

Vorhaben

InCoRAP

**Intentionsbasierte kooperative Roboterhandlungsplanung und
Werkerunterstützung in Fabrikumgebungen**

Titel

Abschlussbericht Teil I: Kurzbericht

Förderkennzeichen

01IW19002

Zuwendungsempfänger

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
Trippstadter Straße 122, D-67663 Kaiserslautern

Ausführende Stelle

DFKI GmbH – Innovative Fabrikssysteme, Kaiserslautern

Projektleiter

Prof. Dr.-Ing. Martin Ruskowski

Bewilligungszeitraum

01.01.2020 – 31.06.2024

Autoren

Parsha Pahlevannejad, Michael Sintek, Juan Carlos Saborio Morales,
Nicolas Großmann

Erstellungsdatum

30. August 2024



Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IW19002 gefördert. Dem Muster in Anlage 2 zu Nr. 4.1 NKBF 2017 folgend, handelt es sich bei diesem Dokument um Teil I, einen Kurzbericht des Vorhabens. Die Teile II und III liegen als gesonderte Dokumente vor.

Vision & Herausforderungen: InCoRAP erforscht die Mensch-Roboter-Interaktion, indem es die Intentionen des Menschen berücksichtigt. Dadurch kann der Roboter seine Handlungen an die erkannten Absichten des Werkers anpassen, was eine fließende Zusammenarbeit ermöglicht. Anders als bestehende Ansätze, die oft zu Unterbrechungen führen, basiert InCoRAPs Assistenzsystem auf einem umfassenden Modell der Umgebung. Dieses Modell nutzt Sensordaten und Prozessinformationen, um menschliche Intentionen zu erkennen und die Roboteraktionen entsprechend zu planen. Der Forschungsschwerpunkt liegt auf der Erkennung menschlicher Intentionen, der Pflege eines multisensorischen Modells und dessen Nutzung für situationsgerechte Unterstützung und Aktionsplanung.

AP1 - Projektmanagement: Das InCoRAP-Team traf sich regelmäßig in wöchentlichen Meetings (virtuell), um die Zusammenarbeit zwischen den verteilten Forschungsbereichen zu koordinieren. Pandemiebedingt konnten in den ersten zwei Projektjahren keine Vor-Ort-Integrationsarbeiten stattfinden, was eine 6-monatige Projektverlängerung erforderlich machte. Diese zusätzliche Zeit wurde genutzt, um verbleibende Integrationsarbeiten abzuschließen, Evaluationsphasen durchzuführen und wissenschaftliche Veröffentlichungen vorzubereiten.

AP2 - Anforderungsanalyse: In AP2 wurde die Anforderungsanalyse durchgeführt und dazu für die beiden Szenarien (Smart Maintenance und Manual Assembly) die entsprechenden User Stories formuliert. Weiterhin wurde ein Katalog erstellt, der Datenquellen (insbesondere diverse Sensoren), aber auch Tools/Systeme und Datenformate umfasst, die potenziell im Projekt eingesetzt werden können. Dieser Katalog zusammen mit der Bewertung der einzelnen Elemente für unser Projekt diente dann als Grundlage für den Entwurf der Systemarchitektur und der dabei vorgesehenen Schnittstellen. Kern dieser Systemarchitektur ist ein Digital Twin auf Basis des Spieleditors Unity, dessen Daten die Basis des Umgebungsmodells sind.

AP3 - Hierarchisches semantisches kontextsensitives Umgebungsmodell:

In diesem Arbeitspaket wurde ein ontologiebasierter digitaler Zwilling erarbeitet, der die Basis für das semantische Umgebungsmodell ist. Dazu wurde eine Ontologie entwickelt, die alle Elemente der Umgebung beinhaltet, also Gebäude, Räume, Produktionsanlagen, Fabrikelemente, mit Tags, Zoomleveln und räumlichen Beziehungen. Basis dieser Ontologie ist dabei die Building Topology Ontology (BOT) vom W3C.

AP4 - Intentionserkennung: Im Rahmen von InCoRAP wurde ein formales Modell und ein Prototyp zur Intentionserkennung entwickelt, das es Roboterassistenten ermöglicht, mit menschlichen Arbeitern zusammenzuarbeiten und deren Bedürfnisse vorherzusehen. Ein "aktiver IR"-Ansatz, der Beobachtungen und eigenständige Aktionen kombiniert, wurde mithilfe eines stochastischen Arbeitersimulators und eines POMDP-Entscheidungsmodells umgesetzt. In Zusammenarbeit mit der University of SIBIU wurde ein flexibler Ansatz entwickelt, der auch mit minimalen Trainingsdaten arbeitet und sich an individuelle Arbeiter anpasst. Zur Verbesserung der Intentionserkennung wurden zudem verschiedene Objekterkennungsmodelle, darunter YoloV5, getestet und evaluiert.

AP5 - Situationspezifische Werkerassistenz: Im Rahmen des Arbeitspakets wurden Studien zur Mensch-Roboter-Interaktion durchgeführt, die eine Anwendung zur Kooperation mit einem mobilen Roboter mithilfe von Augmented Reality realisieren. Hierbei wurden montagebezogene Contextual Affordances modelliert und je nach Interaktionsart als aktiv oder passiv kategorisiert, wobei sie über ein HoloLens-Display dem Werker angezeigt werden. Ein Prototyp zur Erfassung und Fusion von "Human-Centric" Sensordaten wie Eye-Tracking und Hand-Tracking für die Intentionserkennung wurde entwickelt. Der Roboter nutzt diese Daten, um in Echtzeit auf die Ziele des Arbeiters zu reagieren und durch Erkennung, Planung und Aktionen die Arbeitsabläufe zu unterstützen.

AP6 - Integration menschlicher Intentionen in die Handlungsplanung des Roboters: In diesem Arbeitspaket wurde die Roboterhandlungsplanung und -ausführung mithilfe des hierarchischen hybriden Planers CHIMP und der Roboterkontrollarchitektur ROSPlan implementiert, die auf den Roboter Mobipick integriert wurden. Da ROSPlan von Haus aus keine hierarchische Planung unterstützt, wurde eine Methode entwickelt, um CHIMP in ROSPlan zu integrieren. Zusätzlich wurde ein POMDP-basierter Planer als "Orchestrator" eingebunden, der auf Grundlage von Beobachtungen Aufgaben auswählt, die der hierarchische Planer koordiniert und am Roboter ausführt. Diese Planungsphasen interagieren mit den anderen Systemkomponenten, um entweder Informationen zu sammeln oder physische Aktionen zu steuern.

AP7 - Implementierung der Roboterfähigkeiten: Wie bei AP6 wurden die verschiedenen Funktionen, die das autonome Verhalten unseres Roboterassistenten ermöglichen, in verschiedenen Softwareschichten und Modulen implementiert, die hauptsächlich über die ROS-Oberfläche kommunizieren. Dazu gehören die Objekterkennung durch ein YOLO-Netzwerk unter Verwendung von Onboard- und abgesetzten Kameras, die Unterstützung von semantischen Daten aus dem „Digitalen Zwilling“, die Schätzung der Objektlage über Nvidia DOPE, die Planung der Absichtserkennung über Rageplan (unseren POMDP-Planer) und die Steuerung der Roboternavigation und -manipulation auf hoher Ebene über UP und die Embedded Systems Bridge.

AP8 - Realisierung der Use Cases: In AP8 wurden CAD-Modelle der Produktionsmodule und Objekte in den Digital Twin integriert. Für das "Smart Maintenance"-Szenario wurde eine Lösung zur Simulation von Anomalien entwickelt, bei der ein Heizlüfter den Energieverbrauch simuliert und eine Drittanbietersoftware ein Signal zur Fehlerbehebung sendet. Im "Manual Assembly"-Szenario wurden Sichtlagerkästen für die Lagerung von 3D-gedruckten Komponenten des Demo-Produkts "Noppenstein-LKW" genutzt. Zwei Varianten des Endprodukts wurden definiert, die sich in Farbe und Zusammensetzungsmethode unterscheiden. Die Genauigkeit der Sensordaten für das Werkerassistenzsystem (Eye-Tracking, Objekterkennung, Handtracking) wurde mit mehreren Mitarbeitern getestet. Die Daten wurden auf einem Edge-Device erfasst und für die Intentionserkennung und Systemevaluation gespeichert.

AP9 – Evaluation: In AP9 wurde die Evaluation des InCoRAP-Projekts durchgeführt. Diese umfasste eine schriftliche Befragung der Teilnehmenden sowie die Erhebung und Auswertung von Sensordaten. In der ersten Phase wurde ein Stresserkennungsmodell mit dem Open-Source-Datensatz WESAD entwickelt, wobei Empatica E4-Sensoren zur Datenerfassung eingesetzt wurden. Die Teilnehmenden wurden verschiedenen Aufgaben ausgesetzt, um neutrale und stressige Zustände zu erzeugen. Ein Deep-Learning-Modell auf Basis von CNN-LSTM wurde zur binären Klassifizierung (Stress vs. Nicht-Stress) trainiert. In der zweiten Phase fand die endgültige Evaluation in Osnabrück im "Manual Assembly"-Szenario statt. Die Ergebnisse zeigten, dass die Roboterassistenz als hilfreich empfunden wurde und die empfundene Arbeitsbelastung sowie die KI-ermittelten Stresswerte signifikant verringert wurden. Die Roboterunterstützung erwies sich somit als effektiv zur Reduktion von Arbeitsbelastung und Stress.

Publikationen & Kooperationen: Insgesamt wurden durch das InCoRAP-Team 11 Publikationen und 4 Studentische Arbeiten veröffentlicht. Neben anderen Forschungsbereichen und Demo-Lab (SmartFactory-KL) kooperierte das Team insb. mit RPTU (Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau), Universität Osnabrück und University of SIBIU.

Vorhaben	InCoRAP Intentionsbasierte kooperative Roboterhandlungsplanung und Werkerunterstützung in Fabrikumgebungen
Titel	Abschlussbericht Teil II: Eingehende Darstellung
Förderkennzeichen	01IW19002
Zuwendungsempfänger	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH Trippstadter Straße 122, D-67663 Kaiserslautern
Ausführende Stelle	DFKI GmbH – Innovative Fabriksysteme, Kaiserslautern
Projektleiter	Prof. Dr.-Ing. Martin Ruskowski
Bewilligungszeitraum	01.01.2020 – 31.06.2024
Autoren	Parsha Pahlevannejad, Michael Sintek, Juan Carlos Saborio Morales, Nicolas Großmann
Erstellungsdatum	30. August 2024

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IW19002 gefördert. Dem Muster in Anlage 2 zu Nr. 4.1 NKBF 2017 folgend, handelt es sich bei diesem Dokument um Teil II, eine eingehende Darstellung des Vorhabens. Die Teile I und III liegen als gesonderte Dokumente vor.

Inhaltsverzeichnis

1	Verwendung der Zuwendung	4
1.1	AP1: Projektmanagement	5
1.2	AP2: Anforderungsanalyse	5
1.3	AP3: Common Hierarchical Semantic Context-Sensitive Environmental Model	6
1.3.1	Environment model from static data sources (AP 3.1)	6
1.3.2	Update from dynamic sources (AP 3.2)	8
1.3.3	Semantic Zoom (AP3.3)	8
1.4	AP4: Intentionserkennung	9
1.4.1	Erstellung von Aufgabenmodellen und Integration mit dem Umgebungsmodell (AP 4.1)	9
1.4.2	Mapping des Blickverhaltens auf Aufgaben und Umgebungsobjekte (AP 4.2)	10
1.4.3	Intentionsschätzung auf verschiedenen Abstraktionsebenen (AP 4.3)	11
1.4.4	Fusion der Schätzungen (AP 4.4)	18
1.5	AP5: Situationsspezifische Werkerassistenz	19
1.5.1	Framework für die Zusammenarbeit zwischen Werkern und Robotern (AP 5.1)	19
1.5.2	AR-Visualisierung des Prozess- und Umgebungszustands, einschließlich der Visualisierung der Assistenz durch den Roboter (AP 5.2)	20
1.5.3	Dynamische Affordances und Handlungsempfehlungen (AP 5.3)	20
1.5.4	Innovative Eye-Tracking-Interaktion mit AR (AP 5.4)	21
1.6	Integration menschlicher Intentionen in die Handlungsplanung des Roboters (AP6)	21
1.6.1	Handlungsplanung für ein kollaboratives Team (AP 6.1)	21
1.6.2	Planausführung und Ausführungsüberwachung (AP 6.2)	22
1.6.3	Anpassung der Pläne als Reaktion auf geänderte Intentionen (AP 6.3)	23
1.7	Implementierung der Roboterfähigkeiten (AP7)	24
1.7.1	Lokalisierung und Navigation in der Umgebung (AP 7.1)	24
1.7.2	Wahrnehmung der Umwelt (AP 7.2)	25
1.7.3	Ausführung von Manipulationsaufgaben (AP 7.3)	26
1.8	Realisierung der Use Cases (AP8)	26
1.8.1	Smart Maintenance Szenario (AP 8.1)	26
1.8.2	Manual Assembly Szenario (AP 8.2)	27
1.9	Evaluation (AP9)	28
2	Wichtigste Positionen des zahlenmäßigen Nachweises	30

3	Notwendigkeit und Angemessenheit der geleisteten Arbeit	30
4	Verwertbarkeit	31
4.1	Wirtschaftliche Erfolgsaussichten.....	31
4.2	Wissenschaftliche Erfolgsaussichten	32
4.3	Wissenschaftlich-wirtschaftliche Anschlussfähigkeit.....	32
5	Bekannt gewordener Fortschritt	33
6	Veröffentlichungen	34
7	Abbildungsverzeichnis	35
8	Tabellenverzeichnis.....	36
9	Abkürzungsverzeichnis.....	37
10	Literaturverzeichnis	38

1 Verwendung der Zuwendung

In Übereinstimmung mit dem Projektarbeitsplan wurden alle Forschungsschwerpunkte erfolgreich abgeschlossen. Dies umfasste den Aufbau eines hierarchischen semantischen kontextsensitiven Umgebungsmodell, Intentionserkennung für Situationspezifisches Werkerassistenzsystem, Integration menschlicher Intentionen in die Handlungsplanung des Roboters, Implementierung der Roboterfähigkeiten und Realisierung zwei verschiedener Use-Cases. Außerdem wurde das gesamte System in zwei Phasen/Stufen evaluiert. Die dargestellten Arbeiten umfassen alle vom DFKI geleisteten Arbeiten im Vorhaben InCoRAP. Abbildung 1 gibt den Überblick über die Aufteilung der Arbeitspakete.

	2020				2021				2022				2023				2024		Effort/PM				
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	IFS	SDS	PBR	QL	OA
WP1 Project management																			20	4	4	4	32
Project organization																			12	2	2	2	18
Reporting																			8	2	2	2	14
WP2 Requirement analysis																			5	7	3	2	17
User stories																			2	3	1	2	8
Data source catalog																			2	2	1	0	5
Perspectives overview and interrelation documentation																			1	2	1	0	4
WP3 Common hierarchical semantic context-sensitive environment model																			3	21	2	0	26
Environment model from static data sources																			0	8	0	0	8
Update from dynamic sources																			3	5	2	0	10
Semantic zoom																			0	8	0	0	8
WP4 Intention recognition																			26	5	2	0	33
Creation of task models and integration with environment																			8	4	2	0	14
Mapping gaze behavior to tasks and environment objects																			6	1	0	0	7
Intention estimation on different levels of abstraction																			7	0	0	0	7
Fusing estimations																			5	0	0	0	5
WP5 Situation-specific worker assistance																			24	5	2	0	31
Framework for worker-robot collaboration																			6	4	2	0	12
AR visualization of process and environment state																			8	1	0	0	9
Dynamic affordances and action recommendation																			4	0	0	0	4
Innovative eye-tracking interaction with AR																			6	0	0	0	6
WP6 Integration of worker intentions in robot task																			2	0	16	0	18
Planning and reasoning for a collaborative team																			2	0	6	0	8
Plan execution and monitoring																			0	0	6	0	6
Reacting to changed intentions by adapting plans																			0	0	4	0	4
WP7 Implementation of robot capabilities																			0	0	11	0	11
Localization and navigation in the environment model																			0	0	3	0	3
Perception of the environment																			0	0	3	0	3
Execution of manipulation tasks																			0	0	5	0	5
WP8 Realization of use cases																			6	2	4	0	12
Smart maintenance scenario																			3	1	2	0	6
Manual assembly scenario																			3	1	2	0	6
WP9 Evaluation																			2	2	2	16	22
Evaluation of the quality of human-robot collaboration																			1	1	1	10	13
Evaluation wrt. the overall project goals																			1	1	1	6	9
WP10 Communication/knowledge transfer/exploitation																			8	2	2	2	14
Overall																			96	48	48	24	216

Abbildung 1: Gantt-Diagramm der geplanten Arbeit im Projekt

1.1 AP1: Projektmanagement

Das Ziel von WP1 bestand darin, einen reibungslosen Projektablauf sicherzustellen. Dies umfasste die Koordination der Arbeiten der einzelnen Konsortialpartner und ihrer Zusammenarbeit, die Überwachung des Projektfortschritts sowie die regelmäßige Aktualisierung des Projektplans. Zudem gehörten die Berichterstattung an das SAB, die Einholung von deren Feedback sowie die Erstellung halbjährlicher Statusberichte und des Abschlussberichts dazu. Darüber hinaus wurden Workshops und Code Camps organisiert und durchgeführt.

1.2 AP2: Anforderungsanalyse

Im ersten Projektjahr wurde die Anforderungsanalyse durchgeführt und die entsprechenden User Stories für die beiden Szenarien (Smart Maintenance und manuelle Montage) formuliert. Darüber hinaus wurde ein Katalog erstellt, der Datenquellen, aber auch Tools/Systeme und Datenformate enthält, die im Projekt potenziell genutzt werden können. Dieser Katalog diente dann zusammen mit der Bewertung der einzelnen Elemente für unser Projekt als Grundlage für die Gestaltung der Systemarchitektur und der dabei vorgesehenen Schnittstellen, siehe

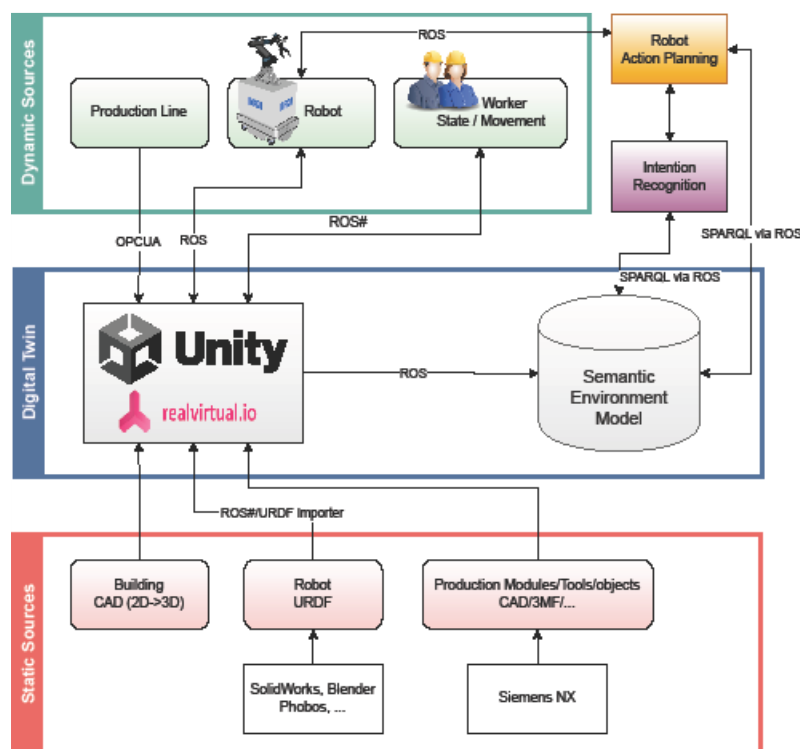


Abbildung 2 - InCoRAP-Systemarchitektur und die Interaktion des Umgebungsmodells mit anderen Systemkomponenten

In diesem Diagramm ist der Kern der Systemarchitektur ein Digitaler Zwilling, der aus einem 3D-Modell der Produktionsumgebung (aufgebaut in der Spiel-Engine Unity3D) und einem Semantischen Umgebungsmodell besteht. In der unteren Schicht sind statische Wissensquellen abgebildet: Gebäude, statische Informationen zu den Robotern und Produktionsmodule, die mit verschiedenen CAD-Tools

erstellt und dann in Unity3D eingespeist werden. Unter Verwendung dieser statischen Wissensquellen wurde eine 3D-Darstellung der Produktionsumgebung (SmartFactory-KL¹-Labor) in Unity3D erzeugt.

In der oberen Schicht werden die dynamischen Quellen der Produktionslinie, des Roboters und des Arbeiters dargestellt. Darüber hinaus greifen die Komponenten der Roboteraktionsplanung und der Absichtserkennung über SPARQL-Abfragen und das ROS-Protokoll auf das semantische Umgebungsmodell zu.

Scenario 1 – Smart Maintenance Scenario: In diesem Szenario erhält der Arbeiter einen Wartungsauftrag von einem ERP-System, bei dem es sich entweder um eine Routinewartung oder um eine Fehlersuche aufgrund eines Systemausfalls handeln kann. Um den aufgetretenen Fehler zu beheben, muss ein defektes Teil innerhalb der Produktionslinie ausgetauscht werden, was für den Werker aufgrund verschiedener möglicher Einschränkungen (z. B. arbeitet der Werker gerade in einer Produktionslinie) mühsam sein kann. Dies könnte für den Arbeiter aufgrund verschiedener möglicher Einschränkungen mühsam sein (z. B. arbeitet der Arbeiter gerade in einer Produktionszelle und hat keinen Zugang zu den Ersatzteilen). Mithilfe von Augmented-Reality-Brillen wird der Arbeiter dabei unterstützt, die Produktionsmaschine zu identifizieren, an der der Fehler aufgetreten ist, und herauszufinden, wie er auf die defekte Hardware zugreifen bzw. sie ersetzen oder reparieren kann. Mit Hilfe der Absichtserkennung wird der nächste erforderliche/mögliche Wartungsschritt auf der AR-Brille visualisiert. Abhängig von der aktuellen Situation und dem Wartungsschritt, den der Arbeiter durchführt, wird das benötigte Ersatzteil von einem AGV gegriffen und transportiert. Daher ist eine robuste Mensch-Roboter-Kollaboration notwendig.

Scenario 2 – Manual Assembly Scenario: Das Hauptziel in diesem Szenario ist es, den Benutzer (Arbeiter) bei der manuellen Montage eines Produkts zu unterstützen. Die aufeinanderfolgenden Schritte, um das Produkt zusammenzubauen, werden auf der AR-Brille in Abhängigkeit von den Ergebnissen des Moduls zur Absichtserkennung visualisiert. Die Absichten des Arbeiters werden mit Hilfe von Objekterkennung und Eye-Tracking erkannt. Anhand der Kombination von erkannten Objekten kann das Assistenzsystem die ausgeführten Schritte und die nächsten möglichen Aktionen bestimmen. Ein AGV nutzt die Ergebnisse der Absichtserkennung, um die notwendigen Montageteile für den nächsten Schritt zu ermitteln und zum Werker zu bringen.

1.3 AP3: Common Hierarchical Semantic Context-Sensitive Environmental Model

1.3.1 Environment model from static data sources (AP 3.1)

Im Rahmen des semantischen Umgebungsmodells wurden zunächst die statischen Daten (Gebäude, Räume, Produktionsanlagen) betrachtet. Als Basis-Ontologie für die semantische Darstellung des Umgebungsmodells haben wir die Building Topology Ontology (BOT²) [RLSP20] identifiziert. Diese Ontologie wurde vom W3C als Ergebnis der Linked Building Data Community Group veröffentlicht und eignet sich daher besonders für die Darstellung von (Fabrik)Gebäuden und den darin befindlichen

¹ www.smartfactory.de

² <https://w3c-lbd-cg.github.io/bot/>

Elementen sowie deren räumlichen Beziehungen. Diese BOT-Ontologie wurde dann erweitert, um Funktionen zu unterstützen, die in dem Projekt für die obengenannten Szenarien benötigt werden: Fabrikobjekte mit Tags und globalen Zoomstufen, Begrenzungsrahmen und räumliche Beziehungen (enthält, disjunkt, schneidet, innerhalb). Für letztere wurde insbesondere die Java-Bibliothek `spatial4j`¹ verwendet.

Auf der Grundlage der Baupläne des SmartFactory-KL-Labors wurde ein 3D-Oberflächenmodell erstellt, das in Abbildung 3 dargestellt ist. Anhand der Schichtinformationen wurden verschiedene Bauteilgruppen wie Wände, Türen und Säulen aus den detaillierten Konstruktionszeichnungen ausgewählt. Weiterhin wurden 3D-Modelle der Demonstrationsmodule des SmartFactory-KL-Labors in Unity3D importiert. Die 3D-Modelle der betroffenen Fabrikobjekte wurden mit Hilfe der CAD-Software Siemens NX² mit Tags und Layern annotiert. Die hierarchischen Beziehungen zwischen den Objekten wurden hauptsächlich vor dem Import modelliert. Daraus konnten durch Extrusion separate 3D-Hüllen erstellt werden, die für die Simulation und für die Umgebungsdarstellung verwendet werden. Dies wurde unter zwei Masterarbeiten an der RPTU (Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau) veröffentlicht [BL23].

Weiterhin haben wir ein System zur Übertragung der Daten des Digitalen Zwillings in einen Triple-Store (als technische Implementierung des Umgebungsmodells). Für die Initiierung und Aktualisierung der Daten im Triple Store unter Berücksichtigung der dynamischen Veränderungen in der Umgebung (z.B. Roboterbewegungen) haben wir ein System entwickelt, das ausschließlich auf Event-Streaming unter Verwendung von ROS basiert.

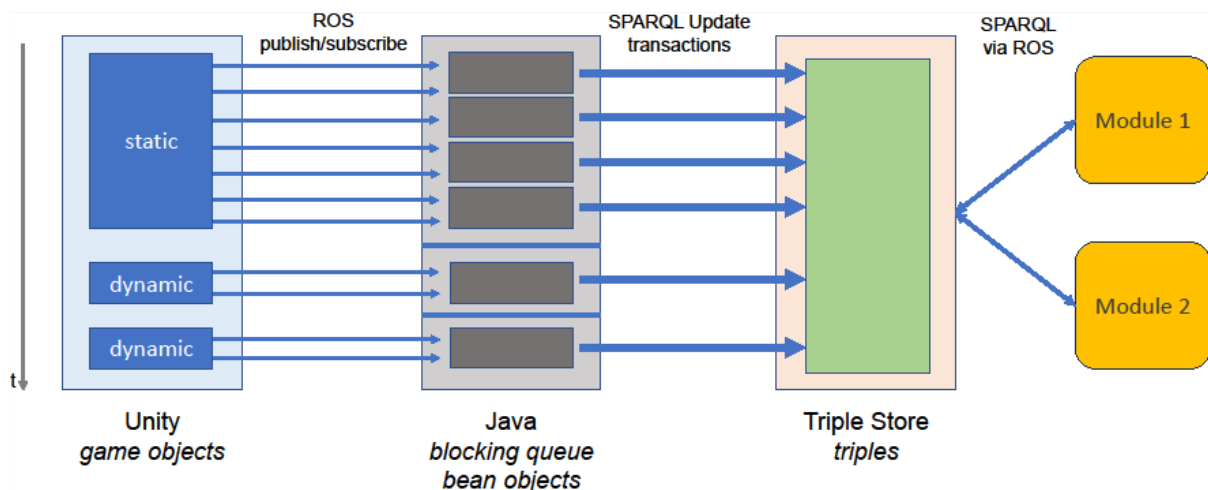


Abbildung 3 - Füllen des Umgebungsmodells aus statischen und dynamischen Quellen über ROS

¹ <https://github.com/locationtech/spatial4j>

² <https://www.plm.automation.siemens.com/global/en/products/nx/>

1.3.2 Update from dynamic sources (AP 3.2)

Wie in Abbildung 3 dargestellt, werden die Spielobjekte aus Unity über ROS publish/subscribe an ein Java-basiertes System gesendet, und zwar sowohl die statischen als auch die dynamischen Quellen. Durch die Verwendung einer Java-Blockierwarteschlange identifiziert das System zusammenhängende Blöcke von Spielobjekten (entweder alle statischen Objekte oder dynamische Objekte die einer Bewegung des Roboters/Arms oder Arbeiters entsprechen). Diese Spielobjekte werden zunächst in Bean-Objekte kopiert, so dass aus Effizienzgründen im nächsten Schritt nur die endgültige Version dieser Objekte über SPARQL-Update-Transaktionen in den Triple Store kopiert werden kann. Hier verwenden wir die Bibliothek RDFBeans¹, um die Erstellung der Inhalte des Triple Stores zu erleichtern.

Außerdem wurde auch ein logischer Formalismus entwickelt, der zur Darstellung von Prozessmodellen verwendet werden kann. Der Formalismus basiert auf den Ideen der HTNs (hierarchische Aufgabennetze), wurde aber einerseits vereinfacht (Bedingungen verwenden nur boolesche Logik) und andererseits erweitert (Darstellung von „Affordance“). Dieser Formalismus wurde dann für Prozessmodelle für unsere (dynamischen) Anwendungsfälle verwendet.

1.3.3 Semantic Zoom (AP3.3)

Die Roboterplanung benötigt je nach Aufgabenstellung unterschiedlich detaillierte Sichten auf die Objekthierarchie im Umgebungsmodell. Diese Ansichten werden durch semantischen Zoom berechnet, für den Konzepte entwickelt wurden. Auf der Seite des Digitalen Zwillinges wird jedes Objekt sowohl mit Tags als auch mit globalen Zoomstufen annotiert, die dann semantisch im Umgebungsmodell dargestellt werden. Dies wird in [BL23] ausführlich beschrieben. Zur Laufzeit wird auf diese semantischen Repräsentationen dann über SPARQL-Abfragen zugegriffen, die entweder auf globale Zoomstufen zugreifen können oder spezifische Situationen widerspiegeln. So wird z. B. der Handarbeitsplatz (aus Szenario 2) als Hindernis betrachtet, wenn er von einem durch die Fabrik navigierenden Roboter ausgesehen wird, aber wenn der Roboter einen Gegenstand am Arbeitsplatz ausliefert, müssen sowohl der Roboter als auch der Arbeitsplatz in einer feineren Zoomstufe betrachtet werden, siehe Abbildung 4. Für diese situationsspezifischen Zoomstufen können oft keine globalen Zoomstufen verwendet werden, sondern der Zoom muss dynamisch berechnet werden, wobei Objektgrößen, Tags, Objektteilhierarchien und räumliche Beziehungen zu berücksichtigen sind.

Diese dynamischen Berechnungen werden über SPARQL 1.1 durchgeführt. Leichtgewichtige Inferenzen wie Abschlüsse für SKOS-Hierarchien werden über SPARQL 1.1 Property Path-Ausdrücke gehandhabt, siehe [3].

Außerdem haben wir das SPARQL Update Protocol (das HTTP-basiert ist) als ROS-Dienst verpackt, so dass andere Module für die gesamte Kommunikation nur ROS verwenden, auch mit dem Triple Store. Um andere Module nicht zu zwingen, Kenntnisse über konkrete SPARQL-Abfragen zu haben, können vordefinierte parametrisierte SPARQL-Abfragen über den SPARQL-ROS-Dienst ausgeführt werden.

¹ <https://rdfbeans.github.io/>

Wir haben auch eine OWL-Ontologie zur Darstellung von kontextuellen Affordanzen entwickelt, die zusammen mit Zoomstufen verwendet werden. Dank der Verwendung von OWL und seiner hohen Ausdruckskraft konnten verschiedene komplexe Abfragen mit OWL-Inferenzen („DL Query“ in Protégé) realisiert werden.

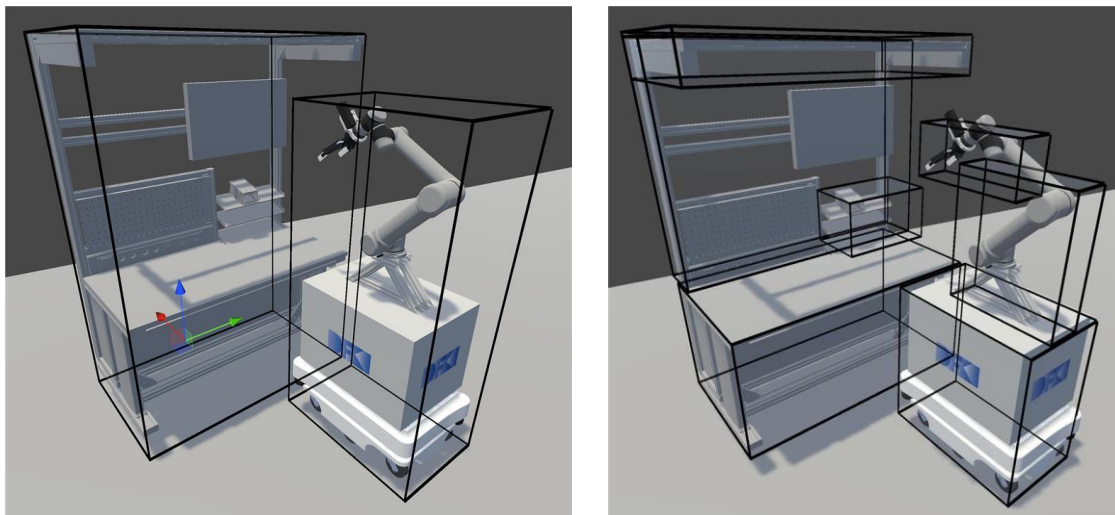


Abbildung 4 - Situation Specific Zoom Levels (Beispiele)

1.4 AP4: Intentionserkennung

Dieses Arbeitspaket konzentrierte sich auf das Verständnis und die korrekte Modellierung des Verhaltens eines automatischen Assistenten unter Verwendung von „Absichtserkennung“, die auch die Richtung verwandter Arbeitspakete erhellt. Wir definierten Absichtserkennung als „die Fähigkeit eines Agenten, die Handlungen eines anderen Agenten zu antizipieren“ und daher informierte Entscheidungen zu treffen, die ihre Ziele beeinflussen (siehe [FZ19]). Für die Zwecke dieses Projekts haben wir uns auf die kollaborative Dimension von IR (im Gegensatz zu z.B. Wettbewerb).

Wir haben eine beträchtliche Anzahl von Arbeiten untersucht, darunter passive IR-Ansätze (der Roboter oder Agent beobachtet nur, greift nicht ein) [RG10, RG11, Gei15, SGB+14], die recht verbreitet sind, sowie aktive Ansätze bei denen das Nachdenken über Absichten zu aktivem Verhalten führt (siehe z.B. [AB19]). Unser Assistenzroboter ist per Definition ein aktiver Teilnehmer. Wir bezeichnen das allgemeine Problem der Absichtserkennung als IR und den gerichteten, aktiven Zielerkennungsansatz als AGR.

1.4.1 Erstellung von Aufgabenmodellen und Integration mit dem Umgebungsmodell (AP 4.1)

Ursprünglich haben wir eine unterstützende Architektur für IR vorgeschlagen, bei der Eye-Tracking-Daten und Bilder von einem Hololens 2 AR-Gerät erfasst und dann von einem Objekterkennungsmodul verarbeitet werden. Anschließend wurde ein erster Entwurf für IR erstellt, der aus Aufgabenmodellen und probabilistischen grafischen Modellen besteht, die den hierarchischen Aufgabenplaner auf dem Roboter informieren können. Die Idee war, gültige Übergänge durch Anforderungsprüfung zu bewerten und diese Informationen zu nutzen, um mögliche Aktionen des Arbeiters zu antizipieren.

In Zusammenarbeit mit der „Lucian Blaga“-Universität in Sibiu haben wir ein Modell zur Vorhersage von Montageschritten im Szenario der manuellen Montage entwickelt [PWG+23, PMW+23], das die Ergebnisse der Objekterkennung aus [TWP+21] nutzt. Das resultierende Markov-Modell ist eine Erweiterung früherer Forschungsarbeiten ([GPPZ20] und [PGM+22]), die an der Universität von Sibiu durchgeführt wurden. Der Montageprozess wurde zunächst in Form einer Markdown-Beschreibung dokumentiert und später automatisch in eine RDF- und JSON-basierte Ontologie/Wissensbasis übersetzt. Ein graphenbasierter Ansatz wurde verwendet, um die verschiedenen Zustände der Baugruppe zu verstehen und wie sie miteinander in Beziehung stehen. Das Baugruppeninformationsmodell dient als Basiswahrheit und ermöglicht es dem Schrittvorhersagemodell, den gesamten Zustandsraum zu erkunden, bevor ein Training durchgeführt wird [PWG+23]. Dieses Modell berücksichtigte jedoch nicht die verrauschte, unsichere und stochastische Natur sowohl des Arbeiters als auch der Umgebung. Die Aufgabenmodelle der Arbeiter wurden dann erweitert, um die Dynamik der Arbeiter bei der Lösung ihrer Aufgaben darzustellen, und so formalisiert, dass sie auch als Arbeitersimulator in einem größeren, stochastischen Arbeiter-Roboter-Modell funktionieren. Das resultierende AGR-Modell, das auf einem partiell beobachtbaren Markov-Entscheidungsprozess (POMDP) basiert, ermöglicht es dem Roboter, Aktionen auf der Grundlage einer Hypothese über den aktuellen Plan des Arbeiters auszuwählen, die Informationen über mögliche Ziele des Arbeiters und nicht nur über die unmittelbar nächsten Aktionen enthält. Das AGR-POMDP wertet mehrere mögliche Verläufe (Folgen von Aktionen und Beobachtungen) aus und bestimmt die beste Aktion, die in jedem Zeitschritt auszuführen ist. Durch die Online-Lösung des Modells (Verschachtelung von Handeln und Planen) kann der Roboter auf Veränderungen in der Umgebung, neue Bedürfnisse und unerwartete Aktionen der Arbeiter reagieren. Weitere Einzelheiten sind in der zugehörigen Veröffentlichung zu finden. Zusammenfassend lässt sich jedoch sagen, dass wir das aktive Zielerkennungsproblem in ein Belohnungsmaximierungsproblem umgewandelt haben, bei dem nützliche (oder belohnende) Aktionen diejenigen sind, die dem Arbeiter helfen, seine Ziele zu erreichen. Zum Beispiel kann der Arbeiter nur dann erfolgreich Objekte zusammenbauen und Maschinen inspizieren, wenn die notwendigen Werkzeuge und Teile zur Verfügung stehen, und diese Bedingungen kann der Roboter im Voraus erfüllen. Das POMDP-AGR-Modell wird in AP4.3 ausführlicher erläutert.

In der letzten Projektphase haben wir unsere Modelle validiert und getestet, um die Aufgaben der Arbeiter genauer abzubilden und die Fähigkeit des Roboters zu verbessern, mit geeigneten Aktionen zu reagieren. Diese wurden in verschiedenen Experimenten statistisch und empirisch validiert und stützen die Behauptung, dass unser System einen innovativen Ansatz zur Unterstützung und Reaktion auf die Aufgaben des Arbeiters. Diese Aufgabenmodelle können durch andere Arbeitermodelle ersetzt werden, um entweder andere Arten von Arbeitern oder ganz andere Aufgaben darzustellen. Die von uns vorgeschlagene Methodik kann jedoch helfen, neue Probleme zu identifizieren und zu modellieren, die ebenfalls einen Roboterassistenten erfordern.

1.4.2 Mapping des Blickverhaltens auf Aufgaben und Umgebungsobjekte (AP 4.2)

Augenbewegungen sind für die Verarbeitung visueller Informationen unerlässlich und können anhand ihrer Eigenschaften und Funktionen in verschiedene Typen eingeteilt werden. Das Hauptziel dieses Unterpakets besteht darin, die Augenbewegungen des Arbeiters zu verfolgen und die Fixations mit den Aufgaben, die der Arbeiter ausführt, und den Objekten in der Umgebung zu verknüpfen. Zu diesem Zweck haben wir uns auf Fixations und Microsaccades konzentriert, um den Interessenbereich (ROI) des Arbeiters zu ermitteln. Fixationen sind die häufigste Augenbewegung und können verwendet werden,

um Rückschlüsse auf kognitive Prozesse und Aufmerksamkeit zu ziehen, während Microsaccades sind sehr kleine, unwillkürliche Sakkaden, die während der Fixierung auftreten. Auch wenn unsere Augen während einer Fixation scheinbar unbeweglich sind, gibt es in Wirklichkeit immer wieder kleine Augenbewegungen. Diese führen jedoch nicht zu einer Verlagerung des Blicks von dem, was gerade fokussiert wird, und werden daher als dieselbe Fixierung betrachtet. Da die HoloLens 2 mit einem Eye-Tracking-Modul ausgestattet ist, haben wir die HoloLens 2 verwendet, um Augenbewegungsdaten zu sammeln. Um die Eye-Tracking-Daten mit Aufgaben und Umgebungsobjekten zu verknüpfen, wurde der Blickpunkt des Arbeiters (POI) mit den Umgebungsbeobachtungen (d. h. den durch den HoloLens-Kamerastream erkannten Objekten) kombiniert. Wenn der Blickpunkt für eine bestimmte Zeit mit dem erkannten Objekt übereinstimmt, deutet dies darauf hin, dass der Arbeiter mit dem entsprechenden Objekt interagieren wird. Wenn die Beobachtung mit dem nächsten möglichen Schritt im Aufgabenmodell übereinstimmt, können wir davon ausgehen, dass der Arbeiter diesen Schritt ausführen wird. Zum Beispiel kann eine Übereinstimmung zwischen einem erkannten Objekt und dem Blick des Arbeiters darauf hindeuten, dass der Arbeiter es zusammenbauen wird (unter Verwendung der Informationen aus dem Aufgabenmodell). Alle Informationen und Sensordaten werden über ROS verarbeitet, was es ermöglicht, dass alle anderen Softwarekomponenten (wie das Umgebungsmodell) über den Zustand des Arbeiters und die Beobachtungen informiert werden.

1.4.3 Intentionsschätzung auf verschiedenen Abstraktionsebenen (AP 4.3)

Der Schwerpunkt dieses Arbeitspakets liegt auf der Nutzung der Umgebungsbeobachtung (erreicht durch Analyse der Objekterkennung, Eyetracking oder anderer Arten egozentrischer Sensordaten) zur Absichtserkennung. In InCoRAP beziehen sich Intentionen auf mögliche (Teil-)Zielbedingungen und die damit verbundenen Aktionen. Der Ausgangspunkt für die Intentionserkennung (IR) ist die Wahrnehmung, die das Erkennen von Objekten und der Umgebung umfasst. Um jedoch einem Agenten zu ermöglichen, über die Ziele und Bedürfnisse eines anderen nachzudenken, sind weitere Abstraktionen notwendig. Über die Objekterkennung hinaus benötigen wir detaillierte Informationen über die aktuellen Aktivitäten oder Handlungen des Arbeiters. Diese können durch die Kombination verschiedener Datenquellen gewonnen werden, wie z.B. die Position und Lage von Objekten, die Gesten des Arbeiters, Handhaltungen und möglicherweise auch die Blickrichtung. Die Aktivitätserkennung, also das Wissen über einzelne Handlungen eines Agenten, bildet die Grundlage für die Intentionserkennung.

Zu diesem Zweck haben wir einen Prototyp eingesetzt, um menschliche egocentrische Sensordaten zu sammeln. In diesem Setup wurde die HoloLens-Kamera für die Objekterkennung genutzt, kombiniert mit Eye-Tracking und Hand-Tracking, um die Interaktionen des Arbeiters mit der Umgebung zu bestimmen.

Abbildung 5 veranschaulicht unseren Ansatz zur Objekterkennung für winzige Montageteile. Die Arbeit wurde unter dem [TWP+21] publiziert. In dieser Arbeit wurden synthetische Trainingsdaten basierend auf [GGSGM+20], [ERAB18] generiert und das Modell wurde mithilfe von CAD-Modellen verschiedener Objekte mit realen Hintergründen der Fabrikumgebung trainiert. Diese Methode verbessert den mAP der Erkennung kleiner Objekte und erreicht 9,4 FPS, was nahezu Echtzeit entspricht.

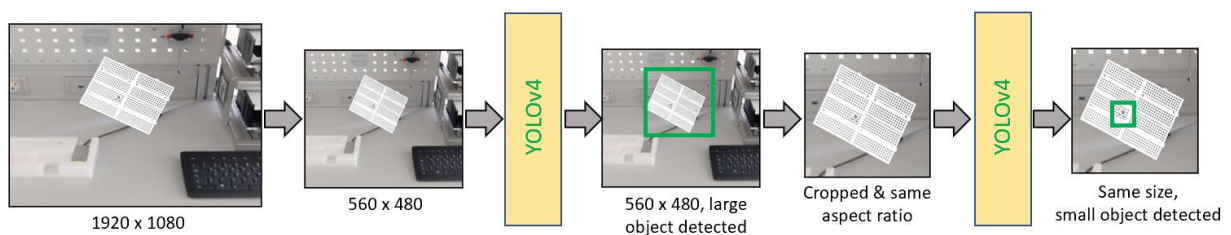


Abbildung 5 - Two step hierarchical detection using YOLOv4.

Zusätzlich haben wir uns auch intensiv mit dem Bereich des Continual Learning befasst und untersucht, wie sowohl synthetische als auch reale Datensätze zur Verbesserung der Erkennungsfähigkeiten innerhalb derselben Objektklassen beitragen. Zusätzlich haben wir den Einfluss zunehmender intra-klassen Variationen auf die Präzision des Modells für bereits gelernte Klassen analysiert. Continual Object Detection bietet Skalierbarkeit, schnelle Anpassungsfähigkeiten und Effektivität im Bereich der Objekterkennung. Dieser Ansatz zielt darauf ab, ein vielseitiges OD-Modell zu entwickeln, das in der Lage ist, eine wachsende Anzahl von Objektklassen zu berücksichtigen und gleichzeitig die Erkennungsleistung für bereits bestehende Klassen zu verbessern. Darüber hinaus reduzieren die geringeren Speicher- und Energieanforderungen für das Training neuer OD-Modelle den Ressourcenverbrauch, was diese Methode nachhaltiger macht. Durch diese Methodik verbessern Objekterkennungssysteme ihre Präzision und Effizienz, insbesondere angesichts der zunehmenden Komplexität und Vielfalt der Objektklassen. Das Konzept des Continual Learning unter Verwendung von synthetisch generierten Daten für den OD-Ansatz wird in Abbildung 6 veranschaulicht. Diese Studie ist noch im Gange.

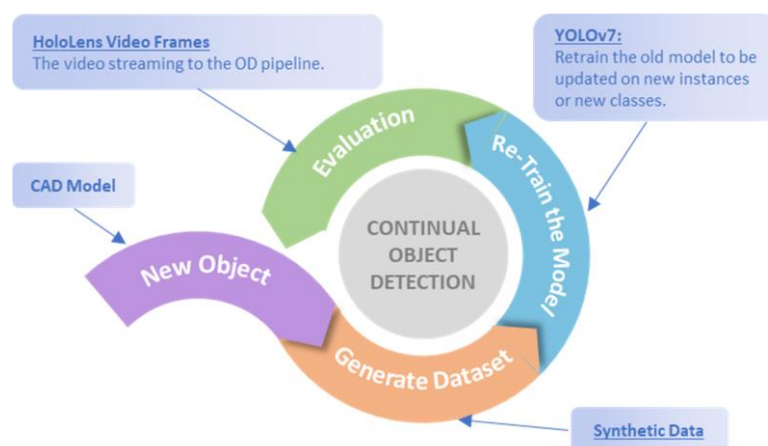


Abbildung 6 - Die Pipeline demonstriert den sequentiellen Prozess der Generierung synthetischer Daten, der Durchführung von Training und Tests mithilfe der HoloLens2 World Camera mit dem Ziel, eine kontinuierliche Lernpipeline für Modellaktualisierungen einzurichten.

Technisch gesehen wechselte unsere Deep-Learning-Architektur von YOLOv4 zu den neueren Versionen YOLOv5 und YOLOv7. Dieser Wechsel wurde durch mehrere Schlüsselfaktoren motiviert. Erstens sind YOLOv5 und YOLOv7 bequem innerhalb des PyTorch-Frameworks verfügbar, was ihre Eignung für Manipulation und Forschungszwecke erhöht. Zweitens bieten diese neueren Versionen von YOLO eine bessere Zeit- und Gesamteffizienz.

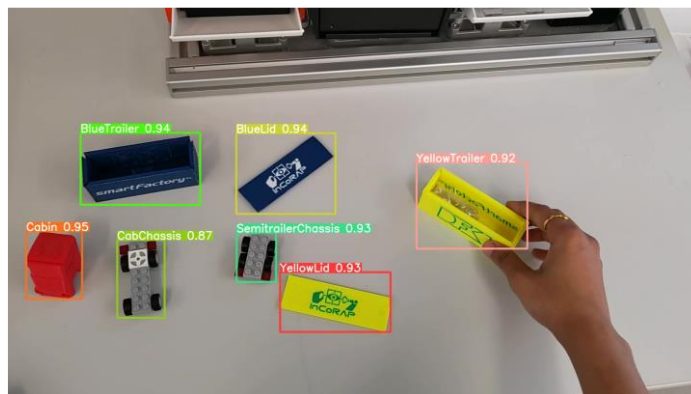


Abbildung 7 - Objekterkennung trainiert und getestet durch YoloV5 für Bauteile des "Manual Assembly" Szenarios

Datensatzfusion für die Objekterkennung:

Die Forschung im Bereich der Objekterkennung (OD) untersucht prominent den Einfluss der Kombination von realen und synthetischen Daten auf die Präzision von OD-Modellen. Ein weiterer verwandter Untersuchungsbereich betrifft das kontinuierliche Lernen, wobei der Fokus darauf liegt, OD-Modelle durch die schrittweise Bereicherung klassenspezifischer Proben zu verfeinern. Als Beitrag zu diesem Bereich strebt unsere Studie an, das Potenzial zur Verbesserung der Präzision von OD-Modellen durch die Integration von realen und synthetischen Datensätzen zu bewerten.

Wie bereits erörtert, bietet synthetisch erzeugte Daten aus CAD-Modellen klare Vorteile, wie etwa die automatische Kennzeichnung und Generierung von Proben, was die Notwendigkeit manueller Annotation umgeht. Dennoch stellt die mögliche Diskrepanz zwischen synthetischen Daten und realen Szenarien ein bedeutendes Problem dar, das die Wirksamkeit des Modells beeinträchtigen könnte. In Anbetracht dieses Problems konzentriert sich unsere Untersuchung darauf, diese Lücke zu überbrücken.

Unser experimentelles Framework umfasst vier unterschiedliche Methoden:

- Training auf vollständig realem Datensatz: Die erste Methode beinhaltet das Training des OD-Modells ausschließlich mit realen Daten über insgesamt 50 Epochen. In dieser Trainingskonfiguration wurde die YOLOv5s-Architektur sorgfältig auf einem umfassenden Datensatz aus der realen Welt über 50 Epochen trainiert. Nach diesem intensiven Trainingsregime wurde die Leistung des Modells gründlich anhand eines realen Datensatzes bewertet. Diese Methode beleuchtet das Verhalten des Modells, wenn es auf einen begrenzten Datensatz trifft, der jedoch die Merkmale des Testdatensatzes eng widerspiegelt. Eine solche Untersuchung liefert wertvolle Einblicke in die Generalisierungsfähigkeit des Modells in Szenarien, in denen die Datenverfügbarkeit begrenzt, aber kontextuell relevant ist (Tabelle 1).
- Training auf vollständig synthetischem Datensatz: In diesem Versuch wurde, wie auch bei den anderen Experimenten, die YOLOv5s-Architektur unter Verwendung des rein synthetischen Datensatzes über 50 Epochen trainiert. Anschließend wurde das trainierte Modell eingefroren, und das Testen wurde, wie bei den anderen Experimenten, mit dem realen Datensatz durchgeführt. Der reale Datensatz wurde so ausgewählt, dass er dem erforderlichen Szenario möglichst nahekommt. Die Ergebnisse dieses Experiments, die ebenso wie die anderen mit realen Daten bewertet wurden, finden sich in Tabelle 2.
- Hybrides Training (Real-Synthetic): Der dritte Ansatz umfasst das Training des Objekterkennungsmodells mit realen Daten über die ersten 25 Epochen, wonach die Parameter des

Netzwerks eingefroren werden. Anschließend wird das Modell für weitere 25 Epochen mit synthetischen Daten erneut trainiert. In diesem experimentellen Ansatz wurde die YOLOv5s-Architektur eingesetzt, die zunächst ausschließlich auf realen Daten für 25 Epochen trainiert wurde. Die vortrainierten Gewichte, die aus dieser ersten Trainingsphase mit realen Daten gewonnen wurden, dienen anschließend als Grundlage für den Trainingsprozess mit synthetischen Daten über weitere 25 Epochen. Die Bewertung der Modellleistung erfolgte erneut mit realen Daten und lieferte Einblicke in die Anpassungsfähigkeit und Robustheit des Modells entlang des synthetisch-realen Kontinuums (Tabelle 3 - Modell wurde zuerst mit realer Daten und dann synthetischer Daten trainiert und anhand realer Daten getestet).

- **Hybrides Training (Synthetisch-Real):** bei dieser Methode wird das Objekterkennungsmodell zunächst über 25 Epochen mit synthetischen Daten trainiert. Anschließend werden die Netzwerkparameter eingefroren, und das Modell wird in weiteren 25 Epochen mit realen Daten weiter trainiert. In dieser Phase des Experiments wurde die YOLOv5s-Architektur eingesetzt, um das Modell mithilfe synthetischer Daten über insgesamt 25 Epochen zu trainieren. Im zweiten Schritt wurde das vortrainierte Modell aus der vorherigen Phase verwendet, um es für weitere 25 Epochen mit realen Daten zu trainieren. Die Effektivität des Modells wurde dann durch eine Bewertung mit realen Daten getestet, um seine Leistung in praxisnahen Szenarien zu evaluieren (Tabelle 4).

Class	Images	Instances	P	R	mAP50	mAP50-95
All	50	115	0.759	0.912	0.867	0.658
BlackLid	50	4	0.554	1	0.945	0.682
BlackTrailer	50	9	0.838	1	0.984	0.906
Führerhaus	50	18	0.95	1	0.995	0.719
GlueGun	50	2	0.791	1	0.995	0.895
Deckel_Vierecke	50	6	0.725	0.667	0.805	0.424
Trailer_Body_White	50	6	0.293	0.833	0.323	0.192
CAB_Chassis	50	36	1	0.8	0.914	0.726
SemiTrailer_Chassis	50	34	0.926	1	0.974	0.717

Tabelle 1 - Modell wurde anhand realer Daten trainiert und anhand realer Daten getestet

Class	Images	Instances	P	R	mAP50	mAP50-95
All	50	115	0.598	0.698	0.686	0.411
BlackLid	50	4	0.238	0.25	0.264	0.11
BlackTrailer	50	9	0.829	1	0.984	0.66
Führerhaus	50	18	0.947	0.889	0.938	0.657
GlueGun	50	2	0.341	0.5	0.249	0.0502
Deckel_Vierecke	50	6	0.768	1	0.995	0.581
Trailer_Body_White	50	6	0.369	0.667	0.589	0.272
CAB_Chassis	50	36	0.939	0.427	0.766	0.454
SemiTrailer_Chassis	50	34	0.353	0.853	0.704	0.502

Tabelle 2 - Modell wurde anhand synthetischer Daten trainiert und anhand realer Daten getestet

<i>Class</i>	<i>Images</i>	<i>Instances</i>	<i>P</i>	<i>R</i>	<i>mAP50</i>	<i>mAP50-95</i>
<i>All</i>	50	115	0.818	0.708	0.819	0.529
<i>BlackLid</i>	50	4	0.904	0.75	0.777	0.48
<i>BlackTrailer</i>	50	9	0.863	1	0.995	0.779
<i>Führerhaus</i>	50	18	0.788	0.944	0.937	0.615
<i>GlueGun</i>	50	2	0.848	1	0.995	0.697
<i>Deckel_Vierecke</i>	50	6	1	0	0.543	0.262
<i>Trailer_Body_White</i>	50	6	1	0	0.361	0.191
<i>CAB_Chassis</i>	50	36	0.47	0.972	0.961	0.673
<i>SemiTrailer_Chassis</i>	50	34	0.671	1	0.985	0.638

Tabelle 3 - Modell wurde zuerst mit realer Daten und dann synthetischer Daten trainiert und anhand realer Daten getestet

<i>Class</i>	<i>Images</i>	<i>Instances</i>	<i>P</i>	<i>R</i>	<i>mAP50</i>	<i>mAP50-95</i>
<i>All</i>	50	115	0.921	0.904	0.96	0.73
<i>BlackLid</i>	50	4	0.929	0.75	0.912	0.792
<i>BlackTrailer</i>	50	9	0.886	1	0.984	0.889
<i>Führerhaus</i>	50	18	0.982	1	0.995	0.738
<i>GlueGun</i>	50	2	0.922	1	0.995	0.796
<i>Deckel_Vierecke</i>	50	6	1	0.844	0.995	0.602
<i>Trailer_Body_White</i>	50	6	0.708	0.833	0.866	0.566
<i>CAB_Chassis</i>	50	36	1	0.805	0.962	0.74
<i>SemiTrailer_Chassis</i>	50	34	0.944	1	0.968	0.714

Tabelle 4 - Modell wurde zuerst mit synthetischer Daten und dann realer Daten trainiert und anhand realer Daten getestet

Diese experimentellen Konfigurationen bieten Einblicke, wie das Objekterkennungsmodell auf reale Daten und Variationen in der Datenzusammensetzung reagiert. Außerdem beleuchten sie die Anpassungsfähigkeit des Objekterkennungsmodells an synthetische Daten, die einen repräsentativen Teil des Datensatzes darstellen.

Während der anschließenden Diskussionsphase unserer Forschung konzentrieren wir uns auf zwei Hauptkonzepte. Zunächst untersuchen wir die Auswirkungen der Verwendung von realen und synthetischen Daten während des Modelltrainings. Darüber hinaus befassen wir uns mit dem Einfluss des kontinuierlichen Lernens und der Aktualisierungen zwischen Klassen auf die Generalisierungsfähigkeit der Objekterkennung, insbesondere für die Klassen im Zusammenhang mit dem Szenario „Manual Assembly“. Die Ergebnisse zeigen, dass das Training des Objekterkennungsmodells ausschließlich mit realen oder synthetischen Daten zu suboptimaler Leistung führt. Besonders hervorzuheben ist, dass das Modell, das mit einer Kombination aus realen und synthetischen Daten trainiert wurde (wie in Tabelle 5 mit einem beeindruckenden mAP50-Wert von 0,96 dargestellt), eine überlegene Präzision bei der Erkennung von Klassen im Zusammenhang mit dem Szenario „Manuelle Montage“ aufweist. Dies deutet darauf hin, dass synthetische Datensätze, die aus CAD-Modellen abgeleitet werden, zwar den Labeling-Prozess für Objekte, die Orakel erfordern,

vereinfachen, jedoch die Einbeziehung realer Daten notwendig ist, um eine zufriedenstellende Objekterkennungsleistung zu erzielen. Es ist erwähnenswert, dass unsere CAD-Modelle in unserer Forschung den realen Objekten stark ähneln.

Auf einer anderen Ebene thematisieren wir das kontinuierliche Lernen für die Objekterkennung. Unsere Experimente, die sequenzielles Training mit abwechselnden real-synthetischen und synthetisch-realen Daten umfassen, liefern bemerkenswerte Erkenntnisse. Wie in den Ergebnissen aus Tabelle 4 ersichtlich, führt die Einbeziehung realer Daten zur Aktualisierung des mit synthetischen Daten trainierten Modells zur besten Gesamtleistung des Objekterkennungsmodells. Vergleicht man dieses Ergebnis mit dem Training ausschließlich auf realen Daten, das einen mAP50-Wert von 0,867 erzielt, zeigt der Ansatz des sequentiellen Trainings auf synthetischen und anschließend realen Daten die optimalste Leistung.

Auf der nächsten Ebene können wir über eine Abfolge wahrgenommener menschlicher Aktivitäten und Umweltbeobachtungen nachdenken, indem wir das Problem der Absichts- oder Planerkennung als (umgekehrtes) Planungsproblem betrachten, bei dem die Eingabe eine Abfolge wahrgenommener Aktionen und die Ausgabe ein Ziel ist, das diese Abfolge erklärt [RG10, RG11, SRU16, FZ17]. Ein Großteil der Arbeiten in dieser Richtung basieren auf klassischen Planern, die Determinismus und vollständige Beobachtbarkeit annehmen, obwohl auch einige Arbeiten durchgeführt wurden, um fehlende Informationen zu berücksichtigen [SRU16] und die Zielerkennung auf probabilistische und teilweise beobachtbare Bereiche auszudehnen [RG11].

Unser AGR-Problem ist von Natur aus teilweise beobachtbar, daher haben wir uns für das POMDP-Darstellungsmodell entschieden. Bestehende Ansätze haben sich mit AGR befasst, indem sie faktorisierte POMDPs [AB19] verwendet und Vereinfachungen mit gemischter Beobachtbarkeit eingeführt haben [MB21]. Stattdessen haben wir jedoch ein Modell vorgeschlagen, das erweiterte POMDP-Zustände mit Arbeiter- und Beobachervariablen verwendet, bei denen keine explizite Faktorisierung zugunsten einer flachen Darstellung vorgenommen wird, was bedeutet, dass keine manuelle Vorverarbeitung erforderlich ist. Dieses Modell eignet sich außerdem gut für die Online-, generative MCTS-Planung und profitiert von POMDP-Planungsverbesserungen wie dem relevanzbasierten Ansatz unseres eigenen Planers RAGEPlan. Das vorläufige Modell und die aktuellen Ergebnisse wurden kürzlich auf dem ICAPS 2023-Workshop zum Thema „Plan-, Aktivitäts- und Absichtserkennung“ (PAIR) [SH23] vorgestellt. Dieses Dokument enthält auch eine kurze analytische Diskussion über die Herausforderungen der Absichtserkennung und die Rolle der Zielschätzung.

Abbildung 8 veranschaulicht die verschiedenen Ebenen und Arten von Informationen, die die Intentionserkennung (IR) unterstützen. Auf der untersten Ebene ermöglichen Sensordaten und (rauschbehaftete) Beobachtungen dem POMDP-Planer, sein Weltmodell zu verfeinern und Werte zu berechnen, um die nächstbeste Aktion zu identifizieren, basierend auf einer Hypothese über den Arbeiterplan, die durch das interne Aufgabenmodell (oder den Arbeitersimulator) gestützt wird. Ein passiver Ansatz würde hier enden, indem er die Wahrscheinlichkeiten verschiedener Arbeiterpläne oder -aktionen auflistet, aber wir verwenden diese Informationen, um den Wert der eigenen Aktionen des Roboters in Bezug auf die Erreichung gemeinsamer Ziele zu berechnen. Auf der höheren Ebene wird, sobald eine Aktion identifiziert und ausgewählt wurde, diese zu einem neuen „Ziel“, das der Roboter physisch ausführen kann. In den meisten Fällen führen diese Aktionen zu neuen Beobachtungen und der Zyklus beginnt von Neuem.

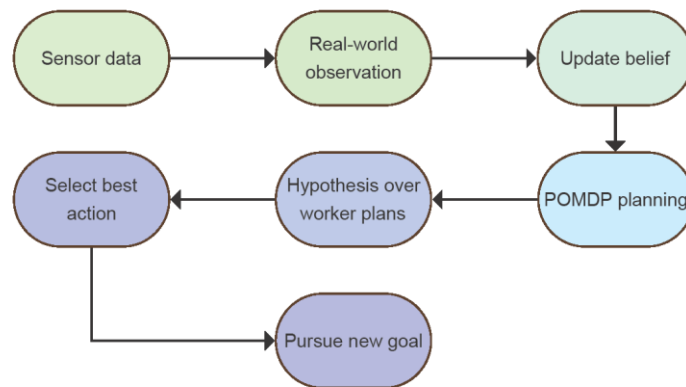


Abbildung 8 - Verschiedene Abstraktionsebenen zur Absichtserkennung

Eine grundlegende Herausforderung bei dieser Art von Problemen besteht darin, rauschbehaftete Beobachtungen effektiv zu nutzen, um den tatsächlichen Zustand der Welt zu approximieren (in POMDPs durch den Glaubenszustand), aber in unserer Version der Ziel- und Intentionserkennung müssen wir auch über Aktionen mit verzögerten Belohnungen nachdenken. Wenn der Roboter feststellt, dass der Arbeiter bald ein bestimmtes Werkzeug benötigen könnte, und im Voraus handelt, um dieses Werkzeug bereitzustellen, wird der wahre Nutzen dieser Aktion erst später wahrgenommen, wenn der Arbeiter tatsächlich dieses Teilziel erreicht. Werkzeuge bereitzustellen ist nur auf der Basis von Bedarf und Intention sinnvoll und ist kein Ziel an sich. Intern führt dies zu Problemen mit der Dimensionalität von POMDPs und der Komplexität der Planung in großen Räumen. Diese und andere Probleme können durch Relevanzschätzung [SH19b, SH19a] angegangen werden, und wir argumentieren, dass diese Techniken gut für AGR geeignet sind. Zusammenfassend haben wir eine Grundlage geschaffen, um teilweise beobachtbare AGR-Probleme online anzugehen, aber interessante offene Fragen bleiben, die durch den Ausbau unserer Beiträge behandelt werden könnten.

Innerhalb des Zeitrahmens dieses Projekts erzielten wir vielversprechende Ergebnisse bei der Lösung von AGR-POMDPs, die weit verbreiteten Planer wie den uniform zufälligen POMCP [SV10] übertreffen. Es ist zu beachten, dass die Aktionen auf POMDP-Ebene noch recht abstrakt sind (z.B. „Multimeter bringen“) und nicht direkt den physischen Fähigkeiten spezifischer Roboter entsprechen. Dies sind Aufgaben auf hoher Ebene, die von einem Aufgabenplanungs- und Ausführungssystem implementiert und ausgeführt werden müssen, das die entsprechenden Navigations-, Greif- und zusätzlichen Wahrnehmungsschritte identifiziert, steuert und überwacht. Das bedeutet, dass der AGR-POMDP auch Funktionen ausführt, die in den Bereich der Zielerkennung fallen könnten, aber diese Ziele reagieren auf neue Bedingungen, die durch die Beobachtung eines Arbeiters und die Schätzung seiner möglichen Bedürfnisse ausgelöst werden.

Die Leistung der Intentionserkennung mittels POMDP-Planung wurde durch eine Reihe von relativ gründlichen statistischen Experimenten getestet (in unserem AGR-Papier). Zusätzlich führten wir mehrere Experimente in realen Szenarien mit allen verfügbaren Modulen an Bord des Mobipick-Roboters durch, die es uns ermöglichten, einige der zugrunde liegenden Modelle zu verfeinern und die Fähigkeit des Roboters zu verbessern, auf neue Bedürfnisse zu reagieren. Diese Ergebnisse zeigen, dass unsere Plattform einem Arbeiter effektiv helfen kann, seine Aufgaben und Ziele zu erfüllen, indem sie verschiedene Arten von Informationen aus unterschiedlichen Quellen nutzt. Darüber hinaus half der

Ansatz des Projekts zur Intentionserkennung bei der Gestaltung der Mensch-Roboter-Interaktionstests in der Evaluierungsphase.

1.4.4 Fusion der Schätzungen (AP 4.4)

Wir haben mehrere Möglichkeiten identifiziert, verschiedene Informationsebenen zu integrieren:

- Objekterkennungs- und Eye-Tracking-Informationen können in Kombination mit dem Umgebungsmodell verwendet werden. Beispielsweise kann der digitale Zwilling die Position und Ausrichtung des Arbeiters verwenden, um ihm bei der Navigation oder beim Auffinden eines fehlerhaften Produktionsmoduls zu helfen.
- Das Umgebungsmodell enthält Daten (wie einen hierarchischen Objektbaum), die mit der Objekterkennung kombiniert werden können, um die Schätzung der Objektposition und -ausrichtung zu verbessern.
- Der digitale Zwilling und das Umgebungsmodell können die Entitäten und Mengen von Ersatzteilen speichern und verwendet werden, um die Verfügbarkeit eines Ersatzteils für den nächsten Schritt abzuschätzen/zu erkennen.

In InCoRAP können wir zwei Hauptakteure unterscheiden, die miteinander interagieren: ein menschlicher Arbeiter und ein robotischer Assistent. Der Arbeiter erhält Unterstützung vom Roboter sowie von digitalen Werkzeugen wie einem Umgebungsmodell und einem AR-Gerät. Der Roboter wiederum unterstützt den Arbeiter, indem er dessen Bedürfnisse antizipiert und erfüllt. Die oben beschriebenen Schritte betreffen in erster Linie die Bedürfnisse des Arbeiters, und obwohl sie gemeinsame Ziele haben, sind die Aufgaben, die Dynamik der Welt und die Menge an Informationen, die dem Roboter zur Verfügung stehen, grundlegend unterschiedlich.

Wir haben zusätzliche Integrationsstufen – oder Pfade – eingerichtet, um das gesamte robotergestützte Assistenzsystem zu unterstützen. Unser Vorschlag zur Erfüllung der Anforderungen des Roboters besteht aus einem modularen System, das aus Schichten besteht, die denen bekannter Architekturen ähneln, die üblicherweise bei der Gestaltung intelligenter Systeme verwendet werden, vgl. [MA02]. Zusammengefasst:

- In der Wahrnehmungsphase (Perception) erkennen wir Objekttypen und -positionen mittels Objekterkennung, die mit Hand- und Blickverfolgungsdaten (Eye-Tracking) kombiniert wird. Diese Daten dienen als Eingabe für das Aufgabenmodell, um das Interaktionsniveau des Arbeiters zu verstehen und folglich dessen Intentionen zu erkennen. Der AGR-Planer erwartet symbolische, abstrakte Informationen, die bei Bedarf durch ein Modul bereitgestellt werden (das sich noch in der Entwicklung befindet), das die kombinierten Erkennungsdaten mit höherwertigen Informationen über Objekttypen, Arbeiteraktionen usw. verknüpft.
- In der Planungs- und Überlegungsphase (Planning & Deliberation) befasst sich der AGR-Planer mit den Ergebnissen und Wahrscheinlichkeiten von hochrangigen Roboteraktionen, einschließlich der Auswirkungen der Informationssammlung und deren Beitrag zur Zielerreichung. Informationen aus der Wahrnehmungsphase leiten diesen Planungsprozess und ermöglichen es dem zugrunde liegenden POMDP, sein Arbeitsmodell (d. h. den Glaubenszustand des Roboters) zu verfeinern. Die Beobachtung und Unterstützung eines menschlichen Arbeiters ist unserer Meinung nach von Natur aus probabilistisch und nur teilweise beobachtbar. Diese Phase generiert hochrangige Aktionen, die unmittelbare „Ziele“ für den Roboterassistenten darstellen.

- In der Aufgabenplanung und Ausführungsphase (Task-planning and Execution) werden Informationen über Ziele verwendet, um ein Planungsproblem für die Robotik zu erstellen und die entsprechenden Schritte auszuführen, die solche unmittelbaren Ziele verwirklichen. Während dieser Phase findet ein unabhängiger Zyklus aus Wahrnehmung, Sensorinformationsverarbeitung und Aufgabenplanung statt (wie es in der Robotik typisch ist), um Werkzeuge oder zusätzliche Gegenstände zu lokalisieren und aufzuheben.

Abbildung 9 zeigt einen funktionalen Überblick über unseren AGR-Ansatz, der zur weiteren Verdeutlichung die Ergebnisarten aus dem „Manual Assembly“ Szenario verwendet. Während dieses Prozesses kann der Roboter auch vom Umgebungsmodell profitieren, soweit es die vollständig beobachtbare, deterministische Dimension des Problems erlaubt. Zum Beispiel kann der digitale Zwilling dem Roboter die Positionen von Teilen und Werkzeugen mitteilen, wenn das System eine „Bring“-Aktion ausführt.

Gegen Ende des Projekts konzentrierten wir uns auf das Testen und Verbessern des Informationsflusses durch die verschiedenen Ebenen oder Schichten, was eine kurze Überprüfung der Art und Weise beinhaltet, wie wir Informationen erfassen, darstellen und dem POMDP-Planer zur Verfügung stellen, und wie der POMDP-Planer Aktionen an die Ausführungs- und Überwachungsmodule weitergibt. Wir waren in der Lage, sehr erfolgreiche Demonstrationen mit allen Komponenten durchzuführen, die zum Zeitpunkt der Tests verfügbar waren. Der Einsatz eines POMDP-Planers als Online-„Orchestrator“ oder Koordinator von Roboterarbeiten ist nicht nur effektiv, sondern auch innovativ, und wir gehen davon aus, dass sich dieser Ansatz auf künftige Projekte auswirken wird, die sich mit Absichtserkennung, gemischter Initiativplanung und kollaborativer Robotik im Allgemeinen beschäftigen.



Abbildung 9 - Funktionale Ansicht des Active Goal Recognition-Frameworks im Szenario „Manual Assembly“

1.5 AP5: Situationsspezifische Werkerassistenz

1.5.1 Framework für die Zusammenarbeit zwischen Workern und Robotern (AP 5.1)

Für dieses Arbeitspaket wurden die Anforderungen an das Framework für die Mensch-Roboter-Kollaboration in mehreren Workshops mit dem gesamten Konsortium aufgenommen, eine Architektur

für dieses Framework entwickelt und die Schnittstellen zwischen der Intentionserkennung, dem semantischen Umgebungsmodell und der Roboteraktionsplanung innerhalb des Frameworks auf konzeptioneller Ebene festgelegt. Die Kommunikationsprotokolle konnten jedoch während der Entwicklungsphase geändert werden. Es wurde festgelegt, dass das semantische Umgebungsmodell das zentrale Element für die Zusammenarbeit sein sollte, das alle für die Zusammenarbeit notwendigen Informationen enthält. Diese Informationen werden u.a. durch die Daten des Roboters, der AR-Brille des Werkers und der dynamischen Veränderungen in der Produktionslinie ständig aktuell gehalten und können von den einzelnen Komponenten mittels SPARQL-Abfragen bedarfsgerecht abgefragt werden.

In verschiedenen Integrationsworkshops in der SmartFactory-KL wurde die Integration der verschiedenen Softwarekomponenten im Projekt diskutiert. Das Assistenzsystem sollte das Verhalten des Arbeiters durch verschiedene Sensordaten, wie z.B. Objekterkennung und Eye-Tracking, überwachen und analysieren, die zusammen nützliche Informationen oder eine Beobachtung für den Roboter darstellen. Auch die Schnittstelle zwischen dem Roboter und dem Arbeiter und die Art und Weise, wie der Roboter die Beobachtungen erhält, wurde definiert. Zusätzlich organisierten wir zwei Integrationsveranstaltungen (eine in Kaiserslautern, eine in Osnabrück), bei denen wir die Manipulations- und Greiffähigkeiten des Mobipicks testeten, indem wir InCoRAP-Objekte als Reaktion auf (simulierte) Umweltbeobachtungen brachten. Darüber hinaus wurden mehrere studentische Forschungsprojekte zum Thema „Human Robot Interaction via Augmented Reality“ durchgeführt. Im Rahmen dieser Studentenprojekte wurde ein Rahmenwerk entwickelt, das eine effektive Zusammenarbeit zwischen Menschen und Robotern durch den Einsatz einer AR-Anwendung ermöglichen soll.

1.5.2 AR-Visualisierung des Prozess- und Umgebungszustands, einschließlich der Visualisierung der Assistenz durch den Roboter (AP 5.2)

In diesem Arbeitspaket wurde eine AR-Anwendung entwickelt, in der die Prozessschritte aus dem definierten Task-Modell sowie der Zustand der Produktionsmodule auf der HoloLens visualisiert werden. Hierfür wurden verschiedene industrielle Kommunikationsprotokolle wie OPCUA getestet. Anschließend wurde eine Verbindung mit der ROS-Umgebung des Roboters hergestellt. Dies geschah, um die laufenden Aktionen und den aktuellen Status des Roboters innerhalb der AR-Anwendung anzuzeigen. Diese Einbindung könnte sich als vorteilhaft erweisen, um den Arbeiter während der Aufgabenausführung durch den Roboter auf dem Laufenden zu halten und die Sicherheit der Zusammenarbeit zwischen Arbeiter und Roboter zu erhöhen.

1.5.3 Dynamische Affordances und Handlungsempfehlungen (AP 5.3)

Ein weiterer wichtiger Punkt ist die Erfassung und Visualisierung von dynamischen Affordances für den Arbeiter: Abhängig von der aktuellen Situation und dem gewählten Pfad im Ausführungsmodell können sich die Handlungsmöglichkeiten für bestimmte Objekte in der Umgebung ändern. Dies sollte für den Werker entsprechend visualisiert werden. Dazu wurde in einem ersten Schritt ein Katalog aller Objekte in den Szenarien erstellt, die für den Werker relevant sind. Diesen Objekten wurden dann die Aktionen zugeordnet, die der Werker mit ihnen durchführen kann. Im nächsten Schritt wurde dieser Katalog mit der Absichtserkennung für das Szenario der manuellen Montage verknüpft, um dem Werker nur die in der aktuellen Situation relevanten Aktionen (bzw. die nächstmöglichen Aktionen) anzuzeigen. Der Katalog wird auch bei der Erstellung der Aufgabenmodelle verwendet (AP4). Die Implementierung ist im

Gänge, um die durchführbaren Aktionen für den Arbeiter zu visualisieren, abhängig von den Ergebnissen der Absichtserkennung (d.h. Umgebungsbeobachtungen, die aus der Objekterkennung und den räumlichen Eingaben des Menschen wie Blickpunkt und Handgelenke gewonnen werden).

1.5.4 Innovative Eye-Tracking-Interaktion mit AR (AP 5.4)

Die Objekterkennung identifiziert Objekte ohne Berücksichtigung des Ausmaßes der Interaktion. Durch die Einbeziehung des Blicks in das System können wir daher feststellen, ob der Arbeiter tatsächlich mit dem erkannten Objekt interagiert. Zu diesem Zweck wurde eine Überlappung des Blickpunkts und des BoundingBox (aus der Objekterkennung) des erkannten Objekts über eine bestimmte Anzahl von Frames analysiert. Mit anderen Worten, wenn der Fokus des Arbeiters mit dem erkannten Objekt übereinstimmt, was auf eine Region von Interesse (Region of Interest) hinweist, bedeutet dies, dass der Arbeiter beabsichtigt, mit dem Objekt zu interagieren.

1.6 Integration menschlicher Intentionen in die Handlungsplanung des Roboters (AP6)

1.6.1 Handlungsplanung für ein kollaboratives Team (AP 6.1)

In diesem Arbeitspaket wurde der CHIMP-Planungsbereich zunächst um einige Arbeiteraktionen erweitert. Es wurde erwartet, dass der Arbeiter sich an einen Ort begibt und dort auf die Lieferung eines Objekts wartet, und der Planer würde dies berücksichtigen. Wir zogen in Erwägung, in beiden Szenarien zusätzliche Arbeiteraktionen als Teil des Planungsbereichs aufzunehmen.

Die genannten Aktionen erleichtern die Ausführung von Roboterarbeiten, wie z. B. das Bringen von Werkzeugen und Teilen, was ein wichtiger Teil der Zusammenarbeit ist, bei der es darum geht, wie der Arbeiter auf den Roboter reagiert. Es gibt jedoch noch eine weitere Ebene der Zusammenarbeit, die sich aus der richtigen Interpretation der Bedürfnisse des Arbeiters auf der Grundlage von Beobachtungen ergibt. Wie oben erläutert, wurde ein Modell für die Aktionen des Arbeiters in den AGR-Planer aufgenommen, damit dieser Hypothesen über den Raum der Pläne des Arbeiters aufstellen und Schlussfolgerungen über die Auswirkungen und Wahrscheinlichkeiten seiner eigenen Aktionen sowie über die Zielbedingungen ziehen kann. Dies ermöglicht es dem Roboter, kontextbezogene Informationen zu nutzen und auf verschiedene Arbeiter sowie sich ändernde Bedingungen und neue Anforderungen zu reagieren, solange die grundlegenden Bausteine (z. B. Objekte, Arbeitstätigkeiten und Aufgaben) mit ausreichender Genauigkeit erkannt werden können. Hier geht es darum, wie der Roboter auf den Arbeiter reagiert.

In verschiedenen Tests und internen Vorführungen haben wir die Fähigkeit des Roboters getestet, auf die Handlungen der Arbeiter zu reagieren - mit sehr guten Ergebnissen. Alle unsere Planer konnten die Eingaben rechtzeitig verarbeiten und brauchbare Ergebnisse liefern (sowohl auf der IR-Ebene als auch auf der Ebene der Robotersteuerung). Die größte Einschränkung beim Versuch, mit einem schnellen Arbeiter Schritt zu halten, war in der Praxis die Roboterplattform selbst, die sich sowohl aus technischen als auch aus Sicherheitsgründen relativ langsam bewegt. Verschränktes Planen und Handeln ist ein interessantes und vielversprechendes Thema für sich, und unser Ansatz bildet einen Rahmen, der zusätzlich das aktuelle und wichtige Thema der Mensch-Roboter-Kollaboration berücksichtigt.

1.6.2 Planausführung und Ausführungsüberwachung (AP 6.2)

Im Laufe des Projekts fügten wir eine experimentelle HTN-Unterstützung in ROSPlan [CFL+15] hinzu und schlossen Vorkehrungen für die von CHIMP generierten Pläne ein (dargestellt als Esterel-Graph oder als hierarchischer Plan). Ein einfacher sequentieller Planausführer und -monitor (plexmo) wurde implementiert, um den ROSPlan-Plan-Dispatcher zu ersetzen.

Als sich die Bedingungen änderten, wechselten wir von ROSPlan zu einer Kombination aus der Unified Planning Library (UP) [MABM+22] und einer kundenspezifischen, für eingebettete Systeme entwickelten Bridge [SSS+23]. Zusammen bieten sie Unterstützung bei der Planung und Ausführung eingehender Aufgaben, z. B. aus dem Modul zur Absichtserkennung. Abbildung 10 veranschaulicht diesen Rahmen. CHIMP ist noch nicht in UP integriert, daher verwenden wir den HTN-Planer Aries [BM23], bei dem ein sequentieller oder nebenläufiger Planausführer zur Laufzeit ausgewählt werden kann. Die Ausführung kann mit Hilfe eines Abhängigkeitsgraphen ähnlich dem ROSPlan Esterel Plan Dispatch gehandhabt werden. Mit diesem Technologiepaket kann der Roboter erkennen, ob ein Arbeiter ein bestimmtes Werkzeug benötigt (oder ob er keine Teile mehr hat), und eine übergeordnete Aufgabe erstellen. Diese Aufgabe wird ausgeführt, indem er zu einem bestimmten Ort fährt, das benötigte Werkzeug oder die fehlenden Teile aufnimmt, zurück zum Arbeiter navigiert und eine Übergabebewegung ausführt. Dieser Schritt wurde anhand simulierter Beobachtungen getestet, die einer realen Situation entsprachen und für den Roboter nicht von denen zu unterscheiden sind, die mit dem InCORAP-Modul zur Objekt- und Aktivitätserkennung gewonnen wurden (zum Zeitpunkt des Tests war es nicht verfügbar). Mehrere Elemente in der Planüberwachung und Ausführungspipeline wurden ebenfalls verbessert, einschließlich der Fähigkeit, sich von Aktionsfehlern zu erholen, indem Statusmeldungen (über physische Roboterarbeiten) zurück an den POMDP-Planer geliefert werden. So kann zum Beispiel eine „Bring“-Aktion nicht erfolgreich sein, wenn das entsprechende Teil nicht gefunden werden konnte oder das Greifen aufgrund mechanischer Probleme fehlgeschlagen ist. Auf diese Weise kann das AGR-POMDP diese Informationen nutzen und entscheiden, ob ein neuer Versuch unternommen oder ein neues Ziel verfolgt werden soll.

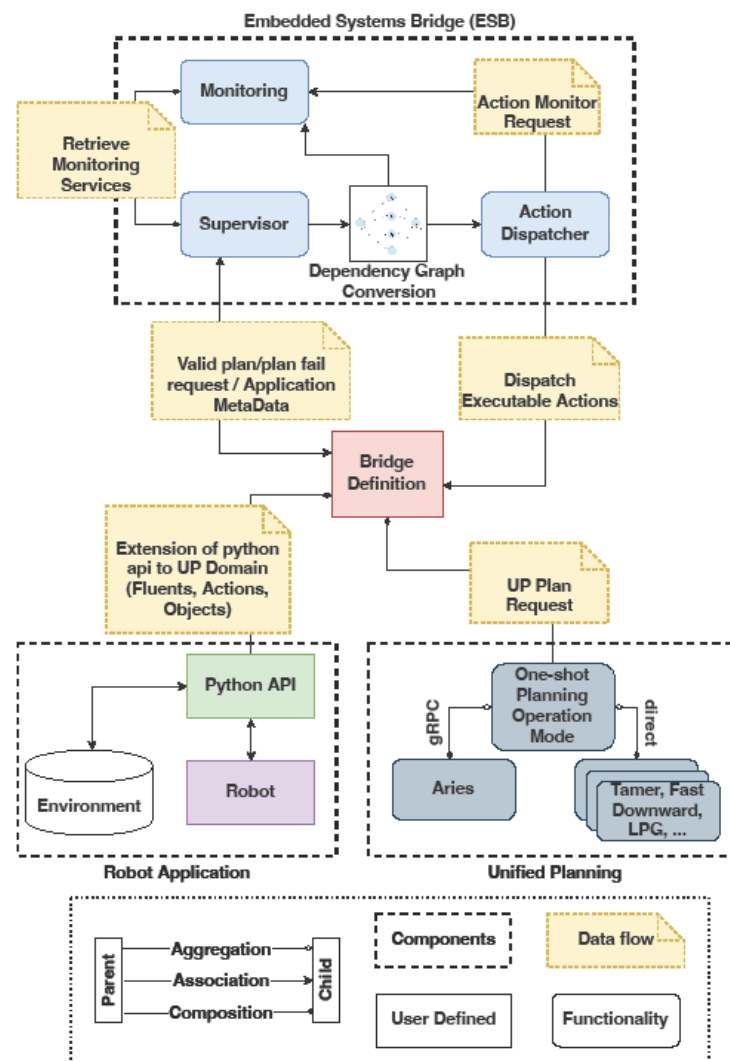


Abbildung 10 Architektur des Planungs- und Ausführungsrahmens

1.6.3 Anpassung der Pläne als Reaktion auf geänderte Intentionen (AP 6.3)

Zunächst wurden Ziele in die ROSPlan-Wissensbasis eingefügt und zur Erstellung eines Plans verwendet. Ein Ziel bestand aus Aufgaben wie dem Abholen eines vom Arbeiter benötigten Objekts, aber es waren weitere Arbeiten erforderlich, um sich an Änderungen und allgemeine Unsicherheit anzupassen.

Wir haben unser System so erweitert, dass es mit verschiedenen Arten von Ungewissheit umgehen kann, z. B. mit nicht-deterministischen Handlungsergebnissen (sowohl für den Roboter als auch für den Arbeiter) sowie mit dem Umstand, dass der Roboter nur Zugang zu teilweisen oder unvollständigen Informationen hat. Anstatt das gesamte Problem im Voraus zu lösen, wird ein POMDP online gelöst, indem Planung und Handlung ineinander übergehen, so dass der aktuelle Plan immer auf dem neuesten Stand ist. In unserem Modell versucht der Roboter, die nächstbeste Aktion zu finden, die auf der wahrgenommenen Abfolge von Beobachtungen basiert und durch sein internes Arbeitermodell und Zustandsübergangsmodell unterstützt wird.

Die sich daraus ergebende Strategie (oder der Plan) berücksichtigt die Möglichkeit, dass Handlungen fehlschlagen könnten, ob Teile oder Werkzeuge benötigt werden und was eine bestimmte Tätigkeit der Mitarbeiter über ihre Ziele aussagt. In der Praxis bedeutet dies, dass unterschiedliche Aktivitäten der Arbeiter und unterschiedliche Umgebungsereignisse (die als Sequenzen von Beobachtungen wahrgenommen werden) zu unterschiedlichen optimalen Handlungen führen werden. Da die Absichten (Intentionen) als mögliche Zielzustände, die der Arbeiter verfolgt, geschätzt werden, wird eine Änderung der Pläne (oder Intentionen) auf der Arbeiterseite bereits bei der Bewertung der nächsten auszuführenden Aktion berücksichtigt. Natürlich beschränken wir uns darauf, den Raum der Pläne - und damit der Absichten - auszuwerten, der bereits im Aufgabenmodell des Arbeiters berücksichtigt wurde.

In den verschiedenen Experimenten und Demos war der Roboter in der Lage, sich an die Bedürfnisse der Arbeiter anzupassen und nützliche Aktionen auszuführen, sofern geeignete Beobachtungen gemacht wurden. Wie bereits erwähnt, lösten wir die InCoRAP-Aufgaben mit guten Ergebnissen unter Verwendung simulierter Beobachtungen. Der AGR-Planer und die Komponenten der Robotersteuerung wurden jedoch auch in Kombination mit einem Platzhalter-Wahrnehmungssystem auf der Grundlage einer ähnlichen Aufgabe getestet. Unter Verwendung eines ähnlichen Aufbaus entwickelten wir ein einfaches Wahrnehmungssystem mit Kameras und einem Objekterkennungsmodul, das den Roboter mit Echtzeitschätzungen von Arbeiteraktionen und Objekteigenschaften versorgte. Der Mobipick-Roboter war mit Hilfe von InCoRAP in der Lage, die Bedürfnisse eines menschlichen Arbeiters richtig einzuschätzen und auf sie zu reagieren. Dieses Platzhalter-Wahrnehmungssystem, das für interne Tests verwendet wurde, kann leicht gegen ein anderes trainiertes Erkennungsmodul ausgetauscht werden.

In unseren Tests hat der Roboter, wenn der Arbeiter sich in einer Phase seines Plans befand, in der er ein bestimmtes Teil benötigte, zunächst geprüft, ob das Teil verfügbar war, bevor er es brachte. Wenn der aktuelle Plan des Arbeiters unsicher war, beobachtete der Roboter weiterhin andere Teile und gab ihnen möglicherweise Vorrang, bis er anhand der Handlungen des Arbeiters feststellen konnte, ob ein fehlendes Teil ebenfalls notwendig ist. Mit anderen Worten: Der Roboter versucht, mehr über die Absichten des Werkers und die verfügbaren Ressourcen zu erfahren, bevor er handelt.

Obwohl dieses Verhalten aufgrund unserer statistischen Leistungsanalyse erwartet wurde, bot die Entwicklung und Prüfung an Bord unseres physischen Roboters eine weitere Sicherheitsebene und half, die Vorteile unseres Rahmens zu veranschaulichen. Auf diesen Grundlagen kann ein vollständig autonomer Ansatz aufgebaut werden, vorausgesetzt, der AGR-Planer hat Zugang zu einem zuverlässigen Wahrnehmungssystem und geeigneten Aufgabenmodellen für den Arbeiter. Wir erwarten, dass diese Arbeit zu neuen Projektvorschlägen in der planbasierten Robotik beitragen wird.

1.7 Implementierung der Roboterfähigkeiten (AP7)

1.7.1 Lokalisierung und Navigation in der Umgebung (AP 7.1)

Analog zur Erstellung der statischen Objekte des in Abschnitt 1.3 beschriebenen Umgebungsmodells wurde mit Gazebo eine Simulationsumgebung aufgebaut, die Objekte aus den zuvor beschriebenen Szenarien enthält, und anhand einer in der physischen SmartFactory-KL aufgenommenen Karte validiert. Lokalisierung und Navigation des Mobipick-Roboters wurden parametrisiert und sowohl in der Simulation als auch physisch auf dem DFKI-PBR-Testgelände getestet. Transformationen zwischen Map-

Frame und Base-Frame sowie Gelenkwinkel wurden mit ROS#¹ an die Umgebungsdarstellung in Unity gesendet. Zur Steuerung des Roboters in der Umgebung verwendeten wir die Robot API, eine Schnittstellschicht zwischen dem Executor und der Bewegungssteuerung des Roboters, die einfache Aufrufe für Bewegung und Steuerung bereitstellt.

Die Robotersimulationsumgebung mit dem Namen „Mobipick Labs“ steht der Community zur Verfügung und wurde in einem ICAPS-Workshop vorgestellt [LGS+23b]. Diese Tools vereinfachen die Nutzung eines Robotersystems mit minimalem Aufwand und ohne die Notwendigkeit eines hohen Maßes an Fachkenntnissen, vor allem dank der Roboter-API, die geeignete Abstraktionen schafft.

Die Lokalisierungs- und Navigationssoftware wurde in allen Tests, die während der Integrationsworkshops durchgeführt wurden, weiter verwendet. Ein interessanter Punkt ist, dass wir, wie erwähnt, unsere Plattform auch in einem anderen, aber verwandten Bereich getestet haben und die entsprechenden Anpassungen einer bestehenden Aufgabe (LKW-Montage) relativ trivial waren, was die Benutzerfreundlichkeit der von uns gewählten Tools unterstreicht. Sowohl die Roboter-API als auch die Simulationsumgebung „Mobipick Labs“ sowie die im Rahmen dieses Projekts vorgenommenen Verbesserungen werden weiterhin für weitere Forschungs- und Demonstrationszwecke genutzt. Diese Ressourcen sind auch für den Rest der Gemeinschaft verfügbar.

1.7.2 Wahrnehmung der Umwelt (AP 7.2)

Wir haben DOPE (Deep Object Pose Estimation [TTS+18]) verwendet, um Objektposes zu schätzen, die anhand von angepassten CAD-Modellen der in den Anwendungsfällen beschriebenen Objekte trainiert wurden. Diese Informationen werden zur Aktualisierung der Objektposes in der Umgebungsdarstellung verwendet. Wir gehen davon aus, dass für jede Objektklasse nur eine Instanz existiert. Bei mehrfachem Vorkommen wäre eine Verankerung und ein Spawning der Objekte in der Umgebungsrepräsentation notwendig. Aus diesem Grund integrierten wir auch den Dynamic Anchoring Agent, der im Rahmen des Projekts CopDa (gefördert von BMBF) entwickelt wurde.

Mit Hilfe der Photogrammetrie haben wir fotorealistische CAD-Modelle neuer Objekte (wie z. B. eine Heißklebepistole und Farbkästen) erstellt und das DOPE-System mit synthetischen Daten trainiert. Das tiefe neuronale Netz von DOPE erkennt die Objekte im 3D-Raum, indem es nur RGB-Daten als Eingabe verwendet, obwohl anschließend eine geringfügige Anpassung der Objekthöhe erforderlich ist, für die Tiefeninformationen benötigt werden. Unter Verwendung der erkannten Objektposes empfängt ein automatisiertes Modul zur Generierung von Fließtexten die Boundingboxen der Objekte und erzeugt mit Hilfe der Kollisionsprüfer der Regale symbolische Beziehungen zwischen den Objekten in der Umgebung, z. B. *in(multimeter, blue_box)*, *on(Relais, Table_1)*. Schließlich erzeugen wir aus den Sensordaten zwei weitere Fluents: die semantische Position des Roboters und die Haltung des Arms, z. B. *robot_at(workstation_1)*, *arme_has_posture(home)*.

¹ <https://github.com/siemens/ros-sharp>

1.7.3 Ausführung von Manipulationsaufgaben (AP 7.3)

Die Planung und Ausführung von Manipulationsaufgaben mit dem Mobipick-Roboter wurde mit Hilfe des ROS-Frameworks MoveIt! [CSCC14] mit einem benutzerdefinierten Greifplanungsmodul namens Grasplan¹ durchgeführt, das aus dem EU-Projekt APRIL stammt. Mit Hilfe des Aufgabenkonstruktors von MoveIt! [GHRZ19] wurde ein Aufgabenserverknoten für die Roboter-API implementiert, in dem einfache Aktionen zu einer Aufgabe hinzugefügt werden können, wie z. B. die Bewegung zu einer bestimmten Position im kartesischen Raum oder Kombinationen mit einer vollständigen Greifaktion. Wir haben sowohl „Pick-and-Place-Container“ als auch „Hand-Over“-Verhalten implementiert. MoveIt! führt auch eine gewisse Kollisions- und Beschränkungskontrolle durch. Greifkonfigurationen wurden für alle InCoRAP-Objekte aufgezeichnet und der Mobipick-Greifer wurde in Grasplan integriert. Die Objekte können über den MoveIt-Pick-Action-Server gegriffen werden, der von einem eigenen Pick-Wrapper-Knoten aufgerufen wird. Die Posen werden von der Wahrnehmungskomponente empfangen und in einem Beobachtungspuffer zwischengespeichert (weitere Details in [LGS+23b]). Ein Systemdemopapier [LGS+23a] fasst diesen Prozess zusammen und enthält ein begleitendes Video².

Wie bei den anderen Modulen, die Roboterfähigkeiten implementieren, wurden das Greifen und die Manipulation in verschiedenen Tests mit zufriedenstellenden Ergebnissen validiert. Eine mögliche Einschränkung besteht darin, dass anspruchsvollere Konfigurationen, wie z. B. die Durchführung einer sanften „Übergabe“ oder das vorsichtige Ablegen eines Objekts auf verschiedenen Tischen, ein erhebliches Maß an manueller Feinabstimmung erfordern können. Dies ist natürlich eine Gelegenheit für weitere Forschungsarbeiten, die solche Aufgaben automatisieren können. In seiner jetzigen Form verfügt der Mobipick-Roboter über alle notwendigen Fähigkeiten, um die InCoRAP Aufgaben erfolgreich durchzuführen.

1.8 Realisierung der Use Cases (AP8)

1.8.1 Smart Maintenance Szenario (AP 8.1)

Der Anwendungsfall wurde so definiert, dass er am Storage&Assembly Modul im SmartFactory-KL-Labor demonstriert werden kann. In diesem Szenario wurde eine Anomalie im Stromverbrauch einer Maschine über eine Heizung simuliert. Die Erkennung von Anomalien wird von einer Software eines Drittanbieters durchgeführt, die von der Empolis Group³ entwickelt wurde und als Dienst oder Agent in einem Multiagentensystem betrachtet werden kann. In [RHH+20] wird beschrieben, wie Agenten in der neuen SmartFactory-KL-Demonstrationsanlage verwaltet werden. Dementsprechend wurden in mehreren

¹ <https://github.com/aprilprojecteu/grasplan>

² <https://youtu.be/4-GgOg2nuGE>

³ <https://www.empolis.com/en/>

Treffen mit Empolis die Möglichkeiten eines so genannten *Knowledge Graph* und dessen Verwendung für die Beschreibung von Anomalieerkennung und Wartungsschritten diskutiert. Abbildung 11 veranschaulicht, wie die verschiedenen Module kommunizieren, um den Arbeiter in diesem Szenario zu unterstützen. Das Knowledge-Graph-System verwendet die Energieverbrauchsdaten, um Änderungen im Verhalten des Produktionsmoduls zu überwachen. Das System bestimmt auch die durchführbaren Wartungsschritte, um die aufgetretene Anomalie zu beheben. Das Assistenzsystem für den Arbeiter nutzt die Informationen, um den Arbeiter durch eine AR-Anwendung auf der HoloLens zu führen. Die Absichten des Werkers helfen dem Werkerassistenten und auch den Wissensgraphen, die aktuellen, tatsächlich laufenden Schritte zu erkennen. Wurde die Störung behoben, wird dies dem ERP-System signalisiert.

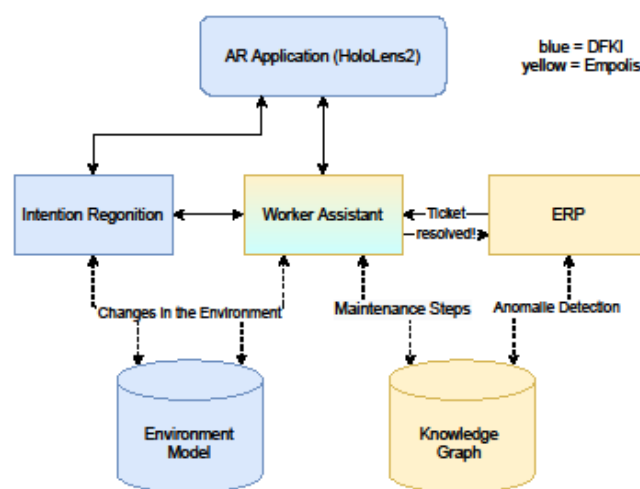


Abbildung 11 - Kommunikationsdiagramm der Softwarekomponenten zur Erkennung von Anomalien und des Arbeitsassistenten

1.8.2 Manual Assembly Szenario (AP 8.2)

Das Hauptziel in diesem Szenario ist es, den Benutzer (Arbeiter) bei der manuellen Montage eines Produkts in einer von mehreren möglichen Reihenfolgen zu unterstützen. Die Absichten des Arbeiters werden mit Hilfe von Objekterkennung und Eye Tracking erkannt. Das Assistenzsystem ist in der Lage, aus der Kombination von erkannten Objekten und dem Interessengebiet des Werkers die erledigten Aufgaben/Schritte und die bevorstehenden möglichen Aktionen zu bestimmen. Ein AGV nutzt die Ergebnisse der Absichtserkennung, um die notwendigen Montageteile für den nächsten Schritt zu bestimmen und sie zum Arbeiter zu bringen.

Zu Beginn des Projekts entschieden wir uns für einen sogenannten „Breadboy“ als Endprodukt dieses Anwendungsfalls. Diese Konfiguration bestand aus diversen elektronischen Komponenten. Während der Umsetzungsphase stießen wir auf eine Reihe von Problemen, wie z. B. komplexe Verfahren für den Arbeiter und eine lange Montagedauer. Wir haben jedoch einen Artikel [TWP+21] über die Erkennung sehr kleiner Objekte mithilfe eines zweistufigen Ansatzes zur Objekterkennung veröffentlicht. Daher haben wir uns entschlossen, das Bauprodukt in einen Miniatur-LKW umzuwandeln (Abbildung 12), der aus Noppensteinen und 3D-gedruckten Komponenten besteht. Wir haben auch zwei verschiedene Versionen des Lastwagens, jede mit unterschiedlichen Eigenschaften wie Farbe und zusätzlichen Merkmalen wie LED-Leuchten. Dies fügt dem Szenario ein herausforderndes Element hinzu und fordert

das Assistenzsystem auf, Informationen zu sammeln, bevor es handelt. Da das neue Produkt einen völlig neuen Satz von Montageteilen hat, mussten wir ein neues maschinelles Lernmodell mit YOLOv5 trainieren, das auf synthetischen Datensätzen und realistischen Bildern der Teile basiert. Es wurden verschiedene Kombinationen von Datensätzen getestet, um die genauesten Ergebnisse bei der Objekterkennung zu erzielen. Wie bereits erwähnt, ist die Identifizierung von elektrischen Komponenten mit dem OD-Ansatz aufgrund von Faktoren wie Größe und ähnlichen Layouts mit unterschiedlichen Funktionen sehr komplex. So weisen beispielsweise Widerstände extrem schmale farbige Linien auf, die verschiedene Widerstandscodes kennzeichnen, und Kabel sind nicht starr. Diese Herausforderungen lenken nicht nur die Ziele unseres Projekts ab, sondern wirken sich auch auf die Forschung zu synthetischen/realen Datensätzen aus. Folglich haben wir eine Strategie übernommen, die sich auf das gesamte Montageszenario konzentriert und den Miniatur-LKW als primäres Objekt mit verschiedenen Komponenten verwendet. Dieser Ansatz stellt zwei Hauptherausforderungen dar. Zunächst geht es um die Erkennung winziger Bauteile innerhalb des Zielobjekts. Zweitens werden die Bauteile durch 3D-Druck hergestellt, was zu CAD-Modellen führt, die den realen Proben stark ähneln. Daher stehen CAD-Modelle der Komponenten zur Generierung synthetischer Datensätze zur Verfügung.

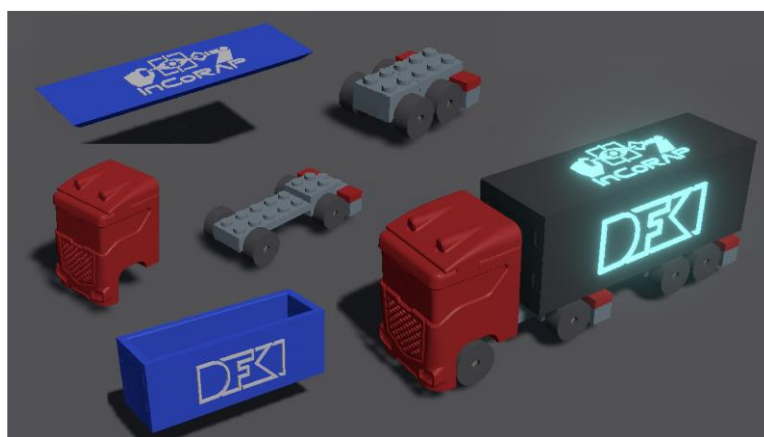


Abbildung 12 - Miniatur-LKW und seine Montageteile, maßgeschneidert für das Szenario der manuellen Montage

Natur. Diese Herausforderungen lenken nicht nur von den Zielen unseres Projekts ab, sondern wirken sich auch auf die Erforschung synthetischer und realer Datensätze aus. Daher haben wir uns für eine Strategie entschieden, bei der wir uns auf das gesamte Montageszenario konzentrieren und den Miniatur-LKW als Hauptobjekt mit verschiedenen Komponenten verwenden. Dieser Ansatz birgt zwei wesentliche Herausforderungen. Zum einen geht es um die Erkennung winziger Komponenten innerhalb des Zielobjekts.

1.9 Evaluation (AP9)

Um unser intentionsbasiertes Robotik-Assistenzsystem zu evaluieren, wurde ein Benutzertest mit 20 Teilnehmern durchgeführt. Sowohl eine qualitative als auch eine quantitative Bewertung wurde vorgenommen [WPS+24]. Die Teilnehmer führten Montageaufgaben in zwei Fällen durch: Fall 1 ohne Roboterunterstützung und Fall 2 mit Roboterunterstützung. Das gesamte Stressniveau der Teilnehmer in beiden Fällen wurde berechnet, um zu untersuchen, ob die Roboterunterstützung dazu beiträgt, die Stresslevel der Teilnehmer zu senken. Außerdem haben wir die wahrgenommene Benutzerfreundlichkeit

des Systems sowie die empfundene Arbeitsbelastung der Teilnehmer mithilfe von selbst ausgefüllten Fragebögen gemessen.

Messung des Stressniveaus: Die Hypothese ist, dass Arbeitnehmer im Fall 2 weniger Stress ausgesetzt sind als Arbeitnehmer im Fall 1. Die experimentelle Studie wurde in zwei Hauptphasen unterteilt. In den folgenden Absätzen werden die Phasen 1 und 2 erläutert.

- Phase 1: In dieser Phase haben wir ein Stresserkennungsmodell entwickelt, das einen Open-Source-Datensatz verwendet – WESAD [SRD+18] –, bei dem die Teilnehmer die Armbanduhr Empatica E4¹ verwenden. Die Teilnehmer wurden gebeten, bestimmte Aufgaben zu erledigen, um verschiedene affektive Zustände wie neutralen oder Grundzustand und Stresszustand hervorzurufen. Der WESAD-Datensatz ordnet physiologische Signale von Empatica E4 diesen affektiven Zuständen zu. In unserer Arbeit haben wir Grund-, Stress- und Vergnügensbezeichnungen im Datensatz verwendet, die jeweils als Null, Eins und Zwei kodiert sind. Für die binäre Klassifizierung wurden Grund- und Vergnügensklassen zu einer einzigen Klasse (ohne Stress) mit der Bezeichnung Null und Stress als weitere Klasse mit der Bezeichnung Eins kombiniert.

¹ <https://www.empatica.com/en-eu/research/e4/>

2 Wichtigste Positionen des zahlenmäßigen Nachweises

Eine Übersicht über die wichtigsten Positionen des zahlenmäßigen Verwendungsnachweises ist in Tabelle 5 zu sehen. Weitere Details gemäß Nr. 4.1 NKBF 2017 finden sich in der Anlage zu diesem Bericht.

Position	Gesamtvorkalkulation (€)	Gesamtnachkalkulation (€)
837 Personalkosten	1.405.939,00	1.383.322,00
838 Reisekosten	43.850,00	16.792,00
847 Abschreibung auf vorhabenspezifische Anlagen	22.008,00	21.472,23
850 sonstige unmittelbare Vorhabenkosten	9.278,00	8.075,41
860 Verwaltungskosten	614.255,00	678.527,13
881 gesamte Selbstkosten des Vorhabens	2.095.330,00	2.108.189,62

Tabelle 5 - Positionen des zahlenmäßigen Verwendungsnachweises

3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Notwendigkeit und Angemessenheit der semantischen Umgebungsmodellierung im InCoRAP-Projekt sind zentral für die Verbesserung der Zusammenarbeit zwischen Mensch und Roboter in komplexen modernen Produktionsumgebungen. Die semantische Modellierung der Umgebung ermöglicht es dem System, nicht nur physische Objekte zu erkennen, sondern auch deren Bedeutung und Funktion innerhalb des Arbeitsprozesses zu verstehen. Dies ist besonders wichtig, da die bloße Erkennung von Objekten ohne Verständnis ihrer Rolle und Beziehung zu anderen Komponenten zu suboptimaler Unterstützung führen kann. Durch die Integration eines detaillierten und kontextualisierten Modells der Umgebung kann der Roboter präzisere Entscheidungen treffen, welche Aufgaben er als Nächstes ausführen sollte, und somit den Arbeiter gezielt unterstützen. Diese Fähigkeit reduziert die kognitive Belastung des Arbeiters und verbessert die Effizienz des gesamten Arbeitsprozesses.

Die Angemessenheit der semantischen Umgebungsmodellierung zeigt sich in der Effektivität der entwickelten Lösungen innerhalb des Projekts. Die Kombination von CAD-Modellen, realen Sensordaten und der semantischen Modellierung ermöglicht es dem Roboter, in Echtzeit präzise auf die Bedürfnisse des Arbeiters einzugehen und die Umgebung intelligent zu interpretieren. Das Ergebnis ist ein robuster und adaptiver Assistenzmechanismus, der nicht nur die Interaktion zwischen Mensch und Roboter optimiert, sondern auch die Fehlerquote senkt und die Produktivität steigert. Die Modellierung und Integration von Umgebungswissen hat sich als entscheidend für das Verständnis und die effektive Nutzung der Produktionsumgebung erwiesen und stellt sicher, dass die Unterstützung des Roboters sowohl effizient als auch zielgerichtet ist.

In der heutigen Industrie sind menschliche Arbeiter oftmals hohen kognitiven und physischen Belastungen ausgesetzt, was zu Fehlern, Stress und verminderter Produktivität führen kann. Die Einführung eines intentionsbasierten Assistenzsystems zielt darauf ab, diese Herausforderungen zu bewältigen, indem es die Fähigkeit von Robotern verbessert, menschliche Absichten zu erkennen und

proaktiv Unterstützung zu bieten. So wird nicht nur die Arbeitslast reduziert, sondern auch die Effizienz der Produktionsprozesse gesteigert. Die Arbeit, die in InCoRAP geleistet wurde, adressiert diese Herausforderungen durch die Entwicklung innovativer Technologien wie Eye-Tracking, Hand-Tracking und fortschrittlicher Planungsalgorithmen, die es dem Roboter ermöglichen, in Echtzeit auf die Bedürfnisse des Arbeiters zu reagieren.

Die Angemessenheit der im Projekt geleisteten Arbeit liegt in der strukturierten Herangehensweise und der Integration modernster Technologien. Durch die Kombination von Sensordatenfusion, künstlicher Intelligenz und Robotik wurde ein robustes System geschaffen, das nicht nur theoretisch fundiert, sondern auch praktisch anwendbar ist. Die verschiedenen Experimente, einschließlich der Testläufe mit echten Teilnehmern, zeigen, dass die entwickelten Lösungen praktikabel sind und zu messbaren Verbesserungen in der Stressreduzierung und Arbeitsunterstützung beitragen. Darüber hinaus unterstützt die Verwendung von Technologien wie ROS und POMDP eine flexible und skalierbare Implementierung, die sich auf unterschiedliche Anwendungsfälle und zukünftige Erweiterungen übertragen lässt. Die Ergebnisse belegen, dass die geleistete Arbeit einen signifikanten Beitrag zur Verbesserung der Mensch-Roboter-Interaktion und der Produktionsprozesse leistet.

4 Verwertbarkeit

Die Ergebnisse des InCoRAP-Projekts bieten eine wertvolle Grundlage für die Weiterentwicklung von intelligenten Assistenzsystemen in der Produktion, indem sie fortschrittliche Technologien wie AGR-POMDP, Objekterkennung, Eyetracking und kontextbasierte digitale Zwillinge miteinander vereinen. Das AGR-POMDP-Modell ermöglicht eine effektive Planung und Entscheidungsfindung in dynamischen Umgebungen, indem es probabilistische Überlegungen und die Unsicherheit in der Wahrnehmung berücksichtigt. Die Kombination von Objekterkennung und Eyetracking liefert präzise Daten über die Interaktionen des Arbeiters mit seiner Umgebung und ermöglicht es dem System, gezielt auf die Bedürfnisse des Arbeiters zu reagieren. Der kontextbasierte digitale Zwilling ergänzt diese Technologien, indem er eine detaillierte und dynamische virtuelle Darstellung der Produktionsumgebung bereitstellt, die kontinuierlich aktualisiert wird, um aktuelle Veränderungen widerzuspiegeln. Diese integrierte Herangehensweise verbessert nicht nur die Genauigkeit der Wahrnehmung und der Reaktionsfähigkeit des Roboters, sondern auch die Anpassungsfähigkeit und Effizienz des gesamten Assistenzsystems, was zu einer signifikanten Steigerung der Produktivität und Reduzierung der Arbeitsbelastung führt.

4.1 Wirtschaftliche Erfolgsaussichten

Das entwickelte Werkerassistenzsystem und der digitale Zwilling aus dem InCoRAP-Projekt konnten bereits während des laufenden Vorhabens in anderen Projekten wie Twin4Trucks und KI4ETA (vom BMBF) gefördert) als grundlegende Technologien eingesetzt und weiterentwickelt werden. Die Kombination dieser Technologien erlaubt eine verbesserte Mensch-Roboter-Interaktion und eine optimierte Echtzeit-Überwachung von Produktionsprozessen. Durch die Weiterentwicklung des digitalen Zwillings, der sowohl kontextbasierte als auch sensorbasierte Daten integriert, kann die Fertigung von LKW-Komponenten effizienter gestaltet werden. Das Assistenzsystem ermöglicht eine präzise Unterstützung der Werker bei komplexen Montageaufgaben, was die Fehlerquote reduziert und die Produktivität steigert. So könnten die in InCoRAP erprobten Technologien direkt zur Steigerung der Produktionsflexibilität und Kosteneffizienz im Rahmen von Twin4Trucks beitragen.

Die in InCoRAP gesammelten Erfahrungen in der Augmented Reality-Integration zur Unterstützung von Werker, bei Monitoring, Maintenance und Prozessoptimierung in der Produktionsumgebung, konnten in BMWi-geförderte KI4ETA angewendet werden.

Durch die wissenschaftlichen und technischen Ergebnisse in InCoRAP konnte das Handarbeitsplatz-Modul (Eine Station für die Manuelle Assembly-Prozesse, die durch Werker durchgeführt werden können) in der SmartFactory-KL-Lab weiterentwickelt und mit state-of-the-Art Technologien wie z.B. Handtracking and Eyetracking ausgerüstet werden.

4.2 Wissenschaftliche Erfolgsaussichten

Im Laufe des Projekts hat das Team zahlreiche Studienarbeiten zu Themen wie, Human-Robot-Collaboration via Augmented Reality, Werkerassistenzsysteme und semantische Umgebungsmodellierung betreut.

Außerdem wurden die in InCoRAP erzielten Ergebnisse zum Thema Werkerassistenzsystem für das vom BMBF geförderte Projekt MILKI-PSY vorgestellt. Hierfür gab es mehrere Austauschrunden über die wissenschaftlich-technische Resultate in InCoRAP.

Darüber hinaus arbeiteten die Mitglieder des InCoRAP-Teams aktiv mit zahlreichen Forschern, Projekten und der Industrie zusammen. Neben anderen Forschungsbereichen und Transfer Labs wie SmartFactory-KL sind hier insb. Researcher/wissenschaftliche Mitarbeiterinnen der RPTU (Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau), Universität Osnabrück und University of SIBIU zu nennen.

4.3 Wissenschaftlich-wirtschaftliche Anschlussfähigkeit

Die in InCoRAP aufgebaute Kompetenz und die Fähigkeit des Systems könnte durch ein Foundational Model verbessert werden, um komplexe menschliche Intentionen in verschiedenen Szenarien zu erkennen und zu verstehen. Durch die Verwendung eines vortrainierten Foundational Models, das auf einer großen Menge an Daten basiert, kann InCoRAP eine genauere und kontextbewusstere Interpretation von Sensordaten und menschlichen Handlungen erzielen. Dies würde die Interaktion zwischen Mensch und Roboter optimieren, indem es dem Roboter ermöglicht, schneller auf die Bedürfnisse des Werkers zu reagieren und seine Aktionen besser an den aktuellen Kontext anzupassen. Darüber hinaus könnte es helfen, die Notwendigkeit umfangreicher Trainingsdaten zu reduzieren und das System an eine Vielzahl neuer Situationen und Aufgaben anzupassen.

Außerdem kann man die in InCoRAP entwickelte Grundlage für die Unterstützung des Werkers und Human-Robot-Collaboration als Basis mit einem Multimodal Language Model kombinieren, um eine verbesserte Intentionserkennung, kontextuelle Entscheidungsfindung, interaktive Kommunikation und Assistenz sowie Echtzeit-Überwachung und Fehlererkennung zu realisieren. Um weitere Forschungsprojekte in dem Bereich voranzutreiben, kann die dargestellte Forschungsfrage eine interessante und innovative Ausgangsbasis bieten.

5 Bekannt gewordener Fortschritt

Während der Durchführung des InCoRAP-Vorhabens wurden dem Zuwendungsempfänger (ZE) bedeutende Fortschritte auf dem Gebiet der Werkerassistenz und der Robotik, insbesondere in der Zusammenarbeit von Mensch und Maschine, bekannt. Zahlreiche Fortschritte wurden sowohl in der kollaborativen Robotik als auch in der intelligenten Produktionsunterstützung erzielt, die auf anderen Projekten und Forschungsinitiativen basierten. Dazu zählen neue Entwicklungen im Bereich der Sensordatenerfassung durch Wearables wie Eye-Tracking und die Nutzung von Augmented Reality (AR) zur Unterstützung der Montageprozesse.

Ausserdem hat InCoRAP auch bedeutende Fortschritte im Bereich der Objekterkennung erzielt, insbesondere durch den Einsatz von synthetischen und realen Datensätzen für industrielle Anwendungen.

Zusätzlich wurde auf dem Gebiet der Künstlichen Intelligenz zur Intentionserkennung (Intention Recognition, IR) erhebliche Fortschritte erzielt, wie es in anderen Forschungsprojekten im Bereich des maschinellen Lernens beobachtet wurde. Speziell die Verknüpfung von Hand-Tracking, Blickverfolgung und objekterkennenden Algorithmen hat neue Möglichkeiten eröffnet, um die Effizienz von Produktionsprozessen durch präzisere Aufgabenmodellierung und Robotersteuerung zu erhöhen. Dies führte auch zu einem besseren Verständnis und zur Implementierung von digitalen Zwillingen, die als zentraler Bestandteil des InCoRAP-Projekts genutzt wurden. Der Transfer von Technologien aus angrenzenden Forschungsbereichen in das Projekt förderte eine starke Interdisziplinarität und sorgte für wichtige Synergien im Forschungsfortschritt.

6 Veröffentlichungen

- [BL23] Jonas Brozeit and Max Lauer. Design und Integration zusätzlicher semantischer Annotationen in CAD-Modellen und ihre Nutzung in einem semantischen Umgebungsmodell. Technical report, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU), July 2023. Supervisors: Parsha Pahlevannejad and Michael Sintek.
- [LGS+23a] Oscar Lima, Martin Günther, Alexander Sung, Sebastian Stock, Marc Vinci, Amos Smith, Jan Christoph Krause, and Joachim Hertzberg. Mobipick Labs System Demonstration. In *The 33rd International Conference on Automated Planning and Scheduling*, 2023.
- [LGS+23b] Oscar Lima, Martin Günther, Alexander Sung, Sebastian Stock, Marc Vinci, Amos Smith, Jan Christoph Krause, and Joachim Hertzberg. A Physics-Based Simulated Robotics Testbed for Planning and Acting Research. In *ICAPS Workshop on Planning and Robotics (PlanRob 2023)*, 2023.
- [MABM+ 22] Andrea Micheli, Alexandre Arnold, Arthur Bit-Monnot, Luigi Bonassi, Luca Framba, Alfonso Emilio Gerevini, Selvakumar Hastham Sathiya Satchi, Malte Helmert, Felix Ingrand, Luca Iocchi, et al. Unified Planning: A Python Library Making Planning Technology Accessible. In *International Conference on Automated Planning and Scheduling System Demonstrations, 2022, 2022*.
- [PMW+ 23] Stefan-Alexandru Precup, Alexandru Matei, Snehal Walunj, Arpad Gellert, Christiane Plociennik, and Constantin-Bala Zamfirescu. Collaborative exploitation of various AI methods in adaptive assembly assistance systems. In *(to appear) 10th Information Technology and Quantitative Management (ITQM 2023)*, Oxford, United Kingdom, August 2023. Elsevier, 2023.
- [PWG+23] Stefan-Alexandru Precup, Snehal Walunj, Arpad Gellert, Christiane Plociennik, Jibinraj Antony, and Martin Ruskowski. Recognising Worker Intentions by Assembly Step Prediction. In *(to appear) 28th IEEE International Conference on Emerging Technologies and Factory Automation (IEEE ETFA 2023)*, Sinaia, Romania, September 2023, 2023.
- [RHH+20] Martin Ruskowski, Arnold Herget, Jesko Hermann, William Motsch, Parsha Pahlevannejad, Aleksandr Sidorenko, Simon Bergweiler, Alexander David, Christiane Plociennik, Jens Popper, Keran Sivalingam, and Achim Wagner. Production Bots für Production Level 4: Skill-basierte Systeme für die Produktion der Zukunft. *atp magazin*, 62:62–71, September 2020.
- [SH23] Juan Carlos Saborio and Joachim Hertzberg. Towards Intention Recognition for Robotic Assistants Through Online POMDP Planning. In *(to appear) ICAPS 2023 Workshop on Plan, Activity, and Intent Recognition (PAIR)*, 2023.
- [SSS+ 23] Selvakumar Hastham Sathiya Satchi Sadanandam, Sebastian Stock, Alexander Sung, Felix Ingrand, Oscar Lima, Marc Vinci, and Joachim Hertzberg. A Closed-Loop Framework-Independent Bridge from AIPlan4EU's Unified Planning Platform to Embedded Systems. In *33rd International Conference on Automated Planning and Scheduling (ICAPS 2023) PlanRob Workshop*, 2023.
- [TWP+21] Hooman Tavakoli, Snehal Walunj, Parsha Pahlevannejad, Christiane Plociennik, and Martin Ruskowski. Small Object Detection for Near Real-Time Egocentric Perception in a Manual Assembly Scenario. In *The E!8!!! International Workshop on Egocentric Perception, Interaction and Computing (EPICS CVPR2021)*, June 2021.
- [WPS+ 24] Snehal Walunj, Parsha Pahlevannejad, Michael Sintek, Juan Carlos Saborio Saborio, Prema Garg, Hooman Tavakoli, Christiane Plociennik, Nicolas Grossmann, Ansgar Bernardi, Martin Ruskowski, Joachim Hertzberg, and Andreas Dengel. Context-aware Robotic Assistance for Workers using Intention Recognition and Semantic Digital Twin. In Jaap van den Herik and Ana Paula Rocha, editors, *Lecture Notes on Computer Science (LNCS)*, Cham, 2024. Springer International Publishing. [To be published].
- [WSP+ 23] Snehal Walunj, Michael Sintek, Parsha Pahlevannejad, Christiane Plociennik, and Martin Ruskowski. Ontology-based Digital Twin Framework for Smart Factories. In *(to appear) 31st International Conference on Information Systems Development (ISD 2023)*, Lisbon, Portugal, 30 August to 1 September 2023, 2023.

7 Abbildungsverzeichnis

Abbildung 1: Gantt-Diagramm der geplanten Arbeit im Projekt.....	4
Abbildung 2 - InCoRAP-Systemarchitektur und die Interaktion des Umgebungsmodells mit anderen Systemkomponenten.....	5
Abbildung 3 - Füllen des Umgebungsmodells aus statischen und dynamischen Quellen über ROS	7
Abbildung 4 - Situation Specific Zoom Levels (Beispiele)	9
Abbildung 5 - Two step hierarchical detection using YOLOv4.	12
Abbildung 6 - Die Pipeline demonstriert den sequentiellen Prozess der Generierung synthetischer Daten, der Durchführung von Training und Tests mithilfe der Hololens2 World Camera mit dem Ziel, eine kontinuierliche Lernpipeline für Modellaktualisierungen einzurichten.....	12
Abbildung 7 - Objekterkennung trainiert und getestet durch YoloV5 für Bauteile des "Manual Assembly" Szenarios	13
Abbildung 8 - Verschiedene Abstraktionsebenen zur Absichtserkennung.....	17
Abbildung 9 - Funktionale Ansicht des Active Goal Recognition-Frameworks im Szenario „Manual Assembly“	19
Abbildung 10 Architektur des Planungs- und Ausführungsrahmens	23
Abbildung 11 - Kommunikationsdiagramm der Softwarekomponenten zur Erkennung von Anomalien und des Arbeitsassistenten	27
Abbildung 12 - Miniatur-LKW und seine Montageteile, maßgeschneidert für das Szenario der manuellen Montage	28

8 Tabellenverzeichnis

Tabelle 1 - Modell wurde anhand realer Daten trainiert und anhand realer Daten getestet	14
Tabelle 2 - Modell wurde anhand synthetischer Daten trainiert und anhand realer Daten getestet	14
Tabelle 3 - Modell wurde zuerst mit realer Daten und dann synthetischer Daten trainiert und anhand realer Daten getestet	15
Tabelle 4 - Modell wurde zuerst mit synthetischer Daten und dann realer Daten trainiert und anhand realer Daten getestet	15
Tabelle 5 Positionen des zahlenmäßigen Verwendungsnachweises	30

9 Abkürzungsverzeichnis

3MF: 3D Manufacturing Format

AGR: Active Goal Recognition

AGV: Automated Guided Vehicle

AR: Augmented Reality

BOT: Building Topology Ontology

CAD: Computer-aided design

CAE: Computer-aided engineering

CAM: Computer-aided manufacturing

CHIMP: Conflict-driven Hierarchical Meta-CSP Planner

DOPE: Deep Object Pose Estimation

ERP: Enterprise Resource Planning

Gazebo: Robot simulation tool

g4a: game4automation, now realvirtual.io

HDDL: Hierarchical Domain Definition Language

IR: Intention Recognition

Kafka: Event streaming platform

Mobipick: Mobile manipulator with mobile base and robot arm

Movelt!: ROS framework for motion planning and manipulation

MQTT: Message Queuing Telemetry Transport

OD: Object Detection

OPCUA: OPC Unified Architecture

OWL: Web Ontology Language

POMDP: Partially Observable Markov Decision Process

RDF: Resource Description Framework

REST: Representational State Transfer

Robot API: Interface layer between plan execution and motion control

ROS: Robot Operating System

ROS#: ROS Sharp, C#/.NET support for ROS

Siemens NX: CAD/CAM/CAE system

SPARQL: SPARQL Protocol and RDF Query Language

Unity: 3D Game Engine

URDF: Unified Robot Description Format

10 Literaturverzeichnis

- [AB19] Christopher Amato and Andrea Baisero. Active goal recognition. *arXiv preprint arXiv:1909.11173*, 2019.
- [BM23] Arthur Bit-Monnot. Experimenting with Lifted Plan-Space Planning as Scheduling: Aries in the 2023 IPC. In *2023 International Planning Competition at the 33rd International Conference on Automated Planning and Scheduling*, 2023.
- [Bro96] John Brooke. Sus: a “quick and dirty” usability. *Usability evaluation in industry*, 189(3):189–194, 1996.
- [CFL+15] Michael Cashmore, Maria Fox, Derek Long, Daniele Magazzeni, Bram Ridder, Arnau Carrera, Narcis Palomeras, Natalia Hurtos, and Marc Carreras. Rosplan: Planning in the robot operating system. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 25, 2015.
- [CSCC14] David Coleman, Ioan Sucan, Sachin Chitta, and Nikolaus Correll. Reducing the Barrier to Entry of Complex Robotic Software: a MoveIt! Case Study, 2014.
- [ERAB18] Mohamed Elfeki, Krishna Regmi, Shervin Ardeshtir, and Ali Boji. From third person to first person: Dataset and baselines for synthesis and retrieval. *arXiv preprint arXiv:1812.00104*, 2018.
- [FZ17] Richard Freedman and Shlomo Zilberstein. Integration of planning with recognition for responsive interaction using classical planners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.
- [FZ19] Richard G Freedman and Shlomo Zilberstein. A unifying perspective of plan, activity, and intent recognition. In *Proceedings of the AAAI Workshops: Plan, Activity, Intent Recognition (Honolulu, HI)*, pages 1–8, 2019.
- [Geil5] Christopher Geib. Lexicalized reasoning. In *Proceedings of the 3rd Annual Conference on Advances in Cognitive Systems ACS*, page 19, 2015.
- [GGSGM+20] Andrija Gajic, Ester Gonzalez-Sosa, Diego Gonzalez-Morin, Marcos Escudero-Vinolo, and Alvaro Villegas. Egocentric human segmentation for mixed reality. *arXiv preprint arXiv:2005.12074*, 2020.
- [GHRZ19] Michael Görner, Robert Haschke, Helge Ritter, and Jianwei Zhang. MoveIt! Task Constructor for Task-Level Motion Planning. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 190–196, 2019.
- [GPPZ20] Arpad Gellert, Stefan-Alexandra Precup, Bogdan-Constantin Pirvu, and Konstantin-Bala Zamfirescu. Prediction-Based Assembly Assistance System. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, volume 1, pages 1065-1068, 2020.
- [MA02] Alexander M. Meystel and James S. Albus. *Intelligent systems: architecture, design, and control*. Wiley Interscience, 2002.
- [MB21] Jean Massardi and Éric Beaudry. Active Goal Recognition using Intention Aware Motion Planning. In *Plan, Activity and Intent Recognition (PAIR) 2021*, 2021.
- [PGM+22] Stefan-Alexandru Precup, Arpad Gellert, Alexandru Matei, Maria Gita, and Constantin-Bala Zamfirescu. Towards an Assembly Support System with Dynamic Bayesian Network. *Applied Sciences*, 12(3), 2022.
- [RGIO] Miguel Ramfrez and Hector Geffner. Probabilistic plan recognition using off-the-shelf classical planners. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence (AAAI 2010)*, pages 1121–1126. Citeseer, 2010.
- [RGII] Miguel Ramirez and Hector Geffner. Goal recognition over POMDPs: Inferring the intention of a POMDP agent. In *IJCAI*, pages 2009–2014. IJCAI/AAAI, 2011.
- [RLSP20] Mads Holten Rasmussen, Maxime Lefrançois, Georg Schneider, and Pieter Pauwels. BOT: the Building Topology Ontology of the W3C Linked Building Data Group. *Semantic Web*, November 2020.
- [SGB+14] Gita Sukthankar, Christopher Geib, Hung Bui, David Pynadath, and Robert P Goldman. *Plan, activity, and*

intent recognition: Theory and practice. Newnes, 2014.

- [SH19a] Juan Carlos Saborio and Joachim Hertzberg. Efficient Planning Under Uncertainty with Incremental Refinement. In *Proc. of the 35th Conference on Uncertainty in Artificial Intelligence, Tel Aviv, Israel, July 22-25, 2019, UAI'19*, page 112, 2019.
- [SH19b] Juan Carlos Saborio and Joachim Hertzberg. Planning Under Uncertainty Through Goal-Driven Action Selection. In Jaap van den Herik and Ana Paula Rocha, editors, *Agents and Artificial Intelligence*, pages 182—201, Cham, 2019. Springer International Publishing.
- [SRD+ 18] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. Introducing WESAD, a Multi-modal Dataset for Wearable Stress and Affect Detection. pages 400—408, 10 2018.
- [SRU16] Shirin Sohrabi, Anton V Riabov, and Octavian Udrea. Plan Recognition as Planning Revisited. In *IJCAI*, pages 3258—3264. New York, NY, 2016.
- [SVIO] David Silver and Joel Veness. Monte-Carlo Planning in Large POMDPs. In *Advances in Neural Information Processing Systems 23*, pages 2164—2172, 2010.
- [TTS+18] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. Deep object pose estimation for semantic robotic grasping of household objects. *arXiv preprint arXiv:1809.10790*, 2018.