

Effiziente und flexible Mensch-Roboter-Interaktion für die Logistik der Zukunft

# Interaktionsfähiger mobiler Roboter in der Logistik

J. Berger, S. Lu, C. Pest

**ZUSAMMENFASSUNG** Interaktive mobile Roboter bieten Potenzial, Effizienz und Ergonomie in der Logistik zu steigern. Dieser Beitrag beschreibt die Entwicklung, Implementierung und Evaluation eines Interaktionssystems, das auf die Anforderungen eines Logistikumfelds abgestimmt ist. Basierend auf einer modularen Architektur kombiniert das Robotersystem Interaktionstechnologien mit robuster Aufgabenverwaltung. Eine Evaluation im Anwendungsfall der Maßartikellagerung zeigt die Vorteile intuitiver Mensch-Roboter-Interaktion und Optimierungspotenziale, insbesondere in Sprachsteuerung und Tätigkeitsdetektion.

## STICHWÖRTER

Mensch und Technik, Handhabungstechnik, Industrieroboter

**Interactive mobile robot in logistics – Efficient and flexible human-robot interaction with a view to the logistics of the future**

**ABSTRACT** Interactive mobile robots allow for enhancing efficiency and ergonomics in logistics. This paper presents the development, implementation, and evaluation of an interaction system tailored to the requirements of a logistics environment. Based on a modular architecture, the robotic system combines interaction technologies with robust task management. An evaluation in the use case of storing customized items highlights the advantages of intuitive human-robot interaction and identifies potential for optimization, particularly in speech control and activity detection.

## 1 Einleitung und Motivation

In modernen Logistikumgebungen, wird weiterhin der Großteil der Arbeiten durch das Personal durchgeführt, wie der Anwendungsfall zeigt. Um die Mitarbeitenden zu entlasten, können mobile Roboter eingesetzt werden [1]. In diesem Kontext werden mobile Roboter als eine Kombination aus einem fahrerlosen Transportsystem (FTS) und einem kollaborativen Roboter definiert [2]. Mobile Roboter sind in der Lage, Transport- und Handhabungsaufgaben zu übernehmen. Sie können sich autonom im Raum bewegen, Objekte erkennen und deren Position erfassen.

Jedoch zeigt sich, dass die reine Automatisierung in der Logistik nicht ausreicht, wenn keine Interaktionsfähigkeiten vorhanden sind [3, 4]. Die menschliche Fachkraft bleibt unverzichtbar, vor allem bei komplexen oder variierenden Aufgaben, die Flexibilität und schnelle Anpassung erfordern. Eine Zusammenarbeit zwischen Menschen und Robotern bietet entscheidende Vorteile: Sie steigert die Effizienz, entlastet die Mitarbeitenden und schafft eine höhere Flexibilität im Arbeitsablauf [5].

Ein mobiler Roboter, der mit Menschen interagieren kann, hebt diese Vorteile auf eine neue Stufe, indem er spezifisch auf die Bedürfnisse und die Arbeitsprozesse der Mitarbeitenden ein geht. Als Teil des Forschungsprojekts „FORobotics“ [6] wurde ein interaktionsfähiger mobiler Roboter entwickelt. Dieser Roboter kann beispielsweise durch eine einfache Geste gestoppt werden, ermöglicht das Abfragen des Akkustands und erlaubt es den Mitarbeitenden, dem Roboter Aufgaben wie „Hole mir 3 Schrauben“ zu erteilen. Diese Funktionen fördern eine intuitive und effektive

Mensch-Roboter-Interaktion, welche die Akzeptanz und Effizienz in der kollaborativen Arbeit erhöht.

Der Ansatz des interaktionsfähigen Roboters wurde im Projekt „MeMoRob“ zugrunde gelegt und mithilfe von Methoden einer nutzerzentrierten Entwicklung angepasst und erweitert. Um die Interaktion zwischen Menschen und Robotern im logistischen Umfeld effektiv zu gestalten, müssen sowohl die Bedürfnisse der Mitarbeitenden als auch die Anpassungsfähigkeit des Systems berücksichtigt werden. Dies stellt sicher, dass der Roboter flexibel auf die spezifischen Anforderungen der Logistik reagieren kann und eine optimale Unterstützung bietet.

Im Beitrag wird anhand eines praktischen Anwendungsfalls in der Logistik gezeigt, wie ein solcher interaktionsfähiger mobiler Roboter entwickelt wird.

## 2 Anforderungen und Problemstellung

### 2.1 Analyse der Umgebung und der Arbeitsaufgabe

Die Firma Juzo stellt medizinische textile Hilfsmittel her. Die Tätigkeiten im Logistikzentrum sind hauptsächlich manueller Art, abgesehen von einer Maschine zum Falten von Kartons. Die Haupttätigkeiten sind Logistikprozesse wie Einlagerung und Kommissionierung. Im Beitrag wird der Schwerpunkt auf die Einlagerung gelegt, insbesondere von Maßartikeln, die größtenteils in Kunststoffbeuteln mit Druckverschluss verpackt sind (Bild 1). Die Beutel haben unterschiedliche Gewichte und sollen in Regale (Bild 2) eingelagert werden.



Bild 1. Beutel für Maßartikel. Foto: Fraunhofer IGCV

Der Prozessablauf für die manuelle Einlagerung von Maßartikeln im Logistikzentrum gestaltet sich wie folgt:

1. Ankunft der Kiste auf dem Förderband
2. Abholung einer oder mehrerer Kisten mit einem Rollbrett
3. Bewegung zu Maßregallager
4. Einlagerung jeweils eines Maßartikels in freien Lagerplatz durch Scannen des Barcodes auf Beutel und Scannen des Lagerplatzes
5. Bewegung zu Förderband zur Ablage der Kiste

Diese Aufgaben soll der mobile Roboter übernehmen. Er soll die Kiste selbstständig von der Abholposition abholen, zum Lagerbereich navigieren, den Barcode des Artikels sowie den Lagerplatz scannen und die Kiste nach der Einlagerung zurückbringen. Die im Logistikzentrum tätigen Fachkräfte bewegen sich im gleichen Bereich wie der Roboter.

## 2.2 Beschreibung der Anforderungen

Um die Anforderungen an das Gesamtsystem zu identifizieren, wurden neben der Prozessanalyse Interviews mit Lagerfachkräften durchgeführt, um einerseits ihre Gedanken zum Einsatz des Roboters und andererseits Anforderungen an die Interaktionssysteme ableiten zu können. Das genaue Vorgehen für die Interviews ist in *Berger et al. [1]* zusammengefasst.

Die **Tabelle** beschreibt ausgewählte Anforderungen an das Interaktionssystem, mit dem die Fachkräfte im Lager den Roboter steuern, mit ihm kommunizieren und Probleme beheben können.



Bild 2. Regal für Maßartikel. Foto: Fraunhofer IGCV

Durch Beachtung dieser Anforderungen soll sichergestellt werden, dass das System benutzerfreundlich, effizient und robust ist.

Das Interaktionssystem soll ein Tablet enthalten, mit dem die Mitarbeitenden intuitiv Aufträge eingeben, pausieren oder fortsetzen können. Zudem soll es den Status des Roboters und des aktuellen Auftrags anzeigen. Die Anforderungen an den Informationsaustausch beschreiben, dass relevante Informationen, wie der Batteriestatus oder die verbleibende Dauer bis zum Ende eines Auftrags, abgefragt und durch das Interaktionssystem ausgegeben werden können.

Ein wichtiges Element eines mobilen Robotersystems ist die Auftragsverwaltung. Das Interaktionssystem soll es ermöglichen, Aufträge aufzunehmen, zu pausieren oder abzubrechen. Ein Auftrag kann etwa die „Einlagerung von Maßartikeln“ oder „Fahren zur Ladestation“ umfassen. Solche Aufträge bestehen aus mehreren spezifischen Aufgaben, wie etwa „Fahren“ und „Handhaben“. Über eine Sprachsteuerung kann der Roboter direkt angesprochen werden, was die Bedienung vereinfacht. Bei Problemen unterstützen Fehlermeldungen und Anleitungen zur Fehlerbehebung die Fachkräfte, und es besteht die Möglichkeit, Hilfe anzufordern. Die Umgebungserkennung erlaubt es dem System, auf die Bewegungen und Gesten von Personen zu reagieren, mit dem Ziel, die Interaktion sicher zu gestalten. Darüber hinaus kann das System flexibel für neue Aufgaben konfiguriert werden, sodass seine Einsatzmöglichkeiten an sich ändernde Anforderungen angepasst werden können.

**Tabelle.** Anforderungen an das Interaktionssystem.

Tablet	Das Tablet soll intuitiv bedient werden können. Über das Tablet sollen Aufträge eingegeben werden können. Ein Bedienungsgerät, das die aktuellen Informationen zum Auftrag oder Stand des Roboters anzeigt.
Auftrag	Neuen Auftrag eingeben
	Auftrag pausieren
	Wiederaufnahme der pausierten Aufgabe
	Abbrechen des Auftrags
	Roboter soll Lager verlassen
Informationsaustausch	Informationsabfrage: z.B. Dauer bis zur Fertigstellung des Auftrags, Batteriestatus Informationsausgabe: Antwort auf Informationsabfragen
Ansprache	Roboter mit Schlüsselwort ansprechen Sprachsteuerung per Button aktivieren
Fehlermanagement und Warnung	Fehlermeldung Anleitung zur Fehlerbehebung Hilfe anfordern
Umgebungserkennung	Erkennung der Tätigkeiten und Geste von Personen
Flexibilität	Leicht konfigurierbar und anpassbar für neue Aufgaben

### 3 Interaktionsfähigkeiten des mobilen Roboters

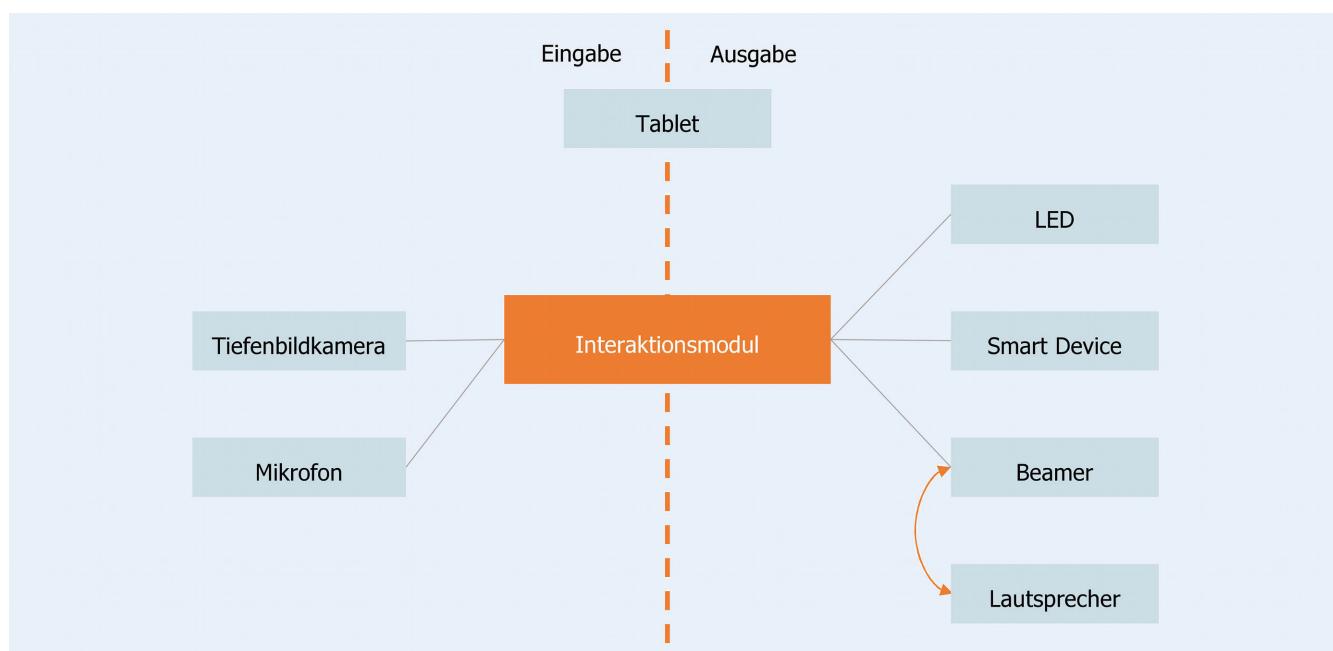
#### 3.1 Konzeption des Interaktionsmoduls

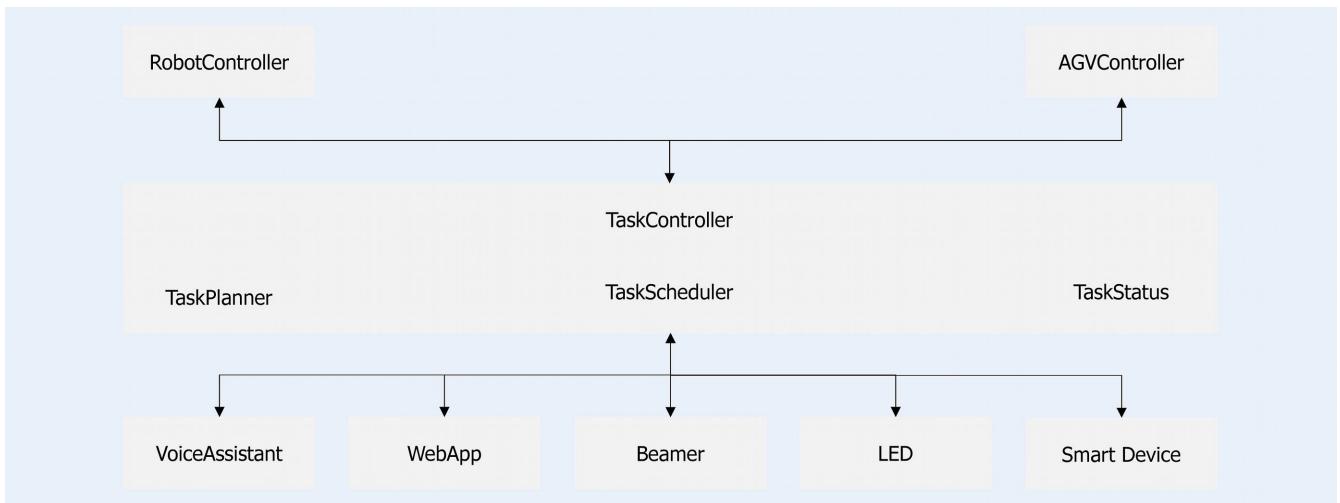
Basierend auf den erarbeiteten Anforderungen wurde das Konzept des Interaktionsmoduls erstellt, um eine benutzerfreundliche und effiziente Interaktion zwischen den Mitarbeitenden und dem Roboter im Lager zu gewährleisten.

#### 3.1.1 Beschreibung der erforderlichen Komponenten

Das Interaktionsmodul besteht aus verschiedenen Komponenten, die in Ein- und Ausgabegeräte unterteilt sind, wie im **Bild 3** dargestellt.

Zu den Eingabegeräten gehören Tablet, Tiefenbildkamera und Mikrofon. Diese Geräte erlauben die Interaktion durch Eingaben, zum Beispiel das Einsprechen von Befehlen oder die Erfassung

**Bild 3.** Interaktionsmodul mit seinen Komponenten (Tablet und Smart Device drahtlos, andere festmontiert). *Grafik: Fraunhofer IGCV*



**Bild 4.** Modularer Aufbau des Interaktionsmoduls. *Grafik: Fraunhofer IGCV*

der Umgebung durch die Kamera. Die Ausgabegeräte umfassen LED-Leuchten, Smart Device, Beamer und Lautsprecher. Mithilfe der Ausgabegeräte werden visuelle oder akustische Rückmeldungen gegeben. Das Tablet dient sowohl als Eingabe- als auch als Ausgabegerät, da es nicht nur Befehle entgegennimmt, sondern auch Informationen anzeigt. Wie zu erkennen, gibt es mehrere Ausgabegeräte, die die gleichen Informationen anzeigen können. Dies dient dazu, dass in unterschiedlichen Situationen verschiedene Ausgabemöglichkeiten zur Verfügung gestellt werden können.

### 3.1.2 Zuordnung der Funktionen in der Anforderung zu den Interaktionskomponenten

Für die Konzeption des Moduls wurde zunächst betrachtet, welche Funktionen erfüllt werden müssen und daraus abgeleitet, welche Interaktionstechnologien eingesetzt werden können. Zur Anzeige von Informationen, wie etwa Fehlermeldungen, kann ein Tablet verwendet werden sowie mit einem Beamer Informationen auf den Boden projiziert werden. Über die Sprachtechnologie können sowohl Ein- als auch Ausgaben gemacht werden. Aus den bisherigen Projekten ging hervor, dass der Einsatz von Gesten nur für bestimmte Zwecke, zum Beispiel das Anzeigen des „Stopp-Signals“ sinnvoll verwendbar ist [6].

Abgeleitet aus den Anforderungen, wurden neben den genannten Systemen auch die LED-Leuchte und das Smart Device eingesetzt. Eine LED-Leuchte am Roboter soll einfach über den Zustand des Roboters aufklären, was von den zukünftigen Nutzern in den Interviews als wichtig angesehen wurde. So soll eine LED-Leuchte am Roboter etwa angelehnt sein an die Betriebsleuchte einer Anlage und beispielsweise im Fehlerfall „Rot“ leuchten. Das Smart Device dient dazu, den Roboterdienst – eine speziell geschulte Person – bei der Überwachung des Roboters zu unterstützen, indem es Echtzeitinformationen über den Zustand des Roboters und Fehlermeldungen bereitstellt. Dank der WLAN-fähigen Smartwatch kann sich der Roboterdienst flexibel vom Roboter entfernen, andere Aufgaben übernehmen und dennoch Benachrichtigungen empfangen. So bleibt der Roboterdienst stets informiert, kann Meldungen direkt auf der Smartwatch einsehen und bei Bedarf zum Roboter zurückkehren. Diese Flexibilität erhöht die Effizienz und verbessert die Benutzerfreundlichkeit.

des Systems. Eine Tiefenbildkamera wird für die Gesten- und Tätigkeitserkennung eingesetzt.

## 3.2 Umsetzung der Funktionen

Flexibilität als Anforderung lässt sich durch Modularität und den Einsatz eines Konfigurators optimal umsetzen. Ein modular aufgebautes System erlaubt es, einzelne Komponenten unabhängig zu entwickeln, zu aktualisieren oder zu ersetzen, was Anpassungen an veränderte Anforderungen erleichtert, ohne die gesamte Struktur zu beeinträchtigen. Ein Konfigurator ergänzt dies, indem er eine benutzerfreundliche Gestaltung und Anpassung des Systems an spezifische Bedürfnisse ermöglicht, ohne tiefgehende technische Eingriffe.

Im Folgenden wird zunächst die Implementierung der Gesamtstruktur beschrieben; danach werden die einzelnen Module vorgestellt, gefolgt von einer Darstellung des Konfigurators.

### 3.2.1 Gesamtstruktur des Interaktionssystems

Ein modularer Aufbau (Bild 4) des Interaktionsmoduls wurde entwickelt, um die Interaktion mit dem mobilen Roboter zu ermöglichen.

Dazu wurden verschiedene Module erstellt, die über MQTT-Nachrichten miteinander verbunden sind. MQTT steht für „Message Queueing Telemetry Transport“ und ist ein leichtgewichtiges Kommunikationsprotokoll, das speziell für den Einsatz in IoT-Anwendungen entwickelt wurde. Mit MQTT können Module Nachrichten veröffentlichen (publish) und Nachrichten abonnieren (subscribe). Wenn ein Modul eine Nachricht veröffentlicht, können andere Module, welche diese Nachricht abonniert haben, diese empfangen und entsprechend darauf reagieren. Dies ermöglicht eine flexible und skalierbare Architektur für die Verbindung verschiedener Module.

### 3.2.2 TaskController

Der TaskController läuft im Hintergrund. Er besteht aus drei Untermodulen, nämlich dem TaskPlanner, dem TaskStatus und dem TaskScheduler. Der TaskPlanner ist dafür verantwortlich, Aufgaben zu planen und zu koordinieren. Er verwaltet die

Aufgabenliste, organisiert die Reihenfolgen der Unteraufgaben und stellt sicher, dass sie entsprechend den vorgegebenen Kriterien und Prioritäten ausgeführt werden.

Der TaskScheduler steuert sowohl das FTS als auch den Roboter. Er synchronisiert die Bewegungen und Aktivitäten der mobilen Roboter, plant ihre Einsätze nach definierten Abfolgen von dem TaskPlanner und koordiniert ihre Aufgaben. So ermöglicht er eine transparente Überwachung und Nachverfolgung des Aufgabenstatus für eine effiziente Interaktion durch verschiedene Module.

Der TaskController ist mit dem RobotController und dem AGVController verbunden. Dank der Modularität des TaskControllers ist es möglich, verschiedene Arten von mobilen Robotern anzuschließen. Diese Flexibilität ermöglicht es, die Interaktion mit den mobilen Robotern entsprechend zu konfigurieren. So kann der TaskController nahtlos mit unterschiedlichen Robotertypen zusammenarbeiten und ihre Bewegungen, Aktionen und Aufgabensteuerung effizient koordinieren.

### 3.2.3 VoiceAssistant

Das Modul VoiceAssistant ermöglicht die Spracheingabe und Sprachausgabe. Dabei kann der Benutzer Befehle erteilen, um bestimmte Aktionen durchzuführen oder Informationen über den Status des Robotersystems abzufragen. Das Modul umfasst eine Hotword-Erkennung, die auf ein bestimmtes Aktivierungswort reagiert, um das System in den Zuhörmodus zu versetzen. Zudem ist eine Natural Language Understanding (NLU)-Komponente integriert, welche die Bedeutung der gesprochenen Befehle analysiert und interpretiert. Für eine flexible Hotword-Erkennung wurde die regelmäßig aufgezeichnete Sprache hinsichtlich der ungefähren Länge des Hotwords überprüft. Anschließend wurde das Hotword mithilfe des Levenshtein-Maßes zeichenweise mit den erkannten Wörtern verglichen.

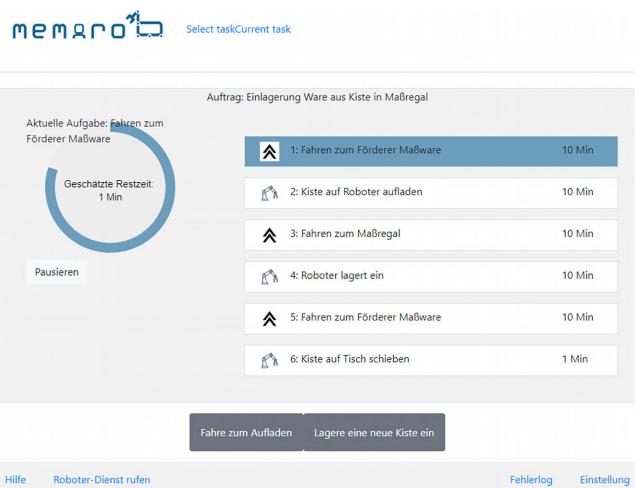
Für die Implementierung der NLU-Komponente wird „Rasa“ verwendet, ein Open-Source-Framework für die Entwicklung von Conversational-AI-Systemen. Die extrahierten Absichten durch die NLU-Komponente werden an den TaskController weitergeleitet, um entsprechende Antworten oder Aktionen auszulösen.

### 3.2.4 WebApp

Das Modul „WebApp“ ist eine Webanwendung, die speziell für Tablets entwickelt wurde. Diese Webanwendung ermöglicht die Verwaltung von Aufträgen und Fehlern, die Konfiguration neuer Anwendungsfälle und die Erstellung von Anleitungen zur Fehlerbehebung. Es bietet die Möglichkeit, Aufgabenbeschreibungen im JSON-Format zu laden und anzuzeigen. Jede Aufgabe kann aus mehreren Unteraufgaben bestehen, wobei die Anwendung den Bearbeitungsstand der einzelnen Unteraufgaben visualisiert (**Bild 5**). Außerdem können Benutzer Unteraufgaben flexibel pausieren oder abbrechen. Durch die flexible Struktur kann das Modul auch neue konfigurierte Aufgaben anzeigen. Die Webanwendung wird mit Flask umgesetzt.

### 3.2.5 Beamer, LED-Leuchten und Smart Device

Der Beamer projiziert den aktuellen Status und die Antworten auf Fragen auf den Boden. Dafür wurde eine C#-Desktop-Anwendung erstellt. Der Beamer abonniert Nachrichten via MQTT von



**Bild 5.** Beispiel WebApp-Ansicht. *Grafik: Fraunhofer IGCV*

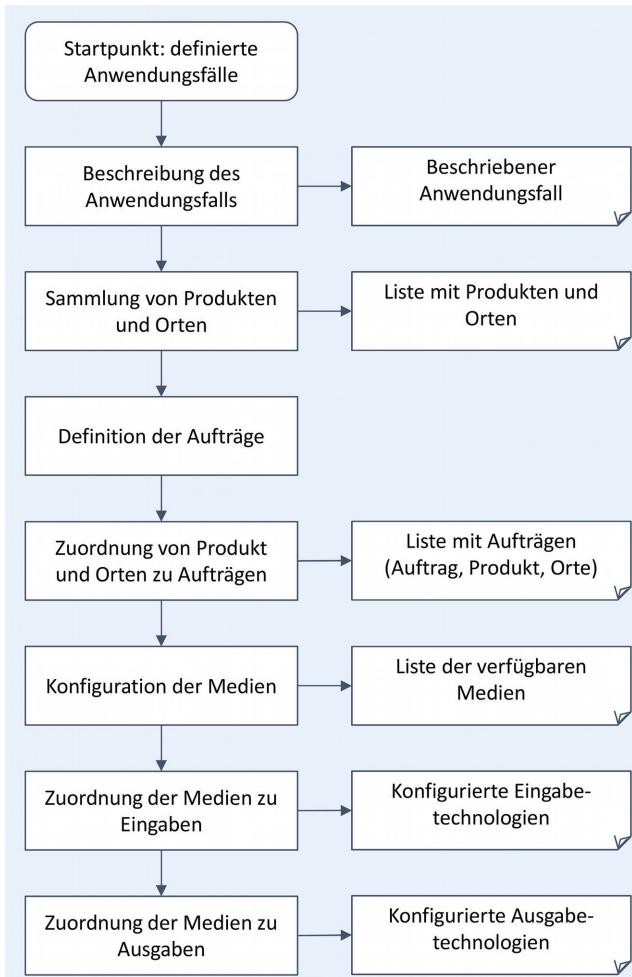
anderen Modulen und zeigt sie an. Zur Ansteuerung der LED-Leuchten wird ein Raspberry-Pi verwendet, der über die GPIO-Pins eine direkte Steuerung über MQTT-Nachrichten ermöglicht. Die Smartwatch-Anwendung wird mithilfe eines Samsung Galaxy umgesetzt, wobei der Fokus auf der Fehlermeldung liegt. Die Kommunikation erfolgt über die bereits erwähnten MQTT-Nachrichten, die in HTTP-Requests umgewandelt und an die Smartwatch gesendet werden. Die Übertragung erfolgt über WLAN. Nach dem Empfang der Nachrichten werden diese auf der Smartwatch angezeigt.

### 3.2.6 Erkennung der Tätigkeiten und Gesten von Personen

Die Tätigkeitserkennung dient dazu, die Interaktion des Roboters mit den Mitarbeitenden effizienter zu gestalten, indem der Roboter auf die Bewegungen und Handlungen der Mitarbeitenden reagieren kann. Dies könnte in zukünftigen Szenarien zur Vermeidung von Konflikten zwischen menschlichen und robotischen Tätigkeiten genutzt werden. Zur Erkennung von Tätigkeiten und Gesten von Personen wurde der RGB (Rot-Grün-Blau)-Stream einer Tiefenkamera genutzt, da vorge trainierte Modelle und umfangreiche Datensätze für RGB-Daten verfügbar sind.

Angelehnt an Niemann et al. [7] wurden fünf Tätigkeiten definiert, die erkannt werden sollen: 1) Stehen, 2) Gehen, 3) Gehen mit Wagen, 4) Handhabung (Objekt) und 5) Unbekannt. Diese Erkennung basiert auf dem LARA-Datensatz [7].

Zusätzlich wurde ein eigener Datensatz erstellt, um die Übertragbarkeit auf den spezifischen Einsatzbereich zu überprüfen. Zur Implementierung wurden vorge trainierte Open-Source-Modelle von OpenMMLab und PyTorchVideo eingesetzt. Zum einen das Modell „VideoMAEv2“ [8], das auf Vision Transformern (ViT) [9] basiert und zum anderen das X3D-Modell [10], das auf Convolutional Neural Networks (CNNs) setzt, um Effizienz und Genauigkeit zu vergleichen. Das Modell „VideoMAEv2-base“ erzielte im Einsatzbereich die höchste Genauigkeit von 84 %, während die X3D-Modelle zwar schneller, jedoch weniger genau waren. Zur Gestenerkennung wurde das Jester-Datenset [11] mit dem X3D-XS-Modell verwendet, um die „Stopp“-Geste zu erkennen.



**Bild 6.** Prozessablauf zur Konfiguration neuer Aufgaben im Interaktionssystem. *Grafik: Fraunhofer IGCV*

### 3.2.7 Konfigurator des Interaktionssystems

Wie in den Anforderungen beschrieben, soll das Interaktionssystem so gestaltet sein, dass es sich leicht für neue Aufgaben konfigurieren lässt. Dazu wurde ein Prozessablauf definiert, der in **Bild 6** dargestellt ist.

Auf der rechten Seite jedes Schritts befindet sich das Ergebnis des jeweiligen Schritts, das in Form einer JSON-Datei gespeichert wird. Diese JSON-Dateien dienen zur strukturierten Dokumentation und Speicherung der Konfiguration. Sie werden in nachfolgenden Prozessen bei der Validierung von Eingabedaten und bei der Ergänzung und Aktualisierung bestehender Konfigurationen eingesetzt. Der Ablauf wird in einer WebApp als Schritt-für-Schritt-Anwendung umgesetzt. So kann der Benutzer das Interaktionssystem intuitiv und benutzerfreundlich konfigurieren, indem er die einzelnen Schritte nacheinander durchläuft und die erforderlichen Einstellungen vornimmt.

Das Diagramm beginnt mit der Definition der Anwendungsfälle. In diesem Schritt werden die spezifischen Anforderungen und Szenarien festgelegt, die das Interaktionssystem unterstützen soll. Anschließend folgt die Beschreibung des Anwendungsfalls, in der die detaillierten Prozesse und Abläufe ausgearbeitet werden. Ein Anwendungsfall kann beispielsweise sein, dass ein Roboter in einer Lagerumgebung eingesetzt wird und Kisten transportiert.

Danach erfolgt die Sammlung von Produkten und Orten, die im System berücksichtigt werden müssen. Orte für einen Anwendungsfall könnten zum Beispiel das Förderband als Abholort und die Montagestation als Zielort sein.

Im nächsten Schritt erfolgt die Definition der Aufträge. Dabei werden konkrete Aufgaben und Ziele festgelegt, die das Interaktionssystem ausführen soll, basierend auf den zuvor definierten Produkten und Orten. Diese Aufträge spezifizieren, welche Produkte von welchem Ausgangsort zu welchem Zielort transportiert oder verarbeitet werden sollen. Beispiele sind der Transport eines Motors vom Lager zum Montagetisch oder das Bewegen einer Kiste von der Abholposition beim Förderband zum Maßregal. Danach werden die Produkte und Orte den jeweiligen Aufträgen zugeordnet und die Informationen in einer strukturierten Datenform gespeichert, um die Konfiguration klar und nachvollziehbar zu dokumentieren. Das Ergebnis wird zusätzlich in der Benutzeroberfläche angezeigt, um dem Benutzer eine transparente und überprüfbare Übersicht der konfigurierten Aufträge und deren Parameter zu bieten.

Im darauffolgenden Schritt wird das Kommunikationsmedium ausgewählt, wie beispielsweise ein Tablet, Sprache, Beamer oder Gestenerkennung. Die Auswahl des Mediums hängt vom Robotersystem ab, da unterschiedliche Systeme unterschiedliche Kommunikationsanforderungen haben können. Anschließend werden die Medien jeweils der Eingabe und Ausgabe zugeordnet. Die Zuordnung der Medien erfolgt zunächst für grundlegende Befehle wie Starten, Pausieren, Weitermachen oder Abbrechen eines Auftrags. Zusätzlich können die Medien anhand der spezifischen Anforderung eines Auftrags angepasst werden. Beispielsweise kann der neue Auftrag „Hole Kiste von Förderband“ dem Medium Sprache zugeordnet werden, sodass der Auftrag über die Sprachsteuerung eingegeben werden kann.

## 4 Evaluation

Das entwickelte System wurde im Logistikzentrum von Juzo eingesetzt und evaluiert. Während der Evaluation übernahm der Roboter die Aufgabe der Einlagerung, wobei Mitarbeitende die Möglichkeit hatten, mit dem Roboter zu interagieren. Ziel der Evaluation war es, die Interaktionsfähigkeit und Benutzerfreundlichkeit des Roboters zu bewerten.

Die Evaluation erfolgte in zwei Phasen: Zunächst testeten Ingenieure die Funktionalität des Systems, um sicherzustellen, dass die technischen Anforderungen erfüllt werden. Anschließend wurde ein Usability-Test mit vier Lagerfachkräften durchgeführt, um die Benutzerfreundlichkeit und Akzeptanz der Mensch-Roboter-Interaktion im praktischen Einsatz zu bewerten.

Die Evaluation der Funktionalität bestätigt, dass die Gesamtstruktur des Systems, basierend auf einer modularen Architektur und der Kommunikation über MQTT, robust und effektiv ist. Die Verwendung von MQTT als leichtgewichtiges Protokoll erlaubte eine zuverlässige und latenzarme Kommunikation zwischen den Modulen. Der TaskController konnte Aufgaben effizient planen und ausführen. Der Beamer erfüllte seine Funktion, indem er Statusinformationen und Antworten wie vorgesehen projizierte. Zusätzlich zeigten LED-Leuchten Statusinformationen an, etwa ob sich der Roboter in Bewegung oder im Fehlerzustand befindet. Im Fehlerfall wurde dies ebenfalls über das Smart Device angezeigt. Die WebApp ermöglichte es den Nutzenden, Aufträge wie die Einlagerung von Maßartikel oder das Fahren zur Ladestation

direkt zu starten. Auch das Pausieren und Fortsetzen von Teilaufgaben war möglich. Zusätzlich konnte über die WebApp eine neue Aufgabe für das Interaktionssystem konfiguriert werden. Außerdem konnten Statusinformationen und Fehlermeldungen über die Smart Watch angezeigt werden.

Zur Evaluation der Tätigkeitserkennung wurden die Modelle „X3D“ und „VideoMAE“ auf dem LARA-Datensatz fine-tuned und anschließend auf einem eigenen erstellten Datensatz evaluiert. Der Datensatz umfasste 148 Videos pro Tätigkeit. Das Modell „VideoMAEv2-base“ erzielte dabei eine größte Genauigkeit von 84 %. Dabei war die Erkennung der Tätigkeit „Stehen“ am ungenauesten, besonders in der Abgrenzung zur Tätigkeit „Gehen mit Wagen“. Dies liegt vermutlich an ähnlichen visuellen Merkmalen, insbesondere wenn der Wagen nur minimal bewegt wird oder die Bewegungen kaum sichtbar sind.

Die Evaluation hat gezeigt, dass die Anwendung großer, vortrainierter Modelle im robotischen Kontext eine effiziente Lösung darstellt. Sie ermöglicht auch bei kleinen Datensätzen eine hohe Genauigkeit, da die Modelle bereits allgemeine Merkmale aus großen Trainingsdatensätzen erlernt haben und diese effektiv auf spezifische Anwendungsfälle übertragen können. Allerdings kann eine falsche Erkennung von Tätigkeiten die Handlungsfähigkeit des Roboters erheblich beeinträchtigen. Dies verdeutlicht die Notwendigkeit, spezifischere Trainingsdaten zu erstellen, um die Erkennungsgenauigkeit weiter zu verbessern. Bei der Erkennung der „Stopp“-Geste waren die Ergebnisse nicht robust, da auch andere Bewegungen fälschlicherweise als „Stopp“ erkannt wurden.

Bei der Evaluation der Funktionalität des VoiceAssistants wurde festgestellt, dass die Erkennung des „HotWords“ problematisch war. Wenn das System aktiviert war, zeigte es eine deutlich bessere Fähigkeit, die Benutzerintentionen zu erkennen. Die Hauptursachen lagen vermutlich in Umgebungsgeräuschen, Variationen in der Aussprache und der Sensitivität der HotWord-Erkennung, wodurch das System entweder nicht zuverlässig reagierte oder fälschlicherweise aktiviert wurde. Dies lässt vermuten, dass eine Anpassung an unterschiedliche akustische Bedingungen sowie ein KI-Modell speziell für die HotWord-Erkennung erforderlich sind, um die Zuverlässigkeit des VoiceAssistants zu verbessern.

Nach der Evaluation der Funktionalität wurden Usability-Tests durchgeführt. Zu Beginn der Evaluation der Benutzerfreundlichkeit erhielten die Mitarbeitenden eine kurze Einführung in Form eines fünfminütigen Video-Tutorials. Dieses stellte die grundlegenden Funktionen des Roboters sowie die verschiedenen Bedienmöglichkeiten vor, darunter das Tablet, die Projektion und die Sprachsteuerung. Ziel des Tutorials war es, Berührungsängste gegenüber dem Roboter abzubauen und die wichtigsten Bedienoptionen kompakt zu erklären. Zudem wurden essenzielle Sicherheitsaspekte wie die Notausfunktion und deren Verwendung detailliert erläutert, um eine sichere Interaktion mit dem System zu gewährleisten.

Jeder Proband erhielt vier Aufgaben, die mit dem Interaktionssystem durchgeführt werden sollten. Dabei wurden die Leistung des Systems sowie das Verhalten der Probanden während der Aufgabenerfüllung beobachtet und dokumentiert. Die Aufgabenstellungen sowie die jeweiligen Ergebnisse werden im Folgenden detailliert dargestellt und diskutiert.

#### 1. Starten das Auftrags „Einlagerung Maßartikel“:

Ziel dieser Aufgabe war es, den Prozess der Auftragsinitialisierung über die WebApp zu testen. Die Probanden mussten die Einlagerung eines Maßartikels starten und den Fortschritt über

die WebApp verfolgen. Eine Person konnte die Aufgabe erfolgreich abschließen, eine weitere hatte geringfügige Probleme, die aber mithilfe von Hinweisen gelöst werden konnten. Zwei Personen konnten die Aufgabe nicht lösen. Die Benutzerführung der WebApp war für einige Teilnehmende nicht intuitiv genug. Auf der ersten Seite der WebApp standen vier Optionen zur Verfügung: „Tutorial Konfigurator“, „Interaktionskonfigurator“, „Tutorial“, „Durchführung von Aufgaben“. Es stellte sich heraus, dass nicht alle erkannten, dass „Durchführung von Aufgaben“ der richtige Einstiegspunkt war, was zu Verwirrung führte. Eine klarere Beschriftung, visuelle Hervorhebung oder ein kurzer Hinweis könnten helfen, die Navigation zu verbessern.

#### 2. Roboter pausieren und weitermachen über das Tablet:

Eine Person konnte die Aufgabe erfolgreich lösen, eine weitere stieß auf geringfügige Probleme, welche jedoch mit Hinweisen behoben werden konnten. Zwei Personen waren nicht in der Lage, die Aufgaben zu bewältigen. Das Hauptproblem war ähnlich wie bei der ersten Aufgabe: Die Teilnehmenden konnten die Seite zur Übersicht aller Teilaufgaben nicht erreichen und hatten Schwierigkeiten, die richtige Navigation zu finden. Um dies zu verbessern, sollte die Navigation eindeutiger gestaltet und der Zugriff auf die Übersichtsseite klarer hervorgehoben werden.

#### 3. Roboter zum Laden schicken über die Sprachsteuerung:

Keine der teilnehmenden Personen konnte die Aufgabe erfolgreich lösen. Zwar war allen Teilnehmenden bewusst, dass der Roboter mit einem „Hotword“ wie „Hallo Robo“ angesprochen werden muss, aber einige Personen wussten nicht, dass das System zunächst aktiviert werden muss, bevor der Befehl „Fahren zur Ladestation“ eingegeben werden kann. Dieses Konzept, ähnlich wie bei den Sprachassistenten „Alexa“ oder „Google Assistant“ im privaten Umfeld, führte bei einigen zu Verwirrung, da sie davon ausgingen, dass das System direkt auf das Hotword reagieren würde. Um die Nutzerfreundlichkeit zu verbessern, sollte der Aktivierungsschritt klarer kommuniziert werden. Zusätzlich war einigen Personen nicht klar, wo das Mikrofon installiert ist, weshalb Befehle nicht in der Nähe des Mikrofons gesprochen wurden, was die Erkennung weiter erschwerte. Es sollte klar kommuniziert werden, dass das System erst aktiviert werden muss und dass Sprachbefehle in der Nähe des Mikrofons eingesprochen werden sollten, unterstützt durch visuelle Hinweise oder eine kurze Anleitung zur Nutzung.

#### 4. Fehlerfall „Hindernis im Weg“ beheben:

Ziel dieser Aufgabe war es, die Reaktion der Teilnehmenden und die Unterstützung des Systems im Fehlerfall zu testen, wenn ein Hindernis im Weg steht. Dabei wurde den Personen ein Tutorial als Schritt-für-Schritt-Anleitung auf dem Tablet angezeigt. Drei Personen konnten die Aufgabe erfolgreich lösen, während eine Person auf Schwierigkeiten stieß, die nur mit zusätzlicher Unterstützung überwunden werden konnten. Obwohl ein Tutorial über die WebApp auf dem Tablet verfügbar war, konnten die meisten Teilnehmenden die Aufgabe ohne dessen Nutzung erfolgreich abschließen. Dies liegt vermutlich daran, dass die Vorgehensweise bereits im Einführungstutorial erklärt wurde. Dies deutet darauf hin, dass eine ausführliche Erklärung aller relevanten Aufgaben und Funktionen vor dem eigentlichen Einsatz hilfreich ist, um Unsicherheiten zu minimieren und die Erfolgsquote weiter zu erhöhen.

Allgemein empfanden alle Teilnehmenden die Bedienung des Roboters nach einer kurzen Eingewöhnungszeit als einfach und intuitiv.

Besonders positiv wurden die Projektion und die visuellen Hinweise auf dem Boden sowie am Roboter selbst bewertet, da sie effektiv dabei halfen, Probleme zu erkennen und den Workflow zu unterstützen. Die Sprachsteuerung wurde grundsätzlich als nützlich angesehen, allerdings zeigten sich hier Schwächen in der Zuverlässigkeit, die zu Frustrationen bei den Nutzenden führten. Zudem wurde angemerkt, dass das Design des Tablets optimierungsbedürftig ist, um die Benutzerfreundlichkeit zu erhöhen.

Verbessert werden könnte es etwa durch die Ergänzung zusätzlicher Informationen für Buttons, einer optimierten Größe der Buttons, klareren Beschriftungen sowie visuellen Hervorhebungen. Eine ausführliche Einführung oder Schulung vor der Nutzung könnte die Bedienung weiter erleichtern und zu einer höheren Zufriedenheit beitragen.

## 5 Zusammenfassung und Ausblick

Der vorliegende Beitrag zeigt die Entwicklung und Evaluation eines interaktionsfähigen mobilen Roboters für die Logistik. Ziel war es, logistische Prozesse wie die Einlagerung von Maßartikeln zu automatisieren und durch eine intuitive Mensch-Roboter-Interaktion die Effizienz und Akzeptanz des Systems zu steigern.

Ausgehend von einer umfassenden Anforderungsanalyse wurden Interaktionsmodule konzipiert und in einem modularen System umgesetzt. Das Herzstück bildet der TaskController, der die Planung und Ausführung von Aufgaben koordiniert. Ergänzt wird dies durch verschiedene Eingabe- und Ausgabemodule, wie eine WebApp, Sprachsteuerung, Beamer und LED-Anzeigen, die eine flexible und benutzerfreundliche Interaktion ermöglichen. Spezifische Technologien, wie die Gestenerkennung und die Nutzung vortrainierter Modelle für Tätigkeitsanalysen, wurden erfolgreich integriert, um die Interaktionsfähigkeit weiter zu optimieren.

Die Evaluation des Systems im Logistikzentrum der Firma Juzo zeigte eine hohe Funktionalität der entwickelten Module. Die Nutzer bewerteten besonders die visuellen Hinweise des Roboters positiv, während die Sprachsteuerung aufgrund von Erkennungsproblemen als verbesserungswürdig angesehen wurde. Die Tätigkeitserkennung zeigte eine grundlegende Funktionalität, wies jedoch in spezifischen Szenarien, wie der Unterscheidung ähnlicher Tätigkeiten, Optimierungsbedarf auf. Die Benutzerfreundlichkeit wurde in Usability-Tests evaluiert, wobei klarere Navigationsstrukturen und präzisere Anleitungen als Verbesserungspotenzial identifiziert wurden.

Die Ergebnisse zeigen, dass noch Optimierungspotenzial bei den Interaktionsfähigkeiten sowie der Flexibilität des Systems besteht, der interaktive mobile Roboter dann jedoch eine wertvolle Ergänzung in der Logistik darstellen kann.

### Literatur

- [1] Berger, J.; Schumacher, S.; Mehrer, J. et al.: User-centered analysis for the implementation of industrial mobile robots. Procedia CIRP 126 (2024), pp. 224–229
- [2] Berger, J.; Lu, S.: A Multi-camera System for Human Detection and Activity Recognition. Procedia CIRP 112 (2022), pp. 191–196
- [3] Ghodsian, N.; Benfriha, K.; Olabi, A. et al.: Toward designing an integration architecture for a mobile manipulator in production systems: Industry 4.0. Procedia CIRP 109 (2022), pp. 443–448
- [4] Rothmeyer, F.; Spitzer, F.; Herrmann, S. et al.: Akustische Basissignale für mobile Logistik-Roboter. Zeitschrift für wirtschaftlichen Fabrikbetrieb 119 (2024) 10, S. 728–734
- [5] Berg, J.; Lottermoser, A.; Richter, C. et al.: Human-Robot-Interaction for mobile industrial robot teams. Procedia CIRP 79 (2019), pp. 614–619
- [6] Berger, J.; Colceriu, C.; Blank, A. et al.: Abschlussbericht: FORobotics – mobile ad-hoc kooperierende Roboterteams. Augsburg: Universität Augsburg/Fraunhofer IGCV 2021, doi.org/10.24406/igcv-n-624794
- [7] Niemann, F.; Reining, C.; Moya Rueda, F. et al.: LARa: Creating a Dataset for Human Activity Recognition in Logistics Using Semantic Attributes. Sensors (Basel, Switzerland) 20 (2020) 15, #4083, doi.org/10.3390/s20154083
- [8] Wang, L.; Huang, B.; Zhao, Z. et al.: VideoMAE V2: Scaling Video Masked Autoencoders with Dual Masking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 14549–14560
- [9] Kolensnikov, A.; Dosovitskiy, A.; Weissenborn, D. et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations, 2021, Arxiv, doi.org/10.48550/arXiv.2010.11929
- [10] Feichtenhofer, C.: X3D: Expanding Architectures for Efficient Video Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 203–213
- [11] Materzynska, J.; Berger, G.; Bax, I. et al.: The Jester Dataset: A Large-Scale Video Dataset of Human Gestures. IEEE/CVF International Conference on Computer Vision Workshop, 2019, pp. 2874–2882



Dr.-Ing. Julia Berger

julia.berger@igcv.fraunhofer.de

Foto: Fraunhofer IGCV

Shuang Lu, M.Sc.

Christian Pest, M.Sc.

Fraunhofer-Institut für Gießerei-, Composite- und Verarbeitungstechnik IGCV  
Am Technologiezentrum 10, 86159 Augsburg  
Tel. +49 821 / 90678-0  
www.igcv.fraunhofer.de

### FÖRDERHINWEIS

Das Projekt MeMoRob (DIK0358/01) wurde durch das Bayerische Staatsministerium für Wirtschaft, Landesentwicklung und Energie gefördert.

### LIZENZ



Dieser Fachaufsatz steht unter der Lizenz Creative Commons Namensnennung 4.0 International (CC BY 4.0)