Hochschule Reutlingen
Reutlingen University

Uwe Kloos, Natividad Martínez, Gabriela Tullius (Hrsg.)

Wissenschaftliche Vertiefungskonferenz
HUC - That's IT

Tagungsband der Informatik-Konferenz

an der Hochschule Reutlingen

Vom 21. November 2018

Impressum

Anschrift:

Hochschule Reutlingen
Reutlingen University
Fakultät Informatik
Human-Centered Computing
Alteburgstraße 150
D-72762 Reutlingen

Telefon: +49 7121 / 271-4002
Telefax: +49 7121 / 271-4042

E-Mail: wvk@reutlingen-university.de
Webseite: wvk.reutlingen-university.de

Organisationskomitee:

Prof. Dr. rer. nat. Uwe Kloos, Hochschule Reutlingen
Prof. Dr.-Ing habil. Natividad Martínez, Hochschule Reutlingen
Prof. Dr. rer. nat. Gabriela Tullius, Hochschule Reutlingen

Steven Cybinsky
Annika Dausend
Michael Essich
Sina Frommer
Elif Hizli
Gamze Karatas
Tim Laine
Nils Lüdeke
Dominik Müller
Christian Pitters
Tobias Schultze

Hochschule Reutlingen
Reutlingen University

Copyright: Hochschule Reutlingen, Reutlingen 2018
Herstellung und Verlag: Hochschule Reutlingen
ISBN 978-3-00-061377-7

Vorwort

Zur diesjährigen Wissenschaftlichen-Vertiefungs-Konferenz (WVK) darf ich Sie herzlich begrüßen! Jedes Jahr organisieren Studierende der Reutlinger Informatik die Konferenz mit äußerst hohem Engagement, um Ihnen eine Auswahl origineller wissenschaftlichen Arbeiten vorzustellen. Dabei stellen sich forschende Studierende im Rahmen ihres Studiums der Herausforderung, ihre Ergebnisse, die sie beispielsweise durch eine gewählte wissenschaftliche Vertiefung im Rahmen ihres Bachelor- oder Masterstudiums erhielten, darzustellen. Die Teilnehmer erfüllen nicht nur den Anspruch, die Ergebnisse ihrer Arbeit in schriftlicher Form anschaulich auszuarbeiten, sondern auch interaktiv vor interessierten Besuchern zu verteidigen und zu diskutieren. Dass dies eine wertvolle Erfahrung für alle Beteiligten ist, steht außer Frage. Es ist zu hoffen, dass über die bevorstehende Abschlussarbeit hinaus die Teilnehmer verstärkt Freude am wissenschaftlichen Arbeiten finden. Vermehrt schlagen Absolventen sogar den Weg einer Promotion im Anschluss Ihres Studiums ein.

Die WVK bietet ein Forum für Studierende, während des Studiums sowohl Ergebnisse ihrer Arbeit professionell einem interessierten Publikum zugänglich zu machen, als auch die Arbeiten anderer kritisch zu beurteilen und im Rahmen einer Konferenzorganisation zusammenzustellen. Dadurch erwerben die Studierenden weitere wichtige Qualifikationen, die über den Erwerb fachlicher Kenntnisse hinausgehen.

Das Motto der diesjährigen WVK heißt „That's IT“. Zur Zeit prägt ein Welle der Digitalisierung die verschiedensten IT-Branchen. Fast jeder IT-Innovation liegt der Einsatz künstlicher Intelligenz zu Grunde. Zudem steht mehr als zuvor Interdisziplinarität für erfolgreiches Arbeiten an den Schnittstellen von etablierten Bereichen. Als Besucher der Konferenz erwarten Sie wieder Beiträge aus den unterschiedlichsten Bereichen, wie z.B. dem Maschinellen Lernen zur visuellen Zustandserkennung des Menschen, neuen Ansätzen für Assistenzsysteme zur Unterstützung von verlorengegangenen Sinneswahrnehmungen, aber auch neuen Ansätzen zur Erzeugung und Nutzung von Virtuellen Realitäten und Nutzerstudien im Zusammenhang mit modernen IT-Systemen.

Den Veranstaltern und Beitragenden der Konferenz gilt größte Anerkennung. Ich wünsche allen Beteiligten und Besuchern eine interessante Konferenz mit angeregtem Austausch.

Prof. Dr. Cristóbal Curio Lehr- und Forschungsbereich Kognitive Systeme,
Vizedekan für Forschung an der Fakultät Informatik

Inhaltsverzeichnis

Tim Laine
Konzeption einer Fallstudie zur Evaluierung von SPARK 5

Sina Frommer
Optimierung des Informationstransfers bei der sensorischen Substitution - Eine theoretische Betrachtung auf Basis der Informationstheorie 15

Vanessa Willenbrock
Mouse-Tracking als Indikator für den Lernerfolg 25

Katharina Pavić
Modellierung der agilen Softwareprozesse in einem Großprojekt mit Dienstleistern und Analyse der Herausforderungen der agilen Transformation 35

Annika Dausend
Aktuelle Methoden zur Segmentierung von Laparoskopieinstrumenten 45

Michael Essich
Entwicklung eines Simulationsmodells eines Time-of-Flight Sensors 55

Maxim Stoljar
Evaluierung von Einsatzszenarien für Augmented Reality im Rettungsdienst 65

Christian Pitters
Vergleich von Plattformen zur Umsetzung einer V2C Kommunikation auf Basis von Automotive Grade Linux 75

Nils Lüdeke
Verfahrensüberblick zur Object Pose Detection im 3D Raum 85

Pia Laubacher
Untersuchung klinisch validierter Assessments zur Abstraktion und Ableitung eines logischen Modells 95

Emanuel Geiger
Entwicklung eines Low Fidelity Prototyps für die Erforschung von Akzeptanzkriterien für Sprachassistenzsysteme 105

Claudia Ryniak
Klassifikation und Evaluierung von Signalen zur Erkennung chirurgischer Aktivitäten im OP 115

Dina Kurbanismailova
Usability Methoden am Beispiel der Fallstudie eines B2B Unternehmens 125

Dominik Müller
Komponentenbasierte Analyse und Animation Retargeting von Gesichtsdynamiken 135

Gamze Karatas
Studie zum Vergleich der Sprachassistenzsysteme Siri und Google Assistant bezüglich des Einsatzes als Fahrassistenzsystem 145

René Blänsdorf
Methoden der Anomaliedetektion mit Fokus auf Logfileanalyse 155

Tobias Schultze
Kollaborative VR-Plattformen für Virtual-Learning-Environments 165

Lucas Hermann
Linked knowledge – Managing information like the human brain 175

Frauke Hanna Grieble
Möglichkeiten zur Repräsentation von Emotionen bei der Kommunikation 185

Konzeption einer Fallstudie zur Evaluierung von SPARK

Tim Laine
Hochschule Reutlingen
Tim.Laine@Student.
Reutlingen-University.DE

Abstract
Diese Arbeit beschreibt die Konzeption einer Fallstudie (Erhebung), mit welcher das reservierungsbasierte Parkleitsystem SPARK validiert werden kann. Für eine Validierung von SPARK ist eine Verkehrserhebung notwendig, welche anhand des Vorgehensmodells der Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) durchgeführt wird. Dabei werden eine Grundlagenkonzipierung und ein Vergleich technischer Geräte für den Aufbau der nötigen Messanlage durchgeführt. Die Ergebnisse und Planung dieser Arbeit legen den Grundstein für eine weitere Durchführung der Evaluierung.

Schlüsselwörter
SPARK, smart Parken, Verkehrsleitsysteme, Grundlagenkonzept

CCS-Concepts
- General and reference → Empirical studies
- General and reference → Measurement
- Hardware → Sensors and actuators

Betreuer Hochschule: Prof. Dr. Marcus Schöller
Hochschule Reutlingen
Marcus.Schöller@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Tim Laine

1 Einleitung
Derzeit werden deutschlandweit pro Parkvorgang eines PKWs etwa 6 Minuten für die Parkplatzsuche aufgewendet [1]. Dabei wird bereits heute versucht, den Verkehr schneller zu den freien Stellplätzen zu leiten. Dazu werden Verkehrsleitsysteme eingesetzt, welche allerdings häufig nicht in der Lage sind, den Status eines einzelnen Stellplatzes zu erkennen sowie die individuelle Routenführung eines Fahrzeuges zu diesem freien Stellplatz durchzuführen. Die lange Suchdauer für einen Stellplatz findet sich auch auf dem Parkplatz der Hochschule Reutlingen wieder. Dieser verfügt dabei über kein Verkehrsleitsystem, wie es bereits in vielen Innenstädten zu finden ist. Mittels SPARK soll sich die Suche nach einem freien Stellplatz auf dem Hochschulparkplatz vereinfachen. So kann ein Fahrer bereits vor Fahrtantritt die Verfügbarkeit eines Stellplatzes prüfen. Sollte der Stellplatz zwischenzeitlich belegt sein, kann dies auch während der Fahrt geprüft werden. So kann frühzeitig eine alternative Parkmöglichkeit angefahren werden. Weiterhin kann auf direktem Weg zu einem freien Stellplatz gefahren werden und die langwierige Suche nach einem freien Stellplatz entfällt.

Zukünftig soll auch die Reservierung von Stellplätzen vor Fahrtantritt in SPARK integriert werden. Dies ist vor allem im Zusammenhang mit den Ladestationen für Elektrofahrzeuge interessant.

Daher ist SPARK ein reservierungsbasiertes Verkehrsleitsystem, welches zum Ziel hat Fahrzeuge auf direktem Weg zu einem geeigneten freien Stellplatz zu leiten.

1.1 Projektumfeld

Im Masterstudiengang Human Centered Computing wurde ein Prototyp zur Überwachung des Hochschulparkplatzes entwickelt. Aus „Smart PARKen“ wurde das Akronym SPARK abgeleitet, welches für den Projekt- und Systemnamen verwendet wurde. Hauptziel des Projektes ist es, das System SPARK zu entwickeln und zu erproben. SPARK soll es ermöglichen, in Echtzeit die Auslastung des Hochschulparkplatzes zu erfassen und darzustellen. Dabei soll es möglich sein, jeden Stellplatz eindeutig zu identifizieren und dessen Belegungsstatus zu erkennen.

Im ersten Proof Of Concept (POC) soll eine Reihe an Parkplätzen mit Sensoren ausgestattet werden. Diese vom Projektteam entwickelten Sensoren erkennen jeweils den Status eines Stellplatzes. Langfristig sollen weitere Sensoren von Drittherstellern unterstützt werden.

Im Rahmen des Projektes wurde die gesamte Plattform von den Sensoren über ein cloudbasiertes Backend bis hin zu einer App für den Endanwender entwickelt. Der Endanwender ist in diesem Fall der Fahrer, welcher einen Stellplatz sucht. Zusätzlich benötigen Administratoren und Verwaltung einen Zugriff auf die Daten, um die Parkplatzsituation zu überwachen und zu verbessern. Das System SPARK soll sich zu einer offenen Plattform entwickeln, welche es Drittanbietern ermöglicht eigene Sensoren und Anwendungen zu integrieren.

Die Sensoren erfassen im ersten Schritt keine personenbezogenen Daten. Das heißt es wird nur der Stellplatz (frei oder belegt) erfasst. Später sollen weitere Daten erhoben werden um mittels künstlicher Intelligenz die Parkplatzsituation zu verbessern.

So können beispielsweise Fahrgemeinschaften vorgeschlagen werden oder die Belegung des Parkplatzes anhand des tagesaktuellen Stundenplans vorausgesagt werden.

Ziel von SPARK ist es, ein reservierungsbasiertes und intelligentes Verkehrsleitsystem bereitzustellen, welches den durch die Stellplatzsuche verursachten Verkehr reduziert. Dies wird ermöglicht, indem der Verkehr gezielt auf freie Stellplätze geleitet wird. Dabei findet die Überwachung auf Stellplatzebene statt, so dass eine Verkehrsleitung auf Mikroebene erfolgen kann.

1.2 Ziel dieser Arbeit

Um die Verbesserung der Parkplatzsituation, welche durch SPARK erzielt wird, eindeutig zu bestimmen, bedarf es einer geeigneten Messmethode. Mit dieser Messmethode müssen vor und nach der Einführung von SPARK Messungen durchgeführt werden. Anschließend kann daraus die geplante Verbesserung empirisch belegt werden.

In dieser Arbeit wird dazu eine Fallstudie konzipiert, welche die derzeitige Parkplatzsituation erfasst. Dazu wird von der Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) ein Vorgehensmodell (Abbildung 1) empfohlen, welches die generellen Arbeitsschritte bei einer Verkehrserhebung beschreibt. Das Vorgehensmodell deckt dabei den gesamten Prozess ab und reicht von der Konzeptionierung über die Durchführung bis hin zur Dokumentation.

Diese Arbeit befasst sich mit den ersten beiden Schritten aus dem Vorgehensmodell. Diese sind die Grundlagenkonzipierung sowie die Erhebungsvorbereitung. Dabei wird ein Fokus auf die Evaluation der technischen Geräte gelegt. Ziel ist es, verschiedene Technologien zu vergleichen und die passende Technologie auszuwählen, welche für die spätere Erhebung verwendet wird.

2 Stand der Technik

In diesem Kapitel wird zu Beginn der Stand der Technik der Technologien zur Verkehrszählung erläutert. Anschließend wird auf den Aufbau der Messanlagen eingegangen.

Aktuelle Technologien zur Verkehrszählung zielen darauf ab, den Verkehr möglichst zeitnah und automatisiert zu erfassen [2]. Dabei wird zwischen verschiedenen Technologien unterschieden. „Die am häufigsten eingesetzten Geräte bei Zählungen im Kfz-Verkehr sind Induktionsschleifen, Radar- und Laserdetektoren […]“ [3]. Zusätzlich gibt es Kombinationsdetektoren, welche verschiedene Technologien wie zum Beispiel Doppler-Radar, Ultraschall oder Passiv-Infrarot miteinander kombinieren [3]. Diese automatisierten Messanlagen bringen jedoch alle ihre eigenen Nachteile mit, so dass zur Verkehrszählung häufig Hilfskräfte zur manuellen Verkehrszählung eingesetzt werden [2]. Zusätzlich haben sich in den letzten Jahren auch verschiedene Kameratechnologien etabliert [2].

All diese Messanlagen setzen im gewissen Umfang bauliche Maßnahmen voraus. Dazu kommen Umwelteinflüsse wie Salz, Feuchtigkeit und Sonneneinstrahlung. Diese sorgen dafür, dass die Messanlagen wartungsintensiv sind und Ungenauigkeiten in der Messung aufweisen können. [2]

Induktionsschleifen arbeiten mit einem Magnetfeld, welches sich verändert sobald sich ein Fahrzeug über der in den Boden eingelassenen Spule befindet. Dadurch ist die Induktionsschleife sehr genau. Für den besten Halt werden Induktionsschleifen im Boden eingelassen. Temporäre Induktionsschleifen, welche auf die Oberfläche geklebt werden, sind durch die ständigen Reifenbewegung nicht für den Dauereinsatz geeignet [4].

Radaranlagen versuchen Fahrzeuge aufgrund von Mustern im Radarbild zu erkennen. Dazu muss die Größe der zu erkennenden Objekte im Voraus bestimmt werden. [4]

Laserdetektoren basieren auf dem Prinzip einer Lichtschranke. Wird diese unterbrochen, so wurde ein Objekt erkannt. Jedoch kann die Art des Objektes nicht bestimmt werden.

Ultraschalldetektoren arbeiten nach dem Prinzip der Ausbreitung von Schall. So werden „Ultraschall-Schallwellen“ ausgestrahlt, welche von Objekten reflektiert werden. Die Reflektion wird durch den Detektor empfangen. [5] So kann das Vorhandensein eines Objektes, sowie ein Längenmaß (wie Höhe oder Breite) ermittelt werden.

Kameratechnologien versuchen auf Bildmaterial Fahrzeuge zu klassifizieren. Dabei treten aber vor allem Probleme beim Datenschutz und der Erkennung bei schlechtem Wetter auf. [2]

Neben den Messanlagen an Straßen können auch sogenannte On Board Units (OBU) in Fahrzeugen zur Verkehrszählung genutzt werden. Diese OBUs können fahrzeuggebunden oder mobil sein. Fahrzeuggebundene OBUs sind mit dem Fahrzeugbussystem verbunden und können auf die Fahrzeugdaten zugreifen. Mobile OBUs sind beispielsweise Smartphones, welche von den Fahrern mitgeführt werden. [2] Der Nachteil dieser OBUs ist, dass nicht jedes Fahrzeug damit ausgestattet ist.

Alle automatisierten Messanlagen unterliegen drei Fehlerquellen: fehlerhafte Bedienung, fehlerhaftes Gerät oder fehlerhafte Datenaufbereitung. Diese Fehler sind zum Teil auch der Heterogenität der Fahrzeuge geschuldet. [3] Zusätzlich kann auch die Witterung einen negativen Einfluss auf die Erkennung haben. Um eine qualitativ hochwertige Erkennung zu erreichen, müssen diese Fehler auf ein Minimum reduziert werden.

Neben der Technik zur Verkehrserhebung gibt es auch verschiedene Methodiken, welche im Folgenden kurz erläutert werden:

„Verkehrserhebungen dienen der systematischen Beschaffung von Informationen.“ [6] Um diese durchzuführen unterschiedet man nach [6] zwischen zwei Erhebungsmethoden:

Verkehrstechnische Erhebungen: Bei dieser Erhebung wird die Ortsveränderung beobachtet. Diese ist vor allem in der Verkehrsplanung relevant.

Verhaltensbezogene Erhebungen: Es wird die Art und das Maß der Teilnahme am Verkehr erfasst. Dabei wird der Gesamtverkehr betrachtet und Wechselbeziehungen dargestellt.

Zusätzlich wird an dieser Stelle das Vorgehensmodell für Verkehrserhebungen der FGSV als anzuwendenden Stand der Technik gesehen. Auch wenn in dieser Arbeit nur die ersten beiden Arbeitsschritte durchlaufen werden, soll die folgende Abbildung einen Überblick über das gesamte Vorgehensmodell bieten:

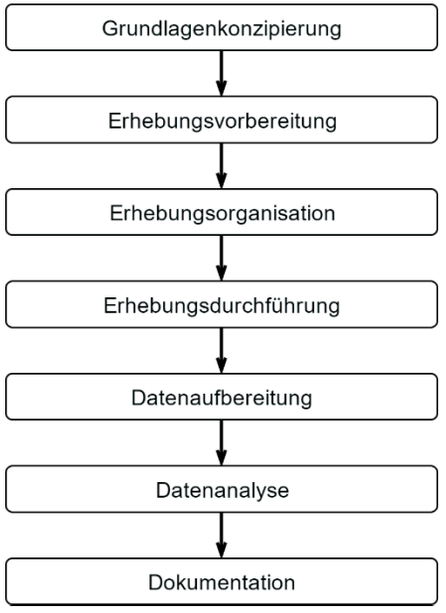


Abbildung 1: FGSV Vorgehensmodell [7]

3 Grundlagenkonzipierung

Der erste Arbeitsschritt aus dem Vorgehensmodell der FGSV ist die Grundlagenkonzipierung. Dabei werden der Erhebungsanlass und das Untersuchungsziel definiert. Zusätzlich werden die Aussagegüte, Randbedingungen sowie zu erfassende Sekundärstatistiken bestimmt. Des Weiteren werden Anforderungen an die Erfassungseinheit (-en), weitere Merkmale der Erfassung sowie den Datenschutz gestellt. Aus dem Bereich der Statistik sollen auch Hypothesen an die Erhebung formuliert werden. [7]

3.1 Erhebungsanlass

Der Anlass der Verkehrserhebung dient der Evaluierung von SPARK. Es soll demnach gezeigt werden, welche Auswirkungen SPARK auf die Parkplatzsituation an der Hochschule Reutlingen hat. Dazu müssen verschiedene Erhebungen durchgeführt werden. Zum einen muss die Situation vor und nach der Integration von SPARK erfasst werden, um einen Unterschied feststellen zu können. Zusätzlich kann eine Erhebung zur Evaluierung von Subsystemen durchgeführt werden, um fortlaufende Evaluierungsergebnisse zur Verfügung zu haben.

3.2 Untersuchungsziel

Das erste Ziel der Untersuchung/Erhebung ist es festzustellen, ob durch die Einführung von SPARK der Verkehr, welcher einen freien Stellplatz sucht, reduziert wird. Daher muss der Verkehr an den Einfahrten zu den Parkplätzen gezählt werden, um festzustellen wie viele Fahrzeuge auf einen Parkplatz fahren. Durch die Menge der Fahrzeuge und die zur Verfügung stehenden Stellplätze kann berechnet werden, wie viele Fahrzeuge einen Parkplatz befahren, obwohl die Kapazitätsgrenze erreicht ist.

Als zweites Ziel soll festgestellt werden, ob das gesamte Verkehrsaufkommen auf dem Parkplatz reduziert wird. Im Gegensatz zum parkplatzsuchenden Verkehr, soll dabei eine Reduzierung der gesamten Fahrzeuge stattfinden.

3.3 Aussagegüte

Die Aussagegüte bestimmt den Grad der Beschaffenheit des Ergebnisses. Da eine Evaluierung von SPARK nur mit korrekten Daten stattfinden kann, sind die Daten in sehr hoher Güte zu erstellen. Konkret bedeutet dies, dass Fehlerkennungen auf ein Minimum reduziert werden müssen. Als Beispiel seien hier Induktionsschleifen genannt, welche aufgrund von Witterung wie Eis Ausfallquoten von bis zu 2/3 aufweisen [2]. Daher müssen fehlerhafte Datensätze vermieden oder erkannt werden. Da eine Güte von 100% im Realbetrieb sehr unrealistisch ist, wird eine Güte von 95% festgelegt. Das bedeutet, dass bei 100 Messungen lediglich 5 fehlerhafte Messungen auftreten dürfen.

3.4 Randbedingungen

Zu den Randbedingungen zählen im Kontext von SPARK das zur Verfügung stehende Budget sowie die Beschränkung auf temporäre Messanlagen. Das Budget für die gesamte Messanlage wird auf 500€ festgelegt, welches für Material zur Verfügung steht.

Da vorerst keine baulichen Maßnahmen durchgeführt werden können und dürfen, muss die Messanlage temporär aufgebaut werden können.

Eine weitere Randbedingung ist die Witterungsbeständigkeit der Messanlage. Da diese ungeschützt im Außenbereich eingesetzt wird, muss diese vor allem gegen Wasser geschützt werden.

3.5 Sekundärstatistiken

Sekundärstatistiken liefern Aussagen aus den Daten, welche nicht dem primären Untersuchungsziel dienen. Aus den erhobenen Daten kann zusätzlich ein Füllstand des Parkplatzes errechnet werden. Über die Kapazität und einen Zeitstempel bei Befahren des Parkplatzes kann errechnet werden, wann die Kapazität eines Parkplatzes erreicht ist. Daraus kann abgeleitet werden, wann in der Zukunft Engpässe zu erwarten sind.

3.6 Erfassungseinheit(-en)

Bei der Erkennung eines Fahrzeuges müssen folgende Werte mit den entsprechenden Einheiten erfasst werden:

- DateTime für Zeitstempel (Sekundengenau)
- String für ID der Einfahrt
- String für ID des Sensors
- String für die Fahrtrichtung (rein oder raus)

3.7 Hypothesen

Um im weiteren Verlauf statistische Methoden wie einen Hypothesentest anzuwenden, sollen laut FGSV zu Beginn Hypothesen aufgestellt werden. Diese Hypothesen sollen mit der Durchführung der Erhebung überprüft werden. Jedoch handelt es sich bei der Erhebung vorerst um eine Vorstudie, welche zum Ziel hat einen Überblick über die Parkplatzsituation zu erhalten. Daher werden nachfolgend statt Hypothesen Leitfragen aufgestellt:

F1: Wie hoch ist das Verkehrsaufkommen?

F2: Lässt sich durch den Einsatz von SPARK der Verkehr auf dem Parkplatz reduzieren?

F3: Um welchen Prozentsatz lässt sich der Parkplatzverkehr durch SPARK reduzieren?

3.8 Merkmale

Neben den bereits genannten Zielen und Bedingungen, ist ein weiteres Merkmal dieser Verkehrserhebung, dass eine Integration in die bestehende SPARK Infrastruktur erfolgen muss. Dazu muss die Messanlage in der Lage sein, die erfassten Datensätze an eine REST-Schnittstelle zu senden.

3.9 Datenschutz

Nicht erst seit dem Inkrafttreten der neuen EU-Datenschutzgrundverordnung spielt das Thema Datenschutz eine wichtige Rolle in Deutschland. Denn „grundsätzlich ist der Fahrzeughalter der Dateneigentümer und besitzt die Verwendungshoheit“ [2] der Fahrzeugdaten.

Darüber hinaus können die Fahrzeugdaten in *gemeinnützige anonyme Daten* oder *fahrzeugspezifische Daten* kategorisiert werden [2]. Zu den gemeinnützigen Daten gehören unter anderem, die für diese Erhebung wichtigen Positionsdaten sowie ein Zeitstempel. Diese können in anonymisierter Form gemeinnützig verwendet werden [2]. Konkret bedeutet dies, dass eine Verwendung zulässig ist, solange keine Personendaten mit den Positions- und Zeitdaten verknüpft werden.

3.10 Anforderungen

Um die verschiedenen Technologien vergleichen zu können, müssen Anforderungen aufgestellt werden, gegen welche die Technologien geprüft werden. Messanlagen lassen sich nach der FGSV durch insgesamt fünf Merkmale unterscheiden. Zu den Merkmalen der FGSV werden jeweils Anforderungen an unsere Messanlage für SPARK definiert, welche aus den vorhergegangen Kapiteln abgeleitet wurden:

Physikalische Arbeitsweise (Messprinzip)

Unterscheidet die zugrundeliegende Technologie und Arbeitsweise der Messanlage wie beispielsweise Laser oder Radar. [8]

A1 Erkennung: Die Messanlage muss in der Lage sein ein Fahrzeug korrekt zu erkennen.

Verkehrstechnische Verwendung

Unterschiedet zwischen verschiedenen Verkehrskenngrößen wie beispielsweise Anwesenheitserkennung, Zählen, Richtungs- und Fahrzeugerkennung. [8]

A2 Fahrtrichtung: Die Messanlage muss die Fahrtrichtung eines Fahrzeugs erkennen können. Dies dient der Unterscheidung von Betreten und Verlassen des Parkplatzes.

Mess- und Ausgabegrößen

Unterscheidet die Grundinformationen, welche von der Messanlage generiert werden [8]. Dies kann beispielsweise ein Zeitstempel oder eine Objekterkennung auf einem Videobild sein.

A3 Messgröße: Die Messanlage muss Daten in Form eines Zeitstempels liefern. Pro erkanntem Fahrzeug soll ein Zeitstempel generiert werden. Die Fahrtrichtung ist über den jeweiligen Sensor bekannt.

Bautechnische Anordnung

Unterscheidet die Anordnung und den Aufbau der Messanlage in, auf, neben oder über der Verkehrsfläche. [8]

A4 Aufbau: Die Messanlage muss temporär zur Messung errichtet werden können. Nach der Messung muss diese entfernt werden können.

Wirkungsbereich

Unterscheidet den räumlichen Wirkungsbereich, in welchem die Messanlage arbeitet. [8]

A5 Reichweite: Die Messanlage muss eine Einfahrt von 10 Meter breite abdecken können

Neben den Anforderungen, welche von den von der FGSV definierten Merkmalen abgeleitet wurden, werden weitere Anforderungen definiert:

A6 Integration: Die Messanlage muss in die bestehende Infrastruktur integriert werden können. Dazu muss der Zeitstempel einer Erkennung an eine REST-API übermittelt werden.

A7 Witterung: Die Messanlage muss auch bei schlechter Witterung wie Regen, Schnee oder starker Sonneneinstrahlung funktionsfähig sein.

A8 Datenschutz: Die Messanlage darf keine personenbezogenen Daten erheben und verarbeiten.

A9 Kosten: Die Messanlage darf Kosten von insgesamt 500€ nicht überschreiten.

A10 Güte: Die Messanlage muss eine Zuverlässigkeit von 95% haben. Das heißt, dass von 100 Erkennungen 95 korrekt als Fahrzeug klassifiziert werden müssen.

4 Erhebungsvorbereitung

Im zweiten Arbeitsschritt wird das Konzept aus dem ersten Schritt weiter konkretisiert. Dabei wird in dieser Arbeit ein Fokus auf die Technologieauswahl gelegt.

4.1 Erhebungsmethode

Bei dieser Erhebung handelt es sich um eine *verkehrstechnische Erhebung*. Dabei werden die Ortsveränderungen von Fahrzeugen beobachtet [6].

4.2 Zählformular

Die FGSV beschreibt diesen Punkt sehr universell. Ein Zählformular kann auf der einen Seite ein Formular aus Papier sein und auf der anderen Seite eine Datenbank. Letzteres ist bei SPARK der Fall. Sensoren erfassen Fahrzeuge automatisiert und die Messanlage meldet die Daten in Echtzeit an das SPARK-Backend, welches die Daten in einer Datenbank abgelegt.

4.3 Überprüfung technischer Geräte

Dieses Unterkapitel vergleicht die verschiedenen Systeme und Technologien Induktionsschleifen, Radar, Laser, Ultraschall, Kameras und Infrarot. Diese Auswahl ergibt sich aus den Anforderungen 1 und 2, da nach [7] diese Technologien Fahrzeuge sowie die Fahrtrichtung erkennen können.

Hersteller von existierenden messanlagen arbeiten in der Regel mit Ihrem eigenen Backend, so dass eine Anpassung der Implementierung seitens der Hersteller nötig ist, um das System in SPARK zu integrieren [4]. Daher ist für eine Integration in SPARK der Eigenbau einer Messanlage nötig.

Für den Vergleich der Technologien werden diese in Tabelle 1 mit einer Skala von 1 bis 10 bewertet. Dabei ist die beste Bewertung eine 10. Die Daten wurden aus Experteninterviews ([4]), Literaturrecherche ([7], [8]) und dem Bau von Prototypen erfasst und entsprechend der Anforderungen und dem Kontext von SPARK bewertet.

Neben der Bewertung wird im Folgenden kurz auf die jeweils wichtigsten Merkmale eingegangen.

4.3.1 Induktionsschleifen

Eine Induktionsschleife eignet sich ideal um Verkehr zu zählen. Jedoch wird ein temporärer Aufbau durch die Reifenbewegung und die dadurch entstehenden Kräfte vom Boden gelöst [4]. Doch auch fest in den Boden eingelassene Induktions-schleifen haben durch „[…] Frostbruch und Verspannungen innerhalb der Fahrbahnstruktur sehr hohe Ausfallquoten von bis zu 2/3.“ [2]

Aufgrund der Komplexität entstehen hohe Kosten bei der Anschaffung bestehender Induktionsschleifen. Diese Messanlagen bieten bereits eigene Systeme zur Weiterverarbeitung der Daten und können dadurch nicht ohne weiteres in SPARK integriert werden. [4]

4.3.2 Radar

Als Besonderheit beim Radar ist hervorzuheben, dass ein Verkehrsfluss von etwa 3 Km/h sowie gute Witterung die Voraussetzungen für eine hohe Güte der gemessenen Daten sind. Weiterhin sind die erfassten Daten sehr komplex. Ein bestehendes Radarsystem ist in der Anschaffung teuer und nicht ohne weiteres in SPARK integrierbar. [4]

4.3.3 Laser

Eine Laserlichtschranke benötigt ein durchgehendes optisches Sichtfeld zum Photowiderstand. Ist dieses durch eine schlechte Witterung wie Schnee unterbrochen, werden fehlerhafte Erkennungen verursacht. Daher ist ein Einsatz bei schlechter Witterung nicht möglich.

Bei der Güte der gemessenen Daten müssen Abstriche gemacht werden, da die Laserlichtschranke sehr empfindlich auf Störungen im Erfassungsbereich reagiert. So werden bereits bei kleinsten Unterbrechungen Objekte erkannt.

Dem kann softwareseitig entgegengewirkt werden, indem ein Fahrzeug nur bei einer längeren Unterbrechung erkannt wird.

4.3.4 Ultraschall

Da ein Ultraschallsensor ein freies Sichtfeld auf das zu erkennende Objekt benötigt, kann er durch schlechte Witterung wie Schnee blockiert werden. Dabei ist er jedoch nicht so fehleranfällig wie eine Laserlichtschranke, was kleine Störobjekte im Erfassungsbereich angeht.

Jedoch ist der Prototyp derzeit nicht gegen Wasser und Feuchtigkeit geschützt. Diesem Problem kann mit einem wasser-dichten Ultraschallsensor begegnet werden.

4.3.5 Kamera

Mittels des von einer Kamera erzeugten Bildes lassen sich Fahrzeuge erkennen. Die erhobenen Daten sind Bilddaten, welche in der Auswertung sehr komplex sind.

Da Bilddaten erhoben werden schneidet die Kamera beim Datenschutz am schlechtesten ab. Auch die Kosten sind bei Kameras im Vergleich zu den anderen Technologien sehr hoch. Dabei liefert eine Kamera Daten in hoher Güte, wenn die Witterung entsprechend gut ist. [4]

4.3.6 Infrarot

Im Projekt SPARK wird bereits ein Infrarotsensor für die Erkennung des Status eines Stellplatzes eingesetzt. Daher ist die Integration eines Eigenbaus ohne weiteren Aufwand möglich.

Der größte Nachteil ist jedoch die Sonneneinstrahlung, welche die Messung „überbelichtet“. Somit ist die Erkennung eines Fahrzeuges nicht mehr möglich. Dieses Problem tritt bereits bei Tageslicht auf.

Tabelle 1: Sensor-Vergleichsmatrix

Anforderung	Induktion	Radar	Laser	Ultraschall	Kamera	Infrarot
A1 Erkennung	10	7	10	10	10	7
A2 Fahrtrichtung	7	10	7	7	10	7
A3 Messgröße	10	3	10	8	3	7
A4 Aufbau	5	5	8	8	5	5
A5 Reichweite	10	10	10	10	10	10
A6 Integration	2	2	7	7	2	8
A7 Witterung	6	6	5	7	6	4
A8 Datenschutz	10	10	10	10	0	10
A9 Kosten	3	3	10	10	3	10
A10 Güte	8	6	4	7	6	3
	71	**62**	**81**	**84**	**55**	**71**

4.3.7 Technologieauswahl

Nachdem die zur Verfügung stehenden Technologien beschrieben und ausgewertet wurden, wird nun die für SPARK passende Technologie ausgewählt. Mit allen Technologien lassen sich Fahrzeuge sowie deren Fahrtrichtung erkennen. Zusätzlich erlauben es alle - mehr oder weniger -, bei einer Erkennung einen Zeitstempel zu generieren. Unterschiede treten vor allem beim Einsatz unter schlechten Witterungsbedingungen auf. So ist die Induktionsschleife die einzige Möglichkeit, wetterunabhängig zu Messen. Alle anderen Technologien benötigen ein direktes und freies Sichtfeld zum Fahrzeug. Messanlagen auf Basis von Radar, Infrarot oder Kameras sind sehr komplex in der Herstellung. Dementsprechend hoch sind auch die Preise bei der Beschaffung bestehender Systeme. Dies führt dazu, dass diese drei Technologien an dieser Stelle ausscheiden. Die Induktionsschleife ist in der Herstellung auch sehr komplex. Allerdings lässt sich diese bereits für etwa 3.000€ betriebsbereit beschaffen und ist damit deutlich günstiger als die drei bereits ausgeschiedenen Systeme [4]. Eine Laserlichtschranke, ein Ultraschallsensor oder eine auf Infrarot basierende Messanlage konnte sich als Prototyp preislich bewähren. Diese sind mit vertretbarem Aufwand und geringen Kosten selbst herzustellen. Infrarot wird allerdings im Freien schnell durch Sonneneinstrahlung negativ beeinflusst. Die Laserlichtschranke ist sehr anfällig gegenüber kleinsten Störobjekten im Erfassungsbereich. Dies führt dazu, dass diese von kleinen Blättern oder ähnlichen nicht kontrollierbaren Gegenständen ausgelöst wird, was eine Fehlerkennung zur Folge hat. Daher wird an dieser Stelle die Verwendung einer Ultraschall-Messanlage empfohlen, da diese im Vergleich zur Laserlichtschranke robuster gegenüber kleinen Störungen ist.

Nach der detaillierten Überprüfung technischer Geräte, folgen im weiteren nächsten Vorgehensschritte der FGSV.

4.4 Stichprobe

In der ersten Vorstudie kann noch keine Stichprobengröße definiert werden. Wenn im weiteren Verlauf Hypothesen definiert werden, ist für den Hypothesentest jedoch eine Stichprobengröße zu bestimmen.

4.5 Zeit und Dauer

Um ein Gesamtbild der Parkplatzsituation zu erhalten, ist eine Messung an allen Wochentagen rund um die Uhr nötig. Um Tage mit besonders hohem/niedrigen Verkehrsaufkommen kompensieren zu können ist die Messung 2 Wochen am Stück durchzuführen.

4.6 Räumliche Abgrenzung

Der Parkplatz, welcher überwacht werden soll ist in zwei Bereiche aufgeteilt. Den größeren P2 und den kleineren P3. Über die Einfahrt 1 lassen sich beide Bereiche erreichen. Somit teilt sich die Einfahrt 1 in zwei mögliche Richtungen. Über die Einfahrt 2 lässt sich der Parkplatz P3 erreichen. Dies ist in Abbildung 2 veranschaulicht. Gemessen werden darf dementsprechend nur an den Einfahrten 1.1, 1.2 und 2.

Abbildung 2: Zufahrten

Insgesamt werden drei Messanlagen benötigt, welche jeweils eine Einfahrt abdecken. Jede Messanlage verfügt über zwei Sensoren, welche jeweils einen Fahrstreifen einer Einfahrt abdecken. Die Haupteinfahrt 1 wird nicht ausgestattet, da sich diese in zwei kleinere Einfahrten aufteilt. Somit lässt eine Messung an Einfahrt 1 keinen Rückschluss an den tatsächlichen Parkbereich zu.

4.7 Datengenauigkeit

Unter der Datengenauigkeit kann auch der Detailierungsgrad der Daten verstanden werden [7]. Dieser ist vor allem auch vom Datenschutz abhängig. Daher werden lediglich die folgenden Informationen zu einem Datensatz erfasst:

- Zeitstempel
- ID der Einfahrt
- ID des Sensors
- Fahrtrichtung

5 Fazit

In dieser Arbeit wurden die Grundlagenkonzipierung sowie die Erhebungsvorbereitung zur Evaluierung von SPARK durchgeführt. Ziel der Evaluierung ist es, die Auswirkungen des reservierungsbasierten Verkehrsleitsystems SPARK auf die Parkplatzsituation zu messen. Dazu wurde das Vorgehensmodell einer Verkehrserhebung der FGSV als Grundlage herangezogen. Basierend auf diesem Vorgehensmodell wurden die ersten beiden Arbeitsschritte - Grundlagenkonzipierung und Erhebungsvorbereitung - in dieser Arbeit durchgeführt. Bei der Grundlagenkonzipierung wurden vor allem der Anlass sowie das Ziel der Erhebung konkretisiert. Anschließend wurden daraus Anforderungen an Technologie und die Erhebung abgeleitet.

Mit diesen Anforderungen wurde im zweiten Arbeitsschritt - der Erhebungsvorbereitung - ein Vergleich verschiedener Technologien durchgeführt. Zu den verglichenen Technologien gehören Laser, Ultraschall, Radar, und Kameras. Dieser Vergleich hatte zum Ziel, die am besten geeignete Technologie auszuwählen. Nachdem die Unterschiede der Technologien herausgestellt wurden, ist Ultraschall die empfohlene Technologie. Für den Vergleich wurden Prototypen gebaut und verglichen.

Zu Radar und Kamera wurden Experteninterviews geführt. Diese Arbeit bietet die Basis für weiterführende Arbeiten an der Evaluierungsmethode. So sind die weiterführenden Arbeitsschritte aus dem Vorgehensmodell durchzuführen. Dies sind vor allem die Durchführung der Erfassung sowie die Analyse der Daten.

Literaturverzeichnis

[1] INRIX. 2017. *Durchschnittliche Suchzeit für Parkplätze in deutschen Großstädten nach Art des Parkens (Stand: 2017; in Minuten pro Fahrt)*. https://de.statista.com/statistik/daten/studie/732264/. Accessed 26 August 2018.

[2] Proff, H., Pascha, W., Schönharting, J., and Schramm, D. 2013. *Schritte in die künftige Mobilität*. Springer Gabler, Wiesbaden.

[3] Hilde. 2010. *Ermittlung von Standards für anforderungsgerechte Datenqualität bei Verkehrserhebungen*. Berichte der Bundesanstalt für Straßenwesen Verkehrstechnik 200. Wirtschaftsverlag NW Verlag für neue Wissenschaft GmbH, Bremerhaven.

[4] Mario Lippoldt. 2018. *Integration einer Verkehrszählung in SPARK*.

[5] Ensminger, D. and Bond, L. J. 2012. *Ultrasonics. Fundamentals, technologies, and applications*. Mechanical engineering series (Boca Raton, Fla.). CRC Press, Boca Raton, FL.

[6] Schmietendorf, G. 2011. *Verkehrsdatenerfassung mit Bluetooth-Detektion: Möglichkeiten und Grenzen*. Technische Universität Dresden, Dresden.

[7] FGSV. 2012. *Empfehlungen für Verkehrserhebungen. EVE*. FGSV 125, Köln.

[8] FGSV. 1991. *Merkblatt über Detektoren für den Strassenverkehr*. FGSV, Köln.

Optimierung des Informationstransfers bei der sensorischen Substitution

Eine theoretische Betrachtung auf Basis der Informationstheorie

Sina Mailin Frommer
Hochschule Reutlingen
sina_mailin.frommer@student.reutlingen-university.de

Abstract

Sensorische Substitutionssysteme dienen der Kompensation verlorener Sinnesmodalitäten. Dabei kommt es zu einem Zusammenspiel zwischen Maschine und menschlicher Sensorik, bei dem oftmals nur wenig Information an die menschliche Wahrnehmung vermittelt werden kann. Diese Arbeit beschäftigt sich mit der Frage, wie der Informationstransfer bei der sensorischen Substitution optimiert werden kann. Durch eine Betrachtung auf Basis der Informationstheorie, die eine abstrahierte Modellierung des Informationsweges ermöglicht, wurden verantwortliche Stellen für einen Informationsverlust identifiziert und Strategien zur Optimierung erarbeitet.

CCS Concepts

• **Information systems** → **Data exchange**;
• **Human-centered computing** → *HCI theory, concepts and models*;

Keywords

Sensorische Substitution, Informationstheorie, Informationstransfer, Modellierung

Betreuer:	Prof. Dr.-Ing. Cristóbal Curio Hochschule Reutlingen cristobal.curio@reutlingen-university.de
Co-Betreuerin:	Salma Maath Ahmad Thalji Hochschule Reutlingen salma.thalji@reutlingen-university.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen

Copyright 2018 Sina Mailin Frommer

1 Einleitung

Die Wahrnehmung über unsere Sinnesorgane ist entscheidend um sich in der Welt zurecht zu finden – sie umfangreich wahrzunehmen. Fällt eine Sinneswahrnehmung weg, zum Beispiel durch einen Unfall, reduziert sich signifikant die Menge an Informationen die wir von unserer Umwelt mitbekommen. Um diesen Informationsverlust zu kompensieren, werden sensorische Substitutionssysteme eingesetzt. Die Wahrnehmung kann damit, zumindest stückweise, wieder vervollständigt werden. Der Informationstransfer zwischen Maschine und menschlicher Sensorik in derartigen Systemen stellt jedoch eine Herausforderung dar. Sensorische Substitutionssysteme schaffen es zumeist nur, eine sehr geringe Teilmenge der Informationen an die menschliche Wahrnehmung zu übermitteln und erfordern zudem einen hohen Trainingsaufwand.

Die Untersuchung des Informationsweges beim Einsatz eines sensorischen Substitutionssystems kann darüber Aufschluss geben, an welcher Stelle potentiell ein Informationsverlust auftritt. In der vorliegenden Arbeit wird eine derartige Betrachtung, auf Basis der Informationstheorie von Shannon [1], durchgeführt und so allgemeine Prinzipien zur Optimierung des Informationstransfers bei der sensorischen Substitution erarbeitet. Dabei wird in den ersten beiden Kapiteln zunächst das Prinzip der sensorischen Substitution näher beleuchtet und ein Grundverständnis für die relevanten Aspekte der Informationstheorie geschaffen. Nachdem daraufhin in Kapitel 3 der aktuelle Forschungsstand auf dem

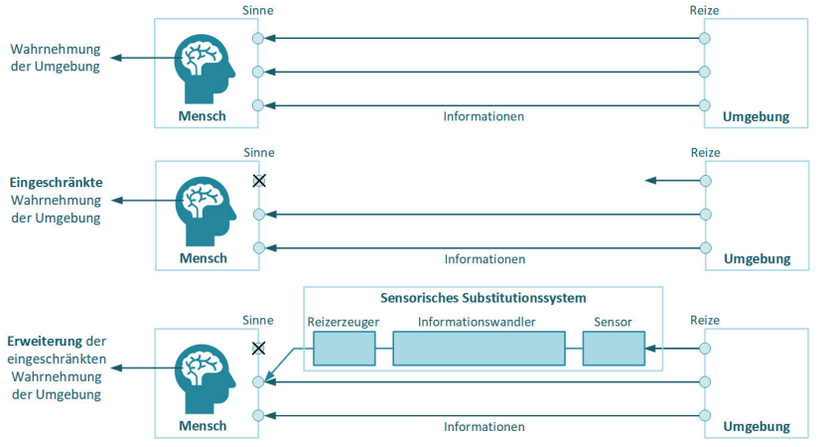


Abbildung 1: Darstellung des Ablaufs bei der Wahrnehmung der Umgebung und der Einschränkung bei Wegfallen einer Sinnesmodalität. Visualisierung der verbesserten Wahrnehmung durch den Einsatz eines sensorischen Substitutionssystems.

Gebiet der sensorischen Substitution unter informationstheoretischen Aspekten dargelegt wurde, folgt die Betrachtung des Informationsweges in Kapitel 4 und 5. Es wird dabei zunächst der Informationsweg beim Einsatz von sensorischen Substitutionssystemen auf Basis der Informationstheorie modelliert und darauffolgend analysiert. Kritische Stellen beim Informationstransfer werden damit identifiziert und Strategien und Ansätze zur Optimierung entwickelt und dargelegt.

2 Das Prinzip der sensorischen Substitution

Um seine Umgebung wahrnehmen zu können, besitzt der Mensch verschiedene Sinnesorgane, wie Augen, Ohren oder die Haut. Diese agieren wie Schnittstellen zur Umgebung und reagieren auf spezifische Reize, deren Signale sie aufnehmen und an das Gehirn zur Verarbeitung weiterleiten [2]. Aus den so gewonnenen Informationen konstruiert das Gehirn ein Gesamtbild, eine Abbildung der Umgebung, die Wahrnehmung [2]. Abbildung 1 visualisiert dieses Prinzip. Die Wahrnehmung ist für Lebewesen entscheidend um sich in ihrer Umgebung zurecht zu finden und mit dieser zu interagieren [2]. Bei Ausfall eines Sinnesorgans, siehe ebenfalls Abbildung 1, fallen die darüber geleiteten Informationen weg. Aus den übrigen Informationen kann nur ein lückenhaftes Bild der Umgebung konstruiert werden und die Interaktion mit dieser wird erschwert. Um die Wahrnehmung in diesem Fall zu verbessern, können sensorische Substitutionssysteme eingesetzt werden. Sie nehmen die Umgebungsreize der gestörten Sinnesmodalität wahr und wandeln sie in Reize einer funktionierenden Sinnesmodalität um [3]. Das System macht dem Menschen damit die verlorenen Informationen über einen intakten Sinn wieder zugänglich.

Ein sensorisches Substitutionssystem besteht typischerweise, wie in Abbildung 1 dargestellt, aus den drei Komponenten Sensor, Informationswandler und Reizerzeuger [3]. Dabei nimmt der Sensor die Reize der verlorenen Sinnesmodalität auf und gibt diese Informationen an die Wandlerkomponente weiter [3]. Sie bildet die Informationen auf wahrnehmbare Reize eines intakten Sinnes ab. Durch die dritte Komponente, den Reizerzeuger, werden

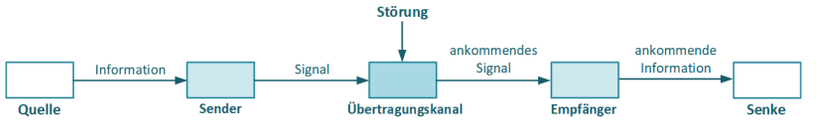


Abbildung 2: Blockdiagramm eines Kommunikationssystems, allgemeine Darstellung nach Shannon [1].

die Reize mechanisch erzeugt und so die Informationen an den funktionierenden Sinn geleitet [3]. Das auf diese Weise stimulierte Sinnesorgan leitet die Informationen über die Nervenbahnen an das Gehirn, dessen Aufgabe es ist, sie zu entschlüsseln und korrekt in das Gesamtbild einzufügen [3]. Die Plastizität des Gehirns macht es dabei durch Training möglich, sich an neuartige Informationen anzupassen [4]. Es lernt sie zu entschlüsseln, wodurch sich das Gesamtbild vervollständigt und die Wahrnehmung verbessert [4].

3 Einführung in die Informationstheorie

Die Informationstheorie bietet eine theoretische Grundlage zur Betrachtung von Übertragungswegen bei der Informationsübermittlung in Kommunikationssystemen [5]. Sie wurde 1948 von Claude E. Shannon begründet, der in seiner Publikation „*A Mathematical Theory of Communication*“ [1] erstmals den Informationsbegriff quantifizierte [5]. Information ist nach Shannon der Gewinn von neuen Erkenntnissen aus einer Quelle [6]. Er misst also den Informationsgehalt einer Nachricht an dessen Grad an Neuigkeit für den Empfänger. Für Shannon spielt dabei die Bedeutung der übertragenen Information keine Rolle, es geht vielmehr um die verlustfreie Übertragung der Information und die Bestimmung der Möglichkeiten und Grenzen eines Übertragungsweges [1]. Dafür modelliert er ein allgemeines Modell zur Informationsübertragung in Kommunikationssystemen, welches in Abbildung 2 dargestellt ist. Ein Kommunikationssystem besteht demnach aus zwei Kommunikationspartnern, einer Quelle und einer Senke. Die Quelle besitzt eine Information, die sie an die Senke kommunizieren möchte [1]. Um die Information zu senden nutzt die Quelle ein Signal als Träger [6]. Das Signal wird über einen Übertragungskanal an die empfangende Instanz der Senke geleitet [6]. Aus dem eingehenden Signal wird die Information extrahiert und erreicht auf diese Weise die Senke [6]. Auf dem Übertragungskanal können Störungen auf das Signal einwirken, sodass das ausgesendete Signal nicht dem empfangenen entsprechen muss und damit ein Informationsverlust auftreten kann [6].

Informationsquellen können in zwei Klassen eingeteilt werden, kontinuierliche und diskrete Quellen. Quellen verfügen dabei über eine Menge möglicher Ereignisse, auch Zeichenvorrat oder Alphabet genannt. Diskrete Quellen haben nur eine endliche Menge an Zeichen zur Verfügung und erzeugen somit auch nur eine begrenzte Anzahl an Ereignissen, zum Beispiel Wörter aus einem Alphabet. Kontinuierliche Quellen haben hingegen einen unbegrenzten Zeichenvorrat und können so eine unendliche Anzahl an möglichen Ereignissen in einem vorgegebenen Bereich erzeugen, beispielsweise Amplitudenwerte bei einem analogen Signal. [6]

Die Abfolge der Ereignisse einer Quelle können entweder unabhängig oder voneinander abhängig sein, wie zum Beispiel die Reihenfolge der Buchstaben in einem sinnvollen Wort [6]. Es kann aber auch eine übergreifende

Abhängigkeit zwischen zwei oder mehreren Quellen bestehen, sogenannte Verbundquellen, deren Ereignisse sich gegenseitig bedingen [6]. Je nach Art der Quelle variiert der Informationsträger, das Signal, das durch den Sender erzeugt wird, sowie der Übertragungskanal. Signale diskreter Quellen werden über einen diskreten Kanal übertragen, der eine endliche Anzahl an Signalzuständen übertragen kann [5]. Analoge Kanäle übertragen beliebige Zustände und dienen zur Übertragung von kontinuierlichen Signalen [5]. Die Kapazität eines Kanals gibt dabei an, wie viele Informationen pro Zeit ein Kanal übertragen kann, diese ist abhängig von seiner Breite und Beschaffenheit [6]. Desto höher die Kapazität, desto schneller läuft die Informationsübertragung ab [6].
Ein Informationssystem kann auch mehrere unterschiedliche Übertragungskanäle beinhalten, die nacheinander durchlaufen werden. Dabei muss zwischen den Übertragungsschritten eine Signalwandlung stattfinden. Je mehr Kanäle auf dem Weg genutzt werden, umso höher ist die Gesamteinwirkung der Störungen. [5]

4 Stand der Forschung

Im Bereich der sensorischen Substitution gibt es in der Forschung einige Ansätze die sich in Beziehung zu informationstheoretischen Aspekten, wie die Wahl des Übertragungskanals, verfügbare Kanalkapazitäten oder auch Kanalstörungen, stellen lassen. Eine Arbeit, die die sensorische Substitution von einer abstrakteren Ebene beleuchtet und diese Aspekte anspricht, stammt von Kristjánsson et al. In ihrer Publikation *„Designing sensory-substitution devices: Principles, pitfalls and potential“* [7] stellen sie sich der Frage, wie insbesondere visuelle Informationen substituiert werden können und welche Knackpunkte bei dem Entwurf eines sensorischen Substitutionssystems Beachtung finden sollten. Dabei ziehen sie verschiedene Sinnesmodalitäten in Betracht und beleuchten die Bandbreite des visuellen, auditiven und taktilen Kanals. Sie kommen, basierend auf den Arbeiten von Wurtz und Kandel [8] und Proulx et al. [9], zu einer Relation der Kanalkapazitäten. Die sensorischen Kanäle unterscheiden sich demnach in ihrer Bandbreite, wobei der auditive Kanal ein Drittel und der taktile Kanal ein Viertel der Kapazität des visuellen Kanals besitzen. Des Weiteren weisen sie darauf hin, dass die sensorischen Kanäle bereits durch ihre eigentliche Funktion beansprucht sind. Ihre Kapazität zur sensorischen Substitution ist entsprechend eingeschränkt und die eingehenden Signale können sich gegenseitig stören [7]. Sie, wie auch Loomis et al. in ihrer Arbeit *„Sensory substitution of vision: Importance of perceptual and cognitive processing“* [10], kommen zu dem Schluss, dass es deshalb entscheidend ist, nur die, für den Einsatzzweck relevantesten, Informationen zu übertragen. Kristjánsson et al. empfehlen zur Optimierung zudem eine multimodale Lösung, die mehrere sensorische Kanäle für die Substitution nutzt, um so jeden einzelnen Kanal weniger zu beanspruchen [7].
Eine Arbeit von Bialek et al. mit dem Titel *„Bits and brains: Information flow in the nervous system“* [11] konzentriert sich auf die neuronale Informationsverarbeitung unter Aspekten der Informationstheorie. Obwohl in dieser Arbeit die sensorische Substitution keine Rolle spielt, liefert sie wichtige Gesichtspunkte zur Beschaffenheit der sensorischen Kanäle. Lebewesen nehmen ihre Umwelt durch die Veränderungen über die Zeit wahr, wie zum Beispiel Änderungen des Schalldrucks [11]. Diese kontinuierlichen Informationen werden über die Sinneszellen in diskrete Sequenzen von identischen elektrischen Impulsen umgewandelt, die sich nur bezüglich ihres zeitlichen Auftretens unterscheiden und so die Information kodieren [11]. Bei sensorischen Bahnen handelt es sich demnach laut [11] um diskrete Kanäle und, zumindest bei der nicht eingeschränkten Wahrnehmung, um eine kontinuierliche Quelle. Auf

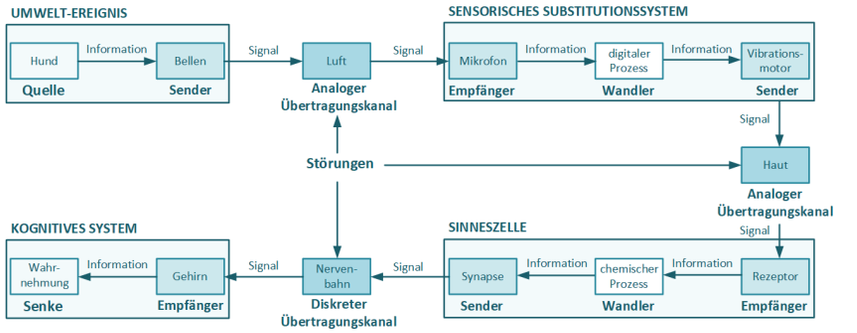


Abbildung 3: Modellierung des Informationsweges beim Einsatz eines sensorischen Substitutionssystems als Blockdiagramm am Beispiel eines Substitutionssystems, das einen auditiven Ausfall durch Substitution mit dem taktilen Sinn kompensiert.

Basis dieses Forschungsstandes wird nun im Folgenden der Informationsweg bei der sensorischen Substitution modelliert.

5 Modellierung des Informationsweges bei der sensorischen Substitution

Wird der Informationsweg beim Einsatz eines sensorischen Substitutionssystems betrachtet, so ergibt sich das in Abbildung 3 dargestellte Blockdiagramm. Es ist nach dem Schema von Shannon [1] modelliert und beinhaltet zwei Zwischeninstanzen, das sensorische Substitutionssystem und die Sinnesmodalität, welche über mehrere Übertragungskanäle hintereinander durchlaufen werden.
Die Quelle dieses Kommunikationssystems bildet die Umwelt, aus der eine Information an die Senke, die Wahrnehmung des Menschen, kommuniziert werden soll. Zum leichteren Verständnis wurde ein Beispiel gewählt, bei dem akustische Signale, das Bellen eines Hundes, durch das sensorische Substitutionssystem in taktile Ereignisse gewandelt werden. Das Bellen erzeugt als Sender das Signal, den Schalldruck, das über den ersten Übertragungskanal an das sensorische Substitutionssystem geleitet wird. Es handelt sich in diesem Fall nach der Definition von [6] um einen analogen Übertragungskanal, der Hund kann in einem bestimmten Frequenzspektrum eine unendliche Anzahl an Ereignissen erzeugen. Im speziellen Beispiel ist dieser Übertragungskanal die Luft, die das Signal in Form der Schalldruckänderungen überträgt. Beim Einsatz eines sensorischen Substitutionssystems übernimmt der Sensor die Aufgabe das Signal zu empfangen. Im konkreten Beispiel würde dies durch ein Mikrofon geschehen.
Ein Wandler übersetzt die erhaltene Information in die Sprache der anzusteuernden Sinnesmodalität. Der Sender des sensorischen Substitutionssystems generiert aus dieser Information die Signale, in diesem Beispiel Bewegungen durch einen Vibrationsmotor, und gibt sie über die Deformation der Haut, einen weiteren analogen Übertragungskanal, an den taktilen Sinn des Menschen. Die Rezeptoren der Sinneszellen in der Haut reagieren als Empfänger, auf diese Signale und es findet in ihnen eine Wandlung statt, ein mechanisch-chemischer Prozess, der die Information in die Sprache der Nerven übersetzt [12]. Die resultierenden Signale, elektrische Impulse, werden von den Sinneszellen an den Synapsen, dem Sender der Sinneszelle, erzeugt und über die Nervenbahnen an das Gehirn geleitet [13]. Bei diesem Kanal handelt es sich,

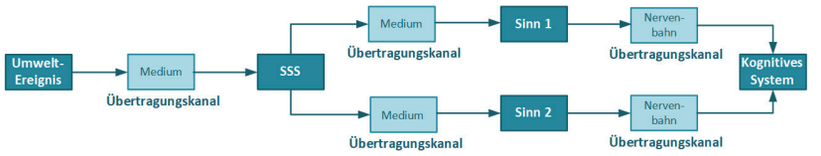


Abbildung 4: Abstrahierte Modellierung des Informationsweges beim Einsatz eines sensorischen Substitutionssystems (SSS) das zwei Sinne zur Übertragung nutzt.

wie bereits im vorherigen Abschnitt beschrieben, nach [11] um einen diskreten Übertragungskanal. Der Empfänger des kognitiven Systems, das Gehirn, nimmt die Signale entgegen, sie werden interpretiert und die Wahrnehmung der Information entsteht.
Werden mehrere Sinneskanäle zur Übertragung der Information eingesetzt, ergibt sich das in Abbildung 4 dargestellte Bild. Die Information wird, wie auch zuvor, von der Umwelt über einen Kanal an das sensorische Substitutionssystem gegeben. Bei diesem handelt es sich in diesem Fall nun um eine Verbundquelle, da die daraus hervorgehenden Ereignisse, die an die unterschiedlichen Sinne gegeben werden, in Bezug zueinander stehen. Die Information wird parallel über die Übertragungskanäle an die Sinneszellen der verschiedenen Modalitäten geleitet. Von dort aus wird sie ebenfalls parallelisiert an die Senke, das kognitive System geleitet. Pro weiterem eingesetzten Sinn zur Substitution kommen zwei zusätzliche Übertragungskanäle zum Einsatz.

6 Optimierung des Informationstransfers bei der sensorischen Substitution

Basierend auf der vorangegangenen Modellierung des Informationsweges beim Einsatz eines sensorischen Substitutionssystems, folgt an dieser Stelle eine Analyse der Komponenten und deren Zusammenhänge. Das Ziel dabei ist es, Stellen im Kommunikationssystem zu identifizieren, an denen ein Informationsverlust auftreten kann. Darüber hinaus sollen Limitierungsfaktoren bestimmt werden, die den Transfer beschränken und Strategien zur potentiellen Optimierung aufgezeigt werden.
Durch den Einsatz eines sensorischen Substitutionssystems, das eine Sinnesmodalität zur Substitution nutzt, wie in Abbildung 3 dargestellt, fügt sich im Vergleich zur normalen sensorischen Wahrnehmung eine Zwischeninstanz und ein zusätzlicher Übertragungskanal in das Kommunikationssystem ein. Dies sind das sensorische Substitutionssystem und der Kanal zwischen sensorischem Substitutionssystem und der anzusteuernden Sinneszelle. Zudem ändert sich die Instanz der Sinneszelle, die die Informationen entgegen nimmt. Die Zwischeninstanz und der zusätzliche Übertragungskanal erhöhen die Dauer für die zeitliche Übertragung der Informationen. Zu spät eintreffende Informationen verlieren ihre Gültigkeit und es wird dem kognitiven System erschwert sie zu interpretieren. Eine sehr schnelle Verarbeitung der Signale durch das sensorische Substitutionssystem ist daher für einen hohen Informationstransfer unerlässlich.
Durch das sensorische Substitutionssystem als Zwischeninstanz ergibt sich die Möglichkeit Störungen des Signals auf dem ersten Übertragungskanal zu korrigieren. Beispielsweise, indem zusätzlich redundante Daten aus der Umwelt aufgenommen, analysiert und zur Korrektur genutzt werden oder durch eine Logikkomponente im sensorischen Substitutionssystem, welches Fehler detektiert und korrigiert. Eine dieser Möglichkeiten sollte für einen optimierten Informationstransfer realisiert werden, um eine Fehlerfortpflanzung über die Übertragungswege hinweg zu verhindern. Die Wandlerkomponente im sensorischen Substitutionssystems kann jedoch auch für einen Informationsverlust verantwortlich sein, da der digitale Prozess bei der Übersetzung eine Quantisierung des analogen Signals erfordert.
Um einwirkende Störungen auf dem zweiten Übertragungskanal, dem zwischen sensorischem Substitutionssystem und den Sinneszellen, zu minimieren, können Redundanzen verwendet werden. Im konkreten Kommunikationssystem, bei dem eine Sinnesmodalität durch die erzeugten Signale angesteuert wird, können Redundanzen zum Beispiel durch örtliche Unterschiede geschaffen werden. Bei einer taktilen Stimulation könnte das Signal beispielsweise gleichzeitig zwei unterschiedliche Stellen am Körper stimulieren.
Die Störungen auf dem dritten Übertragungskanal, den Nervenbahnen, werden durch die Fähigkeit des Gehirns zur Vervollständigung und durch die Redundanzen zu den anderen Sinnen korrigiert. Durch den zusätzlichen Übertragungskanal, der durch den Einsatz des sensorischen Substitutionssystems im Kommunikationssystem dazu kommt, erhöht sich im Gesamten das Störpotenzial, welches auf die Signale einwirkt. Durch die genannten Möglichkeiten, die Korrektur im sensorischen Substitutionssystem und der Einsatz von Redundanzen im erzeugten Signal, könnten diese Einwirkung reduziert werden.
Der Einsatz von Redundanzen in den Signalen führt auf der anderen Seite jedoch zu einer höheren Beanspruchung der Sinneszellen als Empfänger. Die Empfänger der Instanzen begrenzen nicht nur den Übertragungskanal durch ihre Wahrnehmungsbereiche, sondern zudem den gesamten Transfer durch ihre Kapazität. Dabei sind besonders der Empfänger der Sinneszellen und der des kognitiven Systems für eine Limitierung entscheidend, da diese im Gegensatz zum Empfänger des sensorischen Substitutionssystems nicht frei dimensionierbar sind. Die Empfänger der Sinneszellen werden bereits durch ihre eigentliche Funktion beansprucht, sodass ihre Kapazität eingeschränkt ist. Bei gleichzeitig eingehenden Signalen können also Informationen verloren gehen. Signale die zeitlich zu kurz hintereinander eintreffen, können aufgrund der Beschaffenheit von Sinneszellen ebenfalls unter gehen, da sie eine gewisse Regenerationszeit zwischen Stimulationen benötigen [14]. Die Entscheidung welcher Sinn zur Substitution genutzt wird, sollte von der benötigten Kapazität und der spezifischen Physiologie der zugehörigen Sinneszellen abhängig gemacht werden. Die Verfügbarkeit der Kapazität hängt maßgeblich von dem Einsatzzweck ab. Wird das Substitutionssystem zur visuellen Kompensation genutzt, beispielsweise zur Übermittlung von Informationen über die Gestik des Gegenübers in einem Gespräch, so ist es unter diesem Gesichtspunkt nicht ratsam den auditiven Kanal zur Übertragung zu nutzen, da dieser durch den Einsatzzweck *Gespräch* aller Voraussicht nach stark beansprucht sein wird.
Das Gehirn als Empfänger des kognitiven Systems wird ebenfalls durch weitere Sender, die anderen Sinne des Menschen, mit Signalen beliefert. Dies ist zum einen, wie bereits erwähnt, aufgrund der Redundanzen hilfreich, zum anderen können die Signale sich gegenseitig stören oder übertönen. Der Aufmerksamkeitsfokus kann sich dabei auf Kosten des substituierten Signals ändern, gerade wenn viele Signale durch die anderen Sinne gleichzeitig eintreffen. Es kann an dieser Stelle zu einem Informationsverlust kommen. Um einen höheren Informationstransfer zu erlangen, kann es sinnvoll sein, nur dann Signale zu senden, wenn auch Kapazität dafür vorhanden ist. Eine situative Anpassung des Informationstransfers durch das sensorische Substitutionssystem wäre eine denkbare Option. Dies hängt allerdings auch von der Relevanz der substituierten Informationen ab, ob zum Beispiel ein Verzicht möglich ist. Dadurch, dass ein wegfallender Sinn den Einsatz des sensorischen Substitutionssystems nötig macht, ist prinzipiell mehr Kapazität auf Sei-

ten des kognitiven Systems vorhanden. Ob diese zusätzliche Kapazität genutzt werden kann, hängt eventuell auch davon ab wie gut die Hirnbereiche des wegfallenden Sinns mit dem des substituierten verknüpft sind.
Um die Redundanz zu den eintreffenden Signalen der anderen Sinne optimal auszunutzen, ist nicht nur der zeitliche Faktor entscheidend sondern auch der örtliche. Der Bezug der Informationen zueinander entsteht durch die Bewegung des Menschen durch die Umwelt [15]. Der Mensch selbst steuert den Bereich der Umwelt, den er wahrnimmt, indem er sich zu etwas hin- oder von etwas wegbewegt. Er interagiert, erkundet und bestimmt dabei wohin er schaut oder was er berührt. Es entstehen Abhängigkeiten zwischen Aktionen und eingehenden Signalen. Das sensorische Substitutionssystem muss sich also für einen optimierten Informationstransfer mit dem Menschen mitbewegen und durch ihn steuern lassen, sodass Beziehungen in den Informationen für das kognitive System entstehen.
Werden zwei oder mehr Sinne zur sensorischen Substitution eingesetzt, wie in Abbildung 4 modelliert, ergeben sich, durch die Beschaffenheit der Instanz des sensorischen Substitutionssystems als Verbundquelle, weitere Einflüsse auf den Informationstransfer. Die Information kann auf zwei Arten durch das sensorische Substitutionssystem auf die Kanäle verteilt werden, entweder beide Kanäle erhalten redundante Informationen oder die Informationen werden aufgesplittet. Werden sie aufgesplittet, ist die Menge an Signalen auf den einzelnen Kanälen geringer. Um den Störungen auf den Kanälen entgegen zu wirken, sollten die Signale, wie auch bei dem einfachen Modell zuvor, Redundanzen enthalten. Wird die Information auf die zweite Art übertragen, also die selbe Information auf mehreren Kanälen, sind die Redundanzen in den Signalen selbst nicht nötig. Der Störeinfluss im Gesamten ist durch die parallele Struktur der Übertragungswege nicht größer als bei der Substitution über nur einen Sinn. Die Belastung der einzelnen Sinne ist in beiden Fällen geringer als bei der Substitution über nur einen Sinn. Die Abhängigkeiten der Signale können die Interpretation für das kognitive System erleichtern und auch eine Fehlerkompensation ermöglichen, was den Informationstransfer steigert. Eine multimodale Lösung kann sich also lohnen, insofern Kanäle mit ausreichend Kapazität verfügbar sind. Die Dauer für die Übertragung kann jedoch je nach Übertragungskanal variieren. Das zeitverzögerte Eintreffen der Signale könnte die Interpretation erschweren, da die Beziehung zu den eintreffenden Signalen der anderen Sinne verloren geht. Eine schnelle Verarbeitung durch das sensorische Substitutionssystem und eine feine zeitliche Abstimmung für die Stimulation der verschiedenen Sinne ist in diesem Fall wichtig.
Die wohl kritischste Stelle im Informationstransfer beim Einsatz eines sensorischen Substitutionssystems bildet die Dekodierung der Information aus den Signalen durch das kognitive System. In einem Kommunikationssystem benötigt eine Senke eine Art Wörterbuch über die Bedeutung von eingehenden Signalen [6]. Da bei der sensorischen Substitution die Informationen über einen anderen Sinn geleitet werden, werden Signale übertragen für die noch kein Eintrag im Wörterbuch existiert. Das Gehirn muss erst lernen die neuen Signale zu interpretieren, was den Trainingsaufwand für ein sensorisches Substitutionssystem erklärt. Um den Informationstransfer von vorne herein zu gewährleisten und so den hohen Trainingsaufwand zu umgehen müsste auf existierende Verknüpfungen zurückgegriffen werden. Je nach Einsatzzweck können derartige Verknüpfungen existieren.
Um die Mimik unseres Gegenübers deuten zu können, ahmen wir unterbewusst minimal dessen Gesichtszüge nach und können sie so leichter interpretieren [16]. Diese Verhaltensweise, die Mimikry genannt wird [16], ist eine solche Verknüpfung. Fällt der visuelle Sinn weg und der Einsatzzweck eines sensorischen Substitutionssystem ist es, die Informationen über die Mimik an den Menschen weiter zu geben, könnte durch Stimulationen im Gesicht der taktile Sinn angesprochen werden. Die Gesichtsregionen würden entsprechend zu der Mimik des Gegenübers stimuliert und potentiell durch die vorhandene Verknüpfung intuitiv verstanden werden. Spiegelneuronen, spezielle Nervenzellen im Gehirn, bieten eine ähnliche Verknüpfung. Sie zeigen die selbe Aktivität, egal ob wir eine Handlung selbst durchführen oder sie bei jemand anderem beobachten [17]. So kann bei einem Wegfall des taktilen Sinns, durch die visuelle Darstellung einer Berührung ein Gefühl der Berührung hervorgerufen werden [17]. Um den Informationstransfer in sensorischen Substitutionssystemen zu steigern und den Trainingsaufwand zu minimieren ist es ratsam, nach derartigen Verknüpfungen zu suchen und sie zu nutzen. Problematisch könnte diese Strategie sein, falls dem Probanden von Geburt an diese Sinneswahrnehmung fehlt und deshalb eine derartige Verknüpfung eventuell nie entstanden ist. In diesem Fall wäre Zeit und Training nötig um die neue Sprache zu lernen.
Eine letzte Strategie, um den Informationstransfer beim Einsatz von sensorischen Substitutionssystemen zu optimieren, ist die Wahl welche Informationen das System an den Sinn weitergibt. Nach Shannons Auffassung der Information, als Gewinn von neuen Erkenntnissen, könnte es eine denkbare Option sein, nur Unterschiede zu übertragen. Demnach hätten die Signale, die das sensorische Substitutionssystem weiter gibt, einen sehr hohen Informationsgehalt und die Belastung würde sich reduzieren, da zwei gleiche Ereignisse, die aufeinander folgen nur einmal übertragen werden würden. Dies entspräche auch der natürlichen Wahrnehmung des Menschen, der nach [11] seine Umwelt durch die Veränderungen über die Zeit wahrnimmt.

7 Fazit

Durch die Abstraktion im Modell konnte der Informationstransfer für den Einsatz von sehr unterschiedlichen sensorischen Substitutionssystemen beschrieben und analysiert werden. Dabei wurden Stellen im Kommunikationssystem identifiziert, die einen Informationsverlust bedingen. Neben den Übertragungskanälen, zählen hierzu die Empfänger der Sinneszellen und des kognitiven Systems, sowie die Wandlerkomponente im sensorischen Substitutionssystems und das Wörterbuch des kognitiven Systems. Zudem wurden Limitierungsfaktoren für den Informationstransfer detektiert. Besonders der bestehende Informationsfluss durch die reguläre Funktion der Sinne reduziert deren Kapazität für den Informationstransfer durch das sensorische Substitutionssystem. Dieselbe Problematik besteht bei dem kognitiven System, welches die Signale aller Sinne verarbeitet. Einen Limitierungsfaktor stellen zudem die Verknüpfungen im kognitiven System zwischen den verschiedenen Sinnen dar.
Aus der Betrachtung gingen die folgenden potentiellen Strategien für eine Optimierung des Informationstransfers hervor (SSS = sensorisches Substitutionssystem):

- Schnelle Verarbeitung durch SSS
- Fehlerkorrektur durch SSS
- Verlust minimieren bei Quantisierung
- Redundanzen in den Signalen schaffen
- Situative Anpassung des SSS
- SSS durch Proband steuerbar machen
- Fokus auf den Einsatzzweck
- Existierende Verknüpfungen nutzen
- Übertragung von Unterschieden

Die erarbeiteten Ergebnisse sollen künftig als Grundlage zur Entwicklung intuitiverer und effizienterer sensorischer Substitutionssysteme dienen. Die aufgezeigten Strategien könnten einen Anreiz für neue Wege der Optimierung des Informationstransfers bei der sensorischen Substitution darstellen und müssen in nachfolgenden Experimenten auf ihre Tauglichkeit geprüft werden.

Literatur

[1] C. E. Shannon, "A mathematical theory of communication", *BELL SYST TECH J*, Jg. 27, Nr. 3, S. 379–423, Juli 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[2] S. Frings und F. Müller, *Biologie der Sinne - Vom Molekül zur Wahrnehmung*. Springer Spektrum, 2014. DOI: 10.1007/978-3-8274-2273-6.

[3] P. Bach-y-Rita und S. W. Kercel, "Sensory Substitution and the human-machine interface", *TRENDS COGN SCI*, Jg. 7, S. 541–546, 2003. DOI: 10.1016/j.tics.2003.10.013.

[4] S. Maidenbaum, S. Abboud und A. Amedi, "Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation", *NEUROSCI BIOBEHAV R*, Jg. 41, S. 3–15, 2014, Multisensory integration, sensory substitution and visual rehabilitation. DOI: 10.1016/j.neubiorev.2013.11.007.

[5] P. A. Höher, *Grundlagen der digitalen Informationsübertragung*. Vieweg+Teubner Verlag, 2011, ISBN: 978-3-8348-0880-6.

[6] D. Schönfeld, H. Klimant und R. Piotraschke, *Informations- und Kodierungstheorie*. Vieweg+Teubner Verlag, 2012. DOI: 10.1007/978-3-8348-8218-9.

[7] Á. Kristjánsson, A. Moldoveanu, Ó. I. Jóhannesson, O. Balan, S. Spagnol, V. V. Valgeirsdóttir und R. Unnthorsson, "Designing sensory-substitution devices: Principles, pitfalls and potential", *RESTOR NEUROL NEUROS*, Jg. 34, S. 769–787, 2016. DOI: 10.3233/RNN-160647.

[8] R. H. Wurtz und E. R. Kandel, "Central Visual Pathways", *Principles of Neural Science*, Jg. 4, S. 523–545, 2000.

[9] M. J. Proulx, J. Gwinnutt, S. Dell'Erba S.and Levy-Tzedek, A. A. de Sousa und D. J. Brown, "Other ways of seeing: From behavior to neural mechanisms in the online visual control of action with sensory substitution", *RESTOR NEUROL NEUROS*, Jg. 34, Nr. 1, S. 29–44, 2015. DOI: 10.3233/RNN-150541.

[10] J. M. Loomis, R. Klatzky und N. Giudice, "Sensory substitution of vision: Importance of perceptual and cognitive processing", *ASSIST TECHN RES SER*, S. 162–191, Jan. 2012.

[11] W. Bialek, M. DeWeese, F. Rieke und D. Warland, "Bits and brains: Information flow in the nervous system", *PHYSICA A*, Jg. 200, Nr. 1, S. 581–593, 1993. DOI: 10.1016/0378-4371(93)90563-J.

[12] J. Dudel, R. Menzel und R. Schmidt, *Neurowissenschaft: Vom Molekül zur Kognition*. Springer Berlin Heidelberg, 2013, ISBN: 9783642564970.

[13] W. von Seelen und K. Behrend, *Cognitive Systems Monographs: Principles of Neural Information Processing*. Springer International Publishing Switzerland, 2016. DOI: 10.1007/978-3-319-20113-9.

[14] H. Reichert, *Neurobiologie*. Thieme, 2000, ISBN: 3-13-745302-X.

[15] H.-O. Karnath und P. Thier, *Kognitive Neurowissenschaften*. Springer Medizin, 2012. DOI: 10.1007/978-3-642-25527-4.

[16] E. Prochazkova und M. E. Kret, "Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion", *NEUROSCI BIOBEHAV R*, Jg. 80, S. 99–114, 2017. DOI: 10.1016/j.neubiorev.2017.05.013.

[17] M. J. Banissy, R. C. Kadosh, G. W. Maus, V. Walsh und J. Ward, "Prevalence, characteristics and a neurocognitive model of mirror-touch synaesthesia", *EXP BRAIN RES*, Jg. 198, Nr. 2, S. 261–272, Sep. 2009. DOI: 10.1007/s00221-009-1810-9.

Mouse-Tracking als Indikator für den Lernerfolg

Vanessa Willenbrock
Reutlingen University
Vanessa.Willenbrock@Student.Reutlingen-University.DE

Abstract

Bei der Weiterentwicklung von eLearning-Systemen, stellt sich auch die Frage nach der Beurteilung der Lernenden. Feedback ist ein entscheidender Bestandteil des Lernprozesses. Zumeist erfolgt die Beurteilung von Lernenden allein auf ihren Ergebnissen. Der Prozess bis zur Entscheidungsfindung bleibt unberücksichtigt. In einigen Studien konnte bereits gezeigt werden, dass Mouse-Tracking einen Einblick in den Denkprozess der Nutzer gewährt. Bei Unsicherheit führen die Nutzer eine Bewegung aus, die zunächst in Richtung der falschen Antwort oder sich neutral zwischen zwei Antwortmöglichkeiten bewegt. Wenn sie sich ihrer Antwort sicher sind, führen sie hingegen eine zielgerichtete Mausbewegung zur richtigen Antwort hin aus. eLearning-Systeme könnten dies erkennen und dem Lernenden entsprechendes Feedback geben und weitere Übungen anordnen, um Unsicherheiten zu eliminieren. Diese Erkenntnisse sollen auf komplexere Anwendungsfälle im Bereich eLearning übertragen werden. Beispielsweise könnte es Mouse-Tracking ermöglichen, bei einer kieferorthopädischen Modellanalyse, Hinweise darauf zu geben, ob ein Nutzer geübt darin ist oder noch Übung benötigt.

Schlüsselwörter

eLearning, Mouse-Tracking, Learning, Assessment, Simulation

CCS Concepts

• Applied Computing → Education • Applied Computing → Life and medical sciences

1 Einleitung

Lernen ist vor allem dann effektiv, wenn entsprechendes Feedback ausgegeben wird [1] und der Nutzer aus seinen Fehlern lernen kann. Bei der Nutzung von eLearning-Systemen stellt sich daher die Frage nach der Beurteilung der Lernenden. Während es in der Schule oder im Studium die Aufgabe des Lehrers ist Klausuren zu bewerten und zu benoten, sollte in eLearning-Systemen die Beurteilung automatisch erfolgen, um dem Nutzer sofort sein Feedback ausgeben zu können. In einigen Bereichen, wie bei Multiple-Choice-Fragen, ist die Beurteilung recht unkompliziert. Es ist klar definiert welche Antworten richtig und welche falsch sind. Doch durch eLearning und die damit verbundene Technik erschließen sich neue Möglichkeiten zur Beurteilung der Lernenden. So bildet die ergebnisorientierte Beurteilung nicht ab wie der Denkprozess zur

Betreuer Hochschule: Prof. Dr. Oliver Burgert
Hochschule Reutlingen
Oliver.Burgert@Reutlingen-University.de

Betreuer Firma: Dr. Neltje Piro &
Prof. Dr. Dr. Bernd Lapatki
Universitätsklinikum Ulm
Klinik für Kieferorthopädie und Orthodontie

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen

Copyright 2018 Vanessa Willenbrock

Lösungsfindung ausgesehen hat. War die richtige Antwort womöglich nur Zufall? Hat der Nutzer zwar die richtige Antwort gefunden, war sich jedoch unsicher? Eine Antwort auf diese Fragen könnte ein Mouse-Tracking-Tool liefern. Freeman et al. [2] konnten mit ihrer Studie bereits zeigen, dass die Mausbewegungen Aufschluss über die Denkprozesse des Nutzers liefern. „Hand in motion reveals mind in motion“ ist die Erkenntnis zu der sie dabei gelangten. Insbesondere Unsicherheit bei der Entscheidungsfindung lässt sich sehr gut über Mausbewegungen erkennen [3–9] und würde so die Möglichkeit bieten, dem Nutzer ein differenzierteres Feedback über seine Ergebnisse beim Lernen zu liefern. Gerade bei eLearning-Systemen, die am PC ausgeführt werden, ist es ein leichtes Mouse-Tracking einzubinden. Mouse-Tracking ist eine nicht reaktive Untersuchungsmethode [10, S. 165]. Im Vergleich zu Eyetracking ist der Vorteil, dass zum einen keine zusätzliche Hardware benötigt wird und zum anderen der Nutzer in seiner gewohnten Umgebung beobachtet werden kann. Gleichzeitig fühlt dieser sich dabei jedoch nicht beobachtet [10, S. 165].

In dieser Arbeit soll zunächst aufgezeigt werden, wie der aktuelle Stand der Wissenschaft im Bereich von Mouse-Tracking aussieht und im Anschluss darauf eingegangen werden, wie dahingehend der Lernerfolg gemessen werden könnte. Am Ende der Arbeit werden die gewonnenen Erkenntnisse auf ein Anwendungsszenario aus dem Bereich eLearning in der Kieferorthopädie übertragen.

2 Stand der Wissenschaft – Mouse-Tracking

Mouse-Tracking wird in unterschiedlichen Bereichen eingesetzt. Letztlich haben die nachfolgend vorgestellten Anwendungen ein gemeinsames Ziel: den Nutzer besser zu verstehen. Es geht zum einen um die Usability, beispielsweise auf Webseiten und zum anderen darum, Denkprozesse des Nutzers durch Mouse-Tracking abzubilden.

Eine konkrete Suche nach Arbeiten mit Mouse-Tracking-Untersuchungen im Bereich der Simulationen oder 3D-Programmen zur Aufzeichnung und Auswertung von Mouse-Tracking-Daten bei 3D-Fragestellungen brachte keine relevanten Suchergebnisse. Es wurde der Suchstring *("mouse trajectories" OR „mouse tracking“ OR mousetracking OR "mouse-tracking") AND (threedimension* OR 3D OR simulation)* verwendet, um mit den gängigen Suchmaschinen ScienceDirect, IEEE, ACM und PubMed nach diesen Arbeiten zu suchen. Der Suchstring wurde entsprechend den Anforderungen der jeweiligen Suchmaschine angepasst.

2.1 Mouse-Tracking bei Usability-Studien

Ein weitverbreiteter Einsatzzweck für Mouse-Tracking-Tools sind Usability-Studien. Dadurch, dass bereits in einigen Studien gezeigt wurde, dass ein starker Zusammenhang zwischen Augen- und Mausbewegungen besteht, während auf einer Website gesurft wird [4, 11–13], wird es oft für Webdesigns eingesetzt. Zudem ist es sehr leicht, das Mouse-Tracking in die Website zu integrieren, da diese zumeist von einem PC aufgerufen wird.

Unterschiedliche Anbieter haben sich deshalb vor allem auf webbasierte Mouse-Tracking-Tools spezialisiert. Ausgewertet werden diese Programme oft über Heatmaps, die das Nutzerverhalten abbilden sollen. Arroyo et al. [14] verwenden eine Heatmap, die über die Größe der Kreise die zeitliche Intensität und über die Intensität des Kreises die Aktivität in diesem Bereich widerspiegeln. Ein Beispiel hierfür ist in Abbildung 1 zu sehen. So kann bei der Auswertung zum einen betrachtet werden, wie lange sich ein Nutzer mit der Maus in diesem Bereich aufgehalten hat, zum anderen aber auch analysiert werden, wo er besonders aktiv war und mit der Website interagiert hat. Andere Tools bieten zusätzlich an das Mouse-Tracking als Film aufzuzeichnen, sodass dieser zur Auswertung in Echtzeit abgespielt werden kann [15]. Mausbewegungen werden dann als farbige Linien eingezeichnet, um die Nachvollziehbarkeit zu erhöhen [15].

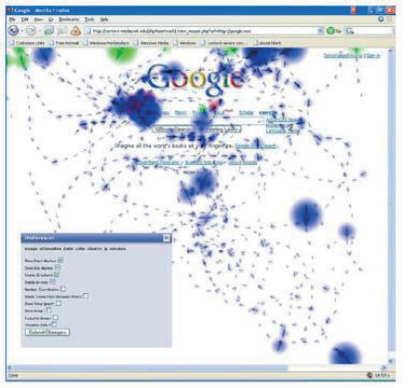


Abbildung 1: Heatmap aus dem Mouse-Tracking-Tool von Arroyo et al. [14]

Durch diese Untersuchungen können beispielsweise Fragen beantwortet werden, wie „Wie schwierig ist es den Link zu finden, der gesucht wird?“.

Dadurch, dass Nutzer in der Regel die Maus dahin bewegen, wo ihre Aufmerksamkeit hingeht, lässt sich nachvollziehen, welchen Bereichen einer Webseite besondere Aufmerksamkeit gewidmet wurde und welchen nicht. Einige Nutzer verwenden die Maus sogar als eine Art Zeiger beim Lesen des Inhalts einer Webseite [12]. Des Weiteren konnte festgestellt werden, dass langsame Mausbewegungen mit gebogenen Laufbahnen für eine unentschlossene Meinung der Nutzer stehen [6, 9].
Wissen die Nutzer nicht, was sie als nächstes tun oder wie sie etwas tun können, dann bewegen sie die Maus langsamer und machen dabei größere Bögen. Navalpakkam et al. [16] gehen mit ihrem Tool einen Schritt weiter. Sie wollen nicht nur das Mouse-Tracking aufzeichnen und im Nachhinein auswerten, sondern mit Hilfe des Mouse-Trackings das Nutzerverhalten analysieren und zudem eine Prognose erstellen, wie die User-Experience aussehen wird. So könnte beispielsweise Frustration des Nutzers erkannt werden und Hilfestellungen eingeblendet werden oder aber auch Inhalte so gesetzt werden, dass sie die individuelle Aufmerksamkeit des Nutzers erregen, was beispielsweise für Werbung sehr nützlich sein kann. Ein Nachteil des Mouse-Trackings ist jedoch, dass subjektive Empfindungen bei der Interaktion mit der Website nicht abgebildet werden können [15]. Auch wenn der Nutzer gut mit dem Aufbau der Website zurechtgekommen ist, kann durch die Mouse-Tracking-Analyse nicht darauf rückgeschlossen werden, ob dem Nutzer die Website gefällt und ob er sich gerne damit beschäftigt. Eine Antwort auf die User-Experience bleibt deshalb aus. Das Tool „m-pathy“ [15] soll für diese Problematik eine Lösung bieten. Hierbei wird das Mouse-Tracking mit OnSite-Befragungen kombiniert, sodass die Informationen über die User Experience zusätzlich gewonnen werden können.

2.2 Psychologische Studien mit Mouse-Tracking

In eine etwas andere Richtung gehen die psychologischen Studien, die mit Hilfe von Mouse-Tracking durchgeführt wurden. Hierbei geht es insbesondere darum die Denkprozesse der Nutzer zu verbildlichen. „Hand in motion reveals mind in motion“ lautet der Leitsatz, den Freeman et al. [2] 2011 aufstellten. Ihre und weitere Studien konnten bereits zeigen, dass Mausbewegungen Aufschluss über die Denkprozesse, die dahinterstecken, geben können. Während früher eine Bewegung als das Ende eines kognitiven Prozesses angenommen wurde [4], so zeigten unterschiedliche Studien jedoch, dass das nicht der Fall ist. Im Gegensatz dazu stellt die Bewegung selbst einen Teil des kognitiven Prozesses dar und wird immer wieder angepasst. Eine Armbewegung ist nicht von Anfang bis Ende als Antwort bestimmt, sondern wird durch kontinuierlich fortlaufende Denkprozesse beeinflusst [8, 17–20]. Mausbewegungen eignen sich deshalb sehr gut, um Denkprozesse abbilden

zu können. Bewegungen beginnen deshalb schon, bevor eine Entscheidung gefallen ist [2]. Diese Erkenntnisse lassen sich sehr gut auf das Mouse-Tracking übertragen, weshalb einige psychologische Studien sich das Mouse-Tracking zu Nutze gemacht haben, um Denkprozesse abzubilden.

Die Forschungsgruppe von Freeman et al. [4] entwickelte ein Tool namens MouseTracker. Das Tool wurde für Aufgaben entwickelt, bei denen zwei Antwortmöglichkeiten zur Verfügung stehen (siehe Abbildung 2). Es zeichnet die Mausbewegungen während der Entscheidungsfindung auf. Dabei startet der Mouse-Cursor immer an einem zentralen Punkt in der unteren Bildschirmmitte und muss für eine Antwort entweder in die obere linke oder obere rechte Ecke des Fensters bewegt werden.

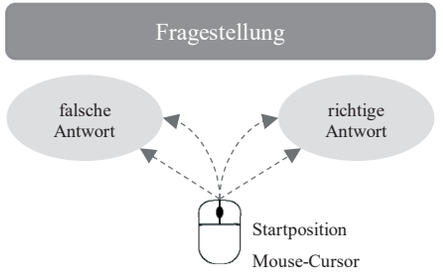


Abbildung 2: Aufbau einer Zuordnungsaufgabe wie von Freeman et al. [4]

Auf diese Weise lässt sich genau analysieren, wie die Probanden ihre Entscheidung treffen, ob sie sich zielstrebig auf eine der Antwortmöglichkeiten zubewegen oder zögerlich zunächst eine der beiden Richtungen vermeiden. Freeman et al. [4] verwendeten ihr Tool selbst, um eine Studie durchzuführen. Es ging dabei um die Zuordnung von Gesichtern zu einem Geschlecht. Sie werteten dabei die zeitliche Dynamik in der Echtzeitwahrnehmung aus und nutzten dazu das Mouse-Tracking. Es wurde untersucht, wie die Probanden auf typische und atypische Gesichter reagieren und wie sie diese einem Geschlecht zuordnen. Sie veränderten dafür beispielsweise weibliche Gesichter und fügten männliche Merkmale hinzu und umgekehrt. Den Probanden wurde dann ein Gesicht präsentiert und die Probanden mussten durch eine Mausbewegung in die rechte oder linke obere Ecke des Bildschirms festlegen, ob sie das Gesicht dem männlichen oder weiblichen Geschlecht zuordnen. Freeman et al. haben dabei festgestellt, dass bei typischen Gesichtern die Maus sehr zielgerichtet zur richtigen Antwort geführt wurde, während bei atypischen Gesichtern die Mausbewegung eher mittig oder sogar in Richtung der falschen Antwort verlief. Erst im Laufe der Bewegung änderte sich die Richtung und wurde erst dann zur richtigen Antwort hin bewegt. Die Probanden haben also eine dynamische aber kontinuierliche Berichtigung der eigenen Interpretation durchgeführt, wie in Abbildung 3 dargestellt.

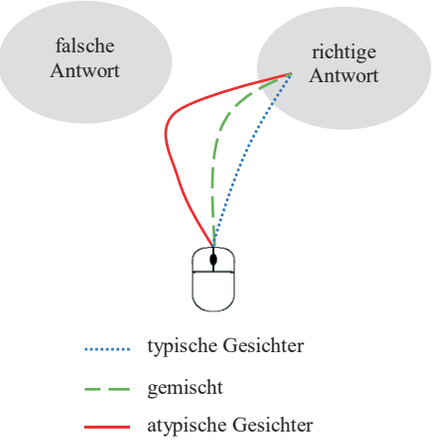


Abbildung 3: Mausbewegungen bei der Zuordnung von typischen und atypischen Fragestellungen, beispielsweise bei Freeman et al. [4]

Ein anderes Gebiet ist die Untersuchung des Sprachverständnisses, was ebenfalls mit Hilfe von Mouse-Tracking untersucht wird. Bei Spivey et al. [8] wird dem Probanden ein Satz wie „Click on the candle“ vorgesprochen, während dem Nutzer zwei Bilder präsentiert werden, bei denen er sich

für eines entscheiden muss. Sie verwenden ebenfalls einen Versuchsaufbau wie in Abbildung 2 dargestellt. Die Aufgaben werden zum Teil so gestellt, dass zu Beginn des Satzes beide Bilder als Antwortmöglichkeit in Betracht gezogen werden können, was ähnlich zu den atypischen Gesichtern oder Tieren ist. Im genannten Beispiel würde das eine Bild eine Kerze („Candle“) und das andere Bild Süßigkeiten („Candy“) zeigen, sodass zu Beginn des Satzes („Click on the can…“) nicht eindeutig ist, welches der beiden Bilder gemeint ist. Dadurch, dass die Wörter so ähnlich sind, kommt es also darauf an, den Satz genau verstanden zu haben, um die richtige Antwort geben zu können. Auch in dieser Untersuchung konnte gezeigt werden, dass bei Fragestellungen, in denen die Antwort nicht eindeutig war, die Maus zunächst in Richtung der falschen bzw. zunächst neutral nach oben geführt wurde, wie in Abbildung 3 ersichtlich. Kam nur eines der Bilder als Antwort in Frage, so wurde die Mausbewegung zielgerichtet in Richtung der richtigen Antwort bewegt.

Eine weitere sprachliche Untersuchung erfolgte auch von Incera et al. [7], die das Verhalten von Bilingualen im Vergleich zu Monolingualen bei einer Stroop-Aufgabe untersuchten. Bei der Stroop-Aufgabe, wie sie in Abbildung 4 dargestellt wird, werden die Namen von Farben in unterschiedlichen Schriftfarben geschrieben. Die Schriftfarbe und der Name der Farbe können dabei übereinstimmen oder unterschiedlich sein.

Abbildung 4: Beispiel für eine Stroop-Aufgabe

In ihrer Studie untersuchten Incera et al., ob monolinguale und bilinguale Menschen, so eine Stroop-Aufgabe unterschiedlich gut bewältigen können oder ob die Aufgabe einer Gruppe schwerer fällt, als der anderen. Sie verwendeten dazu den MouseTracker von Freeman et al. [4] und gestalteten hier einen Versuchsaufbau, bei dem in der Mitte des Bildschirms ein entsprechendes Farb-Wort erschien. Die Probanden sollten dann zuordnen, in welcher Farbe das Wort geschrieben ist und mussten für eine Antwort in die obere linke oder obere rechte Bildschirmecke mit der Maus fahren, in denen die entsprechenden Antwortmöglichkeiten vorgegeben waren. Sie konnten anhand der Mausbewegungen sowohl qualitative als auch quantitative Unterschiede zwischen den beiden Testgruppen aufzeigen. Durch das Mouse-Tracking konnten Sie den gesamten Verlauf der Entscheidungsfindung beobachten, während in vergleichbaren Studien zu Stroop-Aufgaben nur Reaktionszeit oder die Zeit bis zur Antwort gemessen wurde. Dadurch kamen sie zu differenzierten Ergebnissen, sodass sie zeigen konnten, dass die Gruppe der bilingualen Menschen eine größere Initiationszeit benötigte, jedoch die Gruppe der Monolingualen dann einholte und in der Gesamtzeit weniger Zeit benötigte. Sie leiteten daraus ab, dass Bilinguale Experten für Stroop-Aufgaben sind.

3 Lernfortschritt messen

Spätestens seit der Schulzeit ist man immer wieder mit der Fragestellung konfrontiert, wie Lernfortschritt gemessen werden kann. Zur Überprüfung des Lernfortschritts müssen zunächst Lernziele formuliert werden, um dann zu prüfen, ob die Lernenden diese Ziele erreicht haben. Die Überprüfung erfolgt in der Regel mit Hilfe von Tests [21, S. 194]. Im Zuge von eLearning wird die Fragestellung zur Messung des Lernerfolgs spezifischer und komplexer. Eine Aufgabenstellung in einem eLearning-System kann zunächst, ähnlich wie Tests in der Schule, einfach aufgebaut sein. Ebenso kann der Lernstatus allein durch die Beurteilung von richtigen und falschen Antworten erfolgen. Wie zu Beginn dieser Arbeit bereits angedeutet, bietet die Technologie im Bereich eLearning jedoch deutlich mehr Möglichkeiten. Bei

Multiple-Choice-Fragen stellt sich beispielsweise immer die Frage, ob der Lernende die Antwort tatsächlich wusste oder ob diese nur durch Zufall richtig angekreuzt wurde. Zudem ergeben sich durch die Weiterentwicklung der Technologie auch sehr viel komplexere Arten zu lernen, bei denen keine einfache Beurteilung anhand richtiger und falscher Antworten möglich ist. Beispielsweise ist dies im Bereich der Serious Games oder auch der Simulationssoftware der Fall. Mit Hilfe solcher Trainings können Fähigkeiten gezielt verbessert und die Sicherheit der Anwender durch Übung gestärkt werden [22]. Dabei sollen reale Erfahrungen in einer gesicherten Umgebung gesammelt werden [1, 23, 24]. Es ist dabei möglich reale Szenarien zu üben, ohne dabei Menschen oder Gegenstände zu gefährden [25]. Gerade in der Medizin kommen simulationsbasierte Trainings zum Einsatz, um beispielsweise Operationstechniken zu üben und sich die Arbeitsabläufe anzueignen. Durch diese Art der Anwendungen entstehen völlig neue Möglichkeiten für das Lernen, jedoch erfordern sie gleichzeitig auch neue Methoden zur Leistungsbeurteilung. Auch hier ist es notwendig dem Lernenden ein kontinuierliches Feedback über seine Leistungen zu geben, um bestmögliche Lernergebnisse zu erzielen [26].

Dabei können sehr viele Nutzerdaten gesammelt werden. Beispielsweise zeichneten Serrano et al. [27] als GUI-Events Mausklicks, Key Events und weitere Events wie Buttonklicks, Mausbewegungen und ähnliches auf. Sie nutzten diese unterschiedlichen Daten, um daraus Informationen über das Nutzungsverhalten und damit den Lernprozess abzuleiten. So konnten ggfs. ein Spiel oder eine Simulation komplett nachvollzogen werden. Dennoch bleibt die Frage, wie aus den gewonnenen Daten der Lernfortschritt des Nutzers abgeleitet werden kann. Ein Ansatz, um diese Datenmengen verarbeiten zu können ist Learning Analytics. Dabei werden alle anfallenden Daten, die durch die Interaktion mit den online verfügbaren Lerninhalten entstehen, gesammelt und ausgewertet [27]. Bei der Umsetzung von Learning Analytics werden mehrere Schritte verfolgt. Zuerst muss festgelegt werden, welche Daten gesammelt werden sollen, die dann im 2. Schritt aufgezeichnet werden. Nachfolgend werden die gesammelten Daten verarbeitet und in Informationen umgewandelt, auf deren Basis die Bewertung der Lernenden erfolgt.

Insbesondere die Auswertung von Mausbewegungen könnte dabei eine Antwort auf die zuvor erläuterten Fragen und Ansätze liefern. Basierend auf den Ergebnissen aus den psychologischen Studien [4, 7, 8, 16] lässt sich anhand von Mouse-Tracking vor allem die Unsicherheit von Nutzern erkennen. So ließe sich beispielsweise anhand der Mausbewegungen analysieren, ob der Lernende eine Multiple-Choice-Frage durch Zufall richtig beantwortet hat oder ob er sich in seiner Antwort sicher war. Auch für die komplexeren Aufgabenstellungen, wie Spiele oder Simulationen, ließen sich anhand des Mouse-Trackings Informationen über den Nutzer und seine Sicherheit generieren. Das könnte genutzt werden, um zum einen ein differenzierteres Feedback zu erschaffen und zum anderen auch im Lernprozess ggfs. eingreifen zu können. Beispielsweise könnte das eLearning-System erkennen, wenn der Nutzer für seine Aufgaben noch eine hohe Unsicherheit aufweist und ihm vorschlagen, weitere Übungen zu absolvieren.

Wie bereits in Kapitel 2.2 erwähnt geht es in der Studie von Incera et al. [7] darum, mit Hilfe von Mouse-Tracking zu untersuchen wie gut monolinguale Menschen im Vergleich zu bilingualen die Stroop-Aufgabe (vgl. Abbildung 4) bewältigen. In Ihrer Auswertung des Mouse-Trackings gingen Sie im Vergleich zu anderen Studien noch einen Schritt weiter. Sie stellten fest, dass Bilinguale eine kurze Verzögerungszeit einhalten, bevor sie reagieren. Jedoch kommen sie dann meistens zur richtigen Antwort. Im Vergleich dazu reagieren die monolingualen Menschen zwar schneller und

beginnen somit schneller mit der Mausbewegung, jedoch kommen sie dann oftmals zu einem falschen Ergebnis. Daraus schlussfolgerten Sie, dass Bilinguale in Bezug auf Stroop-Aufgaben Experten sind, im Vergleich zu Monolingualen. Sie konnten mit Hilfe des Mouse-Trackings nicht nur eine quantitative Auswertung machen (Bilinguale kamen insgesamt schneller zu einem Ergebnis als Monolinguale), sondern auch eine qualitative Aussage treffen (Bilinguale haben einen anderen Weg ihre Entscheidung zu treffen, als Monolinguale). Anhand dieser Erkenntnisse lässt sich zudem sagen, dass Experten, die zuerst überlegen und dann zielgerichtet auf die richtige Antwort zusteuern, eine geringere Strecke mit der Maus zurückgelegt haben als die Laien, die zwar sofort begonnen haben die Maus zu bewegen, jedoch zunächst auf die falsche Antwort zusteuerten, um dann doch die Maus zur richtigen Antwort zu bewegen. Diese Erkenntnisse könnten ebenso in eLearning-Systemen als Auswerteverfahren integriert werden, um den Lernfortschritt der Nutzer differenziert zu messen.

Jedoch gibt es auch für das Mouse-Tracking, insbesondere in dieser vereinfachten Form, Grenzen. Misst man lediglich die zurückgelegten Strecken in Form von Mausbewegungen, so steht dies nicht im Kontext zur gestellten Aufgabe. Bei der Auswertung ist nicht ersichtlich, wo genau die Mausbewegungen durchgeführt wurden. So könnte es beispielsweise sein, dass ein Usability-Problem vorliegt und der Nutzer mit der Bedienung nicht zurechtkommt und deshalb viele Mausbewegungen macht.

Ebenso wäre es möglich, dass der Nutzer das Programm startet und damit verbunden auch das Mouse-Tracking und sich dann aber dazu entscheidet doch etwas anderes zu machen, im Internet zu surfen oder ähnliches. Dahingehend sollte bei anwendungsübergreifenden Mouse-Tracking-Tools darauf geachtet werden, dass die Daten selektiert werden und lediglich diejenigen Mausbewegungen aufgezeichnet werden, die mit der Aufgabenstellung bzw. der Anwendung verbunden sind, in der die Aufgabe bearbeitet wird. Ebenso muss gewährleistet werden, dass nur diejenigen Mausbewegungen aufgezeichnet werden, die für die Auswertung des Lernfortschritts relevant sind. Beispielsweise müssen die Mausbewegungen sehr genau auf die Phase der Aufgabe abgestimmt sein, in der sich der Nutzer befindet. Hat er zwar das Programm gestartet, aber arbeitet nicht wirklich daran, sondern bewegt die Maus lediglich aus Langeweile oder beispielsweise um sie auf dem Bildschirm wieder zu finden, so sind das Bewegungen, die die Aussage über den Lernfortschritt negativ beeinflussen würden. Hierfür könnten langfristig Algorithmen entwickelt werden, wenn bekannt ist, wie die Mausbewegungen aussehen müssen, um die Aufgabe zu bewältigen und ebenso als Vergleichswerte Aufzeichnungen vorhanden sind, in denen Mausbewegungen durchgeführt wurden, die nicht mit der Aufgabenbewältigung in Zusammenhang stehen.

Ebenso ist es bei der Auswertung von Mausbewegungen wichtig Vergleichswerte zu haben. Ähnlich wie bei Testfragen, muss vorher festgelegt werden was die richtige Antwort ist. Sollen anhand der Mausbewegung Rückschlüsse daraus gezogen werden, wie gut der Nutzer die Aufgabe bewältigt oder wie sicher oder unsicher er dabei ist, so muss zunächst ein Vergleichswert geschaffen werden, der zeigt wie Experten die Aufgabenstellung meistern, welche Strecke sie mit den Mausbewegungen zurücklegen. Dies kann nur schwer auf unterschiedliche Aufgaben übertragen oder umgerechnet werden.

4 Anwendungsszenario – Modellanalyse-Auswertung

Nachfolgend soll ein Anwendungsszenario aus der Kieferorthopädie näher betrachtet werden. Es geht dabei um die Durchführung einer digitalen Modellanalyse in einer dafür vorgesehenen Software. Die Modellanalyse wird dazu eingesetzt den Kiefer und die Zähne eines Patienten zu vermessen, um so

Zahnfehlstellungen zu beurteilen. Darauf basierend kann die kieferorthopädische Behandlung geplant werden [28]. Die Durchführung der digitalen Modellanalyse ist sehr nah am Klinikalltag. Auch hier werden die Modelle bereits in digitaler Form analysiert und vermessen. Dazu wird ein Intraoralscan angefertigt und das Modell in eine dafür vorgesehene Software geladen. In Abbildung 5 ist die Software OnyxCeph zu sehen. Der Nutzer erhält das Modell in verschiedenen Ansichten, während diese durch Drag and Drop drehbar sind.

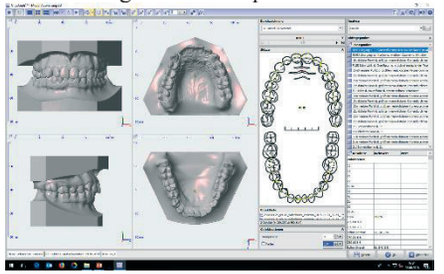


Abbildung 5: Screenshot aus der Software OnyxCeph³ zur Auswertung einer Modellanalyse

Zur Durchführung der Modellanalyse werden dann bestimmte Punkte im Modell gesetzt, die vom Programm vorgegeben werden. Das Programm führt basierend darauf die entsprechenden Abmessungen durch. Es ist wichtig die Punkte sehr genau an den richtigen Stellen zu positionieren, um eine genaue Abmessung zu realisieren. Deshalb wird das Modell häufiger gedreht, um den bestmöglichen Winkel auf den fokussierten Bereich zu erlangen. Wird diese digitale Modellanalyse als eLearning-Anwendung verwendet und in der Lehre eingesetzt, so ist es sinnvoll, eine automatische Beurteilung über den Lernfortschritt zu erschaffen. Neben der Beurteilung wie nah die Punkte im Vergleich zur Musterlösung eines Experten gesetzt wurden, kann gerade in diesem Anwendungsszenario das Mouse-Tracking weitere Informationen über den Nutzer und seinen Lernfortschritt liefern. Zudem wird die Modellanalyse an einem PC mit einer Maus durchgeführt, sodass das Mouse-Tracking ohne großen Mehraufwand angebunden werden kann.

Es wurden Hypothesen entwickelt, die beschreiben, wie sich Experten im Vergleich zu Anfängern in Bezug auf die digitale Modellanalyse verhalten. Die Hypothese lautet, dass Experten weniger Mausbewegungen durchführen als Anfänger bei einer digitalen Modellanalyse. Experten drehen das Modell zudem zielgerichteter und setzen ihre Punkte zielgerichteter, als Anfänger, die sich zum einen unsicher sind, wie sie das Modell drehen müssen, um die bestmögliche Sicht auf die gefragten Punkte zu erhalten und zum anderen unsicher sind, wo genau sie den Punkt zu setzen haben. Diese Hypothesen wurden aus den zuvor beschriebenen Studien abgeleitet. Immer dann, wenn Unsicherheit bei den Probanden auftrat, weil es beispielsweise um eine atypische Fragestellung ging, konnte das in den Mausbewegungen gezeigt werden. Werden die Strecken, die dabei durch die Mausbewegung zurückgelegt wurden, aufsummiert, so wird schnell ersichtlich, dass bei den Studien in denen die Probanden bei Unsicherheit die Maus nicht zielgerichtet auf die richtige Antwort zubewegt haben, auch gleichzeitig weitere Strecken durch die Mausbewegung zurückgelegt haben. Diese Erkenntnis soll deshalb verallgemeinert auf die Modellanalyse-Auswertung übertragen werden und schlussfolgert in der Hypothese, dass Anfänger mehr Mausbewegungen durchführen, als Experten. Drag and Drop Ereignisse, die genutzt werden, um das Modell zu drehen, werden dabei ebenfalls als Mausbewegungen betrachtet.

Durch diese zusätzliche Auswertemethode wäre es dann möglich, den Nutzern ein differenzierteres Feedback zu geben. Es wäre beispielsweise erkennbar, ob jemand zwar schon sehr gute Ergebnisse beim Setzen der Punkte erzielt, aber dennoch Unsicherheit ausstrahlt und könnte diesem Nutzer vorschlagen weitere Übungen durchzuführen, um seine Unsicherheit zu eliminieren.

5 Fazit

Mit dieser Arbeit konnte gezeigt werden, dass Mouse-Tracking bereits in verschiedenen Gebieten zum Einsatz kommt und dabei hilft, den Nutzer besser zu verstehen. Gerade für die Forschung bringt Mouse-Tracking im Vergleich zum Eye-tracking bislang noch den entscheidenden Vorteil mit, dass durch die nicht reaktive Untersuchungsmethode, die Nutzer sich unbeobachtet fühlen und in einer natürlichen Umgebung beobachtet werden können. Zudem bietet sich Mouse-Tracking vor allem deshalb an, da es keiner zusätzlichen Hardware bedarf und in ein bestehendes System integriert werden kann. Aus den psychologischen Studien konnte abgeleitet werden, dass Mouse-Tracking Informationen liefert, um Denkprozesse nachzuvollziehen. So wurde insbesondere gezeigt, dass bei Entscheidungsfindungen, bei denen eine Unsicherheit beim Probanden besteht, diese durch die Mausbewegungen abgebildet werden. Dabei gingen die Mausbewegungen zunächst in eine allgemein gültige Richtung und erst während der Bewegung fällt die finale Entscheidung und die Maus wird zur entsprechenden Antwort hinbewegt. Aus diesem Phänomen wurde in dieser Arbeit abgeleitet, dass die Nutzer, die sich in ihren Entscheidungen sicher sind, weniger Mausbewegungen durchführen, da sie zielgerichtet die Maus auf die Antwort zubewegen. Diese Erkenntnis wurde in einem Anwendungsszenario auf die Auswertung der Modellanalyse in der Kieferorthopädie übertragen. Hierbei führen diejenigen, die bereits Experten in der Modellanalyse sind, weniger Mausbewegungen durch, als diejenigen, die noch dabei sind die Modellanalyse zu erlernen. Die Software zur Auswertung stellt eine Art Simulationssoftware dar, bei der die Nutzer ein Modell des Kiefers erhalten. In diesem Modell müssen bestimmte Punkte gesetzt werden, um Abmessungen zu erhalten. Diejenigen, die in der Modellanalyse geübt sind, wissen potentiell, wann sie wo etwas zu klicken haben und wie sie das Modell drehen und positionieren müssen, um für die entsprechenden Punkte die bestmögliche Sicht zu haben. So wurde die Hypothese abgeleitet, dass sich die aus der Stand der Wissenschaft bereits gewonnen Erkenntnisse auf dieses Anwendungsszenario übertragen ließen.

6 Ausblick

Um die gewonnenen Erkenntnisse und Hypothesen für das Anwendungsszenario zu verifizieren, werden in einer Studie die Mausbewegungen von Experten und Anfängern bei der Modellanalyse verglichen. Dabei können die Mausbewegungen und Drag & Drop-Bewegungen als Strecken aufgezeichnet werden und miteinander verglichen werden. So kann die Hypothese, dass Anfänger mehr Mausbewegungen bei der digitalen Modellanalyse durchführen, als Experten evaluiert werden. Diese Studie wird in einer weiterführenden Arbeit umgesetzt.

7 Literatur

[1] F. Lateef, „Simulation-based learning: Just like the real thing“ (eng), *Journal of emergencies, trauma, and shock*, Jg. 3, Nr. 4, S. 348–352, 2010.

[2] J. B. Freeman, R. Dale und T. A. Farmer, „Hand in motion reveals mind in motion“ (eng), *Frontiers in psychology*, Jg. 2, S. 59, 2011.

[3] R. Dale, C. Kehoe und M. J. Spivey, „Graded motor responses in the time course of categorizing atypical exemplars“, *Memory & Cognition*, Jg. 35, Nr. 1, S. 15–28, 2007.

[4] J. B. Freeman und N. Ambady, „MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method“ (eng), *Behavior research methods*, Jg. 42, Nr. 1, S. 226–241, 2010.

[5] T. A. Farmer, S. A. Cargill, N. C. Hindy, R. Dale und M. J. Spivey, „Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing“ (eng), *Cognitive science*, Jg. 31, Nr. 5, S. 889–909, 2007.

[6] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott und A. M. Wichansky, „Eye tracking in web search

tasks“ in *Proceedings of the 2002 symposium on Eye tracking research & applications*, New Orleans, Louisiana, 2002, S. 51.
[7] S. Incera und C. T. McLennan, „Mouse tracking reveals that bilinguals behave like experts“, *Bilingualism*, Jg. 19, Nr. 03, S. 610–620, 2016.
[8] M. J. Spivey, M. Grosjean und G. Knoblich, „Continuous attraction toward phonological competitors“ (eng), *Proceedings of the National Academy of Sciences of the United States of America*, Jg. 102, Nr. 29, S. 10393–10398, 2005.
[9] I. Hsieh-Yee, „Research on Web search behavior“, *Library & Information Science Research*, Jg. 23, Nr. 2, S. 167–185, 2001.
[10] S. Broschart, *Der Content Faktor: Schreiben Sie Texte, die gefunden und gelesen werden*. Haar: Franzis Verlag, 2017.
[11] M. C. Chen, J. R. Anderson und M. H. Sohn, „What can a mouse cursor tell us more?“ in *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, Seattle, Washington, 2001, S. 281.
[12] F. Mueller und A. Lockerd, „Cheese“ in *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, Seattle, Washington, 2001, S. 279.
[13] B. Pan *et al.,* „The determinants of web page viewing behavior“ in *Proceedings of the 2004 symposium on Eye tracking research & applications*, San Antonio, Texas, 2004, S. 147–154.
[14] E. Arroyo, T. Selker und W. Wei, „Usability tool for analysis of web designs using mouse tracks“ in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, Montreal, Quebec, Canada, 2006, S. 484.
[15] M. Händler, „Wie klicken frustrierte Nutzer?“, http://dl.gi.de/bitstream/20.500.12116/5873/2/H%c3%a4ndler_2011.pdf, 2011.
[16] V. Navalpakkam und E. Churchill, „Mouse tracking“ in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, Austin, Texas, USA, 2012, S. 2963.
[17] R. A. Abrams und D. A. Balota, „Mental Chronometry: Beyond Reaction Time“, *Psychol Sci*, Jg. 2, Nr. 3, S. 153–157, 2017.
[18] J. I. Gold und M. N. Shadlen, „Neural computations that underlie decisions about sensory stimuli“ (eng), *Trends in cognitive sciences*, Jg. 5, Nr. 1, S. 10–16, 2001.
[19] J.-H. Song und K. Nakayama, „Role of focal attention on latencies and trajectories of visually guided manual pointing“ (eng), *Journal of vision*, Jg. 6, Nr. 9, S. 982–995, 2006.
[20] S. P. Tipper, L. A. Howard und G. Houghton, „Action-based mechanisms of attention“ (eng), *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, Jg. 353, Nr. 1373, S. 1385–1393, 1998.
[21] J. Meyerhoff und C. Brühl, Hg., *Fachwissen lebendig vermitteln: Das Methodenhandbuch für Trainer und Dozenten*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.
[22] A. K. Jha, B. W. Duncan und D. W. Bates, „Chapter 45 . Simulator-Based Training and Patient Safety“, https://pdfs.semanticscholar.org/f738/0d061f48c7aec186bcc487380964688dd7c4.pdf, 2001.
[23] T. P. Grantcharov *et al.,* „Randomized clinical trial of virtual reality simulation for laparoscopic skills training“ (eng), *The British journal of surgery*, Jg. 91, Nr. 2, S. 146–150, 2004.
[24] D. M. Gaba *et al.,* „Assessment of clinical performance during simulated crises using both technical and behavioral ratings“ (eng), *Anesthesiology*, Jg. 89, Nr. 1, S. 8–18, 1998.
[25] F. Lateef, „What's new in emergencies, trauma, and shock? Role of simulation and ultrasound in acute care“ (eng), *Journal of emergencies, trauma, and shock*, Jg. 1, Nr. 1, S. 3–5, 2008.
[26] E. L. Meyen, R. J. Aust, Y. N. Bui und R. Isaacson, „Assessing and Monitoring Student Progress in an E-Learning Personnel Preparation Environment“, *Teacher Education and Special Education*, Jg. 25, Nr. 2, S. 187–198, 2016.
[27] A. Serrano, E. J. Marchiori, A. del Blanco, J. Torrente und B. Fernandez-Manjon, „A framework to improve evaluation in educational games“ in *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)*, Marrakech, Morocco, 2012, S. 1–8.
[28] T. Weber, Hg., *Zahnmedizin*. Stuttgart: Georg Thieme Verlag, 2016.

Modellierung der agilen Softwareprozesse in einem Großprojekt mit Dienstleistern und Analyse der Herausforderungen der agilen Transformation

Katharina Pavić
Hochschule Reutlingen
Katharina.Pavic@Student.Reutlingen-University.DE

Abstract
Die vorliegende Arbeit beschäftigt sich mit der agilen Transformation, wie sie im Umfeld eines Produktdatenmanagement-Großprojektes mit Dienstleistern umgesetzt worden ist und welche Schwierigkeiten, Herausforderungen und Probleme mit ihr einhergehen. Dafür werden zunächst die dem Projekt zugrundeliegenden Frameworks Scrum und Large-Scale-Scrum in der Theorie betrachtet, bevor die tatsächlich gelebte Umsetzung im Projekt beschrieben und analysiert wird. Bei der Analyse werden ausschließlich die Schwierigkeiten und Probleme, nicht jedoch die gelungenen Aspekte betrachtet. Die Schwierigkeiten und Herausforderungen werden in die Kategorien technisch, prozessual und kulturell unterteilt, wobei die kulturellen Schwierigkeiten die größte Herausforderung darstellen.

Schlüsselwörter
Agile Softwareentwicklung, agile Transformation, Herausforderungen, LeSS, PDM

CCS-Concepts
• Software and its engineering ~ Software design engineering • Software and its engineering ~ Agile software development

1 Einleitung
Produktdatenmanagementsysteme (PDM) sind aus der Industrie nicht mehr wegzudenken [1]. Auch in der Automobilindustrie sind sie ein wichtiger Pfeiler, ohne die die Entwicklung von Fahrzeugen in ihrer heutigen Komplexität nicht mehr möglich wäre. Ein Fahrzeug besteht aus einer großen Menge unterschiedlichster Einzelteile und zugehöriger Daten [2]. Jedes dieser Einzelteile hat einen eigenen Lebenszyklus, der persistent - auch weit über die Entwicklungs- und Wartungszeit hinaus - dokumentiert werden muss.

Diese in einem Fahrzeug gegebene Funktionsvielfalt würde ein gewöhnliches PDM-System nicht abbilden können und würde man versuchen ein allumfängliches PDM-System zu entwickeln, wäre es sehr komplex und die Effizienz und User Experience würden darunter leiden. Deshalb entschied die Daimler AG, für verschiedene Komponenten und Datenarten verschiedene PDM-Systeme zu entwickeln und diese in einer zentralen Plattform (Systemverbund) zu integrieren und zugänglich zu machen. Der User hat dadurch das Gefühl, ein einziges System zu bedienen, und muss sich keine Gedanken machen, wo er welche Daten eintragen muss und welches PDM-System er dafür öffnen muss.

Betreuer Hochschule: Prof. Dr. Oliver Burgert
Hochschule Reutlingen
Oliver.Burgert@Reutlingen-University.de
Betreuer Firma: Daniel Diedrich
Daimler AG
daniel.diedrich@daimler.com

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Katharina Pavić

Jeder User bekommt nur Zugriff auf die Systeme oder Applikationen, die er für die Erfüllung seiner Aufgaben in Abhängigkeit von seiner Rolle benötigt und wird nicht mit zusätzlichen Funktionen überhäuft. Durch die Eingliederung in einem zentralen Systemverbund wird auch die Bedienbarkeit und Gestaltung vereinheitlicht, was dem Nutzer die Handhabung und Einarbeitung vereinfachen soll. Aktuell sind im Systemverbund 17 verschiedene Applikationen enthalten. Durch die modulare Architektur ist der Systemverbund problemlos erweiterbar. So ist auch das zentrale PDM-System zur Dokumentation und Pflege von elektrischen und elektronischen Komponenten- und Systemdaten (PEEK) ein Teil des Systemverbundes.

Bisher wurden sowohl der Systemverbund als auch alle enthaltenen Applikationen und Systeme nach dem Wasserfallmodell entwickelt. Im PEEK-Projekt erfolgte vor ca. 1,5 Jahren der Umstieg auf die agile Softwareentwicklung. Um die aktuellen, durch die agile Transformation im PEEK-Projekt hervorgerufenen Schwierigkeiten und Herausforderungen zu identifizieren und in Zukunft priorisieren und beheben zu können, wurde diese Arbeit verfasst. Sie ist wie folgt aufgebaut: Zunächst wird die Theorie des agilen Rahmenwerks, nach dem im PEEK-Projekt gearbeitet wird, betrachtet, danach wird ein Überblick über die organisatorische Struktur im PEEK-Projekt gegeben und anschließend die tatsächlich gelebten Prozesse betrachtet, wofür Interviews mit Mitarbeitern geführt wurden. Zuletzt werden die Probleme und Herausforderungen gesammelt, die mit der agilen Transformation einhergehen.

2 Agile Softwareentwicklung

Agile Softwareentwicklung gewinnt stetig an Popularität [3]. Durch die inkrementelle und iterative Vorgehensweise wird die Entwicklungsdauer verkürzt [3]. Mit der Zeit wurden einige agile Vorgehensmodelle entwickelt – Kanban und Scrum sind die bekanntesten und gebräuchlichsten agilen Rahmenwerke [4]. Allen agilen Methoden ist eines gemein: sie orientieren sich am agilen Manifest [5] und den damit einhergehenden zwölf Prinzipien der agilen Softwareentwicklung.

Da im PEEK-Projekt nach Scrum, bzw. nach Large-Scale Scrum (LeSS) gearbeitet wird, werden diese beiden Frameworks näher betrachtet.

2.1 Scrum

Kenntnisse über die Prinzipien und Regeln von Scrum werden angenommen. Deshalb wird hier nur eine kurze Zusammenfassung gegeben, mit den für diese Arbeit wichtigsten Aspekte in Kombination mit Abbildung 1, in der die Schlüsselrollen, Prozesse, Artefakte und Ereignisse von Scrum abgebildet sind.

Scrum ist ein Framework, in dem ein interdisziplinäres, selbstorganisiertes Team (s. Entwicklungsteam in Abbildung 1) von drei bis neun Personen, ein Produkt inkrementell

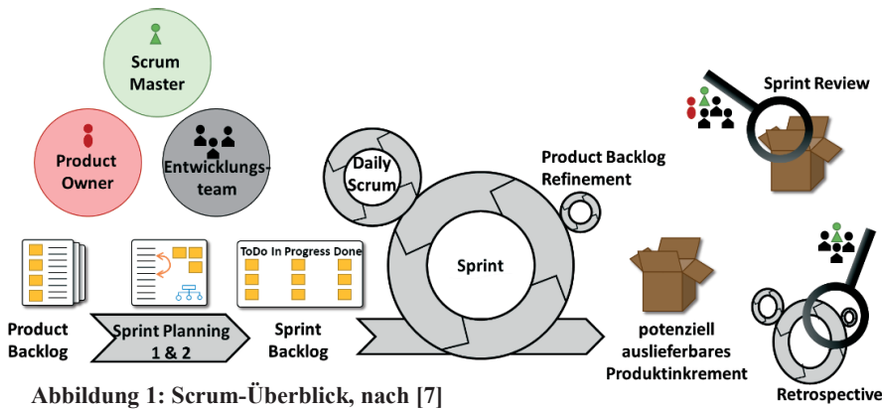


Abbildung 1: Scrum-Überblick, nach [7]

und iterativ entwickelt [6]. Das Resultat jedes Sprints (üblich sind 2-Wochen-Sprints) ist ein potenziell auslieferbares Inkrement [6]. Kontinuierliches Lernen, Prüfen und Anpassen des Produktes und der Prozesse sind essentiell für Scrum [7] und finden sich in dessen Prinzipien wieder [6]. Ein Scrum-Team besteht aus drei Einheiten: dem Product Owner (PO), dem Scrum Master und dem Entwicklungsteam [6]. Der PO ist verantwortlich für die Maximierung des Produktwertes und das Management des Produkt Backlogs [6]. Der Scrum Master hilft dem Entwicklungsteam, dem PO und der Organisation, die Theorie, Regeln und Werte von Scrum zu verstehen und anzuwenden [6]. Das Entwicklungsteam besteht aus Experten aus unterschiedlichen Disziplinen, um am Ende des Sprints das Sprintziel erreichen zu können [6].

2.2 LeSS

2.2.1 Grundsätze

LeSS ist Scrum und kein verbessertes oder erweitertes Scrum. Es geht darum, Scrum so einfach wie möglich in einen umfangreicheren Kontext einzubinden. Transparenz wird durch klare Definition of Done, kurze Zyklen, Zusammenarbeit und keiner Angst vor Fehlern erreicht. Der Grundsatz *More with Less* bedeutet: mehr verantwortungsbewusste Teams, und dabei wenige(r) Rollen. Mehr kundenorientierte Teams und nützliche Produkte, indem es wenige(r) Artefakte gibt. Mehr Eigenverantwortung des Teams für den Prozess und mehr bedeutende Aufgaben, indem es wenige(r) definierte Prozesse gibt. [7]

Bei LeSS wird durch die kontinuierliche Verbesserung Perfektion angestrebt: Nach jedem Sprint entsteht ein kostengünstiges, fehlerfreies Produkt, das Kunden begeistert und die Umwelt und das Leben verbessert. Lean Thinking ist wichtiger Bestandteil von LeSS und kann nur erreicht werden, wenn Manager Lean Thinking anwenden und lehren. Auch der Respekt für Menschen und die Motivation, den Status-Quo in Frage zu stellen und sich kontinuierlich zu verbessern, sind essentiell. Empirische Prozesssteuerung wird durch kontinuierliche Überprüfung und Anpassung von Produkten, Prozessen, Verhaltensweisen und organisatorischen Strukturen sichergestellt. Insbesondere in Großprojekten ist es wichtig, die Warteschlangentheorie zu verstehen. Dieses Verständnis kann auf die Verwaltung von Warteschlangengrößen, work-in-Progress-Grenzen, Multitasking, Arbeitspaketen und Variabilität angewendet werden. [7]

Der Fokus liegt auf dem gesamten Produkt, denn der Kunde möchte ein abgeschlossenes Produkt und keine Einzelteile. Deshalb gilt: *ein* Produkt Backlog, *ein* Product Owner, *ein* auslieferbares Inkrement, *ein* Sprint – egal ob 1 oder 50 Teams beteiligt sind. Ziel ist es, dass jeder das gesamte System (nicht nur Teile) durch Systemmodellierung sieht, versteht und optimiert. LeSS ist kundenzentriert, indem Wert und Verschwendung aus der Sicht des Kunden identifiziert und eliminiert werden und zudem die Wartezeit reduziert wird [7].

2.2.2 LeSS Framework

Es gibt zwei Arten von LeSS: smaller LeSS und LeSS Huge – LeSS Huge ist eine vergrößerte Version von smaller LeSS und das wiederum ist eine vergrößerte Version von Scrum. [7]

smaller LeSS:

Auf Scrum-Team-Ebene ändert sich an der Konstellation und der Verantwortung auf den ersten Blick wenig: Das Entwicklungsteam ist weiterhin selbstorganisiert, interdisziplinär, co-located (alle arbeiten am selben Tisch) und langlebig. Es hält weiterhin seine Daily Scrums ab, hat sein eigenes Sprint Backlog und seine eigene Sprint Retro. [7]

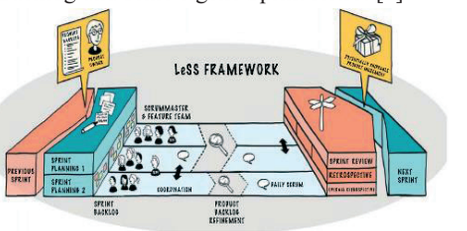


Abbildung 2: smaller LeSS Framework aus [7]

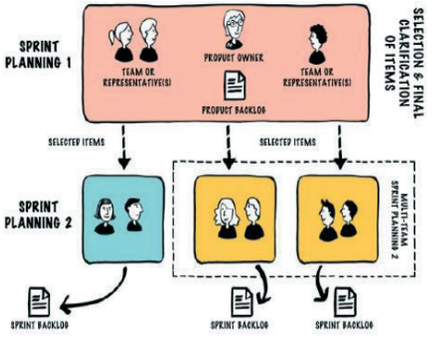


Abbildung 3: Sprint-Planning, aus [7]

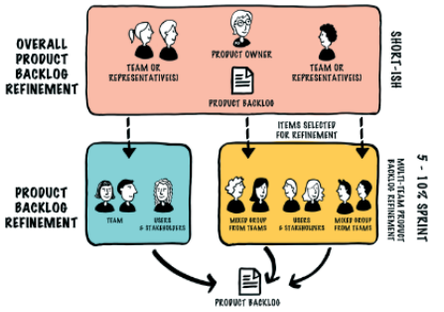


Abbildung 4: Backlog-Refinement, aus [7]

Da es in LeSS mehrere Teams gibt (s. Abbildung 2), nehmen im Sprint Planning 1 nicht mehr alle Teammitglieder eines Scrum Teams teil, sondern lediglich ausgewählte Repräsentanten des Teams, der Scrum Master und der PO. Ebenso wie beim Overall Product Backlog Refinement (OBLR) – es entspricht dem Product Backlog Refinement (PBLR), jedoch sind die Items weniger detailliert. Das PBLR und Sprint Planning 2 finden weiterhin auf Teamebene ohne den PO statt – es sei denn es gibt eng verwandte Themen zu anderen Teams - dann werden PBLR und Sprint Planning 2 mit den entsprechenden Teams gemeinsam durchgeführt. Dennoch hat jedes Team weiterhin ein eigenes Sprint Backlog und kein gemeinsames (Abbildung 3) im Gegensatz zum Product Backlog: hier wird alles, was im OBLR und im PBLR erschaffen wird, in einem gemeinsamen Product Backlog verwaltet (Abbildung 4). [7]

LeSS Huge:

LeSS Huge funktioniert wie smaller LeSS, mit dem Unterschied, dass es mehrere „Areas“ gibt (s. Abbildung 5). Eine Area entspricht Abbildung 2 – also einem Zusammenschluss mehrerer Teams, mit einem gemeinsamen PO in der Verantwortung. Jede Area hat einen eigenen Area Product Owner (APO), der die Rolle eines POs nach smaller LeSS übernimmt und gemeinsam mit dem PO und anderen APOs das Product Owner Team bildet. Ein APO ist für maximal sieben Scrum Teams verantwortlich. Pro Area gibt es ein Area Product Backlog, das in das Product Backlog integriert wird. [7]

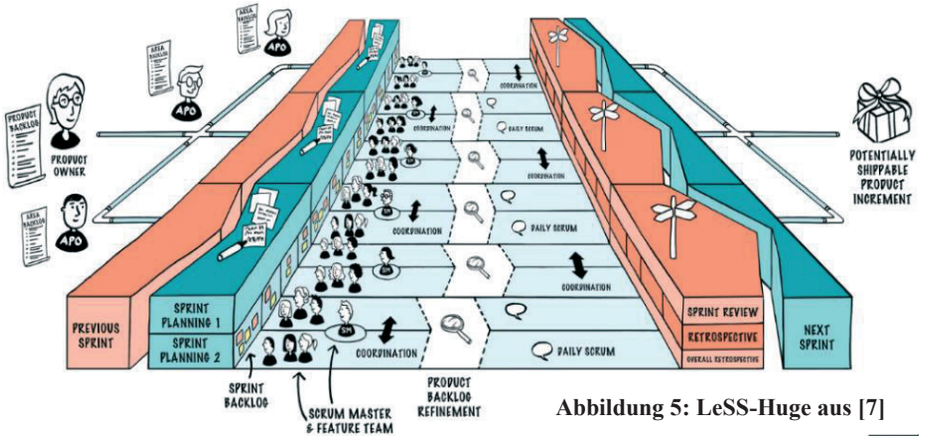


Abbildung 5: LeSS-Huge aus [7]

3 Organisatorische Struktur im PEEK-Umfeld

Um das Verständnis für die im Projekt ablaufenden Prozesse zu vereinfachen, ist es nötig einen groben Überblick über die organisatorische Struktur und einige Begrifflichkeiten zu geben.

3.1 Bereiche

Für PEEK sind drei Bereiche der Daimler AG von Bedeutung, da hier die wichtigsten Schnittstellen bestehen: Research und Development (RD), IT-Datenverarbeitung (IT-DV) und IT-Betrieb (IT-B). Dabei liefert IT-DV Software an RD und IT-B stellt IT-DV, die Infrastruktur zur Verfügung.

RD, auch „Fachbereich“ genannt, ist der Bereich, in dem die Fahrzeuge entwickelt werden – für PEEK relevant sind die elektronischen und elektrischen Komponenten, die im Fahrzeug verbaut werden. RD ist Kunde vom PDM-Systemverbund. Jedoch benötigt nicht jeder Kunde aus RD Zugriff auf alle Applikationen des PDM-Systemverbundes. Abhängig von seinen Aufgaben hat ein RD-Mitarbeiter Zugriff auf eine oder mehrere Applikationen.

IT-DV ist der Bereich, in dem die großen und strategisch wichtigen Applikationen des PDM-Systemverbundes gewartet und weiterentwickelt werden. Wie zuvor erwähnt, sind im Systemverbund aktuell 17 Applikationen verfügbar. Jede Applikation wird von einem oder mehreren Teams gewartet und weiterentwickelt - so auch PEEK.

Die IT-B kümmert sich um die Betriebssteuerung des Rechenzentrums und die Infrastruktur. Für das PEEK-Projekt ist von Bedeutung, dass IT-B den Server mit der Produktiv-Umgebung (PROD) zur Verfügung stellt und wartet.

3.2 Teams

Innerhalb von IT-DV sind drei Teams an der Entwicklung von PEEK involviert: Systementwicklung (SE), Support und operative Prozessverantwortung (SOP) und konzeptionelle Prozessverantwortung (KPV). Während alle Mitglieder des Teams SE ausschließlich für PEEK zuständig sind, sind aus den Teams SOP und KPV nur wenige dem PEEK-Projekt zugeteilt – manche von ihnen arbeiten parallel auch an anderen Projekten. Insgesamt sind ca. 20 Personen aus IT-DV für die Entwicklung von PEEK verantwortlich. Hierbei ist die Aufgabenverteilung wie folgt:

SOP hat direkten Kundenkontakt, sammelt die Anforderungen und kümmert sich um die Erstellung des Handbuchs, Schulungen der User und den Support bei der Bedatung. Neben der Verantwortung für das Test- und Defect-Management, hat KPV die Aufgabe, die Anforderungen, die aus dem Fachbereich kommen, zu analysieren und festzustellen, welche Systeme/Applikationen an den neuen Prozessen oder Prozessänderungen beteiligt sind. Sie sind applikations-, prozessübergreifend und über alle Elektronik und Elektrik- und Software-Dokumentationsprozesse (E/E & SW-Dokuprozess) hinweg zuständig für die grobe Lösungsausarbeitung der Prozesse.

Nur wenn PEEK von einer Prozessänderung/ - Erweiterung oder einem neuen Prozess betroffen ist, kommt SE ins Spiel. SE hat die Aufgabe, aus den von KPV erstellten applikationsübergreifenden Lösungsausarbeitungen, applikationsspezifische Epics und User Stories zu generieren. Projektleitung, (Software-) Architektur, Application Management (AppMan) und Betrieb sind ebenfalls Teil von SE. Aktuell sind sieben Dienstleiter im PEEK-Umfeld beschäftigt und somit sind insgesamt über 100 Personen an der Entwicklung und dem Betrieb von PEEK beteiligt. Ein so großes Projekt sprengt den Rahmen der für ein Scrum-Team üblichen Größe von drei bis neun Mitgliedern [6].

4 Agile Softwareentwicklung in PEEK

Um die Prozesse besser zu verstehen, wird zunächst ein Überblick über den gesamten Prozess ab der Anforderungserstellung (Demand) der Kunden aus dem Fachbereich bis hin zum Go-Live (Delivery) betrachtet.

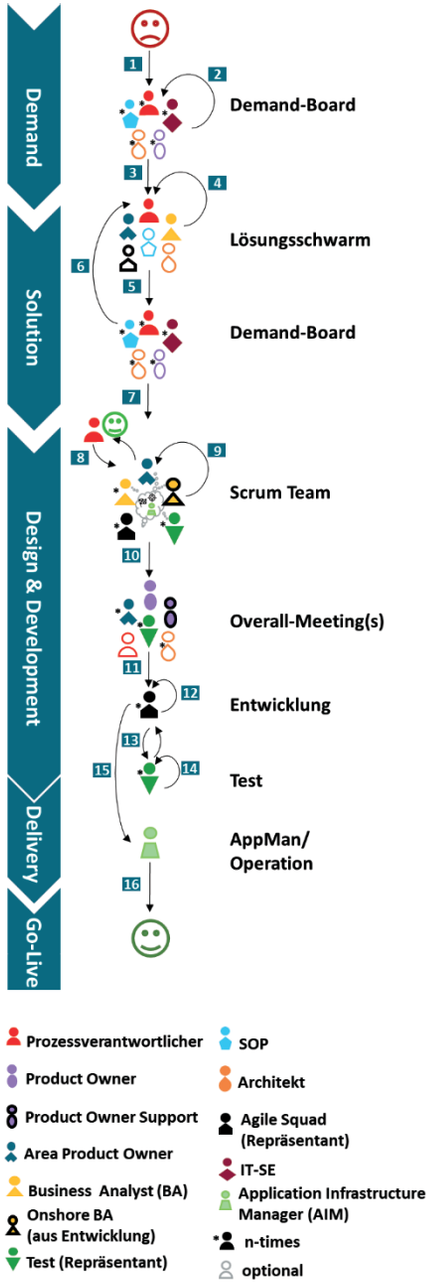


Abbildung 6: High-Level Demand to Delivery -Prozess

4.1 Demand to Delivery

Wie bereits erwähnt, sind sehr viele Personen an der Entwicklung von PEEK beteiligt. Dementsprechend gibt es viele verschiedene Meetings mit unterschiedlichen Aufgaben, Zielen und Konstellationen. In Abbildung 6 ist eine Übersicht über den Gesamtprozess mit den wichtigsten Meetings, Rollen und Konstellationen gegeben. Die Informationen zum Prozess wurden dem internen Wiki, (Confluence [8]) entnommen, andernfalls wurden abweichende Quellen gekennzeichnet. Der Prozess wird nachfolgend beschrieben:

Zu Beginn gibt es einen User im Fachbereich, der eine Anforderung (Demand) hat – in der Regel handelt es sich um Bedarfe aufgrund sich ändernder fachlicher Rahmenbedingungen oder Änderungen an einem bestehenden Prozess [9]. Er meldet den Demand seinem IT-DV-Ansprechpartner (1), üblicherweise einem Kollegen aus dem Team SOP, und beschreibt ihn fachlich gemeinsam mit ihm. Die Demands aus dem Fachbereich werden zentral in Confluence/JIRA gesammelt und regelmäßig vom Fachbereich priorisiert. Sobald der Demand fachlich vollständig beschrieben und priorisiert ist, wird er im Demand-Board vorgestellt (2). Das Demand-Board ist ein Gremium, das alle Bedarfe an E/E & SW-Dokuprozesse [9] sammelt, verwaltet und delegiert. Hier sitzen nicht nur die Verantwortlichen von PEEK, sondern auch die Verantwortlichen aller für den E/E & SW-Dokuprozess relevanten Applikationen.

Nun entscheidet das Demand-Board, welcher Applikation der Demand zugewiesen wird, in diesem Fall PEEK, und benennt einen Prozessverantwortlichen (PV). Der PV wählt, abhängig vom Thema des Demands, Fachexperten aus, mit denen er einen Lösungsschwarm bildet (3). In einem Lösungsschwarm sind üblicherweise ein PV (aus KVP), ein APO (aus SE) und ein Domänenexperte (DE, Dienstleister, der fachlicher Experte für das jeweilige Thema ist) vertreten. Themenabhängig werden auch Architekten, Entwickler oder Supporter benötigt. Der Vorteil dieser Konstellation ist, dass sowohl der DE, als auch der APO später im Scrum-Team die technische Lösung erarbeiten und das Thema aus fachlicher Sicht bereits gut kennen.

Im Lösungsschwarm werden mehrere fachliche Lösungen erarbeitet (4). Die Dauer und Regelmäßigkeit der Meetings des Lösungsschwarms ist abhängig vom Thema und der Komplexität des Demands. Sobald der Lösungsschwarm eine oder mehrere zufriedenstellende Lösungen erarbeitet hat, stellt der PV diese im Demand-Board vor (5). Es kann vorkommen, dass das Demand-Board noch Rückfragen zu der Lösung hat oder diese als unvollständig betrachtet (6) – dann muss der Schwarm die Lösung weiter detaillieren oder andere Lösungen in Betracht ziehen. Empfindet das Demand-Board die Lösung als ausgereift, werden der Demand und die fachliche Lösung im Normalfall direkt im jeweiligen Scrum-Team (ggf. auch im Overall-Meeting) vorgestellt (7). Im Scrum-Team werden Epics und User Stories entwickelt, detailliert und priorisiert und die Sprintinhalte geplant (9). Regelmäßig nehmen PV und der User, der den Demand gestellt hat, an den Reviews teil und prüfen somit den Fortschritt des Demands (8).

Die fertigen Epics und User Stories werden im Overall-Sprint Planning geprüft, das Overall-Backlog optimiert (10), der nächste Sprint final festgelegt und zur Entwicklung freigegeben (11). Die Entwicklungsteams sind nicht nur für die Programmierung der User Stories zuständig (12), sondern auch für Testing, Dokumentation, Testautomatisierung und die Erstellung des Datenbankmodells. Nach jedem Sprint liefert die Entwicklung eine lauffähige und abgesicherte PEEK-Version aus (13), die die letzten zwei bis drei Sprints vor Go-Live einem Verbund- und Fachprozesskettentest unterzogen wird (14). Sind diese erfolgreich, resultiert die Auslieferung an den Betrieb und die Releaseinhalte werden den Kollegen im Support und der Benutzerhilfeerstellung zur Verfügung gestellt (15). Team SOP kümmert sich um die Informationsverteilung beim Kunden und unterstützt den Rollout neuer Funktionalitäten. Zuletzt erfolgt das Go-Live (16) und der User kann die geforderten Funktionen nutzen. Dieser Prozess ist, wie aus der Benennung hervorgeht „High-Level“, weshalb Details über die Zusammensetzung und Arbeit in den Scrum-Teams und dem Overall-Team nicht ersichtlich sind. Auf diese Aspekte soll nachfolgend näher eingegangen werden.

4.2 Scrum & Overall

Die Prozesse im PEEK-Projekt wurden aufgrund der Größe an LeSS angelehnt. Dabei ist der Prozess eine Kombination aus smaller LeSS und LeSS Huge.

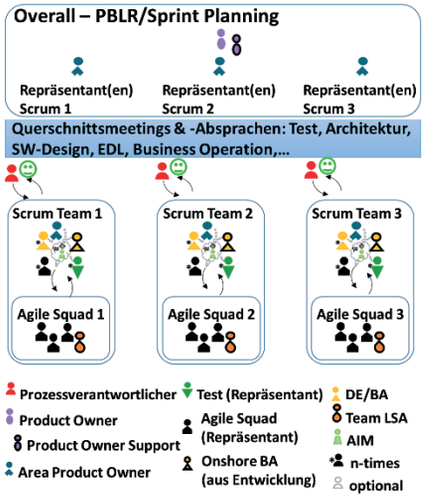


Abbildung 7: Hierarchie: Overall - Scrum Team - Agile Squads

Aktuell sind fünf Scrum-Teams und fünf Agile Squads an der Entwicklung beteiligt. Ein Agile Squad ist jeweils einem Scrum-Team zugeordnet. Die Konstellation und Zusammenarbeit ist schematisch in Abbildung 7 zu sehen. Ein Scrum-Team besteht aus einem APO, einem Onshore Business Analysten (BA) mindestens einem DE, einem Agile Squad-Repräsentanten und mindestens einem lokalen Test- und Defectmanager. Jeder im Scrum-Team hat die Vision, das Big Picture, AppMan und Betrieb bei der Entwicklung der User Stories im Kopf. Der Agile Squad entwickelt die User Stories und garantiert deren Definition of Done. [8]

Der PV und der Demand-Steller aus dem Fachbereich nehmen an den Review-Meetings des Scrum-Teams teil, überprüfen dabei den Fortschritt des Demands und geben Feedback dazu. Über die Scrum-Teams hinweg gibt es bei Bedarf und bei komplexeren Themen Querschnittmeetings und Absprachen zwischen den Fachexperten (Architekten, Domänenexperten, Testern, APOs, …). Das Entwicklungsteam (Agile Squad) sitzt in Indien und einige Mitglieder nehmen remote als Repräsentanten an den Scrum-Meetings teil. Ein Agile Squad besteht aus unterschiedlichen Rollen der Softwareentwicklung (Programmierer, Tester, BA, Lead Solution Architekt (LSA) usw. - die genaue Konstellation variiert von Squad zu Squad, da jeder Squad unterschiedliche Aufgabengebiete hat) und einem Scrum Master. Sie bekommen viel Input vom Onshore BA, der festes Mitglied des Scrum-Teams ist und an der Ausdetaillierung von User Stories beteiligt ist. [8]

Die DEs, lokalen Test- und Defectmanager, das gesamte Agile Squad, der Onshore BA und der Product Owner Support sind externe Dienstleister (EDL), wobei das Agile Squad, der Onshore BA und der PO-Support aus einem einzigen Unternehmen stammt. Der PO-Support ist der Hauptansprechpartner für den PO, wenn er mit dem EDL für die Entwicklung kommuniziert und fungiert intern beim EDL als PO für die Agile Squads. [8]

5 Schwierigkeiten und Herausforderungen

Bisher wurden sowohl der Systemverbund als auch alle enthaltenen Applikationen und Systeme nach dem Wasserfallmodell entwickelt. Es gab zwei Releases im Jahr – eines im Juni und eines im Dezember. Vor ca. zwei Jahren wurde vom Management beschlossen, dass alle IT-Systeme in Zukunft agil entwickelt werden sollen. So startete im PEEK - Projekt vor 1,5 Jahren das erste agile Teilprojekt, das auch erfolgreich umgesetzt wurde. Mit jedem weiteren Release wurden mehr Teilprojekte innerhalb von PEEK in die agile Entwicklung eingegliedert, sodass im letzten Hauptrelease (Juni 2018) das letzte nicht-agile Teilprojekt released und ebenfalls in die agile Entwicklung aufgenommen wurde. Von den ursprünglich zwei Releases jährlich, gibt es im PEEK-Projekt seit diesem Jahr vier reguläre Releases.

Eine agile Transformation geht nicht ohne Probleme einher – diese Erfahrung machte nicht nur das PEEK-Projekt, sondern auch diverse andere Projekte und Unternehmen.

Gandomani et al. [10], die eine große Studie zu den Herausforderungen der agilen Transformation erstellten, aber auch Kalpana et al. [11], die aus eigener Erfahrung über die agile Transformation in ihrem Unternehmen berichten, identifizierten die kulturellen Herausforderungen als die größten. Das rührt daher, dass Gewohnheiten, Denkweisen und Umgangsformern nur mit Zeit, Geduld und der Bereitschaft dazu verändert werden können [10]. Goodmann et al. [12], die ebenfalls in ihrem eigenen Unternehmen die agile Transformation erlebt haben und darüber berichten, sehen als größte Herausforderung das präzise Schätzen der Integrations- und Testphasen. Die Probleme im PEEK-Projekt wurden durch eigene Beobachtungen und Interviews identifiziert und werden unter der Prämisse, dass die Releasefrequenz auf zwölf Releases pro Jahr erhöht wird, betrachtet.

Technisch

Ein zentrales Problem ist, dass die Verbundtestumgebung nicht durchgängig verfügbar ist. Das liegt daran, dass die anderen Systeme/Applikationen aus dem PDM-Systemverbund zum Teil noch nicht auf die agile Entwicklung umgestiegen sind, zum Teil auch daran, dass sie andere Releasezyklen haben als das PEEK-Projekt. Da PEEK einige Schnittstellen zu anderen Systemen hat, ist die durchgängige Verfügbarkeit der Verbundtestumgebung notwendig, um die Releases von PEEK einhalten zu können. Diese Problematik hat fast ein Go-Live von PEEK verhindert. Ein weiteres technisches Problem, das mit dem Testen zusammenhängt, ist die aktuelle Testgeschwindigkeit (2-3 Sprints dauert der Verbundtest). Viele Tests sind bereits automatisiert und die Regressionstests wären potenziell komplett automatisierbar. Die Progressionstests jedoch leider nicht, sodass weiterhin manuelle Tests notwendig bleiben. Aber auch die Deploymentgeschwindigkeit und die damit verbundene Migration der Daten (aktuell 1 – 5 Stunden, abhängig vom Migrationsumfang) ist noch lange nicht optimal für monatliche Releases. Diese Probleme waren auch bei Goodman et al. [12] zu beobachten.

Prozess

Auch im Prozess gibt es einige Hürden zu meistern - bei lediglich fünf Scrum Teams wird in der Literatur smaller LeSS empfohlen und erst ab acht Scrum Teams LeSS Huge [7]. Aktuell wird eine Kombination aus den beiden Frameworks verwendet. Außerdem gibt es keinen ernannten Scrum Master, sondern die APOs übernehmen diese Rolle zum Teil: sie laden zu den Meetings ein, erinnern und beraten die Teammitglieder zu Scrum. Auf der anderen Seite pflegen sie aber auch das Sprint-Backlog und verantworten sich und das Team vor dem PO und Projektleiter. Mit dem PO verhält es sich ebenfalls so: er übernimmt sowohl die Aufgaben des Scrum Masters, als auch die eines POs. Der Scrum-Guide [6], der von den Erfindern von Scrum verfasst wurde, bestätigt, dass die Rollen Scrum Master und PO (APO in LeSS hat ähnliche Funktion) nicht von ein und derselben Person ausgeführt werden dürfen. Da aber ein APO nur ein einziges Scrum Team betreut, was auch den Vorgaben von LeSS widerspricht [7], müssen die genauen Rollenbezeichnungen und –verteilung im PEEK-Projekt überarbeitet werden.

Der Demand-to-Delivery–Prozess erinnert an einen Wasserfallprozess, in dem einige Unterprozesse agil sind. Ein Problem stellt dar, dass aus der Literatur nicht ersichtlich wird, wie systemübergreifende Konzeptarbeiten, Ausarbeitungen von Visionen und Big Pictures im agilen Umfeld gehandhabt werden.

Kultur & Mindset

Das größte Problem ist weder technischer noch prozessualer Natur – es liegt in der Kultur und dem Mindset aller Beteiligten, wie bereits Gandomani et al. [10] und Kalpana et al. [11] feststellten. Ähnliche Probleme konnten auch im PEEK-Umfeld beobachtet werden.

Das PEEK-Projekt gibt es bereits seit ca. 15 Jahren und zahlreiche Mitarbeiter waren bereits zu Beginn des Projektes daran beteiligt. Sie haben all die Jahre nach dem Wasserfallmodell gearbeitet und für sie ist die agile Transformation, die damit einhergehende Aufgaben- und Verantwortlichkeiten-Umverteilung, beängstigend. Sie haben Angst vor dem Macht- und Prestigeverlust, Angst davor nicht mehr benötigt zu werden – zum einen durch die vielen Automatisierungen, die mit der agilen Transformation einhergehen und zum anderen durch das Wissen, das sie nun teilen müssen und sie nicht mehr zur wichtigsten Wissensquelle macht.

Wie bereits unter Prozess erwähnt wurde, ist die klare Rollendefinition und –verteilung noch nicht gelungen. Durch die vielen Beteiligten von der Demanderstellung bis zum Delivery und die wechselnden Verantwortlichkeiten (zuerst ist der Fachbereich verantwortlich, danach der PV, dann das Scrum Team, zuletzt der Betrieb), fühlt sich auch niemand durchgängig verantwortlich.

Eine weitere Schwierigkeit stellt dar, dass in letzter Zeit viele neue Mitarbeiter bei KPV, angefangen haben, die sich nicht nur in PEEK einarbeiten müssen, sondern auch in die anderen Systeme, die am E/E & SW-Dokuprozess beteiligt sind. Der Know-How-Transfer geht schleppend voran, da die erfahreneren Kollegen wenig Zeit haben und einige Vorgänger durch neue Projekte abgezogen wurden und nicht mehr verfügbar sind. Da Einzelpersonen von KPV nicht ausschließlich für das PEEK-Projekt arbeiten, geht auch Effizienz beim Wechsel zwischen den Aufgaben verloren [13].

Die Scrum-Teams und Entwickler des Agile Squads sind laut Scrum [6] zu weit vom User

entfernt – in Scrum wird der direkte Austausch zwischen dem Entwicklungsteam und den Stakeholdern bevorzugt. Der Grund hierfür ist, dass bei jeder Schnittstelle Know-How verloren geht und beim Entwicklungsteam nur ein Bruchteil davon ankommt, was der User bzw. Stakeholder tatsächlich erwartet.

Die Organisationsstruktur mit drei verschiedenen intern verantwortlichen Teams, ist ungünstig gewählt. Bei Scrum geht es darum, dass das Scrum Team die Aufgaben vorantreibt, mit dem User interagiert, selbstverantwortlich arbeitet [6] und was noch wichtiger ist: sich fühlt wie ein Team. Die noch immer bestehende Organisationsstruktur innerhalb von PEEK zeigt auch, dass die agilen Werte, die Kultur und das Mindset noch nicht vollständig verstanden und verinnerlicht wurden.

6 Ausblick

Das PEEK-Projekt ist mit der agilen Transformation auf einem guten Weg und vieles wurde schon sehr gut umgesetzt. Dennoch gibt es einige Herausforderungen, die überwunden werden müssen, um die Qualität und Effizienz der Zusammenarbeit zu verbessern. Lösungen und Konzepte für ausgewählte Probleme werden in einer weiterführenden Arbeit erstellt. Diese werden nach Möglichkeit in den Arbeitsalltag eingeführt und sind damit Bestandteil des kontinuierlichen Verbesserungsprozesses. Jedoch muss bedacht werden, dass Veränderungen, die Kultur und Mindset betreffen, Zeit, Geduld und Resilienz auf Seiten der agilen Verfechter fordern.

7 Literaturverzeichnis

[1] J. Stark, Product Lifecycle Management, Kapitel 1.7, Springer, Chem, 2015.

[2] Y. Zheng , „A Knowledge Based Production Line Development Management System," in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China, 2007.

[3] R. M. Chopade und N. S. Dhavase, „Agile Software Development: Positive and Negative User Stories," Mumbai, India, 2017.

[4] I. Hanschke, Agile in der Unternehmenspraxis - Fallstricke erkennen und vermeiden, Potenziale heben, Wiesbaden: Springer Vieweg, 2017.

[5] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland und D. Thomas, „Manifest für Agile Softwareentwicklung," 2001. [Online]. Available: http://agilemanifesto.org/iso/de/manifesto.html. [Zugriff am 12 09 2018].

[6] K. Schwaber und J. Sutherland, „Der Scrum Guide - Der gültige Leitfaden für Scrum: Die Spielregeln, Deutsche Ausgabe," 2017. [Online]. Available: https://www.scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-German.pdf. [Zugriff am 30 09 2018].

[7] C. Larman und B. Vodde, Large-Scale Scrum: More with LeSS, Crawfordsville, Indiana, USA: Addison-Wesley, 2017.

[8] Daimler AG, Team SE, „Collaboration and Process, Confluence (internes Wiki)," 24 09 2018. [Zugriff am 27 09 2018].

[9] Daimer AG, Demand-Board, „Bedarfsworkflow, Confluence (internes Wiki)," 26 07 2018. [Zugriff am 27 09 2018].

[10] T. J. Gandomani, H. Zulzalil und M. Z. Nafchi, „Agile transformation: What is it about?," in *2014 8th. Malaysian Software Engineering Conference (MySEC)*, Langkawi, Malaysia, 2014.

[11] S. Kalpana und S. Jagadish, „Moving from Waterfall to Agile," in *Agile 2008 Conference*, Toronto, ON, Canada, 2008.

[12] D. Goodman und M. Elbaz, „"It's Not the Pants, it's the People in the Pants" Learnings from the Gap Agile Transformation What Worked, How We Did it, and What Still Puzzles Us," in *Agile 2008 Conference*, Toronto, ON, Canada, 2008.

[13] K. I.-J. Ho und J. Sum , „Scheduling jobs with multitasking and asymmetric switching costs," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, AB, Canada, 2017.

Aktuelle Methoden zur Segmentierung von Laparoskopieinstrumenten

Annika Dausend
Hochschule Reutlingen
annika.dausend@student.reutlingen-university.de

Abstract

Bei laparoskopischen Operationen, bei denen ein Endoskopieroboter zum Einsatz kommt, stellt sich die Frage der Steuerung dieses Geräts. Optimal wäre eine automatische Ausrichtung des Endoskops auf die vom Chirurgen verwendeten Operationsinstrumente. Dazu müssen diese zunächst auf den Bildern aus dem Bauchraum erkannt werden. Ziel ist es durch die automatische Robotersteuerung den Assistenten oder den Operateur von einer manuellen Bedienung zu befreien, sodass sie sich auf die Operation konzentrieren können. Die folgende Arbeit stellt verschiedene aktuelle Methoden zur Segmentierung der Laparoskopieinstrumente vor. Dabei wird abgegrenzt zwischen herkömmlichen Bildverarbeitungsmethoden und Methoden des Deep Learnings. Hinzu kommt die Erkennung der aktuellen Operationsphase als zusätzlicher Output neuronaler Netze.

CCS Concepts

• **Computing methodologies** → *Neural networks*; • **Applied computing** → *Health informatics*;

Keywords

Deep Learning, semantische Segmentierung, Laparoskopie

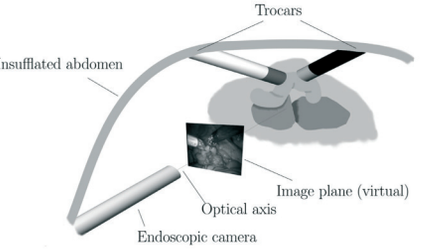


Abbildung 1: Schematische Darstellung einer chirurgischen Laparoskopie [3]

1 Einleitung

Bei einer laparoskopischen Operation handelt es sich um einen minimalinvasiven Eingriff im Bauchraum. Dafür wird dieser zunächst mit einem Gas gefüllt, sodass sich der Operationsraum ausdehnt. Über die kleinen Einschnitte in den Bauch werden Trokare platziert, die die Schnitte abschließen. Dabei handelt es sich um kleine Röhrchen, durch die ein Endoskop und die benötigten Operationsinstrumente eingeführt werden können. In Abbildung 1 ist dieser Aufbau zu sehen.

Ein solcher Eingriff bietet gegenüber einer herkömmlichen offenen Operation einige Vorteile. So ist das Trauma wesentlich geringer aufgrund der kleineren Einschnitte, was auch die Blutungen während der Operation verringert [1]. Des Weiteren werden das Infektionsrisiko, die Schmerzen nach der Operation und die Dauer des Krankenhausaufenthalts reduziert [1, 2].

Betreuer Hochschule: Prof. Dr.-Ing. Cristóbal Curio, Hochschule Reutlingen, cristobal.curio@reutlingen-university.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen

Copyright 2018 Annika Dausend

Allerdings sind diese minimalinvasiven Eingriffe deutlich komplexer, aufgrund der schwierigeren Bedienung der Instrumente und einer eingeschränkten Sicht [1–3]. Durch die Einführung der Instrumente durch den Trokar ist deren Bewegungsfreiheit stark limitiert und es kommt zu einer Umkehr der Bewegungsrichtungen zwischen außerhalb und innerhalb des Bauchraums. Des Weiteren spielt der Faktor Mensch eine wesentliche Rolle. Ein Assistent ist für die Bewegung des Endoskops und damit für die Sicht des Operateurs verantwortlich. Im besten Fall weiß er schon selbst, wohin er die Kamera bewegen muss, damit der Chirurg optimal sieht. Häufig kommt es allerdings zu Kommunikationsproblemen zwischen beiden, wenn der Operateur dem Assistenten Anweisungen zur Positionierung geben muss. Darüber hinaus ermüdet der Assistent im Verlauf der Operation, was zu verwackelten Bildern führen kann.

Als Lösung für diese Probleme werden Laparoskopieroboter eingesetzt. Diese sind entweder nur als Endoskophalter tätig oder führen auch die chirurgischen Instrumente. Je nach Hersteller werden diese über Joystick, Sprache oder mit weiteren Möglichkeiten gesteuert. Im Bereich der Bewegung des Endoskops wäre auch eine automatische Verfolgung der Instrumente möglich. Dadurch würde sich der Operateur sparen, dieses bewegen zu müssen. Um den Instrumenten folgen zu können, müssen, bei reinen Endoskopierobotern, diese auf den Endoskopbildern detektiert werden. Dafür gibt es verschiedene Ansätze von klassischen Bildverarbeitungsmethoden, die die Instrumente z.B. anhand ihrer metallischen Farbe und Form erkennen, bis zu komplexen neuronalen Netzen.

Letztere Variante kann nicht nur für die automatische Steuerung des Endoskops, sondern auch z.B. für die Phasenerkennung während der Operation verwendet

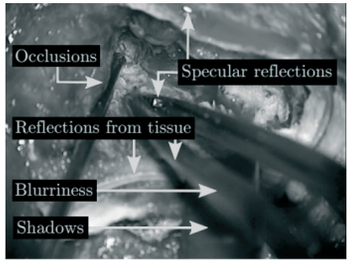


Abbildung 2: Artefakte von Laparoskopiebildern: Verdeckung, Schatten, Reflektion, Unschärfe [4]

werden [5, 6]. Dadurch sind Alarmsysteme, Hilfestellungen und eine Indexierung der Operationsvideos im Nachhinein möglich. Durch die Erkennung der Operationsphase kann die Echtzeitoperationsplanung weiter optimiert werden, was der Ablaufplanung im Krankenhaus zu gute kommt [4]. Gleichzeitig können mit diesem Verfahren die aufgezeichneten Videos im Nachhinein automatisiert mit Informationen beschriftet werden. Dadurch können z.B. Studenten anhand der Videos die Operationsabläufe lernen und zur gleichen Zeit sind sie für die Archivierung aufbereitet.

Bei der Bildverarbeitung sowohl zur Erkennung der Instrumente als auch der aktuellen Phase können auf den Laparoskopiebildern unterschiedliche Artefakten auftreten. Diese sind in Abbildung 2 beispielhaft dargestellt. Vorwiegend handelt es sich um Verdeckungen, Schatten, Reflexionen und Unschärfe. Zu den Ursachen von Artefakten zählen zum Beispiel Blutungen oder Rauch. Als Folge ist die Segmentierung eines Instruments stark beeinflusst und eine herausfordernde Aufgabe. [4]

Neben der minimalinvasiven Operation im Bauchraum und unterhalb der Schädeldecke (Neuroendoskopie), gibt es auch solche Operationen am Auge. Dabei wird mit einem

Mikroskop das Auge als ganzes betrachtet und durch kleine Einschnitte die Instrumente und eine Lichtquelle eingebracht [3].

In dieser Arbeit wird zunächst in Kapitel 2 der State-of-the-Art der herkömmlichen Segmentierung von Instrumenten anhand verschiedener Parameter betrachtet. Anschließend erfolgt eine Vorstellung unterschiedlicher Ansätze im Bereich des maschinellen Lernens zur Segmentierung der Instrumente, ihrer einzelnen Teile und ihres Typs im Abschnitt 3. Im Anschluss erfolgt ein kurzer Einblick in den Bereich der Erkennung von Operationsphasen und der Verwendung von Simulationsdaten als Trainingsgrundlage. Ziel ist es die verschiedensten aktuellen Forschungsansätze zusammenzustellen, um einen guten Überblick zu gewährleisten. Nach Kenntnis der Autorin gibt es bislang keine solche Übersicht, die auch Deep Learning Methoden berücksichtigt.

2 Konventionelle Methoden zur Detektion

Bereits in den 1990er Jahren kam die Fragestellung nach einer automatischen Steuerung von Laparoskopierobotern auf [7]. Mithilfe unterschiedlicher Verfahren wurde damals versucht, die Instrumente auf den Endoskopbildern zu detektieren. Grundsätzlich kann man hierbei zwei verschiedene Varianten unterscheiden. Ein Teil der Forschergruppen verwendet künstliche Marker auf den Instrumenten, sodass sich diese stärker vom Hintergrund unterscheiden. Der andere Teil versucht die Position der Operationswerkzeuge über ihre Form- und Farbeigenschaften herauszufinden. Im Folgenden werden einige Ansätze aus beiden Bereichen näher erläutert.

2.1 Detektion anhand von Markern

Einen ersten Ansatz für die Segmentierung anhand von künstlich auf den chirurgischen Instrumenten angebrachten Markern erarbeiteten Wei et al. [7]. Die Farbe der Marker unterschied sich dabei von denen im Bauchraum vorkommenden Farben, sodass diese über Bildsegmentierung zu erkennen waren. An dieser Stelle ist es besonders wichtig auf die Biokompatibilität der Marker zu achten. Zur Implementierung wurde ein auf dem Markt verfügbarer Laparoskopieroboter (AESOP 1000) verwendet. [7]

Song et al. [2] verwenden in ihrer Forschungsarbeit blaue und grüne Marker auf den Instrumenten. Auch hier sollte durch die gewählten Farben ein möglichst großer Unterschied zu den im Bauchraum vorhandenen Farben gewährleistet sein. Zunächst wird dafür der Marker auf dem Bild segmentiert. Dazu wird das RGB-Bild in den HSV-Raum transformiert und die Farbe von Interesse extrahiert. So entsteht ein Binärbild, das zwischen Marker und Rest unterscheidet. Anschließend kann die Position der Instrumente, definiert als Mittelpunkt des Markers, in Bildkoordinaten bestimmt werden. Eine horizontale bzw. vertikale Bewegung der Kamera erfolgt immer dann, wenn der Marker eine definierte Pufferzone verlässt. Diese ist ebenfalls in Bildkoordinaten angegeben. Bei der verwendeten monokularen Kamera kann die verbleibende Tiefeninformation nur durch Abstandsberechnungen zwischen den Instrumenten auf dem Bild und anschließender Interpretation dieser erfolgen. Allerdings gibt es dadurch einige Situationen, in denen die Abstandsinformationen fehlerhaft sind. Um dies zu korrigieren, haben Song et al. einen weiteren Marker an jedes Instrument angebracht, der immer im Bild bleiben muss. [2]

In Kombination mit Informationen zu der 3D-Position der Trokare, die auf die Bildebene projiziert wird, und künstlichen Markern, bestimmt Zhao [8] die genaue Position der Instrumente im 3D-Raum. Die Marker sind hier in Form von Streifen am Instrument

angebracht. Versuche zeigten, dass diese Methode schnell, robust und präzise ist. Dieser softwarebasierte Ansatz ermöglicht, dass Instrumente und der medizinische Workflow nicht verändert werden müssen. Allerdings gibt es Verbesserungsbedarf im Bereich der Algorithmen zur Farbsegmentierung und Tiefenevaluation. Vorgeschlagen wird hier ein Algorithmus mit adaptiven Grenzwerten zur Farbsegmentierung. [8]

2.2 Detektion ohne Marker

Um das Problem mit der Suche nach biokompatiblen Markern zu umgehen, gibt es Ansätze zur Segmentierung der Laparoskopieinstrumente nur mithilfe von Bildverarbeitung. Dies kann zum Beispiel anhand der metallischen Farbe der Instrumente oder durch ihre Formeigenschaften geschehen.

Im Gegensatz zu [8] verwendeten Voros et al. [9] die Formeigenschaften der Instrumente in Kombination mit Informationen über die Position des Eintrittpunktes. Ihre Ergebnisse zeigten, dass in nahezu Echtzeit die Instrumente erkannt werden können. Die Methode ist folglich schnell und zusätzlich auch robust gegenüber Überlagerungen und Rauch. Die Position der Eintrittspunkte wird zu Beginn der Operation im Roboterkoordinatensystem ermittelt. Dies geschieht während die Instrumente unter Sichtkontrolle eingeführt werden. Um das Instrument zu detektieren wird zuerst nach Kanten mithilfe der Gradientenmethode gesucht. Sind die Gradienten für die drei Farbebenen maximal, befindet sich dort eine Kante. An dieser Stelle werden die Normale und die Orientierung des Kantenpunktes gespeichert. Anschließend werden einzelne Kantenpunkte herausgefiltert, die nicht zu dem Instrument gehören können und deren Normalwert einen Grenzwert nicht übersteigt. Die in diesem ersten Schritt übriggebliebenen Punkte, werden nun für die Berechnung der Achse des Instruments mithilfe der Hough-Methode verwendet. Ist die Achse gefunden, wird auch die Instrumentenspitze gesucht, da diese wichtig für das Tracking ist. Durch das Wissen über die genaue Position der Instrumentenspitze kann diese immer im Zentrum des Bildes gehalten werden oder es können Strategien für das Tracking mehrerer Instrumente entwickelt werden. Die Punkte auf der Achse werden mithilfe des Otsu-Grenzwertes in zwei Klassen (Hintergrund, Instrument) aufgeteilt, von denen jeweils die längsten zusammenhängenden Bereiche ermittelt werden. Daraus wird dann im ersten Bild auf die Position der Instrumentenspitze geschlossen. Sollte das Ergebnis nicht mit der Wirklichkeit übereinstimmen, kann der Operateur an dieser Stelle eingreifen. Im vierten Schritt wird dann aus den möglichen Kantenpunkten die Kante des Instruments ermittelt. Dazu werden sie in zwei Gruppen (oberhalb und unterhalb der Symmetrieachse) aufgeteilt. Eine der beiden Gruppen wird über die Achse gespiegelt, sodass mehr Punkte vorhanden sind. Die Gerade, die am Besten auf die Punkte passt, ist dann die Kante des Instruments. [9]

Ryu et al. [10] verwenden die metallische Farbe der Instrumente, um diese von den Organen zu unterscheiden. Dazu wird der LAB-Farbraum verwendet. Mithilfe von k-means Clustering wird das Bild in Organ und Instrument aufgeteilt. Anschließend wird eine Reihe von Bildverarbeitungsmethoden angewandt, die das Rauschen verringern sollen und schlussendlich die Kanten und das geschlossene Objekt segmentieren. Des weiteren wurde ein Trackingsystem implementiert, das mithilfe von Ähnlichkeitsmessungen und adaptiven Filtern die Instrumente verfolgt. Zusätzlich gibt es eine Warnung vor möglichen Zusammenstößen an den Instrumentenspitzen, wenn diese einander zu nahe kommen. Der Abstand zwischen zwei Instrumenten wird über die euklidische Abstandsformel berechnet. Dieses Warnsystem ist aufgrund der vorhandenen Kanteninformationen aus dem Segmentierungsschritt möglich. Damit ist das Ziel der Verbesserung der Sicherheit gewährleistet, indem unbeabsichtigte Instrumentenkollisionen oder Gewebsverletzungen verhindert werden. [10]

Als Erweiterung der in diesem Abschnitt gezeigten Forschungsansätze ist die Arbeit von Bouget et al. [3] zu sehen. In ihrer Ausarbeitung „Vision-based and marker-less surgical tool detection and tracking: a review of the literature“ stellen sie verschiedenste Ansätze für die Detektion von Instrumenten ohne zur Hilfenahme von Markern zusammen. Dabei werden sowohl Anwendungen im Bereich Laparoskopie als auch in der retinalen Mikrochirurgie und die dazugehörigen Datensets betrachtet. [3]

3 Deep Learning Ansätze

Bei Deep Learning handelt es sich um eine Form des maschinellen Lernens, bei denen besonders tiefe neuronale Netze zum Einsatz kommen. Ziel ist es, dass ein Netz, nachdem es anhand von Trainingsdaten gelernt hat, selbstständig z.B. zwischen Hintergrund und chirurgischem Instrument unterscheiden kann. Im Folgenden werden die von verschiedenen Forschergruppen beschriebenen Methoden zur semantischen Segmentierung von chirurgischen Instrumenten mittels Deep Learning Netzwerken dargestellt. Aufgrund des begrenzten Platzes innerhalb dieser Arbeit ist der Zeitraum, in denen die Publikationen erschienen sind, auf 2017 bis Juli 2018 beschränkt.

3.1 Endoscopic Vision Challenge 2017

Mit einer Version ihrer Deep Learning Architektur sind Shvets et al. [12] die Gewinner der Endoscopic Vision Subchallenge 2017 – Robotic Instrument Segmentation – geworden. Abbildung 3 zeigt beispielhaft einen Ausschnitt aus dem in der Challenge verwendeten Datensatz. Ziel hierbei war es neben der binären Segmentierung in Instrument und Hintergrund auch die einzelnen Teile der Instrumente zu segmentieren. Hinzu kommt die Segmentierung der verschiedenen Instrumentenarten. Auch sollen die durch die Segmentierung gewonnenen Daten in eine Anzeige von Informationen für z.B. Hilfestellungen während der Operation fließen. Als Basis wurde das U-Net verwendet. Die im Folgenden näher vorgestellte Architektur TernausNet verwendet das VGG16-Netz aus der U-Net Familie als Encoder. Anschließend erfolgt ein Upsampling der Featuremaps gefolgt von einer Convolutional Schicht. Dadurch wird die Auflösung des Outputs vergrößert und man erhält als Ergebnis ein Bild der Ursprungsgröße mit pixelweiser Segmentierung. Um das Ergebnis zu verbessern gibt es zusätzlich Verbindungen zu hochauflösenden Featuremaps des Downsamplingteils. Der gesamte Aufbau ist in Abbildung 4 zu sehen. Das Ergebnis zeigt einen Intersection over Union (IoU)-Wert von 83.60% , einen DICE-Wert von 90,1% und eine Dauer von 184ms zur Segmentierung eines Bildes. [12]

3.2 Weitere Arbeiten

Attia et al. [4] haben ein hybrides Deep Learning Netz entwickelt, um neben der semantischen Segmentierung auch eine Phasenerkennung und Flussidentifikation durchführen zu können. Diese beiden Punkte sind besonders bei computerassistierten Interventionen nötig, um u.a. Personal zuzuweisen und automatische Anweisungen bzw. Hilfestellungen während der Operation zu generieren. Das Netz besteht aus einem Recurrent (RNN) und einem Convolutional Neural Network (CNN). Die Netze unterscheiden sich dadurch, dass ein RNN im Gegensatz zu einem CNN eine Feedbackschleife besitzt. Ziel ist es, durch die Kombination eine

Abbildung 3: Ausschnitt aus dem Datensatz (von links: Originalbild, Binäre Segmentierung, Multiklassensegmentierung: Instrumententeile, Multiklassensegmentierung: Instrumentenart) nach [11]

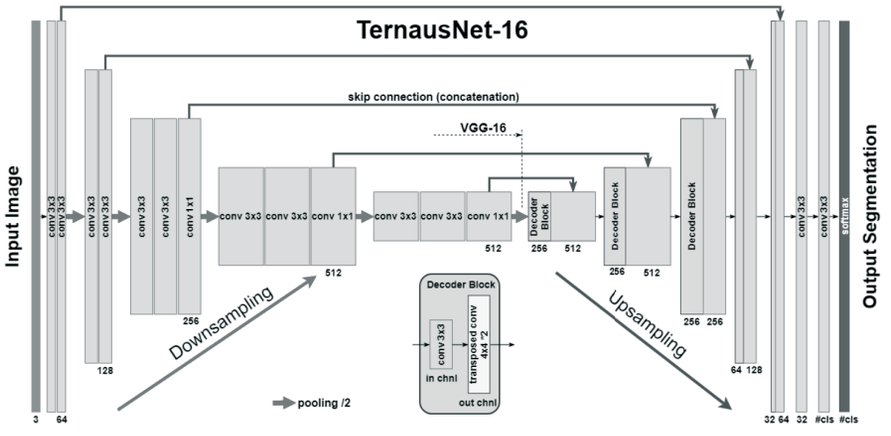


Abbildung 4: Architektur des TernausNet-16 [12]

höhere Genauigkeit bei der Segmentierung in Instrument und Hintergrund zu erreichen. Es ist auch nahezu immun gegenüber den in Abbildung 2 beschriebenen Artefakten und braucht daher keine Dimensionsreduzierung durch eine Farbumwandlung in Graustufen. Das CNN extrahiert in diesem Kontext die Featuremaps und das RNN modeliert die räumlichen Abhängigkeiten zwischen diesen Featuremaps. Dadurch werden kontextuelle Abhängigkeiten zwischen Pixeln ermittelt, die sowohl lokal als auch global gewahrt werden. Zum Lernen dieser Abhängigkeiten wird ein Long Short-Term Memory (LSTM)-Algorithmus verwendet. Dieser arbeitet als ein langes Kurzzeitgedächtnis, das Erfahrungen aus der früheren Iterationen miteinbezieht. Ein beispielhaftes Ergebnis ist in Abbildung 5 zu sehen. Sie erreichten eine Genauigkeit von 93,3%. [4]

Auch die Forschergruppe um Laina [13] segmentiert die chirurgischen Instrumente. Gleichzeitig erkennen sie auch die Position dieser, mithilfe eines Fully Convolutional Residual Network. Die Instrumente werden durch die semantische Segmentierung vom Hintergrund hervorgehoben und für ihre Lokalisierung wird eine Heatmap verwendet. Auf der Map gibt jeder Pixel die Wahrscheinlichkeit an, mit der er sich in der Nähe der

Abbildung 5: Segmentierungsergebnis von [4] in Instrument (dunkel) und Hintergrund (hell), links: Orginalbild, mitte: Groundtruth, rechts: Segmentierungsergebnis

richtigen Landmarke befindet. Dieser Ansatz wurde sowohl mit einem Datensatz der Augenchirurgie als auch mit dem EndoVis Datensatz der Endoscopic Vision Challenge 2016 trainiert und getestet. Als Ergebnis wurde bei der binären Segmentierung eine Genauigkeit von 92,6% und einem DICE-Wert von 88,9% erreicht. Die Segmentierung erlaubt unter anderem auch eine Stelle am Bildschirm zu finden, an die Informationen gesetzt werden können ohne den Operateur zu beeinträchtigen. [13]

Die gleichen Datensätze verwendeten auch Kurmann et al.[14]. Sie empfinden die Erkennung von Instrumenten als Schlüsselrolle zur Gewährleistung der Sicherheit des Patienten. Als Grundlage wurde auch hier das U-Net verwendet, welches man modifizierte und erweiterte. Durch die Erweiterungen ist es nun möglich, dass eine Wahrscheinlichkeit ermittelt wird, mit der ein bestimmtes Instrument im Bild zu sehen ist und eine Map erzeugt wird, auf der die Wahrscheinlichkeitsverteilungen für die Position der Gelenke verzeichnet sind. Ersteres geschieht durch ein Klassifikationslayer am untersten Punkt des Netzwerks. Es besteht aus einem Fully Connected Layer und ist aktiviert durch eine sigmoide Aktivierungsfunktion für eine Ausgabe eines Wahrscheinlichkeitsbereichs. Das Netz besteht insgesamt aus einem vorwärts gerichteten Pfad. Bei der Validierung mit dem retinalen Datensatz wurde eine Genauigkeit über 96,7% mit einem Grenzwertradius von 15 Pixeln erreicht. Anzumerken ist, das bei diesem Ansatz nicht das komplette Instrument segmentiert wird, sondern lediglich die Position der Gelenke im Bild. Darüber hinaus kann das verwendete Netz bestimmen, welche Arten von Instrumenten im Bild zu sehen sind (z.B. Schere oder Greifer). Dabei wurde eine Genauigkeit von 99,9% erreicht. [14]

Um ein Fully Convolutional Network (FCN) auch in Echtzeit verwenden zu können, um Instrumente zu segmentieren, nahmen García-Peraza-Herrera et al. [15] einige Anpassungen vor. Auch sollen dadurch Nachteile durch vernachlässigte Kanten oder auftretende Löcher in den vorhergesagten Masken minimiert werden. Gleichzeitig kommt eine Verlustfunktion zum Einsatz, die für ein ungleichmäßiges Vorkommen der Klassen passt. Dafür wurde der Dice Similarity Koeffizient als Verlustfunktion herangezogen, da dieser bereits als Evaluierung der Segmentierungsqualität verwendet wird. Der Verlust jeder Stufe wird im Deconvolutional Layer für die semantische Segmentierung und die Bestimmung des Verlusts herangezogen. Um die Architektur des Weiteren vor einem Overfitting an das Trainingsset zu bewahren, wurde statt den Rectified Linear Units eine Parametric Rectified Linear Unit (PReLU) verwendet. Dabei handelt es sich um eine nicht-lineare Aktivierungsfunktion, die die negativen Slopes lernt und damit eine end-to-end trainierte Aktivierung bereitstellt. Wird die Gewichtung der negativen Slopes für jede Featuremap bereitgestellt, kommen dadurch eine Menge an zusätzlichen Parametern hinzu, die gleichzeitig die Genauigkeit verbessern. Beim Testen des vorgestellten Netzes (ToolNetH) erreichte die Forschungsgruppe eine Genauigkeit von 81.0% und konnte 28.9 Frames in der Sekunde analysieren. [15]

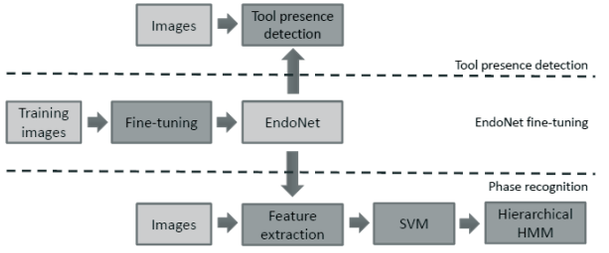


Abbildung 6: Gesamtarchitektur des Ansatzes von[5]

3.3 Mögliche Verbesserungen/Erweiterungen

Im folgenden Abschnitt der Arbeit geht es um verschiedene Aspekte, die das System der Instrumentensegmentierung mittels Deep Learning anpassen und erweitern. Dadurch sollen die Ergebnisse möglichst verbessert und das Einsatzgebiet vergrößert werden.

3.3.1 Trainingsdaten

Statt die neuronalen Netze mithilfe von realen Daten zu trainieren, besteht die Möglichkeit Daten aus Simulatoren zu verwenden. Dabei stellt sich die Frage, ob diese dazu geeignet sind, sie auf die Realität zu übertragen. Diese Fragestellung untersuchten Zisimopoulos et al. [16]. Sie verwendeten einen Simulator für Kataraktoperationen, das sind Eingriffe zur Entfernung der Augenlinse, wenn diese eine Eintrübung besitzt. Die damit erzeugten Daten (ca. 17 000 Bilder) wurden zum Training, zur Validierung und zum Testen verwendet. Die Forschergruppe vergleicht in ihrer Arbeit zwei verschiedene Architekturen. Bei der ersten handelt es sich um ein FCN-VGG-Netz, das die Architektur des ursprünglichen VGGs für semantische Segmentierung durch ein Convolutional Layer anpasst. Als Alternative zu einem FCN-VGG-Netz verwendeten sie ein Conditional Generative Adversarial Network (cGAN), um den Domänentransfer zwischen den simulierten und realen Bildern zu erreichen. Ein GAN besteht aus einem Generator und einem Diskriminator. Ersterer erzeugt Ergebnisse, die letzterer möglichst nicht mehr von echten vorgegebenen Verteilungen unterscheiden kann. Das cGAN besitzt einen Input von Labels als Erweiterung, um die Generierung und Unterscheidung von Bildern zu verbessern. Die Evaluation beider Netzwerke auf realen Daten zeigte eine Genauigkeit von 13% bei cGAN und des FCN-VGGs von 11%. Dieses Ergebnis zeigt, dass eine vielversprechende Erkennung erreicht wurde, aber weitere Arbeit in diese Richtung notwendig ist. Besonders ist dabei auf die Vielfalt der anatomischen Strukturen zu achten, was eine weitere Vergrößerung des Datensets zur Folge hat. [16]

3.3.2 Erkennung von Operationsphasen

Der nächste Schritt, der auf die bereits erfolgreiche Segmentierung von Instrumenten folgen kann, ist die Erkennung der Operationsphasen und der zum aktuellen Zeitpunkt ausgeführten Arbeiten. Diesem Thema widmeten sich unter anderem Twinanda et al. [5] und DiPietro et al. [6]. Letztere verwendeten statt der Bilddaten des Endoskops kinematische Informationen zur Erkennung von Gesten und länger dauernden Handlungen mithilfe von RNNs. Im Falle der Handlungserkennung erreichten sie eine Genauigkeit von 89,5%. [6]

Im Gegensatz dazu stellt das Forscherteam um Twinanda [5] einen Ansatz mit einem CNN, genannt EndoNet, vor, das die Phasen der Operation und die Anwesenheit von Instrumenten automatisch erkennen kann. Der Gesamtaufbau des Ansatzes ist in Abbildung 6 zu sehen. Das EndoNet ist gleichzeitig für die Erkennung der Anwesenheit von chirurgischen Instrumenten und die Extrahierung von Merkmalen zur Erkennung von Phasen zuständig. Für Letzteres ist, nachdem die Features aus dem letzten Fully Convolutional Layer des CNNs übernommen wurden, eine Support Vektor Machine (SVM) und ein Hierarchical Hidden Markov Model (HHMM) Algorithmus zuständig. Der SVM-Algorithmus, den man für die Unterteilung einer Menge in verschiedene Klassen mit möglichst großem Abstand zu den Rändern verwendet, berechnet die Wahrscheinlichkeiten mit der die Phasen vorliegen. Dabei wird allerdings der zeitliche Aspekt nicht betrachtet. Dazu wird dann das HHMM verwendet. Dabei handelt es sich um einen systematisierten unüberwachten Ansatz zur Modellierung von multiscalaren Strukturen (z.B. Sprache) mit der Besonderheit, dass jeder Hidden State des Modells wieder ein eigenständiges HHMM ist, das als Ergebnis eine Sequenz liefert. In diesem Fall werden in einem Level die Abhängigkeiten zwischen den Phasen und in dem anderen Level innerhalb der Phasen modelliert. EndoNet erreicht ein Ergebnis von 89% bei den Phasengrenzen, die innerhalb von 30 Sekunden erkannt werden. 6% der Phasengrenzen werden erst nach einer Zeit von über zwei Minuten erkannt. Dabei ist allerdings auch zu beachten, dass es sich hier um harte Phasengrenzen handelt. In der Realität sind die Übergänge allerdings fließend. [5]

4 Fazit & Ausblick

Anhand der Zusammenstellung der verschiedenen Ansätze, die innerhalb dieser Arbeit vorgestellt werden, ist zu sehen, dass eine Segmentierung von chirurgischen Instrumenten mittels Deep Learning möglich ist. Möchte man dazu Trainingsdaten von Simulatoren verwenden ist noch Forschungsarbeit notwendig, um diesen Prozess zu verbessern. Mithilfe geeigneter Architekturen ist neben der reinen Segmentierung auch die Phasenerkennung denkbar. Dadurch kann bei einer Verwendung in Echtzeit während einer Operation das ungefähre Ende vorhergesagt und die Operationsdaten automatisch annotiert werden.

Nachdem die Instrumente nun richtig segmentiert werden können, gilt es nun geeignete Steuerungssystem zu entwickeln, die diese automatisch verfolgen können, wenn der Operateur dies wünscht. Eine einfache Lösung dafür wäre die Verwendung der Pufferzone, wie von Song et al. beschrieben [2]. Bei der Umsetzung eines Gesamtsystems aus Instrumentensegmentierung, Steuerung und Phasenerkennung dürfen die Gefahren bei dem Einsatz am lebenden Patienten und weitere medizinische Aspekte nicht vernachlässigt werden. Dazu muss man sie genauestens auf Fehler und Risiken überprüfen und testen. Zusätzlich muss ein geeignetes Sicherheitssystem entwickelt werden, dass bei fehlerhafter Segmentierung eingreift.

Literatur

[1] A. Agustinos, R. Wolf, J. A. Long u. a., "Visual servoing of a robotic endoscope holder based on surgical instrument tracking", in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, IEEE, Aug. 2014. DOI: 10.1109/biorob.2014.6913744.

[2] K.-T. Song und C.-J. Chen, "Autonomous and stable tracking of endoscope instrument tools with monocular camera", in *2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, IEEE, Juli 2012. DOI: 10.1109/aim.2012.6266023.

[3] D. Bouget, M. Allan, D. Stoyanov u. a., "Vision-based and marker-less surgical

tool detection and tracking: a review of the literature", *Medical Image Analysis*, Jg. 35, S. 633–654, Jan. 2017. DOI: 10.1016/j.media.2016.09.003.

[4] M. Attia, M. Hossny, S. Nahavandi u. a., "Surgical tool segmentation using a hybrid deep CNN-RNN auto encoder-decoder", in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, Okt. 2017. DOI: 10.1109/smc.2017.8123151.

[5] A. P. Twinanda, S. Shehata, D. Mutter u. a., "EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos", *IEEE Transactions on Medical Imaging*, Jg. 36, Nr. 1, S. 86–97, Mai 2016. DOI: 10.1109/tmi.2016.2593957.

[6] R. DiPietro, C. Lea, A. Malpani u. a., "Recognizing Surgical Activities with Recurrent Neural Networks", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Springer International Publishing, 2016, S. 551–558. DOI: 10.1007/978-3-319-46720-7_64.

[7] G.-Q. Wei, K. Arbter und G. Hirzinger, "Real-time visual servoing for laparoscopic surgery. Controlling robot motion with color image segmentation", *IEEE Engineering in Medicine and Biology Magazine*, Jg. 16, Nr. 1, S. 40–45, 1997. DOI: 10.1109/51.566151.

[8] Z. Zhao, "Real-time 3D visual tracking of laparoscopic instruments for robotized endoscope holder", in *Proceeding of the 11th World Congress on Intelligent Control and Automation*, IEEE, Juni 2014. DOI: 10.1109/wcica.2014.7053773.

[9] S. Voros, J.-A. Long und P. Cinquin, "Automatic Detection of Instruments in Laparoscopic Images: A First Step Towards High-level Command of Robotic Endoscopic Holders", *The International Journal of Robotics Research*, Jg. 26, Nr. 11-12, S. 1173–1190, Nov. 2007. DOI: 10.1177/0278364907083395.

[10] J. Ryu, J. Choi und H. C. Kim, "Endoscopic Vision-Based Tracking of Multiple Surgical Instruments During Robot-Assisted Surgery", *Artificial Organs*, Jg. 37, Nr. 1, S. 107–112, Okt. 2012. DOI: 10.1111/j.1525-1594.2012.01543.x.

[11] Endoscopic Vision Challenge, *Robotic Instrument Segmentation*, https://www.dropbox.com/s/tiwsk0710rx28cu/InstrumentChallenge.ppt?dl=0, (zuletzt besucht am 28.06.2018), 2017.

[12] A. Shvets, A. Rakhlin, A. A. Kalinin u. a., "Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning", März 2018. DOI: 10.1101/275867.

[13] I. Laina, N. Rieke, C. Rupprecht u. a., "Concurrent Segmentation and Localization for Tracking of Surgical Instruments", in *Lecture Notes in Computer Science*, Springer International Publishing, 2017, S. 664–672. DOI: 10.1007/978-3-319-66185-8_75.

[14] T. Kurmann, P. M. Neila, X. Du u. a., "Simultaneous Recognition and Pose Estimation of Instruments in Minimally Invasive Surgery", in *Lecture Notes in Computer Science*, Springer International Publishing, 2017, S. 505–513. DOI: 10.1007/978-3-319-66185-8_57.

[15] L. C. García-Peraza-Herrera, W. Li, L. Fidon u. a., "ToolNet: Holistically-nested real-time segmentation of robotic surgical tools", in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Sep. 2017. DOI: 10.1109/iros.2017.8206462.

[16] O. Zisimopoulos, E. Flouty, M. Stacey u. a., "Can surgical simulation be used to train detection and classification of neural networks?", *Healthcare Technology Letters*, Jg. 4, Nr. 5, S. 216–222, Okt. 2017. DOI: 10.1049/htl.2017.0064.

Entwicklung eines Simulationsmodells eines Time-of-Flight Sensors

Michael Essich
Hochschule Reutlingen
michael_wilhelm.essich@student.reutlingen-university.de

Abstract

In dieser Arbeit wird ein Simulationsmodell eines Time-of-Flight (ToF) Sensors in Unity entwickelt. Dieses Modell soll es ermöglichen, synthetische Trainingsdaten für eine automatisierte Szenenanalyse mithilfe von maschinellen Lernverfahren zu generieren. Im Rahmen dieser Arbeit wird die Funktionsweise von ToF Sensoren erläutert und es werden bereits existierende Simulationstools sowie deren Ansätze zur Simulation eines ToF Sensors betrachtet. Des Weiteren wird der Vorteil der Simulation von Daten im Gegensatz zur Sammlung realer Daten, aber auch deren Herausforderung in Bezug auf eine spätere Anwendbarkeit auf Realdaten erläutert.

CCS Concepts

• **Computing methodologies** → **Machine learning**; **Simulation environments**;

Keywords

Time-of-Flight Sensor, Simulation, Machine Learning

1 Einleitung

Im Bereich des maschinellen Lernens werden große Mengen an Daten benötigt, mit denen z. B. neuronale Netze trainiert werden können. Im Rahmen dieser Arbeit wird deshalb ein Simulationsmodell eines Time-of-Flight (ToF) Sensors entwickelt, mit welchem Tiefendaten von ToF Sensoren simuliert und für eine automatisierte Szenenanalyse mithilfe von maschinellen Lernverfahren exportiert werden können. Die generierten synthetischen Tiefendaten sollen hierbei in einer weiteren Arbeit als Trainingsdaten für die Detektion menschlicher Posen verwendet werden können. Dabei liegt der Fokus nicht auf der physikalisch korrekten Simulation von ToF Sensoren, sondern auf der einfachen Generierung von Trainingsdaten für das maschinelle Lernen.

Zunächst wird die Funktionsweise von ToF Sensoren in Kapitel 2 erläutert und es werden bereits existierende Ansätze zur Simulation von ToF Sensoren betrachtet. Kapitel 3 beschreibt die Entwicklung eines eigenen Simulationsmodells eines ToF Sensors mithilfe der Unity Game Engine, während in Kapitel 4 die Bedeutung von simulierten Daten für das maschinelle Lernen sowie der Wissenstransfer von synthetischen auf Realdaten mit Techniken wie der Domain Randomization oder Domain Adaptation betrachtet wird. Abschließend folgt ein Fazit sowie Ausblick.

2 Stand der Technik

Dieses Kapitel erläutert die Funktionsweise von ToF Sensoren. Des Weiteren werden existierende Ansätze zur Simulation von ToF Sensoren betrachtet.

Betreuer Hochschule:	Prof. Dr.-Ing. Cristóbal Curio Hochschule Reutlingen Cristobal.Curio@Reutlingen-University.de
Mitbetreuer Hochschule:	M.Sc. Dennis Ludl Hochschule Reutlingen Dennis.Ludl@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Michael Essich

2.1 Time-of-Flight Sensoren

Mit ToF Sensoren können 3D Szenen in Echtzeit aufgenommen werden. Dabei liefern diese zum einen ein Tiefen- und zum anderen meist auch ein Amplitudenbild. Während das Tiefenbild die reinen Tiefendaten enthält, entspricht das Amplitudenbild einem mit Infrarotlicht im nahen Infrarotbereich (NIR) belichteten Foto der Szene. [1]

Der NIR Emitter und der Empfänger liegen räumlich nahe beieinander, wodurch Verdeckungen, wie sie bei Systemen auftreten, die mit Triangulation arbeiten, ausgeschlossen sind. ToF Sensoren arbeiten hierbei nach dem Prinzip der Laufzeitmessung eines Lichtsignals. [2]

Die Lichtgeschwindigkeit c beträgt ca. $3 \cdot 10^8 \frac{m}{s}$. Somit ergibt sich die Distanz d mit einer Laufzeit t zu einem bestimmten Objekt durch $d = \frac{c \cdot t}{2}$.

Aufgrund der hohen Geschwindigkeit von Licht besteht die Herausforderung bei ToF Sensoren darin, eine extrem genaue Zeitmessung durchzuführen. Dabei kann zwischen dem Ansatz der Pulse Modulation, der Continuous Wave Intensity Modulation (CWIM) sowie einer Kombination beider Ansätze unterschieden werden. [3]

In den nachfolgenden Abschnitten wird die Pulse Modulation sowie die CWIM genauer betrachtet. Des Weiteren geht Abschnitt 2.1.3 auf Unsicherheiten bei der Messung mit ToF Sensoren sein.

2.1.1 *Pulse Modulation*

Bei der Pulse Modulation wird ein Lichtsignal mit einer bestimmten Dauer bzw. Länge ausgesendet. Des Weiteren wird eine Blende verwendet, welche an die Dauer des Lichtsignals gekoppelt ist und geschlossen werden kann. Die Menge des bis zum Schließen der Blende empfangenen reflektierten Lichts, steht dabei in direktem Zusammenhang mit den Tiefeninformationen. [1]

Die maximal mögliche Distanz d, die bei diesem Ansatz abgedeckt werden kann, hängt somit von der Dauer des ausgesendeten Lichtsignals ab und es gilt $d_{max} = \frac{c \cdot t}{2}$ [1, 2]. Die Messung wird jedoch auch durch die Umgebung (u. a. Dämpfung durch die Luft und Temperatur) beeinflusst [2, 3]. Darüber hinaus ist es schwierig sehr kurze Lichtsignale zu erzeugen, die sehr schnelle Anstiegs- (Zustandwechsel von *aus* zu *an*) und Abfallzeiten (Zustandswechsel von *an* zu *aus*) haben [2].

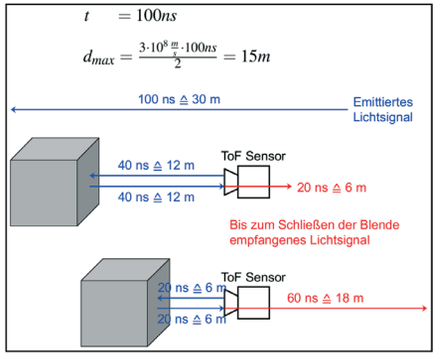


Abbildung 1: Vereinfachte Darstellung der Funktionsweise eines ToF Sensors mit Pulse Modulation

Abbildung 1 stellt die Funktionsweise der Pulse Modulation vereinfacht dar. Es wird von einer Lichtimpulslänge von 100 Nanosekunden und einem daraus resultierenden maximalen Messbereich von bis zu 15 Metern ausgegangen. Dabei ist erkennbar, dass je nach Entfernung des zu messenden Objektes eine unterschiedliche Menge des reflektierten Signals beim ToF Sensor eintrifft.

2.1.2 *Continuous Wave Intensity Modulation (CWIM)*

Im Gegensatz zur Pulse Modulation wird bei der CWIM ein moduliertes kontinuierliches Lichtsignal ausgesendet, wodurch die Anstiegs- und Abfallzeiten der Lichtquelle irrelevant sind. Aufgrund der bekannten Modulationsfrequenz kann die Berechnung der Tiefeninformationen hierbei über die Phase des empfangenen reflektierten Lichtsignals erfolgen, wie dies in Abbildung 2 dargestellt ist. [2]

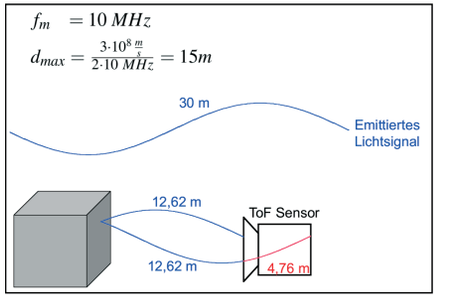


Abbildung 2: Vereinfachte Darstellung der Funktionsweise eines ToF Sensors mit CWIM

Mit einer immer gleichbleibenden Modulationsfrequenz f_m, einer sogenannten homodynen Mischung, und deren Wellenlänge λ_m können Tiefeninformationen eindeutig bis zu einer Entfernung von $\frac{\lambda_m}{2}$ bestimmt werden [3]. Für die maximale Entfernung gilt folglich $d_{max} = \frac{c}{2f_m}$ [1]. Durch die Verwendung mehrerer unterschiedlicher Modulationsfrequenzen, einer sogenannten heterodynen Mischung, und den daraus resultierenden unterschiedlichen d_{max} kann der Messbereich unter Beibehaltung der Genauigkeit und der Eindeutigkeit der Tiefeninformationen vergrößert werden [2].

2.1.3 *Unsicherheiten bei der Messung*

Wie bei gewöhnlichen Kameras kommt es auch bei ToF Sensoren zu Bewegungsunschärfe, wenn sich der Sensor selbst oder ein Objekt in dessen Sichtfeld bewegt. Dies führt vor allem an Stellen, bei denen größere Änderungen der Tiefe bzw. Entfernung auftreten, zu fehlerhaften Tiefeninformationen. Es existieren jedoch diverse Ansätze, durch welche die Bewegungsunschärfe entfernt oder zumindest verringert werden kann. [4]

Darüber hinaus existiert der sogenannte Flying Pixels Effekt, der in Abbildung 3(b) bei

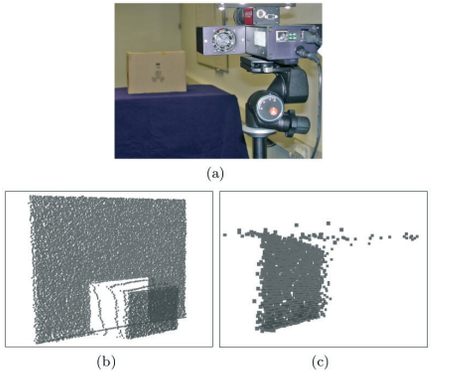


Abbildung 3: Flying Pixels Effekt [5]

geringem Abstand zwischen Objekt und Hintergrund und in 3(c) bei großem Abstand zwischen Objekt und Hintergrund dargestellt ist. 3(a) zeigt den Aufbau der Szene. Ein Pixel des ToF Sensors misst nicht genau einen Punkt auf der Oberfläche eines Objektes, sondern vielmehr einen gewissen Bereich. Kommt es in diesem Bereich zu größeren Tiefen- bzw. Entfernungsunterschieden, wie z. B. an Objektkanten, führt dies zu fehlerhaften Messdaten aufgrund der Mischung verschiedener Phasen des Lichtsignals durch Reflexionen in unterschiedlichen Entfernungen. Diese Artefakte werden Flying Pixels genannt. [5, 6]

Flying Pixels können erkannt und entfernt werden, indem die Anzahl der benachbarten Pixel betrachtet und mit einem bestimmten Schwellenwert verglichen wird [5].

Des Weiteren können Messfehler aufgrund von Reflexionen auftreten, die außerhalb des eindeutigen Messbereichs, wie unter Abschnitt 2.1.2 beschrieben, liegen. Diese Reflexionen treffen wegen der längeren Laufzeit erst bei einer späteren Messung am Sensor ein und führen zu sogenanntem Backfolding. Dabei wird die Distanz zu Objekten als kürzer bestimmt, als diese tatsächlich ist [7]. In Abbildung 3(c) erscheinen die Flying Pixels an der oberen Kante vor dem Objekt, da der Hintergrund außerhalb des Messbereichs des ToF Sensors liegt, was zu Backfolding führt.

2.2 Existierende Simulationen

Es existieren diverse Arbeiten, die sich mit der Simulation von ToF Sensoren beschäftigen. In [8] entwickeln B. Jähne und M. Schmidt ein physikalisches Modell zur Simulation von ToF Sensoren. Meister et al. beschäftigen sich in [9] mit Global Illumination und der bis dahin nicht beachteten Simulation von Mehrwegeffekten (Multipath Effects), die durch (mehrfache) Reflexion von Lichtsignalen entstehen, während M. Keller in [5] eine physikbasierte Echtzeitsimulation für ToF Sensoren und eine Methode zur Akkumulation von Tiefendaten vorstellt.

Aufgrund der freien Verfügbarkeit werden in den nachfolgenden Abschnitten 2.2.1 und 2.2.2 die beiden Simulationstools PMDSim [6] sowie BlenSor [7] betrachtet.

2.2.1 PMDSim

Lambers et al. stellen mit PMDSim[1] [6] eine Anwendung zur realistischen, physikbasierten Simulation von ToF Sensoren vor. Der Fokus liegt nicht auf der Generierung realistischer Tiefendaten, sondern auf der Simulation des Sensors selbst mit dem Ziel, zwei verschiedene Chiplayouts zu evaluieren. [6]

Die Simulation beschränkt sich hierbei auf CWIM basierte Sensoren. Für ein physikalisch korrektes Simulationsmodell werden deshalb ausschließlich physikalische Einheiten verwendet und es wird die individuelle Pixelgeometrie des Sensors sowie dessen Layout modelliert. Des Weiteren werden Effekte, die durch verschiedene Linsen oder auch Bewegung entstehen, sowie der Flying Pixels Effekt im Modell berücksichtigt. [6]

PMDSim geht wie die meisten Ansätze davon aus, dass es sich bei den Oberflächen von Objekten um Lambertsche Reflektoren handelt. Das bedeutet, dass Objekte das Infrarotlicht in alle Richtungen gleichstark reflektieren. Verschiedene Materialien und deren Eigenschaften werden somit nicht berücksichtigt.

Die Entwicklung der Anwendung erfolgte in C++ unter Verwendung von OpenGL. Die Berechnung der Beleuchtung und die Sensorsimulation wird auf der GPU durchgeführt. [6] Unter anderem kann ein Export der Tiefen-, Amplituden-, Intensitäts- sowie Phaseninformationen im CSV-Format erfolgen.

2.2.2 BlenSor

BlenSor[2] ist eine Simulationssoftware für verschiedene Sensoren zur 3D Tiefenmessung, wie LIDAR und ToF, welche auf der Open Source 3D Grafiksuite Blender[3] basiert und sich in diese vollständig integriert. Mit BlenSor ist es sowohl möglich die Testszenarien zu modellieren als auch diese im Anschluss direkt zu simulieren. Der Fokus liegt hierbei nicht auf einer Echtzeit-, sondern Offline-Simulation, wodurch ein hoher Detailgrad erreicht werden kann. Die Simulationsergebnisse können entweder exportiert oder als Mesh der 3D Simulation hinzugefügt werden, wodurch mit diesen auch interagiert werden kann. [7]

Blender erlaubt zwar bereits die Nutzung von Raycasts über dessen Python-API, jedoch mangelt es dabei an Performance, weshalb BlenSor einen Teil des Blender Cores verändert, um mehrere simultane Raycasts durchführen zu können. BlenSor führt die Simulation jedoch nicht auf der GPU, sondern ausschließlich auf der CPU aus. Die Tiefeninformationen werden über die Distanz zu den Positionen, an denen ein Strahl (Ray) auf eine Fläche trifft, bestimmt. Darüber hinaus ist es auch möglich die Materialinformationen zu Flächen, auf denen Rays auftreffen, zu erhalten. [7]

Eine Besonderheit bei BlenSor ist die Möglichkeit Reflexionen und darüber hinaus auch Backfolding zu simulieren. Bei der Simulation von Backfolding werden die gemessenen Distanzen d oberhalb der Hälfte der maximalen Messdistanz in die untere Hälfte nach [7] wie folgt abgebildet:

$$d = \begin{cases} d_{real}, & d_{real} < \frac{d_{max}}{2} \\ d_{real} - \frac{d_{max}}{2}, & ansonsten. \end{cases}$$

Reflexionen werden simuliert, indem Rays je nach Reflexivität des Materials reflektiert und weiter verfolgt werden. Dies führt dazu, dass die Entfernung von Objekten mit reflektierenden Oberflächen als größer gemessen wird, als diese tatsächlich ist, da die Rays durch die Reflexionen eine größere Laufzeit haben. [7]

Für alle von BlenSor simulierten Sensoren wird ein Export der Simulationsergebnisse in einem eigenen Binärdateiformat sowie als PCD-Datei (Point Cloud Data), die von der Point Cloud Library (PCL)[4] verarbeitet werden kann, unterstützt. Wurde als Simulationsmodell *Kinect* gewählt, so können die Ergebnisse des Weiteren als Portable Graymap (PGM) exportiert werden. [10, 11]

3 Entwicklung eines Simulationsmodells eines Time-of-Flight Sensors in Unity

In diesem Kapitel wird die Entwicklung des Simulationsmodells mithilfe der Unity Game Engine beschrieben. Durch die Verwendung von Unity ergibt sich der Vorteil, dass bereits eine umfangreiche API zur Verfügung steht, welche einen Import von 3D Objekten, Animationen sowie Motion Capturing Daten (im Filmbox FBX-Format) erlaubt. Darüber hinaus können Szenen direkt im Unity Editor zusammengestellt sowie bearbeitet werden. Das Simulationsmodell kann des Weiteren in Form eines C#-Scriptes auf einfache Weise in bereits existierende Szenen mit eingebunden werden. Da es sich bei Unity in erster Linie um eine Game Engine handelt, existieren diverse Herausforderungen und Einschränkungen im Bezug auf eine physikalisch korrekte Simulation eines ToF Sensors.

Das Ziel dieser Arbeit ist die Generierung von Trainingsdaten für eine automatisierte Szenenanalyse mithilfe maschineller Lernverfahren, wobei keine physikalisch korrekte Simulation durchgeführt werden soll. Es wäre hierbei ideal, wenn ein neuronales Netz auf den synthetischen ToF Sensordaten trainiert wird, aber soweit generalisiert, dass auch auf Realdaten gute Ergebnisse erzielt werden. Methoden wie die Domain Randomization [12, 13] oder Domain Adaptation [14] versuchen diese Generalisierung bzw. einen Wissenstransfer von einer Quell- auf eine Zieldomäne zu erreichen. Weitere Informationen dazu sind in Kapitel 4 zu finden. Die generierten Daten sollen insbesondere im Bereich der Domain Adaptation verwendet werden.

Das entwickelte Simulationsmodell eines ToF Sensors orientiert sich an der Umsetzung von BlenSor und nutzt ebenfalls Raycasts, um die Distanz vom Sensor zu einem Objekt zu messen. Der Sensor wird dabei in Form einer Kamera realisiert, welcher ein C#-Script zur Simulation des Sensors zugeordnet ist. Dadurch entspricht die Ausrichtung der Kamera auch der Ausrichtung des Sensors. Des Weiteren erlaubt es die von Unity bereitgestellte Methode *ScreenPointToRay* Raycasts von der Kamera aus zu Punkten innerhalb deren Sichtfeld durchzuführen. Die Anzahl an Raycasts und somit die Auflösung des ToF Sensors entspricht hierbei der gewählten Auflösung der Game Preview im Unity Editor. Dadurch können verschiedene Auflösungen des Sensors simuliert werden, wobei das Seitenverhältnis der Kamera (Aspect Ratio), deren Field of View als auch Far Clipping Pane berücksichtigt werden. So wird die maximale Sicht- bzw. Reichweite des ToF Sensors durch die Far Clipping Pane der Kamera bestimmt.

Für jeden Pixel des Tiefenbildes wird die Distanz des Raycasts vom Sensor bis zum Auftreffen an einem Objekt gemessen und gespeichert. Hierbei ist zu beachten, dass Raycasts

[1] https://gitlab.marlam.de/marlam/pmdsim (besucht am: 06.08.2018)

[2] http://www.blensor.org (besucht am: 08.08.2018)

[3] https://www.blender.org (besucht am 08.08.2018)

[4] http://pointclouds.org (besucht am: 08.08.2018)

nicht das Objekt selbst, sondern dessen Collider treffen. Dies ist wichtig, da Objekte, welchen kein Collider zugewiesen ist, von den Raycasts nicht erfasst werden können.
Es existieren unterschiedliche Arten von Collidern wie z. B. Box Collider für Würfel und Quader oder Sphere Collider für Kugeln. Bei geometrisch komplexeren Modellen, wie einem Menschen, muss ein MeshCollider genutzt werden, der sich an das Mesh des Objekts anpasst. Da Unity als Game Engine auf eine hohe Performance ausgelegt ist, werden MeshCollider, nachdem diese angelegt wurden, auch bei einer Änderung des Mesh des Objekts bei Animationen nicht erneut an das veränderte Mesh angepasst. Deshalb ist es bei Objekten, die ihre Form z. B. aufgrund von Animationen verändern, notwendig den MeshCollider bei jedem Frame erneut anzupassen. Dies wurde durch ein C#-Script gelöst, welches komplexeren, animierten Objekten zugewiesen werden kann und die erforderlichen Neuberechnungen des MeshColliders übernimmt.

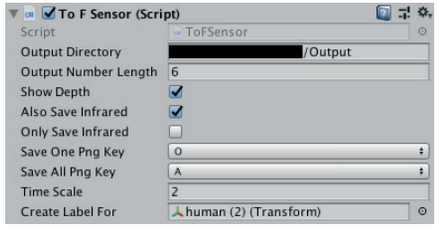


Abbildung 4: Das ToF Sensor Script und dessen Parameter

Die einstellbaren Parameter des ToF Sensor Scripts sind in Abbildung 4 dargestellt. Über frei wählbare Tasten ist es möglich ein einzelnes Bild oder fortlaufende Bildsequenzen aufzunehmen. Sollen hierbei sehr detaillierte Bewegungsabläufe erzeugt und exportiert werden, so muss ein Time Scale kleiner als 1 gewählt werden, da dieser die Geschwindigkeit der Simulation und somit auch der Bewegungsabläufe steuert. Die Tiefendaten werden als zweidimensionales Array, das für jeden Bildpunkt eine Distanz in Metern enthält, gespeichert und in Form einer CSV-Datei in einem entsprechenden Unterordner im gewählten Ausgabeverzeichnis abgelegt.
Zur Visualisierung der Tiefendaten in Unity während der Simulation, wie in Abbildung 5 dargestellt, werden die Distanzen als Graustufen in den RGB Farbraum mit einer Farbtiefe von 8 Bit je Farbkanal konvertiert. Dazu wird für jeden Pixel des Tiefenbildes, also je durchgeführten Raycast, und für jeden Farbkanal (R, G, B) die Berechnung

$$Color = 255 - Round(\frac{distance}{distance_{max}} \cdot 255)$$

durchgeführt. $Distance_{max}$ entspricht dabei der Far Clipping Pane.

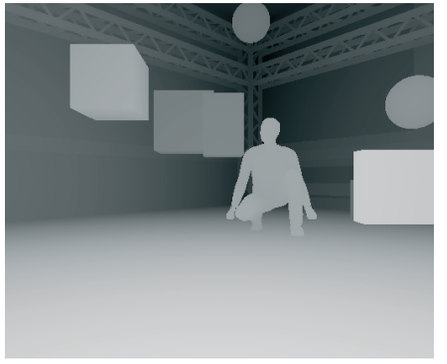


Abbildung 5: Darstellung der Tiefendaten in Unity als Graustufenbild

Die Generierung eines korrekten Amplituden- bzw. Infrarotbildes in Unity stellt eine große Herausforderung dar. Dazu müsste das Verhalten von NIR bei verschiedenen Materialien (Reflexion, Absorption) modelliert und simuliert werden, was Eingriffe in die Renderpipeline von Unity selbst erfordern würde. Aus diesem Grund wurde der Post Processing Stack, welcher für Unity als kostenfreies Asset zur Verfügung steht, als vereinfachte Simulation gewählt. Mithilfe des vom Post Processing Stack bereitgestellten Color Grading können die Farben und die Luminanz des gerenderten Bildes angepasst werden. Somit ist es im einfachsten Fall möglich, wie in Abbildung 6 dargestellt, ein Graustufenbild zu generieren.
Zur Simulation des NIR Emitters wurde ein Spot Light an die Kamera gebunden, sodass die Beleuchtung der Szene ausschließlich von der Position des Sensors ausgeht. Durch eine Änderung der Größe des Lichtkegels, der Intensität des Spot Light oder auch des Color Grading kann der Effekt für das Infrarotbild angepasst werden.

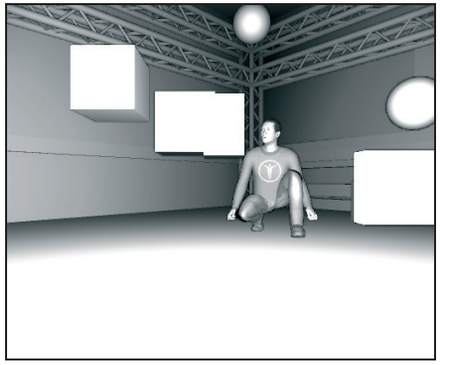


Abbildung 6: Infrarotbild

Im C#-Script erfolgte die Umsetzung des Infrarotbildes durch das Erzeugen einer neuen Kamera zur Laufzeit, welche die Eigenschaften der Hauptkamera kopiert. Dieser Kamera wird anschließend ein Post Processing Behaviour zugewiesen.
Eine Berücksichtigung von verschiedenen Materialien und deren Reflexivität findet nicht statt. Über die Raycasts wäre es zwar möglich das Material des Objektes bzw. der Trefferstelle zu bestimmen, jedoch wäre es hierzu erforderlich über jedes Triangle des Mesh zu iterieren und mit dem Triangle des Treffers zu vergleichen. Darüber hinaus müssten anschließend je nach Reflexivität des Materials erneut Raycasts von der Trefferstelle aus durchgeführt werden, um Reflexionen zu simulieren. Dies ist bei komplexeren Meshes mit einem hohen Rechenaufwand verbunden, weshalb Materialien zunächst nicht berücksichtigt werden.
Ebenso könnte der Flying Pixels Effekt, wie auch in [6] beschrieben, durch Oversampling simuliert werden, indem die Auflösung des ToF Sensors zunächst erhöht und anschließend wieder verringert wird. Bei der Verringerung werden Durchschnittswerte von mehreren Pixeln gebildet, um einen Pixelwert zu erhalten. An Objektkanten existieren größere Unterschiede bei den Distanzen der Pixel, wodurch Flying Pixels simuliert werden können. Da dieser Vorgang jedoch eine Erhöhung der durchgeführten Raycasts und dadurch einen höheren Rechenaufwand erfordern würde, werden während der Simulation keine Flying Pixels simuliert. Dieser Effekt kann jedoch im Nachgang erzeugt werden, indem die Pixelwerte der benachbarten Pixel eines jeden Pixels des Tiefenbildes aufsummiert werden und deren Durchschnitt berechnet wird. Weicht dieser Durchschnittswert stark vom Pixelwert des betrachteten Pixels ab, so bedeutet dies, dass die Distanzen der benachbarten Bildpunkte stark schwanken und an dieser Stelle der Flying Pixels Effekt auftreten kann. Über die Größe der erlaubten Abweichung kann die Stärke des Effekts bestimmt werden. Da bei dieser Methode jedoch keine höhere Auflösung verwendet wird, ist die in [6] beschriebene Methode genauer und entspricht mehr dem physikalischen Ursprung des Flying Pixels Effekts, wie dieser unter 2.1.3 beschrieben ist.
In einer frühen Entwicklungsstufe konnte den Tiefendaten bereits während der Simulation Rauschen hinzugefügt werden. Hierauf wurde jedoch in der weiteren Entwicklung verzichtet, da Rauschen, wie auch der Flying Pixels Effekt, auch im Nachhinein den Tiefendaten hinzugefügt werden kann, ohne dass eine weitere Simulation ausgeführt werden müsste.

Darüber hinaus könnte Backfolding nach dem von BlenSor verwendeten und in Abschnitt 2.2.2 beschriebenen Prinzip implementiert werden, wodurch ein Effekt, wie in Abbildung 3(c) dargestellt, erzielt werden könnte. Da hierbei jedoch auch Distanzen, die über den Messbereich, der durch die Far Clipping Pane begrenzt ist, in die Berechnung miteinbezogen werden müssten, sollte dieser Effekt bereits bei der Simulation berechnet werden. Des Weiteren können Labels bzw. Annotationen zu den Tiefen- und Infrarotdaten in Form von XML-Dateien erzeugt werden. Da die Daten später für die Erkennung menschlicher Posen genutzt werden sollen, beinhalten die Labels zum Zeitpunkt dieser Arbeit die Positionen von 15 Gelenken bzw. Gliedmaßen eines menschlichen Modells.

4 Bedeutung für das maschinelle Lernen

Beim überwachten maschinellen Lernen soll z. B. ein neuronales Netz durch gegebene Eingabe- und den zugehörigen Ausgabedaten eine Funktion erlernen, welche die Eingaben auf die Ausgaben abbildet. Mithilfe dieser Funktion ist es anschließend möglich eine Klassifikation oder Regression mit bisher unbekannten Eingabedaten durchzuführen. [15]
Die gegebenen Eingabe- und zugehörigen Ausgabedaten bilden dabei die gelabelten Trainingsdaten für das neuronale Netz. Betrachtet man z. B. die semantische Bildsegmentierung, bei der Bildinhalte pixelweise verschiedenen Klassen zugeordnet werden, ergibt sich jedoch die Herausforderung eine sehr große Menge an Trainingsdaten, bei der jedem Pixel eines Bildes eine Klasse zugeordnet ist, zu erstellen. Das Labeln, also Erstellen der erwarteten Ausgabe zu jeder Eingabe, ist bei Realdaten sehr zeitaufwändig, weshalb häufig auf die Generierung synthetischer Trainingsdaten zurückgegriffen wird. Dies hat vor allem den Vorteil, dass auch die erwartete Ausgabe zu jeder Eingabe automatisch generiert werden kann. [16]
Eine weitere Herausforderung für das neuronale Netz ist die Generalisierung, sodass dieses trotz des Trainings auf lediglich synthetischen Daten auch gute Ergebnisse auf Realdaten erzielt. Es existieren verschiedene Verfahren, die versuchen dies zu erreichen. Nachfolgend werden die Domain Randomization und Domain Adaptation betrachtet.

4.1 Domain Randomization

Die Idee der Domain Randomization (DR) ist, eine sehr hohe Variabilität in z. B. der Objektanzahl, den Objektpositionen, weiteren Objekten in der Umgebung sowie unterschiedlichen Umgebungen in den synthetischen Trainingsdaten zu generieren, damit das Netz ohne weiteres Nachtrainieren auf Realdaten generalisiert. Dabei soll eine Low-Fidelity-Simulation ausreichen, da sich das Netz durch die hohe Variabilität in den Trainingsdaten auf die wesentlichen Merkmale fokussieren soll. [12, 13]
Tobin et al. nutzen DR in [12] um ein neuronales Netz auf die Erkennung der 3D Position von geometrischen Objekten in RGB Bildern zu trainieren. Dabei konnte deren Hypothese, dass ein vorhergehendes Training auf Realdaten wesentlich für eine spätere Generalisierung ist, widerlegt werden. So konnte mithilfe der DR und ausschließlich synthetischen Trainingsdaten nahezu dieselbe Leistung auf Realdaten erzielt werden, wie mit einem vortrainierten Netz. Im Durchschnitt konnte die 3D Position von Objekten auf den Realdaten bis auf 1,5 cm genau bestimmt werden. [12]

4.2 Domain Adaptation

Domain Adaptation (DA) ist ein spezieller Fall des Transfer Learning (TL). Unter TL versteht man den Wissenstransfer von einem gelösten Problem des maschinellen Lernens auf ein neues, bisher ungelöstes, aber ähnliches Problem. Je nachdem welche Daten über die Quell- oder Zieldomäne vorliegen, kann TL in die Bereiche des induktiven, transduktiven oder unüberwachten TL eingeteilt werden,

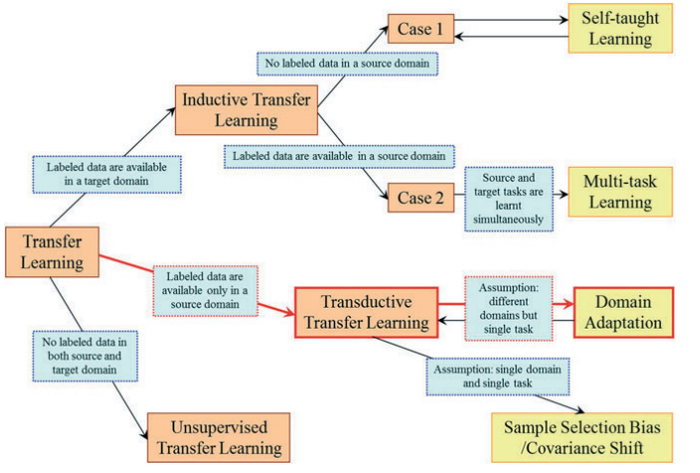


Abbildung 7: Die verschiedenen Arten des Transfer Learning [14]

wobei sich diese Bereiche teilweise weiter aufteilen lassen. DA wird dem transduktiven TL zugeordnet, bei dem gelabelte Daten nur in der Quelldomäne vorliegen (unüberwachte DA) oder nur wenige gelabelte Daten in der Zieldomäne vorhanden sind (teilüberwachte DA). Die Aufgabe sowie die Label sind in beiden Domänen identisch. DA steht hierbei nicht für genau eine, sondern eine Vielzahl an unterschiedlichen Verfahren, bei denen u. a. eine Neugewichtung der Klassen, eine Anpassung des Merkmalsraums oder auch ein Nachtrainieren mit wenigen gelabelten Daten der Zieldomäne erfolgen kann. [14]
Abbildung 7 stellt die verschiedenen Arten des TL und deren Zusammenhänge dar.
Ein Beispiel für DA ist ein von Tzeng et al. in [17] vorgestellter Ansatz, bei dem eine Optimierung hinsichtlich der Domäneninvarianz vorgenommen wird. Dabei werden die Klassenlabel gelernt und es wird gleichzeitig eine Repräsentation gesucht, welche die Quell- und Zieldomäne so ähnlich wie möglich erscheinen lässt. Dies wird durch eine neue Architektur eines Convolutional Neural Networks (CNN) erreicht, die sowohl den Loss der Klassifizierung, als auch den Loss der Domain Confusion, also der Unterscheidbarkeit der Domänen, minimiert. Darüber hinaus werden die Wahrscheinlichkeitsverteilungen der Klassen der Quell- und Zieldomäne genutzt, um das Modell hinsichtlich dieser Verteilung zu optimieren. [17]

5 Fazit und Ausblick

Mithilfe des entwickelten Simulationsmodells eines ToF Sensors können Tiefendaten in Unity auf einfache Weise generiert werden. Die Sensorsimulation kann durch Zuweisung des ToF Sensor C#-Scripts zu einer Kamera in bereits existierende Unity-Szenen integriert werden. Eine Anpassung der Parameter des ToF Sensors, wie z. B. dem Sichtfeld und der Reichweite, kann direkt über die Parameter der Kamera erfolgen.
Zukünftig soll das Simulationsmodell zur Generierung von Trainingsdaten für die menschliche Posenerkennung genutzt werden. In diesem Rahmen wird auch ein Vergleich der Performance eines neuronalen Netzes auf den generierten synthetischen und realen Tiefendaten erfolgen. Darüber hinaus soll auch untersucht werden, wie sich Methoden der Domain Adaptation und Domain Randomization auf

die Erkennungsgenauigkeit auswirken bzw. sich die Leistung optimieren lässt.

Literatur

[1] M. Grzegorzek, C. Theobalt, R. Koch u. a., Hrsg., *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, 2013. DOI: 10.1007/978-3-642-44964-2.

[2] R. Lange, *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*, eng, 2000. Adresse: http://dokumentix.ub.uni-siegen.de/opus/volltexte/2006/178 (besucht am 28. 07. 2018).

[3] B. Jähne, P. Geissler und H. Haussecker, Hrsg., *Handbook of Computer Vision and Applications, Sensors and Imaging*. Academic Press, 1999, Bd. 1, ISBN: 0123797713.

[4] M. Hansard, S. Lee, O. Choi u. a., *Time-of-Flight Cameras*. Springer London, 2013, ISBN: 978-1-4471-4657-5. DOI: 10.1007/978-1-4471-4658-2.

[5] M. Keller, *Real-time simulation of time-of-flight sensors and accumulation of range camera data*, eng, 2015. Adresse: http://dokumentix.ub.uni-siegen.de/opus/volltexte/2016/1022 (besucht am 28. 07. 2018).

[6] M. Lambers, S. Hoberg und A. Kolb, „Simulation of Time-of-Flight Sensors for Evaluation of Chip Layout Variants“, *IEEE Sensors Journal*, Jg. 15, Nr. 7, S. 4019–4026, Juli 2015, ISSN: 1530-437X. DOI: 10.1109/JSEN.2015.2409816.

[7] M. Gschwandtner, R. Kwitt, A. Uhl u. a., „BlenSor: Blender Sensor Simulation Toolbox“, in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin u. a., Hrsg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, S. 199–208, ISBN: 978-3-642-24031-7.

[8] M. Schmidt und B. Jähne, „A Physical Model of Time-of-Flight 3D Imaging Systems, Including Suppression of Ambient Light“, in *Dynamic 3D Imaging*, A. Kolb und R. Koch, Hrsg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, S. 1–15, ISBN: 978-3-642-03778-8.

[9] S. Meister, R. Nair und D. Kondermann, „Simulation of Time-of-Flight Sensors using Global Illumination“, in *Vision, Modeling & Visualization*, M. Bronstein, J. Favre und K. Hormann, Hrsg., The Eurographics Association, 2013, ISBN: 978-3-905674-51-4. DOI: 10.2312/PE.VMV.VMV13.033-040.

[10] M. Gschwandtner. (16. Jan. 2018). Blensor: Export, Adresse: http://www.blensor.org/export.html (besucht am 09. 08. 2018).

[11] ——, (16. Jan. 2018). Blensor: File Format, Adresse: http://www.blensor.org/file_format.html (besucht am 09. 08. 2018).

[12] J. Tobin, R. Fong, A. Ray u. a., „Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World“, *CoRR*, 2017. arXiv: 1703.06907v1.

[13] J. Tremblay, A. Prakash, D. Acuna u. a., „Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization“, *CoRR*, 2018. arXiv: 1804.06516v3.

[14] G. Csurka, „Domain Adaptation for Visual Applications: A Comprehensive Survey“, 17. Feb. 2017. arXiv: 1702.05374v2 [cs.CV].

[15] P. Norvig und S. Russell, *Künstliche Intelligenz*. Pearson Studium, 1. Juni 2012, ISBN: 3868940987.

[16] S. R. Richter, V. Vineet, S. Roth u. a., „Playing for Data: Ground Truth from Computer Games“, 7. Aug. 2016. arXiv: 1608.02192v1 [cs.CV].

[17] E. Tzeng, J. Hoffman, T. Darrell u. a., „Simultaneous Deep Transfer Across Domains and Tasks“, 8. Okt. 2015. arXiv: 1510.02192v1 [cs.CV].

Evaluierung von Einsatzszenarien für Augmented Reality im Rettungsdienst

Maxim Stoljar
Hochschule Reutlingen
maxim.stoljar@student.reutlingen-university.de

Abstract

Der Rettungsdienst ist für die Versorgung medizinischer Notfälle verantwortlich, wobei das Rettungsfachpersonal, sowie Rettungsmittel koordiniert und effizient eingesetzt werden müssen. Der Fortschritt im Rettungsdienstes hinkt dabei den technischen Möglichkeiten hinterher. Im Rahmen dieser Arbeit werden Problemstellungen im Rettungsprozess anhand eines benutzerorientierten Ansatzes ermittelt und Anforderungen an ein System erarbeitet, dass den schwerwiegendsten Schwachstellen im Rettungsdienst mithilfe von Augmented Reality entgegenwirken soll. Anhand der identifizierten Schwachstellen in der Dokumentation, Kommunikation und dem Notarztmangel wird die Entwicklung eines Telenotarztsystems vorgeschlagen.

CCS Concepts

• **Human-centered computing** → *Mixed / augmented reality*; *User centered design*; • **Applied computing** → Life and medical sciences;

Keywords

Rettungsdienst, Augmented Reality,Benutzerzentrierte Entwicklung

Betreuer Hochschule:	Prof. Dr. - Ing. Oliver Burgert Hochschule Reutlingen Oliver.Burgert@Reutlingen-University.de
Betreuer Firma	Richard De Klerk ITK Engineering GmbH Richard.Deklerk@itk-engineering.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen

Copyright 2018 Name des Autors

1 Einleitung

Der Rettungsdienst besteht aus Notarzt (NA), Notfallsanitäter (NotSan), Rettungssanitäter (RettSan), Einsatzleitung (EL) und den vorhandenen Rettungsmitteln, wie beispielsweise Rettungswagen. Im Gegensatz zur medizinischen Patientenversorgung in den Krankenhäusern, hinkt der Fortschritt im Rettungsdienst den technischen Möglichkeiten hinterher. Gerade im Rettungsdienst sollte die technische Aufgabenunterstützung und Prozessoptimierung ausgeschöpft werden, weil gewohnte diagnostische Möglichkeiten nicht zur Verfügung stehen und lebenswichtige Entscheidungen unter Zeitdruck getroffen werden müssen. [1]

In diesem Zusammenhang werden telemedizinsche Anwendungen als möglicher Lösungsansatz für vorhandenen Schwachstellen im Notfall- und Qualitätsmanagement vorgeschlagen [1]–[3]. Wobei eine Vielzahl an Projekten einen positiven Effekt nachweisen konnten, der sich über alle Glieder der Rettungskette erstreckt [1]. Dabei wurden die Teilaspekte Alarmierung Erstversorgung, Transport, Notaufnahme, intensiv Therapie oder operative Behandlung, Rehabilitation und Outcome-Beurteilung betrachtet.

1.1 Problemstellung & Ziel

Die medizinische Versorgung von Patienten hat sich dahingehend entwickelt, dass eine Vielzahl von Notfällen nicht mehr im Krankenhaus sondern präklinisch durch den Rettungsdienst durchgeführt werden, wodurch die Zahl an Rettungsdienst- und Notarzteinsätzen stetig steigt [2], [4]. Zusätzlich nimmt der Ärztemangel in Krankenhäusern immer

stärker zu [5], [6]. In diesem Zusammenhang ist durch zunehmende Kosten im Gesundheitswesen der Bedarf an Notärzten nicht überall und jederzeit zu decken [4]. Daraus folgen große organisatorische Herausforderungen, steigende Qualitätseinbußen und messbare Defizite, die anhand einer verlängerten Eintreffzeit des Notarztes belegt werden können [7]. Die erhöhte Eintreffzeit am Einsatzort führt zu einer verlängerten therapiefreien Zeit, die beispielsweise im Falle eines Herzinfarkts oder Schlaganfalls von hoher Relevanz sein kann [2]. Hinzu kommt die sinkende Zahl dauerhaft besetzter Notarztstandorte in ländlichen Regionen [8].
Neben dem Notarztmangel können während Einsätzen Kommunikationsprobleme zwischen den Einsatzkräften vor Ort und der Einsatzleitung auftreten, was aus einer rein auditiven Beschreibung des Notfalls entsteht. Dabei ist das Problem, dass die Einsatzleitung eine andere Vorstellung des Unfallortes annehmen, weil keinerlei visuelle Kommunikation besteht [9]. Dabei können Informationen verloren gehen, die bei der Entscheidungsfindung essentiell wären. Außerdem ist eine hilfestellende Instruktion durch die Einsatzleitung mithilfe der aktuellen Hilfsmittel nur schwer möglich [9]. Durch das fehlende, eindeutige Verständnis der Situation ist eine optimale Ressourcenverteilung, gute Teamarbeit innerhalb unterschiedlicher Rettungsdienstbereiche sowie mit anderen Berufsgruppen, wie Feuerwehr und Polizei, kaum zu bewältigen [10]. Die Problematik wird besonders in einem Massenanfall von Verletzten (MANV) deutlich, wobei möglichst detailierte Informationen über Bevölkerungsdynamik, Verletztenanzahl, Unfallauswirkungen, besondere Bevölkerungsgruppen, wie Kinder, ältere Menschen und Behinderte, an die Einsatzleitung zur effektiven Einsatzplanung übergeben werden müssen [11].
Die aufgezeigte Problemstellung darf nicht durch geringere Qualitätsanforderungen an den Notarzt oder den Rettungsdienst gelöst werden. Vielmehr müssen hilfestellende Anwendungen entwickelt werden. Das Ziel dieser Arbeit ist bestehende Schwachstellen in der Notfallmedizin anhand einer Befragung innerhalb des Rettungsdienstes zu analysieren. Darauf aufbauend soll ein Anforderungskatalog an die Technik formuliert werden. Abschließend sollen unterschiedliche Lösungsansätze evaluiert werden und die erfolgversprechendste Anwendung für eine spätere Umsetzung ausgewählt werden. Hierbei werden nur Einsatzmöglichkeiten von Augmented Reality (AR) in Rettungsdienst untersucht.

1.2 Aufbau der Arbeit

In Kapitel 2 werden nach Problemstellung und Zielsetzung, relevante Arbeiten vorgestellt, die sich mit den zuvor genannten Schwachstellen im Notfall- und Qualitätsmanagement beschäftigen. Darauf aufbauend wird in Kapitel 3 die Methodik zur Evaluierung von Einsatzmöglichkeiten von AR im Rettungsdienst anhand eines benutzerorientierten Ansatzes vorgestellt. Dabei wird das Vorgehen zur Erfassung des Nutzungskontext beschrieben. Die resultierenden Erkenntnisse zur Problemstellung werden in Kapitel 4 vorgestellt. Basierend auf den Erkenntnisse werden drei besonders schwerwiegende Problemstellungen ausgewählt und Systemanforderungen in Kapitel 5 für die jeweilige Problematik definiert. Abschließend wird in Kapitel 6 eine Fazit gezogen und ein Ausblick gegeben.

2 Stand der Technik

Die Arbeit von Schächinger et al. [1] beschäftigt sich eingehend mit der Analyse der Schwachstellen der Rettungskette, wobei die Problematik in die Bereiche der Informations-, Kommunikations- und Dokumentationsprozesse gegliedert wird [12]. Die Ergebnisse der Analyse des Informationsprozesses zeigen Defizit in den Aspekten: aktueller Standort und Status potentiell einsetzbarer Rettungsmittel, geschilderte Meldebild der Rettungsleitstelle, aktuelle Lage an der Einsatzstelle, Kapazitäten in potentiellen Zielkliniken, zu erwartende Notfallpatienten und das Behandlungsergebnis [12].
Aus der Kommunikationsanalyse resultiert, dass die Kommunikation oft unstrukturiert, ausschließlich synchron und verbal verläuft [1]. Dabei verwendet die Kommunikationsübertragung relativ alte Technik mit oft mäßiger Übertragungsqualität [12] . Die Dokumentation ist aufwendig und redundant, wobei Daten, wie Name, Geburtsdatum und Krankenversicherung eines Notfallpatienten mehrmals erhoben werden [12]. Trotzdem sind diese oft unvollständig, häufig ungenau und schwer auswertbar [12]. Die vorgeschlagene Lösung ist dabei das NOAH-System [1].
Das Forschungsprojekt Med-on-@ix [13] adressiert die Problematik in der Notfallversorgung durch mangelndes Qualitätsmanagement bei steigendem Kostendruck und zunehmendem Notärztemangel. Dabei wurde das System zur Steigerung der Versorgungsqualität und die Einsatzeffizienz entwickelt. Es stellt ein umfassendes notfallmedizinisches Telesystem bereit, welches Notärzten und Rettungsdienstpersonal eine zusätzliche Beratung durch ein Kompetenzzentrum ermöglicht. Abbildung 1 zeigt die Vernetzung des Telenotarztes im Kompetenzzentrum und der Einsatzstelle. Wobei spezielle Kompetenzen, wie beispielsweise Giftnotrufzentrale, durch den Telenotarzt gebündelt werden. Dadurch können die aktuell notwendigen Informationen an die Einsatzstellen übermittelt werden.
Die Arbeiten von Gillis et al. [14], [15] befassen sich mit der erforderlichen Kommunikationsinfrastruktur und der fehlenden Kollaborationstechnologien während eines MANVs. Dabei wurde das Problem von fehlgeleiteten und verzögerten Patientensichtungen bei kritisch Verletzen im Feld entdeckt. Dies resultiert aus einer großen Anzahl von Patienten mit unterschiedlichen Verletzungszuständen, die von einem begrenzten medizinischen Personal behandelt werden muss.
Die Autoren schlagen das “Panacea Glass”-System vor, dass neue Möglichkeiten für das Sichtungspersonal bereitstellt. Dadurch soll eine Koordination mit anderen medizinischen Mitarbeitern auch in einem Katastrophenszenario und mit situationsbezogener Patientenversorgung verfügbar sein. Abbildung 2 zeigt den Systemüberblick, das AR in seinem Umfeld verwendet.
Darauf aufbauend befasst sich Demir et al. [16] mit eine intelligente Benutzerschnittstelle, die eine einfache Verwendung und Informationsanalyse bereitstellen soll. Dazu wurde ein iterative Entwicklung vorgeschlagen, die auf einer erfahrungsorientierten Methodik der Benutzern basiert. Es soll die Koordination zwischen EL und Rettungspersonal unterstützen. Dabei konnte gezeigt werden, dass insbesondere AR und der Einsatz von Smart Glasses, zukunftsweisend für intelligente Anwendungen im Gesundheitswesen mit effektiven und effizienten Kommunikationsmöglichkeiten sein kann. Zu den allgemeinen Designempfehlungen schlagen die Autoren die Integration von Funktionen mit Situationsbewusstsein, synchronisierter Kartenansicht, Interaktion ohne den Einsatz der Hände durch AR- und Smart-Brille vor.
Anhand der vorgestellten Arbeiten wird eine ähnliche Schwachstellenanalyse wie in [1], [13] durchgeführt, die einen Einsatz von AR im Rettungsdienst identifizieren soll. Wobei anstelle von Smart Glasses die Microsoft HoloLens verwendet wird, mit der neuartige Interaktionsmechanismen und Situationsbewusstsein umgesetzt werden können, die als Designvorschläge in [14]–[16] genannt wurden.

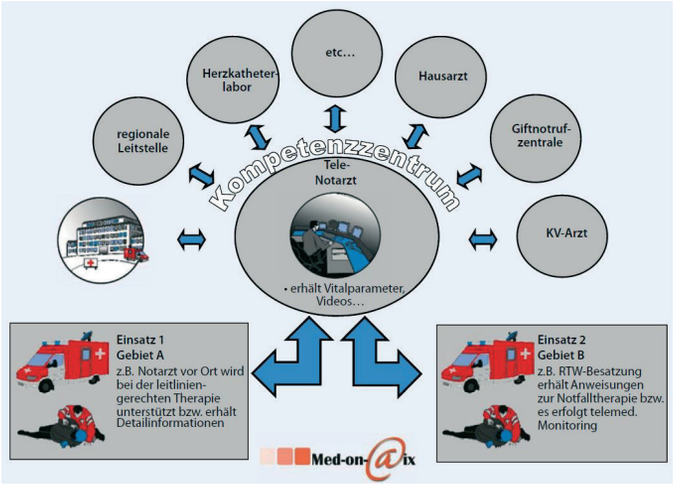


Abbildung 1: ***Ansatz des Med-on-@ix-Systems [13]***

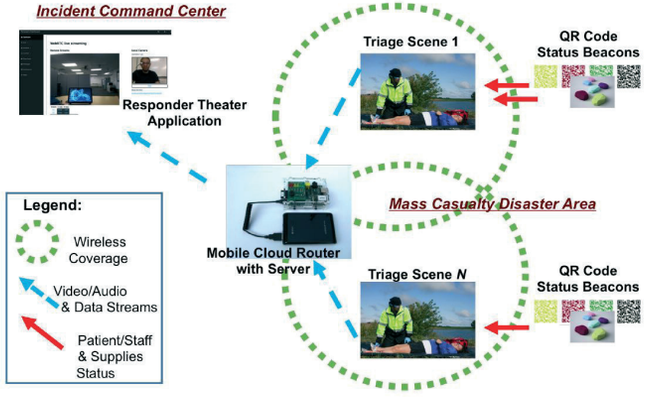


Abbildung 2: ***Übersicht des Panacea Glass Systems [14]***

3 Methodik

Die verwendete Methodik orientiert sich an einem benutzerorientierten Ansatz, dem menschzentrierten Designprozess nach ISO 9241-210 [17], der in Abbildung 3 dargestellt ist. Die Kernelement dieses Ansatzes sind die Darstellung des Nutzungskontext, die Aufstellung der resultierend Anforderungen an das System, sowie die Systementwicklung und die abschließende Evaluierung und Anpassung der Anforderungen oder des Systems. Hierbei betrachtet diese Arbeit nur das Verstehen und Beschreiben des Nutzungskontextes und die Spezifizierung der Nutzeranforderungen.

Damit das Potential durch den Einsatz von AR im Rettungsdienst erfasst werden kann, müssen im ersten Schritt Problemstellungen

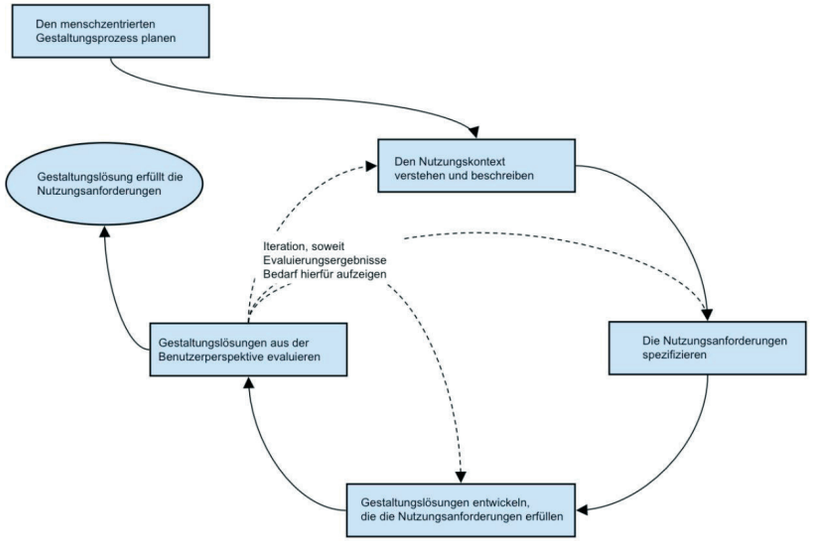


Abbildung 3: ***Schematische Darstellung des meschzentrierten Designprozess nach ISO 9241-210 [17]***

im bisherigen Prozess oder bei der Wahl der Hilfsmittel untersucht werden. Hierzu wurde ein Fragenkatalog für Benutzerinterviews erstellt, der sich an dem Ansatz der qualitativer Marktforschung [18] orientiert. Der Fragenkatalog besteht aus vier Teilen, die sich in folgende Aspekte untergliedern: Fragen zum Rettungsdienst, die sich auf die Tätigkeiten während eines Einsatzes beziehen und deren Schulung. Fragen zum Rettungsdienst, die einen Einsatz von Hilfsmitteln, wie beispielsweise Checklisten oder Nachschlagewerke, unterstützen oder verbessern würden. Fragen zu Problemstellungen während der Kommunikation, Dokumentation und Patientenbehandlung innerhalb eines Einsatzes. Fragen zur Akzeptanz von AR-Brillen während eines Einsatzes.

Die Fragen wurden offen gestellt, damit keine Beschränkung der Antwort, durch Annahmen während der Erstellung des Fragenkatalogs, künstlich erzeugt wurde. Für die Interview-Runde wurden vier Teilnehmer veranschlagt, die die unterschiedlichen Berufsgruppen im Rettungsdienst widerspiegeln und ein Verständnis für den Nutzungskontext schaffen sollen. Die Teilnehmer waren jeweils ein Notarzt, Notfallsanitäter, Rettungssanitäter und eine Person aus der Einsatzleitung. Dabei konnten bis zum Zeitpunkt der Ausarbeitung jedoch nur drei durchgeführt werden, wobei keine Person aus der Einsatzleitung teilnahm.

4 Ergebnis

Die Interviews ergaben bei der Befragung nach der Tätigkeit während des Einsatzes und deren Schulung, dass die Berufsgruppen unterschiedliche Wissensstände besitzen und daher auch verschiedene Befugnisse zur Behandlung der Patienten besitzen. So ist beispielsweise der Rettungssanitäter nicht ausgebildet und rechtlich befugt, einen invasiven Eingriff am Patienten durchführen kann. Wodurch ein Notarzt oder die Anweisung durch

einen Notarzt notwendig ist, um diese Behandlungsmaßnahme durchzuführen.
Zu der Frage nach einem sinnvollen Einsatz zusätzlicher Hilfsmittel wie Checklisten oder Nachschlagewerken wurde diese größten Teils verneint, weil der Informationsaustausch und die Erfahrung innerhalb des Rettungsteams diesen in fast allen Fällen als unnötig erscheinen lässt. Dabei wurde die Möglichkeit zur Auflistung eines Behandlungsablaufs als mögliche Unterstützung aufgezählt, die in zeitkritischen Situation aus Benutzersicht jedoch unpraktikabel wäre. Die einzige Anwendung bei der Informationen nachgeschlagen werden, ist aus Sicht des Notarztes im Falle einer Intoxikation.
Auf die Fragestellung welche Probleme bei der Dokumentation auftreten, wurde bei zeitkritischen Behandlung die Dokumentation nach dem Eingriff genannt, wobei gemessene Werte oder verabreichte Medikamente notiert werden und anschließend in das digitale System eingepflegt werden müssen. Dabei kann es in hektischen Phasen der Behandlung zu Informationsverlust kommen. Besonders gravierend ist das Problem während eines MANVs. Eine weitere große Schwachstelle ist die Kommunikation innerhalb unterschiedlicher Berufsgruppen während eines Einsatzes, wobei die rein verbale Kommunikation zwischen der Einsatzleitung und dem Personal vor Ort eine große Hürde darstellt. Bei einem MANV ist dieses Defizit extrem, weil eine personelle wie auch sachliche Ressourcenknappheit herrscht und diese bei schlechtem Informationsaustausch nicht effizient eingesetzt werden können. Zu der Frage nach Problemstellungen bei der Patientenbehandlung, wurde der Notarztmangel sowie die numerisch fehlenden Rettungsmittel genannt.
Die Akzeptanz von AR-Brillen während eines Einsatzes, wurde bisher als eher gering bewertet. Ein Hindernis soll der vertrauensvolle Bezug zum Patienten sein, der im Laufe der Behandlung aufgebaut werden muss, um Symptome zu identifizieren oder den Patienten zu beruhigen. Dabei wird erwartet, dass die visuell noch untypische Hardware abschreckend wirken könnte. Ein weiterer Aspekt ist der Tragekomfort der Brille, wobei diese das Sichtfeld nicht einschränken darf und auch nicht vom Kopf fallen darf. Zusätzlich wird die Gesten- und Sprachsteuerung als nicht vorteilhaft betrachtet.
Anhand der Auswertung der Interviews konnten drei Problemstellungen identifiziert werden, die in allen Benutzerbefragungen als verbesserungswürdig oder suboptimal beschrieben wurden. Dabei handelt es sich um die Schwachstellen Dokumentation und Kommunikation innerhalb des Notfallmanagements während eines MANVs und dem Notarztmangel bei der Patientenbehandlung innerhalb des Qualitätsmanagements. Der Einsatz von AR soll hierbei helfen, trotz der bisher fehlenden Akzeptanz von AR-Brillen, weil ein immenses Potential zur Problemlösung besteht.

5 Anforderungskatalog

Auf Basis der ausgewählten Schwachstellen wurde jeweils ein Anforderungskatalog für eine mögliche Anwendung aufgestellt. Dabei lassen sich die Anforderungen in funktional und nicht funktional einteilen. Zusätzlich wird anhand der Anwendungstypen Notfall- und Qualitätsmanagement klassifiziert. Tabelle 1 listet die nicht funktionalen und funktionalen Anforderung auf, die alle Anwendungstypen erfüllen müssen. Die dabei aufgeführten Punkte beschrieben die Rahmenbedingungen der Anwendung.
Die Anforderungen N01 bezieht sich auf die Kapselung der Funktionen in geschlossene Bausteine, die eigenständig funktionieren. Hierdurch soll gewährleistet werden, dass identischen Funktionalitäten in unterschiedliche Anwendungen eingesetzt werden können, um so mehrere Anwendungen schneller zu realisieren. N02 beschriebt den Datenaustausch zum Kompetenzzentrum, der über

Tabelle 1: Nicht funktionale (N) und funktionale (F) Systemanforderungen

N01	Modulare Softwarearchitektur
N02	Datenaustausch über WLAN
F01	Intuitive Interaktion mit Feedback
F02	Simple Informationsdarstellung
F03	Kommunikationsaufbau ohne das Voraussetzten vorhandener und intakter Infrastruktur

WLAN erfolgen soll und nicht durch Funk oder sonstige Übertragungsmethoden. Die Anforderungen F01 und F02 beziehen sich auf eine einfache Benutzung, damit die Prozesse nicht aufwendiger werden und dadurch den möglichen Mehrwert eliminieren. F03 definiert den Einsatz der Anwendung in einer unvorhersehbaren und eventuell beschädigten oder überlasteten Kommunikationsinfrastruktur, wodurch die zu entwickelnde Anwendung autark funktionieren muss, um überhaupt eingesetzt werden zu können. Um nicht in Konflikt mit der Anforderung N02 zu stehen, wird davon ausgegangen, dass im Einsatz ein lokales WLAN-Netz aufgespannt ist, welches nicht von der vorhanden Infrastruktur abhängt.
Zusätzlich wurden Anforderungen an die zu verwendende AR-Brille gestellt, die in Tabelle 2 aufgelistet sind.

Tabelle 2: Nicht funktionale (NB) und funktionale (FB) Anforderungen and die AR-Brille

NB01	Portabel
FB01	Kommunikation über WLAN
FB02	Gesten- und Sprachsteuerung
FB04	3D Visualisierung und Interaktion

Die Anforderung NB01 bezieht sich auf die Verwendung der Brille im Einsatz, wobei diese keine zusätzliche Behinderung darstellen darf. Besonders wichtig ist die Größe, das Gewicht und die notwendige Hardware zum Betrieb. FB01 ist durch die allgemeine Anforderung N02 bedingt. Die notwendigen Interaktionsmechanismen werden durch FB02 beschrieben, die es erlauben sollen die Anwendung ohne den Einsatz von Controllern zu steuern. FB04 beschrieb die räumliche Visualisierung und Interaktion, wodurch Markierungen im Sichtfeld des Benutzers dargestellt werden sollen und so eine besser Unterstützung des Einsatzpersonals zu erzielen.

5.1 Notfallmanagement

Das Notfallmanagement wird in die zwei Aspekte Dokumentation und Kommunikation eingeteilt, die eigenständige Module darstellen und dadurch separierte Anforderungen besitzen.

- **Dokumentation**

Im Rahmen der Dokumentation geht es um die Beschleunigung der Aufzeichnung sämtlicher patientenbezogener Informationen, wobei eine möglichst große Automation von Eingabetätigkeiten helfen soll. Tabelle 3 zeigt die Anforderungen an eine solche Anwendung.

Tabelle 3: Nicht funktionale (ND) und funktionale (FD) Anforderungen für die Anwendung zur Dokumentation

ND01	Anbindung eines Chip-Karten-Lesegeräts an die AR-Brille
FD01	Sprache-zu-Text-Konvertierung
FD02	QR-Code-Erkennung
FD03	Darstellung der Patientendaten
FD04	Gesichtserkennung
FD05	Vitalwertmessung
FD06	Sichere Kommunikation

Die Anforderung ND01 dient dazu, dass demografische Daten des Patienten über die Krankenkassen-Chip-Karte bezogen werden können, wobei eine sichere Übertragung gewährleistet werden muss. Die Funktion FD01

bezieht sich auf eine einfachere Eingabe der Tätigkeiten, indem diese über auditive Befehle eingegeben und durch das System transkribiert und richtig eingetragen wird. Dadurch könnte der Verlust von Informationen und Zeit, die während eines MANVs sehr wichtig sind, verringert werden. FD02 bezieht sich auf die Sichtung mithilfe von QR-Code-Tagging. Über den QR-Code sind bisher aufgenommene Daten erreichbar, wodurch das System diesen QR-Code ebenfalls erkennen soll, damit keine redundanten Daten gesammelt werden. Im Zusammenhang von ND01-FD02 ist durch FD03 die benutzerorientierte Visualisierung der Daten beschrieben. FD04 und FD05 beschreiben automatisierbare Vorgänge zur Beschleunigung der Dokumentation, indem zum Beispiel die Gesichtserkennung genutzt wird um die Patientendaten mit einem Bild zu versehen und so eine Verwechslung bei der Behandlung zu verringern. Die automatische Vitalwertmessung anhand von Bilddaten kann ein anbringen von Sensorik am Patienten ersetzten und die Behandlung möglicher Weise beschleunigen. FD06 bezieht sich auf eine sichere Übertragung dieser sensiblen Patientendaten.

- **Kommunikation**

Im Rahmen der Kommunikation geht es um die Verbesserung des Situationsverständisses während der Kommunikation. Dabei soll durch Videoübertragung und hilfestellende Interaktionsmöglichkeiten die effiziente Organisation der Ressourcen ermöglicht werden, sodass eine bessere Entscheidungsgrundlage und damit ein besseres Resultat bei Krisenbewältigung erzeugen soll. Tabelle 4 zeigt die Anforderungen an eine solche Anwendung.
Die Anforderung NK01 bezieht sich auf die Einbindung von GPS-Daten in das System, wodurch das Rettungspersonal lokalisiert werden kann, wodurch ein zielgerichtete Patientensicht im Unfallareal möglich ist. NK02

Tabelle 4: Nicht funktionale (NK) und funktionale (FK) Anforderungen an die Kommunikationsanwendung

NK01	Anbindung eines GPS-Moduls an die AR-Brille
NK02	Zwei getrennte Anwendungen
FK01	Übermittlung des Kamerabildes an die entfernte Anwendung
FK02	Übermittlung des Tons an die entfernte Anwendung
FK03	Übermittlung der Georeferenz an die entfernte Anwendung
FK04	Visualisierung entfernter Eingaben in der AR-Brille
FK05	Übergabe von Georeferenzen an die AR-Brille
FK06	Verbindung mehrere AR-Brillen-Nutzer mit der entfernten Anwendung

bezieht sich auf die Zweiteilung der Anwendung in die Sicht des Rettungspersonals (AR-Brille) und des Einsatzleiters (entfernte Anwendung). Dabei werden Kamerabild, Ton und Georeferenz an die Einsatzleitung übermittelt, dass durch FK01-FK03 definiert wird. Daraufhin kann die Einsatzleitung hilfestellende Maßnahmen in der Sicht des Rettungspersonals visualisieren (FK04, FK05). FK06 bezieht sich auf die Verbindungen mehrerer AR-Brillen-Träger mit der Einsatzstelle. Dabei liefern unterschiedliche Sichten eine bessere Entscheidungsbasis für die Planung, die genutzt werden soll um einen schnelleren Ausgleich des Ressourcendefizits zu Beginn des MANVs zu erwirken.

5.2 Qualitätsmanagement

Die Anwendung zur Behandlung des Qualitätsmanagementproblems bezieht sich auf den Notarztmangel, dem durch ein Kompetenzzentrum ähnlich zu [13] und der Verwendung von AR entgegengewirkt werden soll.

Dabei ermöglicht der Einsatz von AR-Brillen neue Möglichkeiten zur Darstellung der Hilfestellung und bietet somit Verbesserungspotential bei der Behandlung. Tabelle 5 zeigt die Anforderungen an eine solche Anwendung.

Tabelle 5: Nicht funktionale (NT) und funktionale (FT) Anforderungen an die Telenotarztanwendung

NT01	Zwei getrennte Anwendungen
FT01	Übermittlung des Kamerabildes an die entfernte Anwendung
FT02	Übermittlung des Tons an die entfernte Anwendung
FT03	Visualisierung entfernter Eingaben in der AR-Brille
FT04	Entfernte Markierung in der AR-Brille
FT05	Entferntes Abspielen von Videos in der AR-Brille
FT06	Entfernte Einblendung von Zusatzinformationen in der AR-Brille

Ebenso wie NK02 bezieht sich NT01 auf die Entwicklung eines verteilten Systems, wobei das Rettungspersonal die Ar-Brille als Interaktionsmedium benutzt und das Personal im Kompetenzzentrum durch eine entfernte Anwendung kommuniziert. Hierbei werden Kamerabild und Ton an das Kompetenzzentrum übermittelt, dass durch FT01 und FT02 definiert wird. Daraufhin kann das Kompetenzzentrum hilfestellende Maßnahmen in der Sicht des Rettungspersonals visualisieren (FT03- FT06).

6 Fazit & Ausblick

Im Rahmen dieser Arbeit wurden sowohl die Problemstellungen des Notfall- und Qualitätsmanagement anhand der Literatur, wie auch aus der Befragung des Rettungspersonal identifiziert. Hierbei konnten Defizite bei der Dokumentation und Kommunikation während eines MANVs als besonders schwerwiegend bestimmt werden. Bei der Patientenbehandlung wurde der Notarztmangel und der daraus resultierenden Qualitätsverlust als besonders kritisch eingestuft. Für diese Schwachstellen wurde jeweils ein Anforderungskatalog erstellt, der bei der Entwicklung einer Systemlösung unter Verwendung von AR genutzt werden kann. Für eine weiterführende Arbeit wird die Optimierung der Kommunikation während eines MANVs und die Verbesserung der Patientenbehandlung durch einen Telenotarzt betrachtet, weil sich die Systemanforderungen sowohl in funktionalen wie auch nicht funktionalen Anforderungen stark überschneiden. Es wird empfohlen diese Aspekte als erstes zu betrachtet, weil davon ausgegangen werden kann, dass eine Umsetzung einer Anwendung mit leichter Anpassung auf die zweite Anwendung überführt werden kann. Daher ist eine modulare Softwarearchitektur besonders wichtig für die einfache Erweiterung der Applikation.

Literatur

[1] U. Schächinger, S. P. Stieglitz, R. Kretschmer und M. Nerlich, “Telemedizin und Telematik in der Notfallmedizin”, *Notfall & Rettungsmedizin*, Jg. 2, Nr. 8, S. 468–477, 1999, ISSN: 1434-6222.

[2] M. Czaplik und S. Bergrath, “Telemedizin in der Notfallmedizin”, in *eHealth in Deutschland: Anforderungen und Potenziale innovativer Versorgungsstrukturen*, F. Fischer und A. Krämer, Hrsg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, S. 319–333, ISBN: 978-3-662-49504-9.

[3] S. Lackey, R. Shumaker, G. Kurillo, A. Yang, V. Shia, A. Bair und R. Bajcsy, Hrsg., *New Emergency Medicine Paradigm via Augmented Telemedicine: Virtual, Augmented and Mixed Reality*, Springer International Publishing, 2016, ISBN: 978-3-319-39907-2.

[4] A. Gries, M. Helm und E. Martin, "Zukunft der präklinischen Notfallmedizin in Deutschland", *Der Anaesthesist*, Jg. 52, Nr. 8, S. 718–724, 2003.

[5] T. Kopetsch, "Bundesrztekammer-Statistik: Ärztemangel trotz Zuwachsraten", *Dtsch Arztebl International*, Jg. 103, Nr. 10, A–588–, 2006.

[6] W. Martin, "Arbeitsmarkt für Ärztinnen und Ärzte: Der Ärztemangel nimmt weiter zu", *Dtsch Ärztebl International*, Jg. 105, Nr. 16, A–853–, 2008.

[7] D. Stratmann, P. Sefrin, S. Wirtz, A. Bartsch und T. Rosolski, "Stellungnahme zu aktuellen Problemen des Notarztdienstes (Ärztemangel, Arbeitszeitgesetz, DRG)", *Der Notarzt*, Jg. 20, Nr. 03, S. 90–93, 2004, ISSN: 0177-2309. DOI: 10.1055/s-2003-814943.

[8] B. Reimann, B. Maier, R. Lott und F. M. Konrad, "Gefährdung der Notarztversorgung im ländlichen Gebiet", *Notfall & Rettungsmedizin*, Jg. 7, S. 200–204, 2004, ISSN: 1434-6222.

[9] M. Vassell, O. Apperson, P. Calyam, J. Gillis und S. Ahmad, "Intelligent Dashboard for augmented reality based incident command response coordination", in *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Piscataway, NJ: IEEE, 2016, S. 976–979, ISBN: 978-1-4673-9292-1.

[10] E. Kyriacou, I. Nicolaidou, G. Hadjichristofi, S. Kyprianou, S. Christou und R. Constantinou, "Health and rescue services management system during a crisis event", *Healthcare technology letters*, Jg. 3, Nr. 3, S. 205–211, 2016, ISSN: 2053-3713.

[11] Y.-J. Zheng, Q.-Z. Chen, H.-F. Ling und J.-Y. Xue, "Rescue Wings: Mobile Computing and Active Services Support for Disaster Rescue", *IEEE Transactions on Services Computing*, Jg. 9, Nr. 4, S. 594–607, 2016, ISSN: 1939-1374.

[12] R. Maier und W. Röckelein, *An inter-organisational system to support emergency care process chains: The NOAH project*. Lehrstuhl für Wirtschaftsinformatik, 1999.

[13] M. Skorning, S. Bergrath, D. Rörtgen, J. C. Brokmann, S. K. Beckers, M. Protogerakis, T. Brodziak und R. Rossaint, "„E-Health" in der Notfallmedizin – das Forschungsprojekt Med-on-@ix", *Der Anaesthesist*, Jg. 58, Nr. 3, S. 285, 2009. DOI: 10.1007/s00101-008-1502-z.

[14] J. Gillis, P. Calyam, O. Apperson und S. Ahmad, "Panacea's Cloud: Augmented reality for mass casualty disaster incident triage and co-ordination", in *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Piscataway, NJ: IEEE, 2016, S. 264–265, ISBN: 978-1-4673-9292-1.

[15] J. Gillis, P. Calyam, A. Bartels, M. Popescu, S. Barnes, J. Doty, D. Higbee und S. Ahmad, "Panacea's Glass: Mobile Cloud Framework for Communication in Mass Casualty Disaster Triage", in *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, S. Hiroyuki, D. Huang und A. Küpper, Hrsg., Piscataway, NJ und Piscataway, NJ: IEEE, 2015, S. 128–134, ISBN: 978-1-4799-8977-5.

[16] F. Demir, S. Ahmad, P. Calyam, D. Jiang, R. Huang und I. Jahnke, "A Next-generation Augmented Reality Platform for Mass Casualty Incidents (MCI)", *J. Usability Studies*, Jg. 12, Nr. 4, S. 193–214, 2017, ISSN: 1931-3357.

[17] *Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010): Deutsche Fassung FprEN ISO 9241-210:2010 ; Human-centred design for interactive systems (ISO 9241-210:2010);* Berlin.

[18] R. Buber und H. H. Holzmüller, *Qualitative Marktforschung*. Wiesbaden: Gabler, 2007, ISBN: 978-3-8349-0229-0.

Vergleich von Plattformen zur Umsetzung einer V2C Kommunikation auf Basis von Automotive Grade Linux

Christian Pitters
Hochschule Reutlingen
Christian.Pitters@Student.Reutlingen-University.DE

Abstract

Smartphones und das Internet haben eine neue Art der Konnektivität erschaffen, die alle Lebensbereiche der Menschen erreicht. Die Automobilindustrie ist von dieser digitalen Revolution allerdings sehr lange unberührt geblieben. Angetrieben durch das Potenzial, welches durch den Einsatz von IoT und Cloud Computing entsteht, erfährt die Vernetzung in der Automobilindustrie momentan allerdings einen großen Aufschwung mit der neuen Vision, das Fahren mehr zu einem individuellen sicheren Erlebnis zu gestalten. Diese Arbeit analysiert die Vorteile und Möglichkeiten einer solchen Vernetzung und vergleicht verschiedene Plattformen zur Realisierung einer Vehicle to Cloud (V2C) Kommunikation auf Basis von Automotive Grade Linux.

Schlüsselwörter

Vehicle to Cloud, Connected Car, Eclipse Kuksa, AWS IoT, Automotive Grade Linux

Betreuer Hochschule: Prof. Dr. Martínez
Hochschule Reutlingen
Natividad.Martinez@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Christian Pitters

CCS Concepts

• **Computer systems organization** → Cloud computing; • **General and reference** → Surveys and overviews

1 Einleitung

Wir sind in unserem täglichen Leben zunehmend auf Internetdienste angewiesen, sodass es für die meisten nicht mehr vorstellbar ist, auch nur für kurze Zeit komplett offline zu sein. Sobald der Mensch allerdings in das Auto einsteigt, endet die gewohnte Vernetzung, die der „smarte" Benutzer aufgrund der mobilen Geräte normalerweise erfährt [1]. Aus diesem Grund ist es wichtig, die gewohnte Umgebung und das Angebot der gewohnten Dienste auch im Auto zur Verfügung zu stellen. Durch das Internet der Dinge und die damit verbundene Vernetzung verschiedenster Objekte und Sensoren, erhält dieses Thema immer mehr Einzug in das moderne Automobil, sodass dieses zu einem IoT-Gerät (Internet of Things) heranwächst. Ein solches Auto, welches mit seiner Umwelt kommuniziert, wird auch als *Connected Car* bezeichnet. Durch die Möglichkeit zur Kommunikation kann es sich genau an die Bedürfnisse und Stimmungen des Fahrers anpassen.

1.1 Problem & Zielsetzung

Die Vernetzung im Automobil erfährt momentan einen großen Aufschwung. In den letzten Jahren wurde aus diesem Grund vermehrt in die Vernetzung des Automobils investiert. Allerdings ist das Potential aufgrund

vieler Hindernisse, wie beispielsweise fehlender Standards, noch nicht voll ausgereizt. [2] Ziel dieser Arbeit ist deshalb die Einführung in das Thema *Connected Car* sowie die Untersuchung der Bedeutung einer Vehicle to Cloud (V2C) Kommunikation. Zudem sollen verschiedene Möglichkeiten zur Realisierung einer Connected Car Plattform recherchiert werden, um diese miteinander zu vergleichen. Ziel dieses Vergleichs ist es, die Vor- und Nachteile, als auch den Leistungsumfang zu identifizieren, um eine geeignete Plattform für die Realisierung einer V2C Kommunikation am Fahrsimulator der Hochschule Reutlingen zu finden. Als Betriebssystem des In-Vehicle Infotainment Systems (IVIS) soll Automotive Grade Linux (AGL) zum Einsatz kommen.

1.2 Struktur der Arbeit

In dieser Arbeit wird zunächst das *Connected Car* näher betrachtet. Anschließend, in Abschnitt 3 und 4, werden die Vorteile der Cloud sowie die dadurch entstehenden Möglichkeiten im *Connected Car* Szenario analysiert. In Abschnitt 5 wird der aktuelle Stand der Technik dargestellt. Abschnitt 6 gibt eine kleine Einführung in das Open Source Betriebssystem AGL, welches die Basis der In-Vehicle Komponente darstellt. Anschließend wird in Abschnitt 7 ein Kriterienkatalog erstellt, um die, in Abschnitt 8 vorgestellten Plattformen, vergleichen zu können. Der Vergleich und eine darauf aufbauende Diskussion werden in Abschnitt 9 und 10 dargelegt. Der letzte Abschnitt, beinhaltet eine Reflexion der gesamten Arbeit und fasst die wichtigsten Punkte nochmals zusammen.

2 Connected Car

Das Fahrzeug der Zukunft hat das Ziel, die Zeit im Auto effektiver, individueller und komfortabler zu gestalten. Außerdem soll durch die Vernetzung die Sicherheit gesteigert sowie die Zeit beim Reisen reduziert werden. [3]

Dabei gibt es mehrere Arten der Kommunikation, um Autos miteinander zu vernetzen, welche in einem vollständig vernetzten Auto alle miteinander kombiniert werden müssen.

2.1 Kommunikation

Neben der in diesem Artikel betrachteten V2C Kommunikation, wird das Auto in Zukunft auch mit vielen anderen Objekten kommunizieren, was auch als Car to Everything (C2X) Kommunikation bezeichnet wird. Dabei ist es wichtig auch mit den Autos in der Umgebung (C2C) und mit der vorhandenen Infrastruktur (C2I) zu kommunizieren. Dies ist unter anderem für das autonome Fahren von großer Bedeutung, spielt jedoch für den Bereich Infotainment und Individualisierung nur eine untergeordnete Rolle. Welcher Service durch welche Kommunikationsart umgesetzt wird, hängt stark von den Anforderungen des Services ab [3]. Es bedarf einer Unterscheidung, ob ein Service geringe Latenzen braucht, oder ob diese weniger wichtig sind und der Fokus auf die Informationen gelegt wird. Jeder Service hat unterschiedliche Anforderungen egal ob C2C, C2I oder V2C [3].

2.2 Interaktion

Durch das vermehrte Angebot an Connected Services wird die Interaktion mit dem IVIS zunehmend komplexer und die verfügbaren Funktionen dadurch weniger verständlich und intuitiv [4]. Dies widerspricht allerdings dem Ziel, die Ablenkung des Fahrers zu vermeiden [5]. Aus diesem Grund gelten für die Infotaiment Systeme des Autos spezielle Anforderungen an die Benutzerschnittstelle (UI) [6]. Um einer Ablenkung entgegen zu wirken müssen neue UI-Konzepte sowie Interaktionsmöglichkeiten entwickelt werden.

Das Ziel muss sein, die Interaktion zwischen dem Benutzer und dem System auf ein Minimum zu reduzieren. Um dies zu realisieren, können sogenannte Intention-aware-Systeme eingesetzt werden [4]. Diese versuchen das Verhalten des Benutzers vorherzusagen, um auf Basis dieser Vorhersagen das UI dymisch anzupassen oder Applikationen automatisch zu starten [4]. Beispielsweise kann das Auto aufgrund eines eingetragenen Termins im Terminkalender des Fahrers das Navigationsziel automatisch vorschlagen, sodass dieses nur noch bestätigt werden muss. Diese Vorhersagen der Benutzerabsichten gepaart mit einer guten Mensch-Maschine Interaktion kann die Benutzerfreundlichkeit enorm verbessern und so das Maß der Ablenkung reduzieren [4]. Die Steuerung durch Sprachbefehle, basierend auf der Integration bekannter Sprachassistenten, erfährt dabei momentan einen großen Aufschwung.

3 Vorteile der Cloud

Die Menge an Daten, die von den Autos gesammelt werden steigt stetig an, sodass mehrere Gigabyte an Daten pro Stunde anfallen. Um diese zu verwalten, ist der Einsatz von Cloud Computing zwingend notwendig. Es gilt diese Daten zu versenden, zu verarbeiten und zu analysieren, um daraus wertvolle Informationen zu generieren. [2]

Die Cloud ist zudem für die Realisierung von IoT Anwendungen hervorragend geeignet [7]. Da auch das Auto immer mehr zu einem IoT-Gerät heranwächst, ist es naheliegend nun auch im Automobilen-Bereich auf die Cloud zurückzugreifen [7]. Dabei profitieren sowohl die Automobilhersteller, als auch die Verbraucher, durch die dadurch ermöglichten vernetzen Dienste. Zudem ist das Auto dadurch nicht mehr von den verfügbaren lokalen Ressourcen eingeschränkt, da die Cloud theoretisch unbegrenzt Rechenleistung und Speicherkapazität bereitstellt [8].

Ein weiterer Vorteil, den die Cloud mit sich bringt, ist die potentielle Datenmenge. Es gilt: Je mehr Daten der Cloud zur Verfügung stehen, desto mehr Möglichkeiten haben die Automobilhersteller, um die Effizienz, die Sicherheit und das Infotaiment zu verbessern [9]. Deshalb ist es wichtig eine hohe Knotendichte zu erreichen. Dafür müssen allerdings die Daten herstellerübergreifend nutzbar werden [9]. Durch eine erfolgreiche V2C Kommunikation können somit verschiedene Arten von Services realisiert werden.

4 Möglichkeiten durch V2C

Die Vernetzung des Automobils mit der Cloud bietet eine Vielfalt neuer Möglichkeiten für die Automobilhersteller, um den Kunden neue Dienste anzubieten. Im Folgenden werden einige dieser Möglichkeiten vorgestellt.

4.1 Smart Home Integration

Durch den Einsatz von Cloud- und IoT-Lösungen wird es möglich das eigene smarte Zuhause mit dem Auto zu vernetzen [2]. Durch eine Smart Home Integration kann der Fahrer mithilfe des integrierten IVIS oder dem integrierten Sprachassistenten während der Fahrt bereits die Heizung einschalten oder das Garagentor öffnen [2]. Möglich wäre auch eine Automatisierung dieser Prozesse auf Basis der aktuellen Position.

4.2 Standortbasierte Dienste

Standortbasierte Dienste haben das Ziel, abhängig von der aktuellen Position selektive Daten bereitzustellen [2]. Die GPS Daten werden dabei permanent an die Cloud übertragen, um diese zu analysieren. Durch die Aggregation von Zusatzinformationen, wie beispielsweise den Vorlieben des Fahrers oder der aktuellen Uhrzeit, kann ein großer Mehrwert für den Fahrer generiert werden. beispielsweise ist das Auto dadurch in der Lage, dem Fahrer zur Mittagszeit gut bewertete Restaurants in der näheren Umgebung vorzuschlagen oder während einer Reise Informationen zu nahegelegenen Sehenswürdigkeiten auf dem IVIS darzustellen. Allerdings ist es wichtig, den Fahrer nicht mit Informationen zu überfluten.

4.3 Over the Air Updates

Ein Auto hat im Vergleich zum Smartphone eine viel längere Lebensdauer [6]. Aus diesem Grund erscheinen die IVI Systeme in etwas älteren Autos sehr veraltet [6]. Um dies zu umgehen und um die Sicherheit zu gewährleisten, ist es wichtig das System sowie die Dienste und Funktionen, regelmäßig über die Ferne zu aktualisieren. Ein Beispiel für erfolgreiche OTA-Updates bietet Tesla, bei welchem zusätzliche Funktionen, wie eine Einparkautomatik, nach Markteinführung durch ein Update integriert wurde [10].

5 Stand der Technik

Vernetzte Autos sind mittlerweile nicht nur Forschungsgebiet, sondern werden bereits von vielen Automobilherstellern in geringem Umfang angeboten. Durch die aufkommende Relevanz und den fehlenden Standards entsteht momentan allerdings eine undurchsichtige Plattformlandschaft.

Die meisten der neueren Autos in Europa besitzen bereits einen Zugang zum Internet, um den seit 2018 verpflichtenden eCall Service umzusetzen und um die eigenen Connected Car Plattformen zu betreiben [2]. Viele der neueren Mittelklasseautos sowie ältere Autos der Luxusklasse bieten zudem ein integriertes IVIS zur Anzeige von Informationen und zur Interaktion mit dem Entertainment System. Die meisten Hersteller betreiben dabei allerdings ihre eigene Closed-Plattform, um vernetzte Dienste anzubieten [10]. Zudem bieten viele eine Smartphone Integration, durch Apples CarPlay oder Googles Android Auto, um die eigene Plattform um zusätzliche Dienste zu erweitern [11].

Im Bereich der erweiterten Navigation sind von einigen Herstellern bereits viele V2C-Funktionen umgesetzt. BMW liefert beispielsweise zusätzlich zum Kartematerial das aktuelle Wetter, die nächstgelegene Tankstellen und Echtzeit-Verkehrs Informationen.[1] Allerdings werden bislang nur Daten der vernetzen BMWs miteinbezogen.[1]

Nicht alle Hersteller betreiben dafür allerdings ihr eigenes Connected Car-Backend. Viele der Hersteller beziehen die IT-Plattformen von Drittanbietern [10]. So nutzt BMW beispielsweise die Cloud-Plattform von IBM, um die Fahrzeugdaten zu analysieren [10]. Auch ein Fremdbezug von vollständigen Connected Services wird von einigen Herstellern genutzt [10]. BMW und Mercedes nutzen daher für die Anzeige der Prognosen von freien On-Street Parkplätzen den INRIX Dienst „On-Street Parking“ [10].

[1] www.bmw-connecteddrive.de /app/index.html#/portal/store

6 Automotive Grade Linux

Automotive Grade Linux [12] (AGL) ist ein Open Source-Projekt der Linux Foundation, welches 2012 ins Leben gerufen wurde. Durch das Zusammenbringen verschiedenster Stakeholder soll gemeinsam eine offene In-Vehicle Software-Plattform entwickelt werden, welche als De-facto-Industriestandard für das vernetztes Auto dienen soll [12]. Dadurch soll es ermöglicht werden, Anwendungen plattformübergreifend zu implementieren, sodass diese in verschiedene Autos von verschiedenen Herstellern integriert werden können.

Die Besonderheit ist, dass AGL im Gegensatz zu anderen Lösungen versucht die komplette Software im Auto anzusprechen. Deshalb bietet AGL neben dem Infotaiment (IVIS) auch Lösungen für Head-up-Displays, Kombiinstrumente, Telematik, Fahrerassistenzsysteme und autonomes Fahren [12]. Die Idee ist es, eine Unified Code Base (UCB) bereitzustellen, die von Grund auf mit allen Stakeholdern zusammen entwickelt wird [12]. Bei Projektbeginn können Automobilhersteller auf diese UCB aufbauen, welche ca. 70-80% des Gesamtumfangs einnimmt [12] und die anderen 20-30% für ihre individuellen Produktanforderungen, wie beispielsweise einer eigenen Benutzeroberfläche ergänzen. Dadurch soll eine schnelle Bereitstellung gewährleistet werden.

Wie die Untersuchung der aktuellen Situation im Abschnitt 5 ergab, werden die heutigen In-Vehicle Plattformen von jedem Automobilhersteller selbst entwickelt. Dieser Ansatz kann langfristig allerdings nicht zum gewünschten Erfolg führen [2]. Aus diesem Grund ist eine Open Source Lösung, wie die von AGL von großer Bedeutung für die Zukunft. Dies zeigt auch das steigende Interesse der Automobilhersteller, Erstausrüster und IT-Firmen. Nicht umsonst hat AGL bereits über 100 Mitglieder, darunter einige große Autohersteller wie Mercedes-Benz, Toyota und Mazda sowie weitere Big-Player wie Amazon, Intel und Continental [12]. Das AGL das Potential hat ein anerkannter De-facto-Industriestandard zu werden und nicht nur für Forschungs- oder Demozwecke einen Nutzen bietet, zeigt der Einsatz von AGL im neuem Toyota Camry. Auch Mercedes-Benz Vans plant den Einsatz von AGL[2].

7 Vergleichskriterien

Das Auto wird in Zukunft durch die Menge an verbauten Sensoren Unmengen an Daten generieren, sodass die Speicherung und Verarbeitung vor Ort nicht mehr möglich sein wird [2]. Aus diesem Grund ist eine erfolgreiche V2C Anbindung notwendig. Da sich das Auto Aufgrund der Vernetzung immer mehr zu einem IoT-Gerät entwickelt, ist es naheliegend die bekannten und etablierten Anforderungen eines IoT-Szenarios zu adaptieren und für die Umsetzung einer Connected Car Plattform anzupassen. Dafür wurden die in der Tabelle 1 dargestellten Kriterien auf Grundlage verschiedener IoT Plattformen, sowie der in Abschnitt 4 dargestellten Möglichkeiten identifiziert. Die Kriterien wurden dabei in die drei Kategorien *Plattformeigenschaften*, *funktionaler Leistungsumfang* und *Kommunikation* eingeteilt. Diese Kriterien stellen gleichzeitig die Kernanforderungen dar, auf deren Basis anschließend geeignete Plattformen zur Realisierung einer V2C-Kommunikation untersucht werden.

8 Plattformen

Auf Basis der Kernanforderungen haben sich während der Recherche zwei Plattformen herauskristallisiert, welche im Folgendem kurz vorgestellt werden.

8.1 Eclipse Kuksa

Eclipse Kuksa[3] ist ein Open Source Projekt der Eclipse Foundation, mit dem Ziel eine offene Plattform für die Umsetzung eines Car2Cloud Szenarios bereitzustellen. Das Anfang 2017 gestartete öffentlich geförderte europäische ITEA-Forschungsprojekt APPSTACLE, bildet die Grundlage für Eclipse Kuksa. Dabei wird untersucht, wie eine anpassbare, nicht proprietäre und nicht kommerzielle Lösung entwickelt werden kann, um die IoT- und Cloud-Technologie im Fahrzeug zu integrieren. [13]

Was momentan fehlt ist eine offene Plattform, auf welche jeder Automobilherstellerzugreifen kann, um ein V2C-Szenario zu realisieren. Genau hier will Eclipse Kuksa einsteigen und gemeinsam einen etablierten Standard entwickeln, um die individuelle Entwicklung zu stoppen und den Markt für neue externe Anwendungen zu öffnen. Das Kuksa-Ökosystem gliedert sich in die nachfolgend dargestellten drei Bereiche.

8.1.1 In-Vehicle Plattform

Das IVIS ist das zentrale Element im Auto, um dem Fahrer geeignete Services anzubieten und eine Kommunikation nach außen zu gewährleisten. Die Ziele dieser Plattform sind die Installation und der Betrieb von Anwendungen, der Zugriff auf fahrzeuginterne Ressourcen und der Datenaustausch mit dem Cloud-Backend [14]. Kuksa entwickelt dafür allerdings kein eigenes IVIS. In einer umfangreichen State of the Art Analyse[4] wurden verschiedene In-Vehicle Plattformen analysiert und miteinander verglichen, um eine Fahrzeugplattform als Grundlage verwenden zu können. APPSTACLE entschied sich letztendlich für AGL. Dieses nutzt das Yokto Build-System, wodurch AGL um Kuksa spezifische Schichten erweitert werden kann, um die nötigen Tools und Funktionalitäten bereitzustellen.

[2] https://www.automotivelinux.org/announcements/2018/06/19/automotive-grade-linux-powers-new-solutions-for-commercial-and-consumer-vehicles

[3] https://projects.eclipse.org/proposals/eclipse-kuksa

[4] https://itea3.org/project/appstacle.html (Specification of in-car Software Architecture)

Tabelle 1 Kriterienkatalog

Plattformeigenschaften	
Kompatibilität mit AGL	AGL bietet die Möglichkeit der Erweiterung um weitere Schichten, welche zusätzliche Funktionalitäten miteinbringen. Durch eine solche Kompatibilität ist eine vereinfachte Entwicklung möglich.
Open Source	Open Source bietet viele Vorteile. Zum einen sind die Automobilhersteller nicht von anderen Unternehmen abhängig, zum anderen bietet Open Source durch die wachsende Community oft viel mehr Ressourcen.
Kosten, Dokumentation & Support	Eine gute Dokumentation und ein guter Support sind wichtig, um eine reibungslose Entwicklung zu fördern. Die Plattform muss zudem für die Hochschule erschwinglich sein.
Funktionaler Leistungsumfang	
OTA-Updates	Over the Air Updates sind wichtig, um Funktionen und Services aus der Ferne zu aktualisieren, wodurch Sicherheitslücken geschlossen und neue Funktionen angeboten werden können.
App Store	Ein App Store bietet die einfache Möglichkeit für Dritte Anwendungen bereitzustellen, sowie für die Fahrer neue Anwendungen zu installieren.
Echtzeit Datenanalyse	Einige Anwendungsfälle, wie die Erkennung von Anomalien oder ortsbezogene Dienste, erfordern eine Datenanalyse in Echtzeit.
Datenspeicherung	Möglichkeit einer sicheren skalierbaren Speicherung von Fahrzeugdaten
Gerätemanagement	Ein gutes Geräte- bzw. Fahrzeugmanagement ist für die Bereitstellung von Software sowie für das Rechtemanagement erforderlich.
Kommunikation	
Kommunikations-Protokolle	Es ist wichtig eine Art Middleware bereitzustellen, um die Kommunikation von der dahinterliegenden Architektur zu abstrahieren, wodurch eine Unterstützung verschiedener Kommunikationsprotokolle ermöglicht wird.
Bidirektional Kommunikation	Eine bidirektionale Kommunikation ist wichtig, um Telemetriedaten zur Cloud als auch Event Daten zum Auto zu versenden.

8.1.2 Cloud Plattform

Das Cloud-Backend setzt sich aus verschiedenen Projekten des Eclipse-IoT Umfeldes zusammen. Die Kommunikation zwischen dem Auto und dem Cloud-Backend ist der zentrale Punkt einer V2C-Implementierung. Mit dem Projekt Eclipse Hono [15] bietet das Kuksa Ökosystem eine Möglichkeit die Fahrzeuge mit dem Backend zu verknüpfen. Hono arbeitet dabei als eine Art Middleware, welche die *Connected Cars* mit unterschiedlichen offenen Kommunikations-Protokollen ansprechen können [15].

Over the Air Updates sind ein wichtiger Punkt in einem *Connected Car*, um Dienste und Funktionen immer auf dem aktuellsten Stand zu halten [10]. Mithilfe von Eclipse HawkBit [16] ist dies auch mit Kuksa möglich. Dabei können sowohl Funktionen in Form von Apps aktualisiert werden, als auch das Kernsystem [16].

Eclipse Ditto [17] implementiert das Softwaremuster „digital twin“. Dies ermöglicht die Abbildung eines realen Objektes in der virtuellen Welt [13]. Dadurch besteht die Möglichkeit einer nahtlosen Verknüpfung, um beispielsweise den Zustand eines Autos zu kontrollieren und ggf. zu ändern [13].

Die Sicherheit in einer vernetzen Welt und ganz besonders im *Connected Car* Szenario spielt eine wichtige Rolle. Hierbei geht es zum einen um den Datenschutz und nicht zu vergessen, um eine große Maschine die uns mit weit über 100 km/h transportiert. Wird diese nicht abgesichert, kann dies schlimme Folgen nach sich ziehen. Mit Keycloak [18] steht eine Anwendung zur Verfügung, mit welcher durch minimalen Aufwand eine Authentifizierung der Benutzer realisiert werden kann [18]. Eclipse Keti [19] hingegen hat das Ziel eine einfache umfangreiche Zugriffskontrolle für RESTful APIs zu ermöglichen, um den Zugriff von unbefugten zu verhindern [19].

8.1.3 App IDE

Mit Eclipse Che [20] bietet das Kuksa Ökosystem eine IDE, wodurch ein einfaches Entwickeln von Fahrzeugfunktionen, sowie Cloud-Services ermöglicht werden soll. Die IDE ist bereits vorkonfiguriert und beinhaltet alle nötigen APIs und Bibliotheken, welche für die Entwicklung notwendig sind [14]. Außerdem wird ein vereinfachtes Deployment von Fahrzeuganwendungen sowie für Cloud-Services, mithilfe eines vorkonfigurierten Eclipse Che-Stacks realisiert [14].

8.2 AWS AGL

Amazon Web Services (AWS) ermöglicht mithilfe von IoT-Diensten die Realisierung einer serverlosen Connected Car Cloud Plattform, durch den Einsatz von Serverless Computing [21]. Dadurch ist eine Erfassung, Verarbeitung, Speicherung und Analyse von Daten, sowie eine sichere Kommunikation möglich, ohne sich Gedanken über die darunterliegende Infrastruktur machen zu müssen. AWS liefert dafür eine Referenzarchitektur (AWS Connected Vehicle Lösung)[5], sodass die Konzentration auf das Designen und Implementieren der vernetzten Dienste gelegt werden kann. Zudem bietet AWS ein Framework zur Integration von AWS IoT in den AGL Software-Stack[6], sodass eine sichere Kommunikation realisiert werden kann. Im Folgendem werden die wichtigsten AWS Dienste vorgestellt.

8.2.1 AWS IoT

AWS IoT [21] bietet eine IoT Cloud Plattform zur sicheren bidirektionalen Kommunikation zwischen Geräten und der AWS-Cloud sowie Dienste zum Verwalten, Registrieren und Arbeiten mit IoT-Geräten [21]. AWS IoT setzt sich dabei aus mehreren Komponenten zusammen.

Mithilfe des Device Gateways und dem Message Broker kann eine sichere bidirektionale Verbindung aufgebaut werden. Dabei werden die Protokolle http, Websockets und MQTT unterstützt, sodass Nachrichten per Publication/Subscription Modell ausgetauscht werden können. [22] Mithilfe der Rule Engine können die eingehenden Daten basierend auf eigens definierten Regeln gefiltert, transformiert und an bestimmte Prozesse weitergeleitet werden [22]. Dadurch ist es möglich, die Daten mithilfe einer Regel zu überprüfen und diese Daten an die entsprechende Lambda Funktion weiterzuleiten, welche die Daten weiterverarbeitet oder speichert. AWS bietet zudem die Möglichkeit der virtuellen Abbildung eines realen Objekts und implementiert somit die Metapher des digitalen Zwillings [22].

Um die Daten und Dienste ausreichend zu schützen bietet AWS IoT Sicherheits- Identitäts- und Authentifizierungsservices. So wird beispielsweise eine Verschlüsselung an allen Verbindungspunkten garantiert [21]. Außerdem können Geräte authentifiziert und Operationen genehmigt werden, sollten die Geräte die erforderlichen Rechte besitzen.

8.2.2 Greengrass

Mit AWS Greengrass lassen sich AWS Cloud Funktionen offline auf lokaler Hardware ausführen [21]. Dadurch können Daten nah an der Erhebungsquelle verarbeitet werden. Greengrass kann in den AGL-Stack integriert werden[5], um bereits im Auto Daten offline zu verarbeiten und um eine sichere Verbindung mit AWS IoT über MQTT herzustellen. Bei bestehender Internetverbindung werden die Daten wieder mit der Cloud synchronisiert [21].

9 Vergleich

Wie in der Tabelle 2 zu sehen, ist Eclipse Kuksa ein Open Source Projekt und unter der EPL 2.0 lizensiert. Dadurch wird eine erweiterbare offene Entwicklung ermöglicht. Dies bietet Automobilherstellern und Stakeholdern die Möglichkeit einer gemeinsamen Entwicklung, um einen offenen Industriestandard zu realisieren. Die AWS Connected Vehicle Lösung sowie das AWS AGL Framework sind ebenfalls Open Source. Allerdings setzt diese Lösungen trotz allem die

[5] https://aws.amazon.com/de/answers/iot/connected-vehicle-solution/

[6] https://github.com/awslabs/aws-iot-framework-for-agl.

Tabelle 2 Vergleich der Plattformen

	AWS AGL	Eclipse Kuksa
Plattformeigenschaften		
Open Source	Nur, Connected Vehicle Lösung	Ja
AGL Integration	Ja	Ja
Demos	Ja	Noch nicht vorhanden
Kann auf eigener Hardware ausgeführt werden	Nein	Ja
Guidelines	Ja	Nein
Support	Gegen Bezahlung	Community Support
Kosten	12 Monate kostenloses Kontingent, danach gegen Bezahlung	kostenlos
Dokumentation		
AGL Integration	Vorhanden jedoch nicht sehr ausführlich und aktuell	Vorhanden allerdings nicht immer aktuell
Cloud-Dienste	Sehr ausführlich	Sehr ausführlich
Leistungsumfang		
OTA-Updates		
Update von Software	Ja	Ja
Update des Betriebssystems	Ja (nur Greengrass)	Ja
Grafische Benutzeroberfläche	Ja	Ja
Repository	Ja	Ja
Gerätemanagement		
Gruppendefinition	Ja	Ja
Fernüberwachung	Ja	Ja
Fernkonfiguration	Ja	Ja
Gerätestatusverwaltung	Ja	Ja
Entwicklungsumgebung		
Cloud-Service Entwicklung	Ja	Ja
In-Vehicle Entwicklung	Nein (nur Greengrass offline)	Ja
Echtzeit Datenanalyse	Ja	Nein
Datenspeicherung	Ja	Ja
App Store	Teilweise	Ja (in Arbeit)
Kommunikation		
Telemetrie	Ja	Ja
Command&Controll	Ja	Ja
Broadcast	Ja	Ja
Unterstützende Kommunikationsprotokolle		
MQTT	Ja	Ja
HTTP	Ja	Ja
AMQP	Nein	Ja

Verwendung von AWS-Diensten, sowie die Nutzung der AWS Cloud voraus, wodurch eine große Abhängigkeit entsteht, da diese Dienste nicht in andere Clouds portiert werden können. Aufgrund der offenen Entwicklung bietet Eclipse Kuksa zudem eine Community, die sich gegenseitig unterstützen kann, während der technische Support bei AWS nur gegen Bezahlung Unterstützung leistet. Wie in der Vergleichstabelle zu sehen, ist der Leistungsumfang beider Plattformen sehr ähnlich. AWS bietet allerdings mit Greengrass den großen Vorteil, Daten direkt im Auto offline zu verarbeiten. Dadurch werden zum einen Echtzeit-Datenanalysen und zum anderen ein nahtloser Betrieb trotz unzuverlässiger Verbindungen ermöglicht. Beide Plattformen bieten außerdem sehr gute Device Gateways und Message Router, welche unterschiedliche Standardprotokolle unterstützen. Allerdings bietet Eclipse Hono die Möglichkeit eigene Protokoll Adapter zu implementieren und diese dem Projekt hinzuzufügen. Die Community kann das Projekt somit selbstständig erweitern und verbessern.

Mit Eclipse HawkBit und dem AWS Device Management stellen beide ein mächtiges Gerätemanagement-Werkzeug zur Verfügung, mit welchen die Geräte verwaltet und über OTA-Updates ausgerollt werden können. Kuksa sieht zudem die Lösung eines App Stores vor, welcher über Eclipse Che und HawkBit realisiert werden soll. AWS hingegen hat in ihrer Architektur keinen App Store vorgesehen. Allerdings wird die Möglichkeit geboten, mithilfe des API Gateways Dienste von Drittanbietern zu integrieren, sodass sich jeder Hersteller seinen eigenen App Store aufbauen kann.

10 Diskussion

Vernetzte Autos und auch der Bereich IoT sind in der Automobilindustrie ein relativ neues Gebiet, weshalb die angebotenen Lösungen einen dementsprechenden jungen Reifegrad aufweisen. Gerade das Problem von Standards, welches AGL aufgreift, ist momentan stark in Diskussion. AWS hat allerdings viel Erfahrung mit Cloud-Infrastrukturen und IoT-Diensten, sodass diese auf die neuen Szenarien adaptiert werden konnten. Aus diesem Grund stehen für die Realisierung stabile ausgereifte Dienste zur Verfügung. Die Integration von Greengrass in den AGL-Stack ist allerdings weniger ausgereift, weshalb dies momentan noch mit einigen Schwierigkeiten in Form von bestehenden Bugs und fehlender Dokumentation verbunden ist. Das Kuksa Projekt befindet sich dagegen noch mitten in der Entwicklung und ist noch nicht vollständig implementiert. Aktuell ist lediglich die Basisimplementierung umgesetzt, welche es erlaubt AGL-Kuksa Images für den Raspberry Pi3 zu bauen, Daten an die Cloud zu übertragen und Applikationen mithilfe von Eclipse Che zu implementieren. Aus diesem Grund sind bislang keine Demos oder Tutorials vorhanden. Bis Ende Mai 2019 soll das Ökosystem allerdings im vollen Umfang vorhanden sein.

Durch den Einsatz von Serverless Computing bietet AWS viele Vorteile. Einer davon ist, dass sich nicht um das Aufsetzen und Warten der Infrastruktur gekümmert werden muss, sodass die Entwicklung der Dienste im Vordergrund steht. Dadurch sind schnellere Time-to-Market Zyklen möglich. Kuksa hingegen bietet mehr Raum für Konfigurationen und Infrastrukturentscheidungen.

11 Fazit

Die V2C Kommunikation bietet viele Vorteile und Möglichkeiten neue Services in das Auto zu integrieren, um das Auto individuell an den Benutzer anzupassen. Dabei ist es wichtig, durch die Services einen entsprechenden Kundenmehrwert zu erzeugen. Um dies zu realisieren ist allerdings ein offener Standard notwendig, damit Applikationen herstellerweit verfügbar werden. Momentan werden die In-Vehicle-Plattformen allerdings noch in Silos entwickelt. AGL macht einen ersten Schritt in die richtige Richtung mit dem Ziel, gemeinsam eine offene Plattform zu entwickeln. Aus dem Vergleich der Plattformen lässt sich schlussfolgern, dass sich beide für den Einsatz und die Realisierung einer V2C Kommunikation eignen. Allerdings zeigen die durchgeführten Versuche sowie die dargestellten Plattformeigenschaften, das sich der Einsatzzweck durchaus unterscheidet. Aus diesem Grund kann keine eindeutige Empfehlung ausgesprochen werden. Allerdings lässt sich sagen, dass AWS besser geeignet ist, um eine schnelle Realisierung des Backends durchzuführen. Sollen demnach in kurzer Zeit neue Dienste und Funktionen entstehen eignet sich AWS besser für die Realisierung. Kuksa hingegen bietet mehr Freiraum für Konfigurationen, sodass sich diese Lösung mehr dazu eignet, um ein individuelles Backend aufzubauen, welches neben der Cloud auch auf der eigenen Hardware ausgeführt werden kann.

Literaturverzeichnis

[1] Diewald, S., et al. 2011. Mobile Device Integration and Interaction in the Automotive Domain. In *In AutoNUI: Automotive Natural User Interfaces Workshop at the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI 2011)

[2] Coppola, R et al. 2016. Connected Car. Technologies, Issues, Future Trends. *ACM Comput. Surv.*, 3, 1–36. DOI= 10.1145/2971482

[3] Javadi, M. S., et al. 2013. Survey on Inter-Vehicle Communication Applications. Current Trends and Challenges. *Information Technology J.* 12, 2, 243–250. DOI= 10.3923/itj.2013.243.250

[4] Luddecke, D., et al. 2015. Modeling user intentions for in-car infotainment systems using Bayesian networks. In *2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems*. IEEE Incorporated, 378–385. DOI= 10.1109/MODELS.2015.7338269.

[5] Murphy, S., et al. 2013. Advanced service delivery to the Connected Car. In *IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications, 2013. 7 - 9 Oct. 2013, Lyon, France;* IEEE, Piscataway, NJ, 147–153. DOI =10.1109/WiMOB.2013.6673354.

[6] Hueger, F. 2012. Platform independent applications for in-vehicle infotainment systems via integration of CE devices. In *2012 IEEE International Conference on Consumer Electronics - Berlin. Berlin, Germany, 3 - 5 September 2012*. IEEE, Piscataway, NJ, 221–222. DOI=10.1109/ICCE-Berlin.2012.6336502.

[7] He, W., et al. 2014. Developing Vehicular Data Cloud Services in the IoT Environment. *IEEE Trans. Ind. Inf.* 10, 2, 1587–1595. DOI= 10.1109/TII.2014.2299233

[8] Rangarajan, S., et al. 2012. V2C. A secure vehicle to cloud framework for virtualized and on-demand service provisioning.

[9] Lee, C.-H., et al. 2017. A Framework for a Connected Electric Vehicle Cloud to Learn Drivers' Behaviors. In *25th International Conference on Systems Engineering. ICSEng 2017 : Las Vegas, Nevada, USA, 22-23 August 2017 : proceedings*. IEEE, Piscataway, NJ, 405–411. DOI=10.1109/ICSEng.2017.47.

[10] Bosler, M., et al. 2017. Connected-Car-Services. Eine Klassifikation der Plattformen für das vernetzte Automobil. *HMD Praxis der Wirtschaftsinformatik,* 6, 1005–1020. DOI= 10.1365/s40702-017-0336-z

[11] Yogarayan, S., et al. 2016. A Framework on Cloud Based Connected Car Services. In *Information science and applications (ICISA) 2016*, K. J. Kim and N. Joukov, Eds. Lecture notes in electrical engineering 376,1. Springer, Singapore, 11–20. DOI=10.1007/978-981-10-0557-2_2.

[12] *Automotive Grade Linux*. https://www.automotivelinux.org/. Accessed 18.09.2018.

[13] Höttger, R., et al. 2018. *Combining Eclipse IoT Technologies for a RPI3-Rover along with Eclipse Kuksa*. Accessed 18.09.2018.

[14] Eclipse Foundation. 2018. *Automotive Meets IoT: Innovating the Future Vehicle*. www.eclipse.org/community/eclipse_newsletter/2018/july/kuksa.php. Accessed 18.09.2018.

[15] *Eclipse Hono*. www.eclipse.org/hono/. Accessed 13.10.2018.

[16] *Eclipse hawkBit*. www.eclipse.org/hawkbit/. Accessed 18.09.2018.

[17] *Eclipse Ditto*. www.eclipse.org/ditto/. Accessed 19.09.2018.

[18] *Keycloak*. www.keycloak.org/. Accessed 18.09.2018.

[19] *Keti*. projects.eclipse.org/proposals/eclipse-keti. Accessed 22.09.2018.

[20] *Eclipse Che*. www.eclipse.org/che/. Accessed 18.09.2018.

[21] *AWS IoT*. https://aws.amazon.com/de/iot/. Accessed 20.10.2018.

[22] *AWS IoT Doku*. https://docs.aws.amazon.com/de_de/iot/latest/developerguide/what-is-aws-iot.html. Accessed 20.10.2018.

Verfahrensüberblick zur Object Pose Detection im 3D Raum

Nils Lüdeke
Hochschule Reutlingen
Nils.Luedeke@student.reutlingen-university.de

Abstract

Um Menschen mit einer Handlähmung nach einem Schlaganfall zu unterstützen, wird eine Kamera benötigt, welche Gegenstände erkennt und deren Position im Raum in Bezug zur Kamera stellt. Diese Informationen sollen dann einem Hand-Exoskelett weiterhelfen, um beim Greifen eines Objektes zu helfen. In dieser Arbeit werden die möglichen Ansätze zur Object Pose Detection beziehungsweise Recognition vorgestellt und verglichen. Dabei wird eine Methode gesucht, welche Alltagsgegenstände und deren Pose zur Kamera erkennen kann.

CCS Concepts

• **General and reference** → **Evaluation**; • **Computing methodologies** → *Machine learning approaches*; Neural networks;

Keywords

6-Dof, 3D, Object Detection, Pose Detection

1 Einleitung

Die Freiheitsgrade eines Systems, im Englischen Degrees of Freedom oder kurz DoF gennant, beschreiben die Zahl der voneinander abhängigen Variablen bei einer Bewegung. Die Verfolgung der 6 DoF eines planaren Ziels bezieht sich auf dessen Position und Ausrichtung und ist ein wichtiges Ziel im Forschungsbereich der Computer Vision. Mit der Entwicklung von Augmented Reality und Robotik wird die Anforderung, genaue und stabile 6 DoF-Posen für allgemeine planare Ziele zu erhalten, immer wichtiger. Auch für diese neuen Anwendungen ist eine Echtzeit-Systemperformance erforderlich, beispielsweise für fahrende Roboter. In den letzten Jahrzehnten wurden zahlreiche Systeme wie [1] für planare Ziele mit binärem Muster, dem sogenannten Referenzmarker, konstruiert. Obwohl diese Systeme sehr effizient sind, sind sie aufgrund der begrenzten Art von planaren Zielen im Allgemeinen nicht anwendbar. [2] Feature-basierte Methoden und Convolutional Neural Networks (CNN) sind die neuesten Lösungen. Die Kernidee von Feature-basierten Methoden ist es, *n*-Korrespondenzen zwischen dem Zielbild und dem Kamerabild mittels Feature-Matching und Tracking-Algorithmen zu finden. [2] Dabei könnte zum Tracking der überarbeitete SURF von Bay et al. [3] genutzt werden.

Beispielsweise gefolgt von einem RANSAC-basierten Schema [4], welches die Zuverlässigkeit der Korrespondenzen erhöht. Schließlich wird der Perspective-*n*-Point (P*n*P) [5] Algorithmus angewendet, um die endgültige Pose zu schätzen. In den letzten Jahren wurden Systeme vorgeschlagen, die eine Feature-basierte (merkmalsbasierten) Methode anwenden [6, 7]. Dennoch liefern diese Systeme

Betreuer Hochschule: Prof. Dr. Curio, Cristóbal
Hochschule Reutlingen
Cristobal.Curio@Reutlingen-University.de

Betreuer Hochschule: Gulde, Thomas
Hochschule Reutlingen
Thomas.Gulde@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Nils Lüdeke

nicht das richtige Ergebnis, wenn das Zielbild texturlos ist, das Kamerabild verschwommen ist oder der Neigungswinkel zwischen der Kamera und dem Ziel zu hoch ist, da die richtigen Feature Korrespondenzen, die sie gefunden haben, unzureichend sind. Um diese Probleme anzugehen, hat Tseng et al. [8] eine direkte Methode vorgeschlagen. Anstatt Punktkorrespondenzen zu finden, schätzt der Algorithmus die 6 DoF-Posen durch Messung der appearance distance, zwischen dem projizierten Zielbild und dem Kamerabild. Das Ergebnis zeigt, dass es die Feature-basierten Methoden in Bezug auf Genauigkeit und Robustheit übertrifft. Die Laufzeit dieser Methode ist jedoch aufgrund der großen Anzahl von zu berücksichtigenden Posen relativ lang. Da die Berechnung für jede Pose identisch und unabhängig ist, hat der Algorithmus das Potenzial, in hohem Maße parallel ausgeführt zu werden, was die Arbeit von Tseng et al. [2] inspirierte. [2]

1.1 Ziel der Arbeit

Das Ziel dieser Arbeit ist es, einige Ansätze zur 6 DoF Object Pose Detection vorzustellen und diese miteinander zu vergleichen, um einen geeigneten Ansatz für die Realisierung des Hand-Exoskeletts zu finden. Die Methode sollte 6 DoF direct (markerless) Pose Estimation, eine direkte, markerlose Posenschätzung beherrschen und für unterschiedliche Alltagsgegenstände unter verschiedenen Lichtumgebungen arbeiten können. Die verschiedenen Lichtumgebungen sind nötig um bei Dämmerung, schlechtem Licht noch voll funktionsfähig zu sein. Die Erkennung soll auf RGB-Bildern oder 3D Point Clouds basieren, da die Methode zur alltäglichen Nutzung eines Hand-Exoskeletts für Schlaganfallpatienten mit Handlähmung eingesetzt werden soll. Damit soll die Funktionalität der Hand wieder hergestellt werden.

1.2 6 DoF

Ziel der 6 DoF Pose Estimation (Posenschätzung) ist es, die sechs Freiheitsgrade zu bestimmen, die basierend auf der Ausrichtung und Position des Zielobjekts in Bezug auf die kalibrierte Kamera parametrisiert werden. Das Problem ist in Abbildung 1 veranschaulicht. [2]

2 Stand der Technik

Es gibt umfangreiche Literatur im Bereich der Posen- und Objekterkennung, einschließlich Instanz-, Kategorieerkennung, starrer und gegliederter Objekte sowie grober (quantisierter) und genauer 6D Posen. In der folgenden kurzen Übersicht werden Techniken vorgestellt, die sich speziell mit der Erkennung von Instanzen von starren Objekten in überladenen Szenen befassen und gleichzeitig deren 6D-Pose ableiten. Einige der Arbeiten und Methoden wurden bereits in der Einleitung erwähnt.

2.1 Template-basierte Ansätze

Der vielleicht traditionellste Ansatz zur Objekterkennung ist die Verwendung von Templates. Welcher Beispielsweise in der Arbeit von Hinterstoisser et al. [9] in Form eines Frameworks zur Objektverfolgung in Echtzeit mit der Microsoft Kinect zum Einsatz kommt. Dies bedeutet, dass eine starre Vorlage über das Bild gescannt wird und ein Abstandsmaß berechnet wird, um die beste Übereinstimmung zu finden. Als Stand der Technik bei Template-basierten Ansätzen verwendet [9] synthetische Renderings eines 3D Objektmodells, um eine große Anzahl von Templates zu erzeugen, die die gesamte Hemisphäre abdecken. Sie verwenden eine kantenbasierte Abstandsmessung, die für texturlose Objekte gut geeignet ist und verfeinern ihre Posenschätzungen mit Iterative Closest Point (ICP), um eine genaue 6D-Pose zu erzielen. Solche Template-basierten Ansätze können in der Praxis präzise und schnell funktionieren. [10]

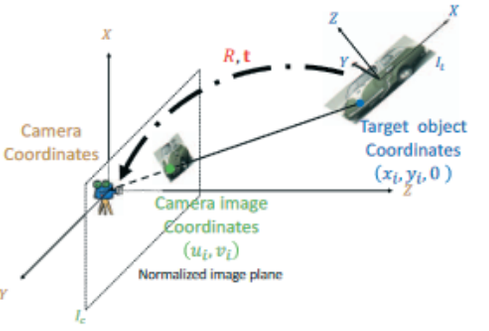


Abbildung 1: Koordinatensystem zwischen Zielobjekt und Kamerabild. [2]

2.2 Point-Cloud-basierte Ansätze

Manchmal sind die von der 3D-Kamera erfassten Daten, die Anzahl der Punkte auf dem Objekt gering. Außerdem ist die Auflösung der Point Cloud Data gering, sodass der Key Point nicht konsistent ist. In der Arbeit "Pose Estimation of Rigid Object in Point Cloud"[11] von Liu et al. , wird mit einem l60 x 120 auflösenden Pointcloud Datensatz, eine Methode präsentiert, die eine Pose Estimation durchführen kann. Dieses Verfahren basiert auf dem ICP Algorithmus. Nach Erhalt der 3D-Punktwolke der Szene werden die Ausreißerpunkte mit einem Filter entfernt, die eine genauere Registrierung des Objekts ermöglichen. Anschließend kann das Objekt aus dem Hintergrund mithilfe des euklidischen Clusteralgorithmus erkannt werden. Wenn ein Objekt erkannt wird, wird die Registrierung zwischen zwei benachbarten Bildern durchgeführt, sodass die 6 DoF Information des Ziels erhalten wird.[11] Zusammenfassend lässt sich sagen, dass der vorgeschlagene Ansatz interessante Ergebnisse für Punktwolken liefert, deren Auflösung gering ist. In diesen Fällen kann es eine effiziente Alternative zu anderen Methoden darstellen, da es keinen key point benötigt. Darüber hinaus nutzt es Tiefeninformationen, die es besser nutzt als die 2D-Pose Estimation [11]. Leider wurden in der Arbeit von [11], keine Angabe über die Menge an erkennbaren Objekten und die Genauigkeit zur Objekterkennung gemacht. Basierend auf dieser Arbeit, wurden eine Menge Verbesserungen durchgeführt, die letzte von Hinterstoisser et al. mit [12]. Dabei wurde die Genauigkeit nochmals verbessert und die genannten Ansätze von Hinterstoisser et al. [9], Brachmann et al. [10] und Drost et al. [13] geschlagen, bevor sie von PoseCNN [14] übertroffen wurden (PoseCNN wird in Kapitel 4 beschrieben).

2.3 Sparse Feature-basierte Ansätze

Eine beliebte Alternative zu Templates sind Sparse Feature-basierte Ansätze. Diese extrahieren interessante Punkte (oft scale-invariant) aus dem Bild, beschreiben diese mit lokalen Deskriptoren (oft affin und illumination invariant) und gleichen sie mit Einträgen einer Datenbank ab. So verwendete Lowe [15] beispielsweise SIFT-Deskriptoren und gruppierte Bilder aus ähnlichen Perspektiven zu einem einzigen Modell. Ein weiteres Beispiel für ein schnelles und skalierbares System ist "MOPED: A scalable and low latency object recognition and pose estimation system"von Martinez et al. [16]. Es hat sich gezeigt, dass Feature-basierte Techniken sich gut auf das Matching über riesige Vokabulare skalieren

lassen[Scalable recognition, Object-retrieval]. Dabei werden folgende Methoden zum Lernen benutzt, Interest Points [17], Deskriptoren [18] und Matching [19]. Trotz ihrer Beliebtheit ist eine große Einschränkung der spärlichen Ansätze für reale Anwendungen, dass sie ausreichend texturierte Objekte benötigen. Der Ansatz von Brachmann et al. [10] kann bei jeder Bilder lernenden 6D-Objekt-Positionsschätzung mit 3D-Objektkoordinaten von 5 Pixeln unabhängig von der Textur angewendet werden und kann die am besten geeigneten Bildmerkmale zu nutzen lernen. Ebenfalls gibt es Literatur zum Thema Kontur- und Formanpassung, die sich mit texturlosen Objekten befassen kann, zum Beispiel von Ferrari et al. [20]. Diese unterscheidet sich jedoch konzeptionell von Brachmann et al. [10]. [10]

2.4 Dense Ansätze

Eine Alternative zu Template-, Point –Cloud- und Sparse Feature-basierten Ansätze sind Dense Ansätze. Im verallgemeinerten Hough-Abstimmungsschema geben alle Pixel eine Stimme in einem quantisierten Vorhersagebereich (prediction space) ab (zum Beispiel 2D-Objektmitte und Maßstab) und die Zelle mit den meisten Stimmen wird als Gewinner gewertet. In Daniilidis et al. [21] und Gall et al. [22] wurde Hough's Voting für Objekterkennung verwendet und konnte grobe Objektpositionen vorhersagen.
In der Arbeit von Brachman et al. [10] wird eine Idee von Gall et al. [22] verwendet, um ein Ziel mit Hough-Voting und Objektsegmentierungen zu trainieren. Im Gegensatz zu [22] hat [10] jedoch eine einfache gemeinsame Verteilung über die Ausgänge (in diesem Fall 3D Objektkoordinaten und Objektklassenbeschriftung) gefunden, um die in [22] vorgeschlagenen Varianten zu übertreffen. Drost et al. [13] arbeitet ebenfalls mit einem Abstimmungsansatz (voting approach) und verwendet Oriented Point Pair Feature, um den Suchraum zu reduzieren. Um eine vollständige 6D-Pose zu erhalten, könnte eine Variante mit [22] helfen, bei der jedes Pixel direkt für eine globale, quantisierte 6D-Pose steht. Die hohe Dimensionalität des Suchraums (und damit der notwendige hohe Quantisierungsgrad) wird jedoch wahrscheinlich zu einer schlechten Einschätzung der Pose führen.
Im Ansatz von Brachmann et al. [10] macht jedes Pixel stattdessen eine kontinuierliche 3D-Vorhersage über seine lokale Übereinstimmung (correspondence) mit einem 3D-Modell. Dies reduziert den Suchraum massiv und ermöglicht das Erlernen einer diskriminierenden Vorhersage einer stark reduzierten Trainingsmenge, da nicht jeder Punkt auf der Oberfläche des Objekts aus allen möglichen Blickwinkeln gesehen werden muss. Ebenso wird gezeigt, dass die 3D Objektkorrespondenzen eine nachfolgende Modellanpassungsphase effizient steuern können, um eine hoch präzise 6D-Objekt-Pose zu erhalten. [10]
Ebenso gibt es Ansätze zur Object Class Detection, die einen ähnlichen Ansatz wie die 3D-Objektkoordinatendarstellung von [10] verwenden. Eines der ersten Systeme ist das 3D LayoutCRF [23], dass die Aufgabe hat, eine dichte Teilbeschriftung (dense partlabelling), die das 3D rigid Object abdeckt, mithilfe eines decision forest vorherzusagen, obwohl sie nicht versucht haben, ein 3D-Modell an diese Labels anzupassen. Danach wurde der Vitruvianische Krümmer (Vitruvian Manifold) [24] für die menschliche Posenschätzung eingeführt und die Szenenkoordinaten (scene coordinate) Regression Forests zur Kamerarelokalisierung in RGB-D-Bildern eingeführt [25]. [10]

3 6D-Object Pose Estimation mit 3D-Objektkoordinaten

Die Arbeit von Brachmann et al. [10] behandelt das Problem der Schätzung der 6D Pose von bestimmten Objekten aus einem einzigen RGB-D-Bild. Es wird ein flexibler Ansatz präsentiert, der mit generischen Objekten umgehen kann, sowohl texturiert als auch texturlos. Das zentrale neue Konzept ist eine erlernte,

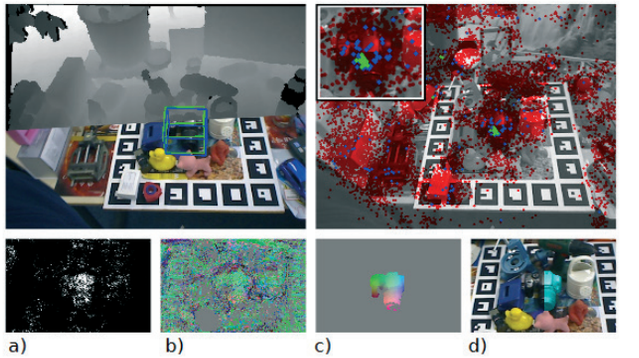


Abbildung 2: Die Methode von Brachmann et al. [10] ist in der Lage, die richtige Pose zu finden, in der eine Template basierte Methode versagt. Das Testbild (a) zeigt eine Situation mit starker Okklusion. Die Pose Estimation (Posenschätzung) von diesem Ansatz ist blau dargestellt. Die Pose, die durch die Neuimplementierung der Methode von Hinterstoisser et al. aus [26] dargestellt wird, ist rot. (b) Die Koordinatenvorhersagen für (a) aus einem Baum, der dem RGB-Würfel zugeordnet und mit p_c, i multipliziert wurde. (c) Ground Truth Objektkoordinaten für eine Abbildung auf den RGB-Würfel. (d) Testbild mit extremen Lichtverhältnissen, die sich vom Trainingsset unterscheiden. Geschätzte Posen werden wie in (a) angezeigt. [10]

zwischengeschaltete Darstellung in Form einer dichten 3D-Object Coordinate Labelling gepaart mit einem dichten Class Labelling. Eine Übersicht ist in Abbildung 2 zu sehen.
Es wurde gezeigt, dass bei einem gemeinsamen Datensatz mit texturlosen Objekten, bei denen Template-basierte Techniken geeignet und auf dem neuesten Stand der Technik sind, dieser Ansatz in Bezug auf die Genauigkeit etwas überlegen ist.Ebenso wurde die Vorteile des Ansatzes im Vergleich zu Template-basierten Techniken auch in der Robustheit gegenüber wechselnden Lichtverhältnissen demonstriert. Zu diesem Zweck wurde ein neuer Ground Truth Datensatz mit 10k Bildern von 20 Objekten erstellt, die jeweils unter drei verschiedenen Lichtverhältnissen aufgenommen wurden. Ebenfalls wird gezeigt, dass der Ansatz mit der Anzahl der Objekte gut skalierbar ist und die Fähigkeit hat, schnell zu laufen. [10]

4 Pose Estimation mit PoseCNN

In der Arbeit von Xiang et. al. [14] wird PoseCNN vorgestellt, ein neues Convolutional Neural Network für die 6D Object Pose Estimation. PoseCNN schätzt die 3D Übersetzung eines Objekts durch Lokalisierung seiner Mitte im Bild und der Vorhersage seines Abstands zur Kamera. Die 3D-Rotation des Objekts wird durch Rückführung in eine Quaternion Repräsentation geschätzt. Ebenso wird eine neuartige Verlustfunktion vorgestellt, die dem PoseCNN ermöglicht mit symmetrischen Objekten umzugehen. Darüber hinaus wurde ein großformatiger Videodatensatz für die 6D-Objekt Pose Estimation mit dem Namen YCB-Video-Datensatz, erstellt.
Dieser Datensatz liefert genaue 6D-Posen von 21 Objekten aus dem YCB-Datensatz, die in 92

Videos mit 133.827 Bildern beobachtet wurden. Ebenso wurden umfangreiche Experimente mit dem YCB-Video-Datensatz und dem OccludedLINEMOD Datensatz durchgeführt. Dies zeigt, dass PoseCNN sehr robust gegenüber Okklusionen ist und mit symmetrischen Objekten umgehen kann. Dabei wird die Pose Estimation nur mit Bildern als Input gefüttert. Bei der Verwendung von Tiefendaten zur weiteren Verfeinerung der Posen erreicht dieser Ansatz Sate of the Art Ergebnisse auf dem anspruchsvollen OccludedLINEMOD Datensatz.

Traditionell wird das Problem der 6D-Object Pose Estimation durch den Abgleich von Feature Points zwischen 3D-Modellen und Bildern gelöst, wie bereits im vorherigen diskutiert. Diese Methoden erfordern jedoch, dass es auf den Objekten reichhaltige Texturen gibt, um Feature Points für den Abgleich zu erkennen. Infolgedessen sind sie nicht in der Lage, mit texturlosen Objekten umzugehen. Mit dem Aufkommen von Tiefenkameras wurden mehrere Methoden zur Erkennung texturloser Objekte unter Verwendung von RGB-D-Daten [9, 10, 27], vorgeschlagen. Bei templatebasierten Methoden wie [9, 28] reduzieren Okklusionen die Erkennungsleistung signifikant. Alternativ können Methoden, die das Erlernen der Regression von Bildpunkten zu 3D-Objektkoordinaten durchführen, um die 2D-3D-Korrespondenzen für die 6D Positionsschätzung [10] zu ermitteln, keine symmetrischen Objekte behandeln. [14]

4.1 Überblick über das Netzwerk

Abbildung 3 veranschaulicht die Architektur des Netzwerks für 6D Object Pose Estimation. Das Netzwerk besteht aus zwei Stufen. Die erste Stufe besteht aus 13 Convolutional Layers und 4 Max-Pooling-Layers, die Feature-Maps mit unterschiedlichen Auflösungen aus dem Eingangsbild extrahieren. Diese Phase ist das Backbone des Netzwerks, da die extrahierten Funktionen über alle Aufgaben des Netzwerks verteilt sind. Die zweite Stufe besteht aus einem Embedding Step, der die von der ersten Stufe erzeugten hochdimensionalen Feature Maps in niedrigdimensionale, aufgabenspezifische Features einbindet. Anschließend führt das Netzwerk drei verschiedene Aufgaben aus, die zur 6D Pose Estimation führen, nämlich Semantic Labeling, 3D Translation Estimation und 3D Rotation Regression. [14]

4.2 Semantic Labeling

Um Objekte in Bildern zu erkennen, wird auf Semantic Labeling zurückgegriffen, bei denen das Netzwerk jedes Bildpixel in eine Objektklasse einordnet. Im Vergleich zu den bisherigen 6D Pose Estimations, die auf die Objekterkennung mit Bounding Boxes [29–31] zurückgreifen, bietet das Semantic Labeling reichhaltigere Informationen über die Objekte und behandelt Occlusions besser. Der Embedding Step des Semantic Labeling Branch, wie in Abbildung 3 dargestellt, besteht aus zwei Feature-Maps mit der channel dimension 512, die von der Feature-extraction-stage als Eingänge erzeugt werden. Die Auflösungen der beiden Feature Maps betragen 1/8 bzw. 1/16 der ursprünglichen Bildgröße. Das Netzwerk reduziert zunächst die Channel dimension der beiden Feature Maps auf 64 mithilfe von zwei Convolutional Layers. Dann verdoppelt es die Auflösung der 1/16 Feature Map mit einem Deconvolutional Layer. Danach werden die zwei Feature-Maps summiert und ein weiterer Deconvolutional Layer wird verwendet. Dieser erhöht die Auflösung um das 8-fache, um eine Feature-Map mit der Originalgröße des Bildes zu erhalten. Schließlich arbeitet ein Convolutional Layer auf der Feature-Map und erzeugt ein Semantic Labeling für Pixel. Die Ausgabe dieses Layers hat n channels mit n die Anzahl der semantischen Klassen. Im Training wird ein Softmax-Kreuzentropieverlust (softmax cross entropy) angewendet, um den Semantic Labeling

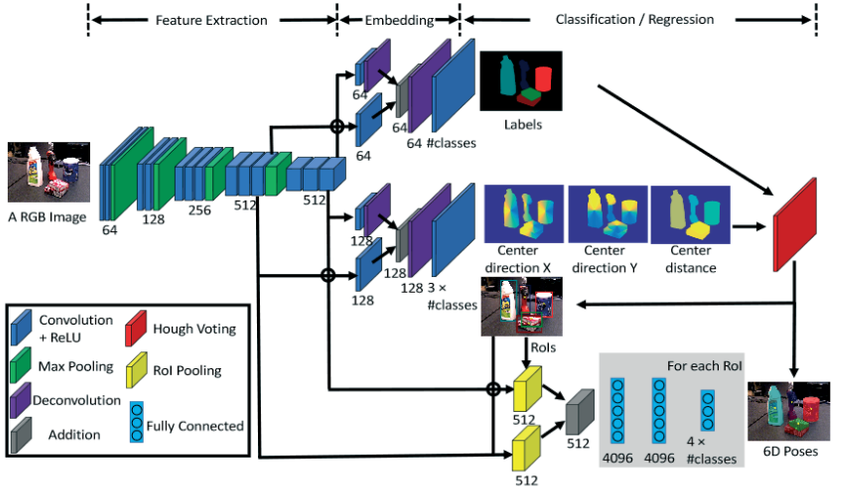


Abbildung 3: Architektur von PoseCNN für die 6D-Object Pose Estimation. [14]

Method	Michel et al. [21]	Hinterstoisser et al. [14]	Krull et al. [17]	Brachmann et al. [3]	Ours PoseCNN Color	Ours PoseCNN+ICP
Ape	80.7	**81.4**	68.0	53.1	9.6	76.2
Can	88.5	**94.7**	87.9	79.9	45.2	87.4
Cat	**57.8**	55.2	50.6	28.2	0.93	52.2
Driller	**94.7**	86.0	91.2	82.0	41.4	90.3
Duck	74.4	**79.7**	64.7	64.3	19.6	77.7
Eggbox	47.6	65.5	41.5	9.0	22.0	**72.2**
Glue	73.8	52.1	65.3	44.5	38.5	**76.7**
Holepuncher	**96.3**	95.5	92.9	91.6	22.1	91.4
MEAN	76.7	76.3	70.3	56.6	24.9	**78.0**

Abbildung 4: 6D Pose Estimation Genauigkeit/Präzision auf dem OccludedLINEMOD Dataset. Die rot gefärbten Objekte sind symmetrisch. Alle Methoden verwenden die Tiefe, mit Ausnahme der Pose CNN-Farbe. [14]

Branch zu trainieren. Während des Tests wird eine Softmax-Funktion ausgeführt, um die Klassenwahrscheinlichkeiten der Pixel zu berechnen. Das Design des Semantic Labeling Branch ist inspiriert von dem vollständigen Convolutional Network in [32] zum Semantic Labeling. Es wird auch in der vorherigen Arbeit für das Scene Labeling [33] verwendet. [14]

5 Fazit und Ausblick

Laut dem Vergleich von [14], zu sehen in Abbildung 4, ist mittels des OccludedLINEMOD Datensatzes das PoseCNN deutlich genauer als die folgenden Ansätze. Das PoseCNN schlägt die Ansätze von Brachmanns et al. [10], Drost et al. [13] und Hinterstoisser et al. [9]. Hierbei erzielte Brachmans [10] 56.6, Hinterstoisser [12] 76.3 das PoseCNN+ICP [14] 78.0. Ebenso übertrifft PoseCNN die Arbeit von Tekin et al. [30] deutlich, insbesondere wenn die Schwelle für den reprojection error klein ist. Dies ist zu sehen in Abbildung 4 von [14]. Tekins CNN basiert auf YOLOv2 und war bisher der Stand der Technik für

den LINEMOD Datensatz. Ein weiterer Vorteil des PoseCNN's ist, dass der Code und der Datensatz auf GitHub zur Verfügung stehen und somit für das Handexoskelett genutzt werden können. Beim Einsatz für das Handexoskeletts ist es von Vorteil, dass PoseCNN auch mit überladenen Szenen zurechtkommt, wie beispielsweise einer Schublade. Jedoch wird in [14] nicht auf den Einfluss verschiedener Lichtverhältnisse eingegangen, was somit noch getestet werden muss.

Literatur

[1] J. Yoon, J. Kim u. a., "A Unified Graphics and Vision Processor With a 0.89 /spl mu/W/fps Pose Estimation Engine for Augmented Reality", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Bd. 21, Nr. 2, S. 206–216, Feb. 2013, ISSN: 1063-8210. DOI: 10.1109/TVLSI.2012.2186157. Adresse: https://ieeexplore.ieee.org/document/6156811.

[2] H. Tseng, P. Wu u. a., "D-PET: A direct 6 DoF pose estimation and tracking system on graphics processing units", in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, Springer Verlag, Mai 2017, S. 1–4. DOI: 10.1109/ISCAS.2017.8050811. Adresse: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=%5C&arnumber=8050811%5C&isnumber=8049747.

[3] H. Bay, A. Ess u. a., "Speeded-up robust features (surf)", *Computer Vision and Image Understanding*, Bd. 110, Nr. 3, 2008. Adresse: https://www.sciencedirect.com/science/article/pii/S1077314207001555?via%3Dihub.

[4] M. A. Fischler und R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Commun. ACM*, Bd. 24, Nr. 6, S. 381–395, Juni 1981, ISSN: 0001-0782. DOI: 10.1145/358669.358692.

[5] Y. Zheng, Y. Kuang u. a., "Revisiting the PnP Problem: A Fast, General and Optimal Solution", in *2013 IEEE International Conference on Computer Vision*, Dez. 2013, S. 2344–2351. DOI: 10.1109/ICCV.2013.291. Adresse: https://ieeexplore.ieee.org/document/6751402.

[6] G. Simon, A. W. Fitzgibbon u. a., "Markerless tracking using planar structures in the scene", in *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, Okt. 2000, S. 120–128. DOI: 10.1109/ISAR.2000.880935. Adresse: https://ieeexplore.ieee.org/document/880935.

[7] M. Schaeferling, U. Hornung u. a., "Object Recognition and Pose Estimation on Embedded Hardware: SURF-based System Designs Accelerated by FPGA Logic", *Int. J. Reconfig. Comput.*, Bd. 2012, 6:6–6:6, Jan. 2012, ISSN: 1687-7195. DOI: 10.1155/2012/368351. Adresse: http://dx.doi.org/10.1155/2012/368351.

[8] H. Tseng, P. Wu u. a., "Direct 3D pose estimation of a planar target", in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, März 2016, S. 1–9. DOI: 10.1109/WACV.2016.7477640. Adresse: https://ieeexplore.ieee.org/document/7477640.

[9] S. Hinterstoisser, S. Lepetit Vincent and Ilic u. a., "Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes", in *Computer Vision – ACCV 2012*, K. M. Lee, Y. Matsushita u. a., Hrsg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, S. 548–562, ISBN: 978-3-642-37331-2.

[10] E. Brachmann, A. Krull u. a., "Learning 6D Object Pose Estimation Using 3D Object Coordinates", in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla u. a., Hrsg., Cham: Springer International Publishing, 2014, S. 536–551, ISBN: 978-3-319-10605-2. Adresse: http://wwwpub.zih.tu-dresden.de//~cvweb/publications/papers/2014/PoseEstimationECCV2014.pdf.

[11] Z. Liu, G. Liu u. a., "Pose estimation of rigid object in point cloud", in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Springer Verlag, Okt. 2016, S. 708–713. DOI: 10.1109/CISP-BMEI.2016.7852801. Adresse: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7852801&isnumber=7852668.

[12] S. Hinterstoisser, V. Lepetit u. a., "Going Further with Point Pair Features", in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas u. a., Hrsg., Cham: Springer International Publishing, 2016, S. 834–848, ISBN: 978-3-319-46487-9.

[13] B. Drost, M. Ulrich u. a., "Model globally, match locally: Efficient and robust 3D object recognition", in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Juni 2010, S. 998–1005. DOI: 10.1109/CVPR.2010.5540108. Adresse: https://ieeexplore.ieee.org/document/5540108.

[14] Y. Xiang, T. Schmidt u. a., "PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes", *CoRR*, Bd. abs/1711.00199, 2018. arXiv: 1711.00199. Adresse: http://arxiv.org/abs/1711.00199.

[15] D. G. Lowe, "Local feature view clustering for 3D object recognition", in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Bd. 1, Dez. 2001, S. I–I. DOI: 10.1109/CVPR.2001.990541. Adresse: https://ieeexplore.ieee.org/document/990541.

[16] M. Martinez, A. Collet u. a., "MOPED: A scalable and low latency object recognition and pose estimation system", in *2010 IEEE International Conference on Robotics and Automation*, Mai 2010, S. 2043–2049. DOI: 10.1109/ROBOT.2010.5509801. Adresse: https://ieeexplore.ieee.org/document/5509801.

[17] S. Holzer, J. Shotton u. a., "Learning to Efficiently Detect Repeatable Interest Points in Depth Data", in *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik u. a., Hrsg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, S. 200–213, ISBN: 978-3-642-33718-5.

[18] S. Winder, G. Hua u. a., "Picking the best DAISY", in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Juni 2009, S. 178–185. DOI: 10.1109/CVPR.2009.5206839. Adresse: https://ieeexplore.ieee.org/document/5206839.

[19] M. Ozuysal, M. Calonder u. a., "Fast Keypoint Recognition Using Random Ferns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 32, Nr. 3, S. 448–461, März 2010, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2009.23. Adresse: https://ieeexplore.ieee.org/document/4760148.

[20] V. Ferrari, F. Jurie u. a., "From Images to Shape Models for Object Detection", *Int. J. Comput. Vision*, Bd. 87, Nr. 3, S. 284–303, Mai 2010, ISSN: 0920-5691. DOI: 10.1007/s11263-009-0270-9.

[21] M. Sun, G. Bradski u. a., "Depth-Encoded Hough Voting for Joint Object Detection and Shape Recovery", in *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos u. a., Hrsg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, S. 658–671, ISBN: 978-3-642-15555-0.

[22] J. Gall, A. Yao u. a., "Hough Forests for Object Detection, Tracking, and Action Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 33, Nr. 11, S. 2188–2202, Nov. 2011, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.70. Adresse: https://ieeexplore.ieee.org/document/5740927.

[23] D. Hoiem, C. Rother u. a., "3D LayoutCRF for Multi-View Object Class Recognition and Segmentation", in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Juni 2007, S. 1–8. DOI: 10.1109/CVPR.2007.383045. Adresse: https://ieeexplore.ieee.org/document/4270070.

[24] J. Taylor, J. Shotton u. a., "The Vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation", in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Juni 2012, S. 103–110. DOI: 10.1109/CVPR.2012.6247664. Adresse: https://ieeexplore.ieee.org/document/6247664.

[25] J. Shotton, B. Glocker u. a., "Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images", IEEE, Juni 2013. Adresse: https://www.microsoft.com/en-us/research/publication/scene-coordinate-regression-forests-for-camera-relocalization-in-rgb-d-images.

[26] S. Hinterstoisser, V. Lepetit u. a., "Model Based Training, Detection and Pose Estimation of Texture-less 3D Objects in Heavily Cluttered Scenes", in *Proceedings of the 11th Asian Conference on Computer Vision - Volume Part I*, Ser. ACCV'12, Daejeon, Korea: Springer-Verlag, 2013, S. 548–562, ISBN: 978-3-642-37330-5. DOI: 10.1007/978-3-642-37331-2_42.

[27] M. Schwarz, H. Schulz u. a., "RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features", in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, Mai 2015, S. 1329–1335. DOI: 10.1109/ICRA.2015.7139363. Adresse: https://ieeexplore.ieee.org/document/7139363.

[28] S. Hinterstoisser, C. Cagniart u. a., "Gradient Response Maps for Real-Time Detection of Textureless Objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 34, Nr. 5, S. 876–888, Mai 2012, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.206. Adresse: https://ieeexplore.ieee.org/document/6042881.

[29] M. Rad und V. Lepetit, "BB8: A Scalable, Accurate, Robust to Partial Occlusion Method for Predicting the 3D Poses of Challenging Objects without Using Depth", *CoRR*, Bd. abs/1703.10896, 2017. arXiv: 1703.10896. Adresse: http://arxiv.org/abs/1703.10896.

[30] B. Tekin, S. N. Sinha u. a., "Real-Time Seamless Single Shot 6D Object Pose Prediction", *CoRR*, Bd. abs/1711.08848, 2017. arXiv: 1711.08848. Adresse: http://arxiv.org/abs/1711.08848.

[31] W. Kehl, F. Manhardt u. a., "SSD-6D: Making RGB-based 3D detection and 6D pose estimation great again", *CoRR*, Bd. abs/1711.10006, 2017. arXiv: 1711.10006. Adresse: http://arxiv.org/abs/1711.10006.

[32] E. Shelhamer, J. Long u. a., "Fully Convolutional Networks for Semantic Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 39, Nr. 4, S. 640–651, Apr. 2017, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2016.2572683.

[33] Y. Xiang und D. Fox, "DA-RNN: Semantic Mapping with Data Associated Recurrent Neural Networks", *CoRR*, Bd. abs/1703.03098, 2017. arXiv: 1703.03098. Adresse: http://arxiv.org/abs/1703.03098.

Untersuchung klinisch validierter Assessments zur Abstraktion und Ableitung eines logischen Modells

Pia Laubacher
Hochschule Reutlingen
Pia.Laubacher@Student.Reutlingen-University.DE

Abstract

Im Rahmen dieser Ausarbeitung werden klinisch validierte Assessmentinstrumente hinsichtlich ihres strukturellen Aufbaus betrachtet und evaluiert. Es werden einzelne Items extrahiert und statistische Skalenniveaus eingebunden, um ein Modell zu generieren, welches softwaretechnische Aspekte umfasst. Somit wird eine Realisierung einer Auswertelogik losgelöst vom User Interface ermöglicht. Daneben zeigt sich, dass Assessments aufgrund kausaler Zusammenhänge von erklärenden Größen und deren Abhängigkeiten untereinander einen multivariaten Charakter aufweisen. Zudem können neu eingehende klinische Assessments durch die Abbildung auf eine gemeinsame Basis in Form der Skalenniveaus eindeutig zugeordnet werden.

Schlüsselwörter

Klinische Assessments, Modellgenerierung, Evaluierung, Abstraktion, Validierung, Skalenniveaus, Softwaremodellierung

CCS Concepts

• Software and its engineering ~ Abstraction, modeling and modularity • Information systems ~ Document representation • Applied computing ~ Life and medical sciences

1 Einleitung

Für die Beurteilung des Gesundheitszustandes oder für die Kontrolle des Therapieverlaufs eines Patienten, z.B. bei der Rehabilitation, gewinnen Assessments immer mehr an Bedeutung [1]. Durch diese systematische Form der Patientenuntersuchung sind standardisierte und valide Messungen hinsichtlich des Zustandes des Patienten im Gegensatz zu rein beschreibenden Berichten möglich [1]. Es werden je nach gewünschter Messung unterschiedliche Assessmentinstrumente verwendet. Die Zahl dieser Instrumente liegt international nach Reuschenbach et al. im fünfstelligen Bereich [2].

In dieser Arbeit wird die zentrale Frage geklärt, ob klinische Assessmentinstrumente hinsichtlich ihres strukturellen Aufbaus auf ein gemeinsames Modell abstrahiert werden können, damit eine softwaretechnische Realisierung einer Auswertelogik losgelöst vom User Interface ermöglicht wird. Daher wird zu Beginn ein Überblick über vorhandene klinisch validierte Assessments mit einer inhaltlichen Kategorisierung gegeben. Anschließend werden diese Assessments hinsichtlich ihres strukturellen Aufbaus evaluiert, um ein zentrales Modell abzuleiten.

Betreuer Hochschule: Prof. Dr. Christian Thies
Hochschule Reutlingen
Christian.Thies@Reutlingen-University.de
Betreuer Firma: Fabian Kühn, M.Sc.
Philips Medizin-Systeme
Böblingen GmbH
fabian.kuehn@philips.com
Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Pia Laubacher

Hierbei wird der erhobene Inhalt vernachlässigt und ausschließlich die Zusammensetzung und Struktur des Aufbaus betrachtet.

2 Klinische Assessments

Unter *Assessment* wird in der Regel ein Prozess verstanden, an dessen Ende eine rationale Entscheidung steht [1]. Diese im Deutschen als „Beurteilung“ oder „Einschätzung“ übersetzte Methode stellt im klinischen Umfeld eine systematische, standardisierte und strukturierte Form der Anamnese und/oder Untersuchung von Patienten dar [1]. Dabei ist zu beachten, dass Informationen nicht nur gesammelt, sondern auch bewertet werden um handlungsleitend zu sein [3]. Neben dem erwähnten Einsatz in der Praxis, spielen klinische Assessments auch in der Forschung eine große Rolle [3]. In Deutschland werden nach Reuschenbach et al. [2] Risikoassessments (z.B. Sturzrisiko), Assessments zur Beurteilung der Pflegebedürftigkeit und des Pflegebedarfs, als auch Assessments zur Einschätzung der Ernährungssituation (z.B. Mangelernährung) stark erforscht. Speziell durch Risikoassessments soll es gelingen, dass Aussagen hinsichtlich der Wahrscheinlichkeit des Auftretens eines Pflegeproblems (z.B. Dekubitus) genauer werden [2]. Es ist zu beachten, dass Assessments eine Situation immer nur vereinfacht und reduziert abbilden [3]. Der Nutzen des Instruments und die Brauchbarkeit des gelieferten Ergebnisses hängen daher neben der Qualität des Instruments auch von den Fähigkeiten und Kenntnissen des Anwenders (=Assessors) ab.

2.1 Assessmenterhebung

Nach Reuschenbach et al. [2] existieren unterschiedliche Methoden, um ein Assessment durchzuführen: Neben Befragungen, zielgerichteten Beobachtungen, physiologischen Messungen (z.B. Blutdruck, Herzfrequenz, Körpertemperatur), Aktenanalysen und Familiendiagnostiken, können auch spezifische Assessmentinstrumente oder Assessmentverfahren (strukturierte Erhebungen mittels Skalen, Testverfahren oder Fragebögen [2]) eingesetzt werden. [2] Hierbei gibt es Instrumente, die einer differenzierten Patientenuntersuchung vorgeschaltet werden, um etwa die Gefährdung von Personen einzustufen [3] (z.B. bei der Triage). Diesen sogenannten Screeninginstrumenten muss eine vertiefende Untersuchung folgen, die die Gefährdung widerlegt oder bestätigt und gegebenenfalls genauer beschreibt [3].

2.2 Arbeitsdefinition

Unabhängig davon, ob Assessmentinstrumente angewendet werden oder nicht, ist jede Einschätzung oder Informationsgewinnung somit ein Assessment. Um das Untersuchungsgebiet für diese Ausarbeitung einzugrenzen, wurden ausschließlich strukturierte Erhebungen mittels Fragebögen, Skalen und Testverfahren näher beleuchtet, die konkret einen Frage-Antwort-Charakter aufweisen.

2.3 Kategorisierung

Im Rahmen der Untersuchung wurden insgesamt 350 klinische Assessments recherchiert und evaluiert. Diese Assessments wurden aus wissenschaftlichen Datenbanken (u.a. PubMed, IEEE, NCBI), Fachliteratur (Onlineressourcen von z.B. Wiley Online Library, Thieme Connect sowie Assessmentvorstellungen in z.B. [1] - [6]) und aus konkreten Kliniken bezogen. Durch inhaltliche Kategorisierung waren Assessments aus unterschiedlichen Gruppen vorhanden (siehe Abbildung 1).

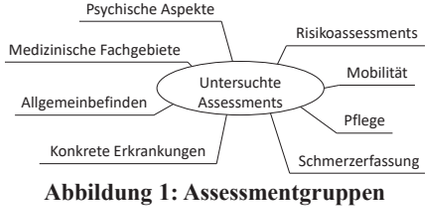


Abbildung 1: Assessmentgruppen

Bei den *Risikoassessments* wurden Assessments beispielsweise hinsichtlich des Sturzrisikos (z.B. Tinetti-Test und Skalen nach Morse, Hendrich und Conley), des Dekubitusrisikos (z.B. Skalen nach Norton und Braden), des Thromboserisikos (z.B. Entzündungen von venösen Gefäßen) und der Ernährungssituation (z.B. Mini Nutritional Assessment) betrachtet. Bei der *Mobilität* fanden sich Assessments bezüglich der Mobilitätsbewertung (z.B. Timed Up and Go Test) und der Selbstständigkeit (z.B. Barthel-Index), aber auch bezüglich der Rehabilitation (z.B. Berg Balance Scale). Bei der *Pflege* wurden Assessments für die Einschätzung der Pflegebedürftigkeit, beziehungsweise des Pflegebedarfs hinzugezogen. In der *Schmerzerfassung* wurde in Selbsteinschätzung (z.B. die in Kapitel 3.1.2.2 Grafische Elemente beschriebenen Skalen) und Fremdbeurteilung (z.B. ECPA-Schmerzskala) unterteilt. *Psychische Aspekte* beinhalteten Assessments beispielsweise für die Ermittlung des mentalen Status (z.B. Mini-Mental-Status-Test) oder für affektive Störungen wie Angst. *Medizinische Fachgebiete* enthielten etwa die Geburtshilfe (z.B. Apgar-Score in der Neonatologie) und die Notfallmedizin/Intensivmedizin (z.B. Glasgow Coma Scale, Emergency Severity Index, Manchester Triage Score). Im *Allgemeinbefinden* waren Assessments unter anderem für Hand Over Reports und bezüglich des generellen, beziehungsweise des oralen Gesundheitsstatus hinterlegt. Den *konkreten Erkrankungen* wurden Assessments für Krankheiten wie beispielsweise Inkontinenz, Herzinsuffizienz oder Demenz zugeordnet.

Es ergab sich nach Abbildung 2, dass Assessments je nach Gruppierung zusätzlich nach unterschiedlichen Kriterien kategorisierbar sind. Für die nachfolgende Abstraktion wurden die recherchierten Assessments hinsichtlich des Aufbaus untersucht, da dies am wichtigsten für einen späteren softwaretechnischen Umsetzungsprozess ist. Die nicht näher betrachteten Kriterien fließen durch spätere inhaltliche Betrachtungen direkt in den Umsetzungsprozess mit ein.

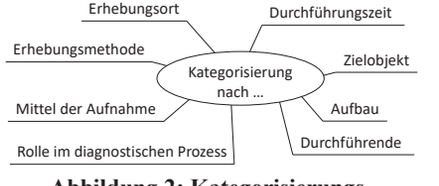


Abbildung 2: Kategorisierungsmöglichkeiten von Assessments

3 Abstraktion und Modellgenerierung

Nachfolgend wird auf Basis der Rechercheergebnisse der Aufbau von Assessments evaluiert. Anschließend erfolgt die Abbildung auf Skalenniveaus, der Transfer in eine Softwaremodellierung und die Gesamtbetrachtung des Modells.

3.1 Aufbau von Assessments

Für die Domäne Assessment zeigte sich während der Evaluation der in Abbildung 3 dargestellte abstrakte Formularaufbau. In den drei Abschnitten *Kopfteil*, *Hauptteil* und *Fußteil* des Formularlayouts finden sich jeweils die Informationsklassen „zu Beurteilender“, „erhobene Daten“ und „Assessor“. Im Kopfteil sind neben allgemeinen Informationen (u.a. Name und Logo der Institution) auch die Identifikationsnummer, der Vor- und Nachname und das Geburtsdatum des Patienten für die Informationsklasse „zu Beurteilender“ hinterlegt. Der Hauptteil ist der gesamte genutzte Bereich für die Informationsklasse „erhobene Daten“, inklusive der daraus resultierenden Bewertung und Einstufung des „zu Beurteilenden“ durch das Assessment. In der Untersuchung werden einzelne Items extrahiert, um das gewünschte Ergebnis (z.B. das Dekubitusrisiko) zu ermitteln und durch das Assessment abzubilden. Es zeigt sich in der Aufbaustruktur, dass die Assessments in den Einfachaufbau und den Mischaufbau aufgeteilt werden können. Im Fußteil sind neben der Informationsklasse „Assessor“ (u.a. Name, Signatur und Rang) auch der Ort der Assessmentaufnahme und das Datum mit Uhrzeit hinterlegt.

3.1.1 Auswertearithmetik

Die erhobenen Daten werden durch die Assessments auf unterschiedliche Weise verarbeitet und bewertet. Dabei werden teilweise Gesamtpunktwerte (=Scores) gebildet, die für eine Einstufung des Patienten dienen (z.B. gefährdet oder nicht gefährdet), beziehungsweise die Hinweise auf das weitere Vorgehen liefern sollen (z.B. Dekubitusrisi-

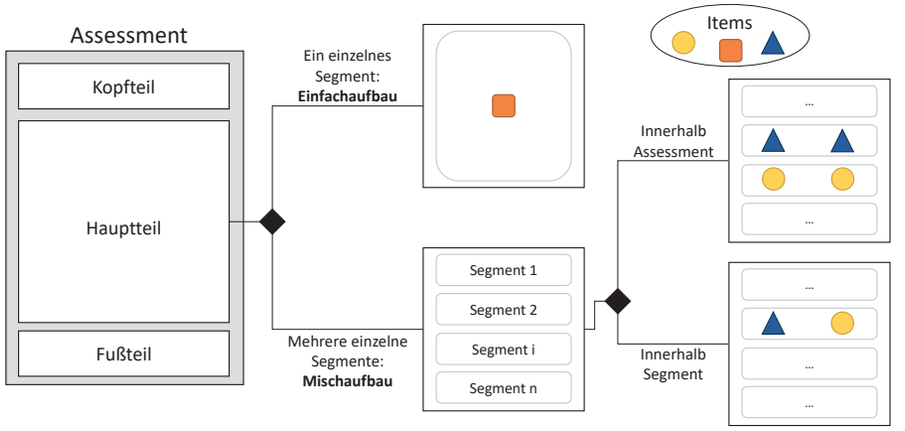


Abbildung 3: Schematischer Aufbau eines Assessments

koskalen) [3]. Im Rahmen dieser Untersuchung wurden drei Möglichkeiten der Scorebildung differenziert: keine Scorebildung, ungewichteter Amountscore und gewichteter Score. Nachfolgend werden diese möglichen Scorebildungen kurz umrissen und im Kapitel 3.1.2 Items durch konkrete Assessments veranschaulicht. Einzelne Instrumente ohne Scorebildung dienen als eine Art Leitfaden für die Informationsgewinnung. Sie sollen dabei helfen, Probleme zu verstehen (z.B. Erfassung und Dokumentation der Ernährungssituation) [3], oder dienen als Trendanzeige, um einen Verlauf festzuhalten. Beim ungewichteten Amountscore sind gegebene Auswahlmöglichkeiten gleichwertig und die Summe der getroffenen Auswahlen wird gebildet. Dahingegen sind gegebene Auswahlmöglichkeiten beim gewichteten Score unterschiedlich gewertet. Hierbei wird die Summe der einzelnen Wertigkeiten der getroffenen Auswahlen gebildet.

3.1.2 Items

Es ließen sich drei visuelle Erhebungsformen (=Items) mit einzelnen Unterformen (=Unteritems) aus den untersuchten Assessments extrahieren: gegebene Texte, grafische Elemente und Freitexteingaben. Auf diese einzelnen Items wird in den nachfolgenden Kapiteln näher eingegangen und es werden jeweils mögliche softwaretechnische Auswertemöglichkeiten gezeigt. Auf weitere Items, die sich im Rahmen der Untersuchung maximal zweimal zeigten, wird nicht näher eingegangen. Hier sei zum Beispiel eine prozentuale Skala in Zehnerschritten genannt, bei der diejenige Prozentzahl eingekreist werden sollte, die am besten die Schmerzlinderung durch Behandlungen zeigt.

3.1.2.1 Gegebene Texte

Bei diesem Item nimmt der Assessor einen kurzen Eintrag vor. Hier existieren die Unteritems Ja/Nein, Ticks und Direktauswahl. Eine Auswertung kann durch die Gabe der höchsten Punktzahl für die höchste Ausprägung erfolgen.

Bei der Reaktion auf gegebene Texte durch Ja/Nein wird der Aussage des Textes zugestimmt, oder die Aussage wird abgelehnt. Es sind immer mindestens diese zwei Optionen vorhanden. Zusätzlich kann ein „Weiß nicht“ als dritte Option verfügbar sein. Bei diesem Unteritem existieren Assessments, bei denen basierend auf der gegebenen Aussage, nur entsprechend „Ja“ oder „Nein“ ausgewählt wird und keine Scorebildung erfolgt. Dahingegen hat beim ungewichteten Amountscore jedes „Ja“, beziehungsweise jedes „Nein“ eine gleichwertige Gewichtung (siehe Abbildung 4). Hier gibt es auch Assessments mit der Bildung eines gewichteten Scores.

Bei der Reaktion auf gegebene Texte durch Setzen eines Kreuzchens oder Häkchens

Score 1 for every category and total at the bottom of the 2 columns		Yes	No
1	Is there a history of any fall in the previous year? **How assessed?** Ask the question.		
2	Is the patient/client on four or more medications per day? **How assessed?** Identify number of prescribed medications.		
	...		
Total			

Level of predicted risk

- 3 – 5 yes's = higher falls risk
  Refer to the falls team
- Less than 3 yes's = lower risk
  If the person has mobility problems consider referral to Community Therapy Services

Abbildung 4: Ungewichteter Amountscore bei Ja/Nein nach [7]

Fall Risk Score Calculation – Select the appropriate option in each category. Add all points to calculate Fall Risk Score. (If no option is selected, score for category is 0)	**Points**
Age (single-select) ☐ 60 – 69 years (1 point) ☐ 70 – 79 years (2 points) ☐ greater than or equal to 80 years (3 points)	
...	
Mobility (multi-select; choose all that apply and add points together) ☐ Requires assistance or supervision for mobility, transfer, or ambulation (2 points) ☐ Unsteady gait (2 points) ☐ Visual or auditory impairment affecting mobility (2 points)	
...	
Total Fall Risk Score (Sum of all points per category)	
SCORING: 6-13 Total Points = Moderate Fall Risk, >13 Total Points = High Fall Risk	

Abbildung 5: Gewichteter Score bei Ticks nach [8]

Risk Factor	Level	Risk Score
Recent Falls	none in last 12 months.....	2
	one or more between 3 and 12 months ago.....	4
	one or more in last 3 months.....	6
	one or more in last 3 months whilst inpatient / resident.....	8
	...	
(Low Risk: 5-11 Medium Risk: 12-15 High Risk: 16-20)	**RISK SCORE**	/20

Abbildung 6: Direktauswahl durch gewichteten Score nach [9]

(=Tick) sind variable Anzahlen an Unterpunkten vorhanden. Zudem ist entweder eine Einfachauswahl oder eine Mehrfachauswahl möglich. Auch hier gibt es Assessments, bei denen kein Score, ein ungewichteter Amountscore und ein gewichteter Score (siehe Abbildung 5) gebildet wird.

Bei der Reaktion auf gegebene Texte kann auch direkt der passende Text, oder der dem Text zugeordnete gewichtete Score (siehe Abbildung 6) ausgewählt werden.

3.1.2.2 Grafische Elemente

Als Unterbausteine werden bei den grafischen Elementen Skalen und Zeichnungen in den Assessments eingesetzt. Durch mathematische Operatoren können die Eingaben des Assessors gegenüber einer gegebenen Information beziehungsweise Referenz abgeglichen und dadurch ausgewertet werden.

Die in Abbildung 7 beispielhaft gezeigten Skalen werden für die Selbsteinschätzung des aktuellen Schmerzzustandes verwendet.

Dahingegen dienen Zeichnungen der Lokalisation (z.B. durch vorderseitige, rückseitige und seitliche Zeichnungen des Menschen) oder als Beurteilungshilfe bei Patienten ohne Kommunikationsfähigkeit (z.B. Deutung der Mimik um mögliche vorhandene Schmerzen beurteilen zu können). Zeichnungen werden auch als zusätzliche Erklärung zu einer Aufforderung oder einer Testdurchführung verwendet (z.B. bei Tests zur Mobilitätsbewertung, bei denen die vom Patienten durchzuführende Übung neben einer textuellen Beschreibung zusätzlich grafisch dargestellt wird).

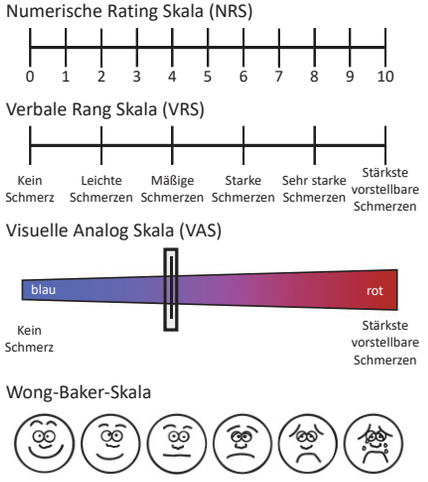


Abbildung 7: Skalen nach [2], [10]

3.1.2.3 Freitexteingaben

Freitexteingaben ermöglichen dem Assessor, einen beliebigen Text in einem dafür vorgesehenen Feld zu verfassen. Für eine Speicherung des Textes sind Datenbanken und für die Auswertung sind entsprechende Text Mining Algorithmen nötig.

Speziell bei Assessments zur Aufnahme psychischer Aspekte werden dem Patienten definierte Fragen gestellt (z.B. „What day of the week is this?“ [4]) und die gegebene Antwort wird festgehalten. Dabei wird überprüft und bewertet, ob die Antwort korrekt ist.

Durch Tests ermittelte und gemessene Werte werden auch durch eine Freitexteingabe festgehalten. Zudem ist eine Kategorisierung beziehungsweise Einstufung des Patienten möglich. So wird beim Timed Up and Go Test die benötigte Zeit für die Durchführung und Vollendung der spezifischen Aufgabe in Sekunden festgehalten und bewertet [6].

Bei Assessments für beispielsweise die Erhebung einer geringen Nahrungsaufnahme (z.B. aufgrund körperlich oder kognitiv bedingter Beeinträchtigung) kann aus einer Liste passender Gründe gewählt werden (z.B. a. Kognitive Überforderung, b. Funktionseinschränkungen der Arme oder Hände, c. Schlechter Zustand des Mundes [11]). Anschließend erfolgt durch eine Freitexteingabe eine Angabe weiterführender Maßnahmen. Somit werden Freitexte auch für eine weiterführende Erklärung oder Handlungsanweisung nach einer gegebenen Auswahl genutzt.

Daneben werden Freitexteingaben für Dokumentationszwecke verwendet. So wird z.B. eine Risiko- oder Krankheitshistorie des Patienten eingetragen, um etwa einen möglichen Trend verfolgen zu können.

3.1.3 Gesamtbetrachtung des Hauptteils

Wird der Gesamtaufbau von Assessments betrachtet, lassen sich diese nach Abbildung 3 in die Arten Einfachaufbau und Mischaufbau unterteilen. Beim Einfachaufbau wird im gesamten Hauptteil nur ein Item verwendet (z.B. in Abbildung 4 ausschnittsweise gezeigt). Dahingegen gibt es beim Mischaufbau mehrere hintereinanderliegende Segmente mit unterschiedlichen Modifikationen und Kombinationen der Items. Hier gibt es wiederum zwei Formen: Bei der ersten Form wird jedes Segment aus einem anderen Item aufgebaut, so dass die Segmente untereinander verglichen einen unterschiedlichen Aufbau haben. Bei der zweiten Form ist ein Segment aus unterschiedlichen Items aufgebaut. Bei beiden Arten können zusätzlich jeweils Abhängigkeiten untereinander existieren. Nach Abbildung 8 generiert jedes Segment einen Wert. Sind untereinander keine Abhängigkeiten vorhanden, fließt jeder Wert für sich in das Ergebnis des Assessments (etwa in einen Score zur Einstufung des Patienten oder eine Pflegeanweisung). Bei existierenden Abhängigkeiten ist der Output des einen Segments der Input des nachfolgenden Segments (z.B. wenn zuerst ein Screening gefordert ist und auf dessen Ergebnis das Assessment fortgeführt werden muss).

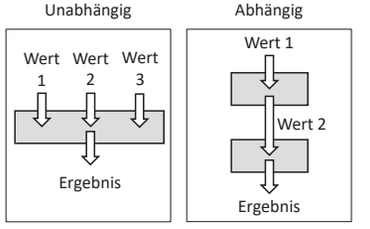


Abbildung 8: Abhängigkeiten innerhalb eines Assessments

3.2 Abbildung auf Skalenniveaus

In der Statistik lassen sich untersuchte Eigenschaften (Merkmale) auf entsprechende Skalenniveaus abbilden [12]. Somit werden Auskünfte über das Informationsniveau, den Aussagegehalt und die erlaubten Rechenoperationen der Merkmalsausprägungen der jeweiligen Merkmale gegeben [12], [13]. In den nachfolgenden Kapiteln werden die einzelnen Skalenniveaus kurz beleuchtet und es erfolgt eine Abbildung der Items auf diese Skalenniveaus, die somit als Grundlage für die Softwaremodellierung dienen.

3.2.1 Nominalskala

Dieses Skalenniveau ordnet Merkmale begrifflichen Kategorien zu [12]. Hierbei ist die Verwendung der Ausprägungen für bestimmte Rechenoperationen nicht anwendbar, da als Skalenwerte lediglich Namen abgetragen sind [13], [14]. Als Beispiel sei hier das Geschlecht eines Patienten genannt [12].

3.2.2 Ordinalskala

Durch eine Ordinalskala können Merkmalsausprägungen in auf- oder absteigender Folge geordnet werden [13]. Hierbei sind arithmetische Operationen nicht zulässig, da die jeweilige „Größe“ und die jeweiligen Abstände zwischen den Ausprägungen nur eine begrenzte Aussagekraft haben, beziehungsweise nicht interpretierbar sind. Nach Schäfer et al. [12] wird z.B. das Maß der sportlichen Betätigung (nie, selten, manchmal, oft, sehr oft) der Ordinalskala zugeordnet. [14]

3.2.3 Intervallskala

Zur Kardinalskala (auch metrische Skala genannt) gehört die Intervallskala. Hier lassen sich bei Merkmalen ohne natürlichen Nullpunkt Aussagen über die Abstände (Intervalle) zwischen den einzelnen Ausprägungen machen. Beispielsweise hat die Körpertemperatur (Grad Celsius) einen willkürlich gewählten Nullpunkt. [12], [14]

3.2.4 Verhältnisskala

Der Kardinalskala zuordenbar ist auch die Verhältnisskala. Hier können Aussagen über Relationen zwischen Merkmalsausprägungen gemacht werden, da im Gegensatz zur Intervallskala ein eindeutig definierter natürlicher absoluter Nullpunkt existiert. Ein Beispiel ist hier das Alter in Jahren. [13], [14]

3.2.5 Auswertung

Nach Tabelle 1 kann jeder der im Kapitel 3.1.2 Items beschriebenen Items durch inhaltliche Betrachtungen der recherchierten Assessments mindestens einer Skala zugeordnet werden. Für die Zuordnung wurden die möglichen Informationen genutzt, die durch die jeweiligen Items erfasst und überliefert werden können. Es kristallisierte sich heraus, dass einzelne Items durch die Zuordnung mehrere Skalenniveaus abdecken. So gibt es beim Unteritem Zeichnungen zwei Zuordnungen: Bei der Nominalskala gibt es Zeichnungen bei denen Eintragungen (z.B. die Schmerzposition) vorgenommen und die Ausprägungen lediglich kategorisierbar sind. Andere Zeichnungen lassen sich der Verhältnisskala zuordnen. Beispielsweise lassen sich beim Uhren-Zeichen-Test nach Watson (zeichnen einer Uhr in einen gegebenen Rahmen) Aussagen über Relationen zwischen den Ausprägungen machen. Hierfür sind entsprechende mathematische Operationen notwendig.

Tabelle 1. Abbildung der Items auf Skalenniveaus mit Einordnung in Datentypen

	Nominalskala	Ordinalskala	Intervallskala	Verhältnisskala
Gegebene Texte				
Ja/Nein	Trifft Aussage zu *(Boolean, Structure)*			
Ticks Ein Unterpunkt	Trifft Aussage zu *(Boolean)*			
Ticks Mehrere Unterpunkte	Auswahl *(List<String>)*	Wie sehr trifft Aussage zu *(List<String>)*	Auswahl *(List<Int>)*	Mengenangabe *(List<Int>)* Auswahl *(List<Int>)*
Direktauswahl	Auswahl *(List<String>)*	Wie sehr trifft Aussage zu *(List<String>)*		Mengenangabe *(List<Int>)* Auswahl *(List<Int>)* Messwert *(List<Int>)*
Grafische Elemente				
Skalen		Zustandsbewertung *(List<String>, List<Image>)*	Zustandsbewertung *(List<Int>)*	
Zeichnungen	Eintragung/Markierung *(Image + Auswertung Int[,])* Erklärung/Beurteilungshilfe *(Image)*			Zeichnungsbewertung *(Image + Auswertung Int[,])*
Freitexteingaben				
True/False	Stimmt Aussage *(String + Auswertung Boolean)*			
Messwerte	Werteintragung *(Integer)*		Werteintragung *(Integer)*	Werteintragung *(Integer)*
Nach Auswahl	Zuordnung und Erläuterung *(List<String>)*			
Dokumentation	Patienteninformationen *(String)*	Klinikinformationen *(Integer)*	Patienteninformationen *(Integer)*	Informationen *(Integer)*

3.3 *Transfer in Softwaremodellierung*

Damit die einzelnen Auswertearithmetiken, Items und Zuordnungen in die Softwaremodellierung überführt werden können, ist als zentraler Schritt der Entwurf von Konstrukten wie Klassen und Interfaces nötig. Hierfür können entsprechende Modellierungen verwendet werden (z.B. mittels UML), um das entworfene Modell softwaretechnisch abzubilden. Dabei lassen sich für die einzelnen Items unterschiedliche softwaretechnische Datentypen in Abhängigkeit vom abgebildeten Skalenniveau auswählen. Hierfür müssen die jeweiligen Ausprägungen der Zuordnungen und die nötigen Operationen für die entsprechenden Auswertungen berücksichtigt werden. Mögliche Datentypen sind in Tabelle 1 in kursiver Schrift nach den jeweiligen Zuordnungen skizziert. Beispielsweise wäre beim Unteritem True/False bei der Zuordnung *stimmt Aussage* neben einem String zur Eingabe auch ein Boolean für die Auswertung der Eingabe denkbar.

Zudem ist nun am Beispiel der Auswertearithmetik „gewichteter Score“ nach Abbildung 9 eine Implementierung der Auswertung der einzelnen Itemklassen und anschließenden Einstufung losgelöst vom User Interface möglich. Durch die Abbildung der Items auf Skalenniveaus ergeben sich die möglichen User Interface Controls für die jeweiligen Items. Somit kann für eine dem Item zugeordneten Frage ein beliebiges der diesem Item zugehörigen Controls für die Darstellung im User Interface verwendet werden. Beispielsweise sind die möglichen User Interface Controls der Unteritems Ja/Nein, Ticks und Direktauswahl beliebig austauschbar und nicht mehr verbindlich, wenn von der zugehörigen übergeordneten Itemklasse gegebene Texte als Eingabeparameter unter anderem ein Set an Booleans gefordert wird. Dies spiegelt sich auch bei Merkmalen wider, die auf eine Verhältnisskala zuordenbar sind. So kann „Alter in Jahren“ (z.B. <60, 60-70, >70) als Ausprägung bei der Zuordnung *Auswahl* beim Unteritem Direktauswahl ein Boolean genutzt werden, da der Patient nicht

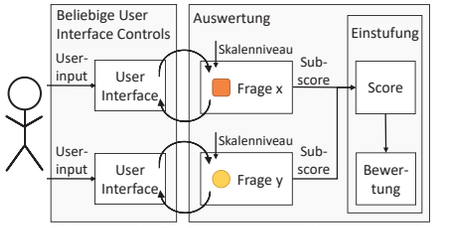


Abbildung 9: Softwaremodellierung bei Auswertearithmetik „gewichteter Score“

zur gleichen Zeit verschiedene Alter aufweisen kann. Hierfür können zusätzliche Metainformationen in die Skalenniveaus eingebettet werden, um neben den möglichen Wertebereichen auch die Einfach- und Mehrfachauswahl gegebener Texte zu regeln.

Im Rahmen dieser Ausarbeitung wird auf diesen Aspekt nicht näher eingegangen. Die Untersuchung und Abstraktion wird mit der nachfolgenden Betrachtung des Gesamtmodells geschlossen.

3.4 *Gesamtbetrachtung des Modells*

Das generierte Abstraktionsmodell kann aus Abbildung 10 entnommen werden. Als erste Stufe wird die Betrachtung des Aufbaus von Assessments angesehen. Hier wird das Assessment durch einzelne Items und unterschiedliche Segmenttypen in der Gesamtbetrachtung des Hauptteils beschrieben. Darauf aufbauend ergibt sich die Abbildung auf Skalenniveaus, woraus Informationsniveaus, Aussagegehalte und Rechenoperationen abgeleitet werden können. Als letzte Stufe wird der Transfer in die Softwaremodellierung angesehen.

Das gesamte Modell weist einen multivariaten Charakter auf. So wird bei klinischen Assessments ein gewünschter Zustand oder ein Risiko des Patienten als Zielgröße erfasst, indem die Zielgröße als Funktion von erklärenden Größen dargestellt wird. Beispielsweise wird das Dekubitusrisiko als Zielgröße durch die Erfassung von erklärenden Größen wie Alter, Hautzustand, Zusatzerkrankungen und Aktivität dargestellt. Durch die kausalen Zusammenhänge der erklärenden Größen und gegebenenfalls anschließenden Berechnungen und Gewichtungen kann somit das Risiko ermittelt werden.

4 Fazit und Ausblick

Für den Bereich der klinischen Assessments wurde durch deren Abbildung auf statistische Skalenniveaus eine gemeinsame Basis definiert. Hier erfolgte durch die Abstraktion eine geprüfte Zuordnung mit Transfer in die Softwaremodellierung, wobei eine software-

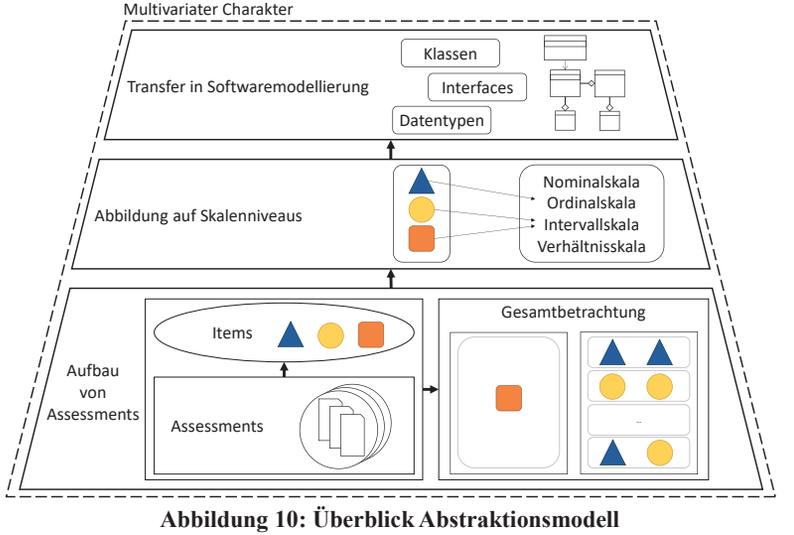


Abbildung 10: Überblick Abstraktionsmodell

technische Realisierung einer Auswertelogik losgelöst vom User Interface ermöglicht wurde. Es lässt sich schließen, dass neu eingehende klinische Assessments durch die Abstraktion auf eine gemeinsame Basis in Form der Skalenniveaus eindeutig zugeordnet und in die Softwaremodellierung transferiert werden können. Basierend auf der gesamten Untersuchung könnte ein *Service-Tool* entwickelt werden. Durch dieses könnten zukünftige Assessments einheitlich etwa per Drag-and-Drop aus gegebenen Items zusammengestellt werden. Hierbei können die Skalenniveaus beispielsweise für die Navigation des Users während einer Assessmentzusammenstellung dienen. Zudem könnte eine Einbettung und Auswertung der aus den Assessments gewonnenen Informationen in bestehende Patientenüberwachungssysteme erfolgen. Hierbei wären die Informationen vor Ort für das Klinikpersonal stets präsent und könnten direkt nachhaltig in die Pflege und Behandlung des Patienten einfließen.

Literaturverzeichnis

[1] M. Wirz, B. Köhler, D. Marks et al., Lehrbuch Assessments in der Rehabilitation, 1. Auflage, Bern: Hans Huber Hogrefe, 2014.

[2] B. Reuschenbach und C. Mahler, Pflegebezogene Assessmentinstrumente: Internationales Handbuch für Pflegeforschung und –praxis, 1.Auflage, Bern: Verlag Hans Huber Hogrefe AG, 2011.

[3] S. Bartholomeyczik und M. Halek, Assessmentinstrumente in der Pflege: Möglichkeiten und Grenzen, 2. Auflage, Hannover: Schlütersche Verlagsgesellschaft mbH & Co. KG, 2009.

[4] I. McDowell, Measuring Health: A Guide to Rating Scales and Questionnaires, Third Edition, New York: Oxford University Press, 2006.

[5] A. Gupta, Assessmentinstrumente für alte Menschen: Pflege- und Versorgungsbedarf systematisch einschätzen, 1. Auflage, Bern: Hans Huber, 2012.

[6] H. Freund, Geriatrisches Assessment und Testverfahren: Grundbegriffe, Anleitungen, Behandlungspfade, 2. Auflage, Stuttgart: Kohlhammer, 2014.

[7] „North West Ambulance Service“. [Online]. Available: https://www.nwas.nhs.uk/media/458865/frat.pdf. [Zugriff am 07. September 2018].

[8] „Johns Hopkins Medicine,“ 2007. [Online]. Available: https://www.hopkinsmedicine.org/institute_nursing/models_tools/JHFRAT_acute%20care%20original_6_22_17.pdf. [Zugriff am 28. September 2018].

[9] Department of Health & Human Services, „Falls Risk Assessment Tool (FRAT),“ 01. Januar 1999. [Online]. Available: https://www2.health.vic.gov.au/about/publications/policiesandguidelines/falls-risk-assessment-tool. [Zugriff am 07. September 2018].

[10] „Deutsches Netzwerk für Qualitätsentwicklung in der Pflege DNQP“. [Online]. Available: https://www.dnqp.de/fileadmin/HSOS/Homepages/DNQP/Dateien/Expertenstandards/Schmerzmanagement_in_der_Pflege_bei_akuten_Schmerzen/Schmerz_NRS_VRS_ VAS.pdf [Zugriff am 14. September 2018].

[11] „Fresenius Kabi“. [Online]. Available: https://m.fresenius-kabi.de/files/digitalFormular_Assessment_Nahrungsmangel_PEMU_2015_09.pdf. [Zugriff am 14. September 2018].

[12] A. Schäfer und T. Schöttker-Königer, Statistik und quantitative Methoden für Gesundheitsfachberufe, Berlin: Springer-Verlag, 2015.

[13] G. Bourier, Beschreibende Statistik - Praxisorientierte Einführung mit Aufgaben und Lösungen, 13. Auflage, Wiesbaden: Springer Gabler, 2018.

[14] A. Kuß, R. Wildner und H. Kreis, Marktforschung: Grundlagen der Datenerhebung und Datenanalyse, 5. Auflage, Wiesbaden: Springer Gabler, 2014.

Entwicklung eines Low Fidelity Prototyps für die Erforschung von Akzeptanzkriterien für Sprachassistenzsysteme

Emanuel Geiger
Hochschule Reutlingen
Emanuel.Geiger@Student.Reutlingen-University.DE

Abstract

Durch die Weiterentwicklung von Sprachassistentsystemen werden diese vermehrt in den Wohnbereichen der Anwender eingesetzt. Während vor einigen Jahren noch zunehmend die jüngere Generation im Blickfeld der Entwickler war, hat sich dies verändert. Nun stehen ältere und körperlich beeinträchtigte Menschen im Fokus der Entwickler. Obwohl dies die zukünftige Zielgruppe ist, gibt es wenig Kriterien oder Wissen über die Zielgruppe und welche Meinungen diese zu Chatbots und Sprachsystemanwendungen haben. Es gibt im Moment wenig Wissen über Akzeptanzkriterien hinsichtlich künstlicher Intelligenz. In dieser Arbeit wurden zwei verschiedene prototypische Sprachsystemanwendungen entwickelt, welche in einer Umfrage evaluiert wurden. Das Ziel der Umfrage war es herauszufinden, welche Aspekte für die Zielgruppe bei der Benutzung von Sprachsystemanwendungen interessant sind.

Schlüsselwörter

Akzeptanzkriterien, Sprachassistenzsysteme, Prototyping, Datenerhebung, Zielgruppenbefragung, Chatbots, Evaluation

CSS-Concepts

• **Human-centered computing~User models**
• Human-centered computing~Usability testing

1 Einleitung

Mit einem Sprachassistenten kann ein Anwender, in einer natürlichen Art und Weise durch eine sprachliche Interaktion mit einem System kommunizieren [1], [2]. In dieser Arbeit werden die Sprachassistenzsysteme und deren Anwendungen innerhalb eines Smart Homes betrachtet. Durch die Einbeziehung der Zielgruppe der älteren Menschen und körperlich beeinträchtigten Personen hat sich das Kauf- und Nutzerverhalten in diesem Bereich geändert, was dazu geführt hat, dass mehr in diesem Gebiet geforscht wird [3], [4]. Es gibt heutzutage sehr wenig Wissen, wie ein Sprachassistent eingesetzt wird. Dies liegt daran, dass es bisher noch sehr wenige Studien über Sprachassistentsysteme gibt [5], [6]. So gibt es auch sehr wenig Wissen über die Benutzerfreundlichkeit, die User Experience und den Nutzen von solchen Systemen [1]. Um einen ersten Überblick über die Nutzung von solchen Geräten zu erlangen, wurde

Betreuer Hochschule: Prof. Dr. Natividad Martinez
Hochschule Reutlingen
Natividad.Martinez@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen

Copyright 2018 Emanuel Geiger

in [5] die Log Daten von 88 Google Geräten ausgewertet.

Ein Nachteil von solchen Geräten wird aus der Sicht des Anwenders darin gesehen, dass dieser für die Interaktion ein mobiles Endgerät besitzen muss [5]. Gleichzeitig ist ein solches System einfach zu benutzen [1]. Die einfache Benutzung ist jedoch nicht nur ein Kriterium, das für den Bereich von Sprachsystemen wichtig ist, sondern ist ein elementarer Bestandteil der Informationstechnologie. Um zu überprüfen, ob die beschriebenen allgemeinen Kriterien auf eine bestimmte Zielgruppe zutreffen, wurden zwei Prototypen entworfen. Diese wurden mit dem Amazon Echo System entwickelt, da diese Hardware, die in den USA die meistinstallierte[1], in einem Smarthome einsetzbar und der Anwendungsfall demonstrierbar ist. Da die in dieser Arbeit entwickelten Prototypen auch in sozialen Einrichtungen getestet wurden, mussten verschiedene Kriterien berücksichtigt werden. Da eine Google Lösung aufgrund der Datenschutzaspekte von vornherein ausgeschlossen werden konnte, wurde sich in einem ersten Schritt für die Entwicklung der Prototypen mit der Amazon Variante entschieden. Mit den Prototypen wurden die Akzeptanzkriterien und auch weitere sprachliche Besonderheiten anhand einer Umfrage erforscht.

Diese Arbeit ist folgendermaßen strukturiert: Im zweiten Kapitel werden die verschiedenen Faktoren und Kriterien, die bereits in der Literatur vorhanden sind, vorgestellt und auf die Zielgruppe, auf die diese Arbeit ausgerichtet ist, eingegangen. Anschließend wird im dritten Kapitel die Methode beschrieben. In diesem werden auch der Versuchsaufbau und die Durchführung aufgeführt. Im vierten Kapitel werden die Ergebnisse der Umfrage dargestellt. Diese werden in dem Kapitel „Diskussion“ aufgegriffen. Die Arbeit wird mit dem Fazit abgerundet.

[1] https://www.statista.com/statistics/898813/us-smart-speaker-installed-base-by-device/ (Stand 01.11.2018)

2 Aktueller Wissensstand

In diesem Kapitel wird der aktuelle Forschungsstand zu den Akzeptanzkriterien von Sprachassistentsystemen und Chatbots anhand einer Literaturrecherche aufgeführt. In einem zweiten Teil dieses Kapitels wird ausgehend von einer Literaturrecherche das Alter der Umfrageteilnehmer definiert. Die Akzeptanzkriterien, die bereits in der Literatur vorhanden sind, orientieren sich entweder an einer jüngeren Zielgruppe, sind allgemein gehalten oder betrachten die Interaktionsmöglichkeiten mit einem Computer. Im Rahmen der vorliegenden Arbeit wurde überprüft, ob die nachfolgend vorgestellten Kriterien für eine bestimmte Zielgruppe einsetzbar sind oder ob hier etwaige Erweiterungen oder Ergänzungen durchgeführt werden müssen.

2.1 Analyse der Zielgruppe

Ein Hauptaugenmerk in dieser Arbeit liegt auf ältere und/oder körperlich beeinträchtigte Personen. Die Möglichkeit, Menschen mit Einschränkungen unbeschränkt Zugang zu allen Bereichen zu gewähren, wurde bereits in der Gleichstellungsregel der United Nations festgeschrieben [9]. Um für die Arbeit eine Alterseinstufung vornehmen zu können, wurde auf mehrere Studien zurückgegriffen, die sich mit dem Alter der Anwender beschäftigen. In [10] wurde eine Umfrage mit 1164 befragten Personen durchgeführt, um zu evaluieren, ob sich die potentiellen Anwender die Kommunikation mit einem Chatbot vorstellen können. In [11] wurde herausgefunden, dass 41% der über 65-jährigen Personen ein Smartphone benutzen und somit in der Lage sind, ein Sprachassistent zu nutzen, da für die Nutzung eines Sprachassistenten ein Smartphone eine Voraussetzung darstellt. Hierzu wurde eine Umfrage mit 1006 Personen durchgeführt. In [10] wurden die Umfrageteilnehmer in die folgenden Altersklassen eingeteilt: 18 bis 24 Jahre, 25 bis 34 Jahre, 35 bis 44 Jahre, 45 bis 54 Jahre und alle Personen, die älter als 55 Jahre sind. Ausgehend von diesen beiden Umfragen wurden in dieser Arbeit differenziert Personen ab einem Alter von fünfundfünfzig Jahren und deren Einstellungen zu Sprachassistentsystemen und künstlicher Intelligenz untersucht. Besondere Eigenschaften, die diese aufweisen mussten, waren eine körperliche Beeinträchtigung, wie beispielsweise das Tragen einer Brille oder schlechtes Gehen. Eine Auswirkung von Beeinträchtigungen können beispielsweise Einschränkungen im Bereich der Lesefähigkeiten oder der Benutzung der Hände sein [8]. In diesem Kontext sind auch Situationen gemeint, in der der Anwender zwar grundsätzlich in der Lage ist, seine Hände zu benutzen, aber keine Hand frei hat, um mit einem bestimmten Objekt zu interagieren [12]. In solch einem Fall kann der Nutzer gestengesteuerte Interaktionsmöglichkeiten nicht anwenden [12].

Durch die in dieser Arbeit entwickelten Prototypen sollen Personen ab fünfundfünfzig Jahren angesprochen werden, die aufgrund von verschiedenen Faktoren, wie Krankheit, Einsamkeit, Unsicherheiten oder beispielsweise einer demenziellen Erkrankung und andere altersbedingte Faktoren, unterstützt werden. Viele ältere Personen leben alleine oder sind einsam [13]. In [13] wurde der Begriff der Einsamkeit und der sozialen Isolation, insbesondere bei älteren Menschen, wissenschaftlich definiert und anhand verschiedener Studien betrachtet. Durch Sprachassistenten existiert eine Möglichkeit, wie Unsicherheiten beseitigt werden können und wie diese Gruppe trotz möglicher Probleme die alltäglichen Aktivitäten aufnehmen und beibehalten können. Zu den alltäglichen Aktivitäten gehören beispielsweise Essen, Körperpflege und sich zu kleiden [14].

2.2 Analyse der Akzeptanzkriterien

Bisher wurden Chatbots und Sprachassistenzsysteme aus einem allgemeinen Blickwinkel betrachtet und keine zielgruppenspezifische Analyse durchgeführt. Ein Vorteil eines Sprachassistenten liegt in der intuitiven Bedienung [1], [2]. Um dies gewährleisten zu können, muss die Bedienung einfach [8] und logisch aufgebaut [15] sein. Durch die sprachliche Interaktion mit den Geräten kann der Anwender auf natürliche Art und Weise mit dem System kommunizieren [1] [2]. Bei der Kommunikation mit einem Sprachassistenten tritt heutzutage immer wieder das Problem auf, dass aufgrund von verschiedenen akustischen Schwierigkeiten das System die Stimme des Anwenders nicht versteht [1] [2]. Dieses Problem lässt sich auch durch fortwährendes Training der Spracherkennung von allen gängigen Sprachsystemen nicht lösen [2]. Durch das Missverstehen des Anwenders kann die Usability sinken. Ein weiteres Problem ist, dass die meisten Sprachassistentsysteme heutzutage nicht in der Lage sind, gleichzeitig auf verschiedene Befehle von unterschiedlichen Anwendern zu reagieren [2]. In [1] wurde herausgefunden, dass die Benutzung eines Sprachassistenten für den Anwender ein aufregendes Erlebnis und eine effiziente Art der Interaktion darstellt.

Als weiterer Faktor, der bei der Entwicklung jeder Software oder Hardware berücksichtigt werden muss, ist, das Produkt so zu entwerfen, dass dieses für den Anwender einen Mehrwert darstellt [16]. Um einen Mehrwert analysieren zu können, müssen in einem ersten Schritt die verschiedenen Bedürfnisse der Anwender evaluiert werden. So wurde in [5] genannt, dass viele Nutzer eines Sprachassistenten von diesem enttäuscht sind, da sie eigentlich andere Funktionalitäten erwarteten und diese aus technischer Sicht nicht möglich sind. Wenn ein Anwender sich mit einem neuen Tool beschäftigt, kann dieser hinsichtlich der Funktionsweise und den damit verbundenen Möglichkeiten irritiert sein [17]. Um darum den Einstieg mit einem Sprachassistenten zu vereinfachen, wurde am Anfang der Anwendungen eine kleine Anleitung entwickelt, die den Anwender durch das Programm führen und ihm somit den Einstieg in diese Technologie erleichtern soll. Ein weiterer wichtiger Aspekt sind die Faktoren Si-

cherheit und Kontrolle, die bei der Interaktion mit einem solchen System eine wichtige Rolle spielen [4]. Ein solches System kann durch Hintergrundgeräusche verschiedene Befehle falsch verstehen, was dazu führt, dass das Gerät entgegen dem Sinn des Anwenders handelt. Ein Sprachassistent kann für delegative als auch kooperative Aufgaben verwendet werden [7]. Unter der Rubrik „Delegation“ wird ein System verstanden, das eine Entscheidung treffen oder aus einer Vielzahl von Optionen das beste Ergebnis heraussuchen kann [7]. Während mit dem kooperativen Aspekt die Unterstützung von bestimmten Benutzergruppen gemeint ist. Diese Begrifflichkeit lässt sich mit dem Begriff „assistive technology“ [7] gleichsetzen. Hierbei wird beispielsweise blinden Menschen mit einem Assistenzsystem ein Überblick über ihre Umgebung ermöglicht [7]. Der Dialog und insbesondere die Antworten, die das System liefert, müssen für den Anwender transparent sein [18]. Die beschriebene Transparenz ist ein allgemeines Kriterium eines jeden Softwareproduktes. In Zusammenhang mit Sprachassistenten bedeutet dies, dass der Anwender die richtigen Antworten zu der richtigen Zeit bekommt. Ein allgemeines Kriterium, welches bei der Entwicklung eines Softwareproduktes berücksichtigt werden muss, ist die dynamische Gestaltung der Software. Dies bedeutet, dass das System spezifisch auf die Wünsche der Anwender reagieren und sich an deren Gewohnheiten anpassen muss. Somit müssen beispielsweise standardisierte Phrasen und Antwortmöglichkeiten minimiert oder am besten ausgeschlossen werden.

3 Methode

In diesem Kapitel wird auf die entwickelten Prototypen und den Versuchsaufbau eingegangen. Im Rahmen dieser Arbeit wurden zwei verschiedene Prototypen entwickelt. Diese wurden so konzipiert, dass die sprachliche Interaktion des Menschen mit der Maschine getestet werden konnte. In den Unterpunkten dieses Kapitels werden die zwei Prototypen jeweils kurz skizziert und anschließend der Versuchsaufbau beschrieben.

3.1 Anwendungsfall Terminerinnerung

In diesem Anwendungsfall geht es darum, den jeweiligen Anwendern eine dynamische Terminerinnerungsfunktion zur Verfügung zu stellen, die es ihnen ermöglicht, in einem beliebigen zeitlichen Abstand vor einem Ereignis erinnert zu werden. Eine solche Funktion gibt es bei den Sprachassistenten, wie Amazon Alexa und Google Home noch nicht, da die Anwendungen dieser Geräte nicht individuell angepasst werden können [2]. Bei dem Sprachassistent von Amazon ist es derzeit nur möglich, sich den gesamten Terminverlauf eines Tages anzuhören oder es auch so einzustellen, dass der gesamte Terminverlauf an einem bestimmten wiederkehrenden Zeitpunkt vorgelesen wird. Es ist jedoch nicht möglich, dass das System den Anwender per Sprache an einen Kalendertermin erinnert. Bei dem Amazon Gerät ist lediglich bei der Rubrik „Erinnerung“ eine Warnung mit einem Signalton und einer kurzen Nachricht möglich, während das Google Home System schon eine Terminerinnerung, durch ein Abrufen der Termine des entsprechenden Smartphones beherrscht.[2] Bei beiden Systemen ist es derzeit nicht möglich, dass der Anwender die Termine mit Hilfe der Stimme ändern oder löschen kann[3]. Derzeit ist lediglich die Eingabe der Termine über die Stimme möglich. So können kurzfristige Erinnerungen bereits abgebildet werden, während dies bei langfristigen noch ein Problem darstellt. Das Ziel dieses Prototyps war es, den Anwendungsfall einer Terminerinnerung zu entwickeln, die sowohl die Ausgabe des entsprechenden Termins als auch die Eingabe über dasselbe Sprachsystem ermöglicht. Der Anwendungsfall sollte die Kalenderfunktion eines mobilen Endgerätes widerspiegeln. So gibt es bei den meisten mobilen Endgeräten einen Kalender, bei dem eingestellt werden kann, dass der Anwender in einem gewissen Zeitabstand vor einem Ereignis informiert wird.

3.2 Anwendungsfall Unterstützung

Das Ziel dieses Anwendungsfalls ist es, Menschen mit einer körperlichen Beeinträchtigung eine Unterstützung in den auditiven, visuellen und sprachlichen Bereichen zu geben. In dieser Arbeit wurde eine prototypische Anwendung für eine auditive Unterstützung entwickelt, welche nachfolgend kurz erläutert wird.

Eine auditive Unterstützung könnte so gestalten werden, dass in einem ersten Schritt der Anwender die Möglichkeit hat, das System lauter oder leiser zu stellen. Im Endeffekt könnte dies dahingehend entwickelt werden, dass das verwendete Sprachsystem so erweitert wird, dass ein mobiler Beamer mit dem Gerät verbunden wird und dem Anwender verschiedene Bereiche, wie zum Beispiel das Gesagte, visualisiert werden.

3.3 Beschreibung des Versuchsaufbaus

In diesem Abschnitt werden die befragten Personen vorgestellt und aufgezeigt, wie der Versuch und die Umfrage aufgebaut waren. Um die Prototypen und die Ideen zu verifizieren, wurden Experten und Anwender befragt. Da es das Ziel dieser Arbeit ist, das Verhalten älterer Menschen gegenüber einem Sprachassistenten zu evaluieren, wurden Personen in sozialen Einrichtungen und in den eigenen vier Wänden befragt. Durch die Experteninterviews konnte ein erster Überblick über die Zielgruppe gewonnen werden und so schon verschiedene wichtige weitere Fragestellungen in die Umfrage mit aufgenommen werden. Insgesamt wurden im Rahmen dieser Arbeit vier Experten aus dem sozialen Umfeld und ein IT-Experte befragt. Aus diesen Bereichen konnte mit Hilfe der durchgeführten Umfrage jeweils unterschiedliche Ansatzpunkte für die Anwenderbefragung in Erfahrung gebracht werden. Nach der Datenerhebung und -auswertung der Expertenumfrage wurden fünfzehn potentielle Anwender unterschiedlichen Alters und Geschlechts befragt. Die Zusammensetzung der geschlechter- und altersspezifischen Merkmale wird in Tabelle 1 aufgezeigt

Tabelle 1 Geschlechter- und altersspezifische Verteilung der Anwender

Alter	Geschlecht	Anzahl
55-64 Jahre[4]	Männlich	2
65-74 Jahre[4]	Weiblich	1
65-74 Jahre[4]	Männlich	3
75-84 Jahre[4]	Weiblich	2
75-84 Jahre[4]	Männlich	2
85-94 Jahre[4]	Weiblich	4
85-94 Jahre[4]	Männlich	1

Diese Alterseinteilung wurde vorgenommen, um mögliche Interessensveränderungen der Zielgruppen herauszufinden und die Umfrage wissenschaftlich auswerten zu können. In einem weiteren Schritt wurden die Daten geschlechtsspezifischen Attributen zugeordnet. So wurden die Umfragen mit acht männlichen und sieben weiblichen Personen durchgeführt. Um den Befragten jeweils ein angenehmes Klima während des Versuchs zu ermöglichen, wurde der Versuch immer an Orten durchgeführt, an denen sich die Perso-

[2] https://updated.de/ratgeber/mit-google-home-durch-den-tag-so-nutzen-sie-kalender-und-erinnerungsfunktionen-30770/ (Stand 13.10.2018)

[3] Getestet am 03.11.2018

[4] Bis zu Vollendung des jeweiligen Lebensjahres

nen wohlfühlten, damit diese sich den Einsatz eines Sprachassistenten im Smart Home Bereich vorstellen konnten. In einem ersten Schritt wurden die Ziele der Umfrage erläutert. Die Anwendung wurde in der Testumgebung vorgestellt. Bei dieser Umfrage ging es in erster Linie darum, wie der Anwender auf die Stimme des Systems reagiert. Als dritter Aspekt wurde die Umfrage durchgeführt. Insgesamt wurden vierzig verschiedene Fragen zu unterschiedlichen Bereichen gestellt, die sowohl offene als auch geschlossene Fragen beinhalteten. So wurden verschiedene Fragen zu der Interaktionsmöglichkeit mit einem Sprachassistenten gestellt, verschiedene sprachliche Besonderheiten und Wünsche der Anwender erfragt und die verschiedenen möglichen Anwendungsfälle der Anwender evaluiert.

4 Ergebnisse der Umfrage

In diesem Kapitel werden zum einen die Probleme und der subjektive Schwierigkeitsgrad, bei der Benutzung eines Sprachassistenten grafisch dargestellt. Je nachdem, ob ein Experte oder ein Anwender befragt wurde, wurden andere Schwerpunkte bei den Antworten gelegt. Nach der Aussage einer Expertin und um die Unbefangenheit der potentiellen Anwender in Bezug zu einem solchen System zu gewährleisten, wurde vorher auf weitere Informationen, wie Hinweise auf mögliche datenschutzrechtliche Risiken in Verbindung mit künstlicher Intelligenz, verzichtet. Auf diese konnte auch ohne Probleme verzichtet werden, da für den Versuch und die damit zusammenhängende Dateneingabe lediglich Beispieldaten verwendet wurden und somit keine zurück verfolgbaren Datensätze, mit deren Hilfe sich Informationen oder eine Meinung des Anwenders abbilden lassen könnte, erhoben wurden. Eine Fragestellung bezog sich auf den subjektiven Schwierigkeitsgrad bei der Bearbeitung der Aufgabe. Die genaue Fragestellung war: „Wie fanden Sie die Bearbeitung der Aufgabe auf den Schwierigkeitsgrad bezogen?“ Hierbei wurden den Anwendern in einem ersten Schritt die Interaktionsmöglichkeiten vorgestellt und in einem weiteren Schritt hatten sie die Möglichkeit die Interaktion mit einem solchen Gerät auszuprobieren. Danach wurden ihnen die in Abbildung 1 genannte Frage gestellt. Die befragten Personen hatten fünf verschiedene Antwortmöglichkeiten „Sehr leicht, Leicht, Mittel, Schwer und Sehr schwer“. Die Ergebnisse dieser Frage wird in Abbildung 1 dargestellt. Mit Hilfe dieser Frage konnte ein erster Blick auf die Möglichkeit der sprachlichen Interaktion geworfen werden. Von den Befragten fanden 80 % der Personen den Schwierigkeitsgrad als „Leicht“ bis „Mittel“. Bei dieser Frage stellt die Rubrik „Sehr schwer“ einen Ausreißer, im Vergleich zu den anderen Antworten dar, da diese nur 13% der gesamten Antworten

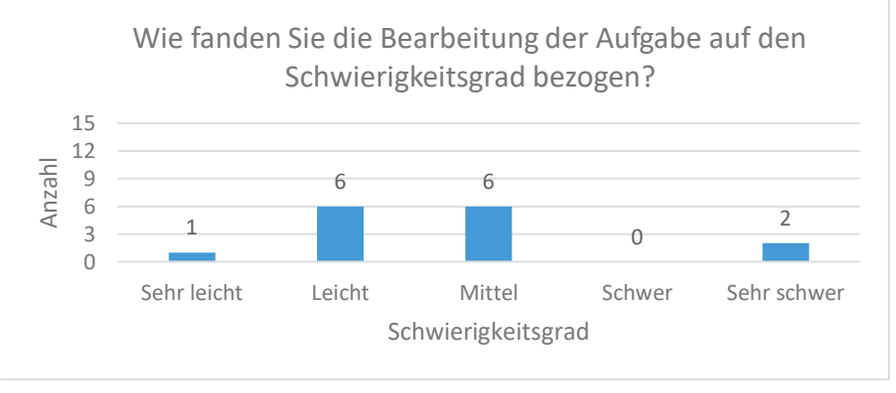


Abbildung 1

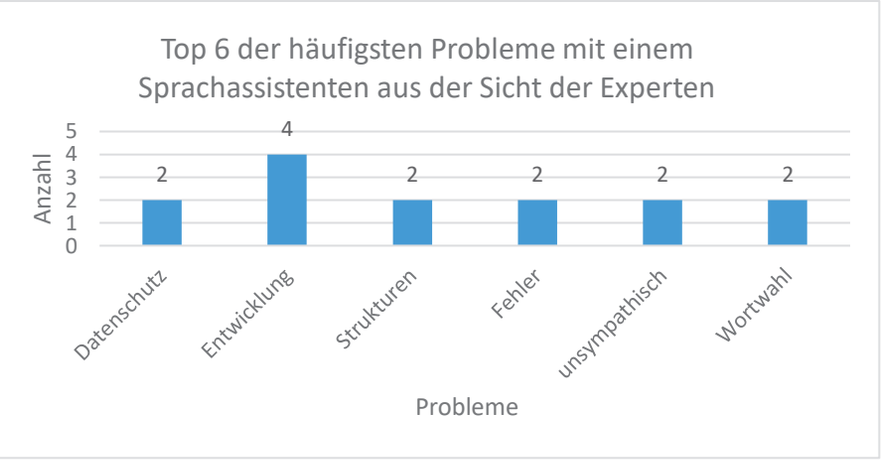


Abbildung 2

einnahmen. Im Rahmen der Umfrage wurden verschiedene mögliche Probleme, die der Anwender bei der Benutzung eines solchen Tools hat, evaluiert. Diese Evaluierung wurde auch bei den Experten durchgeführt. Es wurden unterschiedliche Aspekte zwischen den Aussagen der Experten und der Anwender festgestellt. In Abbildung 2 wird die Sicht der Experten dargestellt. Hierbei ist der Faktor der „Entwicklung“ sehr wichtig. Mit diesem Aspekt ist gemeint, dass so ein System für jeden Anwender individuell entwickelt werden muss. Dieser Aspekt hängt auch sehr stark mit der verwendeten Wortwahl zusammen. Ein anderes Problem, ist, dass insbesondere in sozialen Einrichtungen die Anwender in starren Strukturen leben, die eine Installation eines Sprachassistenten in diesem Bereich erschweren. Im sozial-medizinischen Bereich gibt es strenge Datenschutzrichtlinien, die eine weitere Herausforderung in der Entwicklung darstellen. Ein anderer Faktor ist in Abbildung 2 durch die Rubrik „Fehler“ dargestellt. Mit diesem Faktor sind in erster Linie die falschen Interpretationen gemeint, die ein solches System macht und somit zu einer falschen Handlungsempfehlung führen kann. Hierbei werden jedoch auch die verschiedenen Situationen gemeint, in der ein Sprachassistent einen bestimmten Satz falsch versteht. Aus der Sicht der Anwender wurden jeweils andere Schwierigkeiten gesehen, wie in Abbildung 3 zu erkennen ist. Hier wurden die folgenden Probleme gesehen: Zum einen spielen die technischen Voraussetzungen, die lange Einarbeitungszeit, die Bereitschaft sich mit neuen Dingen zu beschäftigen, die kurze Test-Zeit und die fehlende Differenzierung eines solchen Tools eine wichtige Rolle.

5 Diskussion

In dieser Arbeit wurden verschiedene Umfragen durchgeführt. In Abbildung 1 wird aufgezeigt, dass ältere Personen in der Lage sind, mit einem solchen Tool zu interagieren. Gleichzeitig gibt es verschiedene Probleme, die für eine bessere Interaktion und eine dauerhafte Benutzung durch diese Zielgruppe gelöst werden müssen, wie beispielsweise die datenschutzrechtlichen Bedenken. Diese spielen sowohl in dem privaten Bereich als auch in sozialen und/oder medizinischen Einrichtungen eine große Rolle. Gleichzeitig wollen die potentiellen Nutzer, so die Meinung der Experten, individuell angesprochen werden. Dieser Punkt verbirgt sich unter dem Faktor „Entwicklung“. Dies könnte beispielsweise so gestaltet werden, dass jeder Anwender mit seinem persönlichen Namen angesprochen wird. Ein weiteres Problem ist

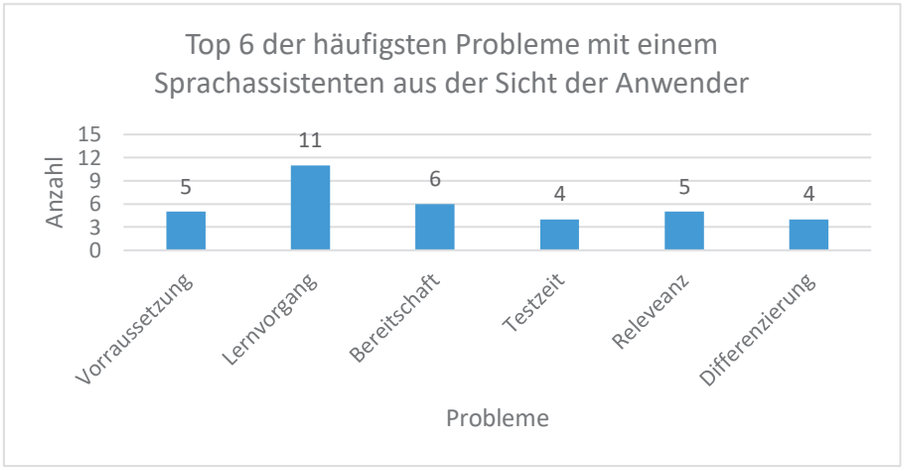


Abbildung 3

die verwendete Sprache, die ein solches System benutzen muss. Gerade in sozialen Einrichtungen muss ein Spagat geschaffen werden, dass die verwendete Wortwahl für den Endanwender geeignet ist und in den Sprachkontext der Betreuer passt. Es ist auch wichtig zu beachten, dass es für eine Entwicklung in dem sozialen, medizinischen Kontext örtliche oder einrichtungsbezogene Wörter gibt, die bevorzugt werden. In sozialen Einrichtungen haben die dortigen Bewohner einen straff organisierten Tagesablauf. Bei einer Entwicklung in diesem Bereich muss die Fragestellung berücksichtigt werden, wie in diesem Tagesablauf ein Sprachassistent platziert werden kann und welche Rolle dieser in Bezug zu den Pflegeverantwortlichen hat. Ein weiteres Problem in diesen Einrichtungen ist, dass ein Großteil dieser Personen nicht oder sehr schwer in der Lage ist, ein solches System zu bedienen. Durch die mechanische Stimme wirkt das System unsympathisch, was die Bereitschaft des Anwenders senkt, mit einem solchen System zu kommunizieren. Ein Faktor, der für den Anwender ein Problem darstellt, insbesondere in einem Altenheim, sind die fehlenden technischen Voraussetzungen. So wird für die Einrichtung eines Sprachassistenten eine WLAN Verbindung und ein Smartphone benötigt. Gerade bei solchen Problemen gab es verschiedene Möglichkeiten, wie dieses ansatzweise gelöst werden könnten. So könnten die technischen Voraussetzungen von der Einrichtung geschaffen werden. Eine weitere Herausforderung der Anwender war der Lernvorgang, der benötigt wird, um sich mit einer solchen Hardware auseinanderzusetzen und ohne Probleme mit ihr zu interagieren. Die Sprachassistentsysteme wie Amazon Alexa, Google Home oder Siri wurden so entwickelt, dass der Anwender die Interaktion nebenher ausführen kann. Damit dies aber funktioniert, muss ein solches Gerät eingerichtet und ähnlich wie bei einem Handy, der Umgang damit geübt werden. Ein weiterer wichtiger Aspekt, der beachtet werden muss, ist die Einarbeitungszeit und der Umgang mit einer neuen Technologie, den die Anwender lernen müssen. Hierbei wurde von den potentiellen Anwendern genannt, dass sie gerne eine Anleitung in schriftlicher Form zu diesen Geräten hätten. Dieser Faktor ist, je nach Anwender, abhängig von der Bereitschaft, sich auf neue Dinge einzulassen. Als weitere Probleme wurden die Faktoren der fehlenden Differenzierung und die zu kurze Test-Zeit des Systems genannt. Die Anwender fanden das Gerät und die damit einhergehende Interaktion sehr gut, sahen aber auch manche Probleme, die erst bei einer langfristigen Nutzung beobachtet werden können: So war eine Fragestellung eines Anwenders, wie sich das System beispielsweise bei einer Uhrumstellung

und bei der Bewältigung alltäglicher Aufgaben verhält. Durch die fehlenden technischen Voraussetzungen und eine zeitaufwändige Einarbeitungszeit gab es auch Anwender, die in einem solchen Gerät keine große Relevanz gesehen haben. Dies lag auch daran, dass die Aufgaben, die ein solches Gerät übernehmen kann, von verschiedenen anderen medialen Geräten und anderen Hilfsmitteln des Anwenders vollständig oder teilweise abgedeckt werden. So benutzen die Anwender heutzutage in den meisten Fällen analoge Kalender, um sich Sachen zu notieren und können über den Fernseher die Nachrichten anschauen. Hierbei sahen die Anwender das Problem der fehlenden Abgrenzung eines Sprachassistenten zu anderen Medien.

6 Fazit

Das Ziel der Umfrage war es herauszufinden, ob es möglich ist, dass diese Zielgruppe mit einem solchen Geräte interagieren kann und welche Faktoren für die Akzeptanz eines solchen Systems wichtig sind. Bei einer weiteren Entwicklung muss darauf geachtet werden, dass die Einarbeitung durch eine einfache Interaktion ermöglicht und so der Lernaufwand reduziert wird. Damit potentielle Nutzer mit einem solchen System interagieren, müssen Anwendungen entwickelt werden, die für diese Personengruppe einen Mehrwert darstellt. Gleichzeitig wäre ein Versuch wichtig, der über einen längeren Zeitraum angesiedelt ist, damit hierbei alltägliche Herausforderungen der Anwender und mögliche Probleme des Systems identifiziert und behoben werden können.

Literaturverzeichnis

[1] Aung Pyae, Tapani N. Joelsson, Investigating the usability and user experiences of voice user interface: a case of Google home smart speaker. In MobileHCI '18 Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Association for Computing Machinery, New York, 127-131, DOI:10.1145/3236112.3236130

[2] Mohammad Hasanzadeh Mofrad, Daniel Mosse, 2018, Speech recognition and voice separation for the internet of things. In IOT`18 Proceedings of the 8th International Conference on the Internet of Things Article No. 8, Association for Computing Machinery, New York, DOI:10.1145/3277593.3277610

[3] Gerhard Leitner, Martin Hitz, Anton J. Fercher, John N. A. Brown, 2013, Aspekte der Human Computer Interaction im Smart Home, HMD Praxis der Wirtschaftsinformatik, Volume 50,6, (Oktober 2013), 37-47, DOI:10.1007/BF03342067

[4] VDI/VDE Innovation+Technik GmbH, Smart Home in Deutschland Untersuchung im Rahmen der wissenschaftlichen Begleitung zum Programm Next Generation Media (NGM) des Bundesministeriums für Wirtschaft und Technologie, Institut für Innovation und Technik (iit), 2010, Seite 9-21, ISBN: 978-3897501652

[5] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, Danielle Lottridge, Understanding the Long-Term Use of Smart Speaker Assistant In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Volume 2, 3 (September 2018), Article No. 91, DOI: 10.1145/3264901

[6] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, Samuel Hardman Taylor, “Alexa is my new BFF: Social Roles, User Satisfaction and Personification of the Amazon Echo. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Association for Computing Machinery, New York, 2853-2859, DOI:10.1145/3027063.3053246

[7] Hartmut Wandke, Elke Wetzenstein-Ollenschläger, Assistenzsysteme: Woher und wohin? In Peissner, M. &

Röse, K. (Hrsg.), Tagungsband UP03. Stuttgart Fraunhofer Verlag. 7-11, ISBN: 978-3-8167-6376-5

[8] Scott R. Klemmer, Anoop K. Sinha, Jack Chen, James A. Landay, Nadeem Aboobaker, Annie Wang, 2000, SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces In UIST '00 Proceedings of the 13th annual ACM symposium on User interface software and technology, ACM New York, NY, DOI:10.1145/354401.354406

[9] S. Iwarsson, A. Ståhl, 2003, Acessibility, usability and universal design – positioning and definition of concepts describing person-environment relationships, Disability and Rehablitation, Volume 25, 2, (Mai 2002), DOI:10.1080/dre.25.2.57.66

[10] YouGov, 2018, „Können Sie sich ganz allgemein vorstellen mit einem „Chatbot“ zu kommunizieren?“ Retrieved 06 Oktober 2018, https://de.statista.com/statistik/daten/studie/872937/umfrage/bereitschaft-zur-kommunikation-mit-chatbots-nach-altersgruppen-in-deutschland/

[11] Bitkom Research, 2017, Anteil der Smartphone Nutzer in Deutschland nach Altersgruppen im Jahr 2017, Retrieved 08 September 2018, https://de.statista.com/statistik/daten/studie/459963/umfrage/anteil-der-smartphone-nutzer-in-deutschland-nach-altersgruppe/

[12] Sharon Oviatt, Phil Cohen, Lizhong Wu, John Vergo, Lisbeth Duncan, Bernhard Suhm, Josh Bers, Thomas Holzman, Terry Winograd, James Landay, Jim Larson, David Ferro, 2000, Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions, Human Computer Interaction, Volume 15, 4, (December 2000), 263 -322, DOI: 10.1207/S15327051HCI1504_1

[13] Dorothea Petrich, 2011, Einsamkeit im Alter Notwendigkeit und (ungenutzte) Möglichkeiten Sozialer Arbeit mit allein lebenden alten Menschen in unserer Gesellschaft, Jenaer Schriften zur Sozialwissenschaft Band Nr. 6, ISSN: 1864-9270

[14] Francisco Javier Carod-Artal, José Luis González-Gutiérrez, José Antonio Egido Herrero, Thomas Horan, Eduardo Varela De Seijas, 2002, Functional recovery and instrumental activities of daily living: follow-up 1-year after treatment in a stroke unit, Brain Injury, Volume 16,3, 207-216 DOI:10.1080/02699050110103337

[15] C. Baber, N. A. Stanton, 1994, Task analysis for error identification: a methodology for designing error-tolerant consumer products, Ergonomics, 1923-1941, DOI:10.1080/00140139408964958

[16] Mitsuo Nagamachi, 1995, Kansei Engineering: A new ergonomic consumer-oriented technology for product development, International Journal of Industrial Ergnomics, Volume 15, 1, 1995, 3-11, DOI:10.1016/0169-8141(94)00052-5

[17] Erik G. Nilsson, 2009, Design patterns for user interface for mobile applications, Advances in Engineering Software, Volume 40,12, (December 2009), 1318-1328, DOI:10.1016/j.advengsoft.2009.01.017

[18] Norman E. Younblodd, Jo Mackiewicz, 2012, A usability analysis of municpal government webiste home pages in Alabama, Government Information Quarterly, Volume 29,4, (Ocotober 2012) 582-588, DOI:10.1016/j.giq.2011.12.010

Klassifikation und Evaluierung von Signalen zur Erkennung chirurgischer Aktivitäten im OP

Claudia Ryniak
Hochschule Reutlingen
Claudia_Isabell.Ryniak@Student.Reutlingen-University.DE

Abstract

Der Operationssaal der Zukunft ist ein voll-integriertes System, in dem komplexe Abhängigkeiten zwischen Geräten, Subsystemen und OP-Personal bestehen. Informationen über den Verlauf der Operation werden gesammelt, um dem Chirurgen entsprechende Assistenz und Hilfestellung zu leisten. Voraussetzung hierfür ist die Kenntnis der aktuellen Situation im Operationssaal. In dieser Arbeit wird ein Gesamtüberblick über aktuelle Forschungsinitiativen im Bereich Aktivitätserkennung im OP gegeben und verschiedene Erkennungsstrategien aus den Bereichen Personenerkennung, Instrumentenerkennung und Erkennung anatomischer Strukturen vorgestellt. Anschließend werden die Erkennungsstrategien und Methoden im Hinblick auf ihren klinischen Einsatz evaluiert.

Schlüsselwörter

Aktivitätserkennung, chirurgischer Workflow, Merkmalsextraktion

CCS Concepts

• **Applied computing~Health informatics** • Information systems~Process control systems

---

Betreuer Hochschule: Prof. Dr.-Ing. Oliver Burgert
Hochschule Reutlingen
Oliver.Burgert@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Claudia Ryniak

1 Einleitung

Die Chirurgie ist ein Teilgebiet der Medizin, welche sich mit Krankheiten und Verletzungen beschäftigt, die eine operative Behandlung zur Diagnose, Vorbeugung und Heilung erfordern. Seit der industriellen Revolution haben die technologischen Fortschritte die Chirurgie zu einer wissenschaftlichen Disziplin gemacht, die in der Lage ist, viele Krankheiten zu behandeln. Diese Fortschritte werden in medizinischen Innovationen, wie beispielsweise modernste chirurgische Instrumente, Navigationsüberwachungssysteme und neuartige Bildgebungstechnologien, umgesetzt. Mit der Verbesserung der Patientenversorgung haben diese Technologien auch die Operationsabläufe verändert [1]. Der Operationssaal der Zukunft ist ein voll-integriertes System, in dem komplexe Abhängigkeiten zwischen Geräten, Teilsystemen und dem OP-Personal existieren. Es werden Informationen über den Verlauf des Eingriffs erhoben und dazu passende Assistenz angeboten [2]. Eine zentrale Rolle dabei spielt der Chirurg, da alles darauf ausgerichtet ist, seine Arbeitsschritte zu unterstützen. Hierfür ist die Kenntnis der aktuellen Situation im OP notwendig [3].

1.1 Problemstellung und Motivation

Die automatische Aktivitätserkennung und damit automatische Assistenz führt zur Effi-

zienzsteigerung chirurgischer Eingriffe, sowohl im Hinblick auf das Wohl der Patienten, als auch in Bezug auf wirtschaftliche Gesichtspunkte. Es können Fehlbedienungen reduziert, Ressourcen optimal eingesetzt und Operationszeiten verkürzt werden. Außerdem kann die Aufmerksamkeit des OP-Teams von der Bedienung der Geräte zurück zum eigentlichen Eingriff gelenkt werden. Zusätzlich lässt sich die automatische Aktivitätserkennung für eine automatische Dokumentation von Eingriffen nutzen und damit personellen Aufwand reduzieren [4]. Um jedoch situationsabhängige chirurgische Assistenz zu ermöglichen, ist es wichtig, den aktuellen Kontext im OP ermitteln zu können. Hierzu muss ein Erkennungssystem entwickelt werden, das alle relevanten Informationen aus dem chirurgischen Prozess sammelt, um damit eine automatische, Eingriffstyp-übergreifende Erkennung chirurgischer Aktivitäten zu ermöglichen.

1.2 Ziel

Der Schwerpunkt dieser Arbeit liegt in der automatischen Erkennung von chirurgischen Arbeitsschritten und leistet damit einen wichtigen Beitrag zur kontextabhängigen chirurgischen Assistenz. Für die automatische Erkennung von Aktivitäten wurden in der Wissenschaft bereits unterschiedliche Methoden behandelt. Diese reichen von der Merkmalsextraktion aus Videoaufnahmen bis hin zum Tracking von OP-Personal, sodass es noch keine ganzheitliche Lösung gibt, sondern nur Teilaspekte aufgegriffen wurden. Aus diesem Grund ist das Ziel dieser Arbeit zunächst die verschiedenen Signale, die zur Aktivitätserkennung beitragen können, zu klassifizieren, sowie Möglichkeiten der Aufnahme und Weiterverarbeitung zusammenzufassen, um einen Gesamtüberblick zu schaffen. Des Weiteren sollen die gesammelten Methoden im Hinblick auf ihre Umsetzbarkeit im klinischen Umfeld und dem chirurgischen Workflow evaluiert werden.

1.3 Vorgehen

Um einen Überblick über den Themenschwerpunkt der Hausarbeit zu schaffen, werden zunächst Grundlagen des chirurgischen Workflows vorgestellt. Dazu wird erst auf den chirurgischen Assistenzprozess und anschließend auf die Mustererkennung eingegangen. Darauf aufbauend wird im Speziellen die Aktivitätserkennung im OP behandelt, indem drei verschiedene Erkennungsstrategien und Möglichkeiten der Aufnahme und Weiterverarbeitung von Signalen vorgestellt werden. Hierzu zählt die Instrumentenerkennung, die Personenerkennung sowie die Erkennung anatomischer Strukturen, die mittels verschiedener Methoden durchgeführt werden können. Zu allen drei Erkennungsstrategien werden Beispiele aus der Wissenschaft und Praxis hinzugezogen und zusammengefasst, um einen Gesamtüberblick über den aktuellen Entwicklungsstand zu geben. Abschließend werden die vorgestellten Erkennungsstrategien und Methoden im Hinblick auf den klinischen Einsatz diskutiert und evaluiert.

2 Chirurgischer Workflow

Ein Workflow beschreibt die Abfolge von Aktivitäten, die von verschiedenen Einheiten oder Agenten im System ausgeführt werden [5]. Mit einer sorgfältigen Analyse des Workflows kann die Verteilung der kognitiven Arbeit und des Informationsflusses in komplexen Umgebungen modelliert werden [6]. Das Konzept des chirurgischen Workflows beschreibt die intelligente Erfassung der Prozessbeschreibung von chirurgischen Eingriffen [7]. In den nachfolgenden Abschnitten wird zunächst auf die Grundlagen zum chirurgischen Assistenzprozess eingegangen und anschließend das Grundprinzip der Mustererkennung näher erläutert.

2.1 Chirurgischer Assistenzprozess

Zur Nutzung der Daten des chirurgischen Workflows, beispielsweise für eine Analyse oder Exploration, bedarf es eine formale Beschreibung des chirurgischen Eingriffs mit Hilfe eines Modells für chirurgische Prozesse. Dieses Modell stellt eine Abstraktion eines chirurgischen Eingriffs dar und kann unter Verwendung geeigneter Ontologien den Arbeitsablauf im Operationssaal in entsprechendem Detailierungsgrad abbilden [8]. Um das kontextabhängige Wissen über den Verlauf einer Operation darzustellen, muss sich die aktuelle Situation im erkannten Workflow wiederspiegeln. Die erhobenen Informationen können weiterfolgend durch Assistenzsysteme genutzt werden.

Abbildung 1 zeigt den chirurgischen Assistenzprozess, dabei steht der chirurgische Prozess am Anfang. Durch die Aktivitätserkennung werden Informationen aus dem Prozess extrahiert und der Workflowverarbeitung zur Verfügung gestellt, welche wiederum die chirurgische Assistenz ermöglicht. Der Kreislauf schließt sich, indem die chirurgische Assistenz auf den chirurgischen Prozess einwirkt.

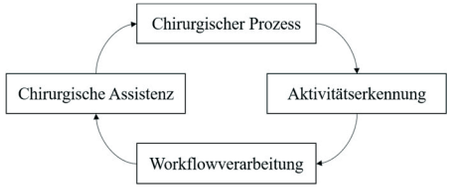


Abbildung 1: Chirurgischer Assistenzprozess [4]

Der Schwerpunkt dieser Arbeit liegt in der Aktivitätserkennung und der damit verbundenen automatischen Erkennung von chirurgischen Arbeitsschritten.

Die Beschreibung chirurgischer Arbeitsschritte soll ähnlich wie bei Geschäftsprozessen den Workflow zu Hilfe nehmen, um chirurgische Prozesse maschinell erfassbar zu machen. Die Autoren in [9] haben eine formale Methode, zur Beschreibung des chirurgischen Workflows auf allen Granularitätsebenen, entwickelt. Dabei werden Beschreibungsmöglichkeiten genutzt, um zum einen individuelle Workflows einzelner Eingriffe zu dokumentieren (individuelles Prozessmodell) und zum anderen ein generalisiertes Modell des Eingriffs zu erhalten (generalisiertes Prozessmodell) [10]. Voraussetzung für die Erkennung von chirurgischen Arbeitsschritten ist eine flexible und allgemein gültige Beschreibung der Vorgänge im OP [4].

2.2 Mustererkennung

Das Erkennen von Aktivitäten im Operationssaal setzt voraus, durch Mustererkennungsmethoden aus vorhandenen Daten sinnvolle und nützliche Informationen zu extrahieren [11, 12]. Die Mustererkennung beschäftigt sich mit der Analyse von Signalen, wie beispielsweise Sensorwerten. Sensorbasierte Signale erfassen einen Teil der physikalischen Welt, in diesem Fall den Operationsaal und die chirurgische Aktivität. Der Mustererkennungsprozess folgt einem grundlegenden Konzept, welches auf verschiedene Mustererkennungsmethoden angewandt werden kann und in Abbildung 2 dargestellt ist.

Der Mustererkennungsprozess beginnt mit der Vorverarbeitung der Signale durch beispielsweise Transformationen oder Entfernen von Artefakten. Anschließend werden bestimmte Merkmale, also Kennzahlen, die ein Signal oder ein Aspekt des Signals charakterisieren, extrahiert. Entsprechend der Art des Signals gibt es Verfahren, die auf die jeweiligen typischen Merkmale abgestimmt sind. Beispielsweise gibt es in der Bildanalyse Verfahren die besonders kontrastreiche Bildbereiche identifizieren. Aus der Merkmalsextraktion resultiert eine Beschreibung des Signals als Merkmalsvektor, dessen Elemente spezifische Kennzahlen enthalten. Dabei werden die wesentlichen Eigenschaften des Signals erfasst und redundante oder unwichtige Aspekte herausgefiltert. Während der nachfolgenden Klassifizierung werden durch vorausgehende Lernprozesse die Merkmale zu Klassen zugeordnet. [13]

Die Lernprozesse teilen sich in zwei Kategorien auf, das überwachte und unüberwachte Lernen. Im überwachten Lernen werden die zu lernenden Klassen vorgegeben, während im unüberwachten Lernen die Klassen auto-

Abbildung 2: Konzept des Mustererkennungsprozesses [4]

matisch aus den vorhandenen Signalen segmentiert werden [14]. Die Kombination beider Lernprozesse ist auch möglich, indem die zunächst automatisch erzeugten Klassen den vorgegebenen Klassen zugeordnet werden.

3 Strategien zur Aktivitätserkennung

Für eine erfolgreiche Erkennungsstrategie müssen Prozessbeschreibung und Erkennungsmethodik aufeinander abgestimmt sein, sodass eine strukturelle und logische Einschränkung verhindert wird. Verschiedene Forschungsgruppen haben diesbezüglich Lösungen erarbeitet, die nachfolgend vorgestellt werden. Die erarbeiteten Ansätze wurden entsprechend der Signalart in Instrumentenerkennung, Personenerkennung und Erkennungen anatomischer Strukturen klassifiziert. Für die Klassifizierungskategorien können unterschiedliche Methoden verwendet werden, um eine erfolgreiche Aktivitätserkennung zu ermöglichen. Diese werden in den nachfolgenden Abschnitten vorgestellt und anschließend in Kapitel 4 evaluiert.

3.1 Instrumentenerkennung

Innerhalb des Operationssaals können verschiedene Instrumente verwendet werden. Je nach Einsatzgebiet werden Instrumente mit bestimmten Charakteristika verwendet, die Aufschluss über die Art des Operationstyps aber auch über die Operationsphase geben können. Aus diesem Grund wird versucht mittels verschiedener Methoden, Informationen über den Zustand und die Nutzung von Instrumenten herauszufinden. In den folgenden Unterabschnitten wird zunächst die RFID-basierte Instrumentenerkennung und anschließend die Erkennung von Systemereignissen des Da-Vinci- Operationssystems – einem roboter-assistierten Chirurgiesystem, mit dem minimalinvasive Operationen durchgeführt werden – vorgestellt.

3.1.1 RFID-basierte Erkennung

Die Idee der Instrumentenerkennung mittels RFID (Radio Frequency Identification) besteht darin die Instrumente mit maschinenlesbaren Merkmalen zu markieren und in vordefinierten Bereichen wiederzuerkennen. Dabei werden bestimmte Regionen im OP als Detektionszonen eingerichtet und dort RFID-Antennen aufgestellt. Wird ein Instrument in eine dieser Zonen bewegt, so kann das RFID-Lesegerät den angebrachten RFID-Transponder lesen und das Instrument registrieren. [4]

Im Ansatz von Meißner [4] wird die RFID-basierte Instrumentenerkennung mit Hilfe von RFID-Tags versehenen Instrumenten sowie mehreren Patch-Antennen und einem RFID-Lesegerät realisiert. Das System kann vier Antennen simultan betreiben und liefert gelesene Daten der RFID-Tags über eine Ethernet-basierte Schnittstelle aus. Zusätzlich besitzt es eine Listen-Before-Talk-Antenne mit der vor einem Lesevorgang festgestellt werden kann, ob das verwendete Frequenzband bereits genutzt wird. Dies ist notwendig, um Interferenzen mit anderen Geräten zu minimieren [4]. Die Patch-Antennen werden im Operationssaal jeweils am Situs und am Instrumententisch angebracht. Da die Abstrahlung der elektromagnetischen Wellen nicht homogen ist, sollten laut Meißner [4] pro Detektionszone mehrere Antennen in orthogonaler Ausrichtung zueinander installiert werden, sodass eine maximale Lesewahrscheinlichkeit gewährleistet wird.

Die von Meißner [4] akquirierten Rohdaten sind eine Menge an Detektionsereignissen, die aus dem Zeitpunkt der Detektion, der Antenne und des Instruments bestehen. Aufgrund von verschiedenen Arten von Interferenzen, wie Reflexion, Absorption und instrumentenspezifischer Abstrahlung kann die Lokalisierung der Instrumente verfälscht werden. Daher betrachtet Meißner [4] ein differenziertes Erkennungsverfahren, das auf der Verwendung von Hidden Markov Modellen (HMM) basiert. HMM‘s können interne Zustände eines Systems über dessen observierbare Symbole ermitteln [15]. Überträgt man dies auf das Lokalisierungsproblem, so stellen sich die internen möglichen Zustände des Modells als Lokalitäten dar und es existiert für jede definierte Zone im Modell ein bestimmter Zustand.

Ziel der Instrumentenerkennung von Meißner [4] ist es nicht nur die Instrumente zu lokalisieren, sondern auch abzuschätzen, ob ein Instrument im aktuellen Zeitpunkt genutzt wird oder nicht. Befindet sich das Instrument im Bereich des Situs, so hat es eine hohe Wahrscheinlichkeit genutzt zu werden. Die Nutzungswahrscheinlichkeit ist demnach eine Verteilung über alle möglichen Instrumente und die Möglichkeit, dass kein Instrument verwendet wird. Das Instrument, das am wahrscheinlichsten genutzt wird, lässt sich über den Maximalwert der Verteilung ermitteln. [4]

3.1.2 Erkennung von Systemereignissen

Ein weiterer Ansatz zur Instrumentenerkennung ist die Erkennung von Systemereignissen. Malpani et al. [16] entwickelten ein Framework für die Erkennung chirurgischer Phasen unter Verwendung von Merkmalen, die aus Systemereignissen des Da-Vinci-Operationssystems gewonnen wurden. Bei den Systemsignalen handelt es sich um leicht verfügbare Ereignisdaten, wie beispielsweise einem binären Signal, das anzeigt, ob ein Instrument aktiv ist. Das Framework zur Aktivitätserkennung besteht aus drei grundlegenden Schritten. Zunächst werden die Systemereignisse über kurze Zeitintervalle zusammengefasst. Anschließend wird die chirurgische Phasenwahrscheinlichkeit für jedes Intervall berechnet, um zuletzt die Segmentierung und Klassifizierung aller chirurgischen Phasen durchzuführen. [16]

Die Methode von Malpani et al. [16] besteht aus der Extraktion ereignisbasierter Merkmale über 90-Sekunden-Schritte und der Zuordnung einer Phasenbezeichnung zu jedem Intervall. Dabei werden drei Klassifizierungstechniken untersucht: Support Vector Machines, Random Forests und temporale künstliche neuronale Netze. Jedes dieser Modelle prognostiziert unabhängig voneinander ein Label für jedes Zeitintervall. Es wird zudem auch die Segmentinferenz untersucht mit einem Ansatz, der auf dem semi-Markov Conditional Random Field basiert, das gemeinsam die Phasensegmentierung und Klassifizierung durchführt. Trotz der Untersuchung mehrerer Modelle mit unterschiedlichen Annahmen haben alle Klassifizierungstechniken eine relativ ähnliche Performance erzielt. Die Support Vector Machine stellte ein einfaches lineares Modell dar, die Random Forrests lernten die wichtigsten Teilmengen von Merkmalen für jede Phase, und die temporalen künstlichen neuronalen Netze modellierten nichtlinear die zeitliche Entwicklung von Merkmalen. [16]

3.2 Personenerkennung

Während einer Operation befinden sich mehrere Personen im Operationssaal. Dazu gehören neben dem Chirurgen auch beispielsweise OP-Schwestern und Anästhesisten. Die Kenntnis über die Bewegungen und Positionen dieser Personen können wertvolle Informationen für die Aktivitätserkennung sein und somit die Zuordnung zu bestimmten Operationsphasen unterstützen. In den folgenden Unterabschnitten werden zwei unterschiedliche Methoden zur Personenerkennung im Operationssaal behandelt. Dazu zählen die Personenerkennung durch Beschleunigungssensoren, sowie durch RGBD-Kameras.

3.2.1 Beschleunigungssensoren

Ahmadi et al. [17] stellen einen Ansatz zur Aktivitätserkennung vor, der auf dem Einsatz von tragbaren Beschleunigungssensoren im Operationssaal basiert. Die Beschleunigungssensoren sind eigenständige, dreiach-

sige Sensoren, die Beschleunigungsdaten erfassen. Die Messdaten wurden zur zentralen Speicherung und anschließenden Analyse in eine dazugehörige Webanwendung hochgeladen. Die Beschleunigungssensoren wurden an der Taille und an beiden Unterarmen des Chirurgen angebracht. Um den strengen Hygienevorschriften im OP gerecht zu werden, wurden die Beschleunigungssensoren unter dem sterilen OP-Mantel angebracht [17].

Für die Aktivitätserkennung wurden Signalgrößen verwendet, die Gravitationseinflüsse effektiv reduzieren und Bewegungen als Energiesignale darstellen. Die Daten wurden in Teilsequenzen aufgeteilt und sowohl im Zeit- als auch im Frequenzbereich als Input für HMM‘s ausgewertet. Charakteristische Bewegungen, die Aktivitäten in der Vertebroplastie darstellen – einem minimalinvasiven medizinischen Verfahren zur Behandlung von Wirbelkörperfrakturen – wie Hämmern, Gehen oder Zementrühren, wurden manuell aus dem Video beschriftet und als HMM-Eingabedaten verwendet. Dadurch wurde die Erkennung der chirurgischen Aktivität bei unbekannten Operationen ermöglicht. Insgesamt wurde eine relativ niedrige Erkennungsrate im realen Szenario erzeugt, dies führte die Autoren zu dem Schluss, dass Beschleunigungssensoren allein nicht ausreichen, um eine zuverlässige Aktivitätserkennung zu gewährleisten. [17]

3.2.2 RGBD-Kamera

OP-Szenevideos erfassen die allgemeinen Aktivitäten der Chirurgen und Kliniker im Operationssaal. Twinanda [1] entwickelte einen Ansatz zur Aktivitätserkennung in dem ein deckenmontiertes Multiview-RGBD-Kamerasystem die Aktivitäten im OP aufzeichnet.

Das Multiview-System wird gewählt, um einen größeren Bereich des OPs zu erfassen. Außerdem wurden RGBD-Kameras verwendet, da diese gleichzeitig Farb- und Tiefenbilder der Szene aufnehmen, die ergänzende Informationen liefern. Farbbilder erfassen das visuelle Erscheinungsbild der Objektoberflächen, während die Tiefenbilder die Abstände zwischen Objekten zum Sensor festhalten. Die Tiefenbilder liefern Informationen, die nicht in den Farbbildern erfasst werden, wie beispielsweise die 3D-Struktur der Szene. Diese Art von RGBD-Sensor liefert die Tiefeninformation unter Verwendung eines Infrarot-Lichtsystems, das aus einem Infrarotprojektor und einer Infrarotkamera besteht. Mit dem Infrarotprojektor werden räumlich-codierte Infrarotmuster projiziert, die auf Objekte in der Szene fallen und anschließend von der Infrarotkamera aufgenommen werden. Die Muster werden dann dekodiert, um den Abstand der Objekte zur Kamera zu bestimmen und zu den Tiefenbildern führt. [1]

Es wurden zwei Kameras angebracht, um die Szene um den Patiententisch herum aufzunehmen und eine Kamera, um den Gerätetisch zu beobachten. Die Durchführung der Aktivitätserkennung an OP-Szenevideos besteht aus den folgenden Schritten: Extraktion visueller Merkmale, Merkmalskodierung und Klassifizierung. Da mit räumlich-zeitlichen Merkmalen gearbeitet wird, wird das visuelle Merkmal aus Bildsequenzen extrahiert, anstatt aus Einzelbildern. Aus jeder Sequenz werden zunächst räumlich-zeitliche Interest Points und tiefe räumlich-zeitliche Interest Points aus Intensitäts- bzw. Tiefensequenzen extrahiert, die dann von Histogrammen des optischen Flusses (HOF) und Tiefenquader Ähnlichkeitsmerkmalen (DCSF) beschrieben werden. Anschließend werden die extrahierten Merkmale aus den Sequenzen aggregiert unter Verwendung eines Feature-Codierungsansatzes, basierend auf dem Bag-of-Words-Verfahren. Hier besteht die Erkennungspipeline aus zwei Klassifizierungsalgorithmen: einer Support Vector Machine (SVM) und einem hierarchischen Hidden Markov Model (HHMM). Das Multiklassen-SVM ist so konzipiert, dass es aus den Merkmalsdarstellungen die Konfidenzwerte berechnet, die die Wahrscheinlichkeit der zu den Phasen gehörenden Sequenzen darstellen. Um die zeitlichen Einschränkungen des chirurgischen Arbeitsablaufs miteinzubeziehen, wird ein zweistufiges HHMM

verwendet. Die oberste Ebene enthält Zustände, die die chirurgischen Phasen und ihre Übergänge modellieren, während die untersten Knoten die phaseninternen Abhängigkeiten darstellen. Alle diese Parameter sind datengesteuert und werden automatisch aus den Daten gelernt. Die Beobachtungen ergeben sich durch die, aus dem SVM erhaltenen, Konfidenzwerte. Die Ergebnisse von Twinanda zeigen, dass sowohl Intensitäts- als auch Tiefenmerkmale wichtige Informationen zur Aktivitätserkennung enthalten und dass die Kombination von Intensitäts- und Tiefenmerkmalen die jeweils einzelnen visuellen Merkmale übertrifft. [1]

3.3 Erkennung anatomischer Strukturen

Das Einbeziehen von anatomischen Strukturen kann ebenfalls den Aktivitätserkennungsprozess unterstützen. Je nach Operationsphase kann das Aussehen dieser Strukturen variieren. Besonders bei Operationstypen die bildgebende Mittel, wie beispielsweise laparoskopische oder mikroskopische Videos verwenden, lassen sich diese Informationen nutzen und bestimmten Phasen zuordnen.

Twinanda [1] hat neben dem bereits vorgestellten Ansatz der RGBD-Kamera, eine Erkennungspipeline mit aus laparoskopischen Videos manuell erstellten visuellen Merkmalen entwickelt, um eine Aktivitätserkennung zu ermöglichen. Der erste Schritt der Aktivitätserkennung durch laparoskopische Videos besteht darin, Merkmale aus den Videobildern zu extrahieren. Hier werden sowohl binäre Werkzeuginformationen als auch visuelle Merkmale verwendet, um die Videobilder darzustellen. Die binäre Werkzeugfunktion enthält die Informationen über die zu einem definierten Zeitpunkt verwendeten Werkzeuge. Formal wird für jedes Videobild ein binärer Vektor definiert, wobei jedes Element des Vektors die Verwendung eines chirurgischen Werkzeugs im Bild darstellt. Für die manuell erstellten visuellen Merkmale werden aus jedem Videobild die gleichen visuellen Merkmale extrahiert. Das Videobild wird anschließend durch einen Merkmalsvektor dargestellt, der die Verkettung aller, aus diesem Bild extrahierten, visuellen Merkmale darstellt. [1]

Der Klassifizierungsschritt besteht wie in Unterabschnitt 3.2.2 aus zwei Klassifizierungsalgorithmen: einer SVM und einem HHMM. Das Multiklassen-SVM wurde entwickelt, um die Feature-Darstellung der Videobilder als Input zu nehmen und die Konfidenzwerte zu berechnen, die die Wahrscheinlichkeiten eines zu den Phasen gehörenden Videobilds darstellen. Um die zeitlichen Einschränkungen des chirurgischen Arbeitsablaufs miteinzubeziehen, wird ein zweistufiges HHMM verwendet. Die oberste Ebene enthält Zustände, die die chirurgischen Phasen und ihre Übergänge modellieren, während die untersten Knoten die phaseninternen Abhängigkeiten darstellen. Die HHMM-Beobachtungen werden durch die aus dem SVM erhaltenen Konfidenzwerte gegeben. [1]

4 Evaluierung der Erkennungsstrategien

Die vorgestellten Methoden aus Kapitel 3 haben gezeigt, dass bereits verschiedene Lösungsansätze für die Aktivitätserkennung im OP betrachtet werden. Um diese zu evaluieren, werden zunächst Anforderungen für die Erarbeitung eines Erkennungssystems behandelt. Anschließend werden alle vorgestellten Methoden entsprechend den Anforderungen bewertet und die Ergebnisse zusammengefasst.

4.1 Anforderungen

Meißner [4] hat bereits wichtige Anforderungen für die Erarbeitung eines Erkennungssystems definiert und in vier Kategorien gegliedert. Die erste Anforderung besteht aus der Möglichkeit, den Ablauf in angemessenem Detailgrad verfolgen zu können, um alle Abstraktionsebenen des chirurgischen Prozesses miteinzubeziehen. Außerdem sollten die verwendeten Algorithmen robust sein um die Variabilität der chirurgischen Handlung

abzudecken. Dazu zählt auch, dass Aktivitäten von unterschiedlichen Personen durchgeführt werden können. Eine weitere wichtige Anforderung ist die Übertragbarkeit des Konzepts auf unterschiedliche Eingriffstypen im realen OP, sodass aufwendige Migrationsarbeiten vermieden werden. Zuletzt sollte das Erkennungssystem möglichst wenig in den OP-Ablauf eingreifen und keine chirurgischen Tätigkeiten behindern. Diese grundlegenden Anforderungen dienen als Basis, um die in Kapitel 3 vorgestellten Ansätze zur Aktivitätserkennung zu evaluieren.

4.2 Evaluierung

Für den Bereich der Instrumentenerkennung wurden Ansätze basierend auf RFID und Systemsignalen vorgestellt. Der RFID-basierte Ansatz ist robust auswertbar und besitzt eine geringe Komplexität. Dennoch setzt dieser Ansatz zusätzliche Sensorik im OP voraus, die optimal angebracht werden muss, da sie momentan nicht im OP vorhanden ist. Dadurch sollte die Auswirkung auf die Patientensicherheit nochmals näher betrachtet werden, sowie Anforderungen an die Geräte, wie beispielsweise Sterilisierbarkeit, konkretisiert werden. Die RFID-basierte Methode lässt sich zwar auf andere Operationstypen übertragen, jedoch setzt dies voraus, dass auch freie Instrumente verwendet werden, die auf dem OP-Tisch gelagert werden. Instrumente und Werkzeuge, wie beispielsweise Endoskop wurden dabei nicht erwähnt. Die Technologie ist zudem abhängig von Umgebungsbedingungen, sodass Personen und metallische Objekte die Ergebnisse negativ beeinflussen können. Des Weiteren werden Prozessschritte, wie die Vor- oder Nachbereitung der Operation nicht beachtet, da sie keinen Bezug zu den Instrumenten haben.

Beim zweiten Ansatz zur Instrumentenerkennung, welcher Systemereignisse eines robotischen Unterstützungssystems genutzt hat, werden wie im RFID-basierten Ansatz Prozessschritte der Operationsvorbereitung und -nachbereitung nicht beachtet. Jedoch sind die Ereignisse nicht unbedingt spezifisch für die Roboterchirurgie und können in chirurgische Modalitäten ohne Operationsroboter aufgenommen werden. Die Nutzung von Gerätesignalen ist an sich ein guter und übertragbarer Ansatz, da sich im OP bereits viele Geräte befinden, allerdings kann das Abgreifen der Signale einige Probleme mit sich bringen. So kann es vorkommen, dass diese Geräte keine offenen Lösungen, wie proprietäre Standards oder offene Schnittstellen besitzen.

Im Bereich der Personenerkennung wurde auf Methoden zur Aktivitätserkennung mit Hilfe von Beschleunigungssensoren sowie RGBD-Kameras eingegangen. Die Beschleunigungssensoren geben Informationen zu den direkten Bewegungen des Chirurgen über den gesamten Operationsprozess. Durch die alleinige Verwendung gehen jedoch weitere Kontextinformationen, wie zum Beispiel Anatomie oder verwendete Instrumente verloren. Außerdem handelt es sich hierbei wieder um zusätzliche Sensorik, die vor jeder Operation direkt am Chirurgen angebracht werden muss. Dabei darf diese den Chirurgen in seiner Arbeit und die Patientensicherheit nicht beeinträchtigen. Sofern dies nicht gefährdet ist, kann man die Methode auch auf unterschiedliche OP-Eingriffe, die eine gewisse Varianz in den Bewegungen besitzen, anwenden.

Der Ansatz, der RGBD-Kameras zur Personenerkennung und laparoskopische Bilder zur Erkennung anatomischer Strukturen zur Hilfe nimmt, ist relativ anfällig in Bezug auf optische Mehrdeutigkeit und Verdeckung, sowie hochkomplex, sodass sich zur Variabilität des Prozesses die Unsicherheit der Methodik addiert. Der Ansatz der laparoskopischen Videos begrenzt sich nur auf wenige chirurgische Spezialfälle, wie beispielsweise laparoskopische Eingriffe, während die Verwendung der RGBD-Kameras sich auf viele OP-Eingriffe übertragen lässt, da hier das gesamte Spektrum des Operationssaals betrachtet werden kann. Für beide Ansätze ist eine Anbringung von zusätzlicher Sensorik nicht zwingend notwendig.

4.3 Ergebnis

Jeder der vorgestellten Ansätze besitzt seine Stärken und Schwächen, sodass eine konkrete Empfehlung für die Verwendung einer spezifischen Methode nicht ausgesprochen werden kann. Eine Gemeinsamkeit haben jedoch alle vorgestellten Ansätze, sie wurden auf bestimmte Einsatzszenarien zugeschnitten oder nur an bestimmten Szenarien evaluiert und erprobt. Aufgrund der hohen Variabilität der OP-Eingriffe wäre es daher von Vorteil, wenn mehrere Methoden in Kombination zum Einsatz kommen, um so ein breites Spektrum an Prozessschritten zu erkennen und miteinzubeziehen.

Um eine präzisere Erkennung von Aktivitäten zu ermöglichen kann beispielsweise die RFID-basierte Instrumentenerkennung gemeinsam mit am Chirurgen angebrachten Beschleunigungssensoren kombiniert werden. Dadurch kann sowohl das aktuell verwendete Instrument als auch die dazugehörige Bewegung ermittelt werden. Des Weiteren könnte die Aktivitätserkennung durch laparoskopische oder mikroskopische Videos gemeinsam mit der Erkennung von Systemereignissen genutzt werden, wodurch nicht nur die anatomische Struktur, sondern auch die Nutzung der sich im Körper befindlichen Instrumente erfasst werden. Für eine ganzheitliche Betrachtung des Operationssaals eignet sich von allen vorgestellten Methoden die Verwendung der OP-Szenenvideos am besten. Da es aber insbesondere um den Patienten herum schwierig ist die Videosequenzen präzise auszuwerten, könnten auch hier zusätzlich die RFID-basierte Instrumentenerkennung oder laparoskopische Videos miteinbezogen werden.

5 Fazit

Die automatische Aktivitätserkennung im OP bildet einen wichtigen Aspekt für Assistenzsysteme und verbindet den chirurgischen Prozess mit der Technik, die eine Assistenz ermöglicht. Wichtig hierbei ist, dass die Erkennung ausreichend Informationen vom Prozess erhält, sammelt, und interpretiert.

In dieser Arbeit wurden Methoden zur Aktivitätserkennung im OP untersucht. Speziell wurde zunächst ein Gesamtüberblick über die aktuellen Forschungsarbeiten geschaffen. Diese wurden anhand der Signalarten in die drei Kategorien Instrumentenerkennung, Personenerkennung und Erkennung anatomischer Strukturen klassifiziert. Hervorzuheben ist, dass sich bereits viele Forschungsteams mit dem Erkennen chirurgischer Prozesse befassen. Allerdings behandeln diese oftmals nur Teilaspekte, die sich auf eine spezielle Technologie und ausgewählte OP-Eingriffe beschränken, wodurch nur Teillösungen entwickelt werden. Im nächsten Schritt wurden die vorgestellten Methoden im Hinblick auf den Detailgrad, die Algorithmik, die Übertragbarkeit und möglicher Beeinflussung evaluiert. Hierbei ist aufgefallen, dass keiner der vorgestellten Ansätze alle vier Anforderungen erfüllt hat und bei jedem diverse Stärken und Schwächen existieren. Es wurde vorgeschlagen, mehrere Signalarten zu kombinieren, um damit ein größeres Aktionsspektrum zu erreichen und den Detailgrad der Erkennung zu erhöhen. Dadurch wäre es zudem möglich mehrere unterschiedliche OP-Eingriffe miteinzubeziehen.

Die automatische Aktivitätserkennung im OP ist ein großer Schritt hin zum Operationssaal, der intelligent und assistierend agiert. Nun gilt es neue kooperierende Ansätze schnell zu erschließen und auch im echten klinischen Umfeld zu testen, um den Weg für eine effizientere und sicherere Chirurgie frei zu machen.

Literaturverzeichnis

[1] A. Twinanda. 2017. Vision-based approaches for surgical activity recognition using laparoscopic and RBGD videos. University of Strasbourg. URL: https://tel.archives-ouvertes.fr/tel-01557522/document.

[2] W. Niederlag, H.U. Lemke, G. Strauß and H. Feußner. 2014. Der digitale Operationssaal. Dresden: Health Academy. ISBN: 9783110335620.

[3] J. Shi, G. Strauß, S. Heininger and T. C. Lueth. 2011. Surgical assistance for instruments' power control based on navigation and neuromonitoring. In: Conf Proc IEEE Eng Med Biol Soc 2011: 2115-8. DOI: 10.1109/IEMBS.2011.6090394.

[4] C. Meißner. 2015. Perspektivenorientierte Erkennung chirurgischer Aktivitäten im Operationssaal. Universität Leipzig. URL: http://ul.qucosa.de/api/qucosa%3A13219/attachment/ATT-0/.

[5] J.J. Brixey, D.J. Robinson, J.P. Turley and J. Zhang. 2010. The roles of MDs and RNs as initiators and recipients of interruptions in workflow. Int J Med Inform 79(6): 109-15. DOI: 10.1016/j.ijmedinf.2008.08.007.

[6] M. Vankipuram, K. Kahol, T. Cohen and V.L. Patel. 2011. Toward automated workflow analysis and visualization in clinical environments. Journal of biomedical informatics 44: 432-40. DOI: 10.1016/j.jbi.2010.05.015.

[7] T. Neumuth, G. Strauß, J. Meixensberger, H.U. Lemke and O. Burgert. 2006. Acquisition of Process Descriptions from Surgical Interventions. In: S. Bressan, J. Küng, R. Wagner (eds) Database and Expert Systems Applications. DEXA 2006. Lecture Notes in Computer Science (4080). Berlin, Heidelberg: Springer. ISBN: 9783540378723.

[8] M. Röger. 2010. Konzeption und Realisierung eines Data Warehouses zur Analyse chirurgischer Workflows. Hamburg: Diplomica Verlag GmbH. ISBN: 9783836634465.

[9] D. Neumuth, F. Loebe, H. Herre and T. Neumuth. 2011. Modeling surgical processes: A four-level translational approach. In: Artif Intell Med 51(3): 147-61. DOI: 10.1016/j.artmed.2010.12.003.

[10] T. Neumuth, P. Jannin, J. Schlomberg, J. Meixensberger, P. Wiedemann and O. Burgert. 2011. Analysis of surgical intervention populations using generic surgical process models. In: Int J Comput Assist Radiol Surg 6(1): 59-71. DOI: 10.1007/s11548-010-0475-y.

[11] R.J. Schalkoff. 2008. Pattern Recognition. In: Wiley Encyclopedia of Computer Science and Engineering. DOI: 10.1002/9780470050118.ecse302.

[12] K. Fu and T. Cover. 1976. Digital pattern recognition. Communication and cybernetics. Springer-Verlag. ISBN: 978-0387102078.

[13] L. Kuncheva. 2004. Combining Pattern Classifiers: Methods and Algorithms. Wiley. ISBN: 9780471660262.

[14] E. Alpaydin. 2008. Maschinelles Lernen. München: Oldenbourg. ISBN: 9783486581140.

[15] A. Malpani, C. Lea, C. Chiung, G. Chen and G.D. Hager. 2016. System events: readily accessible features for surgical phase detection. In: Int J Comput Assist Radiol Surg 11(6): 1201-9. DOI: 10.1007/s11548-016-1409-0.

[16] L. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE 77(2): 257-86. DOI: 10.1109/5.18626.

[17] S.A. Ahmadi, N. Padoy. S.M. Heining, H. Feussner, M. Daumer and N. Navab. 2008. Introducing wearable accelerometers in the surgery room for activity detection. In: Computer- und Roboter-Assistierte Chirurgie. Leipzig. URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=F552A7E91D6F1C61C0B04FB607433017?doi=10.1.1.205.4559&rep=rep1&type=pdf.

Usability Methoden am Beispiel der Fallstudie eines B2B Unternehmens

Dina Kurbanismailova
Hochschule Reutlingen
Dina.Kurbanismailova@Student.Reutlingen-University.DE

Abstract

In der Arbeit wird eine Fallstudie am Beispiel eines B2B Unternehmens vorgestellt, das ein erklärungsbedürftiges Service bietet. Im Rahmen der Fallstudie wird, auf Basis einer Datenanalyse und Personas, versucht, ein Anfrage-Formular zu optimieren. Im Resultat eines multivariaten Tests, in dem eine alte und eine neue Formularversion verglichen werden, hat das optimierte Formular keinen Erfolg gezeigt, weil die Eingabefelder von Kunden nicht zu 100% ausgefüllt wurden. Eine gleich hohe Absprungrate bei den beiden Formularversionen könnte jedoch bedeuten, dass eine höhere Anzahl der Eingabefelder bei der Anfrageerstellung keine negative Auswirkung hat.

Keywords

Business-to-Business, Usability Methoden, Personas, Formular, Datenanalyse

CSS Concepts

- Human-centered computing → User studies; Usability Testing

Betreuerin Hochschule: Prof. Dr. rer nat Gabriela Tullius
Hochschule Reutlingen
Gabriela.Tullius@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Dina Kurbanismailova

1 Einleitung

Aufgrund der vorherrschenden Komplexität finden sich im B2B (Business-to-Business) Bereich oft erklärungsbedürftige Produkte. Die Herausforderung besteht unter anderem darin, dass den Kunden schwerfallen kann, den Nutzen oder das Aussehen des gewünschten Produkts selbstständig zu beschreiben, sie benötigen dabei Unterstützung [2]. Mit diesem Ziel soll ein Dialog ermöglicht werden, durch den die Information über das Produkt gut kommuniziert werden kann und ein erklärungsbedürftiges Produkt nicht als solches wahrgenommen wird. Dies kann über eine gute User Experience (UX) und Usability erreicht werden.

Laut ISO Norm 9241-210 bezeichnet die UX eine Kombination des Markenimages, der Funktionalität, der Systemleistung sowie des Verhaltens eines interaktiven Systems, was zu einem seelischen und physischen Zustand des Benutzers, ausgehend aus seinen früheren Erfahrungen, Fähigkeiten und dem Nutzungskontext, führt [11]. Das Ziel der UX ist das Gesamterlebnis des Nutzers zu optimieren [8] und ihn emotional anzusprechen, damit er die Anwendung bzw. die Webseite positiv und zufrieden verlässt und diese später wieder besucht. Bei der Entwicklung einer UX sollten viele Teildisziplinen involviert werden, d.h. nicht nur Webdesigner und Entwickler, sondern auch Servicemitarbeiter, Produktverantwortliche und Marketingmanager.

Usability ist einer der Bestandteile der UX und bezeichnet das Ausmaß, in dem ein Produkt in einem bestimmten Kontext genutzt werden kann, um bestimmte Ziele effektiv und effizient zu erreichen [11].“ Das Ziel ist ein Produkt möglichst einfach, intuitiv und nutzerfreundlich in der Benutzung zu machen, damit der Anwender sein Ziel schnell und leicht erreichen kann [1].

2 Kontext der Arbeit

Das B2B-Unternehmen, dem die Fallstudie in dieser Arbeit gewidmet ist, bietet Textilveredelung an, was eine hoch individualisierte, daher erklärungsbedürftige Dienstleistung ist. Die Webseite der Firma ist anfragegenerierend, und ruft die Nutzer zum „Jetzt anfragen“ anstatt „Jetzt kaufen“ auf. Jede per Online-Formular erstellte Anfrage wird von einem Vertriebsmitarbeiter bearbeitet und ein individuelles Angebot wird vorbereitet. Bei der Erstellung seiner Anfrage muss der Kunde über mehrere Aspekte entscheiden: Artikel, Menge, die Stoffqualität je nach Verwendungsziel, zusätzlich die Veredelungsart, sowie eine oder mehrere Positionen der Veredelung und Größen. Zusätzlich muss der Interessent eine graphische Datei mit dem gewünschten Muster zur Verfügung stellen. Der Kunde kann dabei eine kreative Idee bezüglich des Veredelungsmotivs haben, die er nicht schriftlich erklären kann.

Auf der Webseite der Firma sind zwei Anfrageformulare zu finden. Das eine Formular ist gut erkennbar als ein Briefsymbol ganz oben auf der Startseite platziert. Das zweite Formular ist ebenfalls auf der Startseite als ein Blockelement mit der Bezeichnung „Express-Anfrage“ platziert und ist ebenfalls nicht zu übersehen.

Die beiden Formulare erfüllen dasselbe Ziel der Anfragesendung und unterscheiden sich lediglich durch die Anzahl der auszufüllenden Felder. Das erste Formular hat folgende Struktur: Kontaktdaten, Nachricht (siehe Abbildung 1). Der Kunde kann in diesem Formular seine Anfrage im Fließtext erstellen.

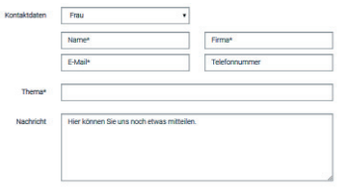


Abbildung 1: Kontaktformular

Das zweite Formular, die Express-Anfrage, verfügt über drei zusätzliche Ausfüllfelder zum Produkt und hat folgende Struktur: Kontaktdaten, Produkt, Nachricht (Abbildung 2).

Abbildung 2: Express-Formular

Der Neubesucher der Webseite findet zwar schnell eine dieser Anfragemöglichkeiten, es wird ihm jedoch nicht erklärt, ob es zwischen diesen Formularen einen weiteren Unterschied gibt. Deswegen fällt seine Entscheidung, welches Formular zu nutzen, spontan. Die weiteren Kanäle, über die eine Anfrage erstellt werden kann, sind eine direkte E-Mail, Anfrage über den Warenkorb, wie in einem traditionellen Online-Shop und per Rückrufwunschoption.

2.1 Problemstellung

Aktuell werden in dem Express-Anfrage-Formular die Produktfelder selten ausgefüllt. Das Formular wird trotz dieser Felder auf dieselbe Weise wie das erste Formular benutzt, d.h. lediglich, um eine Nachricht mit freiem Text zu hinterlassen. Eine umgehende Ausfüllung des Formulars würde dem Vertrieb jedoch die Angebotsvorbereitung erleichtern und die Anzahl der Rückfragen deutlich reduzieren.

2.2 Ziel der Arbeit und Vorgehensweise

Aus dem oben genannten Grund macht es Sinn, das Express-Anfrage-Formular zu optimieren, indem weitere Felder zum Produkt hinzugefügt werden. Die optimierte Express-Anfrage wird abwechselnd mit der alten Version auf der Webseite der Firma geschaltet. Mit Hilfe eines multivariaten Tests (MVT) soll herausgefunden werden, ob das neue Formular effektiver ist. Die Kriterien der Effektivität werden in Abschnitt 7.1.2 definiert. Das Ziel dieser Arbeit ist es die Fallstudie zu beschreiben und die im Rahmen der Studie umgesetzten Usability Methoden darzustellen. Ein weiteres Ziel ist die Frage zu beantworten, ob das Hinzufügen weiterer Felder in das Formular zulässig ist oder ob es die Kunden abschrecken würde.

Die Vorgehensweise bei der Fallstudie entspricht der im März 2010 publizierten ISO-Norm 9241-210, in der die Schritte für benutzerorientierte Gestaltung interaktiver Systeme definiert werden. Der Ablauf berücksichtigt folgende Phasen: den menschlichen Gestaltungsprozess planen, den Nutzungskontext verstehen und spezifizieren, die Nutzeranforderungen detaillieren, Gestaltungsvorschläge erarbeiten und Gestaltungsvorschläge aus der Benutzerperspektive evaluieren [8, 11]. Das Vorgehen in der vorliegenden Fallstudie sieht wie folgt aus.

1) **Quantitative Analyse der Anfragen.** Im Rahmen der Fallstudie wird eingangs die Häufigkeitsverteilung der benutzten Kommunikationskanäle ermittelt. Daraus wird zu sehen sein, wie oft die Besucher das Express-Anfrage-Formular im Vergleich zu den anderen Möglichkeiten bevorzugen.
2) In diesem Schritt werden keine anderen Kommunikationskanäle betrachtet. Nur die Express-Anfrage ist von Interesse. **Quantitative Analyse der ausgefüllten Felder** in den Express-Formularen für dieselbe Zeitperiode wird im nächsten Schritt durchgeführt, um herauszufinden wie ausführlich die Besucher das Formular dieser Art ausfüllen.
3) Neben den quantitativen Daten werden aus den Kundenanfragen **qualitative Angaben** darüber extrahiert, wie gut sich die potenziellen Kunden ihr Endprodukt vorstellen, was sie in ihren Anfragen wissen wollen und wie gut sie sich auskennen.
4) Workshops werden durchgeführt, um Personas zu erarbeiten. Auf Basis der Angaben aus Schritt 3 und der Workshops werden Personas der potentiellen Kunden der Firma erstellt.
5) Ein neues Formular wird auf Basis der Personas, Kundenanfragen und mit Berücksichtigung verschiedener Verhaltenspatterns erstellt.
6) Ein MVT wird durchgeführt, bei dem die alte und neue Version des Express-Formulars abwechselnd geschaltet werden.
7) Eine Analyse und Interpretation der Ergebnisse des MVT soll im letzten Schritt durchgeführt werden.

3 State of the Art

In den existierenden Publikationen, die dem Thema UX im B2B Bereich gewidmet sind, werden oft keine für B2B spezifische Konzepte der Usability und UX entwickelt, sondern häufig auf B2C Konzepte generalisiert [14].

3.1 B2B Bereich

Der B2B Bereich erwirtschaftet meist jedoch höhere Gewinne als B2C Geschäft [6]. Im Gegenteil zu impulsiven B2C Käufen, erfolgen B2B Käufe meistens extensiv. Das bedeutet, Anbieter und deren hochpreisige Produkte werden verglichen und auf Basis mehrerer Kriterien ausgesucht [13]. B2B-Käufe entstehen also nach einem langen komplexen Entscheidungsprozess, an dem mehrere Personen aus verschiedenen Unternehmenshierarchien und mit verschiedenen Funktionen teilnehmen [7, 13]. Dabei sind die Entscheider nicht die späteren Anwender [13]. Aus diesem Grund sollen die Produktinformationen auf der B2B Webseite sowohl

den Kaufentscheidungsträger als auch den tatsächlichen Nutzer des Produktes ansprechen. Viele typische B2B Produkte werden häufig als Teil eines komplexeren Systems gekauft, was eine sorgfältige Prüfung der Kompatibilität erfordert. Die B2B Kunden sollen sicher sein, dass sich ein neu gekauftes Produkt in die bereits vorhandenen Systeme oder Workflows nahtlos eingliedert. Diese Unterschiede vom B2C können auch als Herausforderungen betrachtet werden.

Zu den Vorteilen gehört das Konzept „Habenwollen“ („Starting with Desirability“), dass A. Cooper 2004 eingeführt hat. Das *Habenwollen* sei ein primäres Ziel eines erfolgreichen Produkts und die reine Nützlichkeit spiele dabei eine weniger wichtige Rolle. Ein B2B Kunde würde jedoch eine nicht optimale UX eher in Kauf nehmen, als das im B2C der Fall ist [13].

3.2 UX im B2B Bereich

2016 wurde ein fünfjähriges Forschungsprogramm *Usability and User Experience of complex Systems* [14] beendet, im Rahmen dessen sollten die Unternehmen mit Forschungsorganisationen zusammenarbeiten und die Rolle der UX in ihrer Tätigkeit langsam zu erweitern. Durch diese Forschung ist die Bedeutung der UX in allen Touchpoints offensichtlich geworden. In der letzten Phase des Programms erweitert sich der Fokus langsam auf die Customer Experience (CX) und die Brand Experience (BX). Der Grund dafür ist, dass im B2B Bereich die Kundenbeziehungen weniger produktzentriert sind als in der B2C Branche [14]. In einer Recherche (2004) wurde festgestellt, dass der Wert der CX wichtiger ist, als der Wert des Produkts [16]. Es kann dadurch erklärt werden, dass die Geschäftsbeziehungen, insbesondere in der Schwerindustrie, mehrere Jahrzehnte dauern können [14]. CX bezeichnet das Gesamterlebnis, das ein Kunde mit einem Unternehmen hat, von dem ersten Webseitenbesuch bis zum Vertragsabschluss [15]. BX beschreibt wie Konsumenten eine Marke wahrnehmen aus welchen Gründen und wie sie diese in ihrer Erinnerung abspeichern [3].

Damit ein Wettbewerbsvorteil von UX im B2B erreicht wird, reicht es nicht aus, die UX der Produkte zu verbessern. Die Kette der Unternehmen zwischen dem B2B-Unternehmen und dem tatsächlichen Benutzer kann sehr lang sein und die „User“, deren Experience optimiert werden soll, entstehen an jedem Touchpoint Beschaffung, Messen, Wartung, Webseite, Social Media, auch rechtliche Verträge.

Zum Vergleich, konzentriert sich in der B2C Branche der überwiegende Anteil der UX-Arbeit auf nur zwei Touchpoints zwischen dem Kunden und dem Unternehmen: das Produkt und die Unternehmenswebseite [14]. Bereits während des Forschungsprogramms konnten einige Anzeichen der erhöhten Wettbewerbsfähigkeit gezeigt werden.

Eine Arbeit [17], die sich ebenfalls mit UX im B2B Bereich beschäftigt, stellt drei verschiedenen Phasen der Umsetzung von UX vor: 1) Wissensaufbau und Implementierung, 2) Produktbild und Positionierung und 3) CRM-Richtlinien (Customer Relationship Management). In der ersten Phase werden die richtigen Tools und Methoden zum Sammeln der Nutzerdaten gefunden, also *Nutzer → Dienstleister → Produkt (UX, CX)*. In der zweiten Phase wird das Image der Marke, die Anpassung und Kontrolle der produktbezogenen Erwartungen sowie die Festlegung der passenden Kommunikationsformate und –kanäle definiert, also. *Dienstleister→Produkt (UX, CX)→Nutzer*. In der dritten Phase wird versucht sicherzustellen, dass diese Erwartungen den Erfahrungen der Nutzer und Dienstleister entsprechen. Das wird durch ständige Betreuung, maßgeschneiderte Angebote und UX Optimierung erreicht: *Dienstleister→Nutzer ←Produkt (inkl. UX und CX)* [17].

4 Datenanalyse

UX Messwerte bzw. ausgewertete Daten geben Auskunft über Benutzererfahrung und die Interaktion des Benutzers mit dem Produkt. Sie sollen quantifizierbar und direkt oder indirekt beobachtbar sein [4]. Alle in dieser Arbeit auszuwertenden Daten sind nominal, für dessen Auswertung sich am besten die Funktionen deskriptiver Statistik eignen [4, 17]. Zum Beispiel, mittels der Funktion der Häufigkeitsverteilung wird ermittelt, wie oft eine bestimmte Merkmalsausprägung in einer Stichprobe vorkommt [17].

4.1 Quantitative Analyse der Anfragen

Der Interessent kann seine Anfrage durch fünf verschiedene Kanäle erstellen (Abschnitt 2). Diese fünf Anfragemöglichkeiten sind nominale Daten, deren absolute Häufigkeitsverteilung nun ermittelt werden soll. Die in der Zeitperiode von Januar bis August 2018 erhaltenen Mailboxinhalte werden dazu manuell sortiert und gezählt. Die absoluten Häufigkeiten geben die Anzahl der verwendeten Anfragemöglichkeiten an. Die Summe dieser Werte ergibt die Gesamtzahl der in dieser Periode erstellten Anfragen.

i	Kommunikationskanal	Absolute H	%
1	Kontaktformular	150	31%
2	Express-Anfrage	78	16%
3	Direkte E-Mail-Anfrage	124	26%
4	Warenkorb	72	15%
5	Rückrufwunsch	54	11%
Summe		478	

Tabelle 1. Häufigkeitsverteilung der benutzten Kommunikationskanäle

Aus der Ergebnistabelle 1 ist ersichtlich, dass das einfache Kontaktformular mit dem einzigen Feld für freien Text am häufigsten benutzt wird. Das Express-Anfrage-Formular ist dabei unpopulär.

4.2 Quantitative Analyse der Formularfelder

Da die Express-Anfrage von Interesse ist, wird die absolute Häufigkeitsverteilung der ausgefüllten Felder in diesem Formular ermittelt. Sämtliche Anfragen aus der Testperiode wurden individuell analysiert und alle dort gestellten Fragen wurden aufgeschrieben. Da sich die meisten Fragen vom Sinn her wiederholten, konnten sie nach Kategorien sortiert werden (Tab.3). In 79 abgeschickten Express-Anfrage-Formularen sind nur 39% der Felder “Artikel/Menge” ausgefüllt. Das kann bedeuten, dass der Kunde bei der ersten Anfrage noch nicht entschieden hat, was für ein Produkt er genau braucht. Eine weitere Erklärung wäre, dass der Kunde die Menge erst in Abhängigkeit vom Preis bestellt. Die Tatsache, dass 68% der Kunden ihre Telefonnummer im nicht pflichtigen Feld hinterlassen haben, konnte bedeuten, dass sie ihre Anfrage gern in einem Telefonat klären würden, und dies könnte bedeuten, dass der Veredelungswunsch schwer zu erklären ist.

Express-Anfrage	Anzahl	Anteil
Anfrage, ob die Idee realisierbar ist	3	4%
Felder Artikel/Menge ausgefüllt?	31	39%
Gute Vorstellung	39	49%
Nähanfrage	10	13%
TelefonNr. Hinterlassen?	54	68%
Anzahl der Express-Anfragen allgemein	79	100%

Tabelle 2. Häufigkeitsverteilung der ausgefüllten Felder im Express-Formular

Diese Häufigkeitsergebnisse werden als Grundlage für die Erarbeitung der Personas verwendet.

5 Erarbeitung der Personas

Im Mittelpunkt jeder nutzerorientierten Entwicklung steht der Nutzer [1]. Er muss deshalb so bekannt sein, dass die Konzeption optimal angepasst werden kann. Personas basieren auf realen Menschen und bezeichnen zusammengesetzte Archetype, die auf Verhaltensmuster beruhen, welche während einer Beobachtungsphase gesammelt wurden. Ein Archetyp ist das Gegenteil eines Stereotyps, der dagegen eher Voreingenommenheiten der Entwickler und Forscher, als sachliche Daten, verkörpert [9].

Wie im Abschnitt 3.1 bereits erklärt wurde, nehmen im B2B Bereich häufig mehrere Personen an einem Kaufprozess teil [7]. Andererseits steht hinter jedem Unternehmen, das die Anfrage erstellt, eine Person mit ihren Fachkenntnissen und ihrer individuellen Vorgehensweise am Arbeitsplatz.

Damit die Erarbeitung der Personas präzise und datenbasiert ist, werden die Zielgruppen aus einer in der Firma bereits vorhandenen Marketing-Analyse abgeleitet. Das sind Industrieunternehmen, Werbeagenturen, Handwerkbetriebe und Kleinunternehmen (Handwerker, Einzelhandel, kleine Speditionen, Gastronomie), Dienstleistungsgewerbe, Banken, Ärzte und Vereine.

Personas sollen auf empirisch erhobenen Daten basieren [1]. Zu einem Teil werden die qualitativ ausgewerteten Daten aus der Häufigkeitsverteilungsanalyse als Grundlage genommen (Abschnitte 4.1, 4.2), zum anderen Teil werden für die Persona-Erarbeitung mehrere Workshops durchgeführt, die ebenfalls eine qualitative Auswertung darstellen. Die Erkenntnisse über die Eigenschaften der Nutzer wurden aus 478 Neukundenanfragen abgeleitet. Anhand der Anfragen konnte man sehen, wie tief oder oberflächlich sich der Kunde im Bereich Textilveredelung auskennt, wie gut er vor der Anfrageerstellung auf der Webseite recherchiert hat, ob er schon eine konkrete Idee hatte oder sich über die Möglichkeiten allgemein erkundigen wollte, ob für ihn die Themen Öko, Referenzen, Qualität relevant sind. Bei der Erarbeitung der Personas konnte die Position/Stelle des Anfragenden nicht als aktive Variable benutzt werden, weil es aus den Anfragen es nicht ersichtlich ist, welche Rolle der Anfragende im Unternehmen spielt. Die Personas der Firma unterscheiden sich daher danach, wie viel Ahnung sie in dem Produktbereich haben und über was für Entscheidungsbedürfnisse sie verfügen.

Während des ersten Workshops, an dem die Unternehmensleiterin, ein Vertriebsmitarbeiter und eine UX Designerin teilgenommen haben, wurde für jede Gruppe eine fiktive Person festgesetzt, die für diese Gruppe repräsentativ sein kann. Nach Ähnlichkeiten der Funktionen und Befugnissen wurden endgültig vier voneinander abgegrenzten Personas erstellt. Im Resultat der analysierten Anfragen und der Erfahrung der Workshop-Teilnehmer sind das: ein Einkaufsmanager, eine Sekretärin/Assistentin in einem kleinen oder mittleren Unternehmen, ein Marketing Manager, und ein Geschäftsführer eines kleinen oder mittleren Unternehmens. Das gemeinsame Merkmal für alle Personas ist, dass keiner aus dem Textilbereich kommt. Die prinzipiellen Unterschiede, die erlaubt haben, Personas voneinander abzugrenzen, sind die Position in der Firma, Entscheidungsbefugnis, Selbstständigkeit beim Entscheidungstreffen, usw. Die Auszüge aus der Personas unten lassen das besser veranschaulichen.

Die Assistentin/Sekretärin *ist beratungsoffen, ihr fehlt die Erfahrung. Sie ist nicht risikobereit und trifft fast keine selbstständigen Entscheidungen. Sie hat klare Vorgaben bezüglich Preise, Qualität, Regionalität. Sie weiß nicht genau, welche Veredelungsart sie auswählen soll.*

Die Marketing Managerin *ist, trendorientiert, karriereorientiert. Sie hat bereits eine Idee bezüglich ihrer Wunschprodukte. Sie kann eine Pantone-Farbnummer nennen und eine graphische Datei sofort zur Verfügung stellen.*

Eine Mindestzahl von drei bis maximal zwölf Personas hat sich als hilfreich erwiesen [5]. Jede Persona der Firma beinhaltet unter anderem die folgenden relevanten Punkte: Alter, Familienstand, Bildungsgrad, Position, Nutzungsverhalten, bevorzugte Kommunikationskanal, Gerätenutzung, Arbeitsweise, Verhalten als Kunde, Ziel, Vorgehensweise bei der Suche nach den gewünschten Produkten, Ängste, ideale Servicelösung.

Um Akzeptanz bei allen Stakeholdern des Projektes zu erhalten, werden diese im nächsten Schritt in den Forschungsprozess einbezogen. Die Geschäftsführerin und Vertriebsmitarbeiter haben während des weiteren Workshops bestätigt, dass hinter den erstellten Personas tatsächlich potentielle Kunden stehen, die ihre Anfragen bei diesem Unternehmen normalerweise erstellen.

Nach der Auswertung kann davon ausgegangen werden, dass die Personas auf validen Ergebnissen der qualitativen Daten

beruhen und es wird bei der danach folgenden Studie zu keinen Fehlentscheidungen kommen. Ohne genaue Auswertungen kann ein *Elastic User* entstehen, was letztendlich ein selbstreferentielles Design [9] verursachen könnte, in dem die Entwickler ihre eigene Ziele, Fähigkeiten und Verhaltensmuster im Produkt berücksichtigen.

6 Optimierung des Anfrageformulars

Auf Basis Personas können die Fragen beantwortet werden: Was sucht der Kunde? Was sieht er auf der Webseite? Kann dieser Dienstleister ihm genau das anbieten, was er braucht? Sieht er das sofort? Diese Fragen und Antworten werden berücksichtigt, um so ein Anfrage-Formular zu erstellen, dass nur wenig bis keine Rückfragen seitens des Vertriebs bleiben. Im Formular sollen wichtige Details zur gewünschten Dienstleistung abgefragt werden, die Anzahl der Ausfüllfelder darf jedoch die Kunden nicht abschrecken und keine hohe Absprungsrate verursachen. Die bereits vorhandenen wenigen Punkte wurden im Formular mit folgenden Feldern vervollständigt: „Farbe“, „Ihre Wunschveredelung“ und „Gewünschte Veredelungsposition“. Außerdem wurde die Funktion „Datei hochladen“ hinzugefügt.

Die eben genannten Optionen helfen dem Kunden sein gewünschtes Produkt vollständig zu beschreiben, ohne dass der Bedarf entsteht einen langen unstrukturierten Text im Kommentarfeld dazuzuschreiben. Das heißt, die Produktfelder vervollständigen sich gegenseitig und machen nur in der Kombination mit einander Sinn.

Es wurde versucht, das Design der alten Formularversion wie beispielsweise Platzierung der Labels, beizubehalten. Wenn während des MVT ein Kunde sein Formular vor oder während des Befüllens aus irgendeinem Grund aktualisiert, darf er nicht ein völlig neues Design sehen, was verwirrend wirken würde.

Zusätzlich wurde die Sektion der Kontaktdaten ganz nach unten verschoben. Laut einer nachwiesenen kognitiven Verzerrung *Sunk Cost Fallacy,* wenn im letzten Schritt eines mehrstufigen Formulars nach vertraulichen Informationen gefragt wird, füllen Nutzer diese Felder mit höherer Wahrscheinlichkeit aus, sonst verlieren sie den Fortschritt, der durch das Ausfüllen der vorherigen Schritte erzielt wurde [10].

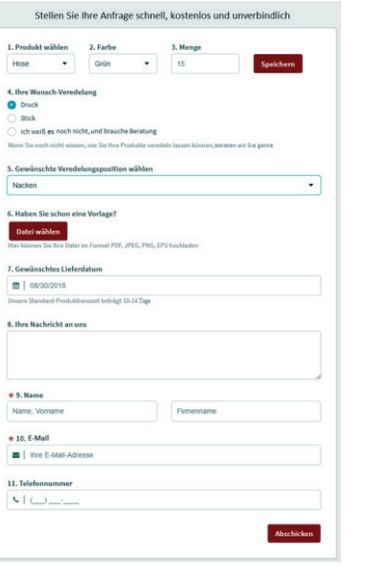


Abbildung 3: Optimiertes Express-Formular

Aus den vorherigen qualitativen Auswetung der Anfragen wurde herausgestellt, dass bei Erstanfragen nie nach Artikelgrößen gefragt wird, weil in dieser Phase der Kunde über diese Information noch selber nicht verfügt, insbesondere wenn es um sehr hohe Bestellmengen geht. Deshalb wird „Größe“ im Formular nicht berücksichtigt.

7 Design des Usability Tests

Das Format der Fallstudie sieht vor, dass **die Teilnehmer** reale Interessenten der Firma, und keine Probanden, sind, die in realen Bedingungen ihre Anfragen erstellen. Sie stellen die realen Benutzergruppen dar.

Die Grundgesamtheit der Teilnehmer wird nicht in weiteren Kategorien unter-

teilt. Ein solches Verfahren entspricht dem Zufallsstichprobenansatz, bei dem jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit in die Stichprobe zu geraten hat [17].

Die Anzahl der Teilnehmer kann in diesem Test nicht vorgesehen oder geplant werden, soll jedoch groß genug sein, um für repräsentative Ergebnisse auszureichen. Je größer der Stichprobenumfang ist, desto näher zu einander werden die obere und untere Grenzen des Konfidenzintervalls [4].

Die Testperiode soll bis zu dem Abgabetermin dieser Arbeit abgeschlossen werden und beträgt einen Monat. Im Fall der nicht repräsentativen Ergebnisse, d.h. bei dem geringen Unterschied zwischen zwei getesteten Versionen wird die Testperiode verlängert. Nach der Verlängerung werden die Ergebnisse nicht mehr in die Arbeit einbezogen werden können.

Im Rahmen dieser Arbeit wird eine **summative Auswertung** des Formulars durchgeführt, bei der die Ergebnisse erst ausgewertet werden, nachdem die Entwicklung des Produkts abgeschlossen ist. Dieser Ansatz wird der formativen Usability gegenübergestellt, bei der das Produkt periodisch während des Entwicklungsprozesses ausgewertet wird.

7.1 MVT

Da im Rahmen dieser Arbeit die Bewertung der Auswirkungen leichter Änderungen (im Formular) stattfindet, ist laut Tullis *Live Website Metrics* das einzige gut geeignete Szenario für die Usability Analyse [4]. Dazu zählt ein multivariater Test als eine Alternative zu einem A/B Test [1].

Der multivariate Test ist eine Testmethode, die sich zur Bewertung mehrerer Varianten des Systems und ihrer Kombinationen eignet. Da die Webseite der Firma wenig Traffic generiert, werden lediglich zwei Varianten des Formulars getestet, wobei jede Variante jeweils 50% des Traffics erhält. Diese gleichmäßige Traffic-Verteilung zwischen den Varianten wird als *Full Factorial Testing* bezeichnet. In einem Fall mit vielen zu testenden Varianten wären für MVT höhere Nutzerzahlen nötig.

Da alle Produktfelder als eine logische Einheit betrachtet werden, werden sie für den MVT alle zusammen eingefügt, und nicht als separate Testvarianten.

Im Fall eines negativen Erfolgs wird eine der Änderungen zurückgesetzt. Zum Beispiel wird die Position der Kontaktdaten wieder nach oben umgestellt. und die abgeänderte „neue“ Version kann mit der alten Version im Rahmen eines weiteren MVT verglichen werden.

7.1.1 Hypothese

Bei der Erstellung der Anfrage füllen die Nutzer die Produktfelder nicht aus, weil diese Angaben nicht vollständig sind, um ihr Wunschprodukt zu beschreiben. Aus diesem Grund halten sie es für sinnvoller ihre Anfrage mit Fließtext zu erstellen. Wenn der Kunde alle für die Bestellung notwendigen Daten ausfüllt (Abschnitt 6) und dazu noch eine Datei mit gewünschtem Muster lädt, kann er seine Anfrage vollständig erklären.

7.1.2 Erfolgskriterien

Für das neue Formular werden messbare Erfolgskriterien definiert:

1. Die neue Formularversion ist effizienter als die aktuelle Version;

2. Die neue Formularversion ist weniger effizient als die aktuelle Version;

3. Der Unterschied zwischen der neuen und der aktuellen Version ist nicht ersichtlich. Das neue Formular zeigt kein Fehlschlagen und keinen Erfolg.

Mittels Logdateien kann die Anzahl der Aufrufe in ein prozentuales Verhältnis zu erfolgreich abgeschickten Anfragen gestellt werden. Je höher der Anteil der abgesendeten Anfragen ist, desto niedriger ist die Absprungrate. Eine gleich hohe Zahl der erfolgreichen Anfragen über die beiden Formularversionen wird bedeuten, dass eine höhere Anzahl der Ausfüllfelder keinen spürbaren Einfluss nimmt.

8 MVT Ergebnisse

Am Ende der Testperiode werden MVT Ergebnisse anhand von Logdateien des Webservers analysiert. Nach der Bereinigung von eigenen IP-Adressen der Firma und anderen sich wiederholenden IP-Adressen (Kunde aktualisiert sein Formular mehrmals) enthält die Logdatei 164 Einträge. Diese Zahl wird durch zwei geteilt.

26.09.2018 - 26.10.2018	Neue Version	Alte Version	Gesamtzahl
Anzahl der erfolgreich abgeschickten Formulare	13	11	24
Anzahl der Aufrufe	167		100%
Verteilung des Traffics %	50%	50%	
Verteilung des Traffics	82	82	

Tabelle 3. Aufrufe der neuen und alten Formulare während der Testperiode

Aus der Tabelle 3 sind zwei wichtige Punkte ersichtlich.

1) Der Unterschied zwischen den alten und neuen Versionen ist sehr gering und die Ergebnisse sind daher nicht repräsentativ. Es ist daraus zu schließen, dass eine einmonatige Testperiode nicht ausreichend ist, um zwei Varianten zu vergleichen.

2) Der Unterschied zwischen den Formularaufrufen und erfolgreich abgeschickten Anfragen ist für die beiden Varianten sehr hoch. Nur 16% und 13% des neuen und alten Formulars sind zu einer Anfrage geworden. Diese gleich hohe Zahl bedeutet einerseits, dass das neue Formular kein Fehlschlagen und keinen Erfolg zeigt (siehe Abschnitt 7.1.2). Andererseits bedeutet das, dass eine höhere Zahl der Ausfüllfelder keinen negativen Einfluss genommen hat.

8.1 Quantitative Analyse der Formularfelder

Unabhängig von dem MVT ist es von Interesse, wie aktiv die Felder in dem optimierten Formular ausgefüllt wurden. In der Tabelle 4 ist es zu sehen, dass „Artikel“ immer in der Kombination mit „Farbe“ ausgefüllt wird, und nicht immer mit „Menge“. Die neuen Felder „Farbe“ und „Veredelung“ sind zu 69% und 62% ausgefüllt, was bedeutet, dass ihre Hinzufügung einen Sinn machte. Die Tatsache, dass die Kontaktdaten in dem Express-Formular ganz nach unten verschoben wurden, hat in dieser Phase des Tests keine bemerkbare Auswirkung auf die Ausfüllung der Felder gezeigt.

Anzahl der Formulare	13	100%
Feld	Anzahl der ausgefüllten Felder	Anteil der ausgefüllten Felder
Wunschliefertermin	12	92%
Name*	13	100%
Firma	10	77%
E-Mail*	13	100%
Telefonnummer	7	54%
Artikel	9	69%
Farbe	9	69%
Menge	8	62%
Veredelung	8	62%
Position	5	38%
Übertragene Datei	5	38%

Tabelle 4. Häufigkeitsverteilung der ausgefüllten Formularfelder

Laut den vordefinierten Erfolgskriterien ist das neue Formular nicht effizient, weil die Felder nicht zu 100% ausgefüllt wurden. Andererseits ist die neue Version effizienter, weil die zusätzlichen Felder nicht weniger aktiv ausgefüllt werden und bei dem gleichen Aufrufanteil das Formular mehr Information über die Anfrage liefert.

9 Diskussion

Die Testperiode wird auf einen weiteren Monat verlängert. Wenn nach einer längeren Testdauer der Unterschied zwischen den zwei Formularen genauso gering bleibt, kann das bedeuten, dass mehr zusätzliche Formularfelder auf den Erfolg der Formularausfüllung keine negative Auswirkung haben. In diesem Fall kann die Hypothese bestätigt werden, dass es in einem Kontaktformular nicht eine möglichst minimale Anzahl der Felder geben muss, sondern genau so viel, dass das gewünschte Produkt vollständig beschrieben bzw. erklärt werden kann.

Als Alternative zur Formularoptimierung ist ein Retrieval-System vorstellbar, das mit Hilfe linguistischer Verfahren aus dem Fließtext den Kundenwunsch ermitteln lässt.

Literaturverzeichnis

[1] Jacobsen, J., Meyer, L. *Praxisbuch Usability & UX*, Rheinwerk Verlag, Bonn 2017. Buch, 512 S. ISBN: 978-3-8362-4423-7

[2] Matys, E. *Erklärungsbedürftige Produkte bewerben und verkaufen* . Dipl.Ing.Erwin Matys, Wien, Austria 2007-2015. Online verfügbar unter: https://www.managementportal.de/pdf2/Erklaerungsbeduerftige-Produkte-2010.pdf ; Besucht am 01.10.2018

[3] Lieven, T.: *Brand Gender. Increasing Brand Equity through Brand Personality.* Macmillan, Cham, 2018. eBook, Online ISBN:978-3-319-60219-6

[4] Tullis, T., Albert, B. *Measuring the user Experience. Collecting, Analyzing, and Presenting Usabilitz Metrics*, Buch, 301 S. Amsterdam, Heidelberg [u.a.],Elsevier Inc, 2013. ISBN: 978-0-12-415781-1

[5] Zeitschriftenheft "T3N: digital pioneers" 2018, Heft 51 „Weniger Arbeiten", 23.02.2018. „Personas für das UX-Design", ASIN:B079YY9DVT

[6] Jamshidi, B., "Web Usability in B2B Websites. User´s Perspective" Master Thesis. Lulea University of Technology. 2008:011 – ISSN:1653-0187-ISRN: LTU-PB-ES—08/011—SE.

[7] Bauer, M., „User Experience (UX) für B2B Websites", Webseite „www.coconew-media.de", 13.03.2017. Online verfügbar unter: https://www.coconew-media.de/user-experience-ux-fuer-b2b-websites; Besucht am 28.09.2018

[8] Richter, M., Flückiger, M. *Usability Engineering kompakt. Benutzbare Produkte gezielt entwickeln*, Springer Vieweg, Berlin Heidelberg 2013. Buch, 179 S. ISBN: 978-3-642-34831-0

[9] Cooper, A., Reimann, R., Cronin, D. *About Face 3. The Essentials of Interaction Design*, Wiley Publishing, Inc., Indianapolis, Indiana, 2007. ISBN: 978-0-470-08411-3

[10] Webseite „VentureHarbour". Artikel „58 Form Design Best Practices & Form UX Examples". Online verfügbar unter:https://www.ventureharbour.com/form-design-best-practices/; Besucht am 27.09.2018

[11] ISO DIS 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland.

[12] Bamberg, G., Baur, F., Krapp, M.: *Statistik. Eine Einführung für Wirtschafts- und Sozialwissenschaftler.* eBook, April 2017. ISBN: 978-3-11-049572-0

[13] Niklas, S. & Laux, M., UX im Konzern – Usability Engineering facing Daily Business. In: Hess, S. & Fischer, 2016. DOI: 10.18420/muc2016-up-0146

[14] Roto, V., Kaasinen, E., Nuutinen, M., Seppänen, M. „UX Expeditions in Business-to-Business Heavy Industry: Lessons Learned", Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. Seiten 833-839. DOI:10.1145/2851581.2851600

[15] Webseite "Rytewiki", "Customer Experience". Online verfügbar unter: https://de.ryte.com/wiki/ Customer_Experience; Besucht am 20.09.18

[16] Sundberg, H-R. (2015). The Role of User Experience in a Business-to-Business Context. (Tampere University of Technology. Publication; Vol. 1278). e-ISBN: 978-952-15-3450-8

[17] Abramov, V.,"Conveying User Experience in Business-to-Business Environment", Master Thesis, 2012, Aalto University. Online verfügbar unter: https://pdfs.semanticscholar.org/0c79/daf6fca15d6fbbdfecef277f64e8b937506e.pdf; Besucht am 30.10.2018

Komponentenbasierte Analyse und Animation Retargeting von Gesichtsdynamiken

Dominik Müller
Hochschule Reutlingen
dominik.mueller@student.reutlingen-university.de

Abstract

In dieser Arbeit wird analysiert, wie ein stabiles und effizientes Tracking und eine komponenten-basierte Übertragung von Gesichtsbewegungen auf einen digitalen Avatar funktionieren kann. Dazu wird anhand der untersuchten Algorithmen im Prozess der Gesichtsdetektion, des Trackings, sowie des Animation-Retargetings eine Software-Lösung geschaffen. Insbesondere die Hauptkomponenten-Ansätze versprechen dabei eine automatisierte und genaue, wenn auch nicht triviale und schwer interpretierbare Transformation von Gesichtsbewegungen.

CCS Concepts

• **Computing methodologies → Motion capture**; *Tracking*; 3D imaging; Shape representations;

Keywords

Markerless Facial Tracking, Retargeting, Facial Action Units, Principal Component Analysis

1 Einleitung

Während Menschen eine sehr gut entwickelte Fähigkeit zur Erkennung und Zuordnung von Gesichtern besitzen, müssen Maschinen diese Fähigkeiten mithilfe digitaler Bildverarbeitung zunächst lernen. Dadurch können menschliche Identitäten erkannt und moderne Anwendungen, wie beispielsweise Face ID auf dem iPhone X [1] oder das ArtSelfie-Projekt von Google Arts and Culture [2], erst möglich gemacht werden.
Insbesondere in den Industrie-Bereichen der 3D-Animation, Computer Generated Imagery, als auch der Spieleindustrie, werden effiziente Algorithmen und Methoden für die Gesichtserkennung und des Trackings, sowie der Übertragung des aufgenommenen Gesichts in einen 3D-Kontext für das sogenannte 'Retargeting' benötigt, um qualitativ hochwertige synthetische Gesichtsbewegungen zu erzeugen. Überzeugende Gesichtsausdrücke in künstlich erstellten Inhalten sind deshalb so enorm wichtig, da Menschen die Fähigkeit besitzen, in Sekundenbruchteilen zwischen echt und künstlich wirkenden Gesichtern zu unterscheiden. Insbesondere der Begriff des 'Uncanny Valley' muss hierbei erwähnt werden, welcher ebenfalls unter Roboter-Ingenieuren relevant ist und die Vertrautheit des Menschen mit Robotern oder Computer-generierten Figuren aufzeigt. Gleichzeitig beschreibt der Begriff, dass diese Vertrautheit drastisch abnimmt, kurz bevor Roboter und CGI den Menschen stark ähneln, während sie wiederum stark zunimmt während sie nahezu menschenähnlich, also auch komplexer, sind. Aus diesem Umstand ergibt sich ein Tal im Graph der Vertrautheit der Menschen mit menschenähnlichen Dingen,

Betreuer: Prof. Dr.-Ing. Cristóbal Curio
Hochschule Reutlingen
Cristóbal.Curio@Reutlingen-University.de
Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen

Copyright 2018 Dominik Müller

welches noch nicht vollständig erklärt werden kann, jedoch hoch relevant in der Forschung ist [3]. In Abbildung 1 ist das 'Uncanny Valley' dargestellt.

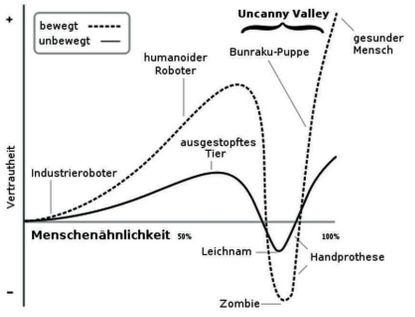


Abbildung 1: © Wikimedia Commons („Mori Uncanny Valley de" von Tobias K.)

1.1 Ziel der Arbeit

Das Ziel der Arbeit ist es, eine Auswahl an sinnvollen Algorithmen und Methoden für das Gesichts-Tracking, sowie das 'Animation Retargeting' zu finden und in einer eigenen Anwendung zu verwenden. Diese Eigenanwendung dient als Machbarkeitsstudie und wissenschaftliche Betrachtung. Hierbei soll ebenfalls darauf geachtet werden, dass die übertragenen Bewegungen des Gesichts auf dem digitalen Abbild im gültigen Bereich von Gesichtsbewegungen liegen.

1.2 Aufbau der Arbeit

Zunächst wird der Stand der Technik betrachtet, wobei als Erstes auf den Bereich der Gesichtsdetektion und die verschiedenen Methoden eingegangen wird. Anschließend folgt eine Betrachtung des Bereichs des Gesichts-Trackings. Um die Wichtigkeit der Gesichtsdetektion und des -Trackings für die Industrie herauszustellen, wird anschließend auf den Bereich des Retargetings eingegangen, welcher davon handelt, die aufgenommenen Aktivitäten auf einen dreidimensionalen Avatar zu übertragen. Ebenfalls wird eine bereits bestehende Lösung für das 'Animation Retargeting', *Dynamixyz* betrachtet.
Final soll eine Eigenanwendung entwickelt werden, welche die in der Arbeit beschriebenen Funktionalitäten abdeckt.

2 Stand der Technik

In diesem Kapitel wird der Stand der Technik für die verschiedenen Teile der Lösung aus Kapitel 3 betrachtet, wobei auf mögliche Algorithmen und Methoden eingegangen wird. Ebenfalls wird die bereits bestehende Komplettlösung *Dynamixyz* dargestellt.

2.1 Gesichts-Detektion

Mit der Detektion von menschlichen Gesichtern innerhalb von 'Frames', also einzelnen Bildern oder Bildausschnitten in Filmmaterial, beschäftigt sich die Wissenschaft schon seit langem. Das ist unter anderem dem Umstand geschuldet, dass das menschliche Gesicht von Menschen selbst extrem schnell erkannt wird, die komplexe Struktur des Gesichts aber dafür sorgt, dass eine digitale Verarbeitung von Gesichtern nicht trivial ist. Deshalb ist die Gesichtsdetektion für automatisierbare Prozesse oder Vorverarbeitungsschritte in der Informatik und dem Bereich des maschinellen Lernens höchst relevant. Eine treffsichere als auch performante Methode der Gesichtsdetektion wird deshalb immer noch aktiv gesucht, wobei die Anforderungen an diesen Gesichts-Detektor Robustheit gegenüber Orientierung, Lichtstimmung und partielle Verdeckung des Gesichts, sowie Performanz, Genauigkeit und Unübersehbarkeit sind.

2.1.1 Kaskaden-Klassifikator

Die erste und performanteste, wenn auch nicht die robusteste Möglichkeit zur Detektion von Gesichtern ist der sogenannte 'Kaskaden-Klassifikator'. Dieser Klassifikator funktioniert mit einem 'Sliding-Window', welches über das Bild fährt und für jede Position die Verhältnisse von Kontrasten anhand eines lokalen Bereichs im Fall des Haarlike-Klassifikators oder im Falle des Local Binary Pattern-Klassifikators LBP die Nachbarschaftstextur des jeweiligen Pixels betrachtet. Beide Modelle funktionieren mit vortrainierten Gewichten aus einem XML-Dokument, welches kaskadenartig durchlaufen wird und mehrere Überprüfungen enthält, um auf eine im Bild vorhandene Form zu prüfen. Aus diesem Grund gibt es diese Klassifikator nicht nur für Gesichter, sondern auch für andere Formen, wie beispielsweise Augen, Mund oder den gesamten Körper eines Menschen [4]. LBP ist im Allgemeinen schneller (10-15 Prozent schneller), als Haar, dafür aber weniger genau [5]. LBP sollte bei einer Verwendung von Kaskadenklassifizierung deshalb die Wahl für ein Embedded System sein, da alle Berechnungen in Integer vollzogen werden, während es bei Haar-Kaskaden Float-Berechnungen sind.
Dadurch, dass diese Kaskaden-Klassifikator jedoch streng auf bestimmte Verhältnisse setzen, wird es für die Detektion einer Form schwer, sobald diese Form gedreht oder gekippt wird. Ein Gesicht muss aufgrund dessen stets mindestens nahezu orthogonal und aufrecht zur Kamera zeigen, sofern es erkannt werden soll.
Für die reine Detektion wird bei den Kaskaden-Klassifikatorn ein 'Sliding-Window' Verfahren eingesetzt, welches mit einer bestimmten Fenstergröße über das Bild fährt und dieses nach den vordefinierten Verhältnissen absucht. Dabei wird das Fenster iterativ vergrößert, bis es die gesamte Größe des Bildes einnimmt. Eine zu kleine Wahl des Vergrößerungsverhältnisses führt daher auch zu langen Durchlaufzeiten.

2.1.2 dlib

Ein anderer Ansatz wird von der der Bibliothek *dlib* ausgeübt. Diese setzt nämlich nicht auf lokale Texturen oder Kontraste, sondern auf sogenannte 'Landmarks'. 'Landmarks' können sich in ihrer Gesamtheit als eine Art Gesichtsmaske vorgestellt werden, wobei einzelne Marker Gesichtsregionen repräsentieren. Der Gesichtsdetektor von *dlib* wurde mittels dem Histogram of Gradients HOG feature in Kombination mit einem linearen Klassifikator, einer Bildpyramide und einem 'Sliding-Window' Verfahren, wie bei der Kaskaden-Klassifikation, zur Detektion geschaffen. Dieser Typ von Objekt-Detektor lässt sich auch generell für bewegliche Objekte einsetzen, wobei ein Detektor auch selbst generiert werden kann [6].

2.2 Gesichts-Tracking

Beim Tracking von Gesichtern muss zunächst zwischen markerbasierter und markerloser Verfolgung von deformierbaren Objekten unterschieden werden. Beim markerbasierten Tracking befinden sich kleine Marker beispielsweise im kompletten Gesicht eines Schauspielers verteilt, welche von einer oder mehreren Kameras erfasst werden. Dadurch können, im Gegensatz zum markerlosen Tracking, auch die Bewegungen der Hautpartien, welche sich auf den Backen oder an der Stirn befinden, gut erfasst und nicht nur geschätzt werden. Dennoch ist die Verteilung der Marker und die Konfiguration relativ schwerfällig. Das markerlose Tracking ist im Gegenzug dazu einfacher zu verwenden. Aufgrund der Freiheit, alle Teile des Gesichts zu analysieren, können beim markerlosen Tracking ebenfalls die Augenbewegungen verfolgt werden. Das markerlose Tracking wird im folgenden beschrieben. Das markerlose Gesichts-Tracking ist eine Fortsetzung des Gedankens der Gesichts-Detektion. Hierbei wird angenommen, dass in mehreren Bildausschnitten hintereinander ein und dasselbe Gesicht zu sehen ist. Dieses Gesicht soll verfolgt werden, wobei die Positionen der Gesichtsmerkmale (Nase, Mund, Augen, usw.) ebenfalls verfolgt werden müssen. *Dlib*, welches bereits zuvor erwähnt wurde, kann ebenfalls für solch ein Gesichts-Tracking verwendet werden. Um ein robustes Face-Tracking

zu gewährleisten, werden 'Appearance Models' trainiert, auf welchen beispielsweise *dlib* basiert. Diese werden im Nachfolgenden betrachtet. In Abbildung 2 ist das Tracking mittels eines *dlib*-Predictors dargestellt.

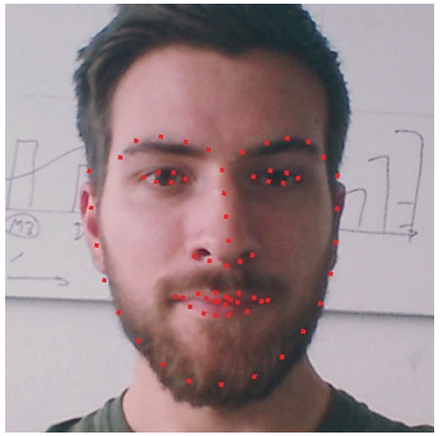


Abbildung 2: Tracking des Gesichts mittels eines *dlib*-Predictors mit 68 'Landmarks'

2.2.1 Active Shape Model Das Active Shape Model (ASM) ist ein punktverteiltes Modell, welches ein deformierbares Objekt repräsentiert. Trainiert wird das Modell mittels einer PCA auf gelabelten Bildern. Das Training wird grob durch folgende Funktion beschrieben:

$$x = \overline{x} + Pb \tag{1}$$

Hierbei wird ein neues Bild durch die durchschnittliche Form bestimmt, wobei die Modell-Punkte iterativ angepasst werden, sodass diese jeweils auf die lokale Nachbarschaft passen. Das Modell ist auf der einen Seite zwar sehr schnell zu trainieren, ist auf der anderen Seite jedoch zu simpel, um neue komplexe Bilder robust zu beschreiben. Das Modell konvergiert dadurch eventuell nicht zu einer guten Lösung [7].

2.2.2 Active Appearance Model Das Active Appearance Model (AAM) ist ein von Cootes et al. [7] entwickeltes Modell zur Charakterisierung von Klassen deformierbarer Objekte, insbesondere von Gesichtern. Das AAM wird über die Beziehung zwischen Fehler und Parameter-Anpassung in Batch-Verarbeitung trainiert. Dadurch resultiert ein lineares Modell:

$$\delta c = A\delta I \tag{2}$$

Das AAM besteht aus drei Schritten, wobei in den beiden ersten Schritten die typischen Modi der Variationen zwischen den Bildern mittels einer PCA erlernt werden [7]:

(1) Es werden mittels eines Shape-Modells die entdeckten 'Landmarks' des Objektes / Gesichts auf eine gemittelte Form gebracht.
(2) Das Ausgangsbild wird mittels Transformation auf dieselbe gemittelte Form gebracht. Dadurch sind die Variationen in unterschiedlichen Bildern auf Unterschiede in der Erscheinung der Objekte / Gesichter zurückzuführen.
(3) Auf einem neuen, unbekannten Bild wird mittels des AAM die beste Anpassung des selbigen gesucht. Dabei kann sich das AAM gemäß der erlernten Modi der Form und der Erscheinung transformieren.

2.3 Animation Retargeting

Beim 'Animation Retargeting' werden Gesichts-Animationen oder andere Bewegungen von lebendigen Körpern auf einen digitalen Avatar übertragen. Oftmals, jedoch optional, ist eine Anforderung an dieses Retargeting eine Übertragung der Bewegung in Echtzeit. Die Animationsdaten können dabei in 2D mit einer herkömmlichen Kamera oder in 3D mit stereoskopischen Kameras aufgenommen werden, wobei bei einer 2D-Aufnahme die Punktewolke der verfolgten Merkmale (Siehe 2.2) auf einen 3D-Avatar projiziert werden muss. Dies geschieht in der Regel durch eine parallele Verfolgung der Merkmale auf dem 3D-Avatar [8]–[10]. Den Schritt der Übertragung der resultierenden Punktewolke, welche die Bewegungen des echten Schauspielers repräsentieren, auf ein digitales Skelett nennt man 'Solving'. Dadurch wird das Skelett wie der Schauspieler bewegt.
Ein 3D-Scan des Gesichts birgt den Vorteil, dass die dreidimensionale Anordnung der Punkte nicht, wie beim 2D-Bild, geschätzt werden muss, da die Daten dafür bereits vorhanden sind. Dabei wird der Scan des Gesichts mit dem 3D-Modell überlagert. Dies kann, wie in [11] ein Iterative Closest Point Algorithmus (ICP) sein.

2.4 Dynamixyz

Dynamixyz ist, je nach Nutzungsschema, eine konfigurierbare Tool-Suite, mit der Gesichtsdetektion, -Tracking und Export erledigt werden können. Die Tool-Suite setzt auf das sogenannte 'Markerless'-Tracking, weshalb sie auch in nicht-professionellen Umgebungen ohne dedizierte 'Head Rigs' eingesetzt werden kann. Die 'Markerless'-Tracking Möglichkeit sorgt im Endeffekt ebenfalls dafür, dass Installationszeit und externe Abhängigkeiten gering gehalten werden, dabei aber ein Haupt-Fokus auf die verwendeten Algorithmen gelegt wird, welche wiederum an Komplexität zunehmen. Es kommen beispielsweise AAM, PCA, Eigen Tracking, Deformable Surface Models und andere Techniken zum Einsatz, um die Gesichtsfeatures in jedem Frame zu verfolgen.
Die Suite ermöglicht es Nutzern, Gesichtsdynamiken aufzunehmen und diese getrackte Session auf einen 3-dimensionalen Avatar zu übertragen.

3 Implementierung

In diesem Kapitel wird eine Eigenanwendung für das Tracking und die Übertragung von Gesichtsbewegungen auf einen digitalen Avatar vorgeschlagen.

3.1 Anforderungen

Um eine eigene Anwendung zu schreiben, welche ein markerloses Tracking eines Gesichts ermöglicht und dieses Gesicht auf einen 3D-Avatar übertragen kann, müssen folgende Module funktionieren:

- Gesichts-Detektion
- (Optional) Gesichts-Normalisierung anhand weniger Referenzpunkte (ohne Emotions-Verzerrung)
- Tracking des Gesichts in jedem Frame
- Export der detektierten Marker als Punktewolke für ein 3D-Bearbeitungsprogramm
- Animation-Retargeting für die Übertragung der Gesichtsbewegungen auf einen dreidimensionalen Avatar und der Möglichkeit zur Neukonfiguration des Gesichtsausdrucks
- (Optional) Erwartete Gesichtsausdrücke, welche aus gegebenen Gesichtsausdrücken folgen

3.2 Umsetzung

3.2.1 Umgebung
Für die eigene Software wird als Programmiersprache Python verwendet, da hierbei eine Einbettung von den nötigen Bibliotheken am leichtesten fällt. Dazu zählen unter anderem

- OpenCV
- dlib
- numpy
- scikit-learn

Für die Aufnahme von Bildmaterial wird die angeschlossene Webcam verwendet, das Bild wird mittels OpenCV abgegriffen.

3.2.2 Detektion
Wie bereits in Kapitel 2 unter verschiedenen Gesichtspunkten benannt, kann die Gesichtsdetektion auf verschiedene Art und Weise erledigt werden. Nach [12] hat sich jedoch *dlib* als besonders effizient herausgestellt. Bei

dieser Bibliothek werden nicht nur die Gesichter positionsweise erkannt, sondern ebenfalls die einzelnen Gesichts-Merkmale als eine Maske vorausgesagt. *Dlib* setzt dabei auf 68 Merkmalspunkte des Gesichts, welche sich in der Gesamtheit 'Face Predictor' nennen. Dieser 'Face Predictor' [6] basiert auf dem klassischen 'Histogram of Oriented Gradients HOG', einem linearen Klassifikator und dem 'Sliding-Window' Verfahren.
Durch Auslagerung des 'Face Predictors' in einen separaten Thread kann eine gute Effizienz und Performanz erreicht werden, wobei die Webcam nicht blockiert wird und die Ergebnisse des 'Face Predictor' im Hintergrund verarbeitet werden.

3.2.3 Normalisierung

Da aus dem Gesicht die intrinsischen Bewegungen erhalten bleiben sollen, also die Emotionen nicht normalisiert werden, ist es stets besser, die Kamera immer orthogonal zum Gesicht zu platzieren, bzw. das Gesicht nicht stark zu bewegen. Jedwede Normalisierung kann dabei für Verzerrungen sorgen. Sollte das Gesicht dennoch normalisiert werden, so muss dies anhand drei statischer Punkte im Gesicht passieren, aus denen eine Ebene zur affinen Transformation [12] erstellt wird. Sobald die komplette Maske auf die Ursprungsposition zurückbewegt wird, kann nämlich keine Analyse mehr stattfinden. Für die affine Transformation wird *OpenCV* verwendet

3.2.4 Tracking

Für das Tracking wird die Vorannahme getroffen, dass *dlib* das Gesicht sofern es orthogonal zur Kamera steht, stets gut verfolgt werden kann. Eine Bedingung, die bei *dlib* stets erfüllt werden sollte, ist die Konvertierung des Webcam-Bilds in Graustufen, für eine bessere Erkennung des trainierten 'Face Predictors'. Um ein durchweg performantes Ergebnis zu erzielen, kann die Detektion für jedes einzelne Bild sicherer gemacht werden, indem zunächst ein Kaskaden-Klassifikator aus *OpenCV* verwendet wird, welcher auf einem separaten Thread läuft und mittels eine 'Haar-Like'- oder 'LBP'-Klassifikator ein Gesicht im aktuellen Bild sucht. Sobald ein Gesicht gefunden wurde, wird vom Klassifikator ein Rechteck zurückgegeben, in dem sich das Gesicht befindet. Sofern mehrere Gesichter gefunden werden, wird das Gesicht verwendet, dessen repräsentierendes Rechteck am größten aller gefundenen Rechtecke ist. Auf diesem Rechteck wird nun der *dlib*-Predictor angewandt. Wichtig, je nach verwendetem Predictor sind nun die verwendeten 'Landmarks', die den gefundenen Gesichtsmerkmalen entsprechen. Diese 'Landmarks' müssen in Kapitel 3.2.5 mit den Merkmalen eines 3D-Gesichtsmodells übereinstimmen.

3.2.5 Export

Um das aufgenommene Video mit den Informationen der Gesichtsbewegungen zu exportieren und ein Retargeting in einem 3D-Programm vorzunehmen muss nun die 2D- oder 3D-Information der verfolgten Gesichtspunkte exportiert werden. Dazu wurde für jedes Bild des Videos die korrespondierenden Positionsinformationen der Punkte gespeichert, welche nun als *numpy*-Array vorliegen. Der Umstand kann mittels 3D-Scans des Gesichtes umgangen werden, doch wird hierfür die entsprechende Ausrüstung benötigt.
Es können aus den vorhandenen 2D-Datenpunkten 3D-Datenpunkte mit der Annahme, dass diese in einem Gesicht liegen, geschätzt werden. Dafür wird [8] als Referenz verwendet. Hierbei wird ein grobes 3D-Gesichtsmodell erstellt, das wie die 'Landmarks' aus Kapitel 3.2.4 ebenfalls die gleichen 'Landmarks' besitzen muss. Das grobe Gesichtsmodell stellt einen 'Mesh' (Polygonnetz zur Modellierung von 3D-Modellen) eines Gesichts dar. Die 2D-Punktewolke der 'Landmarks' kann dadurch mit denen, die auf dem 3D-Modell liegen, ausgerichtet werden [8]. Um die entsprechenden 'Landmarks' im 3D-Modell zu finden, kann ebenfalls der zuvor behandelte *dlib*-Predictor verwendet werden. Dadurch sind, als positiver Nebeneffekt, die 'Landmarks' konsistent. Mit der Annahme, dass das 2D-Kamera-Modell eine schwache Perspektive entlang der Z-Achse darstellt, kann die Projektionsmatrix als $P = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \end{bmatrix}$ beschrieben werden [8].
Die 3D-Information kann nun in gängigen 3D-Bearbeitungsprogrammen als Punktwolke verwendet werden, um ein digitales Skelett zu bewegen. Eine andere Vorgehensweise mittels Zerlegung des verfolgten Gesichts in Komponenten wird unter anderem in Kapitel 3.2.6 betrachtet.

3.2.6 Animation Retargeting

Um für das spätere Retargeting mehrere Möglichkeiten für die Visualisierung des künstlichen 3D-Gesichts zur Verfügung zu stellen, soll ein System basierend auf 'Blend Shapes' geschaffen werden. Dadurch ist es möglich, Gesichtskomponenten einzeln zu bewegen und bereits aufgenommene Gesichter mit neuen Gesichtsanimationen und Emotionen auszuschmücken oder bereits vorhandene Bewegungen zu verstärken. Zwar ist es möglich, das 3D-Modell des Gesichts auf verschiedene Arten und Weisen zu manipulieren, zielführend sind dabei jedoch nur wenige. Beispielsweise kann das 3D-Modell mit Oberflächenbearbeitungswerkzeugen des Bearbeitungsprogramms deformiert werden, hier wird aber nicht darauf geachtet, dass die resultierende Form ein valides Gesicht ergibt [13]. Der Faktor eines validen Gesichts ist besonders relevant, da die Manipulation der Gesichtsbewegungen nur in einem bestimmten Rahmen stattfinden darf, sodass keine Form resultiert, die nicht mehr einem Gesicht ähnelt. Weitere Möglichkeiten zur Manipulation beinhalten die PCA und die lineare Rekombination von verschiedenen 'Blendshape-Targets', welche einen maximalen Ausdruck einer Emotion darstellen.
Eine Darstellung einer durchgeführten PCA auf den maximalen Ausdrücken von Ausdrücken, und die Einordnung dieser Ausdrücke in einen zweidimensionalen Raum, kann in Abbildung 3 eingesehen werden. Hierbei wurden die Gesichtsausdrücke 'Überraschung', 'Lachen', 'Lächeln', 'Kuss', 'Angst', 'Ekel', 'Verachtung', 'Wut', 'Traurigkeit', sowie der Normalausdruck verwendet.
Der hier verfolgte Ansatz involviert eine Erzeugung von 'Blend Shapes' mittels einer PCA, welche bei Anwendung auf die aufgenommenen Daten automatisch ein 'Blend Shape'-Modell zurückgibt. Der Vorteil einer PCA liegt darin, dass das Modell automatisch erstellt wird und die Daten akkurat repräsentiert. Zwar funktioniert die Modellerstellung automatisch, sorgt aber zugleich für ein Modell, das nicht intuitiv von einem Menschen zu bedienen ist [14].
Die resultierenden 'Blend Shapes' stellen die Eigenvektoren des Gesichts (oder auch die Hauptkomponenten) dar [15]. Um Komponenten des Gesichts für die spätere Übertragung zu erzeugen, wird die PCA auf die komplette Animation, also den Videoausschnitt, angewandt. Das Gesicht wird durch 'Vertices', Referenzpunkte, repräsentiert. Die PCA generiert für jede dieser 'Vertices' eine Hauptausrichtung, welche aus dem Eigenvektor mit dem maximalen Eigenwert erzeugt wird. Für jede zwei 'Vertices' im Gesichtsmodell, deren Winkel zueinander innerhalb einer bestimmten Toleranz liegt, wird die selbe Hauptausrichtung angenommen [16]. In Abbildung 4 sind die Gesichtskomponenten sichtbar, die aus der beschriebenen PCA folgen.
Um nun mit den resultierenden Gesichtskomponenten eine Manipulation durchzuführen, müssen die Gesichtskomponenten vom Menschen jedoch verstanden werden. Die resultierenden Komponenten (Siehe Abbildung 3) sind jedoch nicht semantisch. Um ein semantisches Modell für die Manipulation der 'Blend Shapes' zu erzeugen, wird, wie

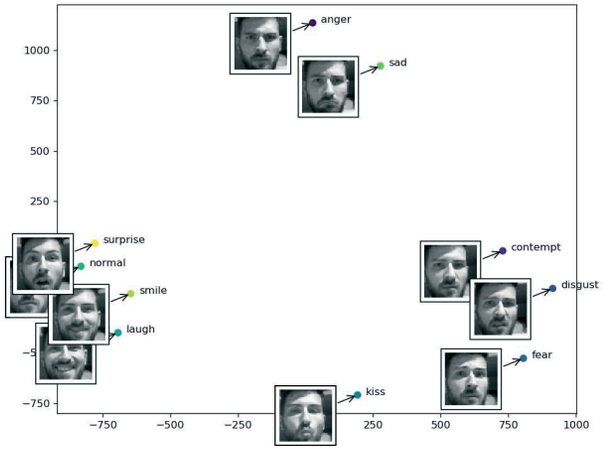
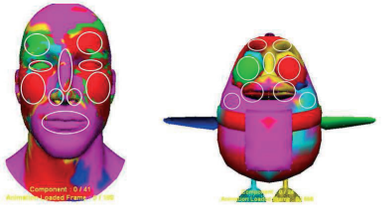


Abbildung 3: Plot einer durchgeführten PCA mit 2 Komponenten auf Maximalausdrücken eines Gesichts



(a) (b)

Abbildung 4: Gesichtskomponenten, die von einer PCA gefunden wurden [16]

in [11], mittels der 'Peak Frames' von Emotionen (Siehe Abbildung 3) und einer jeweiligen Anordnungsmethode der Singularwertzerlegung (SVD) bezüglich des Normalausdrucks ein morphing-fähiges 3D-Gesichtsmodell erstellt [11].

3.2.7 Erwartete Ausdrücke

Mittels einer PCA können ebenfalls erwartete Gesichtsausdrücke aus gegebenen Ausdrücken abgeleitet werden. Dazu müssen Gesichtsausdrücke für jede der zuvor benannten Gesichtskomponenten extrahiert werden, der Start-Frame und der End-Frame eines jeden Ausdrucks bestimmt werden muss - Dadurch kann eine Animationskurve durch die zuvor beschriebene Methode der Hauptausrichtungen konstruiert werden. Die Horizontale Achse des Graphen bestimmt die neutrale Position der Komponente [16]. Hierbei wird sichtbar, dass eine Komponentenbewegung entweder positiv oder negativ sein kann.
Aus dieser Animationskurve kann nun ein erwarteter Ausdruck für einen jeden Gesichtsausdruck gefunden werden, das heißt auf einen Gesichtsausdruck oder Emotion folgt mit hoher Wahrscheinlichkeit ein anderer Gesichtsausdruck. Laut [16] kann eine solche 'Anticipation Expression' in der Animationskurve gefunden werden, indem für jeden Gesichtsausdruck der entsprechende Abschnitt in der Kurve untersucht wird. Für einen bestimmten Gesichtsausdruck auf der Animationskurve können nun beliebig viele Antizipationseffekte, also resultierende Ausdrücke auf der Kurve gefunden werden. Dabei ist es jedoch wichtig, nur einen erwarteten Ausdruck zu wählen, da viele erwartete Ausdrücke in Kombination zu einer abnormalen Verzerrung des Gesichts führen würden, welche die Topologie des Gesichtsmodells missachtet. Die beste Voraussage eines zukünftigen Ausdrucks ist derjenige Ausdruck, der am häufigsten auf allen Gesichtskomponenten aufgetreten ist. Da dadurch aber oft nur kleine Bewegungen des Gesichts resultieren, wird ebenfalls die Differenz beider Ausdrücke mit einbezogen [16].

4 Fazit

Die Implementierung hat gezeigt, dass eine Lösung zur Detektion von Gesichtern und Verfolgung der Bewegungen, sowie einer komponentenbasierten Analyse, sowie eines semantischen Retargetings machbar aber auch mit mehreren Komplexitäten verbunden ist. Ebenso werden Software-Lösungen, wie *Dynamixyz* immer relevanter, da der Bedarf an Übertragungen von Daten aus der echten Welt in virtuelle Umgebungen immer wichtiger werden. Wenige Filme kommen heutzutage ohne computergeneriertes Bildmaterial aus, in virtuellen Umgebungen sind digitale Avatare nicht mehr wegzudenken. Dennoch sind professionelle Lösungen, immer noch mit großen finanziellen Investitionen verbunden, weshalb sie oft auch nur von großen Studios und Firmen verwendet werden können. Wie die Implementierung aber zeigt, ist Software für wenig Budget schon länger denkbar. Höchstwahrscheinlich wird sich der Markt in diese Richtung auch für kleinere Studios, akademische Projekte oder den Hobby-Sektor öffnen, weshalb auch für kleine Produktionen und 'Indy'-Titel in Zukunft mit professionellen Gesichtsanimationen zu rechnen ist. Insbesondere die in Kapitel 3.2.6 gezeigten Zerlegungen in Hauptkomponenten und die in Kapitel 3.2.7 dargestellten erwarteten Gesichtsausdrücke demonstrieren einen besonderen Mehrwert einer Analyse von Gesichtsausdrücken.

5 Glossar

AAM: Active Appearance Model. Verfahren der Bildverarbeitung, mit dem sich besonders gut Klassen deformierbarer Objekte charakterisieren lassen

ICP: Iterative Closest Point. Algorithmus zur Anpassung von Punktwolken aneinander

LBP: Local Binary Patterns. Visueller Deskriptor für die Klassifizierung in der Bildverarbeitung

PCA: Principal Component Analysis. Verfahren der multivariaten Statistik, das dazu dient, umfangreiche Datensätze zu strukturieren, zu vereinfachen und zu veranschaulichen

SVD: Singular Value Decomposition. Darstellung einer Matrix als Produkt dreier spezieller Matrizen, woraus die Singulärwerte der Matrix abgelesen werden können

Literatur

[1] A. Inc. (Juni 2018). iPhone X - Face ID. Aufgerufen am: 30.06.2018, Adresse: https://www.apple.com/de/iphone-x/#face-id.

[2] G. A. Michelle Luo Product Manager und Culture. (Jan. 2018). Exploring art (through selfies) with Google Arts and Culture. Aufgerufen am: 30.06.2018, Adresse: https://www.blog.google/outreach-initiatives/arts-culture/exploring-art-through-selfies-google-arts-culture/.

[3] D. Neumann. (Aug. 2015). Uncanny Valley: Unheimliche Begegnungen der Roboterart. Aufgerufen am: 12.10.2018, Adresse: http://www.roboterwelt.de/magazin/uncanny-valley-unheimliche-begegnungen-der-roboterart/.

[4] Harshdeep Singh, "Object Detection using Haar–like Features: Visual Recognition and Search", Nr. CS 395T, 2008. Adresse: http://www.cs.utexas.edu/

~grauman/courses/spring2008/slides/Faces_demo.pdf.

[5] R. Heymsfeld. (Mai 2016). Haar LBP and HOG - Experiments in OpenCV Object Detection. Aufgerufen am: 27.10.2018, Adresse: http://robotics.hobbizine.com/opencv.html.

[6] dlib. (Juli 2018). face_detector. Aufgerufen am: 04.07.2018, Adresse: http://dlib.net/face_landmark_detection.py.html.

[7] T. F. Cootes, G. J. Edwards und C. J. Taylor, "Active Appearance Models", *IEEE Trans. Pattern Anal. Mach. Intell.*, Jg. 23, Nr. 6, S. 681–685, Juni 2001, ISSN: 0162-8828. DOI: 10.1109/34.927467. Adresse: https://doi.org/10.1109/34.927467.

[8] Jiang, Luo, Zhang, Juyong, Deng, Bailin, Li, Hao, Liu und Ligang, "3D Face Reconstruction with Geometry Details from a Single Image", *IEEE Transactions on Image Processing*, Jg. 27, Nr. 10, S. 4756–4770, 2018, ISSN: 1057-7149. DOI: 10.1109/TIP.2018.2845697. Adresse: http://arxiv.org/pdf/1702.05619v2.

[9] L. Dutreve, A. Meyer und S. Bouakaz, *Feature points based facial animation retargeting*, Jan. 2008. DOI: 10.1145/1450579.1450621.

[10] Ribera, R. B. i., Zell, Eduard, Lewis, J. P., Noh, Junyong, Botsch und Mario, "Facial retargeting with automatic range of motion alignment", *ACM Transactions on Graphics*, Jg. 36, Nr. 4, S. 1–12, 2017, ISSN: 07300301. DOI: 10.1145/3072959.3073674.

[11] C. Curio, M. Breidt, M. Kleiner, Q. C. Vuong, M. A. Giese und H. H. Bülthoff, "Semantic 3D Motion Retargeting for Facial Animation", in *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, Ser. APGV '06, Boston, Massachusetts, USA: ACM, 2006, S. 77–84, ISBN: 1-59593-429-4. DOI: 10.1145/1140491.1140508. Adresse: http://doi.acm.org/10.1145/1140491.1140508.

[12] D. Müller, *Entwicklung einer Lösung zur Detektion, Normalisierung und Erkennung von Ähnlichkeiten zwischen Gesichtern*, Fortgeschrittene Bildverarbeitung, 2018. Adresse: https://drive.google.com/file/d/1ReEytEAuinrFsWIiNhzmHUWYULCeEv-/view?usp=sharing.

[13] J. P. Lewis und K.-i. Anjyo, "Direct manipulation blendshapes", *IEEE computer graphics and applications*, Jg. 30, Nr. 4, S. 42–50, 2010. DOI: 10.1109/MCG.2010.41.

[14] J. Lewis und F. Pighin, "Retargeting: algorithms for performance-driven animation", Jan. 2006. DOI: 10.1145/1185657.1185851.

[15] Matthew Turk and Alex Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Jg. 3, Nr. 1, 1991. Adresse: http://www.face-rec.org/algorithms/PCA/jcn.pdf.

[16] J.-j. Choi, D.-s. Kim und I.-k. Lee, "Anticipation for facial animation", in *In Proceedings of 17th International Conference on Computer Animation and Social Agents*, 2004, S. 1–8. Adresse: https://www.researchgate.net/publication/228947429_Anticipation_for_facial_animation.

Studie zum Vergleich der Sprachassistenzsysteme Siri und Google Assistant bezüglich des Einsatzes als Fahrassistenzsystem

Gamze Karatas
Hochschule Reutlingen
Gamze.Karatas@Student.Reutlingen-University.DE

Abstract

Zwei marktführende Systeme Google Assistant von Google und Siri von Apple werden im Einsatz als Fahrassistenzsystem getestet, vgl. [1]. Dabei wird die Sprachsteuerung beider Systeme benutzt. Das Bereits im Smartphone integrierte Sprachsteuerungssystem ermöglicht dem Nutzer, durch Sprechen, Befehle zu geben, welche dann durch das System erkannt und bearbeitet werden.

Die Arbeit repräsentiert, wie schnell und zufriedenstellend für den Nutzer, die Systeme auf akustische Fragen und unter realen Umwelteinflüssen reagieren. Durch eine Evaluierung werden die Stärken und Schwächen beider Systeme untersucht.

Die methodische Vorgehensweise dieser Forschung setzt sich aus ausformulierten Szenarien mit präzisen Aufgabenstellungen zusammen. Diese werden in unterschiedlichen Szenarien durchgespielt, anschließend bewertet und miteinander verglichen.

Schlüsselwörter

Sprachassistenzsysteme, Siri, Apple, Google Assistant, Google, Usability, Evaluierung

CCS Concepts

• Human-centered computing → Usability testing; HCI theory, concepts and models

Betreuer Hochschule: Prof. Dr. Gabriella Tullius
Hochschule Reutlingen
Gabriella.Tullius@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Gamze Karatas

1 Einleitung

Im folgenden Kapitel wird die Problemstellung ausgearbeitet.

1.1 Problemstellung

Beim Fahren eines Fahrzeugs entsteht eine Interaktion zwischen Menschen und der Maschine. Der Mensch bedient es, indem er Beispielsweise lenkt und zugleich mit den Füßen Gas gibt. Dadurch ist der Fahrer in der Lage das Fahrzeug zu beeinflussen. Durch die Rückmeldung des Fahrzeugs besteht eine gegenseitige Interaktion. Dabei kann die Interaktion zwischen Menschen und Maschine, durch beispielsweise Umwelteinflüsse, wie Gewitter oder Helligkeit beeinflusst und dadurch beeinträchtigt werden vgl. [13, S. 6].

Statistiken zeigen, dass Menschen in Deutschland beim Autofahren nebenher ein Smartphone benutzen.

Dabei werden Aktionen wie das Telefonieren, das Schreiben von Nachrichten, das Navigieren sowie das Bedienen von anderen Elementen durchgeführt, vgl. [3].

Aufgrund dieser Nebentätigkeiten während der Autofahrt, wird die Konzentration des Fahrers vom Straßenverkehr abgelenkt. Hieraus folgt ein erhöhtes Risiko an Verkehrsunfällen. Viele Studien zeigen auf, dass beispielweise das Bedienen und Nutzen eines Smartphones während der Fahrt ein hohes Ablenkungspotenzial darstellt, vgl. [2].

Aus diesem Hintergrund stellt sich Frage, ob die Steuerung des Smartphones per Sprachbefehl ein geringeres Ablenkungspotenzial ermöglicht. Laut der US-amerikanischen Stiftung für Verkehrssicherheit (AAA), lenken Sprachassistenzsysteme den Fahrer ebenfalls ab, vgl. [14]. Dennoch weniger als die manuelle Bedienung eines Bildschirms, vgl. [12]. In diesem Kontext werden die Grenzen ausgewählter Assistenzsysteme getestet und bewertet.

1.2 Zielsetzung

Im Gegensatz zu stationären- werden bei mobilen Geräten oft spontane Suchen und somit schnelle Ergebnisse erwartet, vgl. [7, S.31]. *„Okay Google“* und *„Hey Siri“* sind für viele Smartphone-Nutzer bereits bekannt, ebenso wie ihre Funktion das bedienen ohne Eingreifen per Hand, vgl. [15].

Doch inwieweit unterstützen die Sprachassistenzsysteme den Fahrer? Werden Befehle direkt nach dem Sprechen vom System erkannt oder ist eine zusätzliche manuelle Eingabe auf dem Bildschirm erforderlich? Worin unterscheiden sich beide Systeme im Einsatz als Fahrassistenzsystem?

Um wesentliche Unterscheidungsmerkmale zu erkennen, sowie Schlüsse über das Ablenkungspotenzial ziehen zu können, werden zwei beliebte Sprachassistenzsysteme, Siri und Google Assistant, verglichen. Dafür werden beide Systeme unter realen Bedingungen und natürlichen Einflussfaktoren, als Fahrassistenzsysteme eingesetzt und einander gegenübergestellt.

2 Grundlagen

In diesem Kapitel werden die Grundlagen beschrieben. Zuerst wird Siri vorgestellt und anschließend Google Assistant. Im Kapitel „Tertiäre Aufgaben“, werden Begrifflichkeiten wie primäre, sekundäre und tertiäre Aufgaben erklärt.

2.1 Siri

Siri ist eine Sprachsteuerungssoftware von Siri Inc., welche im Jahre 2010 von Apple übernommen wurde. Seitdem ist es kein eigenständiges Unternehmen mehr. Siri muss nicht vorher installiert werden, da sie bereits im Betriebssystem der IOS integriert ist, vgl. [7].

Durch “Hey Siri“ oder langes Drücken des Home-Buttons wird Siri aktiviert. Öfter gestellte Fragen oder Anregungen werden in der Datenbank gespeichert und dadurch personalisiert. Dies ermöglicht schnellere, optimalere Ergebnisse, vgl. [15].

Auf interessante Anregungen wie z.B. *„Hey Siri, singe mir etwas vor.“,* singt Siri ein Lied vor. Auch bei Fragen wie *„Hey Siri, wie geht es dir?“* bekommt der Nutzer eine angemessene Antwort.

2.2 Google Assistant

Im Jahr 2012 wurde Google Assistant zum ersten Mal auf den Smartphones zur Verfügung gestellt. Entwickelt wurde der Sprachassistent von Google LLC. Bekannt wurde Google Assistant durch die Suchmaschine Google und ist auch auf dem Betriebssystem iOS von Apple verfügbar. Mit *„Ok Google“* kann Google Assistant angesprochen werden, vgl. [7].

Unter anderem können durch Google Assistant Informationen bezüglich Öffnungszeiten von einem Restaurant abgefragt werden. Vergleichsweise einem Dialog, können Fragen, nacheinander gestellt werden, ohne “Ok Google” voranzustellen vgl. [7], [8].

2.3 Tertiäre Aufgaben

Handlungen wie das Schalten, Kuppeln, Anhalten und Bremsen gehören zum Autofahren dazu. Betrachtet man diesen Prozess innerhalb einer Zeitspanne, so sind Wiederholungen von diversen Aufgaben ersichtlich.

Diese Aufgaben lassen sich in primäre, sekundäre und tertiäre unterscheiden. Die primäre Aufgabe, beim Autofahren, wäre die eigentliche Fahrzeugführung. Die Anpassung der Geschwindigkeit oder das Einhalten von Fahrspuren gehören ebenfalls dazu, vgl. [13, S.5].

Sekundäre Aufgaben beschreiben Reaktionen auf Verkehrssituationen. Handlungen wie Schalten oder starten des Fensterwischers zählen dazu. Diese Aufgaben dienen aber nicht zur eigentlichen Fahrzeugführung, vgl. [13, S.6].

Tertiäre Aufgaben zählen im Gegensatz zu primären und sekundären Aufgaben, zu Unterhaltungs- und Komfortzweck. Beispiele dafür wären Musik hören, Telefonieren oder das Bedienen von einem Smartphone. Diese jedoch können den Fahrer vom Fahren ablenken, vgl. [13, S.6ff]. In dieser Arbeit wird das Ablenkungspotenzial der tertiären Aufgaben betrachtet.

3 Leitfragen

In diesem Kapitel werden zwei Leitragen vorgestellt. Das Ziel ist, relevante und explizite Aufgabenstellungen und Szenarien für die Untersuchung zu gewinnen.

Die Verifizierung dieser Fragen erfolgt in den darauffolgenden Kapiteln. Die Leitfragen lauten dabei wie folgt:

1. *Welche Funktionen von Sprachassistenzsystemen werden beim Fahren eines PKWs hauptsächlich genutzt?*
2. *Ab wann wird das Sprachassistenzsystem beim Fahren eines PKWs durch den Lärm beeinträchtigt?*

3.1 Nutzung von Sprachassistenzsystemen

Dieses Kapitel bezieht sich auf die erste Fragestellung und untersucht deshalb die meistbenutzen Anwendungsbereiche der Sprachassistenzsysteme.

Dafür wird eine Statistik (Abbildung 1) herangezogen, welche aktuelle und zukünftig vorstellbare Anwendungsfälle von digitalen Sprachassistenzsysteme aufzeigt.

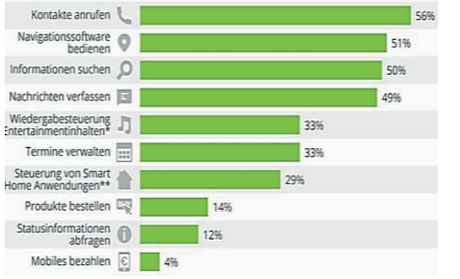


Abbildung 1: Anwendungsbereiche von digitalen Sprachassistenzsystemen [3]

Infolge dieser Statistik lassen sich folgende Aufgabengebiete resultieren:

- Kontakte anrufen
- Navigationssoftware bedienen
- Informationen suchen
- Nachrichten verfassen
- Wiedergabesteuerung von Entertainmentinhalten
- Termine verwalten

Aus der Auflistung ergeben sich insgesamt sechs Fragestellungen für das Testen beider Systeme. Es werden die Anwendungsbereiche ab dem Prozentsatz von 33 und aufwärts in Betracht gezogen, um den zeitlichen Untersuchungsrahmen einzuhalten.

Das britische Informa-Istitut zeigt auf, dass im Jahr 2012 täglich 19,1 Milliarden Nachrichten über Dienste wie WhatsApp

verschickt wurden. Das Verschicken von SMS wurde 17,6 Milliarden Mal pro Tag genutzt, vgl. [18].

Daher wird WhatsApp in der Aufgabenstellung für die Vermittlung von Nachrichten ausgewählt.

Aufgrund der Statistik aus dem Jahr 2016 hört beinahe jeder zweite Nutzer Musik über Anbieter wie Spotify, vgl. [19]. Dies wurde ebenfalls in der Formulierung der Fragestellung aus den Anwendungsbereichen berücksichtigt.

Die Fragestellungen aus den sechs Anwendungsbereichen formulieren sich wie folgt. Zu Beginn wird durch *„Hey Siri“* und *„Ok Google“* die Anwendung gestartet.

1. *„ …rufe Max an.“*
2. *„ …navigiere mich zum Stuttgarter Hauptbahnhof.“*
3. *„ …was heißt auf Englisch Zufriedenheit?“*
4. *„ …schreibe eine WhatsApp Nachricht an Max mit dem Text, wie geht es dir?“*
5. *„ …spiele mir aus Spotify Lieder von Eminem ab.“*
6. *„ …notiere mir einen Arzttermin für morgen auf 15 Uhr.“*

3.2 Spracherkennungsprozess

Bevor die Audiosignale weiterverarbeitet werden, findet eine Vorbereitung der Signale statt. Dabei werden relevante Signale von Störsignalen wie z.B. Rauschen gefiltert.

Anschließend werden die Zeitsignale in kurze und gleichmäßige Frames zerlegt und mit einer Amplitudenhüllenkurve multipliziert. Aus den gewonnenen Feature-Vektoren werden die Phoneme klassifiziert, vgl. [5]. Phoneme sind einer der kleinsten Spracheinheiten, die ein Wort von einem anderen Wort unterscheiden lassen. Aus diesen Phonemen finden eine Worterkennung bzw. ein Vergleich mit dem Vitebi-Algorithmus statt. Dies geschieht mit Hilfe des Hidden Markov Models (HMMs), vgl. [4]. Zunächst wird zwischen einer Einzelworterkennung und einer kontinuierlichen Sprache unterschieden. Anschließend mit dem Vitbi-Algorithmus nach möglichen Kombinationen mit dem Eingangssignal verglichen. Durch eine grammatische oder statistische Analyse wird schließlich aus der Wortmenge heraus, das schlagkräftigste Wort ausgewählt, vgl. [5].

Eine statische Analyse stellt die Methode des Natural Language Processings (NLPs) dar. Dies ist eine Methode, die eine Klassifizierung der Daten ermöglicht und diese dann als Parameter an das HMM übergibt. Dies unterstützt dadurch eine kurzfristige Bearbeitung der Daten. Generell wird diese für Rechtschreibprüfung, Sprachsynthese und Spracherkennung eingesetzt, vgl. [6].

Für eine optimale Nutzung von Siri und Google Assistant kann das System personalisiert werden. Dies geschieht durch die Aufnahme von Stimmproben, bei erstmaliger Nutzung der Systeme, vgl. [15], [16].

Beim Vorbereiten der Signale erfolgt eine Filterung von relevanten und irrelevanten Frequenzen. Das menschliche Ohr kann Töne von etwa 16 bis 16.000 Hz hören. Die Schmerzgrenze liegt bei 20 Pa. Im Gegensatz zum Ohr, nimmt das Mikrofon den Schalldruck unabhängig von der Frequenz auf, vgl. [19, S. 4].

Für die Untersuchung wird an dieser Stelle die zweite Leitfrage herangezogen:

„Ab wann wird das Sprachassistenzsystem beim Fahren eines PKWs durch den Lärm beeinträchtigt?“

Die im Auto auftretenden Störquellen, wie das Radio und offenes Fenster, werden hierfür zum Testen der Sprachassistenzsysteme verwendet, da diese den Schallpegel erheblich erhöhen können, vgl. [20]. Dafür werden Szenarien aufgestellt und die erwähnten Störquellen kombiniert.

Ziel dabei ist es herauszufinden, inwieweit relevante Frequenzen trotz erhöhtem Schallpegel erkannt werden.

Szenario 1: Das Fenster ist geschlossen und die Musik ist aus.

Szenario 2: Das Fenster ist geöffnet und die Musik ist aus.

Szenario 3: Das Fenster ist geschlossen und die Musik ist an.

Szenario 4: Das Fenster ist geöffnet und die Musik ist an.

4 Methodik und Vorgehensweise

In diesem Kapitel werden die Methodik und die Vorgehensweise der Untersuchung beschrieben.

Für die Vorbereitung wurden zwei Leitfragen im Kapitel ‚Leitfragen‘, aufgestellt. Dadurch konnten sechs Aufgabenstellungen und vier Szenarien gebildet werden. Für die Untersuchung wird eine Usability-Evaluation eingesetzt und die nutzerbasierte Methode verwendet.

Vor der eigentlichen Durchführung startet ein Pilottest. Dadurch soll die Klarheit der Aufgaben für einen reibungslosen Ablauf geprüft und sichergestellt werden. Für die nutzerbasierte Evaluation werden sechs Probanden eingesetzt. Dabei dienen die persönlichen Geräte und Fahrzeuge als Testobjekte. Diese werden im Auto der Reihe nach in einer Halterung befestigt und anschließend das Szenario ausgeführt.

4.1 Anforderungen

Die Anforderungen für die Untersuchung sind, ein Internet zugängliches Smartphone, eine Smartphone-Halterung, Verkabelung für die Stromversorgung, eine Applikation, die die Lautstärke misst und ein funktionstüchtiges Auto.

Außerdem ist es wichtig, dass die Probanden einen Kontakt namens Max in der Kontaktliste besitzen, sowie die Applikation WhatsApp und Spotify. Durch dieses Vorgehen aller Probanden können die Aufgabenstellungen immer unter den gleichen Voraussetzungen und Umständen durchgeführt werden.

4.2 Aufbau Usability-Evaluation

Verschiedene Methoden ermöglichen es, die Bedienbarkeit einer Anwendung zu testen. Eine dieser Methode ist der Usability-Test.

Für den Usability-Test stehen eine experten- und eine nutzerbasierte Evaluation zur Verfügung. Für die expertenbasierte Methode werden Experten mit fachlichem Wissen eingesetzt und bei dem nutzerbasierten hingegen die tatsächlichen Nutzer. Allerdings sollte hier im Vorfeld die Zielgruppe klar definiert werden. Außerdem können Methoden wie „Thinking Aloud“ kombiniert werden. Dabei sagen die Nutzer laut was sie denken. In dieser Arbeit wird die nutzerbasierte Methode verwendet. Dabei werden reale Anwender, die ausformulierten Szenarien durchspielen. Zur Veranschaulichung wird der gesamte Prozess visuell dargestellt (Abbildung 2).

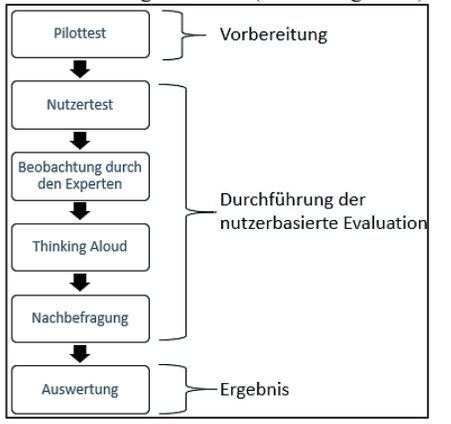


Abbildung 2: Ablaufprozess

Im späteren Teil der Arbeit, werden auf die in Abbildung 2 beschriebenen Prozesse näher eingegangen.

4.3 Durchführung

Bevor die Untersuchung durchgeführt wird, findet der Pilottest statt, um die Verständlichkeit der Aufgaben zu überprüfen. Dabei werden die Aufgabenstellungen nacheinander an Siri und Google Assistant gestellt.

Da diese Phase nicht reibungslos verlief, mussten die Aufgabenstellungen erneut formuliert werden. Von *„…wie komme ich zum Stuttgarter Hauptbahnhof?“* zu *„…navigiere mich zum Stuttgarter Hauptbahnhof“*. Da Google Assistant bei der ersten Formulierung eine Wegbeschreibung und keine direkte Navigation anzeigt. Ansonsten wurden für die Untersuchung keine weiteren Änderungen vorgenommen.

Der Usability-Test wird von einem Usability Experten begleitet. Dabei wurden die Aufgaben, welche als Papierformat vorliegt, laut für den Fahrer ohne das Ansprechen der Sprachassistenzsysteme vorgelesen. Dadurch konnte sich der Teilnehmer auf die Fahrt und auf das Ergebnis des Systems konzentrieren. Anschließend wurde das System von dem Teilnehmer angesprochen und die jeweilige Frage gestellt. Sowohl die nebensächlichen Kommentare (Thinking Aloud) aus der Nutzerseite, als auch die systemseiteigenen Ergebnisse, wurden dokumentiert und für die Auswertung in Betracht gezogen.

Die Abbildungen 3 und 4 zeigen, welche Geräte und welches Fahrzeug dafür eingesetzt wurden.

Siri Nutzer	Proband 1	Proband 2	Proband 3
Smartphone	iPhone X	iPhone 7	iPhone 8
Alter	23	27	24
Geschlecht	w	w	m
Fahrzeug	Audi A1	BMW E36	VW Passat

Abbildung 3: Angaben zum Probanden von Siri

Google Assistant Nutzer	Proband 1	Proband 2	Proband 3
Smartphone	Galaxy S7 Edge	Galaxy S8	Galaxy S9
Alter	25	32	44
Geschlecht	m	m	w
Fahrzeug	VW Fox	VW Golf	BMW E91

Abbildung 4: Angaben zum Probanden von Google Assistant

Die Untersuchung wird dabei von einem Experten mit Fachwissen durchgeführt, vgl. [9]. Währenddessen beobachtet er, um Unterbrechungen auszuschließen. Damit werden für die Probanden die gleichen Bedingungen garantiert, vgl. [10].

Die folgende Abbildung 5, zeigt Ausschnitt der Durchführung.

Abbildung 5: Durchführung der Untersuchung

Die Musik wird bei Szenario drei und vier mit einem Lautstärkenpegel von ca. 80 dB eingestellt.

Da jedes der vier Szenarien von jedem Probanden mit jeweils sechs Fragen durchgeführt wurde, hat jeder Proband insgesamt 24 Anfragen an den Sprachassistenten gestellt.

Die Summe beträgt deshalb bei direkter Erkennung 72 Anfragen pro Sprachassistenzsystem. Der Zeitaufwand betrug dabei ca. 7 Minuten pro Szenario.

5 Auswertung

Für die Auswertung der Ergebnisse werden die weiter unten folgenden Entscheidungskriterien verwendet.

Diese teilen sich in drei unterschiedliche Kategorien, wie hervorragend, mittelmäßig und mangelhaft, auf. Dabei werden folgende Symbole verwendet:

Erfolgreich „+“: das System hat bereits bei der ersten Ansage das optimale Ergebnis geliefert.

Mittelmäßig „0“: Die Ansage muss wiederholt werden und liefert anschließend das optimale Ergebnis.

Mangelhaft „-“: Das System reagiert nicht oder liefert das falsche Ergebnis.

Die Gesamtauswertung wurde tabellarisch festgehalten (Abbildung 6). Notizen, die aus den Kommentaren der Teilnehmer gemacht wurden (Thinking Aloud), werden erst im Kapitel Diskussion und Fazit herangezogen.

Auswertung Siri

	Szenario 1 Fenster zu/ Musik aus	Szenario 2 Fenster auf/ Musik aus	Szenario 3 Fenster zu/ Musik an	Szenario 4 Fenster auf/ Musik an
Aufgabe 1	+ | + |+	+|+| +	-|-|-	-|-|-
Aufgabe 2	+ |+| +	+|+| +	+|+| +	-|-|-
Aufgabe 3	+|+| +	+|+| +	-|-|-	-|-|-
Aufgabe 4	+|+| +	0|+|+	-|-|-	-|-|-
Aufgabe 5	+|+| +	+|+| +	-|-|-	-|-|-
Aufgabe 6	+|+| +	+|+| +	+|+| -	-|-|-

Auswertung Google Assistant

	Szenario 1 Fenster zu/ Musik aus	Szenario 2 Fenster auf/ Musik aus	Szenario 3 Fenster zu/ Musik an	Szenario 4 Fenster auf/ Musik an
Aufgabe 1	+| + |+	+|+| +	+|+|-	-|-|-
Aufgabe 2	+ |+| +	0|+| +	+|+| +	+|-|-
Aufgabe 3	+|+| +	+|+|0	+|+|-	-|-|-
Aufgabe 4	+|+| +	+|+|+	-|-|-	-|-|-
Aufgabe 5	+|+| +	+|+| +	-|+|-	-|-|-
Aufgabe 6	+|+| +	+|+| +	+|+|0	-|-|-

Abbildung 6:Auswertung Siri und Google Assistant

Die Auswertung von Siri und Google Assistant haben bei Szenario 1 eine vollständig positive Messung erreicht.

Das Szenario 2 von Siri zeigt, dass sobald das Fenster geöffnet wurde, Siri nicht direkt Ergebnisse liefert. Deshalb musste ein Proband die Aufgabe 4 wiederholen.

Szenario 2 von Google Assistant hingegen zeigt, dass beim Navigieren und bei der Informationssuche im Internet, der Test von zwei Probanden wiederholt werden musste. Bei Aufgabe 4, beim Schreiben einer Nachricht, wurde Diese beim ersten Mal richtig erfasst, aber nicht vorgelesen. Stattdessen hat Google Assistant gefragt, ob die Nachricht verschickt werden soll. Dies erfordert eventuell eine manuelle Eingabe der Nachricht vom Nutzer, welcher dadurch abgelenkt wird.

Das Szenario 3 von Siri veranschaulicht insgesamt ein durchschnittlich mangelhaftes Ergebnis. Aufgrund der Musik mit ca. 80 dB, wurden die Ansagen von Siri meist nicht erkannt. Beim Navigieren und beim notieren des Arzttermins hat Siri Befehle richtig erfasst und bearbeitet. Ab 90 dB wurde von Siri keinerlei Ansagen wahrgenommen.

In Szenario 3 sind im Gegensatz zu Siri von Google Assistant, wenigere negative Messungen aufgenommen worden. Dies liegt daran, dass Google auch bei 90 dB Befehle wahrnehmen konnte. Einmal musste die Aufgabe 6, der Termineintrag, wiederholt werden.

Bei Siri in Szenario 4 wurde keine der Aufgaben richtig bewältigt. Die Lautstärke beim offenen Fenster und Musik lagen bei 95 dB, wobei die Lautstärke der Musik bei ca. 80 dB lag. Mit Ausnahme von einem Probanden bei Aufgabe 2, scheiterte auch der Google Assistant.

Abbildung 7 veranschaulicht eine visuelle Darstellung der Erbenisse.

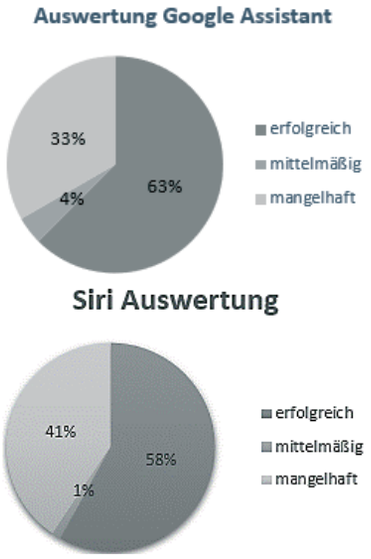


Abbildung 7: Visuelle Darstellung der Auswertungen Google Assistant und Siri

Von insgesamt 72 Bewertungen wurden 42 positive, ein mittelmäßige und 29 mangelhafte festgestellt. Prozentual gesehen sind 58,3% der Aufgaben von Siri erfolgreich umgesetzt wurden, 40,84% der Befehle mangelhaft und 1,38% mittelmäßig.

Bei Google Assistant hingegen wurden von insgesamt 72 Aufgaben, 45 positiv, 3 mittelmäßig und 24 mangelhafte Bewertungen zu erkennen. Prozentual gesehen sind 62,5% der Aufgaben erfolgreich umgesetzt, 33,3% mangelhaft und 4,2% mittelmäßig. Die beiden Diagramme (Abbildung 8) veranschaulichen die Ergebnisse visuell.

6 Ergebnis und Diskussion

Sowohl Siri als auch Google Assistant haben ihre Stärken und Schwächen. Aus dem Grund werden in diesem Kapitel die Notizen während des Tests und die Kommentare der Probanden miteinbezogen.

Wie schon bereits im Kapitel ‚Auswertung' zu lesen ist, zeigt sich Google Assistant im Bereich der Verständlichkeit mit Einfluss der realen Umgebung zuverlässiger als Siri. Dabei liegt die Grenze der Lautstärkenpegel bei Google Assistant bei ca. 95 dB und bei Siri nur bei ca. 80 dB. Siri hingegen hat den Vorteil, dass die Nachrichten nach dem Verfassen vorgelesen werden, dadurch ist vom Nutzer kein Korrekturlesen notwendig. Das Abbrechen kann aber auch durch Sprachsteuerung erreicht werden.

Beide Systeme haben allerdings den Nachteil, dass sie von einer fremden Person angesprochen werden können. Mit Ausnahme des anderen Geschlechts, denn weibliche Nutzer mit hoher Stimme werden von männlichen Nutzern mit tiefer Stimme unterschieden und erkannt.

Wenn ein Kalendereintrag erfasst wird, bietet Google Assistant bei der ersten Eingabe eine Auswahl an Anwendungen mit welchem der Eintrag erstellt werden kann. Sobald sich der Nutzer entscheidet, merkt sich das System die Auswahl der Anwendung und wendet dies ohne erneutes anfragen beim nächsten Kalendereintrag an. Siri hingegen bietet keine weiteren Möglichkeiten und erstellt direkt den Kalendereintrag.

Je nach Fragestellung bzw. Formulierung der Sätze, unterscheidet sich die Bearbeitungsdauer der Anfragen. Beginnt beispielsweise die Satzformulierung *mit „Hey Siri, notiere mir einen Arzttermin für morgen auf 15 Uhr"*, sucht Siri nach einem Kalendereintrag. Beginnt man allerdings mit *„Hey Siri, erstelle mir… "* wird direkt ein Kalendereintrag erfasst. Google Assistant, erfordert bei der Erstellung des ersten Kalendereintrags eine manuelle Bestätigung, während Siri es direkt aufnimmt.

Durch Aufgabe 2 hat sich herausgestellt, dass Siri die Navigation direkt startet, sobald man die Frage *„Wie komme ich zum Stuttgarter Hauptbahnhof?"* stellt. Google Assistant hingegen gibt eine Auskunft darüber und beschreibt die Strecke. Erst mit einer zweiten Ansage wie *„Starte die Navigation",* wird die Navigation fortgesetzt.

Beim Verfassen der Nachrichten werden ebenfalls Möglichkeiten vorgestellt. Die beliebte Musik-Applikation Spotify kann von beiden Systemen aufgerufen werden. Die Wiedergabe einer Musik allerdings ist nur bei Google Assistant möglich. Siri liefert stattdessen die Antwort *„Es tut mir leid, ich kann nichts von Spotify wiedergeben.".*

7 Fazit

Mit dieser Arbeit wurde untersucht, inwieweit die Sprachassistenzsysteme Siri und Google Assistant als Fahrassistenzsystem für den Nutzer zufriedenstellend sind.

Darunter werden Schwächen und Stärken der Systeme verglichen. Dafür wurden explizite Aufgabestellungen und Szenarien aufgestellt und mit dem Usability-Test angewendet. Aus den Resultaten haben sich zahlreiche Stärken und Schwächen beider Systeme ergeben.

In unterschiedlichen Bereichen und Szenarien hat sich herausgestellt, dass die Systeme ähnliche aber dennoch verschiedene Bearbeitungsstile haben. Das Sprachassistenzsystem Siri liefert unkomplizierte und schnellere Ergebnisse als Google Assistant. Dafür reagiert Google Assistant bezüglich der realen Einflüsse robuster und bietet mehr Funktionen und Informationen an. Siri führt manche Befehle direkt aus, während Google Assistant Unterstützung durch zusätzlicher Applikationen benötigt.

Im Rahmen dieser Arbeit wurde festgestellt, inwieweit und in welchen Situationen die Sprachassistenzsysteme den Nutzer von der eigentlichen primären Fahraufgabe ablenken können. Die Ablenkung ereignet sich dadurch aber meist bei Anwendung in zu lauter Umgebung oder durch undeutliche Formulierung der Befehle. Hierbei wird durch erforderliches wiederholen der Frage die Konzentration beeinträchtig. Ebenso die Notwenigkeit Befehle präzisen Formulieren zu müssen kann Ablenkung hervorrufen.

Zusammenfassend lässt sich sagen, dass beide Sprachassistenzsysteme im Einsatz als Fahrassistenzsysteme gebräuchlich, aber dennoch ausbaufähig sind.

8 Literaturverzeichnis

[1] Digital Trends Umfrage zu digitalen Sprachassistenten, 2017, https://www.bvdw.org/fileadmin/user_upload/BVDW_Digital_Trends_Sprachassistenten.pdf (20.09.2018).

[2] N. Rauch; Ein verhaltensbasiertes Messmodel zur Erfassung von Situationsbewusstsein im Fahrkontext; <http://www2.psychologie.uni-wuerzburg.de/izvw/texte/2009_Rauch_Diss.pdf> (01.09.18)

[3] Statista GmbH; statista; Das Statistik-Portal;<https://de.statista.com/statistik/daten/studie/451763/umfrage/groesste-ablenkungen-beim-autofahren-in-deutschland/> (01.09.18)

[4] JP. Haton; Automatic Speech Recognition: Past, Present, and Future; 2004;<https://www.researchgate.net/publication/228687527_Automatic_Speech_Recognition_Past_Present_and_Future>(01.09.18)

[5] A. Eslava; Evolutionäre Algorithmen in der Spracherkennung; GRIN Verlag; 2004; ISBN: 978-3-638-438-919

[6] G. Gobinda Chowdhury; Natural Language Processing; <https://local.cis.strath.ac.uk/research/publications/papers/strath_cis_publication_320.pdf>

[7] Projektbericht zum Studienprojekt „Ok Google... The End of Search as we know it. Sprachgesteuerte Suche im Test"; <http://searchstudies.org/wp-content/uploads/2016/04/OK_Google_2016_Abschlussbericht.pdf> (10.09.18)

[8] Google LLC; USA; <https://assistant.google.com/intl/de_de/platforms/cars/> (15.09.18)

[9] usability-toolkit.de; usability für Web-Projekte; Methoden; Usability-Test http://usability-toolkit.de/usabilty-methoden/usability-test/ (15.09.18)

[10] F. Sarodnick, H. Brau: Methoden der Usability Evaluation – Wissenschaftliche Grundlagen und praktische Anwendung; 2. Auflage; Verlag Hans Hube; 201; ISBN 978-3-456-84883-9

[11] Statista GmbH; statista; Das Statistik-Portal;<https://de.statista.com/statistik/daten/studie/620328/umfrage/anwendungsbereiche-digitaler-Sprachassistenten-in-deutschland/> (10.09.18)

[12] Volvo Car Switzerland; Der Sprachassistent als Beifahrer; 2018; <https://www.volvocars-partner.ch/blog/2018/04/09/der-sprachassistent-als-beifahrer/> (27.10.18)

[13] Ablenkungswirkung tertiäre Aufgaben im PKW systemergonomische Analyse und Prognose; 2004; https://mediatum.ub.tum.de/doc/601919/601919.pdf

[14] News Room;<https://newsroom.aaa.com/2015/10/new-hands-free-technologies-pose-hidden-dangers-for-drivers/> (27.10.18)

[15] Apple; <https://www.apple.com/de/siri/> (27.10.18)

[16] Google Assistant; Holfe für Google Assistant<https://support.google.com/assistant/answer/7394306?hl=de> (27.10.18)

[17] Statista; Das Statistik Portal; <https://de.statista.com/statistik/daten/studie/723295/umfrage/die-beliebtesten-taetigkeiten-bei-der-mobilen-musiknutzung-in-deutschland/> (27.10.18)

[18] DW; Made for minds; <https://www.dw.com/de/totgesagte-leben-länger-die-sms-im-whatsapp-zeitalter/a-17451917> (27.10.18)

[19] Bayerisches Landesamt für Umwel; UmweltWissen – Lärm Lärm – Hören, messen und bewerte<https://www.lfu.bayern.de/buerger/doc/uw_34_laerm_messen_bewerten.pdf> (27.10.18)

Methoden der Anomaliedetektion mit Fokus auf Logfileanalyse

René Blänsdorf
Hochschule Reutlingen
Rene.Blaensdorf@Student.Reutlingen-University.DE

Abstract

Logfiles sind eine interessante Quelle für Informationen über den Zustand von Softwaresystemen. Bisher scheinen sie fast ausschließlich manuell ausgewertet zu werden. Deswegen wurden sechs verschiedene maschinelle Lernverfahren betrachtet, die für eine solche Analyse eingesetzt werden könnten. Dabei haben sich neuronale Netze, Support Vector Machines, Bayes-Klassifikatoren, statistische Methoden und regelbasierte Klassifikatoren als potenziell geeignet erwiesen. Mit entsprechenden Schritten in der Vorverarbeitung wären die ausgeschlossenen Methoden, zumindest teilweise, ebenfalls verwendbar.

Schlüsselwörter

Anomaly Detection, Machine Learning, Log File Analysis

CCS Concepts

• Computing methodologies → Anomaly detection • Computing methodologies → Machine learning approaches • Computing methodologies → Semi-supervised learning settings

1 Einleitung

Die Datenmengen, die tagtäglich von Soft- und Hardwaresystemen verarbeitet werden müssen, steigen seit Jahren immer weiter an. Damit das funktioniert müssen verschiedene Systeme ineinandergreifen. Kommt es in dieser Kette an irgendeiner Stelle zu einem Problem, kann dadurch das ganze System zum Erliegen gebracht werden.

Das Finden und beheben dieser Fehler ist oft zeit- und kostenintensiv, wodurch der Wunsch entsteht sie entweder zu verhindern oder schnellstmöglich beheben zu können. Zhivich und Cunningham versuchten 2009 in einem Paper die hohen Kosten durch Softwarefehler zu benennen [15]. Dabei bezogen sie sich auch auf Werte aus einer alten Umfrage des National Institute of Standards and Technology (NIST). Leider hat einer Recherche zu der Studie ergeben, dass die Zahlen nicht zuverlässig genug sind, um sie als Maßstab verwenden zu können. In dem Paper zeigten sie jedoch auch, dass manche Softwarefehler nicht nur Geld, sondern auch Leben kosten können [15].

Die genauen Kosten für einen Softwarefehler lassen sich allgemein nur enorm schwer bestimmen, da sie sehr stark von der Art des Fehlers und von der

Betreuer Hochschule: Prof. Dr. Christian Decker
Hochschule Reutlingen
Christian.Decker@Reutlingen-University.de
Betreuer Firma: Dr. Kathrin Björkelund
Daimler AG
Kathrin.Bjoerkelund@daimler.com

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 René Blänsdorf

Software, in der sie auftreten abhängen. So kann ein kritischer Fehler im Finanzbereich vermutlich leichter kosten im Millionenbereich verursachen als ein rein grafischer Fehler in der Bedienoberfläche eines Mediaplayers. Da jedoch schwer absehbar ist wie kritisch ein Fehler sein kann, ist es wünschenswert Fehler schnellstmöglich zu erkennen.

Eine mögliche Quelle für entsprechende Analysen sind Logfiles, die von fast allen Softwaresystemen erstellt werden und oft in verschiedenen Detailstufen zur Verfügung stehen. Großer Vorteil der Logfiles ist es, dass sie Informationen liefern ohne, dass die Software oder beispielsweise direkt der Code analysiert werden müssen. In vielen proprietären Softwareprodukten besteht diese Möglichkeit ohnehin nicht, wodurch Logfiles teilweise die einzigen Anhaltspunkte für den Fehler liefern.

Dazu kommt, dass sie hauptsächlich Informationen enthalten, die in den Augen der Entwickler relevant sein könnten. Dadurch wird die Wahrscheinlichkeit erhöht, dass die Informationen, im Falle eines Fehlers, dabei helfen den Fehler zu identifizieren.

1.1 Problemstellung

Die Analyse dieser Logfiles gestaltet sich bei größeren Systemen mit dutzenden Logfiles und teilweise mehreren Megabyte oder gar Gigabyte an reinen Textdaten sehr schwierig.

Besonders die oft noch gängige, händische Analyse der Dateien dauert bei diesen Datenmengen entsprechend lang [3, 5]. Spätestens wenn die Analyse in Echtzeit erfolgen soll, um auf eventuelle Fehler sofort reagieren zu können, ist dies keine Option mehr und ein automatisches System wird benötigt.

Diese Systeme haben theoretisch jedoch nicht nur das Potenzial zu erkennen, wenn kritische Fehler auftreten. Sie sind mit den Informationen aus den Logfiles vermutlich auch in der Lage Vorhersagen über zukünftig auftretende Probleme zu machen.

Auf diesem Gebiet gibt es bereits verschiedene Forschungen, die sich jedoch selten speziell auf Logfiles beziehen [3]. In den meisten Fällen werden maximal Systemlogs betrachtet, beispielsweise um unberechtigte Zugriffe auf das System erkennen oder verhindern zu können. Softwareübergreifende Analysen der dazugehörigen Logfiles scheinen dabei fast gar nicht vorhanden zu sein.

1.2 Zielsetzung

Aus den beschriebenen Problemen ergibt sich, dass eine softwarebasierte Lösung gesucht wird, die in der Lage ist automatisch kritische Fehler in Softwaresystemen zu erkennen. Dabei sollten idealerweise keine weiteren Informationen als die Daten aus den Logfiles benötigt werden.

Zusätzlich sollte die Möglichkeit bestehen die Lösung dafür zu verwenden, oder sie entsprechend erweitern zu können, dass sie auch Fehlervorhersagen treffen kann. Dafür soll ein Vergleich zwischen aktuellen Methoden aus der Anomaliedetektion und ähnlichen Verfahren vorgenommen werden. Idealerweise sollte dabei ein Ansatz ermittelt werden, der sich besonders gut für Logfiles eignet. Maximal sollten jedoch drei Verfahren herauskristallisiert werden, die zur Lösung des Problems besonders gut geeignet sein könnten.

In einer nachfolgenden Thesis des Autors werden die ermittelten Ansätze weiter eingeschränkt und in einer realen Softwarelösung umgesetzt.

2 Stand der Technik

In diesem Kapitel werden aktuelle Methoden der Anomaliedetektion vorgestellt und ein kurzer Einblick in sie gegeben. Dabei wird nur auf maschinelle Lernverfahren eingegangen um Muster in den Daten automatisiert erkennen zu können. Deskriptive Analysen über Grafiken oder ähnliche Methoden werden nicht näher betrachtet, da sie nicht geeignet sind vollautomatisch und nur mithilfe der Logfiles arbeiten zu können.

2.1 Anomalien

Allgemein kann eine Anomalie als eine Abweichung von dem gewohnten oder gewünschten Verhalten definiert werden [5]. Es gibt verschiedene Arten von Anomalien basierend auf dem Zusammenhang indem sie auftreten. Beispielsweise Punktanomalien, Kontextanomalien oder Kollektivanomalien [2].

Zu den Punktanomalien gehören unter anderem einfache Zahlenmengen, bei denen man von einer Anomalie sprechen kann, wenn ein Wert stark von den anderen Werten abweicht [2].

Kontextanomalien sind Anomalien, die nur im korrekten Kontext ersichtlich sind. Dazu zählen beispielsweise Wetterdaten mit zeitlichem Zusammenhang [2].

Kollektivanomalien bestehen aus zusammenhängenden Daten, die im Vergleich zum Rest des Datensatzes eine Anomalie darstellen [2]. Ein Beispiel hierfür wären zeitlich zusammenhängende Messdaten, die immer wieder Schwankungen unterliegen aber plötzlich über einen längeren Zeitraum ungewöhnliche Werte aufweisen [2]. Ein plötzlich stillstehendes EKG beispielsweise.

2.2 Klassifikationsbasierte Methoden

Bei den klassifikationsbasierten Methoden kann zwischen zwei Kategorien unterschieden werden. Modelle mit nur einer Klasse und solche mit mehreren Klassen [2, 14]. Bei der ersten Variante wird jeder Datenpunkt, der nicht zu der als normal definierten Klasse gehört als Anomalie betrachtet [2]. Dadurch ergeben sich effektiv zwei Klassen, die zur Lösung binärer Probleme geeignet sind [14]. Die zweite Variante unterscheidet zusätzlich zwischen verschiedenen normalen Klassen und, sofern vorhanden, zwischen anomalen Klassen [2, 14].

Um Daten klassifizieren zu können wird als erstes ein Model anhand gelabelter Testdaten erlernt [2, 14]. Mit diesem Model werden alle neuen Datenpunkte geprüft und den Klassen zugeordnet, zu denen sie mit der höchsten Wahrscheinlichkeit gehören [2, 14].

Ein wichtiger Aspekt dieser Modelle ist die allgemeine Anwendbarkeit des Modells, sodass auch Datenpunkte korrekt zugeordnet werden können, die bisher unbekannt waren [14]. Wenn nur Datenpunkte korrekt zugeordnet werden können, die bereits bekannt sind würde das die Anwendbarkeit stark einschränken. Eine Datenreduktion könnte dennoch erzielt werden, da nur mit dem Model getestet werden muss und nicht alle bisherigen Datensätze durchsucht werden müssten.

Zu den Klassifikatoren gehören unter anderem Ansätze basierend auf künstlichen neuronalen Netzen (eng.: Artificial neural network, ANN), Bayes-Klassifikatoren, Support Vector Machines (SVM), k-nächsten-Nachbarn (KNN) und regelbasiertem maschinellen Lernen (RBML) [2, 14]. Letzteres versucht aus den vorhanden Daten Wenn-Dann Regeln zu bilden [14].

Je nach verwendeter Methode kann die Trainingszeit vergleichsweise hoch sein [2]. Sofern das Modell nicht anhand der neuen Daten angepasst werden muss geschieht die Klassifikation dafür nach dem Lernen sehr schnell.

2.3 Distanzbasierte Methoden

Die distanzbasierten Methoden zählen zu den am häufigsten verwendeten Methoden der Anomaliedetektion [2, 12]. Sie lassen sich weiter in Methoden unterteilen, die auf dem Nächste-Nachbarn (Nearest Neighbour) Prinzip basieren und solche, die auf Clustering basieren [2, 12].

Distanz entspricht in diesem Zusammenhang der Ähnlichkeit zwischen zwei Datenpunkten. Für numerische Werte kann beispielsweise der euklidische Abstand oder die Mahalanobis-Distanz berechnet werden [2, 4, 12]. Für kategoriale oder multivariate Daten muss die Ähnlichkeit jedoch zuerst mit einer entsprechenden Funktion bestimmt werden, um danach den Abstand berechnen zu können [1, 2, 12].

2.3.1 Nearest Neighbour

Beim KNN-Algorithmus wird die Nachbarschaft eines neuen Datenpunktes betrachtet und anhand dessen eine Zuordnung vorgenommen [14]. Das k ist dabei die Menge an Nachbarn, die zur Bewertung herangezogen werden [14]. Der KNN-Algorithmus gehört damit auch zu den Klassifikatoren, die in Kapitel 2.2 erwähnt wurden.

Um den Algorithmus an verschiedene Anforderungen anpassen zu können sind verschiedene Abwandlungen entstanden. Es gibt unter anderem Methoden, die nur die k nächsten Nachbarn betrachten, die nicht weiter als eine Distanz x entfernt sind [2, 12]. Andere Methoden betrachten die Dichte der Nachbarschaften [2, 12]. Eine geringe Dichte bedeutet in der Regel eine Anomalie [2, 12]. Unterschiedliche Methoden konzentrieren sich dabei auf lokale und globale Dichte, um die Ergebnisse weiter zu verbessern [2, 12].

2.3.2 Clustering

Um Cluster zu bilden werden in der Regel keine gelabelten Daten verwendet, was die Anwendung von unüberwachten Lernmethoden ermöglicht [2].

Eine viel verwendete Methode ist beispielsweise der k-Means-Algorithmus [12]. Dabei legt das k fest wie viele Cluster in den gegebenen Daten vorhanden sein sollen [2, 12]. Im nächsten Schritt werden die Distanzen zum Zentrum des Clusters berechnet und damit der Anomaliewert berechnet [2, 12]. Je größer die Distanz zum Zentrum, desto wahrscheinlicher handelt es sich um eine Anomalie [2].

Ein weiterer Ansatz stützt sich dabei auf die Annahme, dass jeder Datenpunkt, der keinem Cluster zugeordnet werden kann, eine Anomalie ist [2]. Dafür können Clustering Algorithmen wie beispielsweise DBSCAN verwendet werden [6].

Die Annahme, dass sich Anomalien in kleinen oder eher spärlichen Clustern befinden liefert ebenfalls einen Ansatz für verschiedene Methoden, beispielsweise von He et al. [9] und Santos-Pereira und Pires [13].

2.4 Statistische / Wahrscheinlichkeitsbasierte Methoden

Statistische beziehungsweise wahrscheinlichkeitsbasierte Methoden der Anomaliedetektion verwenden Wahrscheinlichkeitsdichtefunktionen (WDF) um mit ihnen Anomalien identifizieren zu können [12]. Sie gehen dabei davon aus, dass Anomalien in den Bereichen von statischen Modellen auftreten, die eine geringe Wahrscheinlichkeit haben [2].

Die Ansätze können dabei in zwei verschiedene Gruppen unterteilt werden. In parametrische und nicht-parametrische Methoden [2].

Parametrische Methoden setzen dabei voraus, dass bekannt ist welche Wahrscheinlichkeitsverteilung für die Daten angenommen werden kann [2]. Viele der Methoden gehen dabei beispielsweise von einer Normalverteilung aus, wie unter anderem der Grubbs Test [8] für Ausreißer [2, 12]. Aber auch Techniken basierend auf Regressionsanalyse und Mischverteilungen werden verwendet [2]. Die Parameter werden dabei aus den vorhandenen Trainingsdaten abgeschätzt [12].

Nicht-parametrische Ansätze treffen hingegen weniger Annahmen über die zugrundeliegenden Daten und generieren die Struktur direkt aus den Daten [2]. Das Model kann daher dynamisch an die Daten angepasst werden [12]. Die einfachste Technik dafür sind Modelle, die auf Histogrammen basieren [2, 12]. Die Kerndichteschätzung (auch Parzen-Fenster-Methode) und negative Selektion finden ebenfalls Anwendung [12]. Letztere basiert dabei auf der Vorgehensweise des menschlichen Immunsystems Antigene zu erkennen [12].

2.5 Rekonstruktionsbasierte Methoden

Diese Methoden sind in der Lage, mithilfe der Trainingsdaten, eigenständig ein Model der Daten abzubilden [12]. Der Rekonstruktionsfehler hängt dabei direkt mit dem Anomaliewert zusammen [12]. Er ergibt sich aus der Distanz zu dem Eingabevektor und dem Vektor, der ihn nach durchlaufen des Models repräsentiert [12].

2.5.1 Neuronale Netze

Neuronale Netze können durch ihre verschiedenen Arten sehr vielfältig eingesetzt werden. Dadurch sind im Laufe der Zeit verschiedene Ansätze entstanden, um sie rekonstruktionsbasiert als Methode zur Neuheitsdetektion oder zur Anomaliedetektion einzusetzen.

Dazu gehören unter anderem Ansätze, die auf mehrlagigen Perzeptrons (MLP), Replicator Neural Networks (RNN), Radial Basis Function Networks (RBF), Hopfield Networks und Growing Neural Gas Networks (GNG) basieren [12]. Es gibt noch einige weitere Ansätze, die bereits untersucht wurden, die Auflistung sollte jedoch eine gute Übersicht jüngerer Forschung widerspiegeln [12].

2.5.2 Unterraum / Spectral

Rekonstruktionsbasierte Methoden, die versuchen eine Identifikation von Anomalien zu erzielen, indem sie versuchen die Trainingsdaten in einen Unterraum mit geringerer Dimensionalität zu projizieren, werden auch Spectral Methoden genannt [12]. Dabei wird davon ausgegangen, dass es möglich ist diese Projektion vorzunehmen und dass sich normale und anomale Daten signifikant voneinander unterscheiden [2, 12].

Viele Methoden in diesem Bereich verwenden die Hauptkomponentenanalyse (eng.: Principal Component Analysis, PCA) in verschiedenen Formen [2, 3, 12]. Bei der PCA wird versucht die unterschiedlichen Merkmale möglichst genau anhand von Linearkombinationen abzubilden. Eine Methode verwendet beispielsweise die kernel PCA, um nicht-lineare Datenverteilungen berücksichtigen zu können [12]. Dabei werden Punkte auf einen höher dimensionalen Merkmalsraum gemappt, um anschließend die PCA darauf auszuführen [12].

Diese Methoden können fast immer auch als Vorverarbeitungsschritt für die anderen Anomaliedetektionsverfahren verwendet werden [11]. Die Reduktion der Dimensionalität wirkt sich dabei in der

Regel vorteilhaft auf die Laufzeit der Verfahren aus.

2.6 Domainbasierte Methoden

Die domainbasierten Methoden versuchen möglichst akkurate Grenzen zwischen den einzelnen Datenpunkten zu ziehen, um sie eindeutig zuordnen zu können [12]. Die Dichte der Klassen spielt dabei keine Rolle, weil alles was sich innerhalb der Grenze befindet automatisch zu der Klasse dazugezählt wird [12].

Methoden dieser Kategorie basieren auf Support Vektor Maschinen (SVM) und ihren unterschiedlichen Abwandlungen [12]. SVMs benutzen eine Hyperebene um die Grenzen zwischen den Klassen über verschiedene Dimensionen hinweg definieren zu können [10, 12]. Dabei werden nur die Datenpunkte verwendet, die am nächsten an der Grenze liegen (Support Vektoren), wodurch der Rest der Datenpunkte nicht weiter betrachtet werden muss.

Für die Anomaliedetektion sind vor allem die One-Class SVM (OC-SVM) und die Support Vector Data Description (SVDD) von Interesse, da sie gut darin sind Ausreißer zu identifizieren [10, 12].

SVDDs arbeiten dabei nicht mit einer Hyperebene, sondern mit einer Hypersphere [10, 12]. Dabei liegt die Annahme zugrunde, dass die Region mit den meisten normalen Datenpunkten mit einer solchen Sphere dargestellt werden kann [10]. Das Problem ist dann nur noch das Zentrum und den Radius der Sphere zu bestimmen [10].

OC-SVMs verwenden eine Hyperebene und suchen die Hyperebene, die den größtmöglichen Abstand zwischen dem Ursprung und den Datenpunkten erlaubt [10, 12]. Unter den entsprechenden Voraussetzungen ist es zudem möglich, dass beide Methoden die Gleiche Lösung liefern [10].

2.7 Informationstheoretische Methoden

Methoden, die auf dem informationstheoretischen Ansatz basieren treffen die Annahme, dass Anomalien signifikante Änderungen am Informationsgehalt des ganzen Datensatzes vornehmen [2, 12]. Um den Informationsgehalten messen zu können werden beispielweise die Entropie und relative Entropie aber auch die Kolmogorow-Komplexität verwendet.

Um die Anomalien identifizieren zu können wird nach einer Teilmenge der Datenmenge gesucht, deren Entfernung die größte Änderung verursacht [2, 12]. Daraus ergibt sich, dass eine möglichst kleine Untermenge mit einer möglichst großen Auswirkung auf den Informationsgehalt des Datensatzes gesucht wird [2].

Um dieses Problem zu lösen gibt es verschiedene Ansätze, da alle Untermengen zu testen einen viel zu großen zeitlichen Aufwand bedeuten würde und praktisch nicht umsetzbar ist [2]. Daher werden beispielsweise heuristische Methoden eingesetzt um einen Wert zu finden, der der optimalen Lösung möglichst nahe kommt [2, 12]. Je nach Verfahren werden dabei die zuvor genannten Messwerte zur Bestimmung des Informationsgehalts des Datensatzes verwendet, um sich ihre speziellen Eigenschaften zunutze machen zu können.

3 Anwendungsgebiete

Zu den Anwendungsgebieten für Anomaliedetektion zählen unter anderem folgende:

- Unberechtigte Eingriffe in Systeme
- Betrugsdetektion
- Medizin und Gesundheitswesen
- Schäden an Industriegeräten
- Bildverarbeitung
- Textdaten
- Sensornetzwerke

Um nur die größeren Bereiche zu nennen. Es gibt jedoch noch viele weitere in denen Anomaliedetektion eingesetzt werden kann [1, 2, 12]. In diesen Bereichen wurden die Methoden jedoch noch nicht so umfangreich eingesetzt wie in den bereits genannten.

Neuheitsdetektion und Softwarefehlervorhersage nehmen eine besondere Stellung ein, da sie spezielle Bereiche der Anomaliedetektion abdecken. Jede Neuheit kann als Anomalie im Vergleich zum bisherigen Verhalten betrachtet werden und jeder Defekt stellt per Definition ebenfalls eine Anomalie dar. Einschränkung ist jedoch, dass bei der Neuheitsdetektion nur semi-überwachte und unüberwachte Lernmethoden zum Einsatz kommen können. Da die Neuheit per Definition noch nicht aufgetreten oder erkannt worden ist können für sie noch keine gelabelten Daten vorliegen.

4 Anforderungen

Aus der genannten Problemstellung und den Zielen aus Kapitel 1.1 und 1.2 ergeben sich folgende Anforderungen an die möglichen Lösungen:

- Das System muss in der Lage sein, Anomalien ohne menschliche Eingriffe erkennen zu können. Lediglich Trainingsdaten dürfen zur Verfügung gestellt werden müssen. Die Lösung sollte dabei maximal semi-überwachte Lernmethoden benötigen.
- Die Lösung muss in der Lage sein, die gegebenen Logfiles in beinahe Echtzeit analysieren zu können.
- Die gegebenen Logfiles müssen auch bei mehreren Megabyte Größe verarbeitet werden können ohne das Resultat relevant zu verschlechtern.
- Unterschiedliche Logfiles von unterschiedlichen Systemen müssen verarbeitet werden können.
- Das System muss Anomalien nur mithilfe der zur Verfügung stehenden Logfiles erkennen können.
- Es ist nicht notwendig, dass die Lösung in der Lage ist den Fehler beheben zu können oder ihn verhindern zu können. Wünschenswert wäre es jedoch, wenn die Lösung in der Lage wäre bei der Reaktion auf den Fehler zu unterstützen.
- Vorhersagen von zukünftigen Fehlern sind optional, die Möglichkeit das System entsprechend zu erweitern sollte jedoch gegeben sein.

Zusätzliche, nützliche Funktionen sind von Vorteil aber nicht notwendig, um die Anforderungen zu erfüllen.

5 Analyse & Evaluation

In diesem Kapitel werden die in Kapitel 2 vorgestellten Methoden in Relation zu den Anforderungen gestellt. Dabei wird genauer analysiert welche Anforderungen sie erfüllen könnten und anschließend eine Lösungsempfehlung gegeben.

5.1 Klassifikationsbasierte Methoden

Ansätze, die mit mehreren Klassen arbeiten sind auf vollständig gelabelte Daten für normales Verhalten und für anomales Verhalten angewiesen. Da

angenommen wird, dass keine gelabelten Daten für anomales Verhalten vorhanden sind scheiden diese Methoden aus.

Die Abwandlungen, die sich auf eine Klasse beschränken und keine gelabelten Anomalen Daten benötigen, sind potenziell jedoch für die Lösung geeignet. Neuronale Netze und SVMs werden in Kapitel 5.4 und 5.5 bereits näher betrachtet. Bayes-Klassifikatoren sind potenziell ebenfalls für die Lösung geeignet. Regelbasierte Methoden könnten auch geeignet sein, allerdings ist es vorstellbar, dass die Komplexität eines solchen Regelsystems sehr groß werden könnte.

5.2 Distanzbasierte Methoden

KNN basierte Methoden können nicht effektiv für hochdimensionale Daten verwendet werden [2, 12]. Aufgrund der Vorgehensweise dieser Methoden ist auch eine große Datenmenge problematisch, wenn ein Resultat in Echtzeit erreicht werden sollen, wodurch diese Methode nicht geeignet ist [2, 12].

KNN und Clustering haben zudem das Problem, dass eine geeignete Methode zur Berechnung der Distanz gewählt werden muss [2, 12]. Da in Logfiles für gewöhnlich sehr viel recht unterschiedlicher Text steht, stellt das ein großes Problem dar. Clustering hat zudem ebenfalls Probleme mit hochdimensionalen Daten [12]. Dadurch sind diese beiden Ansätze als Lösung wenig geeignet.

Dani et al. konnten jedoch zeigen, dass mit entsprechenden Schritten in der Vorerarbeitung, auch Clustering mit k-means erfolgreich verwendet werden kann [5]. Dabei haben sie die Dimensionalität reduziert, um die Nachteile bei hoher Dimensionalität auszugleichen [5].

5.3 Statistische / Wahrscheinlichkeitsbasierte Methoden

Dadurch, dass ein Modell kreiert wird und die Trainingsdaten danach nicht mehr benötigt werden, eignet sich die Methode vergleichsweise gut für die Echtzeitanwendung [12]. Problematisch ist allerdings, dass die Daten bestimmten Wahrscheinlichkeitsverteilungen unterliegen müssen, was oft nicht der Fall ist [2]. Bei hochdimensionalen Daten ist es noch unwahrscheinlicher und schwieriger eine passende Statistik zu wählen [2]. Liegen zudem zu wenige Trainingsdaten vor kann es ebenfalls zu Schwierigkeiten kommen [12]. Die Komplexität ist jedoch vergleichsweise gering und Skalierbarkeit ist gegeben.

5.4 Rekonstruktionsbasierte Methoden

Rekonstruktionsbasierte Methoden haben den Vorteil, dass sie kein Wissen über die zugrundeliegenden Daten voraussetzen [12]. Neuronale Netze sind jedoch sehr anfällig bei der Optimierung ihrer Parameter, die die Struktur des Modells definieren [12]. Wenn sie nicht gut genug sind können auch nur schlechte Ergebnisse erreicht werden [12]. Bei hochdimensionalen Daten ergibt sich dadurch auch hier wieder ein Problem.

Unterraummethoden haben den Nachteil, dass die wichtigsten Attribute identifiziert werden müssen, um ein entsprechendes Mapping vornehmen zu können [12]. Zudem ist die Komplexität der Berechnung der Korrelationsmatrix für das normale Verhalten sehr hoch [12].

5.5 Domainbasierte Methoden

SVMs haben den Vorteil, dass sie nur auf die Daten angewiesen sind, die am nächsten an der Grenze liegen [12]. Die Komplexität im Zusammenhang mit der Kernelfunktion kann bei diesen Methoden jedoch problematisch sein [12]. Zudem ist die Wahl der richtigen Kernelfunktion ebenfalls kritisch und nicht einfach [12]. Die Wahl der richtigen Parameter für die Größe der Grenzregion kann sich ebenfalls als problematisch erweisen.

Es muss ebenfalls bedacht werden, dass SVMs dazu tendieren schlechtere Ergebnisse zu liefern, wenn das Verhältnis zwischen normalen und anomalen Daten zu stark voneinander abweicht [7]. Was wahrscheinlich ein weiterer Grund für den Fokus auf OC-SVMs ist. Trotz allem sind sie potenziell zur Lösung des Problems geeignet.

5.6 Informationstheoretische Methoden

Einer der größten Nachteile des informationstheoretischen Ansatzes ist es, dass Anomalien oft nur erkannt werden können, wenn sie in großer Anzahl auftreten [2, 12]. In den meisten Softwaresystemen stellen kritische Fehler jedoch eher eine Ausnahme dar, sodass nicht genug Anomalien vorhanden sind um die Methode einsetzen zu können. Aus diesem Grund scheidet die Methode in diesem Fall als Lösung aus.

5.7 Ergebnis

Anhand der Analysen dieses Kapitels ergeben sich folgende fünf Verfahren, die eine praktikable Lösung liefern könnten:

- Neuronale Netze
- Suport Vector Machines
- Bayes-Klassifikatoren
- Statistische Methoden
- Regelbasierte Klassifikatoren

Ziel war es die zur Auswahl stehenden Verfahren auf maximal 3 zu reduzieren. Dies ist jedoch anhand der gestellten Anforderungen nicht eindeutig möglich.

6 Fazit und Ausblick

Nach der Vorstellung der einzelnen Methoden und der zugehörigen Analyse konnte gezeigt werden, dass es eine umfangreiche Auswahl an Methoden zur Anomaliedetektion gibt. Zudem eignen sich einige davon scheinbar auch für die Analyse von Logfiles.

In einer auf dieser Arbeit aufbauenden Thesis werden die in Kapitel 5.7 aufgeführten Methoden anhand eines realen Anwendungsfalls näher untersucht. Mit den vorhandenen Logfiles sollte es möglich sein die geeigneten Lösungen zusätzlich einzuschränken, sodass am Ende die beste Lösung implementiert werden kann. Diese Arbeit konnte jedoch eine breite Anzahl an Methoden vorstellen. Dadurch ist es möglich sie als Grundlage für verschiedene Einsatzszenarien zu verwenden ohne durch ein zu konkretes Einsatzgebiete eingeschränkt zu werden.

Die recht große Anzahl an potenziell geeigneten Methoden für die Anomaliedetektion auch in Logfiles, lässt darauf schließen, dass zukünftig weitere Forschung in der Richtung betrieben wird. Dies wird zusätzlich dadurch begünstigt, dass die Bereitschaft maschinelle Lernverfahren im Unternehmen einzusetzen stetig zunimmt. Gleichzeitig werden die entwickelten Verfahren auch immer weiter verbessert.

Literatur

[1] V. Chandola, A. Banerjee, and V. Kumar. 2012. Anomaly Detection for Discrete Sequences: A Survey. *IEEE Trans. Knowl. Data Eng.* 24, 5, 823–839. DOI: https://doi.org/10.1109/TKDE.2010.235.

[2] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly Detection: A Survey. *ACM Comput. Surv.* 41, 3, 1–58. DOI: https://doi.org/10.1145/1541880.1541882.

[3] Xin Cheng and Ruizhi Wang. 2014. Communication Network Anomaly Detection Based on Log File Analysis. In *Rough sets and knowledge technology. 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings*, Duoqian Miao, Witold Pedrycz, Dominik Ślęzak, Georg Peters, Qinghua Hu and Ruizhi Wang, Eds. Lecture Notes in Computer Science, 8818. Springer, New York, 240–248. DOI: https://doi.org/10.1007/978-3-319-11740-9_23.

[4] Linh N. Chu. 2015. *Metric Learning for Software Defect Prediction*. Masterthesis. Monash University.

[5] Mohamed C. Dani, Henri Doreau, and Samantha Alt. 2017. K-means Application for Anomaly Detection and Log Classification in HPC. In *Advances in Artificial Intelligence: From Theory to Practice*, Salem Benferhat, Karim Tabia and Moonis Ali, Eds. Lecture Notes in Computer Science. Springer International Publishing, Cham, 201–210. DOI: https://doi.org/10.1007/978-3-319-60045-1_23.

[6] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. KDD'96. AAAI Press, 226–231.

[7] Song Fu, Jianguo Liu, and Husanbir Pannu. 2012. A Hybrid Anomaly Detection Framework in Cloud Computing Using One-Class and Two-Class Support Vector Machines. In *Advanced data mining and applications. 8th international conference, ADMA 2012, Nanjing, China, December 15-18, 2012 proceedings*, Shuigeng Zhou, Songmao Zhang and G. Karypis, Eds. Lecture notes in computer science. Lecture notes in artificial intelligence, 7713. Springer, Berlin, New York, 726–738. DOI: https://doi.org/10.1007/978-3-642-35527-1_60.

[8] Frank E. Grubbs. 1969. Procedures for Detecting Outlying Observations in Samples. *Technometrics* 11, 1, 1–21. DOI: https://doi.org/10.1080/00401706.1969.10490657.

[9] Zengyou He, Xiaofei Xu, and Shengchun Deng. 2003. Discovering cluster-based local outliers. *Pattern Recognition Letters* 24, 9-10, 1641–1650. DOI: https://doi.org/10.1016/S0167-8655(03)00003-5.

[10] Christoph H. Lampert. 2008. Kernel Methods in Computer Vision. *FNT in Computer Graphics and Vision* 4, 3, 193–285. DOI: https://doi.org/10.1561/0600000027.

[11] Teng Li, JianFeng Ma, and Cong Sun. 2018. Dlog: diagnosing router events with syslogs for anomaly detection. *J Supercomput* 74, 2, 845–867. DOI: https://doi.org/10.1007/s11227-017-2165-9.

[12] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko. 2014. A review of novelty detection. *Signal Processing* 99, 215–249. DOI: https://doi.org/10.1016/j.sigpro.2013.12.026.

[13] Carla M. Santos-Pereira and Ana M. Pires. 2002. Detection of Outliers in Multivariate Data: A Method Based on Clustering and Robust Estimators. In *Proceedings in computational statistics*, Wolfgang Härdle and Bernd Rönz, Eds. Physica-Verlag HD, Heidelberg, 291–296. DOI: https://doi.org/10.1007/978-3-642-57489-4_41.

[14] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. 2019. *Introduction to data mining* (Second edition). Pearson, NY NY.

[15] Michael Zhivich and Robert K. Cunningham. 2009. The Real Cost of Software Errors. *IEEE Secur. Privacy Mag.* 7, 2, 87–90. DOI: https://doi.org/10.1109/MSP.2009.56.

Kollaborative VR-Plattformen für Virtual-Learning-Environments

Tobias Schultze
Hochschule Reutlingen
tobias.schultze@student.reutlingen-university.de

Abstract

Die steigende Popularität von VR ist zunehmend auch in der Industrie spürbar. Längst haben Unternehmen das Potential dieser Technologien erkannt. Neben der Integration in Konstruktionsprozessen durch virtuelle Reviews bieten virtuelle Trainings eine kostengünstige und effiziente Möglichkeit der Kompetenzentwicklung. Durch kollaboratives Arbeiten in virtuellen Umgebungen kann die Wertschöpfung dieser Lernumgebungen gesteigert werden. Einige Softwareunternehmen haben sich auf die Entwicklung von VR Plattformen für die Bedürfnisse dieser Unternehmen fokussiert und bieten spezialisierte Lösungen für den professionellen Einsatz von VR.
Im Rahmen dieser Arbeit werden Anforderungen an kollaborative VR-Plattformen für den Einsatz als virtuelle Lernumgebung definiert und verfügbare Lösungen und Produkte untersucht.

CCS Concepts

• **Human-centered computing** → **Collaborative and social computing systems and tools**; *Visualization systems and tools*; • **Applied computing** → **Interactive learning environments**; **Collaborative learning**; *Computer-aided design*; Business process management systems; E-learning;

Keywords

Virtual-Learning-Environments, Virtual-Reality, Collaborative-Learning, E-Learning

1 Einleitung

Digitale Werkzeuge sind heute im Produktentstehungsprozess kaum mehr wegzudenken. Dank digitaler Technologien wie Computer-Aided-Design (CAD), Computer-Aided-Manufacturing (CAM) oder Computer-Aided-Engineering (CAE) können heute immer komplexere Prozesse, Maschinen und Produkte in immer kürzerer Zeit entstehen. Diese werden oftmals durch Konzepte wie Product-Data-Management (PDM) und Product-Lifecycle-Management (PLM) verwaltet. Der Einsatz dieser digitalen Technologien ermöglicht es zudem, unabhängig des Standortes gemeinsam an Projekten zu arbeiten und die dazugehörigen Daten nahezu in Echtzeit auszutauschen. Unternehmen, Ingenieure und Kunden können so bereits heute von mehr Flexibilität in der Produktentstehung, der Prozessplanung und dem Produktmanagement profitieren.
Der Einsatz digitaler Werkzeuge ist jedoch auch über den Produktentstehungsprozess hinaus attraktiv. Auch im

Betreuer Hochschule:	Prof. Dr. rer. nat. Uwe Kloos Hochschule Reutlingen Uwe.Kloos@Reutlingen-University.de
Betreuer Firma:	Dr. Ing. Matthias Bues Fraunhofer Institut für Arbeitswirtschaft und Organisation Matthias.Bues@iao.fraunhofer.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Tobias Schultze

Vertrieb, dem Service oder der Aus- und Weiterbildung sind diese Technologien attraktiv. So bieten Virtual-Learning-Environments (VLE) beispielsweise das Potential Lernerfolge sowie die Motivation zu steigern (vgl. [1, 2]), Lerneinheiten flexibler zu gestalten oder individuell auf die Bedürfnisse einzelner Anwender anzupassen. Mit Hilfe von Cave-Automated-Virtual-Environments (CAVEs) oder Head-Mounted-Displays (HMD) können immersive Umgebungen genutzt werden, um das Räumlichkeits- und Situationsempfinden bereits ohne die Notwendigkeit physischer Präsens des Lehrgegenstandes zu schaffen.
Im Rahmen dieser Arbeit werden Anforderungen für den produktiven Einsatz einer VR-Plattform für Schulungszwecke definiert. Daraufhin werden verschiedene verfügbaren kollaborativen VR-Plattformen auf eben diese Anforderungen untersucht.

2 Virtual Environments

Als Ivan Sutherland 1965 in seinem Artikel *The Ultimate Display* virtuelle Dinge beschrieb, welche sich exakt wie ihre realen Gegenstücke verhalten, betrachtete er eine virtuelle Umgebung weniger aus einer technischen Sicht. (vgl. [3]) Vielmehr beschreiben diese Virtual Environments (VE) eine computergenerierte, synthetische Umgebung, in derer sich Objekte auf gleiche Weise wie ihre nicht synthetischen Pendants verhalten. Dieses Gedankenmodell legte den Grundstein für das heutige Verständnis virtueller Umgebungen.

2.1 Virtual Learning Environments

Ähnlich wie VEs handelt es sich bei Virtual Learning Environments (VLE) um Computergenerierte Umgebungen, welche eine künstliche (virtuelle) Umgebung abbilden und mit denen der Nutzer interagieren kann. VLEs haben jedoch den Transfer von Informationen und Wissen zum Ziel. [4]
Vorangegangene Studien zeigten, dass mit Hilfe virtueller Lernumgebungen ein höherer Lernerfolg sowie eine höhere Motivation beim Lernen erzielt werden kann. [1, 5, 6] Diese Vorteile bieten besonders im Kontext beruflicher Aus- und Weiterbildung viel Potential. [2, 7]
Besonders immersive virtuelle Umgebungen erlauben durch maßstabsgetreue Abbildungen eine bessere kognitive Verknüpfung zwischen virtuellen Inhalten und realen Erfahrungen. Umgekehrt können in VE erlangte Erfahrungen leichter mit vorhandenem Wissen verknüpft werden. [1, 8]

2.2 Virtual Reality

Obgleich der Begriff *Virtual-Reality* (VR) oftmals für VEs verwendet wurde [1], wird dieser im Rahmen dieser Arbeit in einem technischen Kontext für Immersive VEs, wie sie beispielsweise durch Head-Mounted-Displays (HMDs) oder CAVEs entstehen, verstanden.
Neben digitalen Werkzeugen wie CAD oder PLM erlangt auch VR bei diesen Prozessen zunehmend an Bedeutung. Längst haben einige Unternehmen die Vorteile dieser Technologie erkannt und verwenden diese beispielsweise für digitale Prototypen, Reviews oder zur Simulation. (vgl. [9–11]) Durch die in den vergangenen Jahren zunehmende Weiterentwicklung und die damit gestiegene Zugänglichkeit dieser Technologie, eignen sich HMDs als kostengünstige Alternative zu großen Projektionsanlagen wie CAVES. HMDs bieten eine portablen und somit flexible Einsatzmöglichkeit von Virtual-Environments.

3 Anforderungen

Für den Einsatz einer kollaborativen VR-Plattform als Virtual-Learning-Environment gilt es besondere Anforderungen zu erfüllen. Im Folgenden werden diese Anforderungen definiert und erläutert.

Nutzbarkeit: Auf Seiten der Anwender ist der einfache und intuitive Umgang mit der Technologie von besonders großer Bedeutung. Anwendungen müssen demnach so gestaltet sein, dass Interaktionen mit möglichst geringem kognitivem Aufwand getätigt werden können. Auch Anwender mit weniger ausgeprägter technischer Affinität dürfen durch die Technologie und deren Bedienung nicht in der Durchführung der inhaltlichen Aufgabe behindert werden.
Die Vermittlung des Lerninhaltes muss demnach im Vordergrund stehen. Das bedeutet, dass der Lernaufwand zur Bedienung einer VR-Trainingsumgebung möglichst gering sein muss.

Datensicherheit: Aus Sicht der Unternehmen, welche Kollaborative VR-Plattformen einsetzen möchten, können wiederum besondere Anforderungen von Bedeutung sein. So spielt gerade in der Produktentwicklung die Sicherheit beim Austausch von Daten eine große Rolle. Dieser Austausch von möglicherweise sensiblen Konstruktionsdaten ist jedoch besonders bei einer verteilten Anwendung erforderlich. Die Art und Weise auf die diese Daten jedoch den Nutzern zur Verfügung gestellt werden, sollte einem hohen Sicherheitsstandart entsprechen. Dies gilt sowohl für den generellen Einsatz von kollaborativen VR-Plattformen als auch für den Einsatz für Trainingszwecke, da auch die Prozessinformationen unter Umständen als sensibel eingestuft werden können.

Anwendbarkeit: Für den produktiven Einsatz einer kollaborative VR-Plattform für Trainingszwecke in einem Unternehmen bedeutet dies, dass das nötige know-how sowie der Aufwand für die Inbetriebnahme, Bedienung und Erweiterung möglichst gering sein muss. Letzteres betrifft hierbei die Erstellung neuer und Verwaltung vorhandener Schulungsinhalte.
Für die Erstellung neuer Inhalte sind Autorenwerkzeuge erforderlich. So sollten auch ohne spezifischem Wissen über die technische Umsetzung des VLEs Inhalte erzeugt werden können. Dies hat zum Einen den Vorteil, dass besonders kleinere Unternehmen keine gesonderten Arbeitskräfte für die Erstellung von VR-Inhalten abstellen müssen, aber auch Fachkräfte wie Schulungspersonal das Fachspezifische Wissen auf direktem Weg in die Schulungsprozesse einbringen können. [7] Diese Flexibilität in der Prozesserstellung und Verwaltung ist nötig um die Kosten für den Einsatz einer solchen Technologie gering zu halten.
Zudem gilt sowohl auf Seiten der Anwender als auch der Unternehmen eine einfache Inbetriebnahme der Technologie als erforderlich. Sowohl der Aufwand für das Aufbauen und Einrichten der Hardware, als auch die Inbetriebnahme und Bedienung der Software müssen diese Anforderung erfüllen.

Datenaufbereitung: Da, wie eingangs erwähnt, heute viele Prozesse der Produktentwicklung und Verwaltung bereits digitalisiert sind, bietet dies oft eine gute Datenbasis zur Verwendung dieser verfügbaren Daten für das virtuelle Training.
Vorhandene CAD-, CAM- oder Prozessdatensätze sollten sich mit geringem Aufwand in das VLE transferieren lassen. Eine Kollaborative VR-Lernplattform sollte demnach geeignete Schnittstellen bieten, um diese Daten ohne oder mit geringem Aufwand nutzen zu können. Viele der

etablierten PLM-Systeme bieten bereits Schnittstellen zum Datenaustausch mit Drittsoftware.

4 Analyse

In diesem Kapitel werden drei verschiedene Arten verfügbarer VR-Plattformen genauer untersucht. Dabei werden die jeweiligen Besonderheiten beleuchtet und die Systeme auf die Erfüllung der gesetzten Anforderungen untersucht.
Hierzu wurden die im Folgenden vorgestellten Anwendungen und Plattformen in drei Kategorien unterteilt.

4.1 Komplettlösungen

IC.IDO Das im Jahr 2000 als Ableger des Forschungsinstituts *Fraunhofer IAO* gegründete Unternehmen *IC.IDO*, welches seit 2011 als gleichnamiges Produkt der ESI-Gruppe vertrieben wird, stellt eine modulare VR-Plattform dar. Die ursprünglich als Werkzeug zur Visualisierung von Konstruktionsdaten entwickelte Software wurde um eine Vielzahl an Funktionen und Anknüpfungen sonstiger Softwareprodukte der *ESI-Group* erweitert. Das auf die Entwicklung von Simulationssoftware spezialisierte Unternehmen bewirbt *IC.IDO* mit Anwendungsgebieten wie Virtual-Engineering, -Build, -Service und -Product-Presentation. Neben umfassenden Engineeringtools wie Physik-, Werkzeug- und Ergonimiesimulation, bietet die Software die Möglichkeit kollaborativer virtueller Interaktion wie beispielsweise ortsunabhängiger Produktpräsentationen, Reviews oder Montagetrainings.
Dank der langjährigen Geschichte als Visualisierungswerkzeug ist *IC.IDO* in der Lage selbst große Datenmengen komplexer Konstruktionen auf verschiedenen Plattformen performant zu visualisieren. Hierzu zählen 3D-Stereo-Displays wie CAVES oder Powerwalls aber auch VR und AR Technologien wie einige der verbreiteten HMDs.
Des Weiteren wirbt das Unternehme mit der Unterstützung von Multi-CAD und PDM Systemen zur schnellen Prozessintegration. Dies erlaubt das Laden von Daten in verbreiteten Austauschformaten.
Durch diese direkte Anbindung an etablierte CAD und PDM-Systeme können die wesentlichen Anforderrungen an die Datensicherheit sowie Datenaufbereitung erfüllt werden. Da die *ESI-Group* als erfahrenes und spezialisiertes Unternehmen für Simulationssoftware auf langjährige Erfahrung und eine breite Produktpalette zurückgreifen kann, kann die Anwendbarkeit durchaus hoch eingestuft werden. Die Nutzbarkeit für Anwender wirkt augenscheinlich jedoch etwas schwerfällig.

Vizible *WorldViz* vertreibt ebenfalls eine Komplettlösung für VR Visualisierung namens *Viszible*. Besonderen Fokus legt das amerikanische Unternehmen dabei auf kollaborative Produktpräsentation in virtuellen Umgebungen. Mittels einer Desktopanwendung lassen sich einfach Präsentationen erstellen. *WorldViz* bewirbt; das Tool sei zu bedienen wie Powerpoint. Bei den Folien handle es sich jedoch um dreidimensionale Räume. Zur Erstellung dieser interaktiven 3D Präsentationen bietet das zugehörige Autorenwerkzeug eine Auswahl vorgefertigter Interaktionskomponenten, welche durch Drag and Drop auf die im Raum befindlichen Objekte angewandt werden können.
Die Software erlaubt den Import von 3D Datensätzen in verschiedenen Dateiformaten, Videos, Bild- und Audiodateien sowie PDFs. Ein direkter Import von Konstruktionsdaten (beispielsweise im .jt Format ist jedoch nicht möglich. Für die Aufbereitung der Konstruktionsdaten sind vorgelagerte Prozesse notwendig. *Vizible* dient hauptsächlich der Visualisierung und Präsentation in kollaborativen VEs. Zwar bietet diese Lösung ein leicht zu bedienendes Werkzeug zur Erstellung digitaler Inhalte, die Integration von Prozessinformationen für Trainingszwecke ist jedoch nicht möglich. Somit ist die Anforderung der Anwendbarkeit nicht erfüllt. Die Bedienung der VR-Anwendung geht dafür leicht von der Hand.

TechViz VR Einen etwas anderen Ansatz verfolgt *TechVis VR*. Das plug-in des 2004 in Frankreich gegründeten Unternehmens unterstützt eine Vielzahl verschiedener Programme und ermöglicht die Visualisierung in einer immersiven VR Umgebung. Hierzu greift *TechViz* die Grafikinformationen aus dem *OpenGL* Datenstrom ab und nutzt diesen zur Visualisierung in einem eigenen Renderer. Dies setzt zwar voraus, dass die kompatible Software die offene Grafikbibliothek *OpenGL* nutzt, ermöglicht jedoch die Darstellung auf verschiedenen VE Plattformen wie Poweralls, CAVEs oder einigen HMDs.
So kann beispielsweise die Desktopbasierte CAD-Software *CATIA* mittels *TechViz VR* um deinen VR-Viewer erweitert werden. Laut Angaben des Unternehmens unterstützt das Produkt über 200 verifizierte Anwendungen. Da die Software keinen eigenen Datenimport benötigt, ist auch keine Datenaufbereitung notwendig. Zur Interaktion stellt *TechViz VR* eigene Interaktionskonzepte und Werkzeuge zur Verfügung. Die Werkzeuge der zugrunde liegenden Applikation wie etwa der CAD Software *CATIA* können dabei jedoch nicht verwendet werden. Somit bietet diese Lösung auch nur bedingt bis keine Prozessintegration, wodurch die Software nicht als Werkzeug für Schulungszwecke verwendet werden kann. Kollaboratives Arbeiten unterstützt das VR-plug-in lediglich unter Verwendung der eigenen Visualisierungssoftware *TechViz XL*. Diese bietet jedoch die gleichen Schnittstellen wie *TechVis VR*.

Improov3 Mit *Improov3*, bietet das ebenfalls französische Unternehmen *MiddleVR* ein Werkzeug zur Erstellung kollaborativer VR Sitzungen an. Die Software dient der Visualisierung von Konstruktionsdaten und bietet die erweiterte Visualisierung durch einige vordefinierte Interaktionswerkzeuge. Dies ermöglicht den Einsatz für verschiedene Engineering Anwendungen wie Produktreviews oder -präsentationen.
Die Software unterstützt Konstruktionsdaten in den gängigen Austauschformaten und einen direkten Datenimport ohne vorherige Konvertierung. Eine Darstellung von Prozessdaten wie sie beispielsweise für das Trainieren von Arbeitsabläufen nötig sind, wird von der Software nicht unterstützt.
Neben *Improov3* bietet das Unternehmen ein plug-in für die Game-Engine Unity3D an, die eine einfache Integration verschiedener VR Ein- und Ausgabegeräte ermöglicht. Kollaboratives Arbeiten ist aktuell allerdings nur direkt über *Improov3* möglich.
Beide Plattformen unterstützen eine Vielzahl verfügbarer Hardwarekomponenten für die Ein- und Ausgabe. So können Inhalte beispielsweise auf gängigen HMDs, Caves oder Powerwalls dargestellt, aber auch Eingabegeräte wie *LeapMotion*, *Kinect* und *VICON* verwendet werden.

4.2 Engine SDKs

Derzeit teilen sich einige Unternehmen den Markt für kollaborative VR-Applikationen, welche sogenannte Software-Development-Kits (SDK) vertreiben. Mit Hilfe dieser SDKs können eigenständig Anwendungen

entwickelt werden, welche die SDK spezifischen Komponenten nutzten. Dies hat den Vorteil, dass bei der Entwicklung explizit auf die jeweiligen Anforderungen eingegangen werden kann. Da sich diese SDKs jedoch in erster Linie nicht an Endkunden richten, ist deren Verwendung mit erheblichem Entwicklungsaufwand verbunden. Dies bedeutet, dass Unternehmen für den Einsatz dieser Produkte zusätzliche Ressourcen wie eigene Softwareentwickler oder Dienstleister in Anspruch nehmen müssen. Durch eine eigene Entwicklung können Anforderungen wie Datensicherheit oder Datenimport jedoch speziell auf die Bedürfnisse eines Unternehmens angepasst werden.

Innoactive Hub stellt eine durch das deutsche Entwicklerstudio *Innoactive* entwickelt und vertriebene Lösungen für kollaborative VR-Anwendungen dar. Das Unternehmen bietet hierbei ein SDK zur Integration in die Game-Engine Unity3D an, welches im Kern die Kommunikation und den Datenaustausch unter den verschiedenen Endanwendern (clients) abwickelt. Im Wesentlichen beschäftigt sich das Unternehmen jedoch mit der Entwicklung ihrer Plattform als Dienstleister für Unternehmen.
Neben deiner Kommunikationsschnittstelle werden durch das SDK unterschiedliche vordefinierte Bausteine für die Entwicklung einer VR-Applikationen bereitgestellt. So zum Beispiel vordefinierte Arten der Interaktion, welche der Nutzer etwa mit Bauteilen des geladenen Datensatzes tätigt.
Die so erstellten Sessions können anschließend als ausführbare Datei über das sogenannte *Innoactive Hub* an Nutzer verteilt werden. In Partnerschaften mit Unternehmen übernimmt *Innoactive* auf Kundenwunsch die Entwicklung der Anwendung und stellt diese dann über das *Hub* zur Verfügung.
Arbeitsprozesse welche beispielsweise für Trainings notwendig sind können in der Anwendung zwar hinterlegt werden, sind dann jedoch fest an die über das *Hub* veröffentlichte Sitzung gebunden. Nachträgliche Änderungen an den Prozessen oder Modellen sind somit ohne Verwendung der Game-Engine nicht möglich.

STAGE Das von *vr.on* verfolgt einen sehr ähnlichen Ansatz. Wie bei *Innoactive* bietet *vr.on* ein SDK zur Integration in eine Game-Engine. Aktuell wird das SDK unter Verwendung der *Unreal Engine* unterstützt. Der Support für Unity3D ist für Ende 2018 geplant. *STAGE* verarbeitet auch hier den Datenaustausch und die Kommunikation einzelner Clients mit einem Server. Auch diese Software bietet eine Toolbox mit vorgefertigten Interaktionskomponenten, welche einfach in der Engine zur Entwicklung der Anwendung verwendet werden können. Die Verteilung der Daten und Verwaltung der Sitzungen erfolgt über eine Desktopanwendung, ähnlich dem *Innoactive Hub*.
Das Angebot von *vr.on* richtet sich in erster Linie an Unternehmen, welche die Software zu Review und Präsentationszwecken in kollaborativen virtuellen Meetings abhalten möchten.

Der nicht unwesentliche Entwicklungsaufwand, welcher mit der Nutzung dieser Plattformen und der Erstellung neuer Inhalte verbunden ist, macht den Einsatz von SDKs in Hinblick auf die Anforderung der Anwendbarkeit weniger geeignet. Jedoch bieten sie eine hohe Flexibilität bei der Erfüllung weiterer Anforderungen wie der Datensicherheit und -integration.

4.3 Nicht kommerzielle Lösungen

CoAvatar Im Rahmen des Forschungsprojektes *CoAvatar* beschäftigt sich das Fraunhofer IAO Stuttgart gemeinsam mit dem Max-Plank-Institut für biologische Kybernetik Tübingen ebenfalls mit der Entwicklung einer Kollaborativen VR-Plattform. Diese dient Evaluationszwecken zum Forschungsthema virtueller Avatare und deren Wirkung auf die Nutzer. Die Plattform wird jedoch auch nach den Ansprüchen möglicher Anwender für den produktiven Einsatz in Unternehmen konzipiert und entwickelt. Dabei arbeiten die Teams gemeinsam an einem Protokoll zum sicheren Informationsaustausch in kollaborativen VR-Sitzungen.
So entstand ein Framework, welches ohne die Übermittlung sensibler Daten und Produktinformationen über das Protokoll auskommt. Diese basiert auf der Game-Engine Unity3D und unterstützt aktuell neben den verbreiteten HMDs, die Microsoft Hololens sowie einen Desktop-Viewer.
Neben der einfachen Visualisierung von Modelldaten bietet die Plattform zudem auch die Möglichkeit mit einzelnen Komponenten zu interagieren.

Das Virtual Training-Center (VTC) ist teil der Ergebnisse des vom BMBF geförderte Forschungsprojekt *Glassroom*, bei dem es um den Kompetenzaufbau und -entwicklung in virtuellen Lebenswelten ging. (vgl. [7]) Untersucht wurde der Einsatz von VR und Augmented-Reality (AR) für die Industrie 4.0. Das Konsortium, welches sich aus Anwender aus der Industrie, einem Softwareentwickler und mehrerer Forschungsgruppen aus der Didaktik und IT zusammensetzte, entwickelte gemeinsam Lösungen für den produktiven Einsatz von VR und AR zur Ausbildung und Assistenz von Mitarbeitern.
So entstand unter anderem die VR-Software *VTC*, welche für Schulungszwecke eingesetzt werden kann. Diese ist in der Lage Maschinendatensätze und Prozesse zur Laufzeit zu laden, und den Nutzer an der virtuellen Maschine durch die Arbeitsschritte zu führen. Diese Prozesse werden mit Hilfe eines Autorenwerkzeuges erstellt und verwaltet und können sowohl für die Wiedergabe in VR als auch AR genutzt werden. Gemeinsam mit diesem Tool zur Prozesserstellung *Glassroom Manager* bildet *VTC* ein Paket aus verschiedenen Werkzeugen, welches sowohl die Datenaufbereitung, die Prozesserstellung als auch VR und AR Software zur Wiedergabe der Lektionen umfasst.
Das Werkzeug zur Datenaufbereitung, welches Konstruktionsdaten in .jt darstellen und für die Verwendung in VTC konvertieren kann, wird außerdem in Verbindung mit dem Autorenwerkzeug für die Prozesserstellung und -Verwaltung *Glassroom Manager* verwendet.
Indem die im 3D-Viewer angezeigten Bauteile durch Anklicken direkt mit dem jeweiligen Arbeitsschritt verknüpft werden können, erhält der Nutzer bereits bei Prozesserstellung einen direkten visuellen Bezug zu den verknüpften Konstruktionsdaten. Durch die Möglichkeit der Datenkonvertierung mittels dieses Werkzeuges sind keine weiteren vorgelagerten Prozesse zur Datenaufbereitung notwendig.
Die so erstellten Inhalte können anschließend dynamisch in der virtuellen Lernumgebungen abgespielt werden. Die Verwendung einer Game-Engine und ein damit hohes technisches und spezifisches Know-how sind für die Erstellung neuer Inhalte nicht nötig.
Seit Abschluss des Forschungsprojektes

wurde das immersive VLE *VTC* in Kooperation mit den *Amazone Werken* durch das Fraunhofer IAO weiterentwickelt. Eine direkte Anbindung an PLM und PDM-Systeme ist derzeit noch nicht möglich jedoch aber geplant.
Entgegen der zuvor vorgestellten Lösungen bietet das *VTC* im aktuellen Entwicklungsstand über keine Multi-User Fähigkeit. Durch eine Integration der Kollaborative VR-Plattform *CoVr* in die virtuelle Lernumgebung *VTC* könnte diese jedoch geschaffen werden.

5 Fazit

Die unterschiedlichen Ansätze in der Umsetzung kollaborativer virtueller Umgebungen bieten eine Vielzahl an Möglichkeiten, diese auch für die Digitalisierung in der Aus- und Weiterbildung im beruflichen Kontext zu nutzen. Je nach Anwendungsgebiet sind die Anforderungen an eine solche immersive virtuelle Lernumgebung jedoch sehr unterschiedlich. Für den Einsatz im Produktentstehungsprozess spielen die Visualisierung und Verarbeitung der Konstruktionsdaten eine entscheidende Rolle. Wohingegen bei der Kompetenzentwicklung in zwischenmenschlichen Tätigkeiten wie beispielsweise der Kranken- und Altenpflege nicht nur körperliche Tätigkeiten und Handgriffe im Fokus eines Trainings stehen, sondern vielmehr der Umgang und das Verhalten in gewissen Situationen.
In beiden Fällen ist die Abbildung von Prozessen in einem vom System interpretierbaren Format notwendig. Wo die Digitale Verfügbarkeit von Produktinformationen derweilen alltäglich ist, werden Prozesse und Abläufe oftmals noch immer auf analoge Weise von Mensch zu Mensch weitergegeben. Das bedeutet auch, dass um den produktive Einsatz von VLEs zu Trainingszwecken zu ermöglichen, eine Datenbasis über die Prozessinformationen notwendig ist.
Durch Ansätze wie die im Forschungsprojekt *GLASSROOM* verwendete Prozessmodellierung in *BPMN*, kann eine solche Datenbasis geschaffen werden. Durch die Integration dieses definierten Standard sowie die vorhandene 3D-Datenbasis könnten VLEs generisch und für den Anwender intuitiver gestaltet werden.
Viele der vorgestellten Lösungen bieten umfangreiche Möglichkeiten der Visualisierung. Oftmals wird virtuelles Training als möglicher Anwendungsfall genannt. Praktisch bieten jedoch nur wenige dieser Lösungen Werkzeuge und Schnittstellen für die Integration von Prozessinformationen, welche für die Darstellung von Arbeitsabläufen entscheidend sind.
Im Hinblick auf die gesetzten Anforderungen ist es derzeit nicht möglich eine vorhandene Lösung ohne weitere Anpassungen zur Verwendung als Kollaborative VR-Lernplattform einzusetzen. Die vorgestellten Komplettlösungen zielen oftmals auf den Produktentstehungsprozess und virtuelle Reviews ab. Die Integration von Prozessen wird deshalb nur selten unterstützt. Engine-SDKs bietet eine gute Grundlage zur Entwicklung eine solchen Plattform. Wie das Forschungsprojekt *GLASSROOM* zeigte, lassen sich durch die Integration von standardisierten Prozessbeschreibungen virtuelle Lernplattformen schaffen. *VTC* bietet hierbei die größte Überdeckung mit den Anforderungen. Leider verfügt diese Lösung aktuell über keine Multi-User Fähigkeit. In Verbindung mit der Kollaborationsplattform *CoVr* könnte diese jedoch hergestellt werden.

6 Ausblick

Über den Produktentstehungsprozess hinaus werden schon heute Maschinen oder gesamte Anlagen von *digitalen Zwillingen* begleitet. Diese bilden den aktuellen Zustand einer Anlage digital ab. Diese Datenbasis könnte dazu dienen, bereits vor der Aussendung eines Technikers mögliche Probleme oder Störungen zu erkennen und Lösungsansätze zu entwickeln. Der Einsatz Immersiver VEs zur Darstellung dieses digitalen Zwillings könnte dabei helfen mögliche Probleme noch schneller zu identifizieren. Durch die Verwendung einer kollaborativer VR-Plattformen könnten so auch Techniker vor Ort in ihrer Tätigkeit assistiert und geführt werden. (vgl. [7])

Literatur

[1] R. Lindgren, M. Tscholl, S. Wang und E. Johnson, „Enhancing learning and engagement through embodied interaction within a mixed reality simulation“, *Computers & Education*, Jg. 95, S. 174–187, 2016, ISSN: 03601315. DOI: 10.1016/j.compedu.2016.01.001.

[2] I. V. Alexandrova, M. Rall, M. Breidt, U. Kloos, G. Tullius, H. H. Bulthoff und B. J. Mohler, „Animations of Medical Training Scenarios in Immersive Virtual Environments“, in *2011 Workshop on Digital Media and Digital Content Management*, IEEE, 15.05.2011 - 16.05.2011, S. 9–12, ISBN: 978-1-4577-0271-6. DOI: 10.1109/DMDCM.2011.64.

[3] Ivan E. Sutherland, „The Ultimate Display“, in *Proceedings of the IFIP Congress*, 1965, S. 506–508.

[4] G. Piccoli, R. Ahmad und B. Ives, „Web-Based Virtual Learning Environments: A Research Framework and a Preliminary Assessment of Effectiveness in Basic IT Skills Training“, *MIS Quarterly*, Jg. 25, Nr. 4, S. 401, 2001, ISSN: 02767783. DOI: 10.2307/3250989.

[5] S. E. August, M. L. Hammers, D. B. Murphy, A. Neyer, P. Gueye und R. Q. Thames, „Virtual Engineering Sciences Learning Lab: Giving STEM Education a Second Life“, *IEEE Transactions on Learning Technologies*, Jg. 9, Nr. 1, S. 18–30, 2016, ISSN: 1939-1382. DOI: 10.1109/TLT.2015.2419253.

[6] Y. Rosmansyah und H. Ashaury, „A 3D multiuser virtual learning environment and learning management system“, in *Proceedings of the 2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*, N. A. Mat Zin und I. C. o. E. E. a. Informatics, Hrsg., Piscataway, NJ: IEEE, 2017, S. 1–6, ISBN: 978-1-5386-0475-5. DOI: 10.1109/ICEEI.2017.8312342.

[7] O. Thomas, D. Metzger und H. M. Niegemann, Hrsg., *Digitalisierung in der Aus- und Weiterbildung: Virtual und Augmented Reality für Industrie 4.0*. Berlin: Springer Gabler, 2018, ISBN: 9783662565513.

[8] R. Lindgren und D. L. Schwartz, „Spatial Learning and Computer Simulations in Science“, *International Journal of Science Education*, Jg. 31, Nr. 3, S. 419–438, 2009, ISSN: 0950-0693. DOI: 10.1080/09500690802595813.

[9] J. Drosdol, J. Kieferle und U. Wössner, „The Integration of Virtual Reality (VR) into the Architectural Workflow“, in *21th eCAADe Conference Proceedings. Graz, Austria*, 2003, S. 25–28.

[10] Daimler AG, *Virtuelle Zertifizierung – Virtual Reality bei Mercedes-Benz*, 27.04.2009. Adresse: https://blog.daimler.com/2009/04/27/virtuelle-zertifizierung-virtual-reality-bei-mercdes-benz/.

[11] Audi AG, *Audi testet „Virtual Reality Holodeck“ für schnellere Produktentwicklung*, Ingolstadt, 22.02.2018. Adresse: https://www.audi-mediacenter.com/de/pressemitteilungen/audi-testet-virtual-reality-holodeck-fuer-schnellere-produktentwicklung-9873.

Linked knowledge – Managing information like the human brain

Lucas Hermann
Hochschule Reutlingen
lucas_jan.hermann@
student.reutlingen-university.de

Abstract

A big amount of data is created every day. Nearly everyone acts a part in this data explosion and nearly everyone is affected. This growing amount of data on personal devices such as smartphones, tablets, laptops and PCs is increasingly hard to handle for users. The continued usage of old folder hierarchies makes it hard to archive and find all the data. In contrast our brain is capable of storing an impressive amount of information while maintaining a very good search process. Not only can we find single pieces of knowledge within a thought, but we also get all contextual relevant information – a capability our digital devices lack. This paper takes a closer look at the functioning of our brain and tries to propose helpful concepts how data could be stored, searched and retrieved in a more natural way.

Keywords

Semantic Search, Contextual Data, Linked Data, Personal Information Management, Folder Hierarchy

---

Betreuer Hochschule: Prof. Dr. Tullius
Hochschule Reutlingen
gabriela.tullius@reutlingen-university.de

Betreuer Firma: - - -

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen
Copyright 2018 Lucas Hermann

CCS Concepts

• **Information systems → Data management systems** • **Information systems → Users and interactive retrieval** • **Information systems → Desktop search** • **Human-centered computing → User centered design** • *Software and its engineering → Search-based software engineering* • *Computing methodologies → Semantic networks*

1 Introduction

1.1 The age of data explosion

Data is nothing that we have invented in the 20th/21st century but the amount of data available to humans is something that makes the 21st century very different from all the time before. Mankind has invented an incredible amount of devices that can collect, store, share and process information in digital form. No surprise the amount of processed and stored data has exploded from “2.6 (optimally compressed) exabytes in 1986 (…) to 295 optimally compressed exabytes in 2007” [1]. And the explosion goes on: 2007 Hilbert and López did estimate that the yearly growth of storage capacity lies at about 23%. But where does all the data come from and how could a normal person handle his/her massive amount of personal data and information?

1.2 The origins of data

In [2] Hilbert analyses the reasons of the data explosion and states “that the world’s technological capacity to store, communicate, and compute information has mainly been driven by better, and not merely more technology” [2]. Since this study in

2014 the ongoing research and development of new devices, especially of small sensor based devices like smart watches, fitness trackers and mainly smart sensors for the internet of things has lead to a big amount of new data sources that foster the data explosion. Since the beginning of digitization "network architectures and computing models have oscillated between the use of shared and central resources" [3] starting from first Personal Computers and the use of mainframes to laptops, mobile phones, the cloud and continuing to mixed concepts like edge intelligence (cf. [3]).

1.3 How is Information and data handled in daily life?

In our daily life we interact with a small part of this massive amount of data and we increase the amount by writing text messages, taking pictures and videos, writing notes of our research and so on. And because we produce a lot of data, we also try to organize it. Today most files on private user's hard disks are organized in folders and its sub folders, a system invented in 1958 [4]. Folders allow us to sort the data manually into a hierarchy that represents relation of the files to a topic or each other.

Besides files some other types of data are handled and stored differently: Conversations in our email and messaging clients are stored in databases not in single files and are displayed in the chronological order of the conversation. Pictures on smartphones are often displayed in albums of different days, months, events or locations. Tracks, routes and locations are displayed on maps making it possible to quickly detect distances, groups of locations, etc.

Still it seems to become more complicated to find the file or the information you where looking for. In our minds we often have a good idea of what that information was related to and in which situation it was recorded or send to us. Naturally we seam to be capable of organizing files in a far more intuitive way. In the next chapter we will analyze how information is stored, organized, structured and retrieved in the human brain. Based on this knowledge we will compare existing technological solutions and propose basic guidelines how data should be handled and exposed to the user for a more intuitive and natural interaction and exploration.

2 Semantic Knowledge in the human mind

The human mind exists for the purpose of processing, storing and recalling information. To get an idea about how associated information is linked in the human mind we will take a look at different processes and models. Biological psychology knows that information we perceive goes through multiple areas of the brain for different purposes.

2.1 Sensory memory

The input from our different senses runs through our nervous system into sensory memories where the raw input gets filtered, analyzed and itemized into its temporal, spacial and semantic fragments. Feature extraction, pattern recognition and item detection are tasks that also take place in the sensory memories. Still information stays in the sensory memories for a very short amount of time (less than 1s in the iconic memory for what our eyes perceive). (cf. [5] P.627-628)

2.2 Short-term memory

Preprocessed input from the sensory memories goes to the brains short-term memory where the input is further processed. One important step in processing the information is chunking, where single parts of information (e.g. single letters or notes) were put together to one a more meaningful piece of information (e.g. a word or a melody). While the short-term memory can only hold 5-7 chunks at a moment the amount of information per chunk can become very high. Additionally to chunking the input the short-term memory searches for relevant associations between the chunks and connects them accordingly. Associated information also gets reorganized to create even more meaningful pieces of information. Because of the very limited capacity of the short-term memory only chunks and pieces of information with constantly new activation signals are kept. All other chunks and information is dropped immediately. The new activation signals that keep the chunks in memory can either be new input from the sensory memory or internal feedback from similar information stored in the brain. Chunks that stay in the short-term memory for a certain time and build enough amount of information are stored into the long term memory. (cf. [5] P.627-628)

2.3 Long term memory

Information in the long term memory is differentiated between implicit and explicit knowledge:

2.3.1 Implicit memory

Implicit memory (also known as procedural memory) contains knowledge that is accessed by routine in daily life without conscious awareness. Perceptual tasks like face recognition, basic concepts for understanding situations, classic conditioning and every day's skills like walking are require very little mental effort because they are stored in the implicit memory. Accessing knowledge from the implicit memory intentionally is possible but requires a lot effort and concentration. Because this paper focuses on the conscious retrieval of knowledge and it's contextual information the implicit memory will not be discussed in detail. (cf. [5] P.621-622)

2.3.2 Explicit memory

The explicit memory or declarative memory contains the information that we willingly access in our daily life. Again the explicit memory is differentiated between episodic memory and semantic memory.

The episodic memory stores and links information that mainly is connected to a temporal aspect. It records memories and situations by the time of their experience and puts them in context with the previous and next elements as well as with the knowledge from the semantic memory related to the situation. So when we recall an event from our past the episodic memory makes it possible for us to determine whether event A took place before event B and at which age the event took place. (cf. [5] P.622)

The semantic memory is the basic storage of all the basic pieces of knowledge we consciously access. Singles pieces of information are represented by cell-assemblies of neurons that can compose closed loops of activation, known as reverberatory circuits. Relation between the single pieces of information are maintained by connections that pass activation potentials from one cluster to related clusters (cf. [5] P.622+629). The main part of this complex information network, of our semantic memory resides in the four layers of the hippocampus. New knowledge can be stored in different ways: Formerly silent bus existing connections can be improved by repeated activation signals. Especially in the hippocampus new neurons can be created by a process called "Neurogenesis" to increase the capability of information storage (cf. [5] P.632).

2.3.2.1 Hebbian learning

Another way the brain stores and connects new information is by increasing the transmissive capability of synapses between two neurons that get activated repeatedly.

Cell-assemblies that improve their synaptic connections when firing repeatedly in a short amount of time are named Hebb Synapses after the Canadian scientist Donald O. Hebb how first suggested this principle. Usually Hebb Synapses are made up of so called “reverberatory circuits”, neural cells that are connected to another in the way that they build a closed circuit that repeats to fire round and round again. The specific location, activation frequency and rhythm of this closed-loop cell-assemblies make up single pieces of knowledge. But Hebb’s Rule not only applies for reverberatory circuits but for many neurons in general, “when an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased” ([6] P.62). In a short form Hebb’s rule today is often cited as “Neurons that fire together wire together” ([5] P.631). This has not only been proved to be valid by John Eccles but also is applied in the training of some artificial neural networks [7],[8]. Still this rule is not the universal answer to how our brains are learning new information, but is only one part of the whole picture on the biological level.

2.3.2.2 Models of semantic knowledge

As it currently not yet possible to measure the exact pieces of information stored in the cell-assemblies or their connection, cognitive psychology has developed multiple models to describe how and what items of information are stored in our brains. Some of the more popular models will be compared in the following table. All of them are only models and try to make a description. Experiments have partially confirmed most of them, still none of them is really proven to be exact. These models are our most basic understanding of how humans retrieve contextual information from their brain and therefore describe our most natural way to find specific pieces of information inside the big net of knowledge.

Table 1: Comparison of Models from Cognitive Psychology

Hebbian Learning (cf. [5] P.629ff , [6])	**Spreading activation Network (cf. [9] P. 179ff)**	**Neural Networks (cf. [9] P. 182ff)**	**Schema theory (Barlett) (cf. [9] P.192ff)**	**Script Theory (Schank & Abelson) (cf. [9] P.194ff)**
		Typical connection schema		
From node to every related node. Often forming reverberatory circuits	From node to every related node	From node to every node of lower layer	Schemata can contain “slots” that can be filled with the knowledge from specific other schemata	Scripts link to other scripts that contain parts of the activity
		Storing information		
The single fact/information is stored in the rhythm, location and activation frequency. Related information is connected by outbound synapses that activate the related cells.	One Concept/Word per Node Type of Relation in edge Strength of relation in transmissibility of edge	Knowledge of the net is stored distributed over all weights	Schemata contain concepts of known information Slots are pieces of information that relate to another schema => The combination of schemata completes information	Scripts contain knowledge about processes and activities. One Script can link to/ consist of multiple other scripts that describe parts of that activity
		Training		
Hebbian Learning (cf. [5] P.629ff , [6])	**Spreading activation Network (cf. [9] P. 179ff)**	**Neural Networks (cf. [9] P. 182ff)**	**Schema theory (Barlett) (cf. [9] P.192ff)**	**Script Theory (Schank & Abelson) (cf. [9] P.194ff)**
Connections between neurons were A makes B fire become stronger/more transmissive	Building new edges, creating new nodes, changing transmissibility of edges, changing relation of edges	Changing the weights of the links	New Schemata can be defined Knowledge from existing schemata can be split into a separate schema leaving a slot in the original schema	New scripts and links to other scripts can be created One script can be split into multiple scripts describing parts of the activity
		One time Training possible		
small effect	yes	no	yes	yes
		Strength of Relation between two directly connected nodes		
Transmissibility of the link “neurons that fire together wire together”	Number of edges/links + transmissibility of the links	Transmissibility of the link	Related Schemata are connected, unrelated have no connection (no indication for strength of relation)	Relating scripts are connected, unrelated have no connection (no indication for strength of relation)
		Strength of relation in a context		
Sum of activation that reaches a node	Sum of activation that reaches a node over the edges	Single nodes do not contain information. Context is the values of all output nodes	Context is built from Schemata having slots and linking to schemata that have slots again	Scripts can be related to scripts that again relate to scripts. Helps in understanding related new information by giving a rough idea.
		Special 1		
With every use of an information that helps trigger another information, the connection between both grows more transmissive.	Validity of Information is checked by activating all affected nodes, spreading the activation over the edges till the affected nodes are connected by active edges. Then comparing the relation stored in the connected edges with the information from the initial input.		Similarity with inheritance in Object Oriented Programming: Schemata can extend upon exiting other schemata.	A detailed addition to the model of Schema theory Activation of a script can massively improve understanding of related new information. Missing information in the stream of new information (e.g. a conversation) are easily filled by a related script.
		Relevance for Cognitive Psychology		
Proven to be correct for at least a part of the neurons and synapses in the brain. High relevance.	Interesting concepts with some weak point. Good possible explanation for the “semantic priming effect”	Hard to test with human test subjects. Mainly used as a model in computer simulation and AI.	Interesting because it’s the base of some other more detailed models (like the script model). Additionally it offers a good explanation why perceived information is adapted to personal knowledge.	Is a good explanation of multiple observations. High relevance for cognitive Psychology.

3 The science of personal Information

Besides Cognitive and Biological Psychologies models about knowledge, memory and the human brain, information technology has a lot of relevant scientific work on the topic of personal information management. Computer Scientists have conducted research on the behavior, habit and preferences in managing files and information. Before proposing guidelines for the design and functionality of personal information management software, it is important to include the knowledge already gathered in this field.

In the attempt to get a good overview of the state of research, a list of keywords related to the topic of information management was chosen. Next the four conferences with the most publication matching this keywords were determined. The final list of interesting proceedings what created by searching the database of all four conferences for papers published between 2012 and 2018 matching at least one keyword. The five keywords were:

- Semantic Information Retrieval
- Semantic Context
- Folder Hierarchy
- Folder Structure
- Personal Information Management

The four most relevant conferences were:

- Conference on Human Factors in Computing Systems (ACM CHI)
- International World Wide Web Conference (ACM WWW)
- Conference on Hypertext and Social Media (ACM HT)
- Conference on Information and Knowledge Management (ACM CIKM)

The short web abstracts of all 104 papers matching at least one keyword were read to determine which papers are relevant for the topic of information management. The knowledge gathered from the most relevant papers will be presented next.

Several papers presented research done on the user behaviors, performance and preferences in personal information management. Though all of this research is bases on the widespread file-folder paradigm a lot of the results seem to be applicable for all types of information management.

In [10], [11] and [12] the use of folder hierarchies are subject to research. The results show that there are different types of users sorting their files either in broad and shallow hierarchies or in deep, specific structures. These differences are as well depending on personality traits, mainly on conscientiousness (cf. [11]) but also on the type of work the persons did use the files for (cf. [12]).

Independent of the type of user, folders with only one to three files are common (cf. [12]). This indicates that users try to put their files in folders that fit the specific category of the file – probably to find it faster and to avoid chaos by keeping to many similar files in the same folder.

When given the task to retrieve an certain information, users mainly use navigation for files and seven times more often the search functionality when retrieving emails (cf. [10], [13]).

According to [13] emails are by far easier to search for, because they have one very simple attribute – the sender of the mail. Even if users do not remember the full content of the message they usually know how send the information.

In contrast files have a lot of content, that often is hard to index and search. File names are the only clear attribute and are often not an information the user has when he/she searches for a file created in a specific context or situation. Navigation through folders to the correct sub folder makes it easier to find the file by always selecting the most likely (sub)-folder.

For files that are frequently used, file system paths often are recognized as one “location” found at a specific place in the folder hierarchy instead of a path consisting of multiple elements (cf. [10],[13]). Users need to go through multiple folders to reach this known location, often without an actual need for any of the information from the folders they step through. Knowledge of the locational disposition of a file corresponds to the concept of locational or situational memory that acts an important role in out brain (e.g. episodic memory) when in comes to retrieving our knowledge (cf. [13]).

For both mail inbox and the users desktop there is an exception from the paradigm of sorting information in a hierarchy related to it’s content. Users tend to use both places to store emails and files that remind them of important tasks (cf. [13] P.623) and thereby clutter both places with files of different origin and context.

Regarding the retrieval of information users tend to access files, mail, etc. most often, that they have used recently or that stand in close relation to recently used data. Operating systems and applications try to support the user by giving lists of the last used files, last visited web pages (cf. [10]).

In [14] Nebeling et al. give insight about multi user collaboration online file systems like Dropbox or Google Drive and propose additional views and features to enhance workspace awareness. They show that it’s important to give users and overview of activities in the file system and the specific folder and to help them find moved, renames and deleted files. The proposed and tested changes are “notifications on changes suggested by other users, multi-user aware folder views (…) and shadow file with trace” [14] that point to deleted or moved files.

All of these single pieces of knowledge are important for the proposal of new guidelines for the design and functionality of personal information management tools. At the same time this is only a part of the huge amount of scientific literature covering this subject. Additionally the subject and field of interest is much broader reaching from tagging to information retrieval. In the search for related work a lot more papers were found that are suitable to augment the knowledge in this field. Because of the limited timed nature of this paper only a part of all literature could be read and analyzed.

4 Guidelines for natural and intuitive information management

In the following section basic guidelines for natural and intuitive personal information management are proposed. They include results and knowledge from both biological/ cognitive psychology as well as current research in information technology. Still the knowledge these guidelines are based on is far from complete making them a good basis for future research and changes instead of fixed laws. Still they give a clue why many current information management tools including Mac Finder and Windows Explorer feel tedious, ineffective and difficult to use.

From section 2.3.2 we know that information of different types is connected in the human mind. Concepts of single objects, scripts for behavior in certain situations and temporal information from the episodic memory makes up our knowledge of our past. Location, different persons and task make up projects or other type of contexts. When we search for emails we search for a sender not for the exact subject or date of the mail because we know that the information we need lies in this context.

(1) For intuitive personal information management (short: PIM) a tool that can store, receive and link all types of information is proposed. The tool should

support direct links (e.g. between a file and a person) as well as links to meta-information like tags, projects, etc.

Hierarchical structure matters a lot to the users. As seen in section 3 users tend to navigate to files by the hierarchical structures. Even in the brain scripts and concepts of different objects are understood hierarchically (see section 2.3.2.2). For the current file-folder paradigm hierarchy is the reason it works but also the culprit. Users can sort their files only to one location without copying them. Copied files do not change when one of their copied siblings is changed. The use of symmetrical links (symlinks) could solve this but is not feasible in everyday live .

(2) Because navigation and hierarchy matters an intuitive PIM tool should allow two types of information linking: hierarchical links and associative link that indicate that two entities stand in relation without giving them a hierarchical structure.

(3) As the file-folder paradigm fails because a file can only reside in one folder, a intuitive PIM tool should allow files and any type of information to be linked to and below multiple other information entities, thus making it accessible in every related context.

(4) A PIM tool inspired by the natural way humans process store and retrieve knowledge could also imitate learning like the human brain. Therefore it is proposed that an automated system that helps to tag and link the information entities learns with every new link created. The learning could be inspired by Hebbian learning (section 2.3.2.1). If an information A is linked to information B and then to information C this could be treated like B makes C fire and "what fires together wires together" ([5] P. 631). If the user links another information to B the system should be more likely to offer C as another linking target. This learning could also include inhibition (link-offers the user rejects) and the types of links used.

(5) In the user interface information should be visible with it's linked entities including the possibility to follow the links to hierarchical higher and lower levels as well the associative links to other related entities or contexts of information. By offering this type of interaction users can explore, retrieve and navigate their information. As navigation is still the preferred method of file retrieval (cf. section 3) it is important to offer this way of interaction. Knowledge retrieved in our brain also is often found by remembering corresponding situations, persons, locations, etc. Respectively navigating information and context by its associated knowledge or information is a very natural behavior.

(6) As mentioned in section 3 searching is very seldom used for file retrieval. One reason for this behavior is the fact that file contain no attributes that are easily remembered and used for search. By linking files with other information entities, file relate to usable attributes. Another problem of current file searching functionality is that it is often not accurate enough to find the exact file the user was searching for. In contrast for emails it is sufficient to find find the corresponding conversation in that the email appeared (cf. [13]) to retrieve the information. Because it is easier to find information in it's corresponding context it is proposed that every search leads to a context built of related information instead to a single file. Besides search this context view should be the main way of navigating through information in the PIM tool.

(7) Current research shows that files, emails and web pages that were opened recently are most likely to be required and opened again (cf. section 3). In the human brain information is most easily accessed when the surrounding situation is similar to the situation when the information was stored (cf. [5]). A PIM tool with linked information should use the current geo-location, soon upcoming events and the habits at this time of the day to suggest the user the most relevant contexts in every situation. E.g. when a user is at his university the PIM tool should offer the current study projects as relevant contexts to the user. If there is an upcoming lecture the PIM tool should also suggest all information related to this subject of studies. If the user is at home the system should offer contexts with the private projects and maybe a list if tasks if relevant. Giving the user related contexts based on the versatile information of is current situation could reduce the time needed to access a relevant information obviously.

(8) As shown in section 3 the desktop and email inbox is mainly used to store files and mails that remind of important tasks. Because of new incoming emails "keep pushing the actionable messages down beyond the screen limits, these reminders tend to go out of sight (…)" [13]. Same hold true for the desktop where the amount of important task related files sooner or later clutters the space. Both are problems that arise from the fact that users do use the inbox and desktop for a function they were not designed for. The reason users do this is, that they lack an easily accessible tool that helps them remind tasks, without lots of input needed and with a direct access to the related files, conversation and information. A intuitive PIM tool requires a function that makes it possible to link emails, files, web pages and every other type of information to different tasks. These tasks need to be show to the user on the startup/home-screen corresponding to his situation. A tool like this could avoid misuse of inboxes and would likely have a better performance in reminding of important tasks.

(9) When we process information and input in our brain there are always different areas of knowledge that get activated. As described in section 2.2 our short-term memory can handle between 5 and 7 chunks of information. A PIM tool designed for natural and intuitive use should thereby offer the possibility to access multiple contexts at the same time (either with tabs, split screen or any other GUI paradigm).

(10) As the has his knowledge where ever he is the PIM tool should also offer the up to date information and linking everywhere – On the PC, laptop, smartphone and ideally on every digital device in the world (as soon as the user has authenticated). There must not be any need to copy information from one device to the other to keep it up to date.

From my personal opinion and experience I also recommend the following guidelines (as I would not use a tool that does not adhere to them):

(11) Give the user full control over his data. Give him the possibility to be the hoster of his cloud and to encrypt data before he syncs it onto foreign servers.

(12) Give the user a way to access all data stored in the PIM tool even if he can not start the PIM tool itself (e.g. by building a symlinked folder hierarchy in the background)

(13) Give the user an interface and ways of interaction that feel familiar even if the tool does a lot of things different from all it's predecessors.

5 Conclusion

In this paper, an introduction to data, the data explosion and why it affects personal users in every day's life was given. To find good guidelines for an intuitive PIM that manages information the way humans do naturally, we had a look at the current state of knowledge about the human mind and cognitive models in biological and cognitive psychology. Also current knowledge gained about information and file management was summarized. Based on this knowledge from different fields a set of proposed guidelines

was introduced that explain how an PIM tool could be designed to manage information more like the human mind does. As the knowledge used to develop these guidelines is far from complete, these guidelines should be subject to further research. Still the discrepancy between current PIM tools, everyday file management workflow and the guidelines proposed here should clearly show that a new approach to personal information management is necessary.

To achieve a better understanding of the users needs current PIM software and it's usage patterns could be compared to these guidelines. Additionally knowledge from other scientific fields like User Interface Design, AI and Information Retrieval should be included in the guidelines.

To gain knowledge of the usability and acceptance it is strongly recommend to build the prototype of a PIM tool that follows these guidelines to test it in the field with real users and on their behavior.

References

[1] Martin Hilbert, Priscila López: The World's Technological Capacity to Store, Communicate, and Compute Information, Science (Vol. 332), 2011, DOI: 10.1126/science.1200970

[2] Martin Hilbert: How Much of the Global Information and Communication Explosion Is Driven by More, and How Much by Better Technology?, Wiley, 2014, DOI: 10.1002/asi.23031

[3] Edge intelligence, International Electrotechnical Commission, 2017, URL: https://www.iec.ch/whitepaper/edgeintelligence/?ref=extfooter (accessed 08.10.2018)

[4] G. A. Barnard, L. Fein: Organization and Retrieval of Records Generated in a Large-Scale Engineering Project, ACM, 1958, DOI: 10.1145/1458043.1458058

[5] Niels Birbaum, Robert F. Schmidt: Biologische Psychologie, Springer Medizin Verlag Heidelberg, 7. Auflage, 2010, ISBN: 978-3-540-95937-3

[6] Donald O. Hebb: The Organization of Behavior : A Neuropsychological Theory, Lawrence Erlbaum Associates, Inc, , 2002, ISBN: 0-8058-4300-0

[7] John J. Hopfield: Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences (Vol. 79), 1982, ISSN: 0027-8424

[8] Richard W. Strom: Hebbian Learning in Multilayer Neural Networks, 2007, DOI: 10.1.1.410.3202

[9] Franz J. Schermer: Lernen und Gedächtnis, Kohlhammer, 5. Auflage, 2014, ISBN: 978-3-17-022013-3

[10] Ofer Bergman, Steve Whittaker, at al.: How do we find personal files?: the effect of OS, presentation & depth on file navigation, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12) ACM, 2012, DOI: 10.1145/2207676.2208707

[11] Charlotte Massey, Sean TenBrook, Chaconne Tatum, Steve Whittaker: PIM and personality: what do our personal file systems say about us?, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14), ACM, 2014, DOI: 10.1145/2556288.2557023

[12] Hong Zhang, Xiao Hu: A quantitative comparison on file folder structures of two groups of information workers, Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '14), IEEE Press, 2014, ISBN: 978-1-4799-5569-5

[13] Ofer Bergman, Nimrod Yanai: Personal information retrieval: smartphones vs. computers, emails vs. files, Personal and Ubiquitous Computing (Vol. 22/4), Springer London, 2018, DOI: 10.1007/s00779-017-1101-6

[14] Michael Nebeling et al.: MUBox: Multi-User Aware Personal Cloud Storage, In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15) (Vol.), ACM, 2015, DOI: 10.1145/2702123.2702233

Möglichkeiten zur Repräsentation von Emotionen bei der Kommunikation

Frauke Hanna Grieble
Hochschule Reutlingen
frauke_hanna.grieble@student.reutlingen-university.de

Abstract

Bei der sozialen Interaktion mit anderen Personen helfen uns Emotionen, Informationen zu kommunizieren. Dabei können subjektive Gefühle, Verhaltensintentionen oder Bewertungen einer Situation übermittelt werden. Zentral für diese Übermittlung ist auf der einen Seite die Produktion und auf der anderen Seite das Erkennen und Interpretieren des Emotionsausdrucks. Um eine mögliche Unterbrechung dieser Kommunikation zu umgehen, muss eine passende Repräsentation von Emotionen gefunden werden. Dazu ist eine Repräsentation in Form von Basisemotionen oder Gefühlsdimensionen möglich. Letztere bietet dabei einige Vorteile und kann durch Messungen der Mimik sowie physiologischer Reaktionen des Körpers erhalten werden.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; • **Computing methodologies** → *Modeling and simulation*;

Keywords

Emotionstheorie, nichtverbale Kommunikation, Emotionsausdruck

Betreuer Hochschule:	Prof. Dr.-Ing. Cristóbal Curio Hochschule Reutlingen Cristobal.Curio@Reutlingen-University.de
sowie	Salma Maath Ahmad Thalji Hochschule Reutlingen Salma.Thalji@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
21. November 2018, Hochschule Reutlingen

Copyright 2018 Frauke Hanna Grieble

1 Einleitung

Bei der sozialen Interaktion verlassen wir uns besonders auf die Sprache als Mittel zur Kommunikation. Allerdings werden auch sehr viele Signale und Hinweisreize nicht-verbal mittels Mimik, Gestik, Körperhaltung oder der Stimme an den Interaktionspartner übertragen. Besonders die Mimik einer Person beinhaltet viele wertvolle Informationen, die es dem Partner ermöglichen, emotionale Zustände und damit auch Handlungen und Intentionen in einem bestimmten Kontext zu erkennen und richtig zu verstehen [1].

Um diese Informationen nutzen zu können, müssen sie von dem Interaktionspartner zunächst visuell verarbeitet und anschließend auch im gegebenen Kontext richtig interpretiert werden. Bei dieser Übertragung der Information kann es zu Fehlern oder auch zu Unterbrechungen kommen, wenn die visuelle Wahrnehmung einer oder beider Interaktionspartner gestört ist. Um diesen Verlust der visuellen Information zu kompensieren, können sensorische Substitutionssysteme eingesetzt werden. Diese versuchen Reize einer Sinnesmodalität in Reize einer anderen umzuwandeln, um somit den Wegfall der einen zu kompensieren. Das vorliegende Paper betrachtet genauer den Fall, bei dem ein Wegfall der visuellen Wahrnehmung vorliegt, wodurch die Übertragung von emotional wichtigen Inhalten gestört wird. Bei der Überlegung, wie dieser Wegfall an Informationen kompensiert werden kann, kommt die Frage auf *Wie können diese emotionalen Inhalte repräsentiert werden?*

Diese Frage soll in der folgenden Arbeit genauer betrachtet werden und um diese beantworten zu können, werden zunächst die Fragen geklärt, was Emotionen sind und wie diese bei der interpersonellen Kommunikation genutzt werden. Anschließend werden Möglichkeiten zur Klassifikation von Emotionen genauer analysiert. Dies beeinflusst auch die Art und Weise, wie Emotionen gemessen werden, was in Abschnitt 5 betrachtet wird. Zuletzt werden die Ergebnisse zusammengeführt und die Frage zur Repräsentation erneut aufgegriffen.

2 Definition von Emotionen

Um die Frage beantworten zu können, wie Emotionen repräsentiert werden können, muss zunächst geklärt werden, was Emotionen eigentlich sind und wozu sie genutzt werden.

Emotionen haben einige Eigenschaften und Komponenten, die sie kennzeichnen [1]: Dazu zählt zum einen eine *subjektive* Komponente, was die *Affektivität* von Emotionen beschreibt, also die Tatsachen, dass diese von uns empfunden werden können [1]. Diese Komponente ist per Definition subjektiv und kann daher auch nicht objektiv gemessen werden [2]. Dies trifft ebenfalls auf die *kognitive* Komponente zu. Diese beinhaltet eine Einschätzung und eine Bewertung der Situation, wodurch eine Emotion reguliert und geformt werden kann. Anders ist dies bei der *physiologischen* Komponente, die sich auf die neuronalen und hormonellen Veränderungen bezieht, die beim Erleben einer Emotion ausgelöst werden [2]. Die vierte Komponente ist eine *Verhaltenskomponente*, die sich ebenfalls objektiv beobachten lässt. Gemeint mit dieser Komponente sind Veränderungen im Verhalten, wie beispielsweise die Bewegung bestimmter Gesichtsmuskeln [2].

Besonders die subjektive und die physiologische Komponente der Emotionen können nicht willentlich gesteuert werden, sondern werden *unwillkürlich* ausgelöst [1]. Diese Eigenschaft lässt sich nicht auf die Verhaltenskomponente übertragen, denn diese kann von der subjektiven Komponente abweichen. Allerdings trifft für alle Komponenten von Emotionen zu, dass sie *objektgerichtet* sind [1]. Das bedeutet, dass sie auf ein bestimmtes Objekt oder eine bestimmte Situation bezogen sind, von der sie ausgelöst werden. Auch die Dauer einer Emotion ist in allen Komponenten *zeitlich begrenzt* und an das emotionsauslösende Ereignis, das sowohl von außen aber auch gedanklich auftreten kann, gekoppelt [1].

Mit diesen Eigenschaften begleiten und fördern Emotionen ein zielgerichtetes Verhalten, indem versucht wird positive Emotionen zu erreicht und negative Emotionen zu vermeiden [2]. Dadurch wird dem Organismus eine Anpassung an seine Lebensbedingungen ermöglicht. Dies ist besonders durch die physiologische Komponente gegeben, da diese den Körper beispielsweise durch eine erhöhte Herzfrequenz in einer bedrohlichen Situation in Handlungsbereitschaft versetzt [2].

Nachdem die Frage: *Was sind Emotionen?* geklärt ist, schließt sich nun die Frage an: *Wozu sind Emotionen da?*

Emotionen weisen einige verschiedene Funktionen auf [2]: Zum einen haben sie einen handlungsvorbereitende bzw. handlungsleitende Funktion. Das bedeutet durch die Aktivierung des peripheren Nervensystems wird der Körper auf eine Handlung, wie eine Flucht bei Furcht, vorbereitet [3]. Diese Funktion wird besonders aus evolutionsbiologischer Sicht bestärkt, da vor allem die physiologische Komponente eine Anpassung des Verhaltens an die Umgebung und die Situation fördert [3]. Aber auch aus Sicht der kognitiven und subjektiven Komponente wird Verhalten gefördert oder gehemmt. So wird versucht positive Emotionen zu erreichen und negative zu vermeiden, was in der Motivation bestimmter Verhaltenstendenzen resultiert [2].

Emotionen besitzen jedoch auch eine informative Funktion. Hier wird die Person selbst über Veränderungen der Umwelt und bedeutsame Ereignisse informiert. Dazu werden verschiedene Bewertungsvorgänge durchgeführt [1]. Low-Level-Bewertungen sind Prozesse, die bereits auf einer frühen Verarbeitungsstufe auftreten und finden dabei unbewusst und schneller als bewussten Bewertungen statt. Durch diese Bewertung leiten sie eine physiologische Reaktionen ein [2]. Dadurch werden relevante Ereignisse detektiert und die Aufmerksamkeit auf diese gelenkt. Gleichzeitig informieren Emotionen über die Folgen eigener Handlung oder Verhaltensweise [1].

Emotionen besitzen jedoch nicht nur diese intrapersonelle informative Funktion, sondern auch eine interpersonelle kommunikative Funktion [1]. Diese Funktion von Emotionen reguliert zwischenmenschliche Beziehungen und ist somit essentiell für die Interaktion mit anderen Personen [2]. Daher soll diese Funktion im folgenden Kapitel genauer betrachtet werden.

3 Interpersonelle Kommunikation von Emotionen

Die interpersonelle Kommunikation ist ein Informationstransfer, bei dem Information von einem Sender an einen Empfänger übermittelt werden können. Eine Information, welche bei diesem Transfer nicht verbal über mittelt wird, können Emotionen sein. Dabei kann nur die Komponente von Emotionen diese Information kommunizieren, welche auch nach außen hin wahrnehmbar ist [2]. Bei dieser Komponente handelt es sich um die Verhaltenskomponente, welche sich im Emotionsausdruck niederschlägt. Der Sender, also die Person, welche den Emotionsausdruck darstellt, übermittelt durch diesen Ausdruck Informationen über das subjektive Empfinden, über Verhaltensintentionen oder über die Bewertung einer Situation an eine andere Person, den Empfänger der Information [2]. Diese Darstellung des Emotionsausdrucks geschieht dabei teilweise unbewusst, kann von dem Sender allerdings auch bewusst eingesetzt werden [1]. So muss ein Lächeln nicht zwingend eine positive Befindlichkeit anzeigen, sondern kann auch für soziale Zwecke, wie einer Begrüßung oder Beschwichtigung, eingesetzt werden. Um eine Kommunikation zu ermöglichen, ist es von Seiten des Empfängers nun wichtig, dass der Emotionsausdruck richtig erkannt und anschließend in dem gegebenen Kontext korrekt interpretiert wird.

Die beiden Seiten der Kommunikation, also die des Senders und die des Empfängers, sollen hier betrachtet werden.

3.1 Emotionsausdruck

Bereits Darwin [4] ging 1872 von zwei Schritten der Entstehung des Emotionsausdrucks aus. Demnach entstand dieser zunächst, um den Organismus selbst auf aktuelle Stimuli aus der Umgebung vorzubereiten [3]. Darunter zählen beispielsweise die geweiteten Augen, sowie verstärkte Augenbewegungen bei der Empfindung von Angst [4]. Diese Reaktionen entwickelte sich nach Darwin zu Beginn, um das visuelle Feld zu erhöhen und somit potentielle Gefahrenquellen schneller erkennen zu können. Mit der Zeit fungierten diese Emotionsausdrücke dann jedoch nicht nur zur Anpassung des eigenen Organismus auf Veränderungen aus der Umgebung, sondern wurden auch als Hinweisreize von anderen genutzt [3]. Bei dem Ausdruck von Furcht bedeutet dies, dass andere Personen, die ihn beobachten, wissen, dass eine Gefahr vorhanden ist. Der Emotionsausdruck entwickelte sich daraufhin weiter, um sozial wichtige Informationen aktiv kommunizieren zu können [3].

Die Information, die dabei kommuniziert werden soll, kann ganz unterschiedlich sein. Zum einen kann der eigene emotionale und motivationale Zustand übermittelt werden [2].

So würde ein Lächeln eine positives Befinden ausdrücken. Wie bereits erwähnt kann durch den Emotionsausdruck zum anderen allerdings auch die Bewertung einer Situation übermittelt werden. Dadurch können andere den Ausdruck nutzen, um unbekannte Situationen einzuschätzen [2]. Durch einen Blick auf die Bezugsperson kann ein Kind so beispielsweise in Erfahrung bringen, ob ein Gegenstand ein Spielzeug sein kann, wenn die Bezugsperson lächelt, oder ob Gefahr von diesem ausgeht, wenn der Ausdruck Furcht oder Wut zeigt [3, 5]. Mit dem Emotionsausdruck kann jedoch auch eine Verhaltensintention kommuniziert werden [2]. Eine aggressive Handlung kann beispielsweise durch Mimik und Körperhaltung eindeutig übermittelt werden. Zusätzlich ist es möglich, über den Emotionsausdruck anzuzeigen, in welcher Beziehung man zu dem gegenüber steht und ob sich eine Veränderung in dieser Beziehung ergeben hat [2]. Ein Lächeln kann so beispielsweise den Wunsch, eine soziale Beziehung anzustreben, ausdrücken.

Der Kommunikationsinhalt lässt sich grob in zwei Seiten einteilen [5]: Zum einen gibt es den Emotionsausdruck, der spontan auf ein emotionsinduzierendes Ereignis gezeigt wird und damit unwillkürlich entsteht. Zum anderen kann der Emotionsausdruck bewusst kontrolliert werden, um willentlich einen Informationsinhalt zu erstellen. Bei der neuronalen Produktion dieser verschiedenen Ausdrucksarten können diese Unterschiede jedoch nicht so eindeutig gezeigt werden [5]. Die Amygdala, das limbische System und weitere Systeme, die emotionale Informationen verarbeiten, enervieren in die Basalganglien sowie den frontalen Kortex, die unter anderem in der Produktion von Gesichtsausdrücken involviert sind. Die Basalganglien scheinen jedoch verstärkt Einfluss auf die Produktion spontaner Emotionsausdrücke und weniger auf kontrollierte zu besitzen [5].

Dieser Abschnitt zeigt, dass bei der Kommunikation von Emotionen zwar durchaus das subjektive Empfinden, welches eine unwillkürliche Reaktion auslöst, übermittelt werden kann und soll, dass dies jedoch nicht die einzige Information ist, die über einen Emotionsausdruck übermittelt wird. Daher muss auch der bewusst kontrollierte Emotionsausdruck, welcher nicht zwingend mit einer subjektiv erlebten Emotion übereinstimmen muss, betrachtet werden. Wenn man also die interpersonelle Kommunikation betrachtet, scheint die bloße Repräsentation von Emotionen nicht vollkommen ausreichend zu sein, um die Interaktion korrekt darstellen zu können. Andererseits scheint bei der neuronalen Produktion des Emotionsausdrucks die Unterscheidung zwischen den verschiedenen Informationsinhalten nicht so deutlich zu sein, da ähnliche bzw. die gleichen Strukturen bei der Erstellung eines unwillkürlichen sowie eines bewusst kontrollierten Ausdrucks aktiv sind.

3.2 Erkennen von Emotionen

Auf der anderen Seite des Informationstransfers muss die Information zunächst erkannt und anschließend auch verstanden werden. Um den Emotionsausdruck zu erkennen, wird dieser über die Augen und die visuellen Verarbeitungswege an das Gehirn weitergeleitet. Dort wird es im visuellen Kortex vorverarbeitet. Über den ventralen Pfad wird anschließend die Semantik aus diesen Informationen gewonnen [6]. Dabei gibt es bei der Erkennung von Emotionen zwei Verarbeitungspfade [6]: Der eine Pfad erkennt invariante Aspekte von Gesichtern, um dadurch Gesichter von anderen Objekten zu unterscheiden und um einzelne Individuen zu erkennen. Diese Repräsentation des Gesichts kann dann mit semantischem, episodischem Wissen mit dieser Person verbunden werden. Der andere Pfad ist für die Verarbeitung von dynamischen Eigenschaften, wie Augenbewegungen und auch den Gesichtsausdruck zuständig. Dieser Pfad leitet die Informationen unter anderem an die Amygdala und das lymbische System weiter, wo der Ausdruck analysiert wird.

Dies bedeutet für die Repräsentation, dass das Gesicht des Interaktionspartners nicht erkannt und repräsentiert werden muss, sodass der Emotionsausdruck unabhängig von der Identität repräsentiert werden kann.

Häufig werden bei der Kommunikation das Verhalten und im Besonderen die Gesichtsbewegungen des Kommunikationspartners imitiert. Dies wird *Mimikry* genannt und geschieht unbewusst und ohne aktive Motorsteuerung [7]. Das bedeutet, dass die Gesichtsmuskeln aktiviert werden, welche zu dem emotionalen Gesichtsausdruck des Partners passen [7].

Eine mögliche neuronale Erklärung für dieses Verhalten, sind sogenannte *Spiegelneurone*. Dies sind Nervenzellen, die dasselbe Aktivierungsmuster für die aktive Ausführung einer Bewegung sowie für das bloße Beobachten der Durchführung der selben Handlung aufweisen [6]. Das bedeutet, dass diese Neurone eine Verbindung zwischen eigenen Bewegungen und den Bewegungen anderer herstellen könnten. Dadurch könnten diese Bewegungen kopiert werden [8], um so den Gesichtsausdruck des Interaktionspartners zu imitieren.

4 Klassifikation von Emotionen

Um Emotionen zu klassifizieren, gibt es grundsätzlich zwei Ansätze [2]: Der *kategoriale* Ansatz versucht dabei qualitativ verschiedene Emotionen, wie Trauer, Freude, Wut etc. voneinander abzugrenzen. Das bedeutet, es wird versucht Emotionen als distinkte, klar voneinander unterscheidbare Phänomene zu fassen. Eine Annahme, die dabei gemacht wird ist, dass sich komplexere Emotionen aus einfachen Emotionen, sogenannten Basisemotionen zusammensetzten [2]. Der zweite Ansatz zur Klassifikation ist der *dimensionale* Ansatz. Dieser geht davon aus, dass sich Emotionen hinsichtlich ihrer quantitativen Ausprägung auf verschiedenen Dimensionen einordnen lassen.

4.1 Basisemotionen

Die Theorie der Basisemotionen geht bereits auf Charles Darwin [4] zurück, der Emotionen als ein Produkt der Evolution und der natürlichen Selektion beschreibt. Sie ist damit eine Antwort auf grundlegende adaptive Anforderungen der Verhaltenssteuerung und evolutionär alt. Dadurch wird auch in dieser Theorie angenommen, dass diese Basisemotionen angeborene Reaktionsmuster hervorrufen [6]. So werden beispielsweise bei Überraschung die Augen aufgerissen, um das Sichtfeld zu erhöhen und die visuelle Aufmerksamkeit auf eine größere Fläche zu verteilen [4]. Dies bedeutet allerdings zugleich, dass Basisemotionen angeboren sind und nicht erlernt werden müssen [6]. Unterstützt wird diese Annahme davon, dass Kinder und Personen, welche blind und taub auf die Welt kamen, trotzdem Emotionsausdrücke, wie ein Lächeln, zeigen [1]. Da angenommen wird, dass diese Basisemotionen angeboren und evolutionär alt sind, wird der Emotionsausdruck dieser auch kulturübergreifend gezeigt und verstanden [9].

Die Annahme distinkter, klar voneinander trennbarer Emotionen bedeutet, dass jede Emotion auch unabhängig in ihrer physiologischen Ausprägung ist [6, 10]. Das bedeutet, dass jede Emotion durch einen spezifischen, von den anderen unterscheidbaren neuronalen Verarbeitungspfad bestimmt wird [10]. Diese Annahme konnte jedoch bisher noch nicht belegt werden. So unterscheiden sich die physiologischen Messungen einzelner Basisemotionen je nach der Art des Stimulus, der diese Auslöst [10]. Ebenso ist bisher nicht geklärt, wie aus diesen Basisemotionen komplexere Emotionen zusammengesetzt werden [1]. Uneinigkeit besteht auch darin, welche Emotionen als Basisemotionen zählen. Die Literatur reicht hier zumeist von vier bis neun

verschiedenen Emotionen [6]. Die meisten jedoch beinhalten, die von Ekman 1969 [9] identifizierten und in Abb. 1 dargestellten sechs Basisemotionen zu denen Furcht, Freude, Ärger Trauer, Überraschung und Ekel zählen [6].

Abbildung 1: Die Basisemotionen Trauer, Freude, Wut, Angst, Ekel und Überraschung nach Ekman [9] aus [6].

4.2 Gefühlsdimensionen

Bei der Klassifizierung von Emotionen im dimensionalen Ansatz wird jeder Emotion ein Punkt in einem mehrdimensionalen Raum zugewiesen [6]. Dabei werden zwei oder mehrere kontinuierliche Dimensionen genutzt [6]. Das bedeutet für die Analyse, dass die Daten mindestens ein Ordinalskalenniveau aufweisen müssen. Insgesamt herrscht in der Literatur keine Einigkeit darüber, wie viele Dimensionen benötigt werden, um alle Emotionen abzubilden [6]. Allerdings beinhalten die meisten Theorien Maße für Erregung sowie Maße für die Valenz [6, 11]. Die Erregung beschreibt dabei die physiologische und/oder die subjektive Intensität der Emotion. Die Valenz dagegen beschreibt, wie angenehm bzw. positiv oder negativ die Emotion empfunden wird [2]. Dieser Ansatz geht also davon aus, dass es unabhängige Systeme gibt und dass jeder affektive Zustand durch eine lineare Kombination dieser unabhängiger Systeme beschrieben und anschließend als eine bestimmte Emotion interpretiert wird [10].

Innerhalb der dimensonalen Theorien sind das Circumplex-Modell [12] und das Vektormodell die bekanntesten Modelle [6, 11]. Das Circumplex-Modell besteht aus zwei orthogonal ausgerichteten Achsen, die die Valenz und die Erregung beschreiben. Emotionen können demzufolge in einem Kreis um diese Achse angeordnet werden (siehe Abb. 2) [6, 12]. Das Vektormodell besitzt ebenfalls eine unterliegende Dimension für die Erregung [6]. Die Valenz jedoch besteht aus zwei Achsen mit positiver und negativer Valenz, die im 90° Winkel zueinander stehen und sich in einem neutralen Punkt treffen (siehe Abb. 3) [6]. Die Annahme, die hierbei gemacht wird, ist, dass Emotionen mit einer hohen Erregung in positive und negative Valenz unterschieden werden können und dass dementsprechend neutrale Emotionen nicht intensiv, also mit einer hohen Erregung, empfunden werden können [11].

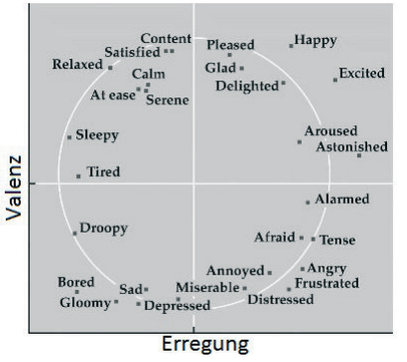


Abbildung 2: Das Circumplex-Modell nach Russell [12] aus [6].

5 Messung von Emotionen

Bei einer funktionierenden interpersonellen Kommunikation verlässt sich der Interaktionspartner bei der Erkennung von Emotionen auf das gezeigte Verhalten des anderen. Wenn nun eine Repräsentation für diese Inhalte erstellt werden soll bedeutet dies, dass

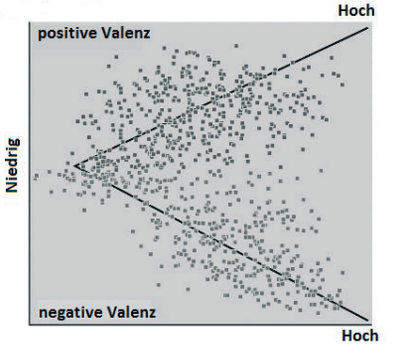


Abbildung 3: Das Vektor-Modell aus [6].

dieses Verhalten gemessen werden muss. Da in der natürlichen Kommunikation die Verhaltensmaße des Interaktionspartners erkannt werden, können diese auch für eine Repräsentation gemessen werden. Das am besten untersuchte Verhaltensmaß ist sicherlich die Mimik [1].

Ein weiteres Maß, das Emotionen objektiv messen kann, sind physiologische Maße [2]. Daher wäre es auch vorstellbar, physiologische Maße wie den Herzschlag, zu messen, um die Kommunikation zu verbessern. Damit könnte ebenfalls eine Repräsentation des emotionalen Zustandes des Interaktionspartners erhalten werden. Daher wird im folgenden nicht nur die Mimik als ein Verhaltensmaß betrachtet, sondern es werden auch physiologische Maße vorgestellt, welche zur Messung während einer Kommunikation eingesetzt werden können.

5.1 Mimik

Die Messung der Mimik kann über zwei Vorgehensweisen erfolgen.

Ein visuelles Maß ist das *facial action coding system* (FACS) [13]. Dabei werden einzelne, visuell erkennbare Muskelbewegungen im Gesicht gemessen und die Kombination dieser Muskelbewegungen erfasst. Diese Muskelbewegungen werden *action units* (AUs) genannt [2]. Der Vorteil dieser Messung ist, dass es sich um eine objektive Messung handelt, bei der lediglich sichtbare Muskelaktivität beschrieben wird, ohne diese hinsichtlich ihres emotionalen Inhalts zu interpretieren. In einem späteren Schritt können die Kombinationen und zeitlichen Veränderungen der AUs anschließend interpretiert werden. Ein Nachteil dieser Methode ist jedoch, dass recht starke Emotionsausdrücke benötigt werden, um die Muskelaktivität detektieren zu können [14]. Diese Maße sind gut erforscht und werden bereits in verschiedenen Anwendungen genutzt, um beispielsweise in der Marktforschung Erkenntnisse über die Einstellungen von Kunden zu erlangen (siehe Garcia-Garcia et.al [15] für ein Review an Technologien).

Die Muskelbewegungen können jedoch auch direkt über beispielsweise die Elektromyographie (EMG) gemessen werden. Hierbei zeichnen im Gesicht platzierte Elektroden die elektrische Aktivität von sich kontrahierenden Muskeln auf [2]. Damit können auch nicht beobachtbare Bewegungen gemessen werden. Ein weiterer Vorteil ist, dass dadurch eine Messung der Stärke der Muskelaktivität möglich ist, wodurch wiederum eine Quantifizierung der Reaktion auf einen Stimulus möglich wird [2]. Ein Nachteil ist allerdings, die Notwendigkeit der Anbringung von Elektroden im Gesicht der Person, welche den Emotionsausdruck darstellt, wodurch eine Verwendung in der natürlichen Interaktion erschwert wird. Im Gegensatz dazu ist eine solche Anbringung an Elektroden oder anderer Messinstrumente an dem Interaktionspartner bei der Messung der AUs nicht nötig, da es sich dabei um ein rein visuelles Maß handelt. Allerdings muss eine visuelle Verbindung zu dem Interaktionspartner bestehen.

5.2 Physiologische Maße

Physiologische Maße können sowohl von dem Zentralen Nervensystem (ZNS) als auch

von dem autonomen Nervensystem (ANS) abgeleitet werden [2]. Messungen des ZNS beinhalten Messungen über die Elektroenzephalografie (EEG), wobei die neuronale Aktivität des Gehirns an der Oberfläche des Schädels gemessen werden, sowie andere bildgebende Verfahren [16]. Bildgebende Verfahren wie die Positronen-Emissions-Tomographie, benötigen jedoch aufwendige Technologien und Auswertungstechniken und sind nicht in natürlichen Szenarien einsetzbar. Messungen über ein EEG sind zwar mit tragbaren Sensoren möglich [16], wodurch ein Einsatz mit einem gewissen Bewegungsfreiraum denkbar wäre. Allerdings sind auch diese auf Grund der komplexen Anbringung, sowie Ungenauigkeiten der Messung, welche durch Artefakte oder Verrutschen verursacht werden können, hauptsächlich für eine Nutzung in experimentellen Umgebungen geeignet [16].
Das ANS besteht aus dem Sympathikus und dem Parasympathikus, welche unabhängig und komplementär sind [6]. Diese regulieren und passen innerkörperliche Vorgänge an, wobei der Sympathikus für eine Leistungserhöhung zuständig ist [6]. Das sympathische System wird bei einer positiven oder negativen Erregung einer Person aktiviert. Dadurch werden der Herzschlag, die Atmung sowie der Blutdruck erhöht [16]. Daher sind bekannte Maße, durch die Emotionen erkannt werden können, Messungen des kardiovaskulären Systems, der Hautleitfähigkeit oder des Atmungssystems [16]. Die Messung des kardiovaskulären Systems ist beispielsweise durch ein Elektrokardiogramm oder durch die Messung des Blutdrucks möglich [2]. Für die Messung dieser Aktivitäten existieren bereits einige Technologien wie Fitnessarmbänder, wodurch dieses Maß für die Verwendung in einer natürlichen Umgebung denkbar wäre. Bei der Messung der Hautleitfähigkeit werden Elektroden an der Handinnenfläche oder dem Finger einer Person befestigt und die Reaktion tritt bereits kurze Zeit (0.4 Sekunden) nach dem emotionsinduzierenden Ereignisses auf [6]. Die physiologische Antwortzeit des Atmungssystems ist jedoch sehr langsam [16], wodurch die Messung dieses nicht für eine Nutzung in einem natürlichen Setting verwendbar ist. Denn hier ist eine Bereitstellung der Information nur sinnvoll, wenn diese in direkte Verbindung mit dem aktuellen Kontext interpretiert werden kann.

6 Diskussion

Für die interpersonelle Kommunikation sind Emotionen wichtig, um nichtverbale Informationen zu übermitteln. Wenn diese Information über ein sensorisches Substitutionssystem übermittelt werden soll, muss diese zunächst gemessen und in entsprechender Weise repräsentiert werden. Betrachtet wurden Repräsentationen von Emotionen im diskreten und dimensionalen Ansatz. Der Vorteil einer Repräsentation durch einen diskreten Ansatz ist, dass diese Theorie bereits auf dem Ausdruck von Emotionen beruht. Die Basisemotionen, welche in diesem Ansatz angenommen werden, sind kulturübergreifend verständlich, wodurch sie für die Kommunikation hilfreich sein könnten. Allerdings muss eine feste Anzahl an Emotionen festgelegt werden, welche repräsentiert werden sollen. Für eine Übertragung an einen Interaktionspartner bedeutet dies, dass für jede dieser Emotionen ein eigenes, klar trennbares Übertragungsmuster erstellt werden muss. Oft werden Emotionen jedoch nicht unbedingt als klar trennbare, isolierte Einheiten gesehen, sondern eher als überlappend und uneindeutig beschrieben [10]. Das bedeutet, man könnte Emotionen eher als ein Spektrum, ähnlich dem der Farben, betrachten. Diese Betrachtungsweise ist durch den dimensionalen Ansatz gegeben, da hier Emotionen durch die Verbindung mehrerer kontinuierlicher Variablen beschrieben werden. Dadurch muss auch keine verbale und konkrete Zuordnung eines emotionalen Zustands zu einer Emotion bestimmt werden. Bei einer Übertragung über ein sensorisches Substitutionssystem würde diese Repräsentation einen gewissen Interpretationsspielraum für den Empfänger lassen, der die Information dann selbstständig in den aktuellen Kontext setzen könnte.
Bei der Messung der Mimik wird die Aktivität von Gesichtsmuskeln gemessen, ebenso wie deren Veränderung über die Zeit. Die Gesamtheit dieser Aktivierungen in Verbindung mit der Veränderung über die Zeit kann anschließend interpretiert werden [2]. Bei der Interpretation der Daten wird in den meisten Anwendungen eine kategoriale Einordnung vorgenommen. Dabei wird eine bestimmte Kombination an Muskelaktivierungen einer bestimmten Emotion zugewiesen [15]. Wie zuvor erwähnt, existieren jedoch einige Probleme bei der Repräsentation des emotionalen Zustands in distinkten Einheiten, die diese Repräsentation als nicht so geeignet erscheinen lassen, wie eine Repräsentation in zwei oder mehreren kontinuierliche Variablen. Es existieren in der Literatur jedoch auch Ansätze, die versuchen, die aus der Mimik gewonnenen Daten in den Dimensionen Valenz und Erregung einzuordnen [17].
Da bei der Kommunikation besonders die Mimik genutzt wird, um emotionale Inhalte zu übermitteln, scheint es auch sinnvoll zu sein, diese visuellen Stimuli zur Messung zu nutzen. Allerdings bieten auch Stimuli, welche nicht visuell gemessen werden, zusätzliche hilfreiche Informationen, um Emotionen zu detektieren. Dies ist zwar bei der Kommunikation von bewusst gesteuerten Emotionsausdrücken nicht möglich, sie helfen jedoch bei der Messung des subjektiven Erlebens, welches unwillkürlich ausgelöst wird. Bei den Methoden zur Messung physiologischer Veränderungen können solche unwillkürlichen Reaktionen des Körpers gemessen werden, welche wiederum eine Aussage über den emotionalen Zustand des Interaktionspartners ermöglichen [2]. Ein Vorteil dieser physiologischen Maße ist, dass sie eine Einteilung in einen dimensonalen Ansatz ermöglichen [15]. So ist beispielsweise die Hautleitfähigkeit ein guter Indikator für den Erregungszustand einer Person [6] und könnte damit genutzt werden um eine Einordnung des emotionalen Zustandes auf der Skala der Erregung vorzunehmen. Andersherum ist eine Bestimmung diskreter Emotionen durch diese Messungen nicht oder nur schwer möglich, da eine Bestimmung unterscheidbarer physiologischer Reaktionsmuster für verschiedene Basisemotionen noch nicht gezeigt werden konnten [10].
Da in dieser Arbeit der Fall betrachtet wird, der eine Beeinträchtigung des visuellen Systems eines oder beider Kommunikationspartner voraussetzt, kann argumentiert werden, dass auch lediglich visuelle Maße als Information genutzt werden sollten. Diese Maße besitzen den Vorteil, dass sie ohne direkten Kontakt bzw. lediglich mit visuellem Kontakt zu dem Interaktionspartner gemessen werden können. Allerdings werden relativ starke Veränderungen der Gesichtsmuskulatur benötigt, um diese zu detektieren. Über dieses Maß können allerdings auch sehr gut willentlich produzierte Emotionsausdrücke erkannt werden. Diese können durch physiologische Maße nicht detektiert werden. Dieses Maß hat dagegen den Vorteil auch Emotionen zu erkennen, wenn der Emotionsausdruck unterdrückt wird und misst somit unwillkürlich ausgelöste Reaktionen des Körpers.
Insgesamt zeigt sich, dass eine Repräsentation von Emotionen in zwei oder mehr kontinuierlichen Dimensionen sinnvoll für eine Übertragung über ein sensorisches Substitutionssystem sinnvoll wäre. Für die Maße eignen sich die Mimik, welche über ein visuelles Maß, wie die AUs, gemessen werden, sowie physiologische Reaktionen des Körpers. Die Verbindung dieser Maße könnte eine interessante Möglichkeit für eine zukünftige Anwendung darstellen.

Literatur

[1] K. Rothermund und A. Eder, *Allgemeine Psychologie: Motivation und Emotion*, 1. Aufl. Prof. Dr. Jürgen Kriz, 2011. DOI: 10.1007/978-3-531-93420-4.

[2] V. Brandstaetter, J. Schüler, R. M. Puca und L. Lozo, *Motivation und Emotion - Allgemeine Psychologie für Bachelor*. Berlin Heidelberg, 2013. DOI: 10.1007/978-3-642-66916-3_6.

[3] A. F. Shariff und J. L. Tracy, "What Are Emotion Expressions For ?", *Current Directions in Psychological Science*, Jg. 20, Nr. 6, S. 395–399, 2011. DOI: 10.1177/0963721411424739.

[4] C. Darwin, *The expression of emotions in man and aminals*. London, 1872, S. 399. DOI: 10.1037/10001-000.

[5] R. J. Blair, "Facial expressions, their communicatory functions and neurocognitive substrates", *Philosophical Transactions of the Royal Society B: Biological Sciences*, Jg. 358, Nr. 1431, S. 561–572, 2003. DOI: 10.1098/rstb.2002.1220.

[6] D. Purves, R. Cabeza, S. A. Huettel, K. S. LaBar, M. L. Platt und M. G. Woldorff, *Principles of Cognitive Neuroscience*, 2. Aufl. Sinauer Associate, 2013, ISBN: 978-0-87893-573-4.

[7] B. Seibt, A. Mühlberger, K. U. Likowski und P. Weyers, "Facial mimicry in its social setting", *Frontiers in Psychology*, Jg. 6, Nr. August, 2015. DOI: 10.3389/fpsyg.2015.01122.

[8] A. F. C. Hamilton, "The neurocognitive mechanisms of imitation", *Current Opinion in Behavioral Sciences*, Jg. 3, S. 63–67, 2015. DOI: 10.1016/j.cobeha.2015.01.011.

[9] P. Ekman, E. R. Sorenson, W. V. Friesen, N. Series und N. Apr, "Pan-Cultural Elements in Facial Displays of Emotion", *Science*, Jg. 164, Nr. 3875, S. 86–88, 1969. DOI: 10.1126/science.164.3875.86.

[10] J. Posnera, J. Russellc und B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology", *Dev Psychopathol*, Jg. 17, Nr. 3, S. 715–734, 2005. DOI: 10.1017/S0954579405050340.

[11] D. C. Rubin und J. M. Talarico, "A Comparison of Dimensional Models of Emotion: Evidence from Emotions, Prototypical Events, Autobiographical Memories, and Words", *NIH Public Access*, Jg. 17, Nr. 8, S. 802–808, 2009. DOI: 10.1080/09658210903130764.A.

[12] J. A. Russell, "A Circumpex Model of Affect", *Journal of Personality ans Social Psychology*, Jg. 39, Nr. 6, S. 1161–1178, 1980. DOI: 10.1037/h0077714.

[13] Paul Ekman Group, *Facial Action Coding System*, https://www.paulekman.com/product-category/facs/ Accessed: 10.09.2018.

[14] K. Wolf, "Measuring facial expression of emotion", *Dialogues in Clinical Neuroscience*, Jg. 17, Nr. 4, S. 457–462, 2015, ISSN: 12948322.

[15] J. M. Garcia-Garcia, V. M. R. Penichet und M. D. Lozano, "Emotion detection: a technology review", *Proceedings of the XVIII International Conference on Human Computer Interaction - Interacción '17*, Nr. October, S. 1–8, 2017. DOI: 10.1145/3123818.3123852.

[16] S. Jerritta, M. Murugappan, R. Nagarajan und K. Wan, "Physiological Signals Based Human Emotion Recognition : A Review", in *IEEE 7th International Colloquium on Signal Processing and its Applications*, 2011. DOI: 10.1109/CSPA.2011.5759912.

[17] R. Niese, A. Al-Hamadi, M. Heuer, B. Michaelis und B. J. Matuszewski, "Machine Vision based Recognition of Emotions using the Circumplex Model of Affect", *2011 International Conference on Multimedia Technology*, 2011. DOI: 10.1109/ICMT.2011.6001887.