HOCHSCHULE RUHR WEST
UNIVERSITY OF APPLIED SCIENCES

INSTITUT INFORMATIK

Internal Report 19-02

DamokleS 4.0

Dynamisches, adaptives und mobiles System zur kontextbasierten Und intelligenten Unterstützung von Mitarbeiter/Innen in der Schwerindustrie

Hochschule Ruhr West
N. Zengeler, M. Grimm, A. Arntz, D. Kessler,
M. Jansen, S. Eimler, U. Handmann

celano GmbH
K. Hermsen, C. Grubert

EUROPÄISCHE UNION
Investition in unsere Zukunft
Europäischer Fonds
für regionale Entwicklung

EFRE.NRW
Investitionen in Wachstum
und Beschäftigung

Inhaltsverzeichnis

1 Einleitung **2**
- 1.1 Zielstellung . . . 3
- 1.2 Ansatz . . . 4

2 Stand der Technik **5**
- 2.1 Kontextmodelle . . . 5
- 2.2 Augmented Reality . . . 6
- 2.3 Personenerkennung und Personenverfolgung . . . 6

3 Anwendungsszenarien **7**
- 3.1 Sicherheit am Arbeitsplatz . . . 8
- 3.2 Produktion . . . 10
- 3.3 Wartung . . . 10
- 3.4 Lagerverwaltung . . . 12

4 Versuchsaufbau **13**
- 4.1 Ablauf . . . 13
- 4.2 Auswertung . . . 14

5 Kontextmodell **14**
- 5.1 Mehrdimensionale Kontextmodellierung . . . 14
- 5.2 Cyberphysikalisches System . . . 17
- 5.3 Positionsbestimmung über Beacons . . . 21

6 Augmented Reality Anwendung **23**
- 6.1 Entwicklungsprozess . . . 23
- 6.2 Navigationsarten . . . 24
- 6.3 Simulierte Umgebung . . . 24
- 6.4 Auswertung . . . 26

7 Kamerabasierte Personenverfolgung **29**
- 7.1 Personenerkennungsmethoden . . . 29
- 7.2 Bildstörungsresistenz . . . 33
- 7.3 Personenverfolgung . . . 37
- 7.4 Thermografieintegration . . . 42

8 Zusammenfassung **46**
- 8.1 Anwendungsnutzen . . . 48
- 8.2 Diskussion . . . 48
- 8.3 Zukünftige Studien . . . 51

Zusammenfassung

Dieser interne Bericht beschreibt die Zielsetzung, Durchführung und Auswertung des Projektes *Damokles 4.0*. Das Projekt zielt darauf ab, neue, digitale Technologien in die Schwerindustrie einzuführen um Produktionsprozesse zu modernisieren. Unter Einsatz neuer Technologien, insbesondere mobiler Geräte, soll ein cyberphyiskalisches System (CPS) eine kontextbasierte und künstlich intelligente Unterstützung der Mitarbeiter in den Bereichen der Schwerindustrie ermöglichen. Hierzu werden typische Anwendungsfälle und die damit verbundenen Szenarien zur Unterstützung der Mitarbeiter auf Basis von neuen, flexiblen, adaptiven und mobilen Technologien, wie Augmented Reality und künstlicher Intelligenz, modelliert. Um den Prototypen einer AR-Anwendung und einer kamerabasierte Personenverfolgung zu entwickeln, hat die Hochschule Ruhr West im kleinen Technikum am Campus Bottrop eine entsprechende industrielle Umgebung simuliert. Die Projektergebnisse zeigen die Anwendbarkeit der vorgeschlagenen Softwareansätze und die Ergebnisse einer Untersuchung der psychologischen Einflüsse auf die Mitarbeiter.

1 Einleitung

Der medienwirksame Begriff Industrie 4.0 bezeichnet die vierte industrielle Revolution, welche durch massive Digitalisierung in allen Geschäftsbereichen charakterisiert wird. Dies bezieht sich auf die Integration von hochwertigen, digitalen Technologien, welche eine neue Qualität in klassische Produktionsprozesse einführen. Mit diesen Technologien können Maschinen nahtlos miteinander verbunden und hohe Interoperabilitäten zwischen Anwendern, Produktionsprozessen und informationsverarbeitenden Systemen erzeugt werden.
Die Schwerindustrie, insbesondere die metallverarbeitenden Unternehmen in Deutschland, stehen vor der Herausforderung eines ständig zunehmenden Wettbewerbs in globalem Maßstab [22]. Verschiedene Einflüsse wirken auf die wirtschaftlichen Überlegungen ein. Der Trend zu einer immer flexibleren Produktion, einem breiteren Produktmix und kleineren Produktionseinheiten ist eine aktuelle Herausforderung für die Stahlindustrie, wachsende Kundenanforderungen an die Produktqualität sind ein weiterer. Darüber hinaus wird die Verbesserung der Energieeffizienz und die Reduzierung der Kohlendioxidemissionen aufgrund von Rohstoffknappheit, steigenden Energiepreisen und höheren Umweltschutzanforderungen immer wichtiger. Um unter diesen Wettbewerbsbedingungen bestehen zu können, müssen die Produktionsprozesse optimiert werden.

Ein wesentlicher Faktor im Produktionsprozess ist die Belegschaft [20]. An jedem Punkt der gesamten Wertschöpfungskette haben die Entscheidungen der Mitarbeiter Einfluss auf den Erfolg der Produktion. Ein System, welches die Mitarbeiter durch die Bereitstellung prozessbezogener Echtzeitdaten unterstützt, kann dazu beitragen, die Arbeitssicherheit zu erhöhen und den Produktionsprozess wirtschaftlicher zu gestalten. Dies trägt zum wirtschaftlichen Erfolg bei und kann zu einer insgesamt verbesserten Qualität der Produktion führen.

1.1 Zielstellung

Ziel des Projektes *DamokleS 4.0* ist es, neue Szenarien und Prozesse zur Unterstützung der Mitarbeiter in der Schwerindustrie auf Basis von neuen, flexiblen, adaptiven und mobilen Technologien zu entwickeln. Die Entwicklung von Systemen und neuen Bedienkonzepten für mobile, kontextbezogene und datenbasierte Unterstützung erfordert eine zuverlässige Erkennung des jeweiligen Kontexts. Als Eingabe für die Kontexterkennung werden die Sensoren der mobilen Geräte sowie externe Sensorvorrichtungen, in diesem Anwendungsfall Kamerabilder und Beacondaten, verwendet.
In Produktionsbetrieben, wie beispielsweise in der Schwerindustrie, gibt es im Allgemeinen widrige Umgebungsbedingungen, die zu berücksichtigen sind und für die Auswertung der Sensordaten erschweren können. Ein weiteres Ziel der Arbeit ist es, neue Lösungen zu finden, die sicherstellen, dass die Nutzung mobiler Geräte unter diesen Bedingungen gewährleistet werden kann. Um die Zielstellung des Projektes zu erreichen ist es von besonderer Bedeutung, die angedachte automatisierte Datenverarbeitung auf einfache, zuverlässige und intuitive Art und Weise bereit zu stellen. Eine zentrale Rolle spielt hierbei die automatische Erkennung des Kontextes einer Situation. Hierbei werden die Rollen und Berechtigungen der Benutzer des Systems als Dimensionen eines Kontextmodells betrachtet [42]. Das Kontextmodell soll diejenigen Informationen liefern, die unter den gegebenen Umständen erforderlich sind und redundante Daten vermeiden. Für die Bereitstellung von Daten muss es entschieden, wer welche Daten wann und wo benötigt. Hierzu müssen die Rollen der Anwender, deren Standort sowie die jeweilige Produktionssituation erkannt und ausgewertet werden. Das Kontextmodell soll alle anfallenden Daten verwalten, sortieren und den verschiedenen mobile Endgeräte zur Verfügung stellen.

1.2 Ansatz

Eines dieser mobilen Geräte, die im Projekt eingesetzt werden, ist ein Augmented Reality (AR) Head Mounted Display. Diese Geräte ermöglichen die Visualisierung von Daten, wie beispielsweise Navigationswegen, direkt in das Sichtfeld (E: *Field of view*, FOV) der Mitarbeiter. Aufgrund der verschiedenen Sicherheitsbedenken in einer industriellen Umgebung ist es jedoch notwendig zu bewerten, wie und auf welche Weise die AR-Unterstützung am effektivsten ist. Zu diesem Zweck simuliert das Projektteam ein industrielles Szenario, in welchem das Projektteam zwei AR-Anwendungen miteinander vergleicht, die jeweils über eine eigene Benutzeroberfläche verfügen, aber den Testteilnehmer die selben Aufgaben stellen.
Ein solches System zur Navigation, beispielsweise zu bestimmten Maschinen oder zu einer schnellen Evakuierungswegen im Notfall, gründet auf einer zuverlässigen Erkennung von Menschen in der jeweiligen Anlage. Ein wesentlicher Bestandteil einer solchen Software ist die Verarbeitung von Kamerabildern, auf welchen die Personen erkannt werden sollen. Mit Hilfe von Methoden der künstlichen Intelligenz liefert das Projekt *DamokleS 4.0* einen ganzheitlichen Softwareansatz zur Lokalisierung und Verfolgung von Personen in industriellen Umgebungen auf der Grundlage von Kamerabildern. Für den Einsatz in der Schwerindustrie hat sich das Projektteam aus Datenschutzgründen dazu entschieden, die Mitarbeiter über tragbaren Smart Devices anstelle einer Gesichtserkennung zu identifizieren.
Die Hochschule Ruhr West untersucht verschiedene Methoden zur Erkennung von menschlichen Körpern in schwerindustriellen Umgebungen [65, 4]. Aus den Experimentalszenarien entsteht eine Arbeitsvideodatensatz in simulierter industrieller Umgebung, welcher zur Auswertung der Personenverfolgungsansätze herangezogen wird. Insbesondere werden Kamerabilder verwendet, um die Fußpunkte der Menschen im Bild zu lokalisieren und somit auf ihre Position in der Umgebung zu schließen. Zur Findung eines genauen und ausfallsicheren Personenerkennungssystem werden programmatisch verschiedene Bildstörverfahren angewandt und die Genauigkeit einiger ausgewählter Detektoren ausgewertet. Die Kenntnis über den Standort des Mitarbeiters sowie seiner Rollen innerhalb des Produktionsprozesses ermöglicht es dem Kontextmodell, die entsprechenden Werte zu verarbeiten und die passenden, individuellen Hilfsinformationen anzuzeigen. Auf diese Weise kann ein Informationsdienst relevante Daten, wie beispielsweise die aktuelle Aufgabe, an die AR-Brille eines Mitarbeiters senden. So erhält der Mitarbeiter die kontextsensitive und rollenspezifische Information, die er zur Zeit benötigt, um beispielsweise den Zustand einer Maschine zu untersuchen, einer konkrete Arbeitsanweisung zu folgen oder einen Evakuierungsweg im Notfall zu finden.

2 Stand der Technik

Das Projekt *DamokleS 4.0* nutzt das Wissen aus vorherigen Projekten und vorangegangener Forschung zur bestmöglichen Gestaltung der Projektergebnisse. Das Projektteam verwendet die Ergebnisse von verwandter Forschung auf dem Gebiet von Kontextmodellen, Augmented Reality Brillen sowie kamerabasierter Personenerkennung und Personenverfolgung. Der Stand der Technik wird hinsichtlich der Nützlichkeit in Bezug auf die Anforderungen des Projektes untersucht und mit alternativen Ansätzen verglichen.

2.1 Kontextmodelle

Bezüglich des Kontextmodells bezieht sich das Projektteam auf die Definition von Dey [12]: „*Der Kontext besteht aus allen Informationen, die zur Charakterisierung der Situation einer Entität herangezogen werden können. Eine Entität ist eine Person, ein Ort oder ein Objekt, das für die Interaktion zwischen einem Benutzer und der Anwendung als relevant angesehen wird, einschließlich des Benutzers und der Anwendung selbst.*“Nach dieser Definition unterscheiden sich die Aspekte der Kontextinformationen je nach Anwendung. Viele Arbeiten nutzen Kontextmodelle für verschiedene Zwecke. Zum Beispiel verwenden Context-Aware Recommendation Systems (CARSs) ein Kontextmodell zur Bereitstellung von interessanten Daten (E: *Points of Interest*, POI) für den Nutzer [2, 57]. Dieser Ansatz findet in den Bereichen des Tourismus, der mobilen Lernressourcen und der Nachrichtenempfehlungen Anwendung [9, 5, 64, 50, 11, 1, 23, 51].
Kontextbezogene Anwendungen verwenden unterschiedliche Ansätze, wie beispielsweise einen mehrdimensionalen Kontextmodellierungsansatz [31], objektrollenbasierte Modelle [19] und ontologiebasierte Modelle [58]. Ein mehrdimensionaler Ansatz, auch als Rich Context Model (RCM) bezeichnet, übertrifft andere Ansätze in Bezug auf die Erfüllung der Anforderungen an Datenreichtum, Flexibilität, Granularität und Leistung [49]. Darüber hinaus verfügen die Kontextmodelle dieses Ansatzes über hohe Leistung und Flexibilität bei der Darstellung von Kontextinformationen, die sich auf verschiedene Detaillierungsstufen beziehen [49]. In diesem Ansatz stellt jede Dimension die Eigenschaften einer Entität dar, wobei verschiedene Kontextsituationen als Vektoren in einem mehrdimensionalen Raum betrachtet werden. Das Modell erkennt den aktuellen Kontextwert mithilfe einer Kontextvektorähnlichkeit.

2.2 Augmented Reality

Es gibt mehrere mobile AR-Navigationsansätze und verschiedene Projekte zur Erforschung von AR im Kontext der Schwerindustrie, jedoch wurde bisher wenig Arbeit geleistet, welche diese beiden Komponenten kombiniert [29]. Mobile Lösungen können sich nicht nur auf das GPS-Signal verlassen, um eine präzise Darstellung zu ermöglichen, da die Signalstärke in Innenräumen sehr begrenzt sein kann [29]. Die genaue Lokalisierung ist jedoch unerlässlich, um das Personal mit genauen Angaben zu versorgen. Frühere Forschungen haben Rucksäcke mit zusätzlicher Hardware verwendet um die Zuverlässigkeit der Navigation zu verbessern, was jedoch die Bewegungsfreiheit des Benutzers stark einschränken kann [30]. Kim und Jun schlagen einen kamerabasierten Ortungsalgorithmus für die AR-Indoor-Navigation vor, der eine Bilddatenbank mit zuvor aufgenommenen Bildern nutzt, um den aktuellen Standort des Benutzers zu ermitteln [25]. Hierzu ist ein externer Computer erforderlich, welcher die Kamerabilder in Echtzeit verarbeitet und die Ergebnisse entsprechend kommuniziert. Rehman und Cao verfolgen einen ähnlichen Ansatz für ein mobiles Framework und vergleichen AR-Geräte mit herkömmlichen Smartphones [46]. Sie kommen zu dem Ergebnis, dass die Nutzer die tragbaren AR-Geräte zwar als präziser empfinden, jedoch in der Arbeitsleistung kaum einen Unterschied zur Nutzung von Smartphones besteht.

2.3 Personenerkennung und Personenverfolgung

Der aktuelle Stand der Technik kennt eine Vielzahl von Methoden zur Erkennung von menschlichen Körpern auf Kamerabildern [8, 10, 13, 15, 43, 44, 45, 47, 52, 59, 62, 63, 67, 68]. Es wurde entschieden, die Untersuchungen auf drei gut erforschte Ansätze zu konzentrieren: die Merkmalsklassifikation mit Gradientenorientierungshistogrammen (E: *Histograms of Oriented Gradients*, HOG) [7, 10, 35, 52, 56, 69], die neuronale Architektur *You Only Look Once* (YOLO) [43, 44, 45] und die *OpenPose* Architektur (OP) [8, 47, 59].
Der aktueller Stand der Forschung untersucht Personenverfolgungstechniken unter verschiedenen Anforderungen mit unterschiedlichen Ansätzen. Zum Beispiel schlägt [53] einen Personenverfolgungsalgorithmus für ein autonomes unbemanntes Luftfahrzeug vor. In diesem Ansatz verfolgt eine Drohne mit einer Überwachungskamera einzelne Personen, was ein sehr flexibles Überwachungssystem mit dem Vorteil von einfacher und zuverlässiger Gesichtserkennung ermöglicht. Im Vergleich zum stationären, kamerabasierten Ansatz kann der Einsatz einer Drohne in einer Industriehalle eher von Nachteil sein, da die fliegende Drohne möglicherweise mit beweglichen Objekten wie Kränen, Fahrzeugen oder sogar anderen Personen kollidieren kann. Um das

Problem der gleichzeitigen Verfolgung mehrerer Personen zu lösen, verwendet [17] die *Slow Feature Analysis* (SFA, [61]). [3] stellt ein Framework zur Personenpersonenverfolgung bereit, das aus vier Modulen zur Verfolgung über kurzzeitige Merkmale sowie Funktionen für unbeaufsichtigtes und überwachtes maschinelles Lernen besteht.
Im Projekt *APFel* hat das Projektteam ein Videoüberwachungssystem zum Schutz kritischer Infrastrukturen ausgearbeitet, welches ausschließlich die HOG-Methode verwendet, kombiniert mit einem Kalman-Tracking-Algorithmus [16]. In diesem Projekt wurde die Softwarearchitektur so konzipiert, dass sie menschliche Sicherheitskräfte dabei unterstützt, verdächtige Personen und Objekte zu erkennen und über die verschiedenen Kameras hinweg zu verfolgen. Damit dies geschehen kann, werden die einzelnen Kamerabilder nicht in zeitlicher Reihenfolge sortiert, sondern in der Reihenfolge der Anwesenheit gesuchter Merkmale. Um die Praxistauglichkeit zu bewerten, wurde das System an zwei Referenzflughäfen implementiert. Als ein Problem stellt sich die riesige Menge an Bilddaten heraus, die in einem großen Kameranetzwerk anfallen. Um die Auswertung zu beschleunigen und somit die Echtzeitanforderung erfüllen wurde die Nutzung einer GPU-Implementierung entschieden. Die im Projekt *APFel* beschriebenen Szenarien ähneln denen des Projekts *DamokleS 4.0* im Hinblick auf die Herausforderungen durch unterschiedliche, teils schwierige Lichtverhältnisse und den hohen Bedarf an schnellen, echtzeitfähigen Algorithmen.

3 Anwendungsszenarien

Gemeinsam mit den Interessenvertretern und assoziierten Partnern der Stahlindustrie wurden reale Produktionssituationen gesammelt und in vier verschiedene Szenarien eingeteilt. Vor diesem Hintergrund werden einige Szenarien vorgeschlagen, die sich auf Arbeitssicherheit, Produktion und Instandhaltung beziehen. Die zugrundeliegenden Geschichten der Szenarien, die alle den Bedarf an kontextbasierter Datenunterstützung beschreiben, sind die Grundlage für die simulierte industrielle Umgebung. Anhand dieser simulierten Umgebung werden die AR-Anwendung sowie die kamerabasierten Personenverfolgungsmethoden evaluiert. Zur technischen Modellierung der Szenarien werden Komponentendiagramme verwendet, welche die angedachten Dienste und Geräte im Szenario abbilden. Abbildung 1 fasst die Gesamtheit der verwendeten Geräte und Dienste zusammen. Nachfolgend werden die Szenarien dargestellt, wie sie nach mehrfacher Überarbeitung von allen Beteiligten für in Ordnung befunden wurden.

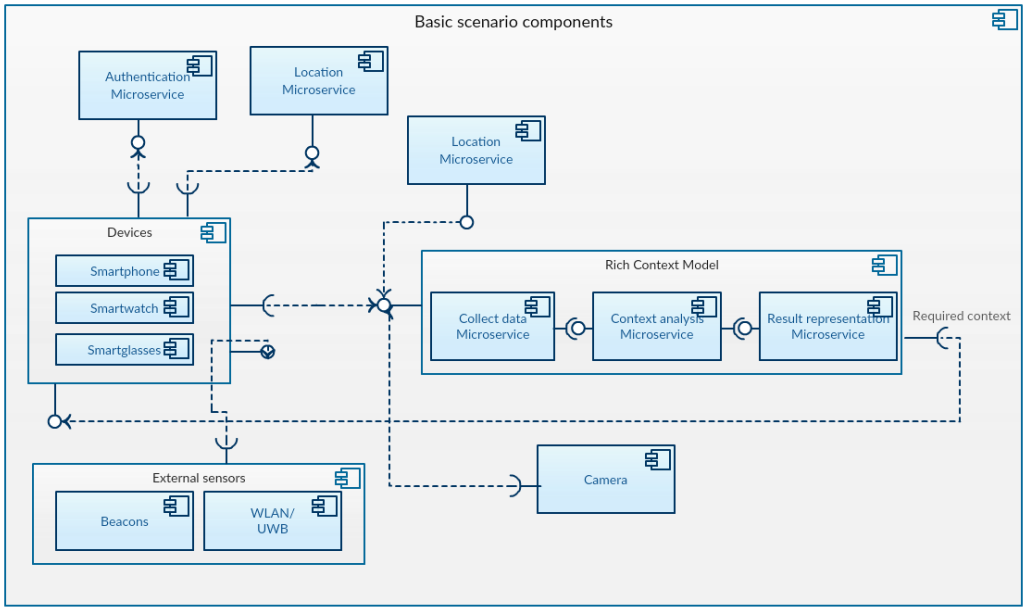


Abbildung 1: Die Geräte und Dienste, die allen Szenarien zu Grunde liegen, sowie deren Verbindungen.

3.1 Sicherheit am Arbeitsplatz

Frank ist bei einem Stahlunternehmen seit 15 Jahren für verschiedene Maschinen verantwortlich. Er kommt früh am Morgen zur Arbeit, zieht seine Smartwatch an, koppelt es an sein Smartphone und authentifiziert sich mit der Rolle „Schlosser “im System. Werner arbeitet für ein Unternehmen, welches externe Dienstleistungen, einschließlich Maschinenwartung, anbietet. Er hat einen Auftrag erhalten und trifft seinen Ansprechpartner Frank in dessen Büro. Dort erhält er eine kurze Sicherheitseinweisung, eine Smartwatch und ein Smartphone. Die Smartwatch ist bereits mit dem Gerät verbunden. Das Smartphone und die Werks-App des Stahlkonzerns wurden bereits auf dem Smartphone gestartet. Werner authentifiziert sich selbst mit der Rolle „externer Schlosser “, steckt das Smartphone in Arbeitsjackentasche und legt die Smartwatch um sein linkes Handgelenk. Danach betritt er die Halle Elf und macht sich auf den Weg zu der Maschine, die er warten soll. Er wird den Sicherheitsbereich auf dem Weg dorthin verlassen. Die sicheren und unsicheren Bereiche sind im System gespeichert. Das System vergleicht wiederholt Werners Positionsdaten mit den Flächendaten und erkennt, dass Werner sich im unsicheren Bereich aufhält. Werners Uhr vibriert und er kann eine Warnmeldung lesen, zugleich wird ihm eine sichere Navigation zu seinem Ziel angeboten. Frank ist sicherheitshalber ein wenig auf Werner zugegangen, aber

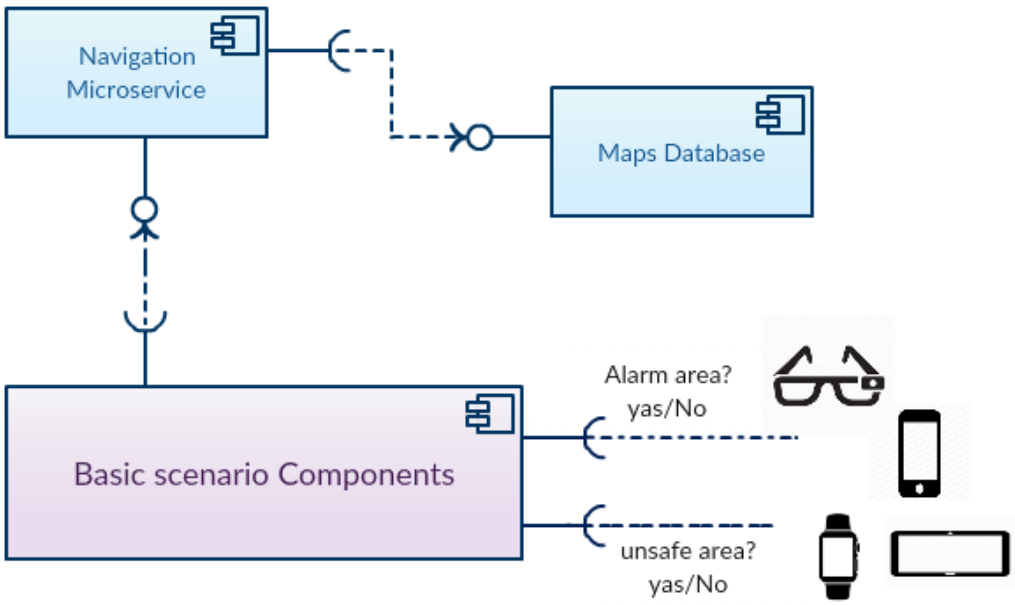


Abbildung 2: Das Arbeitssicherheitsszenario baut auf den grundlegenden Komponenten auf und nutzt zusätzlich eine Navigationsdienst mit einer entsprechenden Kartendatenbank.

bekommt dann vom System die Gewissheit, dass der „externe Schlosser “den unsicheren Bereich bereits verlassen hat.
Abbildung 2 zeigt die vorgeschlagenen abstrakten Komponenten für dieses Arbeitssicherheitsszenario. Die Gerätekomponente repräsentiert die Benutzeranwendung, die in einem Smart Device des Benutzers verwendet wird. Diese Anwendung sammelt Daten von externe Sensoren, beispielsweise Beacons und Wi-Fi. Mit Hilfe dieser Daten und des Ortungsdienstes kann der Mitarbeiter innerhalb der Fabrik den aktuellen Kontext erhalten, die Komponenten des Kontextmodells bieten die erforderlichen Informationen. Die Benutzeranwendung zeigt dem Mitarbeiter diese Informationen an und stellt die erforderlichen Dienste in Abhängigkeit der Rolle des Benutzers zur Verfügung. Beispielsweise wird ein Navigationsdienst aufgerufen, wenn der Mitarbeiter in einem sicherheitskritischen Bereich befindet, und eine Warnmeldung an eine verantwortliche Person gesendet.

3.2 Produktion

Frank nimmt ein Smartphone zur Blechinspektion und geht zum Kühlbett. Er hat von einem seiner Kollegen die Information erhalten, dass ein Stahlblech von einer seiner Maschinen nicht richtig verarbeitet wurde und möchte sich dieses Blech genauer ansehen. Bei dem Stahlblech angekommen, hält er zunächst die Wärmebildkamera seines Smartphones auf die Platte. Das System ermittelt die Temperatur des Bleches, überträgt diese Information zum Kontextmodell, und informiert Frank mit einer Warnung, dass dieses Stahlblech zu heiß zum Betreten ist. Da Frank das Blech zur Zeit nicht im Detail prüfen kann, geht er zurück zur Maschine. Er markiert das Blech als nicht bearbeitet, um es später zu untersuchen.
Abbildung 3 zeigt die vorgeschlagenen Softwarekomponenten für die Belange dieses Produktionsszenarios. Die Kontexmodellkomponenten erhalten den Standort des Mitarbeiters, wie im vorherigen Szenario, sowie die Informationen über die Bleche, also deren Standort, Temperatur, Material, Wärmebild, Zusatzinformationen und dergleichen. Durch den Vergleich von Blechdaten und Mitarbeiterdaten kann das Kontexmodell dem Mitarbeiter wesentliche Informationen anzeigen, beispielsweise wie heiß das Blech momentan ist.

3.3 Wartung

Als Frank an der zu wartenden Maschine ankommt, bemerkt er, dass Öl aus der Maschine austritt. Er vermutet, dass keine Gefahr für andere Mitarbeiter besteht, ist sich jedoch nicht hundertprozentig sicher. Deshalb markiert er den Bereich um die Maschine und auch einen Teil der benachbarten Maschinen auf seinem Smartphone als unsicher. Er erhält dann auf seinem Smartphone die Nachricht, dass sich in diesem neuen unsicheren Bereich noch ein externer und ein interner Mitarbeiter befinden.
Frank überprüft die Maschine, indem er sein Smartphone vor das Gerät hält. Auf seinem Smartphone wählt Frank aus den verfügbaren Maschinendaten die für seine Überprüfung wichtigen Daten aus. Frank bemerkt, dass aufgrund des Ölmangels ein Maschinenteil bereits eine hohe Temperatur hat. Er stoppt die Maschine, damit sie nicht beschädigt wird. Frank überprüft nun, woher das Öl kam und sieht, dass er den Fehler leicht selbst beheben kann. Er schließt das Leck, startet die Maschine neu und gibt den Sicherheitsbereich um die Maschine herum wieder frei.

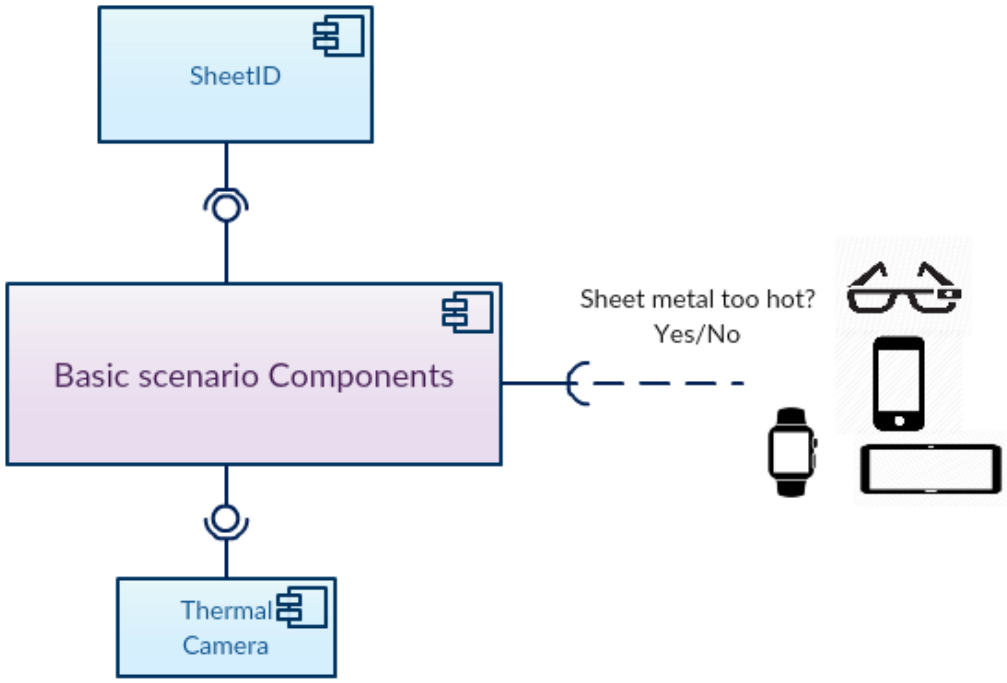


Abbildung 3: Das Produktionsszenario nutzt grundlegenden Komponenten integriert eine Datenbank über die Bleche und Aufnahmen von Wärmebildkameras.

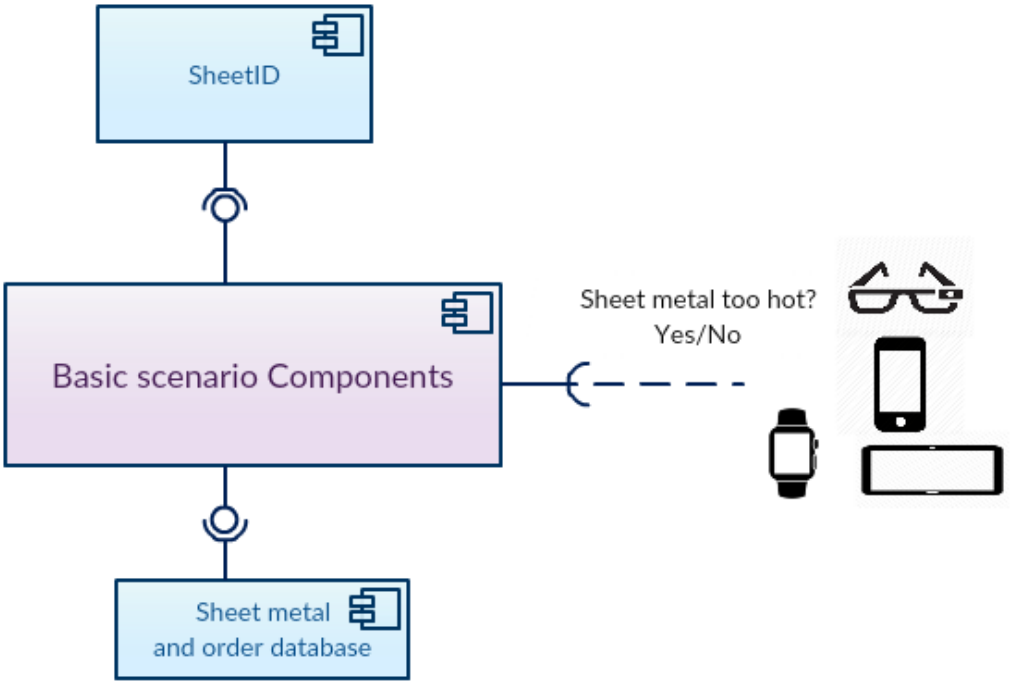


Abbildung 4: Im Lagerverwaltungsszenario werden die grundlegenden Komponenten integriert, sowie eine Datenbank über die Bleche und Lagerbestände eingebaut.

3.4 Lagerverwaltung

Nachdem seine Maschine viel Öl verloren hat, öffnet Frank die Lagerverwaltungs-App auf seinem Smartphone und schaut nach, wie viel Öl für diese Maschine noch im Lager ist. Basierend auf der App sieht er, dass mehrere Fässer zur Verfügung stehen. Er geht ins Lager und holt ein Fass, um den Ölstand in der Maschine aufzufüllen. Das Lagerverwaltungssystem ermittelt, dass ein Fass aus dem Lager entnommen wurde und berechnet, dass immer noch genügend Öl vorhanden ist und nicht nachbestellt werden muss.
Frank holt dann das mobile Tablet für die Blechinspektion und geht zum Kühlbett. Nach einer Wärmeprüfung mit Hilfe des Systems betritt er einen sicherheitskritischen Bereich und das System zeigt ihm automatisch an, auf welchem Blech er steht. Er zeigt sich die Auftragsliste an und dann ermittelt den zu diesem Blech gehörigen Auftrag. Abbildung 4 zeigt das Diagramm der vorgeschlagenen Softwarekomponenten für das Lagerverwaltungsszenario.

4 Versuchsaufbau

Auf Grundlage der ermittelten Anwendungsfälle wird an der Hochschule Ruhr West ein Versuchsablauf errichtet. Als Versuchsstätte dient das kleine Technikum der Hochschule Ruhr West auf dem Campus Bottrop. Dieser Versuch dient zur Evaluation der AR-Funktionalitäten, Personenverfolgung über Kamerabilder und Positionsbestimmung über Beacons sowie der psychologischen Auswertung anhand von Fragebögen. Zur kontextsensitiven Datendarstellung in einer simulierten Industriehalle wird die Microsoft HoloLens AR-Brille zur Innenraumnavigation verwendet. Zur Innenraumnavigation werden zwei unterschiedliche Benutzeroberflächen in der AR-Umgebung eingesetzt. Eine Navigationsart blendet einen farbigen Pfeil ein, die andere Navigationsart mit eine Route zwischen den Wegpunkten. Die Probanden laufen einen vordefinierten Weg ab, der sie zu zwei Stopps an Maschinen führt, wo sie jeweils eine kleine Aufgabe erledigen, gefolgt von einem Gefahrenhinweis und einer Evakuierungssequenz. Vor und nach diesem Parcours füllen die Probanden einen Fragebogen aus, auf dessen Grundlage die psychologischen Einflüsse der verschiedenen Navigationsarten untersucht werden.

4.1 Ablauf

Im Versuchsablauf wird den Probanden zunächst eine kurze Einweisung in die Funktionsweise der Datenbrille gegeben. Nach Aktivierung der Kameras leitet die Brille die Probanden zur ersten Station an der Maschine 1. Dort wird als Aufgabe die Aufforderung eingeblendet, einen Stecker in eine vorgesehene Steckdose zu stecken und sich die Nummer dieser Steckdose zu merken. Sobald diese Informationen angezeigt wurden, läuft eine Zeitschaltuhr von zehn Sekunden herunter, welcher nach Ablauf die Anzeige beendet. Nach Ablauf der Zeitschaltuhr wird die Person zur nächsten Station geleitet, also zur Maschine 2. An dieser Maschine ist ein Fingerabdruckscanner aufgebaut, an welchem die Probanden eine simulierte Authentifizierung durchführen sollen. Des Weiteren ist dort ein Notebook aufgestellt, an welchem Eingaben bezüglich der vorher angezeigten Informationen getätigt werden müssen. Auf dem nun folgenden Weg befindet sich ein heißes Blech. Die Probanden werden vom System mit einem visuellen Warnhinweis in der AR-Umgebung benachrichtigt und an der Gefahr vorbei navigiert. Anschließend wird ein akustisches Warnsignal über Lautsprecher im Raum ausgegeben und eine Benachrichtigung über einen Brand wird im Sichtfeld der Nutzer eingeblendet, woraufhin die Evakuierung zum Ausgang gestartet wird. Der Versuch wird beendet, sobald die Probanden am Ausgang angelangt sind. Nach dem Versuch werden die Probanden gebeten, einen weiteren Fragebogen zu beantworten.

4.2 Auswertung

Die vergleichsweise geringe Teilnehmerzahl von 52 war ausreichend für die psychologischen Auswertungen. Das Durchschnittsalter betrug 28, 9 Jahre mit einer Standardabweichung von 4.3 Jahren. Alle Teilnehmer haben die erste Aufgabe, nämlich das Abgleichen der Seriennummer und das Stecken des Steckers, erfolgreich abgeschlossen. Es gab keine Beschwerden über die Präsentation des Textes. Die zweite Aufgabe wurde nur teilweise abgeschlossen. Der Fingerscanner für die Authentifizierung stand zwar direkt neben dem Arbeitsplatz, an dem die zuvor erworbene Seriennummer eingegeben werden sollte, jedoch konnte kein Teilnehmer den Scanner finden, obwohl er mit einem großen Schild ausgestattet war. Die meisten von ihnen gaben an, sie hätten gedacht, die Säule mit dem Scanner sei Teil des Arbeitsplatzes selbst. Ungefähr die Hälfte der die Probanden haben die Eingabe der Seriennummer in den Computer abgeschlossen. Diejenigen, welche diese Aufgabe nicht erfüllten, gaben an, durch die Überlappung der Projektion des virtuelle Seriennummer und der Eingabemaske gestört gewesen zu sein oder nicht genügend Zeit gehabt zu haben, um die Aufgabe zu erledigen.

5 Kontextmodell

Das Kontextmodell im Projekt *DamokleS 4.0* dient als Beispiel für eine Softwarearchitektur, welche in der Lage ist, Daten verschiedener Quellen zu speichern und auf Anfragen eine entsprechend kontextualisierte Antwort auszugeben. Im Bereich der mobilen Geräte bieten die Sensoren, beispielsweise Beschleunigungssensoren, Gyroskopsensoren, Bluetooth und WiFi die Möglichkeit, Informationen über die aktuelle Situation des Benutzers des mobilen Geräts zu sammeln. Diese Informationen helfen, den aktuellen Kontext des Benutzers für die Empfehlung relevanter Informationen zum jeweiligen Zeitpunkt und Ort zu bestimmen [49]. Nach der Definition von Dey [12] variieren die Aspekte der Kontextinformationen je nach Anwendung der Kontextualisierung. Um beispielsweise den Kontext in mobilen Anwendungen zu erkennen helfen die Sensoren, die Position und Bewegungsrichtung eines mobilen Gerätes und damit eines Benutzers zu bestimmen.

5.1 Mehrdimensionale Kontextmodellierung

Kontextbewusste Anwendungen verwenden verschiedene Ansätze zur Modellierung der relevanten Kontextinformationen. Der Ansatz des Projektteams ist ein mehrdimensionaler Kontextmodellierungsansatz (RCM) [42]. Ein mehrdimensionaler Ansatz bietet Vorteile im Vergleich zu anderen Ansätzen, in

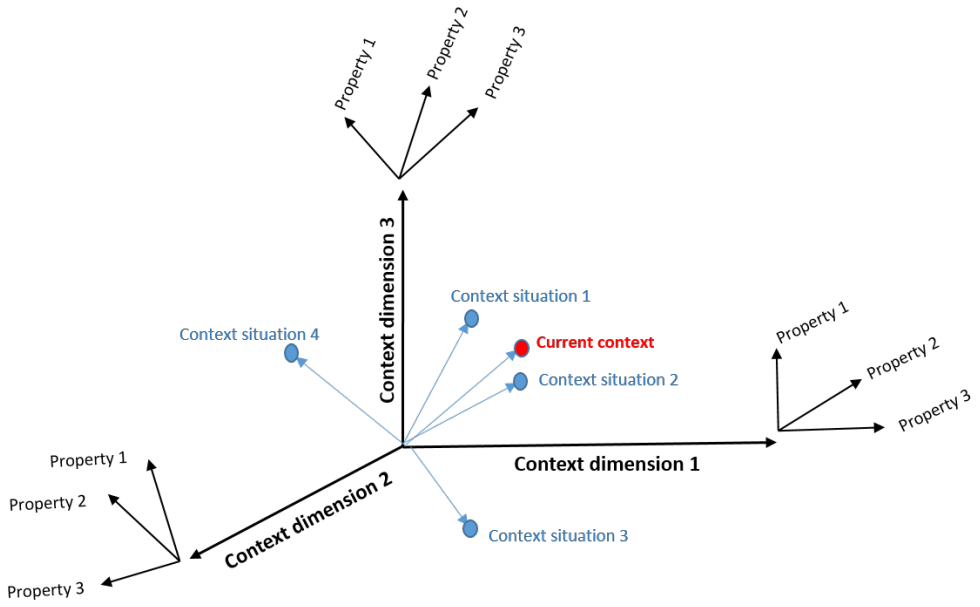


Abbildung 5: Darstellung eines mehrdimensionalen Kontextmodells [21].

Bezug auf die Erfüllung der Anforderungen an Datenreichtum, Flexibilität, Granularität und Leistungsfähigkeit [49]. In diesem Ansatz kann jede Dimension unterschiedliche Eigenschaften einer Entität repräsentieren, beispielsweise einen Gerätekontext, einen Umgebungskontext sowie eine persönlichen Kontext, wobei verschiedene Kontextsituationen als Vektoren in einem mehrdimensionalen Raum repräsentiert werden. Das Modell erkennt den aktuellen Kontext aufgrund seiner Ähnlichkeit mit diesen Vektoren. Abbildung 5 skizziert eine mehrdimensionale Kontextmodellierung.

Das RCM kann einen Bezug von den Sensordaten auf die aufgabenbezogenen Inhalte herstellen, in diesem Fall etwa Informationen über den aktuellen Status einer Maschine. Der reichhaltige Kontextbegriff wird verwendet, um die bereitgestellten Daten zu beschreiben, welche durch verschiedene Sensoren auf mobilen Geräten und möglicherweise externen Quellen wie Webdienste, angereichert werden können. Die kontextbezogen Informationen, die hier verwendet werden, beziehen sich auf die verschiedenen Objekte innerhalb der Fabrik, beispielsweise den Objekttyp, den Standort, die Rolle und den Zustand beziehungsweise die Situation. Zur Erkennung der konkreten Kontexte werden die Einsatzmöglichkeit von Bluetooth- oder WiFi-Beacons zur Lokalisierung von Mobiltelefonen in Fabrikhallen sowie Methoden der Bildverarbeitung und

der künstlichen Intelligenz genutzt.
Die Projektarbeit basiert auf der Frage, wie neue Technologien in der Praxis funktionieren, insbesondere wie im Bereich der Schwerindustrie mobile Geräte eingesetzt werden können, um kontextbasierte und intelligente Unterstützung der Mitarbeiter in den Bereichen die Fertigungsindustrie vor dem Hintergrund der steigenden Anforderungen bei der Planung, Produktion und Wartung von Prozessen zu realisieren.
In dieser Forschung müssen verschiedene Aspekte berücksichtigt werden. Zum Einen muss eine zuverlässige Kontexterkennung gewährleistet sein. Die Grundvoraussetzung dafür ist die robuste Erkennung des jeweiligen Kontextes, also auch im Falle fehlerhafter Sensoren. Des Weiteren muss eine sichere und zuverlässige Datenübertragung gewährleistet sein. Die Dämpfung von elektromagnetischen Wellen durch Geräte, Materialien oder Interferenzen machen es notwendig, Lösungen zu entwickeln, die eine sichere und ununterbrochene Datenübertragung zwischen dem Server und den mobilen Geräte gewährleisten. Zu diesem Zweck wird eine Softwarelösung entwickelt, die ein zuverlässiges Umschalten zwischen der Nutzung von WiFi-Beacons und Mobilfunknetz zur Ortung der Personen erlaubt. Neue Produktionskonzepte werden die Anforderungen an die Interaktion zwischen Mensch und Maschine ändern. Aufgrund der Notwendigkeit einer flexiblen und transparenten Produktion werden mehr Aktivitäten mit höherer Wertschöpfung und Kreativität erforderlich. In dieser Arbeit werden mobile Geräte und Wearables, konkret AR-Brillen, Armbanduhren und Smartphones zur Interaktion vorgeschlagen. Aufgrund der Kontextabhängigkeit werden die Daten vorverarbeitet und mit parallelen Rechenmethoden ausgewertet. Im Gegensatz zur statischen Nutzung von Desktop-Systemen bieten mobile Geräten und Wearables neue, dynamische Interaktionsmöglichkeiten. Zum Beispiel durch das Auswerten der Blickrichtung des Benutzers auf eine Maschine können entsprechende Daten je nach Standort und Blickrichtung dargestellt werden. Mit dem Aufkommen der IT in den Produktionsprozessen rückt der Aspekt der sicheren Datenübertragung, und auch des Datenschutzes, mehr denn je in den Vordergrund. Es ist insbesondere wichtig für Produktionsunternehmen, ihr eigenes Wissen gegen unbefugten Zugriff zu schützen. Gleichzeitig soll ein möglichst breites Spektrum von Anwendungen bereitgestellt werden, welches alle unterschiedlichen Anforderungen berücksichtigt. Spezifikationen auf der Benutzerseite und die verwendeten Systemkomponenten sollten Geräte- und Betriebssystemunabhängig sein. Die drei für die Anwendung wichtigsten Geräteplattformen, Android, iOS und Windows, erhalten besondere Aufmerksamkeit. Die Bibliotheken und Frameworks, die für die Entwicklung verwendet werden, sollten Open Source sein. Die Verwendung offener Standards stellt auch die Wartbarkeit und nachhaltige Nutzbarkeit sicher.

5.2 Cyberphysikalisches System

Getrieben durch das Internet der Dinge, das Internet der Dienste und der Industrie 4.0 nimmt die Bedeutung des Einsatzes mobiler Geräte auch in industriellen Anwendungen zu [34]. In cyberphysikalischen Systemen (CPS) spielt das Kontextbewusstsein eine wichtige Rolle für die Anpassung von Softwareanwendungen im Einklang mit dem aktuellen Kontext innerhalb der Zieldomäne mit einem Minimum an menschlicher Arbeit [34, 66]. Dieses Wissen macht es notwendig, neue Technologien auszuwählen und deren Einsatz im Hinblick auf technische Eigenschaften, Wirtschaftlichkeit, zukünftige Entwicklung, neue Möglichkeiten und Nachhaltigkeit zu bewerten. Bei diesen Untersuchungen stellte sich heraus, dass eine webbasierte Darstellung am Besten geeignet erscheint. Diese Technik ermöglicht den Einsatz auf Desktop-Systemen sowie auf mobilen Geräten. Der Einsatz von webbasierten Clients macht auch eine Erweiterung der Systemarchitektur notwendig. Abbildung 6 zeigt die vom Partner Celano bereitgestellte celCAP 4.0 Infrastruktur. Die klassische Client-Server-Architektur wird durch eine zusätzliche Schicht mit der Middleware (MW) erweitert. Die neue Struktur celCAP 4.0 bildet damit die Grundlage für die Integration von webbasierten Clients, Desktop-Systemen sowie mobilen und intelligenten Geräten.

Neben dem Einsatz von Smartphones oder AR-Brillen für die Navigation ist auch die Verwendung von Smart Watches möglich. Neue Anwendungen für eine Smart Watch auf Basis der Anwendungsfälle, die in den Szenarien beschrieben werden, wurden entwickelt und in das celCAP 4.0-Framework integriert. Abbildung 7 zeigt das Mockup für eine Smart Watch Nutzeroberfläche. Frei konfigurierbare Benachrichtigungen, beispielsweise über den Zutritt in einen unsicheren Bereich, können über Backend-Prozesse an die Smart Watch gesendet werden und den Benutzer per Alarm, beispielsweise ein lauter Ton oder Vibration, über die aktuelle Situation informieren. Die intelligente Uhr kann aber auch dazu verwendet werden, Informationen über den aktuellen Zustand von Maschinen und Produktionsprozessen zu liefern. Diese Informationen werden in Form von Prozesswerten in den Backend-Prozessen des celCAP-Frameworks gespeichert. Um diese Informationen auf der Smart Watch auf intuitive Weise darzustellen, wurde eine neue Smart Watch Benutzeroberfläche entwickelt. Darüber hinaus wurde eine neue Kommunikationsbibliothek integriert, um den Datenaustausch zwischen der Smart Watch, dem zugehörigen Smartphone und der Middleware einschließlich des Kontextmodells zu ermöglichen, wie in Abbildung 8 dargestellt. Mit diesem Konzept können Prozessinformationen abhängig von den aktuellen Situationen, wie beispielsweise aktuellen Ofentemperaturen oder Lebensdauerstunden

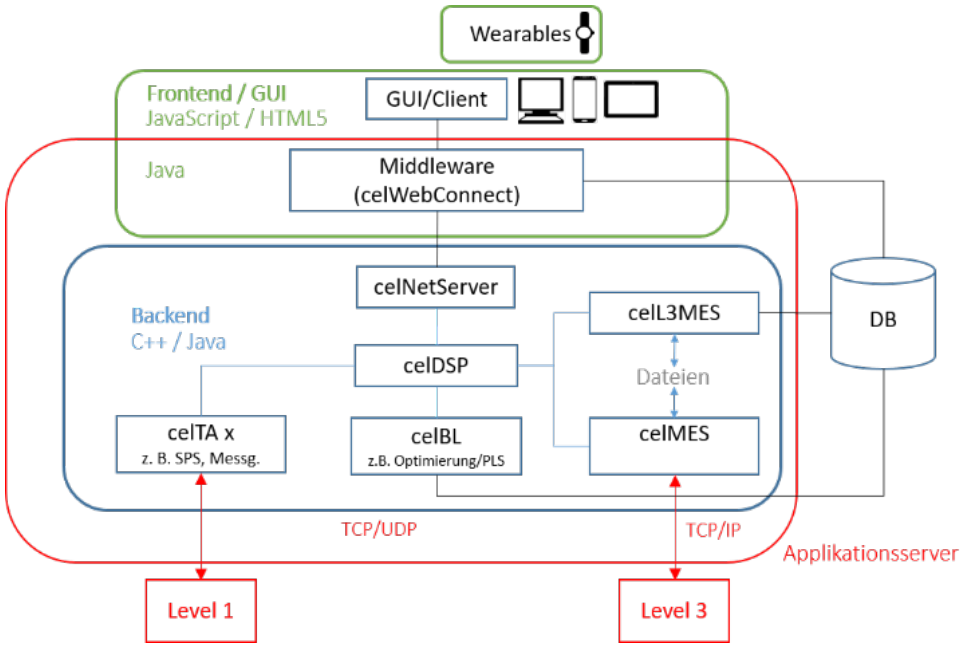


Abbildung 6: Die Software celCAP 4.0 tauscht Daten mit Fremdsystemen aus und kombiniert diese mit eigenen Informationen, wie beispielsweise dem Kontext, um sie für mobile Anwendungen und Wearables zur Verfügung zu stellen [21].

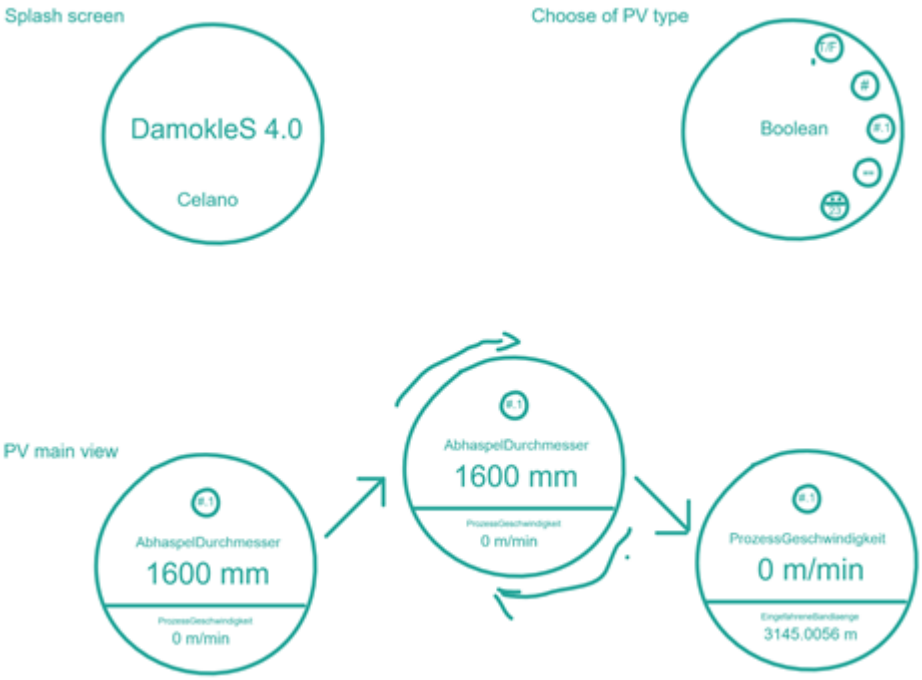


Abbildung 7: Ein Mockup für die Nutzeroberfläche einer beispielhaften Smart Watch Anwendung [21].

von Maschinenteilen, sortiert werden. In Verbindung mit dem Konzept eines zweiten Bildschirms, skizziert in Abbildung 9, können Bediener in Kontrollzentren oder in der Produktion solche Geräte auch als Fernbedienung verwenden, um etwa eine entfernte Desktop-Benutzeroberfläche zu steuern. Die Geräte können verwendet werden, um die Desktop-Benutzeroberfläche zu erweitern, um mehr Inhalte anzuzeigen, die sich auf die aktuelle Situation beziehen, um Benachrichtigungen anzuzeigen oder um in eine detailliertere Prozessansicht zu zoomen. Die Kopplung erfolgt hier über Barcodes, die auf dem Desktop-Bildschirm angezeigt und mit dem Smartphone gescannt werden.

Zur Anbindung an ein CPS werden verschiedene, kleine Dienste (E: *microservice*) in nodeJS und Angular CLI auf Grundlage von JSON-Dokumenten implementiert. Das Kontextmodellierungswerkzeug [38] erlaubt es, Kontextmodelle zu erstellen und zu bearbeiten. Der zugrundeliegende Vektorraumdienst führt eine entsprechende Datentransformation durch und berechnet die Ähnlichkeit von kontextsensitiven Situationen mit einem Vektorabstand [41]. Ein Konfigurationsdienst erlaubt eine Änderung am Kontextmodell nach Authentifizierung mit einer entsprechenden Benutzerrolle [36]. Ein Datensammeldienst

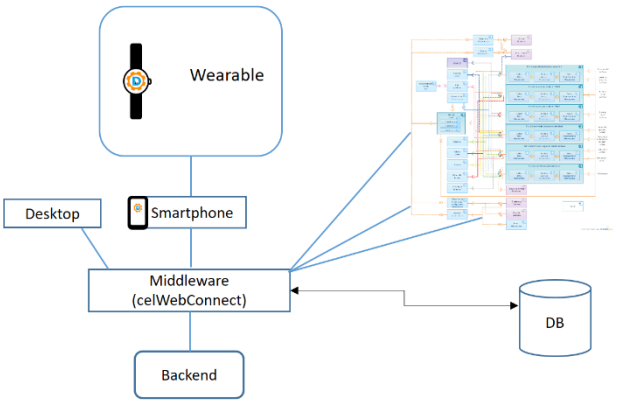


Abbildung 8: Die Kommunikation zwischen Smart Watch und Kontextmodell erfolgt über ein zwischengeschaltetes Smartphone [21].

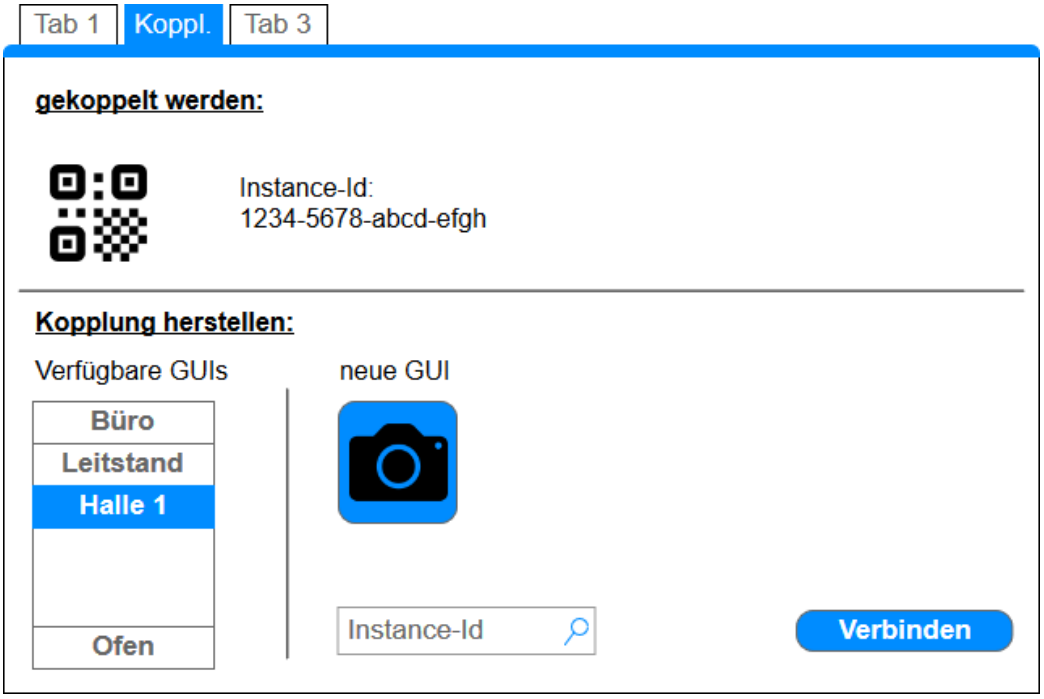


Abbildung 9: Eine beispielhafte Kopplung zur Nutzung eines Smartphones als zweiten Bildschirm für einen Desktop-PC [21].

verwaltet die anfallenden Daten im JSON-Format in einer mongodb [37]. Über eine Ortungsdienst kann, unter Angabe einer Nutzerrolle, die Position von wichtigen Gegenständen in der Umgebung mitsamt einer zugehörigen Textinformation abgefragt werden [40]. All diese microservices können über ein Dockersystem migriert werden [39].
So ist die Innovation von neuen, intelligenten Ansätzen im Bereich der Kontexterkennung ein effektiver Beitrag zur Entwicklung intelligenter und benutzerfreundlicher CPS. Darüber hinaus wird es für neue CPS notwendig sein, die neuesten Softwaretechnologien zu integrieren und aktuelle Hardware zu verwenden, um die gesammelten Daten und Dienste optimal zu nutzen. Mit steigender Leistung und Funktionalität von mobilen Geräten und Sensoren werden kontextbasiere CPS immer effektiver darin, die Benutzer mit mehr Flexibilität zu unterstützen sowie genaue und aussagekräftige Informationen über den aktuellen Stand der Dinge zu liefern [48].

5.3 Positionsbestimmung über Beacons

Ein Beacon ist ein Sender oder Empfänger, der auf Bluetooth oder WiFi basieren kann. Ihre Reichweite beträgt etwa $50m$. Die Datenübertragung zwischen den Geräten verbraucht deutlich weniger Energie als beim Vorgänger Bluetooth, was beispielsweise zu einem geringen Batterieverbrauch bei Smartphones führt. Auch die Transferkosten bleiben niedrig. Große Datenmengen sind für den Austausch allerdings nicht geeignet, da die Übertragungsrate relativ gering ist. Die Beacons selbst werden mit Batterien, oder seltener auch mit direktem Stromanschluss betrieben und an verschiedenen Stellen im kleinen Technikum der Hochschule Ruhr West am Campus Bottrop installiert und auf den selben Start- und Endpunkt geeicht wie die AR-Anwendung und die Kamerakalibrierung. Auf diese Weise ist es praktisch möglich, ihre Position jederzeit zu bestimmen und an den. Für die Untersuchungen zur Navigationsfähigkeit im Projekt *DamokleS 4.0* wird zur Ortung eine Trilateration nach [55] verwendet. Trilateration ist eine Geometrietechnik, um eine Positionen basierend auf den relativen Abständen zu drei Referenzpositionen zu berechnen. Der Trilaterationsansatz arbeitet mit mindestens drei Beacons und die Signalstärke muss stabil sein. Das Problem der Instabilität der Signale konnte für diesen Versuchsaufbau nicht gelöst werden. Auch mit entsprechenden Filtern konnten die Position zuverlässig aus den Daten gewonnen werden. Wie in Abbildung 10 gezeigt, wird die Information aus den Beacons in dieser Umgebung derart verzerrt, dass keine zuverlässige Personenortung möglich erscheint.

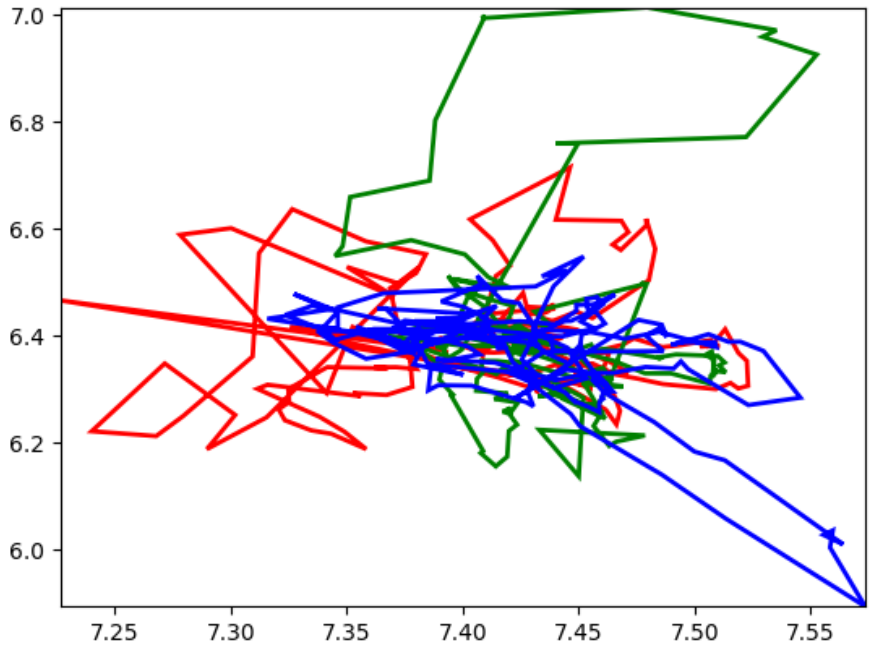


Abbildung 10: Ein Beispielbild für die gescheiterte Personenortung über Beacons im kleinen Technikum der Hochschule Ruhr West am Campus Bottrop mit drei Testpersonen (rot, grün, blau).

Abbildung 11: Beispielhafte positionsabhängige Augmented Reality Information [4]. Links: Die Arbeitsanweisung, eine Seriennummer zu prüfen. Rechts: Die Darstellung einer Evakuationsroute zum Ausgang.

6 Augmented Reality Anwendung

Ein wichtiger Kernbeitrag des Projektes ist die Augmentet Reality (AR) Anwendung. Sie existiert in zwei verschiedene Benutzerobeflächenausführungen, erledigt die Kalibrierung im Raum anhand eines angebrachten Markers und zeigt 2D- sowie 3D-Inhalte an, abhängig von Position und dem Blickfeld des Nutzers. Insgesamt besteht der erweiterte Inhalt aus Textprojektion, Wegpunkt-Symbolen, den Maschinen und dem Gefahrenbereich sowie einer Evakuation. Abbildung 11 zeigt beispielhaft zwei kontextualisierte Situationen aus der AR-Umgebung.

6.1 Entwicklungsprozess

Bei der Gestaltung der Anwendung wurden die Vorrecherchen als Richtlinien herangezogen, um ein effizientes System ohne störende Einflüsse zu gestalten [6]. Teile der Microsoft Richtlinien für Mixed-Reality-Design wurden auch in Kombination mit der ISO-9241 Norm für ergonomische Anwendungen eingesetzt, welche wichtige Aspekte wie die visuellen Darstellungen, auditiven Rückmeldungen und Benutzerinteraktionen mit dem System umfasst [27, 14, 28]. Die Farben wurden hell genug gewählt um in einer industriellen Umgebung gesehen zu werden und auch bei schlechten Lichtverhältnissen einen ausreichenden Kontrast für Anwender mit einer Sehbehinderungen oder Farbenblindheit zu gewährleisten. Es wurden zwei verschiedene Navigationsmethoden zur Auswertung ausgewählt. Die erste besteht aus der Darstellung eines Pfeils, der die Richtung eines aktuellen Wegpunktes zeigt. Die zweite Methode zeichnet eine 3D-Linie auf den Boden, welche die Benutzer den optimalen Weg entlang führt. Diese beiden Techniken wurden ausgewählt, da

sie den traditionellen Navigationssystemen in der Automobilindustrie oder Smartphones ähneln. Die Benutzeraufgaben, die Ikonographie und die Texte waren in beiden Anwendungen identisch und wurden mit Unity 3D und Csharp erstellt.
Aufgrund der Schwierigkeit, ein kontrolliertes Experiment in einer realen industriellen Umgebung zu starten, wurde, wie bereits beschrieben, das kleine Technikum der Hochschule Ruhr West am Campus Bottrop als simulierte Fertigungshalle eingerichtet. Dazu gehören eine reduzierte Beleuchtung, laute, industrietypische Umgebungsgeräusche und verschiedene andere Requisiten, die als Maschinen fungieren und den Teilnehmern den Eindruck einer realen Industrieumgebung vermitteln. Die AR-Anwendung wurde entwickelt, um die Teilnehmer durch das Testgelände zu führen und sie während des Prozesses zu unterstützen. Der Fuktion folgend soll die Anwendung so gestaltet werden, dass sie die Möglichkeit der Indoor-Navigation beinhaltete [28].

6.2 Navigationsarten

Die Darstellung der die Benutzeroberfläche der Navigation wurde in zwei verschiedenen Layouts erstellt, der pfeil- und der routenbasierten Navigation. Beide Navigationsarten haben das selbe Entfernungsetikett und auch die gleichen allgemeinen Anweisungen, die den Teilnehmen während des Versuchs präsentiert werden. Bei der pfeilbasierten Navigation bestand das Layout hauptsächlich aus einem dreidimensionalen Pfeil, der einem Kompass ähnelt und in Richtung des nächsten Ziels zeigt. Der Pfeil war farblich markiert; grün für die richtige und rot für derer entgegengesetzte Richtung, mit einem glatten Übergang zwischen den beiden Zuständen. In der routenbasierten Navigation hingegen zeigt das System eine Navigationsroute, die, basierend auf der die eigene Körpergröße des Benutzers, dreißig Zentimeter unter dessen Augen schwebt, wie in Abbildung 12 gezeigt. Dieser Ansatz soll eine gut sichtbare Navigationslinie bieten, ohne die Teilnehmer zu behindern. Die Projektion des Weges auf das Erdgeschoss der realen Umgebung würde einen großen Teil davon außerhalb des Darstellungsbereichs der HoloLense positionieren. Die angezeigte Linie zeigte einen direkten Weg zum nächsten Ziel und wies auch den Weg zum übernächsten Ziel an.

6.3 Simulierte Umgebung

Während der Studie werden die Teilnehmer zu zwei Arbeitsplätzen navigiert. Die Arbeitsaufgaben werden durch die Projektion textbasierter Aufgaben in das Sichtfeld des Benutzers dargestellt. Die erste Aufgabe erfordert, dass der Benutzer eine projizierte Seriennummer mit der einer realen Maschine

Abbildung 12: Ein Beispielbild aus der Routennavigationsanwendung der Augmented Reality Umgebung.

vergleicht und einen Stecker an die richtige Steckdose anschließt. Die zweite Aufgabe wird später eingeleitet und bittet den Benutzer, sich zunächst mit einem Fingerabdruck zu authentifizieren. Anschließend soll die angegebene Seriennummer auf einem vor den Teilnehmern platzierten, realen Computer eingegeben werden. Ein besonderes Segment des Versuchsaufbaus war der daraufhin folgende, zu vermeidende Gefahrenbereich, dargestellt in Abbildung 14. Ein simuliertes heißes Blech ist auf dem Boden positioniert und die AR-Anwendung zeigt den gefährlichen Raum um das Blech herum in Verbindung mit einer Temperaturwarnung an. Der letzte Teil des Versuchsaufbaus war die Evakuierung. An einem bestimmten Punkt im Parcours wird ein lauter, akustischer Alarm ausgelöst. Gleichzeitig wird die Anweisung gezeigt, der Navigation bis zum nächsten Ausgang zu folgen, wie in Abbildung 15 dargestellt. Dies soll einen Notfall, wie beispielsweise einen Feueralarm, simulieren. Abbildung 13 zeigt den Aufbau des Testparcours samt Position der Kameras.

6.4 Auswertung

Die Ergebnisse eines Probelaufs mit zehn Teilnehmern einer Think-aloud-Technik erlaubt eine erste Auswertung. Die Teilnehmer dieser Think-aloud-Studie erwähnen, dass die Navigation mit dem Pfeil ein großes Sichtfeld ermöglicht, das frei von jeglicher Art von virtueller Darstellung ist. Da der Pfeil relativ klein ist und einen feste Position hat, ist es einfach, die Umgebung zu sehen oder die Kopfposition einzustellen. Darüber hinaus empfinden die Nutzer die Farbcodierung des Pfeils als eine schnelle und einfache Möglichkeit, um zu beurteilen ob die aktuelle Blickrichtung korrekt ist oder nicht. Andererseits wird das 3D-Modell des Pfeils nicht immer eindeutig als ein Pfeil erkannt, der genau nach vorne zeigt und in einigen Fällen mit einem nach hinten gerichteten Pfeil verwechselt. Darüber hinaus ist der Pfeil möglicher Weise nicht ausreichend, um einfach zum Ziel zu gelangen, da der Pfeil keine genaue Route zur Position unter der Überschrift zeigt. Fast alle Probanden erwähnen, dass der Vorteil der Routennavigation in der präzisen Darstellung liegt. Die Präsentation eines exakten Weges am Boden hilft, sich zu orientieren. Insbesondere die Anzeige der anstehenden Wegpunkte wird positiv bewertet. Der wichtigste negative Aspekt ist die Tatsache, dass sich die Route oft mit einem großen Teil der realen Umgebung überschneidet, was es schwierig macht, mögliche Hindernisse zu erkennen. Viele Teilnehmer schlagen Verringerungen die Dicke der Linie bei gleichzeitiger Erhöhung der Transparenz vor, um eine höhere Qualität zu ermöglichen. Darüber hinaus war kein Indikator mehr sichtbar, der den Benutzer zurück zum aktuellen Pfad führen könnte, sobald die Blickrichtung von der Route weg gerichtet war. Dies zwingt die Teilnehmer, sich umzusehen, bis sie es selbst wiederfinden. Die AR-Geräte analysieren jede

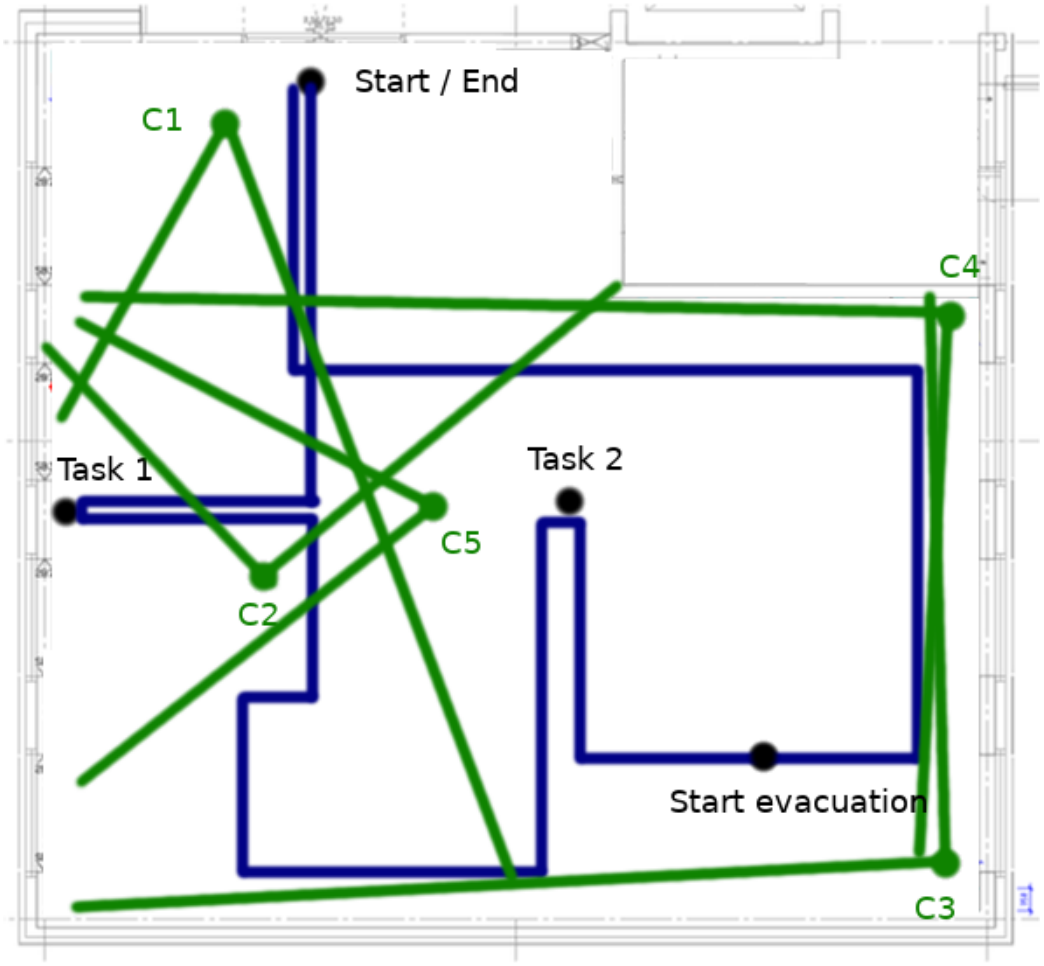


Abbildung 13: Der Testparcours in der simulierten industriellen Umgebung [4]. Die Probanden folgen einer Spur (blau) und absolvieren derer entlang eine Reihe von Aufgaben. Hierbei zeichnen fünf Kameras (grün) Videomaterial auf.

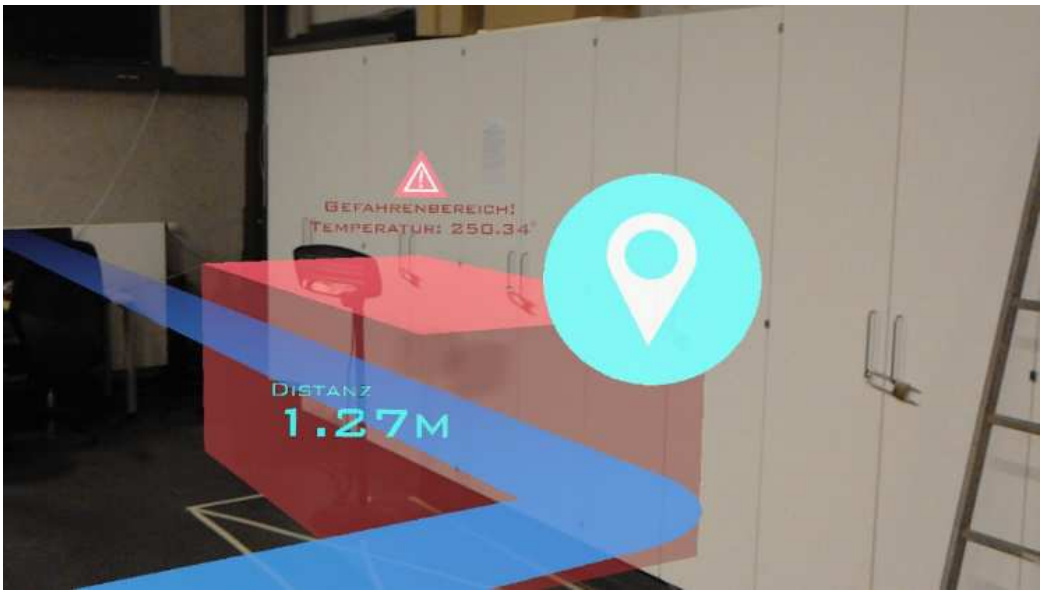


Abbildung 14: Ein Beispielbild vom Gefahrenbereich in der Augmented Reality Umgebung.

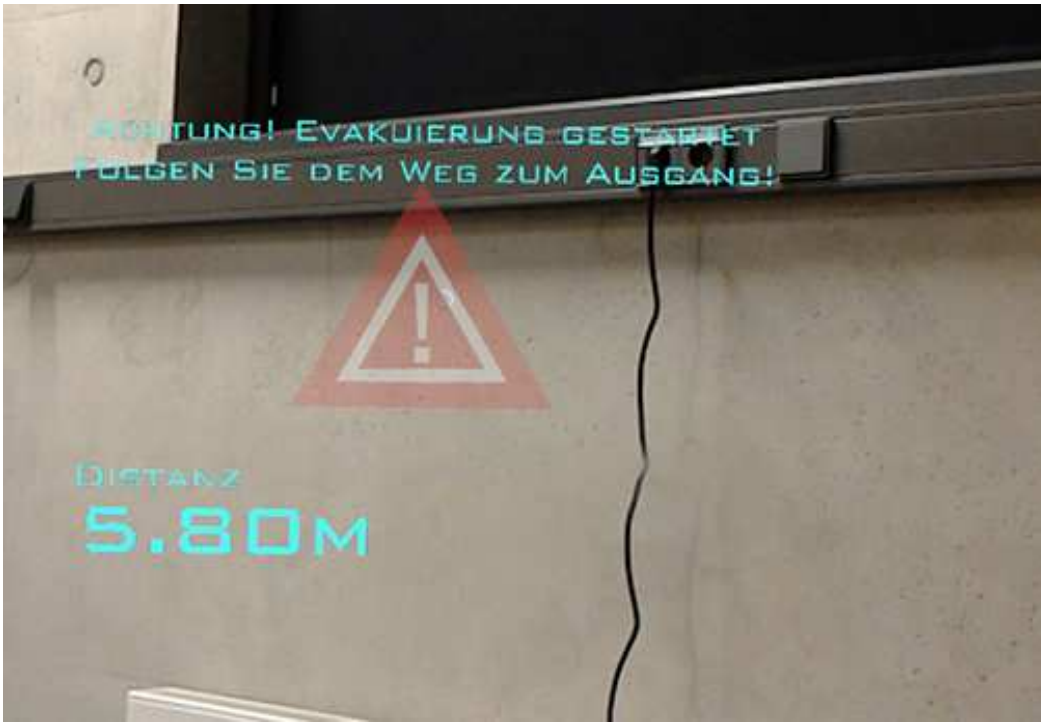


Abbildung 15: Die Evakuierungsmeldung in der Augmented Reality Umgebung.

halbe Sekunde die internen Koordinaten und speicherten die aktuelle Position. Die genaue Lokalisierung der Position und deren Aufzeichnung wird durch die Verwendung des internen Koordinatensystems des AR-Gerätes ermöglicht. Diese Daten sind wichtig für für die Auswertung der Genauigkeiten der Personenortung über Beacons und der kamerabasierten Personenverfolgung. Analysen zeigen, dass die Teilnehmer bei der Nutzung der Routennavigation den beabsichtigten Weg genauer verfolgen als im Vergleich zur Pfeilnavigation. Dieses Ergebnis entspricht den Erwartungen, da die Routennavigation genaue Wege anstelle von allgemeinen Richtungen anzeigt.

7 Kamerabasierte Personenverfolgung

Die kamerabasierte Personenverfolgung bietet ein robustes Detekions- und Lokalisierungssystem als stabile Sensoreingabe für das Kontextmodell. Das Personendetektionssystem muss unterschiedliche Anforderungen erfüllen. Erstens müssen die besonderen Herausforderungen an die digitale Bildverarbeitung in industriellen Umgebungen berücksichtigt werden. Daher müssen Störungen durch Staub, Lichtreflexionen und sich bewegende Objekte, welche die Arbeiter vor den Kameras verstecken können, antizipiert werden. Zweitens muss eine genaue Übersetzung der ermittelten Positionen aus den Bildkoordinaten in ein einheitliches Weltkoordinatensystem gewährleistet werden, was eine genaue Schätzung der Fußpunkte benötigt. Als dritte Anforderung muss das System in Echtzeit auf möglichst kostengünstiger Hardware laufen. Zur graphischen Darstellung der Personenverfolgungsmethoden wurde eine Simulationssoftware entwickelt. Abbildung 16 zeigt einen Screenshot der Simulationsumgebung, die es erlaubt, die aufgezeichneten Positionen mitsamt den zugehörigen Kamerabildern zu untersuchen.

7.1 Personenerkennungsmethoden

Die Untersuchungen fokussieren sich auf drei gut erforschte Detektionssysteme; die Merkmalsklassifikation mit Gradientenorientierungshistogrammen (E: *Histograms of Oriented Gradients*, HOG) und *Support Vector Machines* (SVM) [7, 10, 35, 52, 56, 69], das neuronale Netzwerk *You Only Look Once* (YOLO) [43, 44, 45] und die *OpenPose* Architektur (OP) [8, 47, 59]. Um die Zuverlässigkeit eines Detektionssystem zu beurteilen werden drei wichtige Qualitätsfaktoren bestimmt: die Genauigkeit in Bezug auf die Erkennung des menschlichen Körpers, die Präzision der Fußpunktlokalisierung sowie die Tendenz der fehlerhaften Erkennungen, die entweder fälschlicher Weise Personen erkennen könne, die nicht da sind, oder Personen im Bild nicht

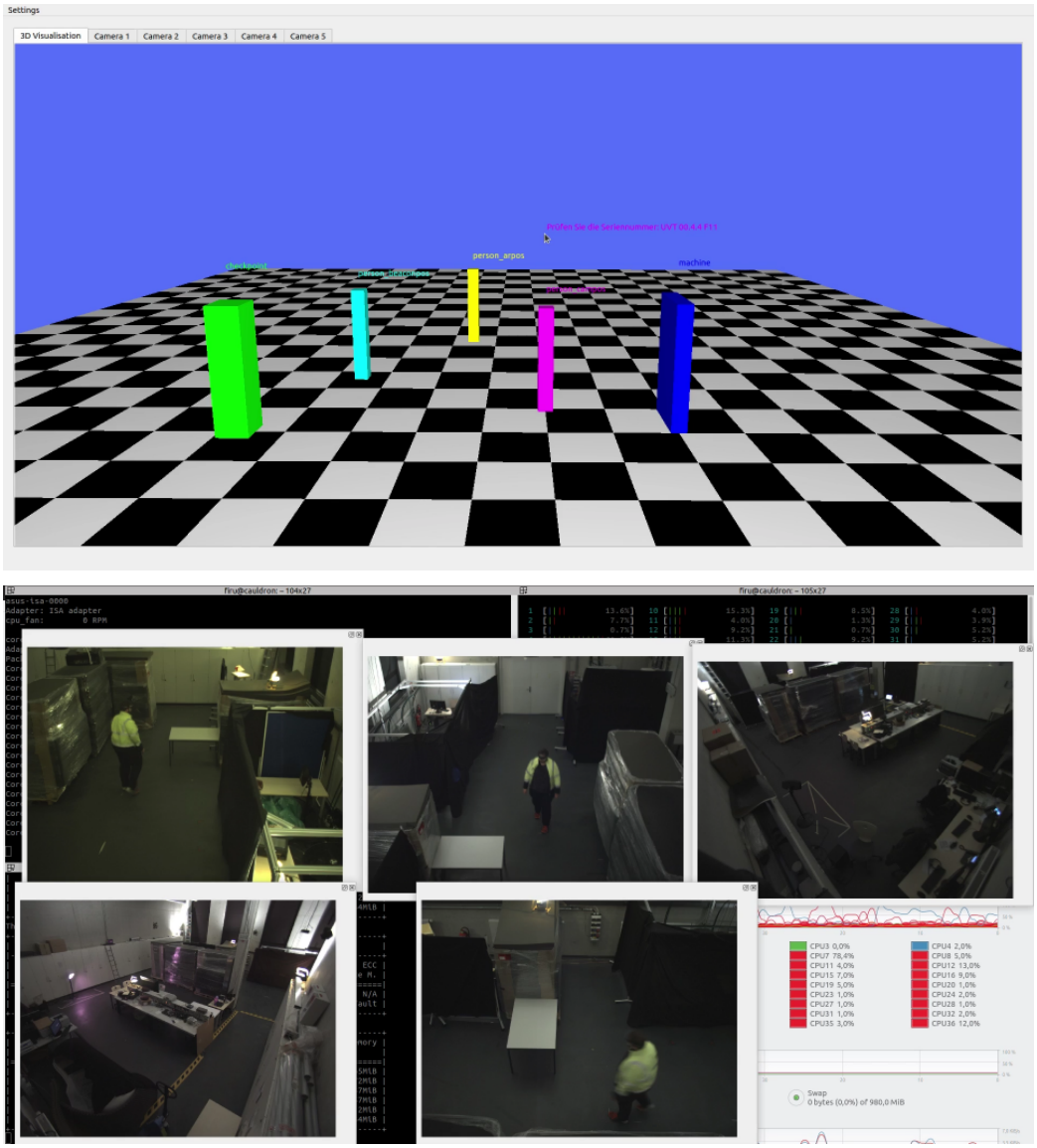


Abbildung 16: Ein Screenshot der Simulationssoftware. Oben: Darstellung der relevanten Positionen als Blöcke auf einem Schachbrettmuster. Unten: die zugehörigen Kamerabilder.

Abbildung 17: Beispielhafte Detektionen der drei untersuchten Methoden auf verschiedenen Einzelbildern aus einer simulierten Arbeitsaufgabe. Links: OpenPose erkennt ein menschliches Körperskelett auf einer, teilweise durch ein Objekt versteckte, Person. Mitte: YOLO erkennt mehrere Personen, aber verfehlt eine Person in der Menge. Rechts: HOG schätzt drei grobe Begrenzungsrahmen drei Personen in einer Menge ab [65].

erkennen können. Diese Eigenschaften werden anhand einer Reihe von Videosequenzen bewertet, die in einer ersten Vorstudie im kleinen Technikum der Hochschule Ruhr West am Campus Bottrop aufgenommen wurden, gemäß des in Abbildung 18 gezeigten Kameraaufbaus. Die aufgenommenen Videosequenzen werden programmatisch verändert, mit steigendem gauß'schen Rauschen, steigender Lichtreflektion sowie variierender Teilbedeckung von Personen im Bild durch Flächen von Salz-und-Pfeffer-Rauschen.
Abbildung 17 zeigt exemplarische Erkennungen der drei untersuchten Methoden auf ausgewählten Beispielbildern. Die Methoden HOG und YOLO zeigende mangelnde Antizipation von verdeckten Personen, so dass die Position des Fußpunktes eher ungenau bestimmt werden kann. Eine beispielhafte Positionsschätzung durch den OP für zeigt hingegen eine gute Antizipation von verdeckten Körpern. Somit kann der Fußpunkt zuverlässig lokalisiert werden, auch wenn die Person hinter einem Hindernis steht, wo die anderen Personenerkennungsmethoden lediglich den Oberkörperteil erfassen würden. Vor dem Hintergrund dieser Annahme werden die besten Ergebnisse vom OP-System erwartet, da es Modellwissen über einen menschlichen Körper beinhaltet und daher Störungen durch Rauschen, Lichtverhältnisse und Teilbedeckung robuster standhalten sollte. Ebenso erwarten Schwierigkeiten bei der Fußpunkterkennung durch den HOG-Algorithmus und die YOLO-Architektur, wenn nur der obere Körperteil sichtbar ist. Von der HOG-Methode wird eine gute Leistung unter normalen Bedingungen erwartet, aber aufgrund fehlender Generalisierung ein schneller Verfall in der Genauigkeit, während die systematischen Störungen fortschreiten.

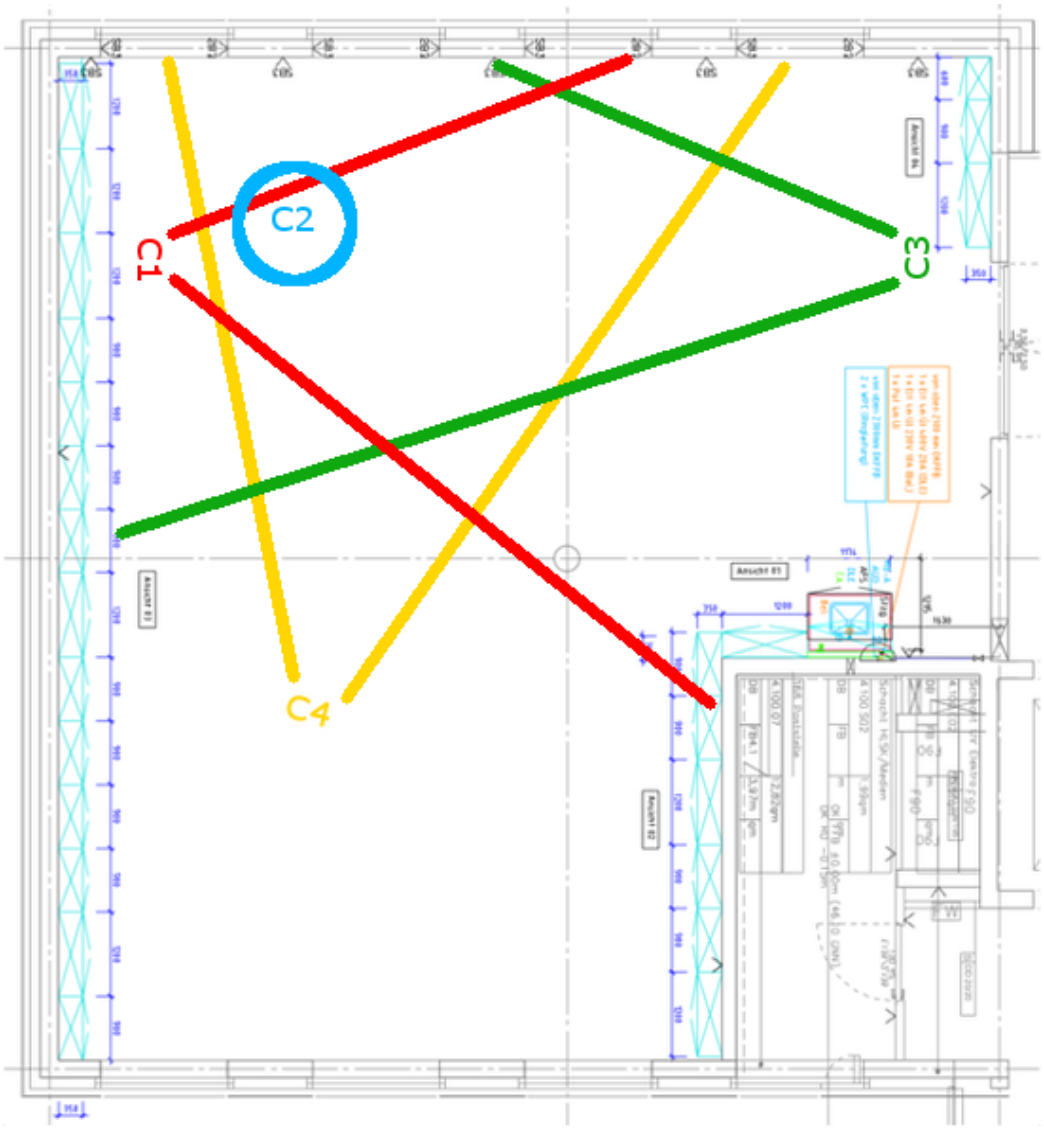


Abbildung 18: Der Aufbau der Kameras für die Testaufnahmen für den Vergleich der Personenerkennungsmethoden. Effektiv genutzt wurden hierbei nur die Aufzeichnungen von C_1 und C_2.

Die HOG-Methode liefert Merkmalsbeschreibungen eines Bildes, die prinzipiell jedes Modell des maschinellen Lernens zur Objekterkennung, oder auch zur Durchführung einer Personenerkennung, anwenden kann [10]. Nach [56] berechnet dieses Verfahren die Gradientenorientierungen für jede Zelle des in ein Gitter aufgeteilten Bildes und bindet sie in ein Histogramm. Die mit der euklidischen Norm normalisierten Zellen dienen als Merkmale für einen Klassifikator, beispielsweise einer SVM [7]. Für die Auswertung wird die OpenCV-Implementierung einer SVM genutzt [24, 54].
Das Netzwerk *You Only Look Once* (YOLO) kann Objekte verschiedener Klassen in Bildern erkennen. YOLO läuft in Echtzeit und kann Menschen sowie eine Anzahl von gelernten Objekte erkennen. Die neuronale Architektur besteht im Wesentlichen aus einer tiefen Faltungskaskade mit gelegentlichen Max-Pooling-Schichten verschiedener Filtergrößen und Pooling-Regionen. Die Ausgabeschicht von dieses Netzwerk liefert ein Gitter aus $S \times S$-Zellen. Für für jede Zelle nennt YOLO die Klassenwahrscheinlichkeit mit einem zugehörigen Konfidenzwert der umgebenden Begrenzungsrahmen [45]. Das Verfahren multipliziert die bedingte Klassenwahrscheinlichkeiten mit den einzelnen Begrenzungsrahmenkonfidenzwerten und erhält einen klassenspezifischen Konfidenzwert für jede Zelle. Die YOLO-Architektur erfährt stetige Verbesserungen [44]. Für diese Experimente wird YOLO in zweiter Version verwendet, wie es in der C++-Bibliothek *darknet* implementiert ist [43].
Die *OpenPose* (OP) Architektur schätzt die Körperhaltungen der Personen im Bild durch die Anwendung von Deep Learning mit Körpermodellwissen [8, 47, 59]. Je nach Wahl des Körpermodells (COCO, BODY25 oder MPI) ändert sich die Laufzeit und Präzision. OP schätzt zweidimensionale Posen mehrerer Personen in einem Bild, indem es mehrere Stufen von gelernten *Part Affinity Fields* (PAF) und *Part Confidence Maps* (PAC) in einem zweiteiligen Graphenabgleichsverfahren kombiniert [8]. Die Echtzeitanwendbarkeit ermöglicht es, mehrere Personen auf einem Echtzeitdatenstrom zu erkennen und diese Informationen für eine präzise Fußpunktortung zu nutzen.

7.2 Bildstörungsresistenz

Um die Zuverlässigkeit der Detektoren zu untersuchen, werden die Eingabebilder mit insgesamt vier Methoden gestört. Abbildung 19 zeigt die durch die Vorverarbeitungsmethoden induzierten Bildstörungen auf ausgewählten Einzelbildern. Um die Widerstandsfähigkeit der Methoden gegen gauß'sches Rauschen zu bewerten, wird Rauschen aus einer Normalverteilung auf die Bilddaten addiert [32]:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

.

Es wird $\mu = 0$ gesetzt und σ in einem Bereich von $[0, 250]$ mit einer Schrittweite von $\delta_\sigma = 25$ erhöht, so dass insgesamt elf Bilder mit unterschiedlichen Rauschpegel für jedes Videobild erzeugt werden.

Um die Leistung unter verschiedenen Lichtverhältnissen zu untersuchen, wird zunächst die Annahme getroffen, dass die Beleuchtungszeit der Kameras und Bildversbesserungs- und Aufhellungsverfahren ein angemessen beleuchtetes Eingabebild herbeiführen können. Zu Untersuchen bleibt die Auswirkung eines Blendlichteffekt auf einer leuchtenden Oberfläche. Hierfür wird ein individueller Blendlichteffekt für jede Kamera mit einer Stärke von 0% bis 100% in Schritten von 10% addiert, wie in Abbildung 20 gezeigt.

Zur Auswertung der Robustheit gegen teilweise Verdeckungen werden die Bilddaten in jedem Begrenzungsrahmen um die Personen in sechs Modalitäten verändert: horizontaler linker, mittlerer und rechter Teil sowie wie das vertikale Ober-, Mittel- und Unterteil. Jede Verdeckungsmodalität deckt ein Drittel des Begrenzungsrahmens entlang der jeweilige Achse mit einer Fläche aus Salz-und-Pfeffer-Rauschen ab.
Um einen Härtefall zu untersuchen, werden die vorherigen drei Störeinflüsse um einen vierten, schwierigsten Testfall kombiniert. Um eine kombinatorische Explosion zu vermeiden, werden nur elf Bilder pro Originalbild zu erzeugen, von denen jedes Einzelne aus einer Addition von Rauschen und Blendeffekt besteht. Hierbei werden Rauschen und Blendlichteffekt in gleichmäßiger Erhöhung hinzugefügt. Für $i \in 0...10$ wird die Stärke des Rauschens auf $\sigma = 25 - i$ und die Lichtstärke auf $(10 - i)\%$ gesetzt. Weiterhin wird eine Kombination von Begrenzungsrahmenbedeckung durch Abdeckung jeder einzelnen Person im unteren Drittel in horizontaler Richtung und entlang der vertikale Mittelachse auf die Bilddaten addiert. Die untere Reihe von Abbildung 19 zeigt die Ergebnisse dieses kombinierten Störungsprozesses.
Zur Findung des zuverlässigsten Personendetektors werden die Videodaten aus einer einfachen Szene in drei Modalitäten genutzt. Zunächst eine einzelne Person, die durch die Halle geht und die Arbeit an einem Computer aufnimmt, dann eine Gruppe von zwei Personen und schließlich eine Gruppe von sechs Personen, die den gleichen Weg gehen, beliebige Arbeiten ausführen und zufällige Wege gehen. Für die ersten beiden Modalitäten trugen die Darsteller typische Berufskleidung. In der Gruppenaufnahme treten drei Personen in Alltagskleidung in Erscheinung. Hierzu wurden etwa viertausend Einzelbil-

Abbildung 19: Beispielbilder der vorverarbeiteten Videobilder. Obere Reihe: Das gauß'sche Rauschen steigt von $\sigma = 0$ (links) auf $\sigma = 250$ (rechts). Zweite Reihe: Die Stärke des blendenden Lichts nimmt zu, von 0% der Lichtkartenaddition (links) bis 100% (rechts). Dritte Reihe: die sechs Verdeckungsmodalitäten. Untere Reihe: die kombinierte Störungen, von $i = 0$ (links) bis $i = 10$ (rechts) [65].

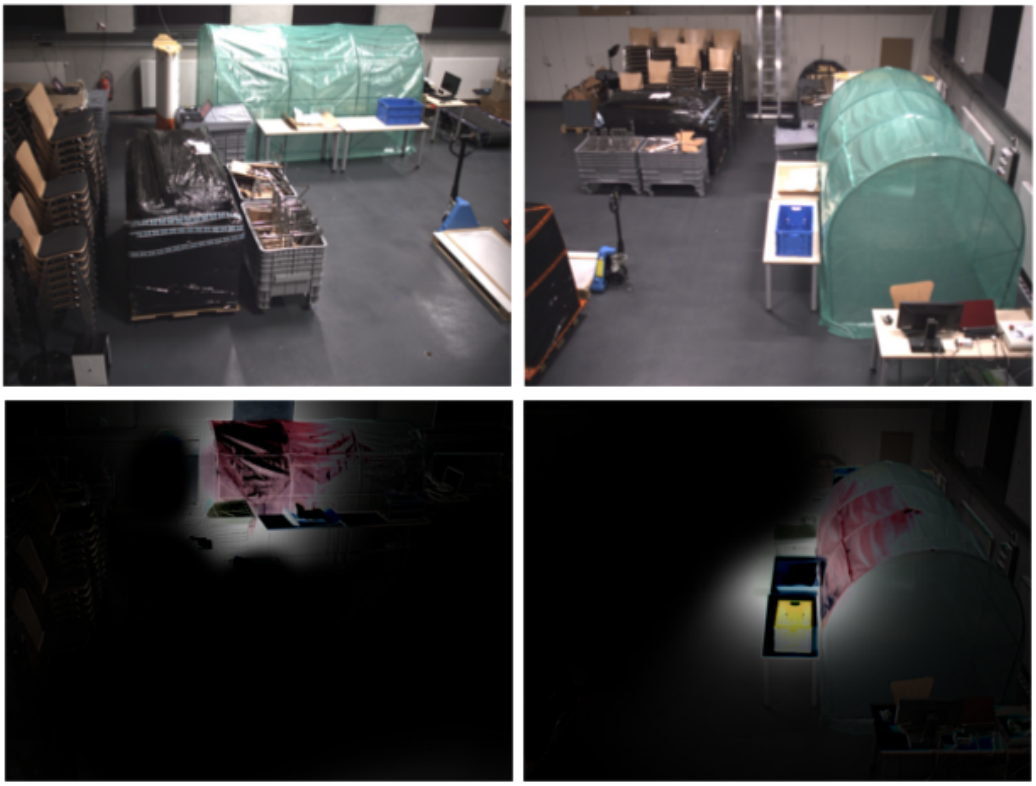


Abbildung 20: Obere Zeile: Hintergrundbilder der Kamera C_1 (rechts) und C_2 (links). Untere Reihe: Entsprechende Blendeffekte für die Bilder der Kamera C_1 (rechts) und C_2 (links) [65].

	HOG	YOLO	OP
FPS	33.5	18.4	11.3
GPU Memory	251 MB	1293 MB	1313 MB

Tabelle 1: Arbeitsgeschwindigkeit und belegte Rechenressourcen der untersuchten Methoden.

der mit zwei Kameras C_1 und C_2 aufgenommen, von denen ein Student des Projektteams insgesamt 3.557 Bilder manuell beschriftet hat, jeweils mit Begrenzungsrahmen um die einzelnen Personen und um ihre Füße herum. Diese Bilder wurden anschließend mit den erläuterten systematischen Störungen manipuliert und somit ein Datensatz mit 142.311 Bildern generiert, an denen die jeweiligen Qualitäten der Methoden gemessen wurden. Um die Methoden zu bewerten, werden auch die jeweils belegten Rechenressourcen ausgewertet, wie in Tabelle 1 dargelegt.
Die manuell beschrifteten Daten bestehen aus Begrenzungsrahmen für die Personen und ihre Füße. Um eine vergleichbare Punktzahl für die Leistungsmessung zu erhalten, wird getestet, ob die Methoden die korrekten Positionen schätzen. Hierzu wird überprüft, ob die vorhergesagten Punkte in den manuell beschrifteten Boxen liegen. Angesichts der vorhergesagten Begrenzungsrahmen wird der Mittelpunkt für die Personenortung und die Mitte der unteren Linie des Begrenzungsrahmens für die Fußpunktortung herangezogen. Zur Auswertung des OP-Systems wird das Modell BODY25 mit Standardparametern verwendet. Mit diesem menschlichen Körpermodell können den Halsschlüsselpunkt für die Personenposition und die Mitte sowohl das rechte als auch das linke Fußgelenk für die Fußpunktposition nutzen.
Wie in Abbildung 21 dargestellt, zeigt das OP-System einen klaren Vorteil in Bezug auf Rauschwiderstand, Lichtresistenz und Körpervervollständigung, insbesondere unter Berücksichtigung der Fußpunktposition. In Bezug auf Teilverdeckungen kann festgestellt werden, dass YOLO insbesondere Personen nicht erkennt, die entlang ihrer vertikalen Mittelachse abgedeckt sind. Die Tendenzen der Fehlerkennungen legen nahe, dass OP im Durchschnitt dazu neigt Personen zu erkennen die nicht da sind, während YOLO dazu neigt, anwesende Persone nicht zu erkennen. In Anbetracht der Härte der kombinierten Störungen, fallen alle untersuchten Methoden schnell in der Leistung, aber OP ist dennoch zuverlässiger als die anderen Methoden.

7.3 Personenverfolgung

Die Softwarearchitektur besteht konzeptionell aus drei verschiedenen Modulen, die auf einem Echtzeitbilddatenstrom aus mehreren Kameras arbeiten.

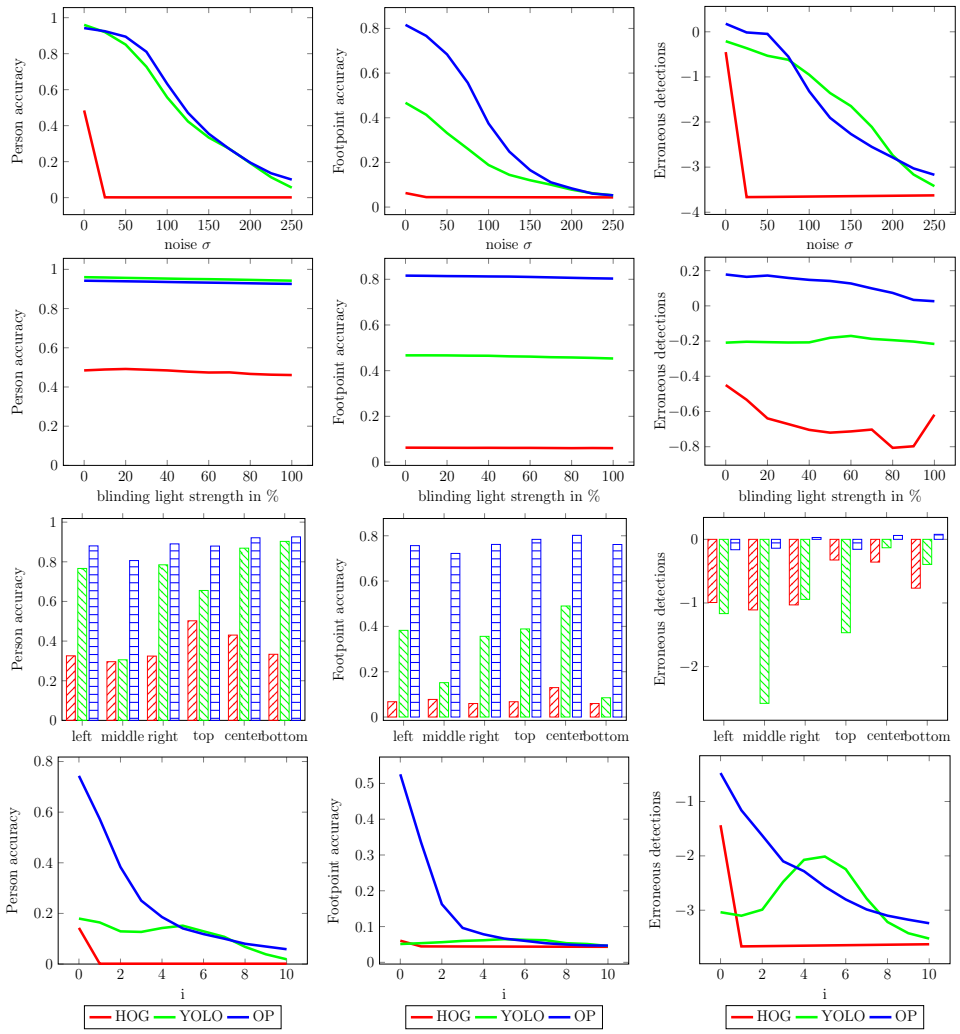


Abbildung 21: Die über alle Sequenzen gemittelten Ergebnisse der Personen- und Fußpunkterkennung [65]. Obere Reihe: Die Beständigkeiten gegen Rauschen. Zweite Reihe: Die Antizipation des blendenden Lichteffekts. Dritte Reihe: Die Ergebnisse zu den teilweise abgedeckte Personen in den sechs Bedeckungsmodalitäten. Untere Reihe: die kombinierten Störungen. Linke Spalte: Genauigkeit der Personenerkennung. Mittlere Spalte: Genauigkeit der Fußpunkterkennung. Rechte Spalte: Anzahl der Fehlerkennungen (< 0: nicht erkannte Personen, > 0: zu viel erkannte Personen, $= 0$: durchschnittlich so viele nicht erkannte wie zu viel erkannte Personen).

Abbildung 22 zeigt das Programmablaufdiagramm der einzelnen Verarbeitungsschritte. Beginnend mit einem Fußpunktortungssystem, für das OP verwendet wird, wird eine Koordinatentransformation der Fußpunkte von Bildkoordinaten in die Weltkoordinaten durchgeführt. Diese stellen die Eingabe für das zweite Modul, einen regelbasierten Sensorfusionsansatz. Das regelbasierte System bereitet die von den Personen zurückgelegten Spuren für den dritten Verarbeitungsschritt vor. Das Dritte Modul, ein Kalman-Filter, glättet die Spuren. Alle Verarbeitungsschritte finden in Echtzeit auf dem gleichen Bildstapel statt, in einem Kompromiss zwischen Stapelgröße und verfügbaren Computerressourcen. Je mehr Bilder die Software sieht, desto genauer wird sie. Auf diese Weise wird die Echtzeitfähigkeit nur von den verfügbaren Rechenressourcen ab. Als Eingabe wird ein Stapel von k Bildern pro Kamera präsentiert, wobei k so gewählt wird, dass das Programm in Echtzeit und so genau wie möglich läuft. Die Erhöhung des Parameters k führt zu einer höheren Genauigkeit auf Kosten einer höheren Rechenzeit. Als Ausgabe liefert das System die Personenstandorte in Weltkoordinaten, die lokal gespeichert oder an ein Kontextmodell gesendet werden können.

Vor der Durchführung der Fußpunktortung werden die Kamerabilder mittels eines adaptiven Histogrammausgleichs verbessert [33]. Unter Verwendung der OP Architektur werden dann die Fußpunkte der Personen in der Kamerakoordinaten lokalisiert. Der Kalibrierungsprozess geht davon aus, dass sich die Person auf einer ebenen Fläche bewegen, so dass eine temporäre konstante Höhenkoordinate $z = 1$ verwendet werden kann, die nach der Transformation wieder verworfen wird. Zur Überführung der Kamerakoordinaten $(x, y, 1)^T$ in Weltkoordinaten $(p_x, p_y)^T$, wird die intrinsische Kameramatrix M, die Rotationsmatrix R und den Übersetzungsvektor d, det aus einem Standardkalibrierungsprozess mit Schachbrettmustern erhalten werden kann, verwendet [26, 18]. Mit einer Hilfsmatrix

$$R' = \begin{pmatrix} R_{0,0}, R_{0,1}, R_{0,2} \\ R_{1,0}, R_{1,1}, R_{1,2} \\ d_0, d_1, d_2 \end{pmatrix} \tag{1}$$

kann eine Koordinatentransformation durchgeführt werden, wie etwa:

$$\begin{pmatrix} p_x \\ p_y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} (R'M)^{-1} \tag{2}$$

Die daraus resultierenden Weltkoordinaten beziehen sich auf den Ursprung des Schachbrettmuster. Für mehrere Kameras, die unterschiedliche

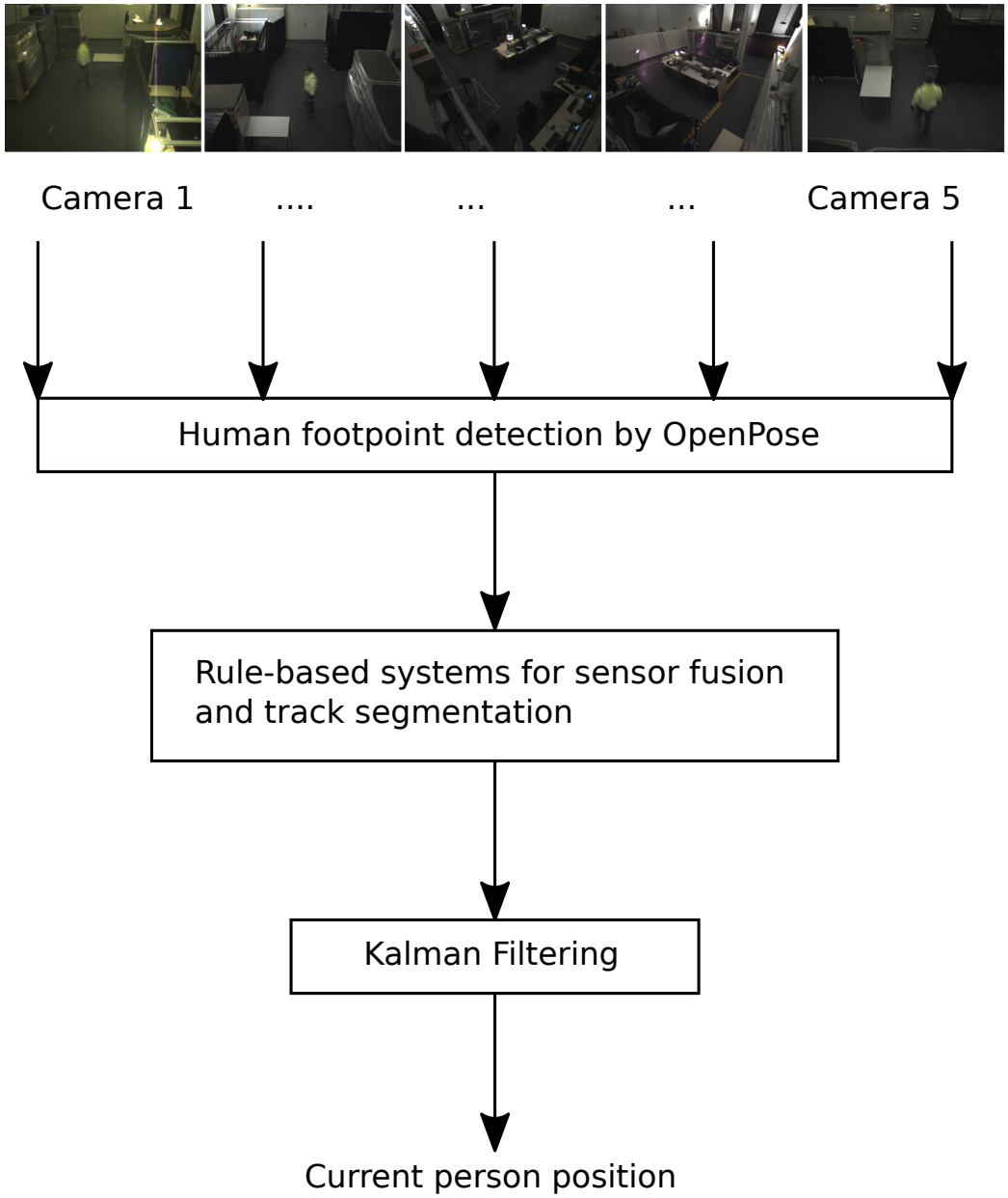


Abbildung 22: Das Programmablaufdiagramm des Personenverfolgungsverfahrens [4]. Die oberste Reihe zeigt Beispielbilder, die aus der gleichen Studie zur gleichen Zeit aufgenommen wurden. Von links rechts: Kameras C_1 bis C_5, wie in Abbildung 13 dargestellt. Die Person bewegt sich gerade in Sichtweite der Kameras C_1, C_2, C_5, aber außer Sichtweite der Kameras C_3 und C_4.

Teile der Umgebung beobachten werden mehrere extrinsische Kalibrierungen durchgeführt und Abstandsvektoren zwischen den verschiedenen Koordinatenursprüngen gemessen.
Das zweite Modul besteht aus regelbasierten Systemen, die zunächst eine automatische Vervollständigung innerhalb der k Bilder durchführt. Wird in einem Bild keine Erkennung gefunden, aber im vorigen und folgenden Bild schon, so wird die fehlende Erkennung durch das geometrische Mittel zwischen den beiden erfolgreichen Detektionen aufgefüllt. Auf diese Weise werden die Erkennungen innerhalb von k Bildern vervollständigt und die Aufrechterhaltung der Echtzeit-Anwendbarkeit für einen optimalen Wert von k gewährleistet. Aus dem Wissen über die Erkennungen auf den Bildern wird aus diesen Wahrheitswerten eine zweidimensionale Matrix gebildet, die aussagt, welche Kameras kohärente Erkennungen innerhalb des Echtzeitfensters liefern. Unter Verwendung von Vorkenntnissen über die Positionen der Kameras in der Umgebung entscheidet nun ein regelbasiertes System, welche Sensordaten zu fusionieren sind. Um die Positionsdaten für die Kalman-Filterung aufzubereiten, wird anschließend einen weiteren Regelsatz angewandt, der jede Position einer eindeutigen Spur zuordnet. Jede einzelne Spur besteht hierbei konzeptionell nur aus einer gleichmäßigen Bewegung, was das Kalman-Filterverfahren vereinfacht.
Um die resultierenden Spuren zu glätten, wird ein Kalman-Filter [60] eingesetzt. Zur Initialisierung des Kalman-Filters wird eine vierdimensionale Matrix für eine gleichmäßige Bewegungsdynamik verwendet, welche die Positionen $(p_x, p_y)^T$ sowie die aktuellen Geschwindigkeiten der Person $(v_x, v_y)^T$ in Bezug setzt zu einer Zeitschrittweite dt, wie sie durch die Aufnahmefrequenz der Kameras vorgegeben ist:

$$F = \begin{pmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

sodass:

$$\begin{pmatrix} p_x \\ p_y \\ v_x \\ v_y \end{pmatrix}_{t+1} = F \cdot \begin{pmatrix} p_x \\ p_y \\ v_x \\ v_y \end{pmatrix}_{t} \tag{4}$$

Zur Initialisierung der Kalman-Filter wird die Einheitsmatrix als Kovarianzmatrix verwendet. Für den Schätzprozess wird über k aufeinander folgende Positionen iteriert und somit die Echtzeitanwendbarkeit erhalten.

Die Videoaufzeichnungen aus der Studie im kleinen Technikum der Hochschule Ruhr West am Campus Bottrop werden verwendet, um die Personenverfolgung auszuwerten. Die Kameras haben Videomaterial mit acht bis zwölf Bildern pro Sekunde aufgezeichnet. Eine Stapelgröße von $k = 4$ Frames wird verwendet, um die Echtzeitanwendbarkeit mit der verwendeten Hardware zu erhalten. Das Personenverfolgungssystem hat eine Reaktionszeit von etwa $500ms$. Zur Auswertung werden die Spuren, die aus den Kameradaten gewonnen wurden, mit den originalen Spuren aus der AR-Anwendung verglichen. Die AR-Anwendung liefert zwei Positionsbestimmungen pro Sekunde, die Kameras liefern also mehr Daten in der gleichen Zeit, da sie mit einer höheren Bildrate aufnehmen. Abbildung 24 zeigt die Spuren, wie sie von der AR-Anwendung bereitgestellt werden, die Rohkamerapositionsschätzungen nach der Sensorfusion, die separierten Spuren sowie die Endpositionen nach der Kalman-Filterung. Die Statistiken über die zurückgelegten Abstände, Dauer und Geschwindigkeiten, zeigt Abbildung 25. Um den Vergleich der mittlere Abweichungen zwischen den geschätzten Kamerapositionen und der Grundwahrheit durchzuführen, wird zunächst das Problem der verschiedenen zeitlichen Auflösungen gelöst, indem für jeden Punkt aus den Kamerapositionen der nächstgelegene Punkt in den Grundwahrheitspositionen gesucht wird.

Der Mittelwert der mittleren Abweichungen zwischen den Spuren nach der Sensorfusion und den Grundwahrheitsspuren beträgt etwa $0,67m$, nach der Kalmanfilterung nur $0,59m$. In den Darstellungen der Spuren in Abbildung 24 zeigt sich eine leichte metrische Verzerrung im Anfangs- und Endbereich, die wahrscheinlich darauf zurückzuführen ist, dass nur die Kamera C_2 diesen Bereich beobachtet. Nach Analyse der Histogramme in Abbildung 25 kann gesehen werden, dass sowohl die zurückgelegten Wegstrecken als auch die durchschnittlichen Geschwindigkeiten nach der Kalman-Filterung der ursprünglichen Verteilung näher kommen. Darüber hinaus zeigt sich, dass in den Evakuierungsspuren, also den Spuren neun bis zwölf, die mittleren Geschwindigkeiten höhere Werte aufweisen.

7.4 Thermografieintegration

Zusätzlich zur Personendetektion auf Bildern der RGB-Kameras untersuchte die Hochschule Ruhr West die Anwendbarkeit der Personendetektionsverfahren auf Wärmebildern von Thermokameras. Hierzu wurden zwei FLIR-Kameras montiert, welche den simulierten Gefahrenbereich sowie den Bereich um Aufgabe Zwei herum beobachten, wie in Abbildung 26 gezeigt. Zunächst untersuchte das Projektteam die Anwendbarkeit der Personendetektoren auf Bildern von FLIR Thermokameras und kommt zu dem Ergebnis, dass OP

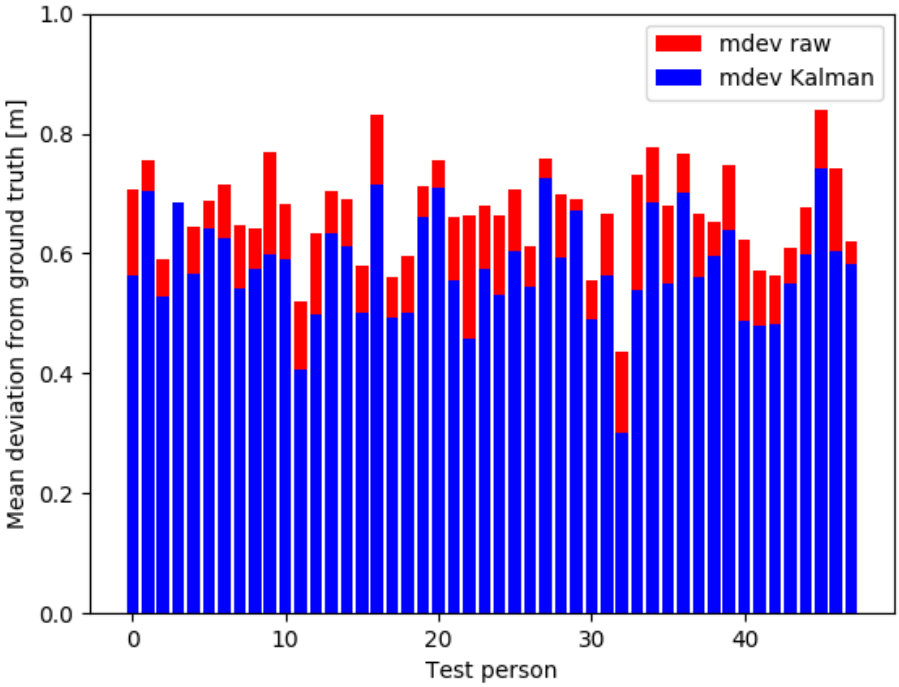
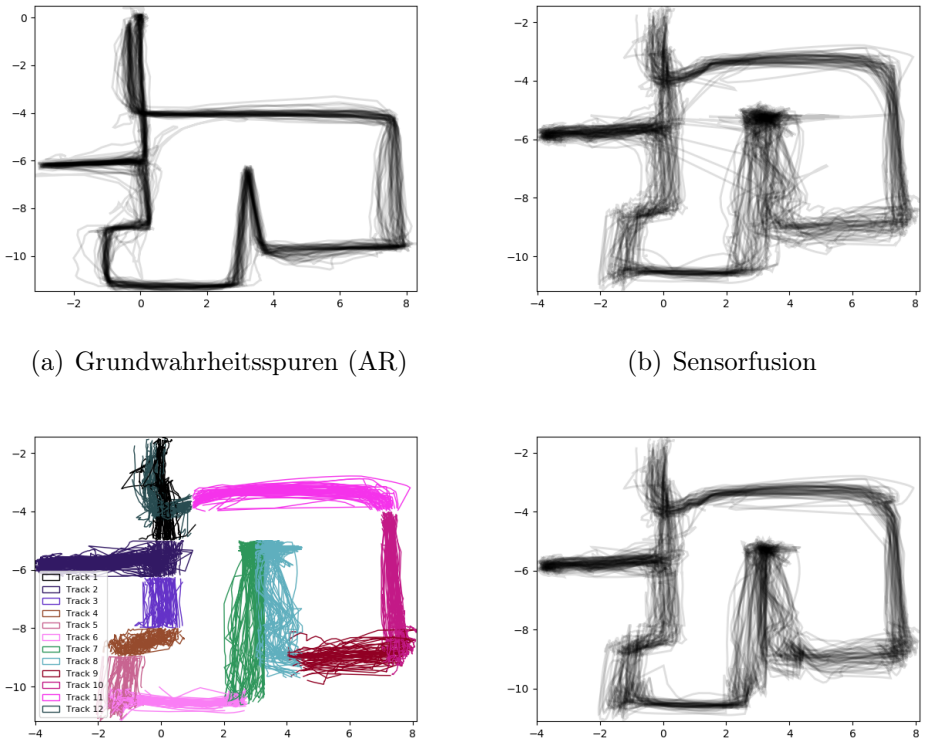


Abbildung 23: Die mittlere Abweichung in Metern zwischen den geschätzten Kamerapositionen und den Grundwahrheitspositionen der AR-Anwendung [4]. Die roten Balken zeigen die mittleren Abweichungen für die Rohkameraschätzungen und die blauen Balken zeigen die mittleren Abweichungen für die kalmangefilterten Positionen.



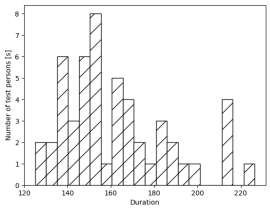
(a) Grundwahrheitsspuren (AR)

(b) Sensorfusion

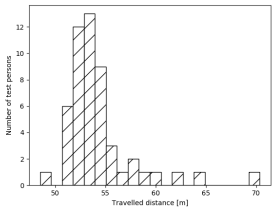
(c) Spurtrennung

(d) Kalmanfilterung

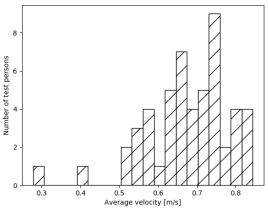
Abbildung 24: Darstellung der zurückgelegten Wege der Probanden [4]. Die Abbildung oben links zeigt die Grundwahrheit über die Spuren, in diesem Falle die konkreten Positionsdaten aus der AR-Anwendung. Das Bild oben rechts die Spuren nach dem Einsatz des ersten regelbasierten Systems zur Sensorfusion. Im unteren linken Plot visualisiert sind die Ergebnisse nach der Spurtrennung. Wie im rechten unteren Diagramm zu sehen ist, sind die Spuren nach der Kalmanfilterung geglättet und ähneln den Positionen aus den Grundwahrheitsdaten.



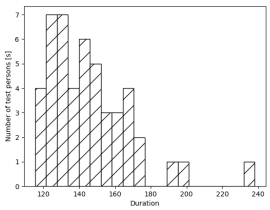
(a) Dauer (AR Data)



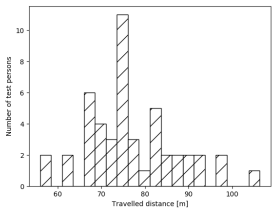
(b) Wegstrecke (AR Data)



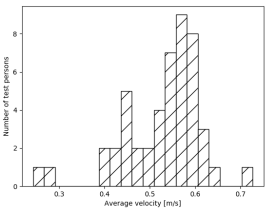
(c) mittlere Geschwindigkeit (AR Data)



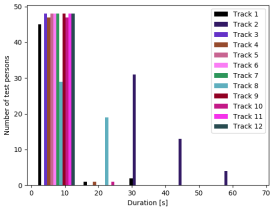
(d) Dauer (Sensorfusuion)



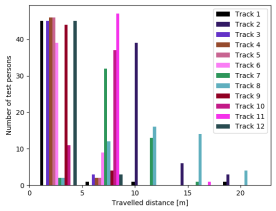
(e) Wegstrecke (Sensorfusuion)



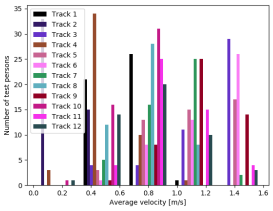
(f) mittlere Geschwindigkeit (Sensorfusuion)



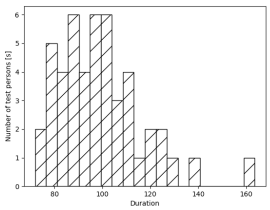
(g) Dauer (Spurtrennung)



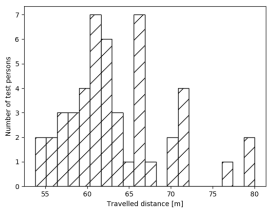
(h) Wegstrecke (Spurtrennung)



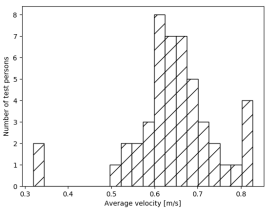
(i) mittlere Geschwindigkeit (Spurtrennung)



(j) Dauer (Kalmanfilter)



(k) Wegstrecke (Kalmanfilter)



(l) mittlere Geschwindigkeit (Kalmanfilter)

Abbildung 25: Die Histogramme zeigen Gesamtstatistiken der Ergebnisse [4]. Von links nach rechts: Dauer in Sekunden, zurückgelegte Wegstrecke in Metern, durchschnittliche Geschwindigkeiten in Metern pro Sekunde. Von oben nach unten: Grundwahrheitsdaten von der AR-Anwendung, Sensorfusion, Spurtrennung und die endgültigen kalmangefilterten Ergebnisse.

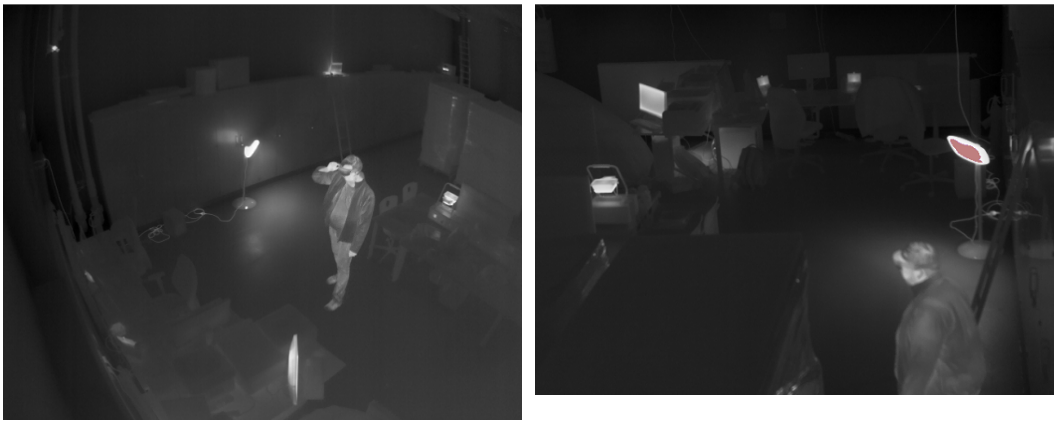


Abbildung 26: Beispielaufnahmen der FLIR Thermokameras.

auch auf Wärmebildern zuverlässig menschliche Körper erkennt, wenn ein entsprechender Parameter $USE_FLIR = 1$ gesetzt wird.
Um die FLIR Thermokameras auf ein Weltkoordinatensystem zu kalibrieren, hat das Projektteam ein spezielles Schachbrettmuster vorbereitet. Auf den schwarzen Feldern dieses Schachbrettmuster hat das Projektteam Aluminiumfolie angebracht und mit einem Wärmestrahler aufgeheizt, sodass die schwarzen und weißen Felder auf einem Wärmebild ein hinreichend kontrastreiches Schachbrettmuster darstellen. Der Standardkalibrierungsprozess hat funktioniert, liefert jedoch Matrizen, welche die Koordinatentransformationen zu den Rändern des Bildes hinaus stark verzerren. Im äußeren Bereich der Bilder betrugen die Abweichungen gegenüber den Originalpositionen bis zu 20 Metern und wurden daher nicht in das regelbasierte System integriert. Die zuverlässige Transformation der Bildkoordinaten aus Wärmebildern in Weltkoordinaten stellt eine Herausforderung dar.

8 Zusammenfassung

Ein Schlüsselmerkmal der Industrie 4.0 ist der Einsatz neuer Technologien zur Erkennung des aktuellen Kontexts, sodass die Mitarbeiter durch kontextbezogene Informationen unterstützt werden können. Dies dient der Beschleunigung der Entscheidungsfindung bei den verschiedenen Prozessen, beispielsweise im Zusammenhang mit Planung, Produktion oder Instandhaltung. Tragbare Geräte, wie Augmented Reality Brillen, bieten die Möglichkeit, die Mitarbeiter in ihrem Arbeitsalltag zu unterstützen. Allerdings stellen gefährliche Umgebung, beispielsweise Anlagen in der Stahlindustrie, besondere Anforderungen an die Benutzbarkeit und Sicherheit.

In diesem Beitrag wird der Softwareansatz für ein kontextbasiertes System von mobilen und intelligenten Geräten vorgestellt, das die Mitarbeiter in der Stahlindustrie unterstützt. Basierend auf Szenarien in den Bereichen Sicherheit, Produktion, Instandhaltung und Lagerverwaltung wurden Konzepte entwickelt und bewertet. Dazu gehören neben der eher technischen Perspektive auch die Entwicklung neuer Anwendungen und Nutzkonzepte für mobile und intelligente Geräte, Methoden zur Lokalisierung und Kontextualisierung, soziale Aspekte wie Nutzerempfinden oder Akzeptanz neuer Konzepte.
Die Position von mobilen Geräten kann über Sensoren wie Bluetooth- oder WiFi-Beacons erfasst werden, um den Benutzer des Mobilgeräts im Inneren der Fabrikhalle zu lokalisieren. Für einen Ortungsdienst hingegen schlägt das Projektteam vor, digitale Bildverarbeitung mit künstlicher Intelligenz zu verwenden, um Personen in der Fabrikhalle zu orten und Fabrikobjekte zu identifizieren. Zur Findung robuster Personendetektoren werden verschiedene maschinelle Lernprozesse zur Erkennung menschlicher Körper bewertet. Um den Herausforderungen von Kamerabildern in der Schwerindustrieumgebung, wie Staub, schwierigen Lichtverhältnissen und teilweise verdeckten Körpern gerecht zu werden, werden entsprechende programmatische Bildstörungen angewandt. Anhand eines Testbildsatzes werden die Genauigkeit dieser Methoden bezüglich der Personenerkennung allgemein bewertet, der Fußpunktabschätzung im Besonderen sowie der Tendenz der fehlerhaften Erkennungen. Weiterhin schlägt das Projektteam eine Methode zur Lokalisierung und Verfolgung der Personen im Bild mit mehreren Kameras vor. Unter Verwendung des OpenPose Netzwerks werden die Fußpunkte der Personen auf jedem Kameranild lokalisiert und entsprechende Transformationen von Bild- in Weltkoordinaten durchgeführt. Mit Vorkenntnissen über die Kameraeinstellungen in der Umgebung wird ein regelbasiertes System zur Sensorfusion eingesetzt. Dann wird ein Kalman-Filter angewandt, um die Nachführung der Spuren zu stabilisieren. Aufgrund einer variablen Bildstapelgröße kann das Verfahren die Genauigkeit erhöhen, indem es mit mehr Bildern versorgt wird. So erlauben zusätzliche Rechenressourcen die Verarbeitung von mehr Information in der selben Zeit, was in höherer Genauigkeit resultiert. Aufgrund der beobachteten metrischen Verzerrung kommt die Hochschule Ruhr West zu dem Schluss, dass für eine zuverlässige Standortbestimmung eine Person von mindestens zwei kalibrierten Kameras gesehen werden muss. Um genaue Verfolgungsergebnisse auf Kamerabildern zu erhalten, empfiehlt das Projektteam aufgrund des Rauschens in solchen Umgebungen die Verwendung zusätzlicher Mittel wie beispielsweise dem Kalman-Filter.

8.1 Anwendungsnutzen

Das redundante Personenverfolgungssystem stellt eine nützliche Anschaffung für die Schwerindustrie dar. Auch wenn einzelne Kameras oder Bluetooth-Beacons ausfallen, können Personen noch zuverlässig genug detektiert werden. Unter Nutzung dieser Sensoreingaben können auch Gefahrenquellen automatisch erkannt und im System vermerkt werden, um automatisiert entsprechende Warnhinweise an die Betroffenen zu geben. Ebenso können die Positionen von Maschinen und anderen Objekten im Raum redundant bestimmt werden, wie Abbildung 27 zusammenfasst. Die gesammelten Echtzeitinformationen dienen der Flexibilisierung des Produktionsprozesses und der Steigerung der Arbeitssicherheit. Testaufnahmen bei einem assoziierten Partner der Stahlindustrie, dargestellt in Abbildung 28, demonstrieren die Anwendbarkeit der untersuchten Methoden in einer realen Industrieumgebung.

8.2 Diskussion

Die vorgestellten Ergebnisse sind ein Beispiel für eine interdisziplinäre, multimethodische Untersuchung. Eine Einschränkung betrifft die erforschte Gruppe, da alle Teilnehmer eine hohe Affinität zur Technik im Allgemeinen hatten. Obwohl dies vielleicht zu einer höheren Akzeptanz der AR-basierten Technologie führen könnte, wurde die Anwendung so konzipiert, dass sie so wenig Eingaben wie möglich vom Benutzer erfordert. Darüber hinaus werden zukünftige Generationen mit ähnlichen Technologien aufwachsen und wird wahrscheinlich weniger Bedenken gegenüber solchen Anwendungen im Vergleich zu heutigen Generationen haben, die keine *digital natives* sind. Das Projekt hat nützliche Erkenntnisse über die gemäß der vereinbarten Szenarien mit AR-Technologien zu erledigenden Aufgaben geliefert.
Für alle künstlichen neuronalen Netze wurden die Standardeinstellungen verwendet, die Feinabstimmung der Hyperparameter der verwendeten neuronalen Modelle kann die Ergebnisse zu verändern. Da keine rekurrenten neuronalen Netze genutzt wurden, wurden die Vorteile, die solche Netze der Verarbeitung von Bildsequenzen bringen können, nicht untersucht. Das Projektteam konnte keine verwandten Arbeiten finden, die Herausforderungen für den Personenortung in schwerindustrieller Umgebung beschreiben, daher kann nach derzeitigem Kenntnisstand keine Aussage darüber getroffen werden, wie gut der Datensatz die realen Szenarien simuliert hat. Für das Navigationssystem wurde noch keine Echtzeitpositionierung im Raum durch das Kontextmodell, Signalstärken oder Videoaufnahmen verwendet. Weiterhin wurde kein Grundriss der Umgebung hinterlegt für die AR-Anwendung hinterlegt und die Route nicht dynamisch berechnet.

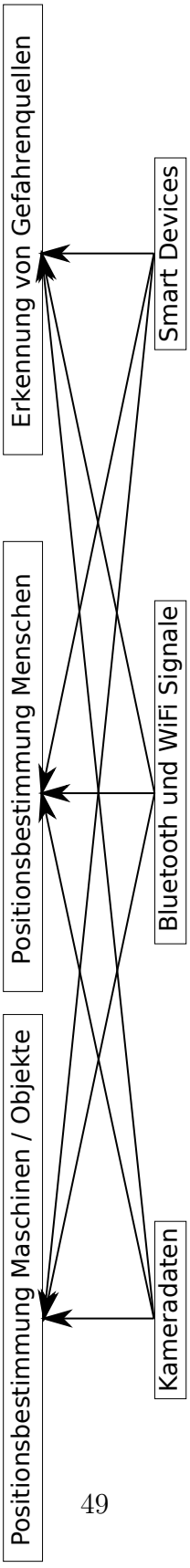


Abbildung 27: Positionen von Menschen, Maschinen und anderen Objekten können mittels redundanter Eingaben bestimmt werden.

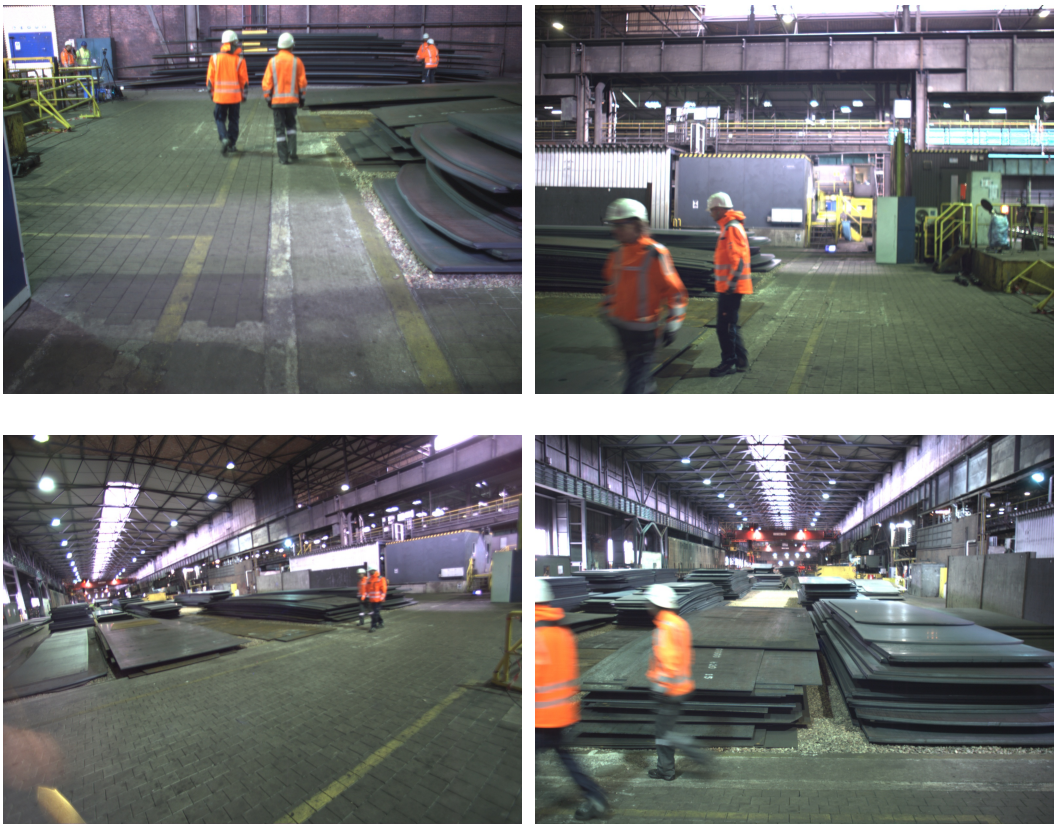


Abbildung 28: Beispielaufnahmen bei einem Industriepartner.

Die Sensorfusion in Form eines regelbasierten Systems basiert auf Vorkenntnissen, aber ermöglicht durch seinen transparenten Regelsatz einfache Änderungen. Das Modell geht davon aus, dass die Personen bewegen sich auf einer flachen Ebene, so dass es nicht möglich ist, unterschiedliche Höhenniveaus zu unterscheiden. Die regelbasierten Systeme, obwohl sie für den Benutzer transparent und einfach zu bedienen sind, lassen die Flexibilität in der Generalisierung auf verschiedene Aufgaben vermissen. Das gleiche Problem ergibt sich für die regelbasierte Spurtrennung. Da die Laborstudie ist nur Videomaterial von Einzelpersonen auf dem Parcours enthält, wurde die Genauigkeit des Personenverfolgungssystems für mehrere Personen nicht bewertet.

8.3 Zukünftige Studien

Die Ortung und Rollenzuweisung mehrerer Personen erfordert eine zusätzliche Identifikation über noch nicht integrierte Modelle wie Gesichtserkennung, Gangerkennung oder Smartphonegeräteidentifikation. Es gibt verschiedene Schlussfolgerungen für zukünftige Studien, die es zu berücksichtigen gilt. Um zu beurteilen, wie sich eine Person fühlt kann während der Nutzung der entwickelten Anwendung ein Fitness-Tracker verwendet werden, um zusätzliche Daten zu sammeln, welche das Kontextmodell bereichern. So können auch weitere Aspekte, beispielsweise der Stress bei der Nutzung des Systems und der Ausführung der Aufgaben, gemessen und ausgewertet werden. Im nächsten Schritt muss diese Bewertung mit den neuen Konzepten für intelligente Uhren und für eine Kombination aus intelligenter Brille und intelligenter Uhr erfolgen. Die Untersuchungen konzentrieren sich auf die Erkennung von Menschen mit auf RGB-Kameras. Entsprechende Aufnahmen mit Wärmebildkameras existieren und können in zukünftigen Studien ebenfalls zur Untersuchung von Personenortungssystemen eingesetzt werden. Auch kann es sinnvoll sein, Methoden des Transferlernens zu untersuchen, um das modellierte Wissen als Ausgangspunkt für weitere Lernprozesse zu nutzen. Zukünftige Arbeiten können die aktuelle Entwicklungen im Bereich der Objektdetektoren bewerten. Um den Ansatz weiterzuentwickeln, könnte eine flexiblere Spurtrennung, beispielsweise durch eine Verstärkungslernagenten, zu einer höheren Flexibilität führen und der dynamischen Routenplanung dienen.

Finanzierung

Das Projekt *DamokleS 4.0* wurde aus Mitteln des europäischen Fonds für regionale Entwicklung (EFRE), der Europäischen Union (EU) und des Bundeslandes Nordrhein-Westfalen finanziert.

Literatur

[1] Sofiane Abbar, Mokrane Bouzeghoub und Stéphane Lopez. "Context-aware recommender systems: A service-oriented approach". In: *VLDB PersDB workshop*. 2009, S. 1–6.

[2] Gregory D Abowd u. a. "Cyberguide: A mobile context-aware tour guide". In: *Wireless networks* 3.5 (1997), S. 421–433.

[3] Sam Amin und Jeff Burke. "OpenMoves: A System for Interpreting Person-Tracking Data". In: Juni 2018, S. 1–4. DOI: 10.1145/3212721.3212846.

[4] Nico Zengeler Alexander Arntz u. a. "Person Tracking in Heavy Industry Environments with Camera Images". In: *S-CUBE 2019 10th EAI International Conference on Sensor Systems and Software*. 2019.

[5] Victoria Bellotti u. a. "Activity-based serendipitous recommendations with the Magitti mobile leisure guide". In: *Proceedings of the sigchi conference on human factors in computing systems*. ACM. 2008, S. 1157–1166.

[6] M. Billinghurst, R. Grasset und J. Looser. "Designing Augmented Reality Systems". In: *Siggraph Comput. Graph., vol. 39*. ACM. 2005.

[7] Hilton Bristow und Simon Lucey. "Why do linear SVMs trained on HOG features perform so well?" In: *CoRR* abs/1406.2419 (2014). arXiv: 1406.2419. URL: http://arxiv.org/abs/1406.2419.

[8] Zhe Cao u. a. "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields". In: *CVPR*. 2017.

[9] Federica Cena u. a. "Integrating heterogeneous adaptation techniques to build a flexible and usable mobile tourist guide". In: *AI Communications* 19.4 (2006), S. 369–384.

[10] N. Dalal und B. Triggs. "Histograms of oriented gradients for human detection". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Bd. 1. Juni 2005, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.

[11] Toon De Pessemier u. a. "A user-centric evaluation of context-aware recommendations for a mobile news service". In: *Multimedia Tools and Applications* 75.6 (2016), S. 3323–3351.

[12] Anind K Dey. "Understanding and using context". In: *Personal and ubiquitous computing* 5.1 (2001), S. 4–7.

[13] P. Dollar u. a. “Pedestrian Detection: An Evaluation of the State of the Art”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.4 (Apr. 2012), S. 743–761. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.155.

[14] *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. https://www.iso.org/standard/52075.html. 11. Juli 2019.

[15] IN) Girish Banavathi Venkata Subba (Bangalore. “METHOD AND APPARATUS FOR HUMAN DETECTION IN IMAGES”. 20180157904. Juni 2018. URL: http://www.freepatentsonline.com/y2018/0157904.html.

[16] Uwe Handmann u. a. “APFel - Fast multi camera people tracking at airports, based on decentralized video indexing”. In: 2 (Jan. 2014), S. 48–55.

[17] Tong Hao u. a. “Multiple person tracking based on slow feature analysis”. In: *Multimedia Tools and Applications* 77 (Sep. 2017). DOI: 10.1007/s11042-017-5218-4.

[18] Janne Heikkila und Olli Silven. “A Four-step Camera Calibration Procedure with Implicit Image Correction”. In: *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*. CVPR '97. Washington, DC, USA: IEEE Computer Society, 1997, S. 1106–. ISBN: 0-8186-7822-4. URL: http://dl.acm.org/citation.cfm?id=794189.794489.

[19] Karen Henricksen, Jadwiga Indulska und Ted McFadden. “Modelling context information with ORM”. In: *OTM Confederated International ConferencesÖn the Move to Meaningful Internet Systems"*. Springer. 2005, S. 626–635.

[20] M. Hermann, T. Pentek und B. Otto. “Design Principles for Industrie 4.0 Scenarios: A literature Review”. In: 2015.

[21] K. Hermsen u. a. “Dynamic, Adaptive and Mobile System for Context-Based and Intelligent Support of Employees in the Steel Industry”. In: *4th ESTAD (European Steel Technology and Application Days)*. Düsseldorf, Germany, 2019. URL: https://www.metec-estad2019.com/files/190619_metec-estad_programmflyer_a5q_web-5.pdf.

[22] A. Garcia Herrero. “Europe in the Midst of China-US Stratetic Economic Competition: What are the European Union’s Options?” In: 2019.

[23] Ilija Ilievski und Sujoy Roy. "Personalized news recommendation based on implicit feedback". In: *Proceedings of the 2013 international news recommender systems workshop and challenge*. ACM. 2013, S. 10–15.

[24] Itseez. *Open Source Computer Vision Library*. `https://github.com/itseez/opencv`. 2015.

[25] Jongbae Kim und Heesung Jun. "Vision-based location positioning using augmented reality for indoor navigation". In: *IEEE Transactions on Consumer Electronics* 54.3 (2008), S. 954–962.

[26] MathWorks. *Camera Calibration*. `https://mathworks.com/`. 23. Feb. 2018.

[27] Mircosoft. *Design for Mixed Reality*. `https://docs.microsoft.com/en-us/windows/mixed-reality/design`. 11. Juli 2019.

[28] Mircosoft. *Spatial Mapping*. `https://developer.microsoft.com/en-us/windows/mixed-reality/spatial\_mapping\_in\_unity`. 11. Juli 2019.

[29] Alessandro Mulloni, Hartmut Seichter und Dieter Schmalstieg. "Handheld augmented reality indoor navigation with activity-based instructions". In: *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*. ACM. 2011, S. 211–220.

[30] Wolfgang Narzt u. a. "Augmented reality navigation systems". In: *Universal Access in the Information Society* 4.3 (2006), S. 177–187.

[31] Robert M Nosofsky. "Tests of an exemplar model for relating perceptual classification and recognition memory." In: *Journal of experimental psychology: human perception and performance* 17.1 (1991), S. 3.

[32] Peyton Z Peebles. *Probability, random variables, and random signal principles*. Bd. 3. McGraw-Hill New York, NY, USA: 2001.

[33] Stephen M. Pizer u. a. "Adaptive Histogram Equalization and Its Variations". In: *Comput. Vision Graph. Image Process.* 39.3 (Sep. 1987), S. 355–368. ISSN: 0734-189X. DOI: `10.1016/S0734-189X(87)80186-X`. URL: `http://dx.doi.org/10.1016/S0734-189X(87)80186-X`.

[34] Davy Preuveneers und Yolande Berbers. "Internet of things: A context-awareness perspective". In: *The Internet of Things: From RFID to the Next-Generation Pervasive Networked Systems* (2008), S. 287–307.

[35] Victor Prisacariu und Ian Reid. *fastHOG - a real-time GPU implementation of HOG*. Techn. Ber. 2310/09. Department of Engineering Science, 2009.

[36] Ziyaad Qasem. *CM Configuration Microservice*. https://gitlab.dsalab.de/ziyaad.qasem/CMConfigurationMicroservice. 11. Juli 2019.

[37] Ziyaad Qasem. *CM Data Collection Microservice*. https://gitlab.dsalab.de/ziyaad.qasem/CMDataCollectionMicroservice. 11. Juli 2019.

[38] Ziyaad Qasem. *Context Modeling Tool*. https://gitlab.dsalab.de/ziyaad.qasem/ContextModelingTool. 11. Juli 2019.

[39] Ziyaad Qasem. *docker-system*. https://gitlab.dsalab.de/ziyaad.qasem/docker-system. 11. Juli 2019.

[40] Ziyaad Qasem. *Location Microservice*. http://195.37.209.152:10251/api/locations/api-docs. 11. Juli 2019.

[41] Ziyaad Qasem. *Multidimensional Vector Space Model Microservice*. https://gitlab.dsalab.de/ziyaad.qasem/MVSMmicroservice. 11. Juli 2019.

[42] Ziyaad Qasem u. a. "Dynamic, Adaptive, and Mobile System for Context-Based and Intelligent Support of Employees in Heavy Industry". In: (Okt. 2018). DOI: 10.1109/ES.2018.00021.

[43] Joseph Redmon. *Darknet: Open Source Neural Networks in C*. http://pjreddie.com/darknet/. 2013–2016.

[44] Joseph Redmon und Ali Farhadi. "YOLOv3: An Incremental Improvement". In: *arXiv* (2018).

[45] Joseph Redmon u. a. "You Only Look Once: Unified, Real-Time Object Detection". In: *CoRR* abs/1506.02640 (2015). arXiv: 1506.02640. URL: http://arxiv.org/abs/1506.02640.

[46] Umair Rehman und Shi Cao. "Augmented-reality-based indoor navigation: A comparative analysis of handheld devices versus google glass". In: *IEEE Transactions on Human-Machine Systems* 47.1 (2016), S. 140–151.

[47] Tomas Simon u. a. "Hand Keypoint Detection in Single Images using Multiview Bootstrapping". In: *CVPR*. 2017.

[48] Alexander Smirnov u. a. "Context Variation for Service Self contextualization in Cyber-Physical Systems". In: *International Conference on Business Information Systems*. Springer. 2015, S. 309–320.

[49] Alisa Sotsenko. "A Rich Context Model: Design and Implementation". Diss. Faculty of Technology, Linnaeus University, 2017.

[50] Alisa Sotsenko, Marc Jansen und Marcelo Milrad. “Implementing and validating a mobile learning scenario using contextualized learning objects”. In: *The 22nd International Conference on Computers in Education (ICCE), November 30, 2014 to December 4, 2014, Nara, Japan*. Asia-Pacific Society for Computers in Education. 2014, S. 522–527.

[51] Alisa Sotsenko, Marc Jansen und Marcelo Milrad. “Using a rich context model for a news recommender system for mobile users”. In: *2nd International Workshop on News Recommendation and Analytics (NRA) in conjunction with 22nd Conference on User Modelling, Adaptation and Personalization (UMAP 2014), July 11, 2014, Åalberg*. Bd. 1181. CEUR. 2014, S. 13–16.

[52] T. Surasak u. a. “Histogram of oriented gradients for human detection in video”. In: *2018 5th International Conference on Business and Industrial Research (ICBIR)*. Mai 2018, S. 172–176. DOI: 10.1109/ICBIR.2018.8391187.

[53] Olarik Surinta und Sanya Khruahong. “Tracking People and Objects with an Autonomous Unmanned Aerial Vehicle using Face and Color Detection”. In: Jan. 2019. DOI: 10.1109/ECTI-NCON.2019.8692269.

[54] *The OpenCV Reference Manual*. 2.4.9.0. Itseez. Apr. 2014.

[55] Federico Thomas und Lluis Ros. “Revisiting trilateration for robot localization”. In: *IEEE Transactions on robotics* 21.1 (2005), S. 93–101.

[56] Carlo Tomasi. “Histograms of oriented gradients”. In: *Computer Vision Sampler* (2012), S. 1–6.

[57] Mark Van Setten, Stanislav Pokraev und Johan Koolwaaij. “Context-aware recommendations in the mobile tourist application COMPASS”. In: *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. Springer. 2004, S. 235–244.

[58] Holger Wache u. a. “Ontology-Based Integration of Information-A Survey of Existing Approaches.” In: *OIS@ IJCAI*. 2001.

[59] Shih-En Wei u. a. “Convolutional pose machines”. In: *CVPR*. 2016.

[60] Greg Welch, Gary Bishop u. a. “An introduction to the Kalman filter”. In: (1995).

[61] Laurenz Wiskott und Terrence J. Sejnowski. “Slow Feature Analysis: Unsupervised Learning of Invariances”. In: *Neural Computation* 14.4 (2002), S. 715–770. DOI: 10.1162/089976602317318938. eprint: https://doi.org/10.1162/089976602317318938. URL: https://doi.org/10.1162/089976602317318938.

[62] Bichen Wu u. a. “SqueezeDet: Unified, Small, Low Power Fully Convolutional Neural Networks for Real-Time Object Detection for Autonomous Driving.” In: *CVPR Workshops*. 2017, S. 446–454.

[63] S. Wu u. a. “Exploiting Target Data to Learn Deep Convolutional Networks for Scene-Adapted Human Detection”. In: *IEEE Transactions on Image Processing* 27.3 (März 2018), S. 1418–1432. ISSN: 1057-7149. DOI: 10.1109/TIP.2017.2779271.

[64] Lina Yang und Yi Yang. “Research for personalized recommendation of learning resource in mobile computing context”. In: *Advanced Multimedia and Ubiquitous Engineering*. Springer, 2015, S. 309–315.

[65] Nico Zengeler u. a. “An Evaluation of Human Detection Methods on Camera Images in Heavy Industry Environments”. In: *2019 IEEE 14th Conference on Industrial Electronics and Applications (ICIEA)*. 2019.

[66] Daqiang Zhang u. a. “Survey on context-awareness in ubiquitous media”. In: *Multimedia tools and applications* 67.1 (2013), S. 179–211.

[67] S. Zhang u. a. “Towards Reaching Human Performance in Pedestrian Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.4 (Apr. 2018), S. 973–986. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2017.2700460.

[68] Shanshan Zhang, Jian Yang und Bernt Schiele. “Occluded Pedestrian Detection Through Guided Attention in CNNs”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Juni 2018.

[69] Qiang Zhu u. a. “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients”. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Bd. 2. Juni 2006, S. 1491–1498. DOI: 10.1109/CVPR.2006.119.

Impressum

Internal Report 19-02

ISSN: 2199-9937

1. Auflage, 01.9.2019

© Institut Informatik, Hochschule Ruhr West

Anschrift

Institut Informatik

Hochschule Ruhr West

Lützowstraße 5

46236 Bottrop