

■ WHITEPAPER

Future Challenges in Logistics and Supply Chain Management

KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

■ WHITEPAPER

KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

Als Teil der Social Networked Industry in der Logistik zielt das „Intelligente Regal“ auf die Anwendung und die Evaluierung neuer, auch indirekter Kommunikationstechniken zwischen Mensch und Maschine und die damit verbundenen Auswirkungen auf die Arbeitsorganisation. In diesem Whitepaper werden mehrere Möglichkeiten der Kommunikation zwischen Menschen und Maschinen vorgestellt und evaluiert.

Future Challenges in Logistics and Supply Chain Management

Die Schriftenreihe »Future Challenges in Logistics and Supply Chain Management« greift aktuelle Herausforderungen auf, beleuchtet Trends und fokussiert neuartige Technologien und Geschäftsmodelle.

Die verschiedenen Ausgaben der Schriftenreihe zeichnen das Zukunftsbild einer innovativen Branche, das von Forschung und Praxis gestaltet und gelebt wird.

WHITEPAPER

KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

Autoren

Arkadius Schier, Fraunhofer IML
Benedikt Mättig, Fraunhofer IML
Lukas Nikelowski, Fraunhofer IML
Christian Hoppe, Fraunhofer IML
Sebastian Roeder, Fraunhofer IML
Jens Leveling, Fraunhofer IML
Dirk Höning, Fraunhofer IML
Timo Erler, Fraunhofer IML

Internet

Das Whitepaper steht Ihnen auch im Internet unter www.innovationslabor-logistik.de zur Verfügung.

DOI

10.24406/IML-N-642964

Herausgeber

Prof. Dr. Dr. h. c. Michael ten Hompel
Prof. Dr. Michael Henke
Prof. Dr.-Ing. Uwe Clausen

Kontakt

Fraunhofer-Institut für Materialfluss
und Logistik IML

Joseph-von-Fraunhofer-Str. 2–4
44227 Dortmund

schriftenreihe@iml.fraunhofer.de

Innovationslabor – Hybride Dienstleistungen in der Logistik

Das Innovationslabor – Hybride Dienstleistungen in der Logistik stärkt die digitale Vorreiterrolle des Standorts Dortmund und fördert eine nachhaltige Weiterentwicklung, indem die Zukunftsfragen der Logistik und der Informationslogistik unter dem Aspekt der Mensch-Technik-Interaktion adressiert und deren Forschungsstand am Standort reflektiert werden.

Dazu entwickeln Wissenschaft und Wirtschaft, bestehend aus Logistikern und Soziologen, gemeinsam technologische Innovationen für eine sozial vernetzte Industrie, die Social Networked Industry. Die Ausgestaltung der Innovationen zu hybriden Dienstleistungen wird durch so genannte Showcases in fünf Bereichen gewährleistet: Handel, Produktionslogistik, Transport, Instandhaltung und Virtual Training.

Das Innovationslabor – Hybride Dienstleistungen in der Logistik ist ein vom Bundesministerium für Bildung und Forschung gefördertes Leuchtturmprojekt. Das interdisziplinäre Forschungsprojekt ist ein Vorhaben des Fraunhofer-Instituts für Materialfluss und Logistik IML in Dortmund, des Fraunhofer-Instituts für Entwurfstechnik Mechatronik IEM in Paderborn sowie der Technischen Universität Dortmund mit der Fakultät Maschinenbau und dem Forschungsgebiet Industrie- und Arbeitsforschung. Darüber hinaus sind zahlreiche Netzwerkpartner in die Forschungsarbeit eingebunden.

■ WHITEPAPER

INHALT

Einführung	1
Konzepte zur Kommunikation zwischen Mitarbeitern und Maschinen	2
Intelligentes Regal	2
Einleitung	2
Interaktionsmodule.	2
SmartRack-WMS-Modul	5
Anwesenheitserkennung	6
Personenidentifikation	11
Natürlich-sprachliche Interaktion	16
Mika	21
Konzept	21
Prototypische Umsetzung	23
Pick-by-Laser	25
Konzept	25
Prototypische Umsetzung	27
Validierung	29
Literaturverzeichnis.	33

Einführung

Die bereits in den vorangegangenen Whitepapers erarbeiteten Ideen basieren auf dynamischen Daten und individuellen Informationen von Mitarbeitern und Maschinen. Um qualitativ hochwertige Daten zu bekommen, wurden neue (auch indirekte) Kommunikationstechniken zwischen Mensch und Maschine auf Basis von Smart Devices getestet. Über berührungslose Sensoren können beispielsweise die physische und psychische Belastung einzelner Mitarbeiter gemessen und ausgewertet werden. Die gewonnenen Daten können als direkter Input für die weitere Planung der Arbeitsorganisation genutzt werden.

Zur Veranschaulichung des Nutzens wurde im Anwendungszentrum eine Regalgasse zur Kommissionierung aufgebaut. Dabei bildet ein Intelligentes Regal die Schnittstelle zwischen virtueller Welt und dem Mitarbeiter. Durch zusätzliche Systeme wie einem Pick-by-Laser-Modul wird die direkte Kommunikation zwischen Kommissionierer und Regalsystem ermöglicht und erlaubt eine individuelle Anpassung an den jeweiligen Mitarbeiter. Vitalwerte der Mitarbeiter werden durch das Intelligente Regal und passende Sensoren gemessen und ausgewertet. Zusammen mit weiteren Daten aus dem internen Social Network können so direkt Maßnahmen (z. B. eine vorzeitige Pause) abgeleitet werden.

Konzepte zur Kommunikation zwischen Mitarbeitern und Maschinen

Intelligentes Regal

EINLEITUNG

Das Intelligente Regal bietet eine interaktive Schnittstelle zwischen Mensch und Warenmanagement. Diverse Sensorik, kameraunterstützte Software und Zugriff zum WMS-System ermöglichen einen angenehmen und schnelleren Kommissionierprozess. Die Mitarbeitererkennung erlaubt es, Kommissionieraufträge individuell zusammen zu stellen und Aufträge Mitarbeitern zuzuordnen. Mithilfe der Spracherkennung ist es sogar möglich bei Problemen bei der Kommissionierung Handlungsoptionen zu geben. So sollen Mitarbeiter entlastet und Arbeitsabläufe im Warenlager optimiert werden. Die Abläufe am Regal können Aufschluss darüber geben wie viele Produkte sich an einer Lagerposition befinden und so Inventur und Dokumentation des Bestands unterstützen. Außerdem könnte durch eine Auswertung der benötigten Hilfestellungen oder Interaktionszeiten ermittelt werden, wie Arbeitsabläufe weiter verbessert und Mitarbeiter bei der Kommissionierung unterstützt werden können.

INTERAKTIONSMODULE

Die entwickelte Hardware des Interaktionsmoduls besteht aus der Anforderung heraus auch komplexere Aufgaben wie Bildverarbeitung/WLAN Kommunikation und Tonaufnahmen zu bewältigen aus einer Kombination von Microcontroller (MCU) und Microprocessor (MPU). An diesen Prozessor sind die notwendigen Peripherieeinheiten angebunden wie Kamera, Mikrofone, Sensorik, WLAN, Speicher, Display, Power-Management und Bedienelemente.

Zur Umsetzung der Anforderungen an das Interaktionsmodul wurden die nachfolgenden Komponenten ausgewählt.

Für die MCU/MPU Einheit wurde ein Prozessor von STMicroelectronics eingesetzt, der STM32MP153 mit dual ARM Cortex®-A7 und Cortex®-M4 Cores welcher die beiden

WHITEPAPER KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

Einheiten MCU/MPU in einem Chip vereint und es ermöglicht ein Linux basiertes Betriebssystem auf den Chip laufen zu lassen um alle höheren Anforderungen wie z. B. WLAN/Bildverarbeitung bewältigen zu können. Dem Prozessor steht 1GB Arbeitsspeicher zur Verfügung und das Betriebssystem wird von einer SD-Karte geladen. Über ein Spezielles Konfigurationssystem lässt sich beim Booten festlegen welche Schnittstellen der MPU oder der MCU zugewiesen werden. An die MPU ist so der Speicher, das WLAN/Bluetooth-Modul, das PowerManagement (PMIC), die 5MP Kamera (OV5640 Chip; siehe Abbildung 2), die 4 MEMS Mikrofone (siehe Abbildung 3) in den Ecken des Labels und die SD-Karte (64GB) angebunden. Die MCU verwaltet die Sensorik (VL53L1 Time-of-Flight Entfernungssensor; siehe Abbildung 4), das Display (2.9" EPD Display monochrom 296x128 Pixel; siehe Abbildung 1) und die Bedienelemente (4 Tasten).



Abbildung 1:
Display des
Intelligenten Regals

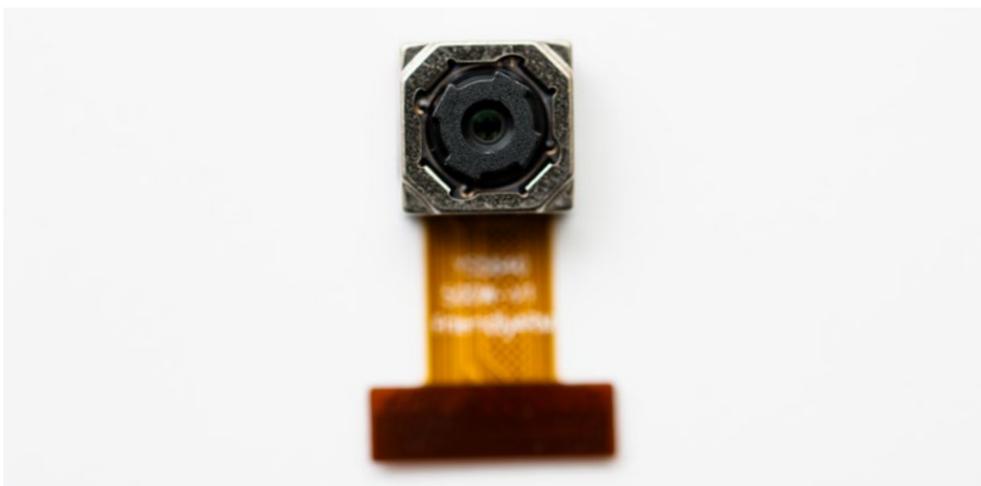


Abbildung 2:
Kamera des
Intelligenten Regals



Abbildung 3:
Mikrofon

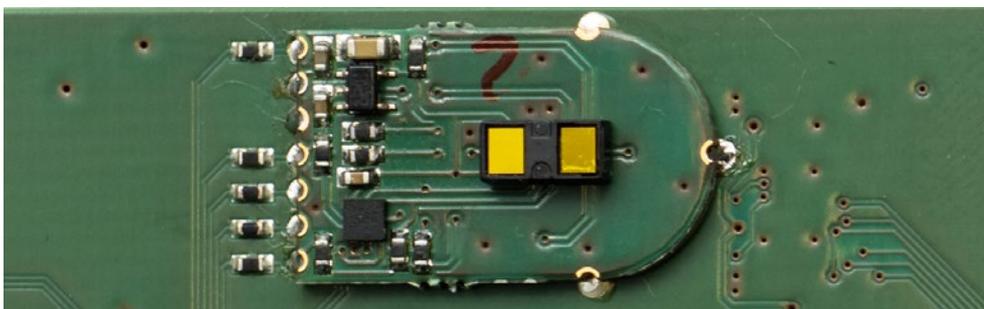


Abbildung 4:
ToF Sensor

Das Label verfügt über einen eingebautem Lithium-Polymer Akku mit 1200mAh der es ermöglicht auch über einen längeren Zeitraum unabhängig von der Schiene (siehe Abbildung 5) zu agieren. Durch die eingebauten LiPo-Charger und Fuel-Gauge IC's weiß das Betriebssystem wie lange es noch ohne externe Spannungsversorgung laufen kann und wann es wieder über die Schiene geladen werden muss.



Abbildung 5:
Energieversorgungs-
schiene

Dank des eingebautem WiFi Modems hat das Label immer eine Kommunikationsmöglichkeit mit externen Servern oder zu anderen Devices. Das Modem verfügt über eine Kombination von WiFi und Bluetooth im 2.4 Ghz Bereich. Durch das integrierte Muster auf der Schiene und den entsprechenden Abnehmer auf der Rückseite des Labels ist es nicht nur möglich das Label zu laden, sondern auch festzustellen ob auf der Schiene seine Position verändert bzw. ob es abgenommen/aufgesetzt wurde. Dank

WHITEPAPER KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

des eingebauten ToF-Sensors kann das Label die meiste Zeit im Standby bleiben und wird nur aufgeweckt, wenn sich eine Person im Ortungsbereich des Sensors befindet. Beginnt ein Benutzer mit der Interaktion können die vier Mikrofone zur Ortung und als Spracheingabe fungieren und zusätzlich der Benutzer über die Kamera identifiziert werden.

Als Betriebssystem fungiert ein maßgeschneidertes Embedded Linux auf Basis von "Yocto". Dieses stellt Schnittstellen zur Kommunikation zwischen der Firmware auf dem Microcontroller (M4) und der Module Controller Software (A7 CPUs) bereit und sorgt für einen geregelten Ablauf dieser Komponenten. Ebenso wird die Kommunikation mit dem Leitstand per WiFi ermöglicht. Zur Erleichterung der Entwicklung und Wartung der Module, sowie zur temporären Vorhaltung größerer Daten (Bilder, Audio) kommen für das System austauschbare SD-Karten als Speichermedium zum Einsatz.

SMARTRACK-WMS-MODUL

Als zentrale Kontrolleinheit dient der sogenannte Leitstand. Bei diesem kann es sich um ein eigenständiges System oder um ein Add-On zu einem bestehenden WMS-System handeln. Der Leitstand kennt alle Teilsysteme und kann jedem dieser Systeme Nachrichten senden oder welche von ihnen empfangen. Dazu stellt er entsprechende Schnittstellen bereit, zum Beispiel über das MQTT-Protokoll. Gleichzeitig kann er auf das WMS-System zugreifen, und ermöglicht so die Kommunikation zwischen diesem und dem Intelligenten Regal.

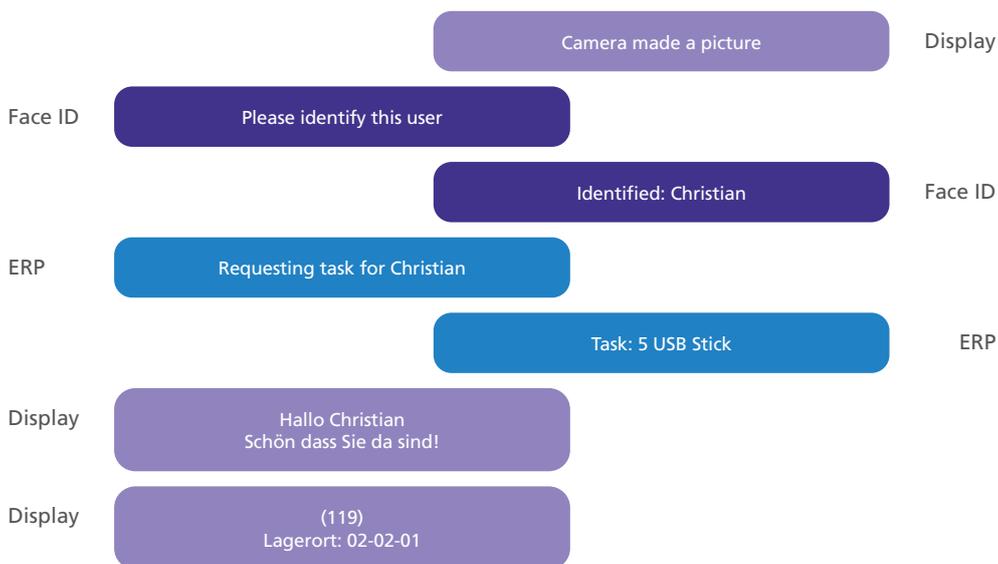


Abbildung 6:
Chat-Verlauf

Die Hauptaufgabe des Leitstandes besteht darin, Nachrichten von Teilsystemen entgegenzunehmen und relevante Informationen an das richtige Ziel weiterzuleiten. Für diesen Zweck existiert ein Verhaltensmodell für jedes Teilsystem, insbesondere für jedes Regal-Modul, welches den aktuellen Zustand beschreibt und die passende Reaktion auf eingehende Nachrichten vorgibt. Dabei werden Nutzer und ihre Aufgaben aus dem WMS geladen, und nach erfolgten Interaktionen die entsprechenden Daten im WMS aktualisiert.

Zusätzlich erlaubt der Leitstand eine Administration des Intelligenten Regals, um beispielsweise neue Teilsysteme zu integrieren oder bestehende Teilsysteme manuell zu kontrollieren, und es können erweiterte Funktionen wie das Anlernen neuer Nutzer (siehe Personenidentifikation) ausgeführt werden.

Abgerundet wird der Leitstand durch eine Weboberfläche, in der die Kommunikation zwischen Teilsystemen in einer für Menschen verständlichen Form, einem „Chat“ wie aus Messenger-Diensten bekannt, dargestellt wird.

ANWESENHEITSERKENNUNG

Bei vielen gleichzeitig unter Volllast betriebenen Modulen besteht die Gefahr, die Energiekapazität der Versorgungsschiene zu überschreiten. Selbst die zum Ausgleich von Spitzenlasten in den Modulen verbauten Akkus können ein solches Energiedefizit bei fehlender Ladekapazität nicht langfristig kompensieren. Darum werden die Module in Phasen, in denen keine Nutzerinteraktion stattfindet, in einen tiefen Energiesparmodus versetzt, während dessen sie lediglich eine energiesparende Anwesenheitserkennung durchführen.

Die Anwesenheitserkennung fungiert als Impulsgeber zur kontrollierten Wieder-Inbetriebnahme des Moduls, sobald eine Person vor dieses tritt. Der ausgelöste Aufwachprozess verläuft dabei in mehreren Phasen, in denen sich die genauen Umstände, in denen ein Modul wieder vollständig zur Verfügung steht, stufenweise präzisieren und gezielt die für die jeweilige Phase benötigte Hard- und Software zugeschaltet wird. Auf diese Weise soll eine potenzielle Interaktionsbereitschaft erkannt und den Erfassungsbereich des Moduls passierende Personen weitestgehend ignoriert werden.

Das Konzept, der Entwurf und die Realisierung der energiesparenden Anwesenheitserkennung durch ein mehrstufiges Aufwachsystem entstanden im Rahmen einer Bachelorarbeit auf einem Modulprototypen. Im Fokus stand dabei insbesondere die Auswahl geeigneter COTS-Sensorik sowie Abwägungen zwischen Energieeinsparung und Aufwachlatenz. Letztere ist vor dem Hintergrund der Nutzerakzeptanz von großer

Bedeutung, da die Anwesenheitserkennung die erste MMI-Schnittstelle zum Endnutzer ist und als Schnittstellenkomponente die vom Leitstand koordinierte Bereitstellung der digitalen Dienstleistungen anstößt.

Konzept

Die Anwesenheitserkennung untergliedert sich in drei Phasen - der Ersterfassung, der Verweilkontrolle und der Personenerkennung. Aufgrund der Projektanforderungen verläuft jede Phase vollständig kontaktlos zum Anwender und ohne dass dieser Hilfsggeräte am Körper trägt.



Abbildung 7:
Ablauf der drei
Phasen der Anwesen-
heitserkennung

Die Ersterfassung dient der Wahrnehmung eines Objektes im kegelförmigen Erfassungsbereich vor dem Modul, während sich das Modul in einem tiefen Energiesparmodus befindet. Dabei ist es vorerst unerheblich, ob es sich bei dem detektierten Objekt um eine Person oder einen Gegenstand handelt, sodass sich die Detektion auf eine binäre Entscheidung reduziert. Auf diese Weise lässt sich die Ersterfassung energiesparend anhand von einfacher COTS-Sensorik realisieren.

Sobald die Ersterfassung anschlägt, wird in der zweiten Phase eine Verweilkontrolle vorgenommen. Diese prüft, ob das durch die Ersterfassung detektierte Objekt über den Zeitraum einer Mindestverweildauer vor dem Modul verbleibt. Dabei werden weniger Hardwareressourcen als die in der Ersterfassung verwendete Sensorik genutzt, sodass der Energieverbrauch nur gering ansteigt.

Die durch die Verweilkontrolle verursachte Verzögerung im Aufwachprozess des Moduls ist aus mehreren Gründen beabsichtigt. Zunächst soll die Verweilkontrolle verhindern, dass ein Modul bei jeder den Erfassungsbereich passierenden Person vollständig aufwacht, wodurch unnötig Energie verbraucht wird und Latenzen bei der Umstellung auf eine neue Person entstehen. Wenn eine Person hingegen mit einem Modul interagieren möchte, hat diese für den Zeitraum der Mindestverweildauer die Möglichkeit, sich dem Modul zuzuwenden. Dadurch verringert sich das Risiko einer Fehlentscheidung in der letzten Phase der Anwesenheitskontrolle, der Personenerkennung.

Die Personenerkennung stellt schließlich fest, ob es sich bei dem durch die Ersterfassung und die Verweilskontrolle detektierten Objekt um einen Menschen handelt. Die Module steuern dabei die dazu benötigte Datengrundlage bei, während die eigentliche Personenerkennung ausgelagert ist und in der Koordinationsverantwortung des Leitstandes liegt. Dieser trifft schließlich auch die Entscheidung über die weiteren Schritte nach der Anwesenheitserkennung und gibt den Modulen entsprechende Folgeanweisungen.

Sensorikauswahl

Zur technischen Umsetzung der Ersterfassung und Verweilskontrolle wurde günstige COTS-Sensorik ausgewählt, die sich energiesparend betreiben lässt, auf die Einsatzumgebung abgestimmt ist und das Potential für weitere Optimierungen und Anpassungen im Praxiseinsatz erlaubt.

Dazu wurde eine Abwägung zwischen den verwendeten Messtechniken, dem Detektionsmedium sowie dem Funktionsumfang und speziellen Eigenschaften der zur Auswahl stehenden Sensormodelle getroffen. Der schließlich gewählte Time-of-Flight (ToF) Sensor VL53L1X ist ein aktiver IR Laser Sensor der Laser-Klasse 1, der anhand von Triangulation begrenzbarere Entfernungsmessungen auf bis zu 4 Meter auf einem einstellbaren Sichtfeld von maximal 27 Grad ermöglicht und über einen breiten Konfigurationsumfang verfügt [1].

Infrarot hat gegenüber Ultraschall und Radiowellen den entscheidenden Vorteil, dass es nicht anfällig gegenüber Umgebungsgeräuschen ist und in Umgebungen mit Metall - wie etwa Metallregalen in Lagern - nicht zusätzlich durch Interferenz gestört wird. Die Messbeeinträchtigungen des VL53L1X durch direkte Lichteinstrahlung werden durch eine Schutzlinse über dem Sensor vermindert und anhand von Filtertechniken durch den Mikrocontroller des Sensors weiter minimiert.

Da es sich beim VL53L1X um einen aktiven IR Sensor handelt - der also selbst IR-Strahlung emittiert - sollte jedoch bedacht werden, dass sich die VL53L1X Sensoren zu nah einander gegenüber angebrachter Module gegenseitig beeinflussen können, was zu Messabweichungen führt. Bei ausreichendem Regalabstand lassen sich die Module jedoch frei platzieren.

Der breite Konfigurationsumfang des VL53L1X erlaubt es hierbei, die Ersterfassung und Verweilskontrolle durch Anpassung unterschiedlicher Parameter auf die jeweilige Einsatzumgebung abzustimmen. Abhängig vom Abstand nebeneinander angebrachter Module lässt sich zum Beispiel das Sichtfeld des Sensors oder die Messdistanz, in denen Aufwachmeldungen erfolgen, sowie die Messrate anpassen. Aufwachmeldungen signalisiert der VL53L1X nach außen hin durch ein abgreifbares Spannungssignal,

das abhängig von der Konfiguration nach jeder abgeschlossenen Distanzmessung oder bei Entfernungsmesswerten erfolgt, die in einem einstellbaren Messwertfenster liegen.

Wahl der Energiesparmodi für CM4 und CA7

Die richtige Wahl des Energiesparmodus für die Module entscheidet, wie viel Energie auf Kosten der Aufwachlatenz eingespart wird. Es muss das Ziel sein, die Energieeinsparung zu maximieren, ohne dabei den Endnutzer durch eine übermäßig lange Aufwachlatenz zu behindern. Insbesondere das Wecken des CA7 und die damit verbundene Wiederinbetriebnahme des hiesigen Linux-Kernels ist zeitaufwendig. Zur schnellen Betriebsfortführung wird der Zustand des Betriebssystems nach dem Suspend-to-RAM Prinzip während des Energiesparbetriebs im RAM abgelegt, der zum Datenerhalt automatisch refreshed wird [2].

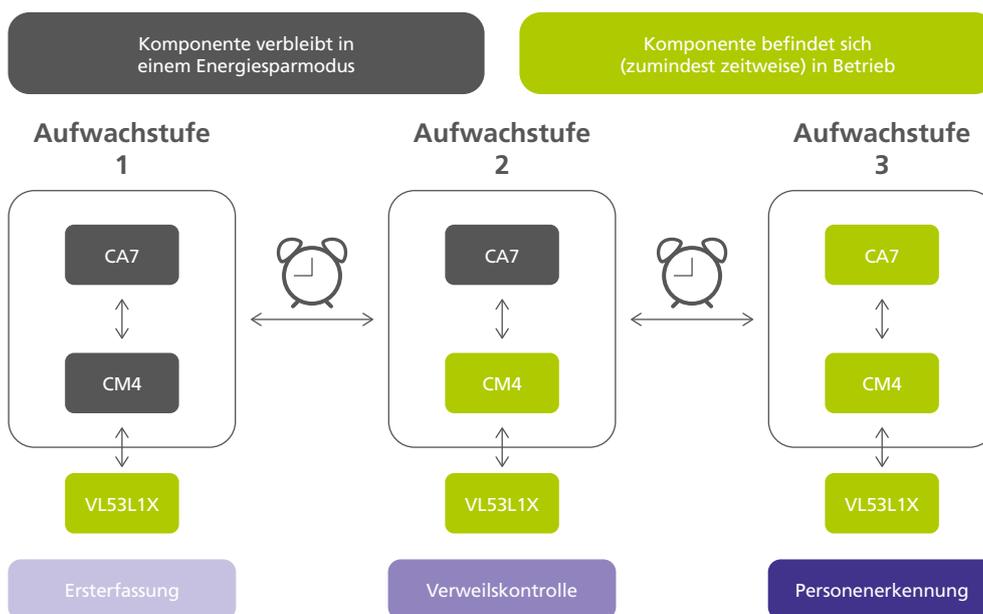


Abbildung 7: Betriebsmodi der Hardwarekomponenten in der jeweiligen Phase der Anwesenheitserkennung

Um weitere Aufwachlatenzen zu vermeiden, werden zudem nur die Clocks von CA7 und CM4 und ausgewählte Hardwareperipherie ausgeschaltet, sodass sich die Aufwachlatenz des CA7 auf rund eine Sekunde beschränkt. Da es sich beim CM4 lediglich um einen einfachen Mikrocontroller handelt und dessen Inbetriebnahme darum weit weniger aufwendig ist, fällt dessen Aufwachlatenz nicht weiter ins Gewicht. Wichtig ist hingegen, dass im Energiesparmodus der Ersterfassung die Hardwareeinheiten zum Wecken des CM4 durch ein Spannungssignal des VL53L1X weiterhin aktiv bleiben.

Umsetzung

Der VL53L1X wird durch den CM4 kontrolliert, welcher diesen in Vorbereitung auf die Phase der Ersterfassung für den Betrieb im Threshold Detection Mode konfiguriert. Dabei wird auf die vom Hersteller bereitgestellte API sowie die für alle auf dem CM4 umgesetzten Implementierungen grundlegenden Hardware Abstraction Layer (HAL) zurückgegriffen. Während der Ersterfassung befindet sich das Modul und damit auch der CM4 in einem Energiesparmodus, nur der VL53L1X führt kontinuierlich im Abstand von jeweils einer halben Sekunde Distanzmessungen durch. Der VL53L1X überprüft jeden Distanzmesswert, ob dieser im eingestellten Schwellwertfenster von 5cm bis 125cm liegt und damit der CM4 über ein Spannungssignal auf einer der Interrupt Lines geweckt werden muss.

Mit dem Eintritt in die Verweilskontrolle konfiguriert der CM4 den VL53L1X um, sodass dieser dem CM4 nun jede abgeschlossene Messung meldet. Damit lässt sich der CM4 zwischen Messungen wieder in seinen Energiesparmodus versetzen und bei jedem neuen Messwert wecken. Das Prüfen, ob der Messwert des VL53L1X im Schwellwertfenster liegt, erfolgt in der Phase der Verweilskontrolle auf dem CM4. Liegen drei aufeinanderfolgende Messwerte außerhalb des Schwellwertfensters, bereitet der CM4 den VL53L1X wieder für die Ersterfassung vor und legt sich daraufhin schlafen. Liegen drei aufeinanderfolgende Messwerte hingegen innerhalb des Schwellwertfenster, weckt der CM4 den CA7. Dazu ist die virtuelle UART Schnittstelle vorgesehen, welche auf Seite des CM4 durch das OpenAMP Framework und auf CA7 Seite durch das RPMMsg Framework realisiert und konfiguriert wurde [3][4]. Dieser für die Inter-Prozessor-Kommunikation vorgesehene Mechanismus dient so zusätzlich als Weckmechanismus für den CA7, der durch das Senden einer Nachricht ausgelöst wird.

Die Aufwachnachricht des CM4 an den CA7 beinhaltet die Aufwachursache, damit der CA7 angemessen auf diese reagieren kann. Nach der Wiederinbetriebnahme des CA7 und Auslesen der Aufwachnachricht veranlasst der CA7 mithilfe des Video4Linux 2 (V4L2) Frameworks die Aufnahme eines 720px x 576px Pixel großen Bildes [5]. Dieses wird in das PGM-Format umgewandelt und als Base64-encoded String drahtlos per MQTT zur Personenerkennung an den Leitstand übermittelt. Damit ist die Anwesenheitskontrolle abgeschlossen und das Modul wartet auf Folgeanweisungen des Leitstandes.

Auswertung

Die Auswertung des Stromverbrauchs der Modulprototypen zu den verschiedenen Phasen der Anwesenheitserkennung ergab eine Energieeinsparung von 80% zwischen Normalbetrieb und dem Betrieb in der Ersterfassung. Durch weitere Optimierungsmaßnahmen,

wie eine Drosselung der Clocks oder eine genauere Kontrolle der Betriebszustände spezifischer Hardwarekomponenten über den Power Management Integrated Circuit (PMIC), besteht jedoch weiteres Energiesparpotential. Die Aufwachlatenz beläuft sich auf drei bis vier Sekunden.

PERSONENIDENTIFIKATION

Konzeptionierung

Um die Nutzung des Moduls weiter für Mitarbeiter zu erleichtern, wurde eine bildbasierte Nutzererkennung entwickelt. Die Hauptmotivation für diese Gesichtserkennung stellt dabei ihre Personalisierungskomponente dar: Sie bildet den Übergang zum Nutzerprofil eines Mitarbeiters und leitet nach Erkennung die Aktivierung der Folge-Systeme wie des Kommissionierassistent und des Dialogassistent ein. Beide greifen auf den identifizierten Namen oder Kürzel zu. Die Gesichtserkennung ist also eng an die anderen Teilsysteme angebunden und bildet dadurch eine wichtige Schnittstelle. Durch ihre Login-Funktion stellt die Gesichtserkennung außerdem einen sekundären Sicherheitsmechanismus dar und verhindert den Zugriff durch unbefugte Personen.

In der Anwendung lässt sich das Problem der Gesichtsidentifizierung als "one-to-many mapping" definieren, also im Falle des Intelligenten Regals als Abgleich einer Momentaufnahme der Person vor dem Modul mit einer Reihe von hinterlegten Daten der authentifizierten Nutzer. Zur Lösung dieser Aufgabenstellung wurden Machine Learning-basierte und mathematische Verfahren eingesetzt, als auch eine eigene Datenbank angelegt. Konkreter wurde das Verfahren der Nutzererkennung in einen mehrschrittigen Prozess aufgeteilt, der aus Gesichtserkennung (der Detektion eines Gesichts im Bild), dem Erstellen von Merkmalsvektoren und schließlich der eigentlichen Identifikation besteht. Die Gesichtsidentifikation an sich wurde dabei als ein Klassifikationsproblem definiert.

Flexible Unterteilung der Sub-Komponenten

Die Unterteilung der Gesichtserkennung in diese einzelnen Komponenten erlaubt für eine gewisse Unabhängigkeit der einzelnen Teilprozesse und bietet damit eine Flexibilität in der Anwendung. Diese Flexibilität wurde zu Nutze gemacht und ein zusätzlicher Anwendungsfall für das Intelligente Regal in Form eines Ad-hoc Anlernens neuer Personen implementiert. D.h. soll ein neuer Mitarbeiter eingearbeitet und für die Nutzung des Intelligenten Regals autorisiert werden, so kann er oder sie direkt über das Interaktionsmodul zur Datenbank der Gesichtserkennung hinzugefügt werden. Dabei wird der Lernprozess der Gesichtsidentifikation automatisch eingeleitet und steht nach wenigen Sekunden auf allen Modulen zur Verfügung.

Um ein Zusammengreifen der einzelnen System-Komponenten und auch die schnelle Bereitstellung eines neu-gelernten Identifikationsmodells auf den Modulen zu ermöglichen, ist eine effiziente Kommunikation notwendig. Hier spielt der Leitstand eine zentrale Rolle, über welchen ein ständiger Austausch stattfindet und der Informationsfluss geregelt wird. In einem der ersten Schritte werden z. B. die Fotoaufnahmen der Anwesenheitserkennung durch den Leitstand an die Gesichtsdetektion weitergeleitet. Genauso werden dann nach der Verarbeitung der Bilder die Ergebnisse der Personenidentifikation auch wieder über denselben Kommunikationsweg den anderen Systemen mitgeteilt, welche daraufhin aktiviert werden oder andere Prozesse einleiten.

Auswahl der Detektions- und Klassifikationsverfahren und Umsetzung des Identifikationsprozesses

Für die Gesichtsdetektion wurde auf ein bestehendes Verfahren mit einem sogenannten Multi-task Cascaded Neural Network (MTCNN) gesetzt. Dabei handelt es sich um drei hintereinander geschaltete Convolutional Neural Networks (CNNs), die jeweils auf einzelne Aufgaben trainiert sind. Als Ergebnis erhält man von dem Netzwerk sogenannte Bounding Boxen, also Rahmen, um die erkannten Gesichter, sowie eine Hervorhebung markanter Punkte im Gesicht wie Augen oder Mundwinkel [6].

Da es bei der initialen Gesichtsdetektion auch der Fall sein kann, dass mehrere Gesichter erkannt werden, z. B. von Personen im Hintergrund, wird nur das größte erkannte Gesicht weiterverwendet. Es wird der Annahme gefolgt, dass der zu erkennende Mitarbeiter direkt vor dem Smart Rack-Modul steht und dadurch sein oder ihr Gesicht den meisten Raum im Bild einnimmt. So wird gewährleistet, dass die richtige Person authentifiziert wird, selbst wenn im Hintergrund eine andere Person auftauchen sollte, was im Betrieb durchaus zu erwarten ist.

Das Training neuronaler Netze benötigt große, annotierte Datensätze und viele Bilder einer Klasse, um verschiedene Klassen gut voneinander unterscheiden und neue Bilder korrekt klassifizieren zu können. Um mit möglichst wenigen Bildern von Mitarbeitern arbeiten zu können, werden sogenannte Face Embeddings und ein vortrainiertes Modell zum Erzeugen dieser eingesetzt. Face Embeddings sind mehrdimensionale Merkmalsvektoren, die prägnante Eigenschaften der auf einem Bild abgebildeten Person einfangen und numerisch repräsentieren. Die Embeddings erlauben es die Eigenschaften verschiedener Personen in einem Vektorraum darzustellen, zu vergleichen und Vektoren der gleichen Person zu gruppieren, bzw. unterschiedlicher Personen zu separieren. Für die Identifikation eines Mitarbeiters kann also ein einfacher Abgleich zu den eingepflegten Merkmalsvektoren vorgenommen werden. Die Transformationsfunktion, die aus Bildern diese konzentrierte Repräsentation erzeugt ist ein weiteres neuronales Netz: FaceNet [7].

Anstatt das FaceNet von Grund auf zu trainieren, wird ähnlich wie bei der Gesichtsdetektion auf ein vortrainiertes Modell zurückgegriffen. So wird sichergestellt, dass auch für einen kleinen Datensatz sinnvolle Embeddings erstellt und relevante Merkmale der Personen eingefangen werden können. Die Parameter des FaceNets wurden auf dem MS Celeb Datensatz, welcher aus tausenden Bildern berühmter Personen besteht, angepasst, so ist das Modell speziell für die Extrahierung markanter Gesichtsmerkmale sensibilisiert und generiert aussagekräftige Embeddings [8].

Der Abgleich findet danach u. a. durch eine Berechnung der Kosinus-Ähnlichkeit statt. Dabei wird die Ähnlichkeit des Anfragebildes zu allen anderen vorhandenen Bildern bekannter Nutzer berechnet. Der Nutzer mit den meisten Treffern und somit der größten Ähnlichkeit wird als Klasse angenommen. Es ist dabei zu beachten, dass die Embeddings nie eine komplette Übereinstimmung erzielen werden, da tagesabhängig Lichtverhältnisse, Distanz zur Kamera und Aspekte wie Frisur oder Make-up immer zu leicht veränderten Merkmalsrepräsentationen führen werden. Zur Stärkung der Prognose wird daher mit einem Threshold gearbeitet, sodass ein Nutzer signifikant häufig erkannt werden muss. Kann kein Nutzer eindeutig bestimmt werden, wird dieses Ergebnis entsprechend an den Leitstand durch "User unbekannt" zurückgegeben.

Als zusätzliche Absicherung für das Ergebnis wird neben dem Kosinus-Ähnlichkeit auch eine Support Vector Klassifikation (SVC) eingesetzt. Dabei werden die einzelnen Nutzer wieder als Klassen angesehen und zwischen allen möglichen Klassenpaaren eine lineare Support Vector Machine (SVM) erzeugt. Das heißt, auch an dieser Stelle werden pro Nutzer die gespeicherten Face Embeddings betrachtet und die Trennlinie durch die SVMs innerhalb des aufgespannten Datenraumes gezogen.

Da die SVC immer eine Klasse ausgibt, kann sie allerdings nicht erkennen, wenn das neue Bild von einer unbekannt Person ist. Selbst in diesem Fall ist sich das SVC-Modell trotz einer Fehlklassifikation sehr sicher über seine Entscheidung.

Aus diesem Grund werden die Ergebnisse der Kosinus-Ähnlichkeit bevorzugt und nur für den Fall, dass beide Ansätze tatsächlich unterschiedliche bekannte Nutzer ausgeben und sich die SVC sehr sicher ist, auf das Ergebnis dieser zurückgegriffen.

WHITEPAPER

KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

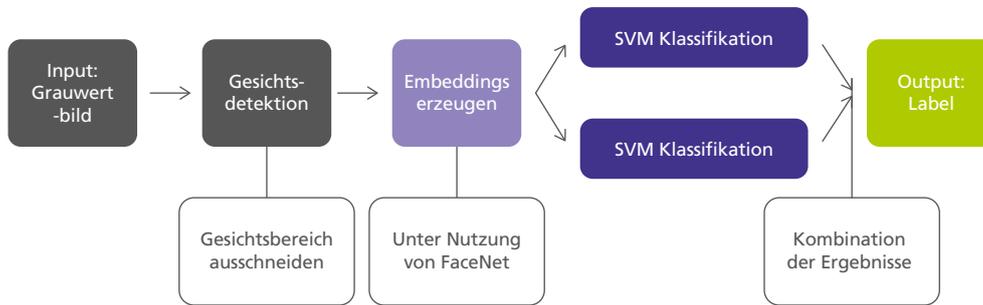


Abbildung 9: Ablauf des Identifikationsprozesses vom Grauwertbild zur Klassifikation

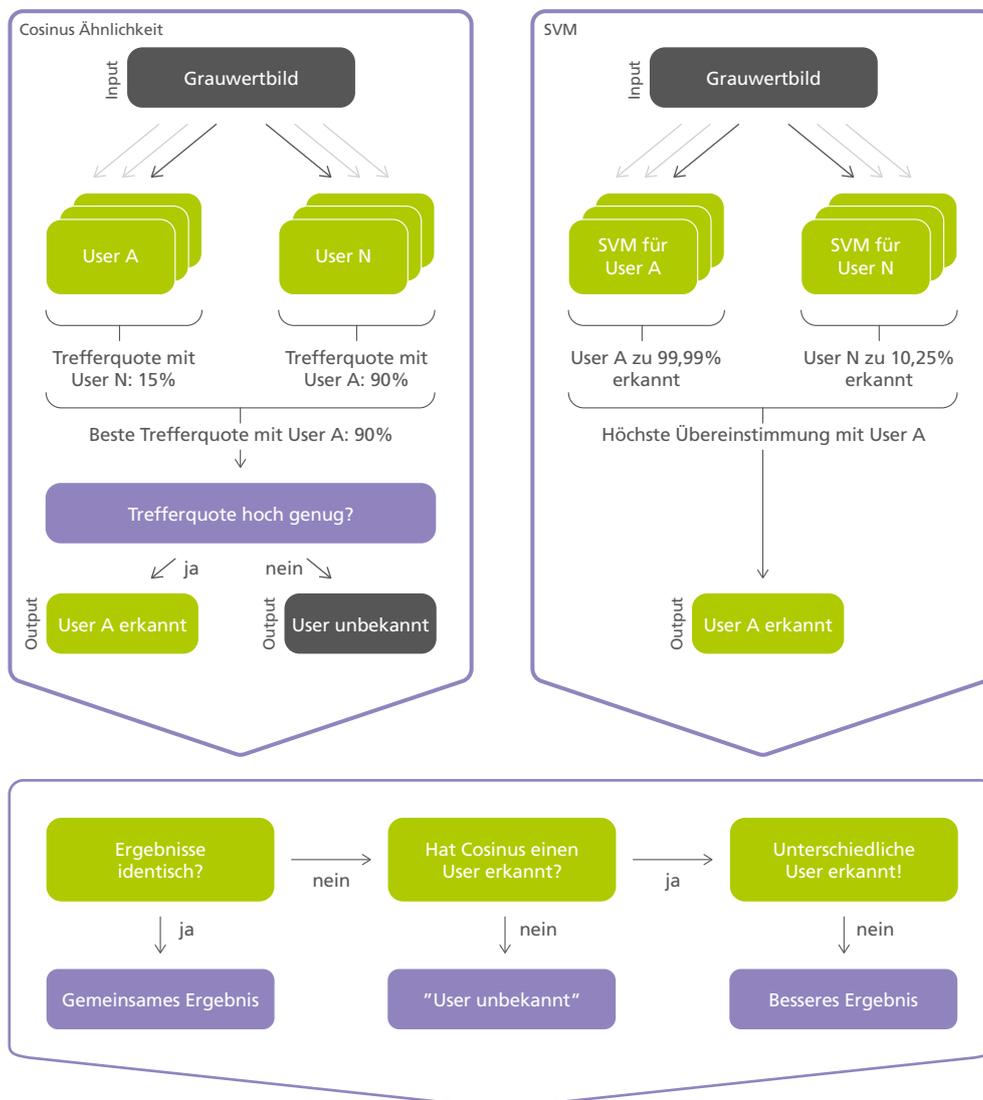


Abbildung 10: Klassifikationsprozess mit den beiden Modellen. Links: Klassifikation über Cosinus-Ähnlichkeit, Rechts: Klassifikation über SVMs, Unten: Zusammenfügen der Ergebnisse

Umsetzung des Anlernprozesses

Wie eben dargestellt nutzen wir zwei Ansätze zur Identifikation einer Mitarbeiterin bzw. eines Mitarbeiters. Daher müssen beide Klassifikatoren angeleitet. Die Datengrundlage

für das Training sind die bereits bei der Identifikation genannten Face Embeddings, die aus den detektierten Gesichtern der Nutzer erzeugt werden.

Da für die Klassifikation über die Kosinus-Ähnlichkeit das Anfragebild ad hoc mit den abgespeicherten Bildern der authentifizierten Nutzer abgeglichen wird, benötigt dieser Ansatz kein tatsächliches Training. Allerdings sollten aus Performanzgründen die Face Embeddings pro Nutzer vorberechnet und abgespeichert werden. Das Training des Support Vector Modells besteht aus dem Training der einzelnen SVMs, die alle Nutzerklassen paarweise trennen. Für das Training müssen nur die Face Embeddings und entsprechende Nutzernamen (als Klassenlabels) übergeben werden.

Ad-Hoc Anlernen neuer Personen

Das Anlernen neuer Personen funktioniert analog zum grade beschriebenen Anlernprozess. Von der neuen Person werden nach Auslösen des Anlernprozesses an der Administrationsoberfläche eine gewisse Anzahl an Bildern aufgenommen und auf jedem Bild nach Gesichtern gesucht.

Sind genügend Gesichter in den Bildern erkannt worden, werden aus diesen Face Embeddings generiert. Sowohl die Gesichter als auch die Embeddings werden dann in der Datenbank mit dem entsprechenden Nutzernamen abgespeichert. Für die SVC muss jetzt nur noch das Modell mit allen Daten neu trainiert werden. Dazu laden wir alle Embeddings aus der Datenbank mit den entsprechenden Klassenlabels (das sind die Nutzernamen) und geben sie zusammen mit den Daten des neuen Nutzers ins Training. Das resultierende Modell wird abgespeichert und bei späteren Identifikations-Anfragen einfach nur abgerufen.

In diesem Prozess ist es sowohl möglich im System neue Nutzer anzulegen als auch die Daten bestehender Personen zu aktualisieren. In den Testdurchläufen hat unser System auf Basis der Face Embeddings sehr zuverlässig funktioniert und hat die Nutzer richtig identifizieren können. Aber dennoch, ist es denkbar, dass durch äußerliche Veränderungen der Nutzer, wie z. B. eine Brille, ein Bart oder deutlich andere Frisur, die neuen Embeddings zu unterschiedlich zu den hinterlegten werden und sich das Ergebnis der Abgleiche verschlechtert. Mit diesem Hintergedanken wurde die Option zur Aktualisierung der Daten eingebaut.

NATÜRLICH-SPRACHLICHE INTERAKTION

Eine Maschine, in diesem Fall das Intelligente Regal, erwartet seine Eingaben in Form von Bits und Bytes. Softwaresysteme kapseln diese Form der Eingabe durch standardisierte Notationen (z. B. JSON). Ferner kapselt das Softwaresystem diese Kommunikation durch Bildschirmmasken, die für den Nutzer leicht verständlich sind, um dem Menschen zu ermöglichen seine Arbeit ausführen zu können, ohne komplizierte Abfragesprachen oder ähnliches zu lernen. Diese Form der Interaktion eignet sich nicht im industriellen Umfeld eines Lagers bei Kommissionier Tätigkeiten.

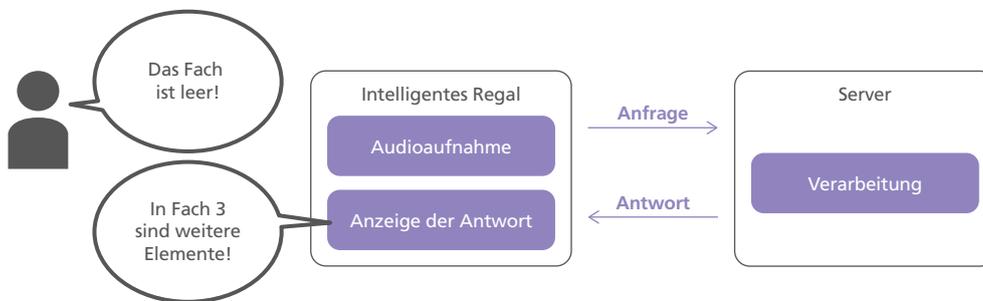


Abbildung 11: Informationsfluss bei der natürlich-sprachlichen Interaktion mit dem Intelligenten Regal.

Die natürlich-sprachliche Kommunikation bietet einen menschenzentrierten Ansatz für diese Herausforderung. Sie ermöglicht dem Menschen die Interaktion mit einer Maschine in seiner gewohnten Kommunikationsart.

Der Mensch kann während der Kommissionier Tätigkeit Anfragen an das Regal stellen, um z. B. Probleme zu melden, weitere Informationen zu bekommen oder Handlungen zu bestätigen. Im Folgenden ist eine Liste mit geplanten Anliegen und jeweils einem Beispiel:

- **Problem:** Das angegebene Fach ist leer.
- **Weitere Informationen:** In welchem Fach finde ich weitere Bauteile dieser Art?
- **Bestätigung:** Ich habe die Schrauben entnommen.

Das Intelligente Regal verarbeitet die Anfrage mit Hilfe des Leitstands und gibt seine Antwort auf dem Display in textueller Form aus. Der industrielle Kontext erschwert eine Ausgabe durch Audio. Der Kommissionierer könnte so Informationen der Antwort falsch, nur in Teilen oder gar nicht aufnehmen. Eine kompakte visuelle Aufbereitung der Antworten garantiert, dass der Kommissionierer alle notwendigen Informationen zum aktuellen Handlungsstrang bekommt.

KONZEPT DER VERARBEITUNG

Die Verarbeitung der Anfrage erfolgt in sechs Schritten. In einem ersten Schritt wird auf dem Modul die Anfrage über Mikrofone aufgenommen. Diese aufgenommene Anfrage wird an den Leitstand übermittelt. Der Leitstand beinhaltet vier Module, die die Anfrage verarbeiten. Die Module sind über MQTT als Publish-/Subscribe-Methode mit dem Leitstand verbunden und können so flexibel ausgetauscht werden.

Zunächst wird im Modul **Sprache-Zu-Text** die Audio-Anfrage in Text überführt, um sie verarbeiten zu können. Die gewählte Sprache zur Spracherkennung ist Deutsch.

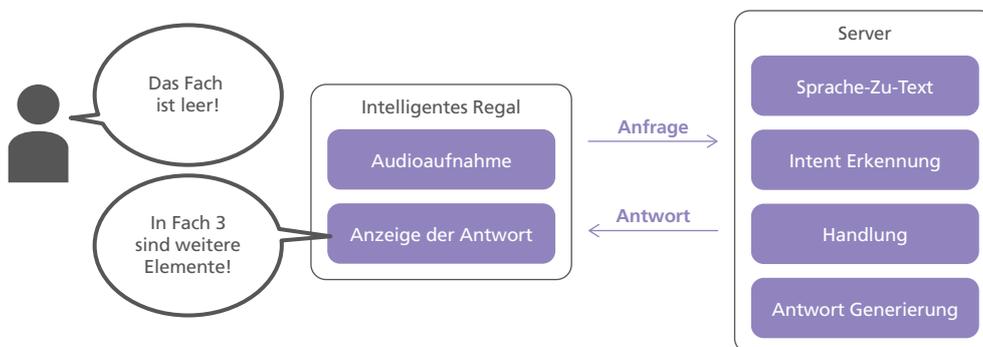


Abbildung 12: Verarbeitung einer Sprachanfrage in sechs Schritten.

In der Intent Erkennung wird die textuelle Anfrage mit einem Modell abgeglichen und einem Anliegen zugeordnet. Das Modell beinhaltet Muster-Datensätze, die die drei Intentionen (Problem, Weitere Informationen, Bestätigung) mit Beispielen untermauern. Für den Intent "Problem" werden noch einmal Untergliederungen vorgenommen. Hier können die Ausprägungen "Fach ist leer", "Inhalt ist fehlerhaft" oder "Falscher Inhalt" unterschieden werden.

Der erkannte Intent wird vom Leitstand interpretiert und eine **Handlung** ausgelöst. In diesem Schritt kommuniziert der Leitstand mit anderen Systemen oder folgt eigenen Logiken. So können z. B. in einem Fehlerfall, weitere Fächer abgefragt werden, die die gleichen Inhalte bieten.

Im Anschluss wird im Modul "**Antwort Generierung**" eine Antwort erstellt, die an das Intelligente Regal zurückgeschickt wird. Die Antwort kann losgelöst von der Handlung (z. B. Entnahme bestätigt) oder abhängig von der Handlung sein (z. B. Weitere Teile in Fach 3). Am Regal wird die Antwort in textueller Form ausgegeben.

Konzept für die lernende Intent Erkennung

Die Intent Erkennung vollzieht einen Abgleich mit einem Modell und klassifiziert die Anfrage in Form von Intents. Die Klassifikation beinhaltet den Intent und eine Wahrscheinlichkeit, mit welcher die Anfrage diesem Intent zugeordnet werden kann. Je umfangreicher das Modell ist, desto genauer können Intents klassifiziert werden. Um vor diesem Hintergrund das Modell stetig zu optimieren, beinhaltet das Modul "Intent Erkennung" eine Komponente, die es ermöglicht, weitere Beispiele während der Nutzung aufzunehmen. In Abbildung 13 ist der Ablauf exemplarisch beschrieben.

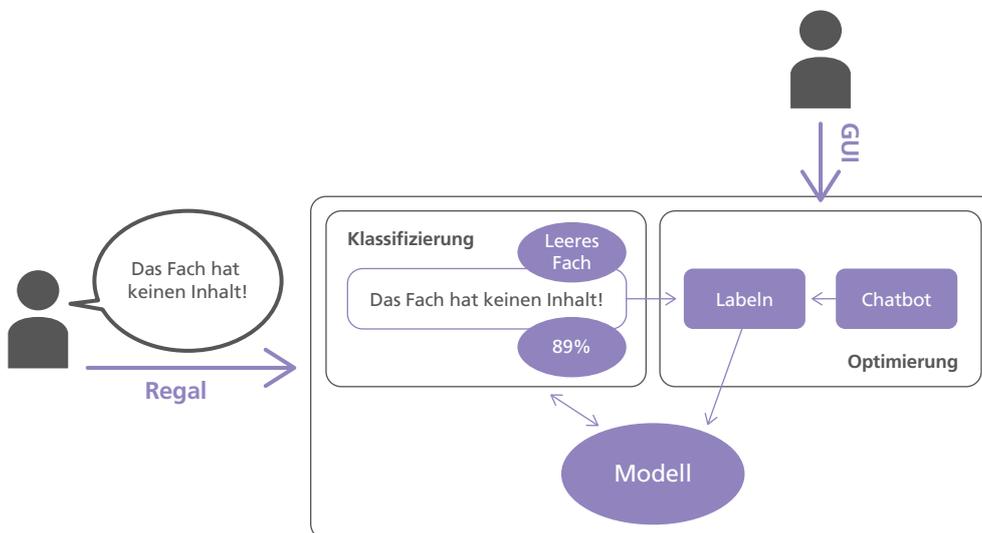


Abbildung 13:
Optimierung des
Sprachmodells durch
Labeln unsicherer
Intenterekenntnisse
über einen Anwen-
der.

Erkennt das Modul eine Anfrage, die nicht 1:1 im trainierten Modell existiert (keine Wahrscheinlichkeit von 100%), merkt es diese Anfrage mit zugehörigem Intent und der erzielten Wahrscheinlichkeit vor. Im aufgeführten Beispiel ist es die Anfrage "Das Fach hat keinen Inhalt". Dieser Satz ist im Modell nicht vorhanden. Die Intent Erkennung gibt aus, dass es sich bei dieser Anfrage zu 89% um das Problem "Leeres Fach" handelt. In einem weiteren Schritt können diese Anfragen durch einen Anwender/ Entwickler analysiert und ins Modell überführt werden. Dieser Schritt wird als "Labeln" bezeichnet. Die Komponente verfügt über eine grafische Oberfläche, sodass die Interaktion für den Anwender geführt vollzogen werden kann.

Neben dem laufenden Betrieb können Anwender und Entwickler über einen Chatbot textuell mit der Komponente kommunizieren, um Inhalte für das Modell zu generieren. Der Ablauf für die Optimierung folgt analog dem zuvor beschriebenen Ansatz.

Umsetzung im Intelligenten Regal

Für die Umsetzung wurden vier der sechs beschriebenen Schritte durch Eigenentwicklungen vorgenommen. Auf dem Modul des Intelligenten Regals wurde die Audioaufnahme und Anzeige der Antwort direkt implementiert. Das Modul ist über MQTT mit dem Leitstand verbunden, der die Koordination übernimmt. Im Leitstand sind die Module "Handlung" und "Antwort Generierung" zusammengefasst implementiert, dass diese für den Anwendungsfall eine direkte Abhängigkeit aufweisen. Für die Module "Sprache-Zu-Text" und "Intent Erkennung" wurden jeweils Open Source Software verwendet.

In der Spracherkennung wurde auf das Framework "DeepSpeech" von Mozilla zurückgegriffen [9]. Mozilla verfolgt hier den Ansatz über neuronale Netze mit Tensorflow. DeepSpeech ist mit dem öffentlichem Projekt Common Voice verbunden, in dem jeder bei der Weiterentwicklung der Sprachmodelle mitwirken kann, in dem er Texte einspricht oder Audioaufnahmen mit Texten abgleicht. Für die Auswahl von DeepSpeech sprachen drei wesentliche Kriterien. Es ist Open Source. Für das intelligente Regal wurde sich auf Open Source festgelegt. Kommerzielle Lösungen sind somit ausgeschlossen. Das zweite Kriterium ist der Privacy-Aspekt. Die Lösungen sollen intern laufen und nicht extern an Server weitergeleitet werden. Dieser Aspekt schließt Lösungen, wie Google, aus. Der letzte Punkt umfasst den Stand der Technik und Performanz. Aus diesem Grund wurden ältere Ansätze, wie PocketSphinx, nicht gewählt. Zu dem Punkt Performanz gehört auch, dass das Intelligente Regal die Sprache Deutsch voraussetzt. Für DeepSpeech gibt es aus der Community ein umfangreiches Modell für die deutsche Sprache. Dieser Ansatz umfasst unterschiedliche existierende Audio-Modelle und ergänzt das übersetzte englische Modell von Mozilla. Dieses Modell wurde für die Umsetzung genutzt. Der Anwendungsfall konnte mit Hilfe dieses Modells nahezu vollständig umgesetzt werden [10, 11].

In der Intent Erkennung setzt das Intelligente Regal auf die Chatbot-Software "rasa" [12]. Rasa besteht aus zwei wesentlichen Teilen. Rasa Open Source ist der eigentliche Chatbot, der die Klassifizierung und die Möglichkeiten zur Modellerstellung bietet. Rasa X ist eine grafische Serveranwendung, die die Funktionen für die Optimierung des Modells im laufenden Betrieb sicherstellt. Die Auswahlkriterien sind analog zur Spracherkennung angesetzt worden. Ansätze von Google oder für den privaten Sektor freigegebene Lösungen konnten nicht genutzt werden. Rasa erfüllte alle Kriterien. Rasa setzt ebenfalls auf neuronale Netze mit tensorflow und bietet umfangreiche Möglichkeiten zum Training und effizienten Klassifizierung. Insbesondere Rasa X bietet hier eine sinnvolle Unterstützung für den aufwendigen Prozess des Trainierens. Rasa ist modular aufgebaut und bietet einige Möglichkeiten zur Erweiterung. Die Module Handlung, Antwort

Generierung und sogar die Koordination kann über den Chatbot übernommen werden. Dieser modulare Ansatz ermöglichte eine einfache Integration in den Leitstand [13, 14].

Ausblick

Rasa X bietet eine Möglichkeit, das Modell für die Intent Erkennung stetig zu optimieren. Ein zukünftiges Ziel könnte sein, diesen Ansatz auf DeepSpeech zu übertragen und den aufwändigen Prozess des Trainings zu vereinfachen und zu optimieren.

Das Intelligente Regal berücksichtigt aktuell noch keinen Umgang mit fehlerhaften Interpretationen. Diese können in den Modulen Sprache-Zu-Text und Intent Erkennung auftreten. An dieser Stelle muss in Zukunft eine Interaktion mit dem Nutzer angestrebt werden, um Eingaben zu korrigieren und fehlerhafte Handlungen zu reduzieren.

Mika

KONZEPT

Die Idee fest installierter bzw. stationärer Assistenzsysteme für die Kommissionierung steht im Kontrast zu mobilen Systemen, wie sie in der Logistik mehrheitlich eingesetzt werden. Während mobile Systeme am Körper (Pick-by-Voice oder Pick-by-Vision) oder in der Hand (Pick-by-Scan über ein mobiles Datenerfassungsgerät (MDE)) getragen werden, lassen sich fest installierte Systeme unabhängig vom jeweiligen Mitarbeiter als ubiquitäres System einsetzen. Sie sind meist auf einen speziellen Anwendungsfall ausgerichtet und erfordern eine Installation in der Infrastruktur des Lagers. Als Beispiel seien hier Pick-by-Light-Systeme genannt, die aufwändig installiert werden müssen und in diesem Zuge entsprechende Investitionen mit sich bringen.

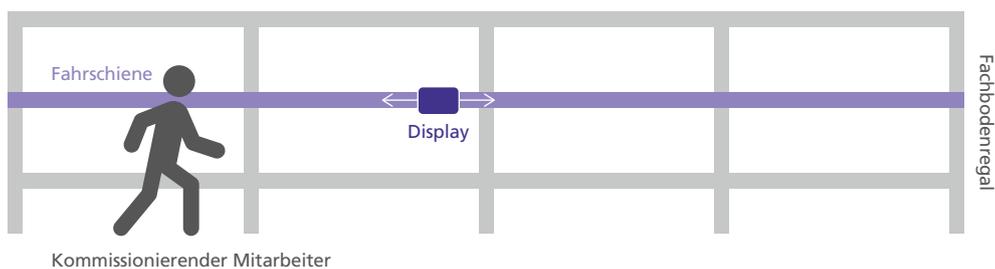


Abbildung 14:
Aufbau des konzeptionierten Assistenzsystems

Im Innovationslabor bestand nun die Zielstellung, ein Assistenzsystem zu konzeptionieren, das die Vorteile eines mobilen mit denen eines stationären Systems verbindet. Konkret bedeutet dies, dass ein System gesucht wurde, das den Mitarbeiter dynamisch im Kommissionierprozess unterstützt, geringe Investitionskosten erfordert und nicht vom Mitarbeiter getragen werden muss. Der Ursprung der hier vorgestellten Idee ist die Annahme, dass Fachbodenregale, wie sie in der Kommissionierung in der Regel zum Einsatz kommen, einen festen Aufbau besitzen, der in Höhe und Länge begrenzt ist. Der Mitarbeiter benötigt im Prozess Informationen zu den zu pickenden Artikeln, deren Menge und den Ort, an dem die Artikel entnommen werden sollen. Während die Darstellung des Artikels sowie der Menge über ein handelsübliches Display erfolgen kann, ergibt sich aus der Anzeige des Pickortes eine Herausforderung. Dem Mitarbeiter muss intuitiv und klar übermittelt werden, wo sich der zu entnehmende Artikel befindet. Stationäre Displays können diese Anforderung nicht ergonomisch erfüllen. Entweder muss sich der Mitarbeiter die Informationen merken oder es müssen mehrere Displays im Regal verteilt werden. Beides ist aus ergonomischen und investitionstechnischen Gründen nicht zielführend. Auf dieser Grundlage baut die Idee auf, ein Anzeigemedium zu

WHITEPAPER KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

konstruieren, das sich horizontal entlang der Regalzeile bewegen kann, um den Mitarbeiter exakt zu dem jeweiligen Entnahmeort zu führen.

Die Idee des mobilen, interaktiven Kommissionierassistenten (mika) beschreibt ein System, das aus verfahrbaren Displayeinheiten (Shuttles) besteht, die den Mitarbeiter durch den Kommissionierprozess – innerhalb einer Regalzeile – führen. Ein Shuttle besteht aus einem Display sowie einer Antriebseinheit, die das Verfahren auf einer Linearführung ermöglicht. Die Linearführung besteht aus einer Fahrschiene sowie einer integrierten Stromschiene, über die das Shuttle mit Energie versorgt wird (siehe Abbildung 14).

Die Shuttles sind so gestaltet und in den Prozess eingebunden, dass jeweils ein Mitarbeiter mit einem Shuttle verknüpft ist. Der Mitarbeiter meldet sich per Gesichtserkennung am Shuttle an und erhält daraufhin eine Anzeige der offenen Kommissionieraufträge. Sobald ein Auftrag gestartet wurde, fährt das Shuttle zum nächsten Entnahmeort. Eine exakte Anzeige des Artikels, des Regalfachs sowie der zu entnehmenden Anzahl wird über das Display dargestellt. Die Bestätigung der Entnahme geschieht durch den Scan der Artikel über einen Handschuhscanner, der sich zuvor automatisch mit dem Shuttle verbunden hat. Muss die Regalzeile gewechselt werden, erfolgt eine automatische Abmeldung des Mitarbeiters mit dem Hinweis, welche Regalzeile als nächstes angesteuert werden soll. In der nächsten Regalzeile erfolgt wiederum eine Anmeldung des Mitarbeiters an einem dort befindlichen Shuttle.

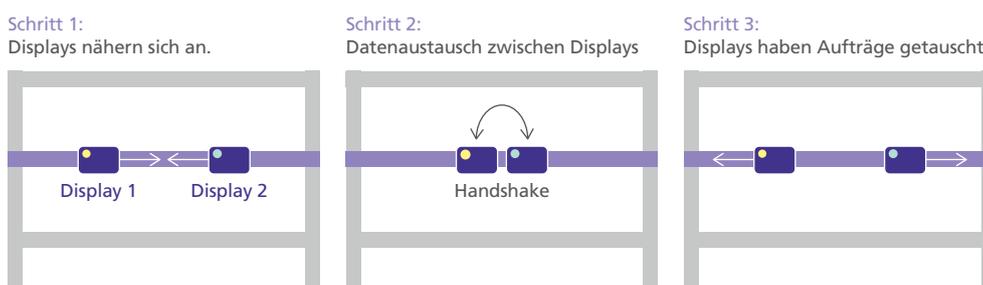


Abbildung 15:
Konzept des Tauschs
von Aufträgen bei
kreuzenden Bewe-
gungen mehrerer
Displays

Ein Teil des beschriebenen Konzeptes umfasst die Möglichkeit, mehrere Shuttles zur gleichen Zeit in einer Regalzeile einzusetzen. Dies wird durch eine Kommunikation der Shuttles untereinander umgesetzt. Kreuzen sich die Wege von Shuttles, wird dies mit der Übergabe des Auftrages – also dem Tausch der Aufträge zwischen den beteiligten Shuttles – umgesetzt. Um dem Mitarbeiter diesen Prozess zu veranschaulichen, erhält jeder aktive Auftrag eine eigene Farbe, die von dem Shuttle dargestellt wird. Hierdurch soll dem Mitarbeiter die Zuordnung zu einem Shuttle verdeutlicht werden (siehe Abbildung 15).

WHITEPAPER KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

Um die Akzeptanz der betroffenen Mitarbeiter zu fördern, wurde die Benutzerschnittstelle über das Display so gestaltet, dass jedes Shuttle an einen menschlichen, persönlichen Assistenten angelehnt ist. Der Mitarbeiter wird von einem interaktiven Avatar begrüßt und über den gesamten Prozess hinweg von diesem Avatar begleitet. Basis dieses Vorgehens bilden Studien, die belegen, dass menschlich anmutende Avatare die Angst der Mitarbeiter vor technischen Systemen reduzieren können.

PROTOTYPISCHE UMSETZUNG

Für die Umsetzung des Konzeptes und unter Berücksichtigung eines industriellen Einsatzes wurde die Konstruktion des Assistenzsystems basierend auf Standardkomponenten durchgeführt. Dies bezieht sich sowohl auf die Fahrschienen als auch auf die Stromversorgung und das Shuttle inkl. Display und Antrieb.



Abbildung 16:
Erster Prototyp des
mika-Systems

Für den prototypischen Aufbau der Linearführung sowie des Antriebes wurden Gleitlager verwendet, die den stabilen Gleichlauf des Shuttles gewährleisten sollen. Der Antrieb erfolgt über ein separates Antriebsrad, das über einen 24V Gleichstrommotor angetrieben wird. Die Stromversorgung wird über handelsübliche Stromschienen gewährleistet, die gegen den versehentlichen Kontakt des Mitarbeiters geschützt sind. Die Steuerung und Orchestrierung aller Komponenten erfolgt über ein Tablet auf Android-Basis, das über WLAN mit anderen Shuttles sowie dem Gesamtsystem (z. B. WMS) kommuniziert. Damit sich das Shuttle auf der Schiene orientieren kann, werden in fest definierten Abständen NFC-Tags auf die Fahrschiene aufgebracht, die über ein intern verbautes NFC-Lesegerät erfasst werden.

Die angezeigten Inhalte des Displays wurden auf grundsätzlichen ergonomischen Grundsätzen aufgebaut und beschränken sich auf die minimal mögliche Informationsdichte im jeweiligen Kontext. Es werden zu jedem Zeitpunkt immer nur die Informationen dargestellt, die der Mitarbeiter im entsprechenden Prozessschritt benötigt. Darüber hinaus findet der mika-Avatar durchgehend Verwendung in der Visualisierung, um den menschlichen Charakter des Assistenzsystems zu unterstreichen und die Akzeptanz zu steigern. Abbildung 16 zeigt den Aufbau eines ersten Prototyps mit einem aktiven Kommissionierauftrag.

Jede Entnahme von Artikeln aus dem Fachbodenregal wird vom Mitarbeiter durch den Scan mit Hilfe eines Scanner-Handschuhs bestätigt. Wurde der korrekte Artikel in der richtigen Menge gescannt, springt das System automatisch zum nächsten Prozessschritt. Wurde ein falscher Artikel gescannt, wird dieser Fehler umgehend über das Display visualisiert.

Pick-by-Laser

KONZEPT

Ein weiterer Ansatz, intelligente Assistenten in bestehende Regalsysteme zu integrieren ist die Kombination aus einem fest installierten und mobilen System. Das Konzept Pick-by-Laser baut auf der Annahme auf, dass Mitarbeiter in der Kommissionierung häufig mit Kommissionierwagen arbeiten, auf denen die Aufträge gesammelt werden und die der Mitarbeiter mit sich bewegt. Idee dieses Konzeptes ist, eine Projektionseinrichtung – das Laser-Pick-Modul – fest auf einem Kommissionierwagen zu installieren. Durch kontextsensitive Projektionen auf den Boden sowie in das Fachbodenregal können Arbeitsanweisungen dynamisch eingeblendet werden.

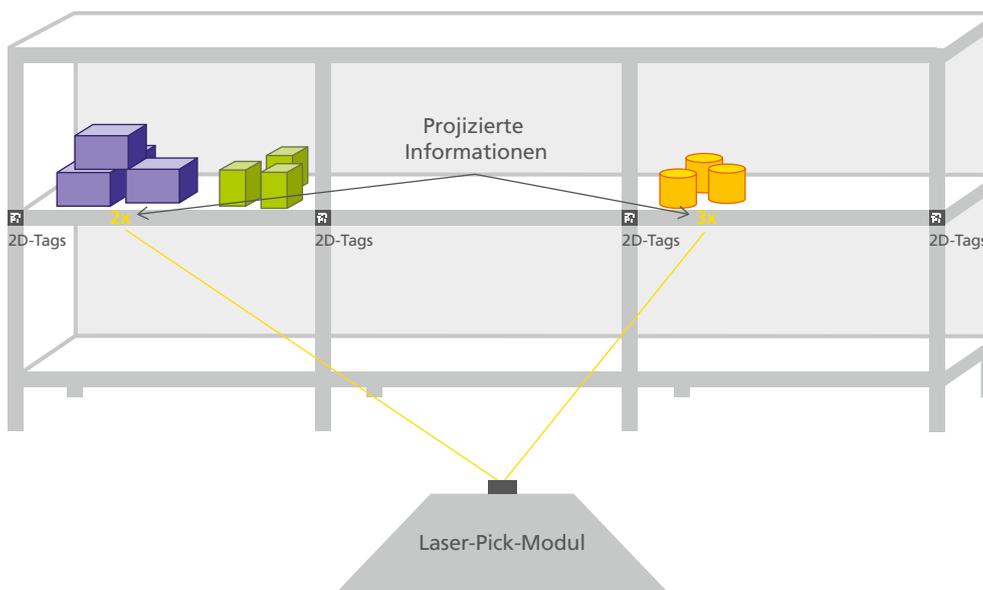


Abbildung 17:
Projektion von
Arbeitsanweisungen
im Fachbodenregal

Grundsätzlich werden in diesem Konzept zwei unterschiedliche Arten der Projektion betrachtet. Auf der einen Seite steht die Führung des Mitarbeiters zu einem bestimmten Regalfach. Diese erfolgt über die Projektion von Richtungsangaben – ähnlich einem Navigationssystem – auf den Boden neben dem Mitarbeiter. Diese Informationen sollen den Mitarbeiter bis zum Entnahmeort leiten. Auf der anderen Seite steht die konkrete Pickanweisung am Regalfach. Erreicht der Mitarbeiter das entsprechende Regalfach, werden Pickanweisungen für jeden Artikel im Sichtfeld des Lasers unmittelbar unter oder in das entsprechende Regalfach projiziert. Das Grundprinzip des beschriebenen Konzeptes wird in Abbildung 17 dargestellt.

WHITEPAPER KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

Die Lokalisierung im Raum erfolgt in diesem Konzept über visuelle Marker, die auf 2D-Codes aufbauen, die an exponierten Stellen im Regal bzw. im Raum angebracht wurden. Über eine interne Karte und die über eine Kamera erkannten Codes kann eine exakte Lokalisierung im Raum erfolgen.

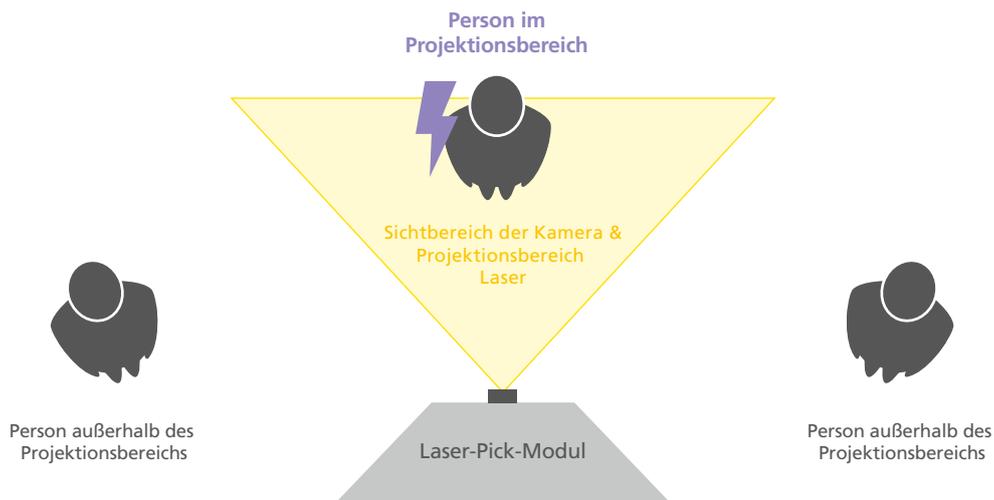


Abbildung 18:
Schutz der Mitarbeiter durch die automatische Deaktivierung der Projektion

Um die Mitarbeiter zu schützen, beinhaltet das beschriebene Konzept eine Sicherheitsfunktion, die über die integrierte Kamera automatisch Menschen erkennt und den Laser deaktiviert, sobald ein Mensch im Sichtfeld identifiziert wurde. Diese Maßnahme wird benötigt, um den Arbeitsschutz im Umgang mit Lasern zu gewährleisten.

PROTOTYPISCHE UMSETZUNG

Für die prototypische Umsetzung des Konzeptes wurde ein Kommissionierwagen mit einem Laser-Pick-Modul ausgestattet. Dieses Modul besteht aus einer Recheneinheit, einem Touch-Monitor, einer Kamera sowie einem Laserprojektor. Durch eine feste Verknüpfung der Kamera mit dem Projektor, kann die exakte Projektion auf Basis der ermittelten Lokalisierungsdaten gewährleistet werden.

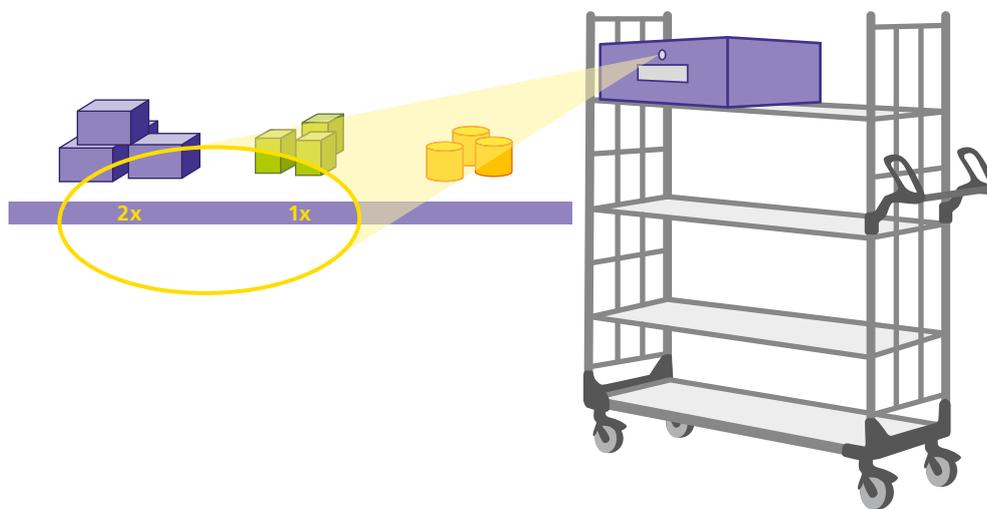


Abbildung 19:
Projektion von Ent-
nahmeanweisungen

Die Lokalisierung im Raum wurde durch den Einsatz sogenannter Aruco-Marker umgesetzt, die so im Raum verteilt wurden, dass sich immer mindestens ein Marker im Sichtfeld der Kamera befindet. Anhand der Erkennung eines Markes lassen sich Rückschlüsse auf den Winkel und die Distanz der Kamera in Relation zum Marker ableiten. Sind mehrere Marker im Sichtfeld steigt die Genauigkeit der Lokalisierung.

Befindet sich das System in einer Regalzeile bzw. in unmittelbarer Nähe zum Entnahmepunkt, erfolgt die Projektion der Arbeitsanweisungen unter oder in das entsprechende Regalfach (siehe Abbildung 19). Befindet sich das System nicht in einer Regalzeile bzw. nicht in Sichtweite des Entnahmepunktes, erfolgt die Projektion der Navigationsinformationen auf den Boden neben den Mitarbeiter (siehe Abbildung 20).



Abbildung 20:
Projektion von
Navigationsinfos auf
den Boden

Validierung

Heutzutage sind verschiedene Assistenzsysteme in der Kommissionierung im Einsatz. Durch das Forschungsprojekt Innovationslabor Hybride Dienstleistungen in der Logistik wird das aktuell verfügbare Systemspektrum durch neue Prototypen wie das Pick-by-Laser System, das Intelligente Regal oder Mika erweitert. In der Validierungsphase geht es nun darum, die am IML entwickelten Kommissioniersysteme mit den bereits etablierten Systemen zu vergleichen und die Vor- und Nachteile herauszustellen.

Die Grundlage der Untersuchung bildet eine intensive Literatur- und Marktrecherche, mittels dieser die Auswahl gängiger Verfahren und deren Bewertungen nach bestimmten Kriterien herausgearbeitet wurde [15, 16, 17, 18]. Bei den Kriterien handelt es sich um Vergleichskriterien, die sich durch unterschiedliche Ausprägungen bei den verschiedenen Kommissioniersystemen kenntlich machen. Hierbei wurden zum einen eigene Kriterien aufgestellt und zum anderen bereits definierte Kriterien aus der Literatur oder durch Produktbeschreibungen festgelegt.

Zur Validierung der Prototypen und zum Vergleich mit etablierten Kommissioniersystemen, wurde eine Validierungsmatrix erstellt (Abbildung 21). Hierbei wurden folgende Systeme zum Vergleich herangezogen [19, 20, 21, 22]:

Pick-by-Paper: Kommissionierinformationen werden per Pick-Liste dargestellt. Vorgänge werden dort händisch festgehalten.

Pick-by-Scan: Kommissionierinformationen werden per Hands scanner dargestellt. Vorgänge werden händisch oder durch Scan festgehalten.

Pick-by-Light: Kommissionierinformationen werden per LED-Modul am Regalfach dargestellt. Vorgänge werden durch Bestätigung am Modul festgehalten.

Pick-by-Voice: Kommissionierinformationen werden per Headset übermittelt. Vorgänge werden durch verbale Rückmeldungen festgehalten.

Pick-by-Projection: Kommissionierinformationen werden per Projektion an den Kommissionierfächern dargestellt. Vorgänge werden durch Interaktion mit dem Projektor festgehalten.

WHITEPAPER

KONZEPTE ZUR KOMMUNIKATION ZWISCHEN MITARBEITERN UND MASCHINEN

Zusätzlich zu den am Markt bestehenden Kommissioniersystemen, wurden die drei erwähnten Prototypen in die Vergleichsmatrix aufgenommen:

Pick-by-Laser: Kommissionierinformationen werden per Lichtpunkte am Lagerfach ausgehend von einem Laser am Kommissionierwagen dargestellt. Vorgänge werden am Wagen festgehalten.

Mika: Kommissionierinformationen werden per verfahrbaren Displays an den Regalzeilen dargestellt. Vorgänge werden durch Knopfdruck am Display festgehalten.

Intelligentes Regal: Kommissionierinformationen werden dem Mitarbeiter an einem Display am relevanten Lagerfach angezeigt. Vorgänge werden durch Knopfdruck oder verbal am Display festgehalten.

	1	2	3	4	5	6	7	8	9	10	11	
Bewertungskriterien	Freie Hände zum Kommissionieren	Bedarf an tragbaren Kommissionierhilfen	Mentale Anstrengung bei Dauernutzung	Intuitive und strikte Prozessführung beim Kommissionieren	System kann Mitarbeiter automatisch identifizieren	System kann aktuellen Standort des Kommissionierers bestimmen	System hebt Kommissionierort visuell hervor	Automatische Bestandsaktualisierung wird durchgeführt	System ist lernfähig (Machine Learning)	Technische Kommissionierhilfe muss pro Lagerfach installiert werden	System ist aufwandsarm skalierbar	0% = erfüllt Kriterien nicht 100% = erfüllt Kriterien voll
Pick-by-Paper	nein	ja	ja	nein	nein	nein	nein	nein	nein	nein	ja	18%
Pick-by-Scan	nein	ja	ja	nein	nein	nein	nein	ja	nein	nein	ja	27%
Pick-by-Light	ja	nein	nein	ja	nein	nein	ja	ja	nein	ja	nein	55%
Pick-by-Projection	ja	nein	nein	ja	nein	nein	ja	ja	nein	ja	nein	55%
Pick-by-Vision	ja	ja	ja	ja	nein	nein	ja	ja	nein	nein	ja	55%
Pick-by-Voice	ja	ja	nein	ja	nein	nein	nein	ja	nein	nein	ja	55%
Pick-by-Laser	ja	nein	nein	ja	nein	ja	ja	ja	nein	nein	ja	82%
Intelligentes Regal	ja	nein	nein	ja	ja	ja	ja	ja	ja	ja	ja	91%
Mika	ja	nein	nein	ja	ja	ja	ja	ja	ja	nein	ja	100%

Legende



positive Eigenschaft des Kommissioniersystems

negative Eigenschaft des Kommissioniersystems

Abbildung 21: Validierungsmatrix der Kommissioniersysteme

Um die genannten Kommissioniersysteme vergleichen zu können, wurden verschiedenste Bewertungskriterien aufgenommen, die unterschiedliche Vor- und Nachteile der Systeme adressieren. Einige dieser Kriterien fokussieren sich auf den Prozess, den der Mitarbeiter bei der Kommissionierung durchführt und darauf, inwieweit das Kommissioniersystem den Mitarbeiter unterstützt oder aber einschränkt. Beispiele hierfür ist der Bedarf an Hilfsmitteln, die am Körper getragen werden müssen und die Freihändigkeit des Mitarbeiters bei der Kommissionierung. Weitere Kriterien beziehen sich auf die Fähigkeiten des Systems z. B. Mitarbeiter zu erkennen, zu lokalisieren oder von durchgeführten Aktionen zu lernen und sich selbst zu optimieren. Der dritte Kriterienbereich bezieht sich auf die Implementierung und Erweiterbarkeit der Systeme. Hierbei wird zum Beispiel bewertet, wie aufwandsarm die Systeme eingeführt oder skaliert werden können. Die Bewertung der Kommissioniersysteme wurde durch Literatur- und Marktrecherche sowie durch Experteninterviews durchgeführt.

Hinsichtlich der Benutzerfreundlichkeit ist bei der Auswertung zu erkennen, dass bei Pick-by-Paper sowie Pick-by-Scan ein Bedarf an tragbaren Kommissionierhilfen besteht. Da es sich zum einen um eine Liste und zum anderen um einen Scanner handelt, sind die Hände beim Kommissionieren nicht frei. Auch wenn die Systeme Pick-by-Voice und Pick-by-Vision ebenfalls tragbare Hilfsmittel nutzen, kommt es aufgrund des Einsatzes am Kopf zu keiner Belegung der Hände. Zusätzlich ist hinsichtlich der Durchführbarkeit zu erkennen, dass sich die Systeme, bei denen die Hände belegt sind, sich nicht durch intuitive und strikte Prozessführung auszeichnen. Ebenfalls ist eine hohe mentale Daueranstrengung bei längerer Nutzung festzustellen. Das Verfahren Pick-by-Vision kann bei einer dauerhaften Anwendung ebenfalls zu mentalen Beschwerden aufgrund der visuellen und physischen Überlastung führen.

Im Bereich der Fähigkeiten der einzelnen Systeme ist auffällig, dass Aspekte, die aus der Industrie 4.0 entspringen wie Machine Learning, automatische Identifikation von Objekten und Fehlern bei den neuen Systemen wie Mika oder dem intelligenten Regal berücksichtigt werden. Bis auf die Kommissionierung mittels Pickliste kann mit jedem System eine automatische Bestandsaktualisierung durchgeführt werden, da die Systeme mit dem Lagerverwaltungssystem interagieren können.

In Bezug auf die Umsetzbarkeit und Skalierbarkeit wird bei den Systemen Pick-by-Light, Pick-by-Projection und dem Intelligenten Regal deutlich, dass die Implementierung mit einem hohen Aufwand verbunden ist. Diese Systeme werden ortsgebunden pro Lagerfach installiert und lassen sich nur mit hohem organisatorischem Aufwand verändern. Von diesen drei Kommissioniersystemen lässt sich nur das Intelligente Regal genauso wie die übrigen Systeme aufwandsarm skalieren, nachdem die Grundlage der Infrastruktur einmal geschaffen ist.

Die Ergebnisse in Abbildung 21 zeigen, dass die neuen Systeme bei der Kriterienerfüllung einen höheren Erfüllungsgrad als bei den konventionellen Systemen aufzeigen. Im Betrachtungsraum der neuen Systeme kann lediglich das Pick-by-Laser Verfahren die automatische Identifikation des Mitarbeiters und das Machine Learning nicht ermöglichen. Dennoch erfüllt das Pick-by-Laser System im Gegensatz zu den bekannten Verfahren höhere Anforderungen. Während das Mika Verfahren am besten abschneidet, ergibt sich beim Intelligenten Regal eine Schwäche durch die Installation des Systems pro Lagerfach.

Literaturverzeichnis

- [1] VL53L1X - Time-of-Flight ranging sensor based on ST's FlightSense technology
URL: <https://www.st.com/en/imaging-and-photonics-solutions/vl53l1x.html#documentation> -
Abgerufen am: 25.04.2020
- [2] AN5109 - STM32MP1 Series using low-power modes
URL: https://www.st.com/resource/en/application_note/dm00449434.pdf - Abgerufen am
26.11.2020
- [3] The OpenAMP Project.
URL: <https://www.openampproject.org/> - Abgerufen am 27.04.2020
- [4] OpenAMP. RPMsg Messaging Protocol.
URL: <https://github.com/OpenAMP/open-amp/wiki/RPMsg-Messaging-Protocol> - Abgerufen am 27.
04. 2020
- [5] LinuxTV.org - Television with Linux.
URL: <https://www.linuxtv.org/> - Abgerufen am 27.04.2020
- [6] K. Zhang, Z. Zhang, Z. Li and Y. Qiao (2016), "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, doi: 10.1109/LSP.2016.2603342.
- [7] F. Schroff, D. Kalenichenko and J. Philbin (2015), "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.
- [8] How to Develop a Face Recognition System Using FaceNet in Keras
URL: <https://machinelearningmastery.com/how-to-develop-a-face-recognition-system-using-facenet-in-keras-and-an-svm-classifier/> - Abgerufen am: 29.03.2021
- [9] Github Repository Mozilla DeepSpeech
URL: <https://github.com/mozilla/DeepSpeech> - Abgerufen am: 29.03.2021
- [10] Github Repository AASHISHAG deepspeech-german
URL: <https://github.com/AASHISHAG/deepspeech-german> - Abgerufen am: 29.03.2021

- [11] German End-to-end Speech Recognition based on DeepSpeech
URL: https://www.researchgate.net/publication/336532830_German_End-to-end_Speech_Recognition_based_on_DeepSpeech - Abgerufen am: 29.03.2021
- [12] Github Repository RasaHQ rasa
URL: <https://github.com/RasaHQ/rasa> - Abgerufen am: 29.03.2021
- [13] Anran Jiao (2020), "An Intelligent Chatbot System Based on Entity Extraction Using RASA NLU and Neural Network", Phys.: Conf. Ser. 1487 012014.
- [14] Bocklisch, T., Faulkner, J., Pawlowski, N., & Nichol, A. (2017). Rasa: Open source language understanding and dialogue management. arXiv preprint arXiv:1712.05181.
- [15] Spee, D., Kretschmer, V. (2018): Kognitive Ergonomie. HUSS-VERLAG GmbH
- [16] Mättig, B., & Kretschmer, V. (2020). Einsatz digitaler Assistenzsysteme in der Logistik 4.0. In Handbuch Industrie 4.0 (pp. 435-459). Springer Vieweg, Berlin, Heidelberg.
- [17] Schlögl, D., & Zsifkovits, H. (2016). Manuelle Kommissioniersysteme und die Rolle des Menschen. BHM Berg-und Hüttenmännische Monatshefte, 161(5), 225-228.
- [18] Spee, D. (2009). Systematik der Kommissioniersysteme. Schietinger and Pulverich [61], 30-54.
- [19] Baechler, A., Baechler, L., Autenrieth, S., Kurtz, P., Hoerz, T., Heidenreich, T., & Kruell, G. (2016, January). A comparative study of an assistance system for manual order picking--called pick-by-projection--with the guiding systems pick-by-paper, pick-by-light and pick-by-display. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 523-531). IEEE.
- [20] Guo, A., Raghu, S., Xie, X., Ismail, S., Luo, X., Simoneau, J., ... & Starner, T. (2014, September). A comparison of order picking assisted by head-up display (HUD), cart-mounted display (CMD), light, and paper pick list. In Proceedings of the 2014 ACM International Symposium on Wearable Computers (pp. 71-78).
- [21] Pick-by-Vision: Vorteile der Picavi Datenbrille auf einen Blick.
URL: <https://picavi.com/nutzen/> - Abgerufen am 06.10.2020.
- [22] Kommissioniermethoden im Vergleich: Pick-by-Vision, Pick-by-Light, Pick-by-Voice und Pick-by-Scan
URL: <https://www.serkem.de/kommissioniermethoden-im-vergleich-pick-by-vision/> - Abgerufen am 06.10.2020.

WHITEPAPER
KONZEPTE ZUR KOMMUNIKATION ZWISCHEN
MITARBEITERN UND MASCHINEN