während die Ansteuerung der Motoren im Gelenkraum erfolgt. Man unterscheidet zwei grundsätzliche Varianten der Regelung: die dezentrale Regelung der Motordrehwinkel im Gelenkraum und die zentrale Regelung der Pose im Arbeitsraum [11]. Dabei ist zu berücksichtigen, dass bei redundanten Seilrobotern r Zwangsbedingungen zwischen den Seillängen $\mathbf{q}$ bestehen. Diese sorgen schon bei kleinen Abweichungen der Seillängen dafür, dass bei einer einfachen dezentralen Regelung der Motordrehwinkel einzelne Seile schlaff werden oder sich das System unkontrolliert verspannt [12]. Daher wird eine zentrale Regelstruktur gewählt, die auf einer Seilkraftregelung statt einer Drehwinkelregelung der Motoren basiert. Diese erfordert zwar aufwändigere Berechnungen, ist aber robust gegenüber Ungenauigkeiten im kinematischen Modell, da auf Größen im Arbeitsraum geregelt wird, statt die Seillängen einzeln zu regeln.

Das in **Abbildung 4** dargestellte Konzept setzt eine Bestimmung der Plattformpose voraus. Statt dabei wie beispielsweise in [13, 14] ein externes Lokalisierungssystem einzusetzen, wird die Pose mithilfe der Vorwärtskinematik aus den Seillängen berechnet. Dazu werden die Seillängenänderungen aus den Encoderwerten $\mathbf{n}_e$ und die absoluten Seillängen aus einer initlialen Nullstellung bestimmt. Da die Lösung der Vorwärtskinematik analytisch nicht möglich ist, kommt ein iteratives Näherungsverfahren [15] zum Einsatz. Um die Anzahl der benötigten Iterationen zu reduzieren und Konvergenz sicherzustellen, wird dabei die Pose des letzten Regelkreiszyklus als initiale Pose gewählt. Alternativ kommt eine geometrischen Abschätzung der initialen Pose nach [16] zum Einsatz. Aus der Differenz $\mathbf{e}$ zwischen der berechneten Istpose $\mathbf{p}_{ist}$ und der vorgegebenen Sollpose $\mathbf{p}_{soll}$ ergibt sich am Ausgang des Reglers eine Last $\mathbf{w}$ im Arbeitsraum mit n Komponenten. Diese soll durch Superposition der m Seilkräfte $\mathbf{f}$ aufgebracht werden. Im Rahmen der Seilkraftverteilung werden diese durch Lösung des Gleichungssystems

$$\mathbf{A}^T\mathbf{f} = -\mathbf{w} \tag{2}$$

bestimmt. Die Strukturmatrix $\mathbf{A}^T$ entspricht dabei der transponierten Jacobimatrix der inversen Kinematik mit

$$\mathbf{A} = \mathbf{J}_{inv} = \frac{\partial \mathbf{q}}{\partial \mathbf{p}}. \tag{3}$$

Für redundante Seilroboter wird das unterbestimmte Gleichungssystem mithilfe der Lösung in geschlossener Form [17] sowie deren Erweiterung [18] derart gelöst, dass die quadratische Seilkraftabweichung zu einem vorgegebenen Seilkraftniveau minimiert wird.

3 Motorstrombasierte Seilkraftvorgabe

Die in der Regelung berechnete Seilkraft f wird für jedes Seil von den Antrieben erbracht. Die Seilkräfte werden dafür häufig über Kraftsensorik gemessen und über eine eigene Regelschleife eingestellt (vgl. beispielsweise [15, 19, 20]). Die Einbindung von expliziter Kraftsensorik bedeutet allerdings eine erhöhte Komplexität des Gesamtsystems bezüglich Hardware und Software und höhere Kosten. Weiter resultieren aus der Messauswertung Latenzen in der Regelschleife, wie sie in [21] diskutiert werden. Im Sinne eines vereinfachten Seilroboteraufbaus wird im Projekt RopeBot auf explizite Kraftsensorik verzichtet. Da kein Getriebe eingesetzt wird und das Seil nur über eine einzelne Umlenkrolle geführt wird, werden die Reibungsverluste im System vergleichsweise klein gehalten. Dies ermöglicht eine Abschätzung der Seilkraft f aus dem Windenradius r und dem Motormoment τ, welches sich wiederum aus dem Motorstrom i_q ergibt. Dabei wird der linea-

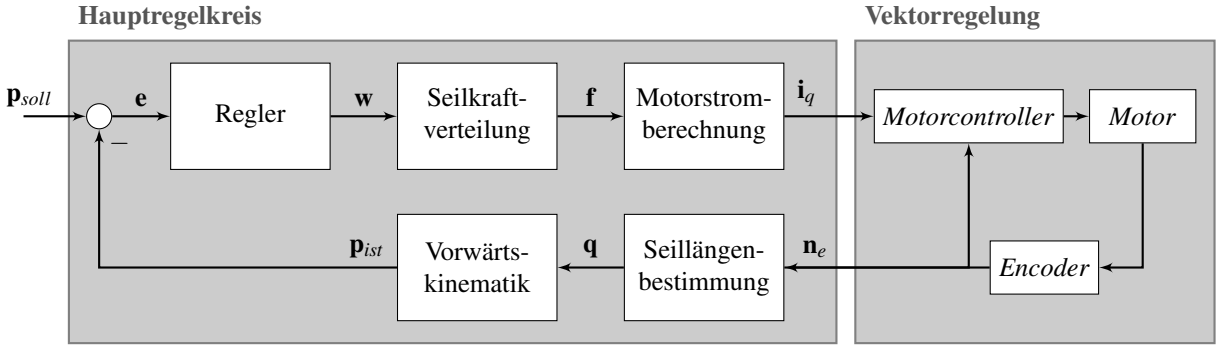


Abbildung 4 Schematische Darstellung des Regelkreises

re Zusammenhang

$$i_q = \frac{\tau}{k_\tau} = \frac{r \cdot f}{k_\tau} \tag{4}$$

mit der Momentenkonstante k_τ des Motors angenommen. Die Ansteuerung der Motoren in den Motorcontrollern basiert auf einer Vektorregelung (Field-Oriented Control). Das dabei in der elektrischen Raumzeigerdarstellung orthogonal zum Rotorfeld ausgerichtete Statorfeld wird dabei vom geregelten Motorstrom i_q bestimmt. In der Realität kommt es allerdings bauartbedingt zu Abweichungen durch die sogenannte Drehmomentwelligkeit. Diese äußert sich bei konstantem Motorstrom i_q in einer definierten Drehmomentabweichung $\Delta\tau(\varphi)$ des resultierenden Motormomentes

$$\tau_{res} = \tau(i_q) + \Delta\tau(\varphi) \tag{5}$$

für jeden Motordrehwinkel φ mit $\tau(i_q)$ aus Gleichung 4. Sie resultiert vor allem aus der magnetischen Wechselwirkung zwischen den Permanentmagneten des Läufers und dem eisenhaltigen Stator.
Mithilfe eines einfachen Messaufbaus lässt sich diese Drehmomentwelligkeit quantifizieren. Dafür wird mit einem Kraftsensor die resultierende Seilkraft f in Abhängigkeit des Motordrehwinkels φ bei konstanter Motorstromvorgabe i_q gemessen. **Abbildung 5** zeigt den gemessenen Kraftverlauf über dem Motordrehwinkel für einen Motor, der im Prototyp des Projektes RopeBot verwendet wird. Es werden Motorströme von $i_q = 2\,\mathrm{A}$ bzw. $i_q = 3\,\mathrm{A}$ vorgegeben. In beiden Messungen treten dabei Seilkraftabweichungen $|\Delta f| \leq 3{,}5\,\mathrm{N}$ mit Standardabweichungen von $\sigma_f \approx 1{,}4\,\mathrm{N}$ auf, die aus den Drehmomentabweichungen $\Delta\tau$ resultieren.

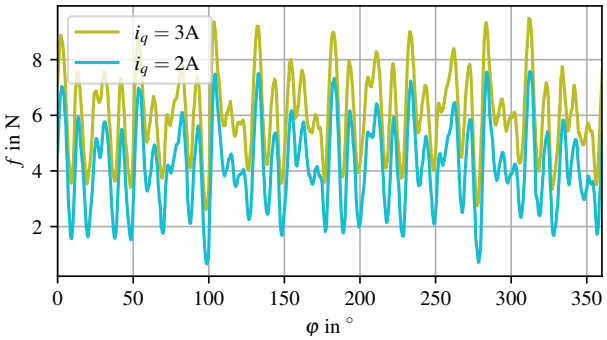


Abbildung 5 Seilkraftverlaufs $f(\varphi)$ über eine Umdrehung bei zwei verschiedenen Motorströmen i_q

Eine Drehmomentwelligkeit mit großen Drehmomentabweichungen $\Delta\tau$ gilt als typisch für kostengünstige bürstenlose Motoren [22]. Im Zusammenspiel aller Seile ergibt sich daraus eine posenabhängige Last, die als Störgröße die Regelung beeinflusst. Je größer die Drehgeschwindigkeit der Motoren, desto geringer ist der Einfluss der Drehmomentwelligkeit, maßgeblich bedingt durch die mechanische Systemträgheit. Eine genaue Positionierung der Plattform bei niedrigen Geschwindigkeiten wird hingegen erschwert.
Die bauartbedingte Drehmomentwelligkeit bewirkt die wesentliche Abweichung vom Motorstrom-Seilkraft-Verhältnis in Gleichung 4 über eine Motorumdrehung. Es gilt dabei $\Delta\tau(\varphi) = \Delta\tau(\varphi + 2\pi n)$ mit $n \in \mathbb{Z}$. Dieser systematische Charakter der Drehmomentwelligkeit ermöglicht eine winkelabhängige Korrektur der Motorströme [22]. Dafür werden die zur Positionshaltung des unbelasteten Motors nötigen Motorströme $\Delta i_q(\varphi)$ über eine Umdrehung gemessenen und gespeichert. Die so ermittelten Werte werden im Motorcontroller vor der Vektorregelung als Offset in Form von

$$i_{q,res} = i_q - \Delta i_q(\varphi) \tag{6}$$

verwendet [23].
Der mit einer entsprechenden Kompensation aufgezeichnete Seilkraftverlauf, ausschnittsweise dargestellt in **Abbildung 6**, zeigt eine deutliche Reduktion der Seilkraftabweichung auf $|\Delta f| \leq 1{,}5\,\mathrm{N}$ und $\sigma_f \approx 0{,}4\,\mathrm{N}$.

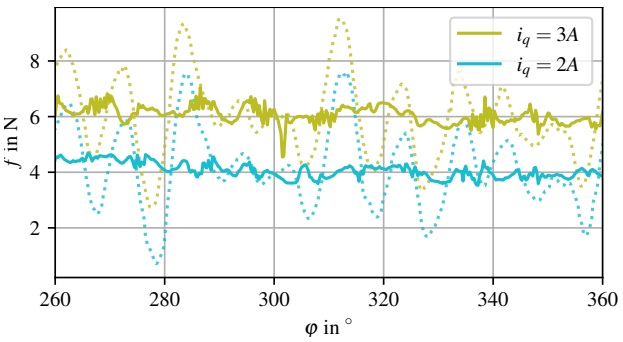


Abbildung 6 Ausschnitt des Seilkraftverlaufs $f(\varphi)$ ohne Kompensation (gepunktet) und mit Kompensation der Drehmomentwelligkeit bei zwei verschiedenen Motorströmen i_q

Eine weitere Strategie, um den Anteil der Drehmomentwelligkeit am resultierenden Motordrehmoment zu reduzieren, besteht darin, die Motoren in möglichst hohen Momentbereichen einzusetzen. Da die Drehmomentwelligkeit $\Delta\tau$ lediglich vom Motorwinkel φ und nicht vom Motorstrom i_q abhängt, sinkt der Anteil entsprechend mit zunehmendem Motormoment. Die Drehmomentwelligkeit ist daher vor allem bei geringen Motormomenten dominant. Als konstruktive Maßnahme führt ein größerer Windenradius bei gleichbleibendem Seilkraftniveau zu höheren Drehmomenten und somit reduziertem Anteil der Drehmomentwelligkeit am Gesamtmoment. Bei der Auslegung der Winden ist es also zweckmäßig, das Maximaldrehmoment der Motoren auszunutzen, indem entsprechend große Winden verwendet werden.

4 Prototyp

Das vorgestellte Konzept wird anhand eines planaren Prototyps mit $m = 4$ Seilen und den $n = 3$ Freiheitsgraden x, y und α demonstriert. Wie in **Abbildung 7** dargestellt, sind die Motorblöcke und Umlenkrollen dieses redundanten Seilroboters an einem Rahmen befestigt, der auf einem Aluminium-Systembaukasten basiert.
Wesentliche Parameter der verwendeten Komponenten sind in **Tabelle 1** angegeben. Neben Motoren, Winden und

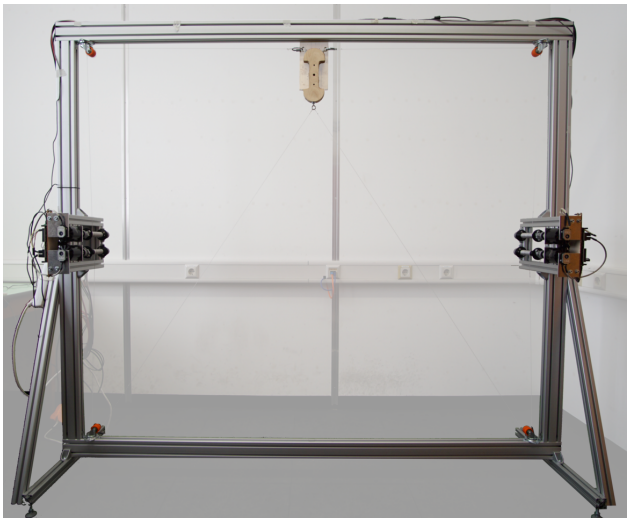


Abbildung 7 Planarer Prototyp mit $n = 3$ Freiheitsgraden und $m = 4$ Seilen. Die Plattform ist zur Nulllagenbestimmung fixiert.

Umlenkrollen kommen unter anderem Kunstfaserseile zum Einsatz, die bei kleinem Seildurchmesser eine hohe Längssteifigkeit aufweisen.
Die in C++ implementierte Regelung ist auf einem Raspberry Pi umgesetzt. Die Kommunikation mit einem Rechner zur Nutzerinteraktion ist per Modbus-Protokoll realisiert. Als Motorcontroller kommen ODrive-Boards [24] zum Einsatz, die über CAN-Bus mit dem Raspberry Pi kommunizieren. Die Kompensation der Drehmomentwelligkeit ist bereits an Einzelmotoren getestet, aber noch nicht für die vier Motoren des Prototyps umgesetzt.

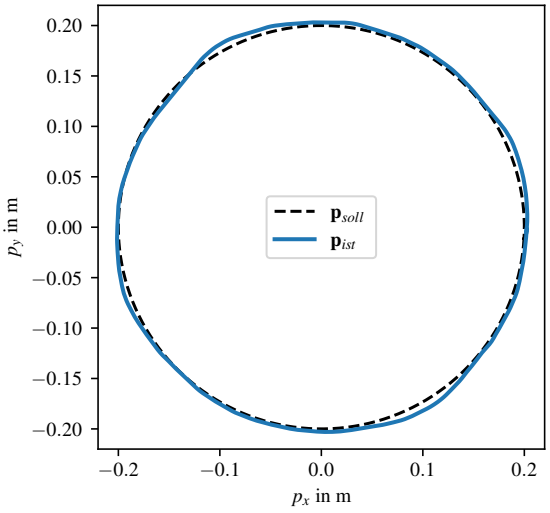


Abbildung 8 Kreisförmige Beispieltrajektorie

Abbildung 8 zeigt das Folgen einer kreisförmigen Trajektorie mit einer Bahngeschwindigkeit von 0,38 m/s während die Verdrehung der Plattform konstant gehalten wird. Wie in **Abbildung 9** gezeigt, treten dabei translatorische Regelabweichungen von $e_p \leq 10\,\text{mm}$ bei einer Abweichung der Orientierung von $e_\alpha \leq 1\,^\circ$ auf.

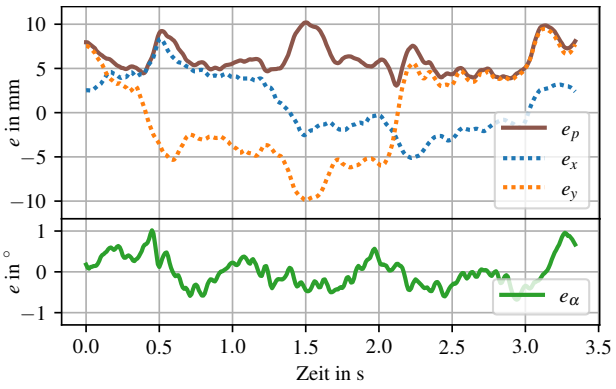


Abbildung 9 Regelabweichungen der Position $e_p = \sqrt{e_x^2 + e_y^2}$ und Rotation e_α während der kreisförmigen Beispieltrajektorie

Die Trajektorie ist umgesetzt als eine Abfolge von Einzelposen. Der verwendete Regler ist ein PID-Regler, der die drei Freiheitsgrade unabhängig voneinander mithilfe der Regelabweichung der Pose einstellt. Eine modellbasierte Vorsteuerung auf Basis der bekannten Beschleunigungen kommt nicht zum Einsatz. So resultiert die dargestellte Positionsabweichung e_p zum Teil aus dem Nachlaufen der Freiheitsgrade x und y. Weitere Einflussfaktoren sind geometrische Abweichungen und die nicht kompensierte Drehmomentwelligkeit.
Die Auswertung der Trajektorie demonstriert, dass das Einstellen von Posen bis auf wenige Zentimeter bzw. Grad genau möglich ist. Der Prototyp dient als Ausgangspunkt für eine Erweiterung des Konzepts auf mehr Freiheitsgrade und Einsatz in konkreten Anwendungen.

Tabelle 1 Parameter der verwendeten Komponenten des Seilroboter-Prototyps

Komponente	*Parameter*
Winde	Durchmesser: 22 mm
Rahmen	Breite: 2 m; Höhe: 1,8 m
Plattform	Breite: 10 cm; Höhe: 24 cm
Seil	*Dyneema* ® 1 mm
Motor	*ODrive D6374 150KV*
Encoder	*CUI Devices AMT-102*
Motorcontroller	*ODrive V3.6*
Steuerrechner	*Raspberry Pi 4 Model B*

5 Zusammenfassung

In diesem Beitrag wird ein Konzept zur Vereinfachung und Modularisierung des Aufbaus von Seilrobotern vorgestellt. Es wird dabei auf komplexe Seilführungen, Getriebe und Kraftsensorik verzichtet. Seilroboterspezifische Bauteile werden additiv aus Kunststoff gefertigt. Dabei sind Seilroboter mit beliebiger Seilanzahl umsetzbar. Die Regelung von Posen und Trajektorien basiert auf einer zentralen Regelung im Arbeitsraum. Um die Seilkräfte möglichst genau über die Motorströme vorgeben zu können, werden Möglichkeiten zur Kompensation der Drehmomentwellig-

keit vorgestellt. Am Beispiel eines planaren, redundanten Seilroboters wird gezeigt, dass mit dem vereinfachten Aufbau einer Trajektorie mit guter Genauigkeit gefolgt werden kann. Eine Verbesserung des Regelungskonzepts lässt eine Reduktion der verbleibenden Abweichungen erwarten. Der vorgestellte Ansatz ermöglicht eine Öffnung der Seilrobotertechnologie für einen breiten Anwendungsbereich bis in die privaten Nutzung hinein.

6 Literatur

[1] Bruckmann, T.; Lalo, W.; Schramm, D.; Hiller, M.: *Design and realization of a highrack storage and retrieval machine based on wire robot technology* Proceedings of the XV International Symposium on Dynamic Problems of Mechanics (2013), S. 771–780.

[2] Pott, A.; Meyer, C.; Verl, A.: *Large-scale assembly of solar power plants with parallelcable robots* 41st International Symposium on Robotics (ISR) and 6th German Conference on Robotics (2010), ISBN 978-3-8007-3273-9, S. 999-1004.

[3] Izard, J.; Gouttefarde, M.; Baradat, C.; Culla, D.; Sallé, D.: *Integration of a parallel cable-driven robot on an existing building facade* Cable-Driven Parallel Robots. Mechanisms and Machine Science vol 12 (2012), ISBN 978-3-642-31987-7, S. 149-164.

[4] Bruckmann, T.; Mikelsons, L.; Pott, A.; Abdel-Maksoud, M.; Brandt, T.; Schramm, D.: *A novel tensed mechanism for simulation of maneuvers in wind tunnels* 33rd ASME Mechanics and Robotics Conference (2009), ISBN 978-0-7918-4904-0, S. 17-24.

[5] Miermeister, P.; Lächele, M.; Boss, R.; Masone, C.; Schenk, C.; Tesch, J.; Kerger, M.; Teufel,H.; Pott, A.; Bulthoff, H.: *The CableRobot simulator: large scale motion platformbased on cable robot technology* IEEE/RSJ International Conference on Intelligent Robotsand Systems (2016), ISBN 978-1-5090-3763-6, S. 3024-3029.

[6] Robot Muralist: `https://www.robotmuralist.com`

[7] Hangprinter: `https://hangprinter.org`

[8] Bloft Design Lab: `https://en.bloft.fi`

[9] Trautwein, F.; Reichenbach, T.; Tempel, P.; Pott, A.; Verl, A.: *COPacabana - A Modular Cable-Driven Parallel Robot* Sechste IFToMM D-A-CH Konferenz (2020), DOI 10.17185/duepublico/71189.

[10] Pott, A; Mütherich, H; Kraus, W; Schmidt, V; Miermeister, P; Verl, A: *IPAnema – A family of cable-driven parallelrobots for industrial applications* International Conference on Cable-Driven Parallel Robots (2013), ISBN 978-3-642-31987-7, S. 119-134.

[11] Paccot, F.; Andreff, N.; Martinet, P.: *A Review on the Dynamic Control of Parallel Kinematic Machines: Theory and Experiments* Interational Journal of Robotics Research 28 (2009), DOI 10.1177/0278364908096236, S. 395-416.

[12] Valasek, M; Belda, K; Florian, M: *Control and Calibration of Redundantly Actuated Parallel Robots*

[13] Chellal, R.; Cuvillon, L.; Laroche, E.: *A kinematic vision-based position controlof a 6-DoF cable-driven parallel robot* Cable-Driven Parallel Robots. Mechanisms and Machine Science vol 32 (2014), ISBN 978-3-319-09489-2, S. 213-225.

[14] Dallej, T.; Gouttefarde, M.; Andreff, N.; Dahmouche, R.; Martinet, P.: *Vision-based modeling and control of large-dimension cable-driven parallel robots* IEEE/RSJInternational Conference on Intelligent Robots and Systems (2012), ISBN 978-1-4673-1737-5, S.1581-1586.

[15] Fang, S: *Design, Modeling and Motion Control of Tendon-Based Parallel Manipulators* Fortschritt-Berichte / VDI : Reihe 8, Meß-, Steuerungs- und Regelungstechnik (2005), ISBN 3-18-507608-7.

[16] Pott, A: *Cable-Driven Parallel Robots: Theory and Application*. Heidelberg: Springer, 2018.

[17] Pott, A; Bruckmann, T; Mikelsons, L: *Closed-Form Force Distribution for Parallel Wire Robots* Computational Kinematics (2009), ISBN 978-3-642-01946-3, S. 25-34.

[18] Pott, A: *An Improved Force Distribution Algorithm for Over-Constrained Cable-Driven Parallel Robots* Computational Kinematics (2014), ISSN 2211-0984, S. 139–146.

[19] Mohammad A. Khosravi; Hamid D. Taghirad: *Robust PID Control of Fully-Constrained Cable Driven Parallel Robots*. Mechatronics 24 (2014), ISSN 0957-4158, S.87-97

[20] Kraus, W: *Force control of cable-driven parallel robots* Stuttgarter Beiträge zur Produktionsforschung 49 (2016), ISBN 978-3-8396-0979-8

[21] Kraus, W; Schmidt, V; Rajendra, P; Pott, A: *System Identification and Cable Force Control for a Cable-Driven Parallel Robot with Industrial Servo Drives* International Conference on Robotics and Automation (2014), ISBN 978-1-4799-3684-7, S. 5921-5926.

[22] Piccoli, M.; Yim, M.: *Cogging Torque Ripple Minimization via Position Based Characterization* Proceedings of Robotics: Science and Systems (2014), DOI 10.15607/RSS.2014.X.042.

[23] GitHub: *Anti-cogging documentation* `https://github.com/odriverobotics/ODrive/blob/devel/docs/anticogging.md` (zuletzt abgerufen am 26.01.2021)

[24] ODrive: `https://odriverobotics.com`