Sachbericht zum Verwertungsnachweis

Teil 2: Ausführlicher Bericht

Vorhabenbezeichnung:

VASCO: KI-basiertes Visuelles ASsistenzsystem zur Computergestützten Orientierungs- und Handlungsfähigkeit für Menschen mit Seheinschränkungen

Teilvorhaben:

Computer Vision, 3D Rekonstruktion, Hardware und Software von VASCO

Autoren:

Frederik Lange, Chao-Yang Fu, Oleksandr Karpenko, Maximilian Singh, Jia-Wei Wang

Berichtsdatum: 04.08.2026

Zuwendungsempfänger:	**Förderkennzeichen:**
DC Vision Systems GmbH Rollnerstr. 59 90408 Nürnberg	13N16532 VASCO
Berichtszeitraum: von: 01.01.2023 bis: 30.06.2026	

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) unter dem Förderkennzeichen 13N16532 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Gefördert durch:

Bundesministerium für Forschung, Technologie und Raumfahrt

Inhaltsverzeichnis

1 Übersicht 3
1.1 Projektbeschreibung und Zielsetzung 3
1.2 Ziele des Gesamtvorhabens 4
1.3 Ziele des Teilvorhabens 5
2 Ergebnisse des Teilvorhabens 6
2.1 Übersicht 6
2.2 Systemarchitektur 7
2.3 Hardware: Feedback und taktiles Display 8
2.3.1 Tragbarer Computer 8
2.3.2 Headset und taktiles Display 10
2.4 Bildverarbeitung und Umfeldmodell 10
2.4.1 Infrastruktur der Bildverarbeitung 10
2.4.2 Tracking, Lokalisierung 11
2.4.3 3D Rekonstruktion 12
2.4.4 Semantische Klassifikation 13
2.4.5 Erkennung und Tracken von Objekten 14
2.4.6 Umfeldmodell 15
2.4.7 Visualisierung des Umfeldmodells 15
2.5 Systemsoftware und Systemintegration 16
2.5.1 Integration des Basis-Systems und Inbetriebnahme der Gesamt-Applikation 16
2.5.2 Operator Schnittstelle 16
2.5.3 Integration des taktilen Armbandcontrollers 17
2.6 Realisierung von Use-cases 18
2.6.1 Greifen und Interaktion 18
2.6.2 Kollisionsvermeidung und 3D-Freiraumerkennung 19
2.6.3 Outdoor-Navigation 20
3 Verwendung der Zuwendung 23
3.1 Darstellung und Begründung der wichtigsten Positionen des zahlenmäßigen Nachweises 23
3.2 Notwendigkeit und Angemessenheit der geleisteten Arbeit 23
3.3 Darstellung des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit der Ergebnisse im Sinne des fortgeschriebenen Verwertungsplans 23
3.4 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen 24
3.5 Veröffentlichungen der Ergebnisse 24
4 Anhang 25

Abbildungsverzeichnis

Abbildung 1: Person mit Langstock in Alltagssituationen (Quelle: BFW Düren) 3
Abbildung 2: Übersicht zu VASCO 4
Abbildung 3: Funktionale Architektur von VASCO 5
Abbildung 4: Blockdiagramm des Gesamtsystems 7
Abbildung 5: Hardwarediagramm des tragbaren Computers 9
Abbildung 6: Links: Aufgebautes Gehäuse des tragbaren Computers, rechts: Mainboard mit GPU, FPGA, SSD und KI-Beschleuniger-Modulen 9
Abbildung 7: Architektur der Bilderfassung und -Verarbeitung 10
Abbildung 8: Slam Kamerapfad und Punktwolke einer Straßenszene 11
Abbildung 9: Relokalisierung in rekonstruierter Umgebung 12
Abbildung 10: Live 3D Rekonstruktion und Visualisierung 12
Abbildung 11: Semantische Klassifikation basierend auf Prompts 13
Abbildung 12: Semantische Klassifikation auf Voxel-Ebene 14
Abbildung 13: 3D Objekterkennung und Tracking. Links: Projektion ins Kamera-Bild. Rechts: Lokalisierung im 3D Umfeldmodell 14
Abbildung 14: Architektur der Verbindung zum Operator 17
Abbildung 15: Dashboard zur Anzeige der Sensordaten für den Operator 17
Abbildung 16: Verarbeitungskette beim Use-case “Greifen und Interaktion” 18
Abbildung 17: Links: Test-setup bei „Greifen und Interaktion“, Rechts: Visualisierung des Umfeldmodells/Dashboard 19
Abbildung 18: Schwierige Szenarien mit Potential für Kollisionen (Quelle: BFW Düren) 19
Abbildung 19: Prinzip der Kollisionserkennung 20
Abbildung 20: Kollisionserkennung auf realen Daten im Innenraum (links) und Außenbereich (rechts) 20
Abbildung 21: Repräsentation der Oudoor-Szene im Umeldmodell 21
Abbildung 22: Aufteilung eines Pfades in Wegpunkte 22
Abbildung 23: Projektion des Umfeldmodells mit Pfadvorgabe in das Kamerabild 22

1 Übersicht

1.1 Projektbeschreibung und Zielsetzung

Die visuelle Wahrnehmung ist für den Menschen der wichtigste Sinn, um die Umwelt zu verstehen bzw. zu interpretieren. Daher ist es für Menschen mit Seheinschränkungen sehr schwer bis unmöglich, am alltäglichen Leben vollumfänglich teilzuhaben. Ohne die Möglichkeit, die Umgebung in Echtzeit visuell zu erfassen, ist eine Interaktion mit dieser nur sehr begrenzt möglich. Beim Gehen ist es beispielsweise neben der Orientierung auch wesentlich, begehbare Bereiche und Hindernisse zu erkennen. Jegliche Interaktion der Hand mit Objekten erfordert, die Beschaffenheit, Position und Lage der Objekte zu kennen. Hier spielt die Hand-Auge-Koordination eine wesentliche Rolle.

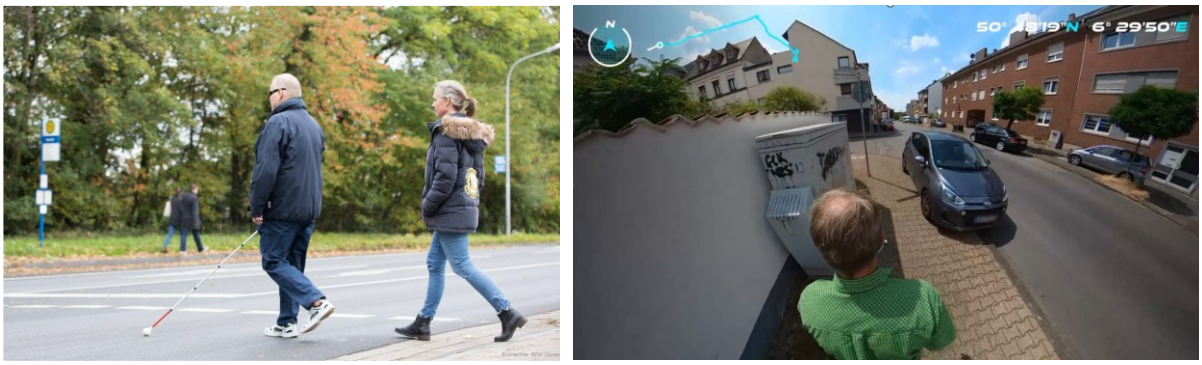

Abbildung 1: Person mit Langstock in Alltagssituationen (Quelle: BFW Düren)

In diesem Projekt wird ein tragbares Assistenzsystem entwickelt, das Menschen mit Seheinschränkungen unterstützt. Wie Abbildung 2 zeigt, soll das Assistenzsystem am Kopf getragen werden und den Anwender in vielfältigen, für das tägliche Leben relevanten Situationen behilflich sein. In einem Headset ist ein 3D-Kamerasystem zur Umfelderkennung integriert. Die Rückmeldung an den Anwender soll insbesondere über ein taktiles Display sowie akustisch über einen Kopfhörer erfolgen. Das taktile Display wird in das Headset integriert und übermittelt dem Anwender über die Stirn in Echtzeit gut differenzierbare, situationsbedingte Informationen. Die Datenverarbeitung und Stromversorgung erfolgt über einen kompakten, tragbaren Rechner.

Die Entwicklung eines tragbaren Assistenzsystems zur Unterstützung von Menschen mit Seheinschränkungen stellt hohe Anforderungen an die eingesetzte Sensorik, die Datenverarbeitung sowie die Integration der verschiedenen Hardware- und Softwarekomponenten des Gesamtsystems. Im Gegensatz zu stationären oder fahrzeuggebundenen Anwendungen müssen sämtliche Komponenten auf einer mobilen und energieeffizienten Plattform betrieben werden, ohne dabei Einschränkungen hinsichtlich der Funktionalität oder der Echtzeitfähigkeit der Algorithmen zu verursachen.

Für die Unterstützung des Anwenders in komplexen und dynamischen Alltagssituationen ist eine kontinuierliche Erfassung und Analyse der unmittelbaren Umgebung erforderlich. Das System muss dabei sowohl statische als auch dynamische Objekte zuverlässig erkennen und deren räumliche Beziehung zum Anwender in Echtzeit bestimmen. Gleichzeitig müssen die erkannten Informationen in einer Form bereitgestellt werden, die eine direkte Integration in die Handlungen und Bewegungen des Anwenders ermöglicht. Daraus ergeben sich hohe Anforderungen an die Latenz der gesamten Verarbeitungskette vom Sensor bis zur Ausgabe der Informationen. Eine wesentliche Voraussetzung für die Durchführung des Vorhabens war daher die Verfügbarkeit einer leistungsfähigen und gleichzeitig energieeffizienten Rechenplattform, die eine Ausführung komplexer Verfahren der Bildverarbeitung und künstlichen Intelligenz auf einem tragbaren System ermöglicht. Insbesondere die Erzeugung von 3D-Informationen aus mehreren Kamerabildern, die semantische Analyse der Umgebung sowie die Szenenanalyse erfordern eine hohe Rechenleistung bei gleichzeitig begrenzter

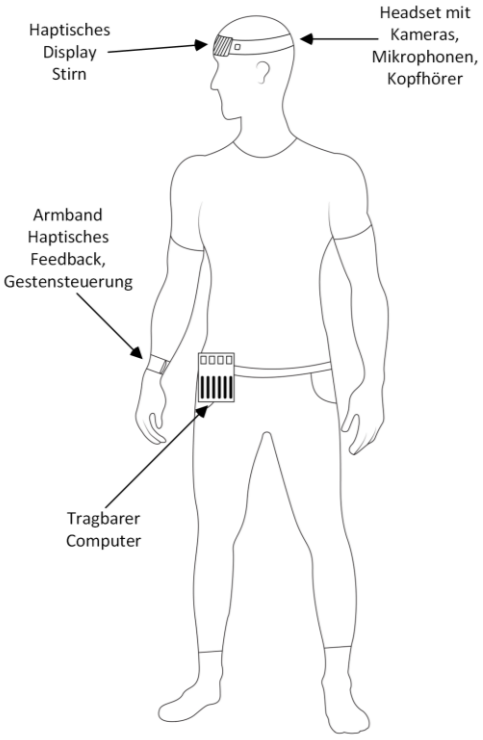

Abbildung 2: Übersicht zu VASCO

Leistungsaufnahme und Baugröße. Darüber hinaus setzt das Vorhaben die enge Zusammenarbeit der verschiedenen Projektpartner voraus, da die einzelnen Teilvorhaben aufeinander aufbauen und über definierte Schnittstellen miteinander verbunden sind. Während im Teilvorhaben der DC Vision Systems GmbH insbesondere die Erfassung und Analyse des Umfeldes sowie die Integration der Softwaremodule adressiert werden, liefern andere Teilvorhaben Beiträge im Bereich der Mensch-Maschine-Interaktion, der taktilen und akustischen Ausgabe sowie der Benutzerstudien und Evaluationen. Die Definition gemeinsamer Datenstrukturen, Kommunikationsschnittstellen und Integrationsmechanismen stellt daher eine wesentliche Voraussetzung für die erfolgreiche Umsetzung des Gesamtsystems dar.
Die Arbeiten des Teilvorhabens basieren auf einer mehrkamerabasierten Erfassung der Umgebung, bei der die aufgenommenen Bilddaten zunächst vorverarbeitet und anschließend zur Erzeugung von 3D-Informationen und semantischen Beschreibungen der Szene genutzt werden. Die resultierenden Informationen werden in einem gemeinsamen Umfeldmodell zusammengeführt, welches als zentrale Informationsbasis für die darauf basierenden Funktionen des Assistenzsystems dient. Die Durchführung des Vorhabens erfolgte im Rahmen der Förderinitiative „Photonik Forschung Deutschland“ im Themenfeld „Photonik für die digital vernetzte Welt“. Das Vorhaben adressiert dabei insbesondere die Bereiche der optischen Sensorik, der schnellen und energieeffizienten Datenauswertung sowie der Nutzung der gewonnenen Informationen zur Unterstützung von Anwendern in komplexen und dynamischen Umgebungen und entspricht damit unmittelbar den Zielsetzungen der Fördermaßnahme.

1.2 Ziele des Gesamtvorhabens

Zentraler Bestandteil des Assistenzsystems ist der Benutzer, der mithilfe der entwickelten, technisch intelligenten Plattform gezielte Wechselwirkungen mit seiner Umwelt ausführen kann. Das Funktionsprinzip und die Datenflüsse sind in Abbildung 3 dargestellt.
Die Bilddaten werden kontinuierlich mit den Kameras aufgenommen und an den tragbaren Rechner übertragen. Auf dem Rechner werden sie durch Verfahren des maschinellen Sehens („Computer Vision“) und der KI ausgewertet, um ein Umfeldmodell in Echtzeit zu erzeugen. Auf Basis dieses Umfeldmodells werden Situationen erkannt und interpretiert. Abhängig vom Arbeitsmodus und der Systemkonfiguration werden dem Anwender gezielte Rückmeldungen gegeben. Das System wird technisch von dem Modul „KI, Aufgabenplanung, zentraler Zustandsautomat” gesteuert. Dieses Modul gibt letztendlich vor, was das System tun soll und welche Rückmeldung an den Anwender gegeben wird. Die wichtigsten Funktionen sind im Folgenden beschrieben. Das Gesamtsystem realisiert folgende Use-cases:
Nahbereichs-Navigation und Kollisionsvermeidung: Mit Hilfe dieser Funktion können sich Anwender sicher in bekannten und unbekannten Umgebungen bewegen. Zu diesem Zweck erstellt das System ein dreidimensionales Modell der unmittelbaren Umgebung des Anwenders. Dabei werden bewegliche und unbewegliche Objekte klassifiziert. Aus der Bewegung des Anwenders und der Objekte werden potenzielle Kollisionen vorhergesagt.
Über das taktile Display soll der Anwender über begehbare Bereiche sowie über potenzielle Kollisionen informiert werden. Höhenunterschiede wie Treppenstufen, Bordsteine etc. sollen durch eingängige Symbolik übermittelt werden. Zu diesem Zweck werden die Informationen

für das taktile Display aus der Geometrie der Szene und einer semantischen Analyse sinnvoll aufbereitet. Diese Funktion soll stets aktiv sein.

Mobilität und Straßenverkehr: Die Funktion zielt darauf ab, dem Anwender Mobilität zu ermöglichen, also eine Orientierung und sichere Bewegung im öffentlichen Raum und im Straßenverkehr. Der Anwender gibt dem System ein Ziel für die Navigation vor. Unter Verwendung verfügbarer und gängiger Navigations-APIs [1] ermittelt das System unter Berücksichtigung der unmittelbaren Umgebung und der sich darin bewegenden Objekte einen sicheren Pfad zu dem gewünschten Ziel. Das Assistenzsystem erkennt und klassifiziert die für den Anwender in der spezifischen Situation wesentlichen Merkmale bzw. Elemente des Straßenverkehrs und übermittelt diese Informationen an den Anwender. Diese Funktionalität basiert auf Nahbereichs-Navigation und Kollisionsvermeidung und erkennt beispielsweise Bürgersteige, Straßen und Ampeln mit Signalzustand als statische Objekte. Als dynamische Objekte werden Fahrzeuge, Fußgänger und weitere Verkehrsteilnehmer erkannt.

Greifen und Interaktion: Das Assistenzsystem unterstützt den Anwender bei der Bewegung der Arme. Sobald die Bildverarbeitung ein Objekt lokalisiert hat und der Anwender dem Assistenzsystem signalisiert, dass er das Objekt greifen möchte, unterstützt das System ihn bei der Greifbewegung. Dazu lokalisiert die Bildverarbeitung die Hand des Anwenders in den Kamerabildern relativ zum Objekt. Über das taktile Feedback wird eine Richtungs-/Bewegungskorrektur übermittelt. Das Greifen und die Interaktion sind bei vielfältigen Aufgaben des täglichen Lebens wichtig. Die dafür erforderlichen grundlegenden Mechanismen sind immer gleich. Das System soll auch potenzielle Kollisionen des Arms mit Gegenständen (Gläser auf einem Tisch etc.) erkennen und den Anwender davor warnen.

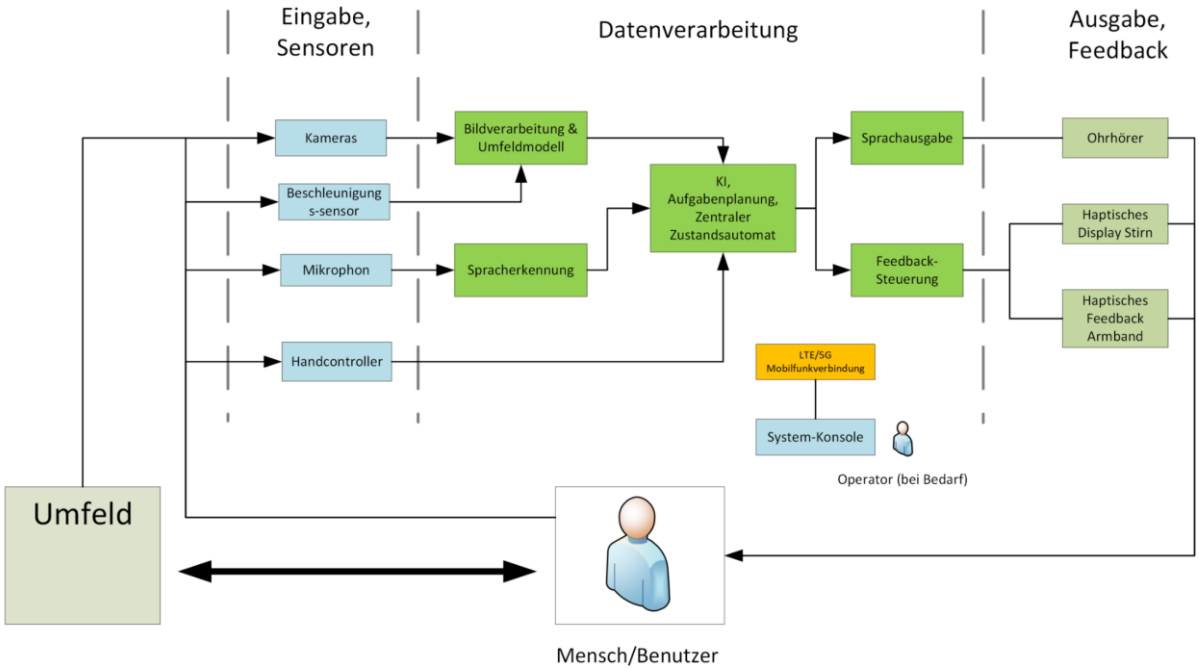

Abbildung 3: Funktionale Architektur von VASCO

1.3 Ziele des Teilvorhabens

Dieses Teilvorhaben hat folgende, die für die Realisierung des Gesamtvorhabens wesentliche Ziele:

Bildverarbeitung, Umfeldmodell: Aus den Bilddaten der Kameras sollen die für das Projekt wichtigen Informationen über das Umfeld in Echtzeit mit geringer Latenz extrahiert werden. Die Bildverarbeitung und die Algorithmen sollen mit möglichst geringer Latenz implementiert werden, um den Anwender in Echtzeit unterstützen zu können. Insbesondere sollen potenzielle Kollisionen frühzeitig erkannt werden, damit dem Anwender rechtzeitig vor einer Gefährdung geeignete Warnhinweise bereitgestellt werden können. Das erzeugte Umfeldmodell soll die für die jeweiligen Assistenzfunktionen relevante Umgebung zuverlässig

erfassen und dabei sowohl den unmittelbaren Sichtbereich als auch bereits zuvor erfasste Bereiche der Umgebung berücksichtigen. Dadurch wird eine kontinuierliche und möglichst vollständige räumliche Beschreibung der Umgebung bereitgestellt, die als Grundlage für Navigation, Kollisionsvermeidung sowie Greifen und Interaktion dient.
Für den Anwendungsfall des Greifens sollen relevante Alltagsobjekte wie Flaschen, Gläser, Geschirr, Besteck oder Lebensmittelverpackungen zuverlässig erkannt und räumlich lokalisiert werden, um eine sichere Unterstützung des Anwenders während der Interaktion mit diesen Objekten zu ermöglichen. Für den Anwendungsfall der Nahbereichs-Navigation und Kollisionsvermeidung sollen begehbare Bereiche sowie relevante statische und dynamische Hindernisse wie Wände, Türen, Personen oder weitere Objekte sicher detektiert und im dreidimensionalen Umfeldmodell repräsentiert werden. Hierzu muss das Umfeldmodell die räumliche Struktur der Umgebung mit ausreichender Genauigkeit und Aktualität erfassen, um eine robuste Kollisionsvermeidung, lokale Navigation sowie die weiteren Assistenzfunktionen zu ermöglichen.
Tragbarer Computer: Dieser muss leistungsfähig genug sein, um die Bildverarbeitung mit der KI und aller weiteren Funktionen in Echtzeit ausführen zu können. Trotzdem muss der Rechner klein, leicht und energiesparend sein. Der tragbare Computer soll eine leistungsfähige und gleichzeitig energieeffiziente Rechenplattform für den mobilen Einsatz bereitstellen. Hierzu müssen sämtliche Verfahren der Bildverarbeitung, der künstlichen Intelligenz sowie die weiteren Assistenzfunktionen in Echtzeit auf einer kompakten, tragbaren Hardware ausgeführt werden können. Gleichzeitig sind eine ausreichend lange Betriebsdauer im Akkubetrieb sowie eine ergonomische Bauform mit geringem Gewicht und hoher Alltagstauglichkeit sicherzustellen.
Anbindung eines Operators an das Assistenzsystem: Der Operator soll bei Bedarf von dem Anwender zuschaltbar sein (siehe Abbildung 3). Schwierige Situationen sollen mit Hilfe des Operators gemeistert werden können. Der Anwender kann dafür mit dem Operator einen Dialog führen. Zusätzlich hat der Operator Zugriff auf die Bilddaten der Kameras und das Umfeldmodell. Hierbei wäre auch denkbar, die bereits vorhandene Operatoren App “BeMyEyes [2]” einzubinden, welche in der Zielgruppe eine hohe Akzeptanz besitzt. Ebenso denkbar ist die Einbindung der verbreiteten Microsoft App “Seeing AI [3]”.

2 Ergebnisse des Teilvorhabens

Dieser Schlussbericht dokumentiert ausschließlich die im Rahmen des Teilvorhabens der DC Vision Systems GmbH durchgeführten Entwicklungsarbeiten und deren Ergebnisse. Beiträge der Verbundpartner werden nur insoweit beschrieben, wie dies zum Verständnis der Gesamtarchitektur, der definierten Schnittstellen sowie der Integration des Gesamtsystems erforderlich ist. Die detaillierte Darstellung der jeweiligen Entwicklungen ist den Schlussberichten der entsprechenden Verbundpartner zu entnehmen.

2.1 Übersicht

Im Folgenden werden die wesentlichen Ergebnisse des Teilvorhabens zusammengefasst. Der Schwerpunkt der Arbeiten der DC Vision Systems GmbH lag auf der Entwicklung der Bildverarbeitung, der Erzeugung eines dreidimensionalen Umfeldmodells sowie der Hard- und Softwareintegration des tragbaren Assistenzsystems. Aufbauend auf diesen Technologien wurden die im Projekt definierten Assistenzfunktionen und Use-Cases schrittweise umgesetzt und validiert.
Im Rahmen des Projekts konnten wesentliche wissenschaftlich-technische Ergebnisse erzielt werden. Hierzu zählen insbesondere die Entwicklung eines leistungsfähigen SLAM-Verfahrens, einer echtzeitfähigen 3D-Rekonstruktion sowie eines kontinuierlich aktualisierten dreidimensionalen Umfeldmodells als zentrale Informationsbasis des Assistenzsystems. Aufbauend darauf wurden Verfahren zur semantischen Szenenanalyse, zur Erkennung und Verfolgung statischer und dynamischer Objekte sowie zur Kollisionsbewertung entwickelt und in eine gemeinsame Softwarearchitektur integriert.
Neben den Verfahren der Computer Vision wurden die Systemsoftware, die Hardwareplattform des tragbaren Computers sowie die erforderlichen Schnittstellen zwischen Sensorik,

Bildverarbeitung und den Ausgabekomponenten entwickelt und zu einem funktionsfähigen Gesamtsystem zusammengeführt. Ergänzend entstand eine leistungsfähige Visualisierung des Umfeldmodells, die während der Entwicklung, Integration und Validierung sämtlicher Systemkomponenten eingesetzt wurde.
Für den Use-Case „Greifen und Interaktion“ wurde eine kontinuierliche Unterstützung des Anwenders beim Erreichen und Greifen von Objekten realisiert. Für den Use-Case „Nahbereichs-Navigation und Kollisionsvermeidung“ wurden robuste Verfahren zur Erkennung freier Bewegungsräume sowie potenzieller Kollisionen im Innen- und Außenbereich entwickelt. Darüber hinaus konnten für den Use-Case „Mobilität und Straßenverkehr“ die wesentlichen Verfahren zur Umfeldwahrnehmung, Szenenanalyse und lokalen Navigation entwickelt und in das Gesamtsystem integriert werden.

2.2 Systemarchitektur

Es wurde eine modularen Systemarchitektur, welche die verschiedenen Hardware- und Softwarekomponenten des VASCO-Assistenzsystems zu einem leistungsfähigen Gesamtsystem zusammenführt, entwickelt. Hierzu wurden die funktionalen Anforderungen analysiert, die einzelnen Funktionen den jeweiligen Systemkomponenten zugeordnet sowie die erforderlichen Hard- und Softwareschnittstellen für die Integration der Beiträge aller Projektpartner definiert.
Die entwickelte Gesamtarchitektur ist in Abbildung 4 dargestellt. Sie beschreibt die funktionale Struktur des VASCO-Assistenzsystems, die Zuordnung der einzelnen Funktionen zu den jeweiligen Hardware- und Softwarekomponenten sowie die definierten Kommunikationsschnittstellen zwischen den Modulen. Dadurch wird die Grundlage für die Integration der Beiträge aller Projektpartner und deren Zusammenführung zu einem gemeinsamen Gesamtsystem geschaffen.

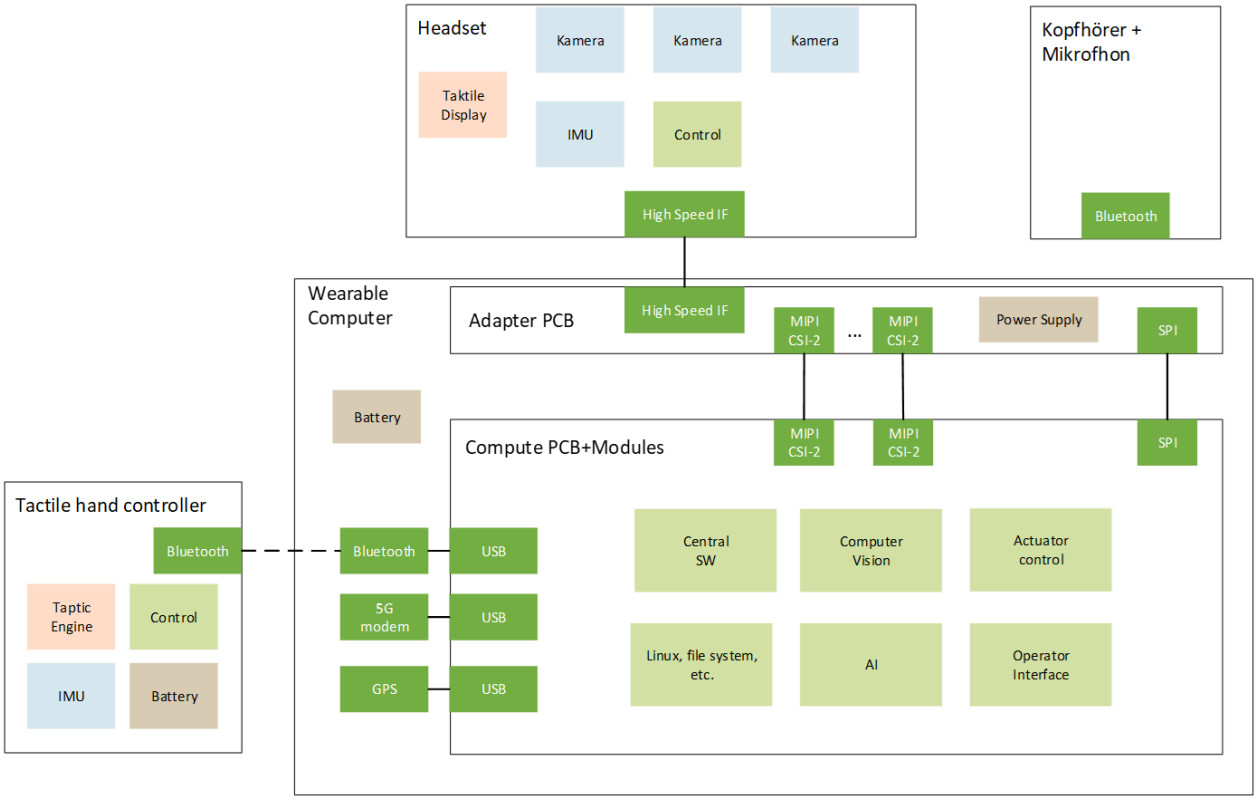

Abbildung 4: Blockdiagramm des Gesamtsystems

Die Erfassung der Umgebung erfolgt über das kamerabasierte Sensorsystem im Headset. Die Bilddaten werden auf dem tragbaren Computer verarbeitet, auf dem sämtliche rechenintensiven Verfahren der Bildverarbeitung, der künstlichen Intelligenz sowie der 3D-

Rekonstruktion ausgeführt werden. Die gewonnenen Informationen werden in einem gemeinsamen dreidimensionalen Umfeldmodell zusammengeführt, das als zentrale Datenbasis für sämtliche Assistenzfunktionen dient. Darauf aufbauend werden unter anderem Funktionen zur Kollisionsvermeidung, Navigation sowie Greifen und Interaktion realisiert. Die Ausgabe der Assistenzinformationen erfolgt über das taktile Display, akustische Ausgabegeräte sowie den Armbandcontroller.
Ein weiterer wesentlicher Bestandteil der Architekturentwicklung war die Definition standardisierter Kommunikations- und Datenschnittstellen zwischen den einzelnen Systemkomponenten. Diese ermöglichen den Austausch von Sensor-, Zustands- und Umgebungsinformationen und bilden die Grundlage für eine modulare Integration der im Projekt entwickelten Hardware und Software. Die gewählte Architektur erlaubt gleichzeitig eine weitgehend unabhängige Weiterentwicklung einzelner Komponenten sowie die spätere Erweiterung des Systems um zusätzliche Sensoren, Algorithmen oder Assistenzfunktionen.
So entstand eine konsistente, modular aufgebaute Systemarchitektur mit klar definierten Funktionsallokationen der einzelnen Komponenten und standardisierten Schnittstellen. Sie bildet die Grundlage für die nachfolgend beschriebenen Entwicklungen der Hard- und Software sowie deren Integration zu einem echtzeitfähigen Assistenzsystem.

2.3 Hardware: Feedback und taktiles Display

2.3.1 Tragbarer Computer

Der tragbare Computer stellt die zentrale Recheneinheit des VASCO-Assistenzsystems dar und wird vom Anwender während des Betriebs am Körper getragen. Auf ihm werden alle Algorithmen und Software-Funktionen ausgeführt. Entsprechend hoch sind die Anforderungen an Rechenleistung, Energieversorgung, Konnektivität sowie der mechanischen Integration.
Der grundsätzliche Aufbau des tragbaren Computers ist in Abbildung 5 dargestellt. Er zeigt die wesentlichen Hardwarekomponenten sowie deren Zusammenwirken innerhalb des Gesamtsystems. Neben der Rechenplattform umfasst der tragbare Computer unter anderem die Energieversorgung, Mobilfunk- und GNSS-Komponenten sowie die Schnittstellen zu Headset, Kameras und den weiteren Ein- und Ausgabegeräten.
Im Rahmen des Projekts wurden die Hardwareplattform sowie das vom Projektpartner entwickelte Adapterboard in Betrieb genommen und zu einer funktionsfähigen Gesamthardware integriert. Darüber hinaus wurde ein Gehäuse für den mobilen Einsatz entwickelt, in das sämtliche Komponenten des tragbaren Computers integriert wurden. Der entwickelte mechanische Aufbau sowie das fertige Gehäuse sind in Abbildung 6 dargestellt.
Bei der Entwicklung standen eine kompakte Bauform, ein robuster Aufbau sowie die vollständige Integration aller Hardwarekomponenten im Vordergrund. Hierfür wurden verschiedene Gehäusevarianten entwickelt und als Funktionsmuster sowohl aus Aluminium als auch mithilfe additiver Fertigungsverfahren aufgebaut und hinsichtlich Einbausituation, Handhabung und Funktion überprüft.
Im Ergebnis entstand eine vollständig integrierte Hardwareplattform, welche sämtliche Komponenten des Assistenzsystems aufnimmt und die erforderlichen Schnittstellen für Sensorik, Energieversorgung und Ausgabegeräte bereitstellt.

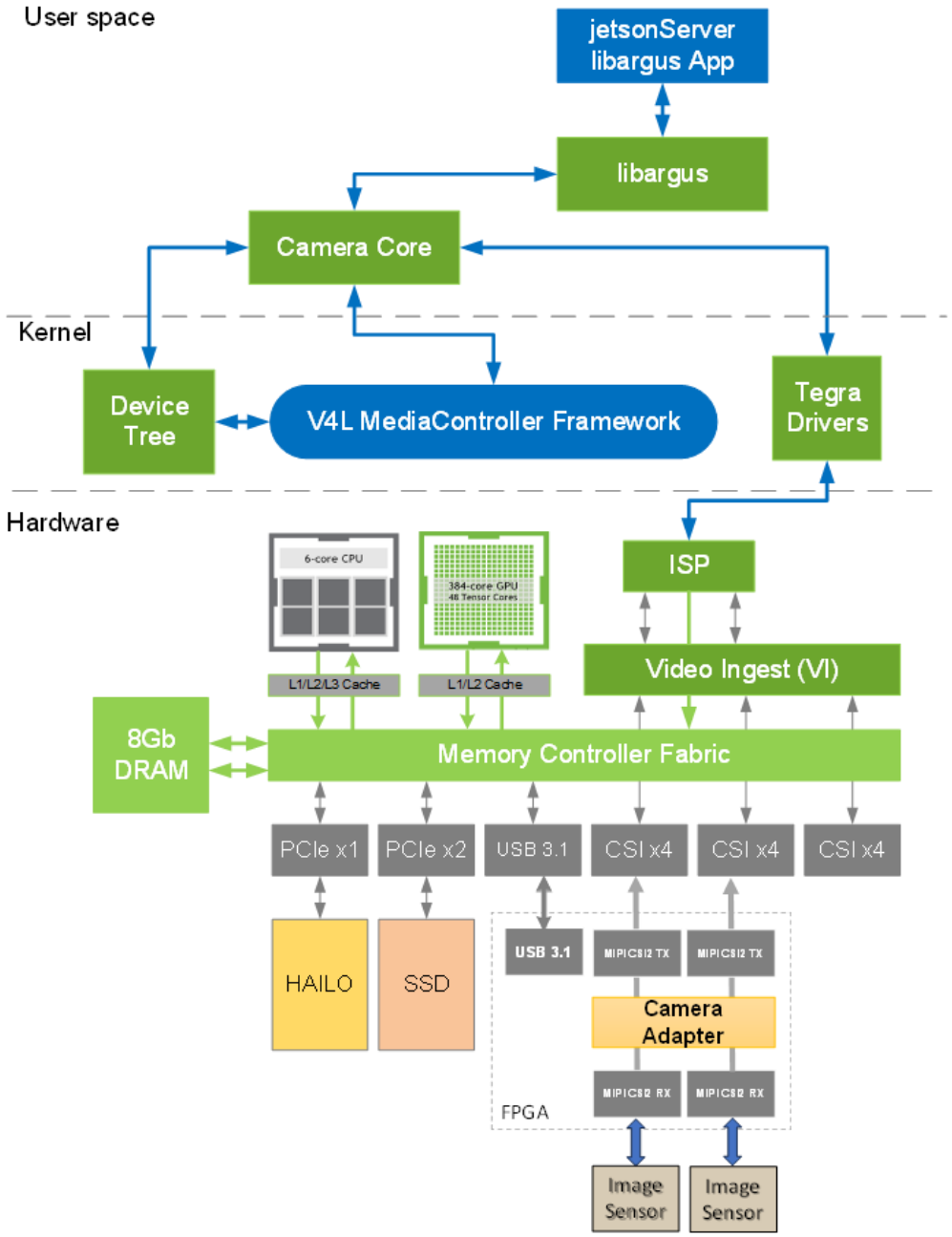

Abbildung 5: Hardwarediagramm des tragbaren Computers

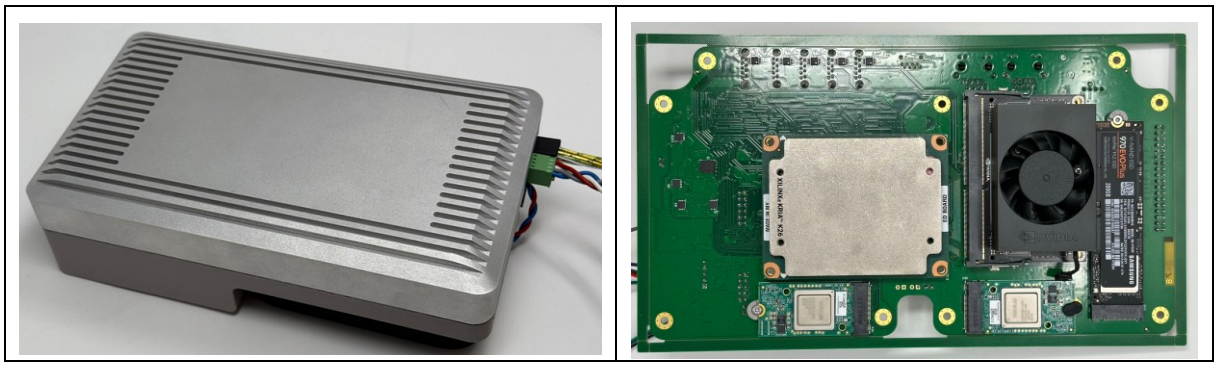

Abbildung 6: Links: Aufgebautes Gehäuse des tragbaren Computers, rechts: Mainboard mit GPU, FPGA, SSD und KI-Beschleuniger-Modulen

2.3.2 Headset und taktiles Display

Das Headset dient als Träger der optischen Sensorik sowie des taktilen Displays und bildet gemeinsam mit dem tragbaren Computer das mobile Assistenzsystem. Die Entwicklung der Hardware erfolgte überwiegend durch die Projektpartner, während DC Vision Systems die Integration der optischen Sensorik sowie die Anbindung an die Bildverarbeitung und das Gesamtsystem unterstützte.
Die Arbeiten umfassten insbesondere die Abstimmung der technischen Anforderungen und Schnittstellen für die Kameramodule sowie die Inbetriebnahme und Integration des Headsets in die Systemarchitektur. Darüber hinaus wurden die Hardware- und Kommunikationsschnittstellen zwischen Headset und tragbarem Computer umgesetzt und für die weitere Systemintegration vorbereitet.
Im Bereich des taktilen Displays wurden die technischen Spezifikationen sowie das Kommunikationsprotokoll zwischen dem tragbaren Computer und der Anzeige gemeinsam mit den Projektpartnern erarbeitet und umgesetzt.

2.4 Bildverarbeitung und Umfeldmodell

2.4.1 Infrastruktur der Bildverarbeitung

Als Basis wurde eine modulare, zur Laufzeit konfigurierbare Softwareinfrastruktur entwickelt. Diese ermöglicht einen durchgängigen Datenfluss von der Sensorerfassung bis zur Weiterverarbeitung in den einzelnen Algorithmen. Die Architektur der Bildverarbeitung ist in Abbildung 7 dargestellt.

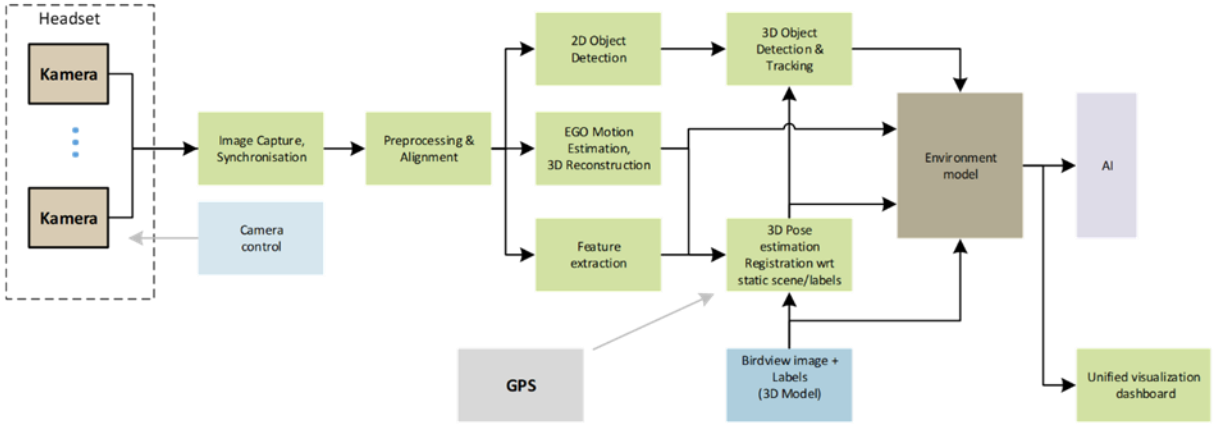

Abbildung 7: Architektur der Bilderfassung und -Verarbeitung

Die Infrastruktur basiert auf einer Pipeline aus Data Providern, Data Processoren und Data Consumern. Data Provider übernehmen die Bereitstellung der Sensordaten und weiterer Eingangsquellen. Hierzu gehören insbesondere die Kameratreiber sowie die Software-Schnittstellen zu den Bildsensoren. Die eigentliche Bildverarbeitung erfolgt innerhalb der Data Processoren, in denen Vorverarbeitungsschritte sowie die verschiedenen Algorithmen der Bildverarbeitung und künstlichen Intelligenz integriert sind. Data Consumer stellen die verarbeiteten Daten den nachgelagerten Systemkomponenten zur Verfügung, beispielsweise der Visualisierung oder dem dreidimensionalen Umfeldmodell.
Ein wesentliches Merkmal der Infrastruktur ist die Möglichkeit, Datenströme sowohl im Live-Betrieb als auch offline zu verarbeiten. Sensordaten können während des Betriebs vollständig aufgezeichnet und anschließend für Entwicklung, Analyse und Tests reproduzierbar wieder in die Verarbeitungspipeline eingespeist werden. Dadurch lassen sich neue Algorithmen unter identischen Randbedingungen entwickeln und bewerten, ohne dass eine erneute Datenerfassung erforderlich ist.
Die gesamte Bildverarbeitungsinfrastruktur wurde in C++ implementiert und unterstützt sowohl den Betrieb auf der eingebetteten Zielplattform als auch auf leistungsfähigen Entwicklungsrechnern. Dadurch können identische Softwarekomponenten für Entwicklung, Test und den späteren Produktivbetrieb verwendet werden. Ergänzend wurden die

Kameraschnittstellen einschließlich der Sensorsteuerung in Betrieb genommen sowie die Vorverarbeitung der Bilddaten, insbesondere das Debayering der Rohdaten, implementiert und die Bildqualität hinsichtlich Schärfe, Kontrast, Dynamikbereich und Farbdarstellung optimiert.

2.4.2 Tracking, Lokalisierung

Die Bestimmung der Kameraposition und -orientierung bildet die räumliche Referenz für sämtliche nachfolgenden Verarbeitungs- und Assistenzfunktionen und ermöglicht es, die über die Zeit erfassten Sensordaten sowie die Ergebnisse verschiedener Algorithmen in einem gemeinsamen dreidimensionalen Koordinatensystem zusammenzuführen. Erst dadurch können Informationen dauerhaft in einem konsistenten Umfeldmodell gespeichert, miteinander verknüpft und von den weiteren Komponenten des Systems gemeinsam genutzt werden. Hierfür wurde ein leistungsfähiges SLAM-Verfahren (*Simultaneous Localization and Mapping*) entwickelt. Es ermöglicht die gleichzeitige Schätzung der Eigenbewegung des Systems sowie den schrittweisen Aufbau einer dreidimensionalen Repräsentation der Umgebung. Ein Beispiel des rekonstruierten Kamerapfades ist in Abbildung 8 dargestellt.

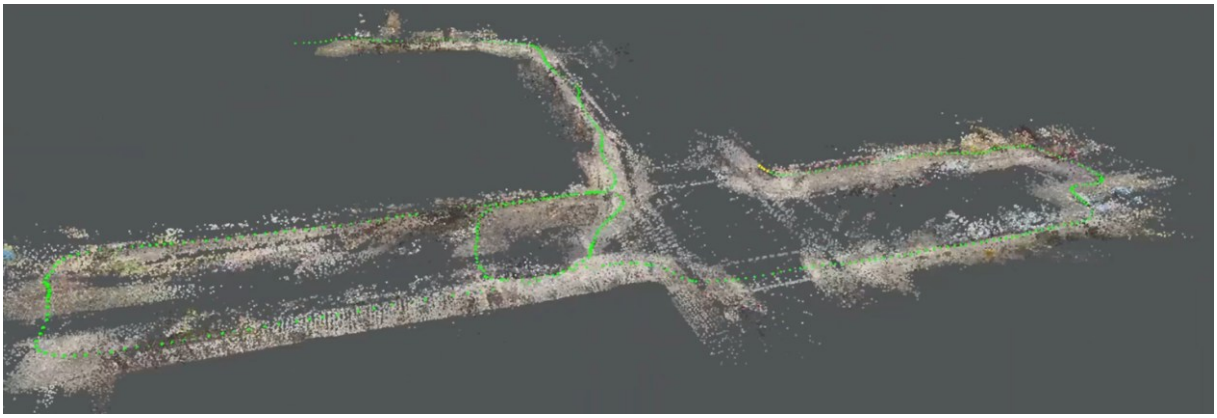

Abbildung 8: Slam Kamerapfad und Punktwolke einer Straßenszene

Das Verfahren arbeitet echtzeitfähig und wurde für den dauerhaften Einsatz in dynamischen Umgebungen entwickelt. Neben der kontinuierlichen Schätzung der Kamerapose wurden Verfahren zur Loop-Closure-Erkennung sowie zur Relokalisierung implementiert. Bereits besuchte Bereiche können dadurch zuverlässig wiedererkannt, Driftfehler reduziert und eine konsistente Lokalisierung auch über längere Einsatzzeiten sichergestellt werden. Gleichzeitig wurde die Verwaltung der internen Szenenrepräsentation so ausgelegt, dass die Größe der Datenbasis auch bei langfristigem Betrieb beherrschbar bleibt und die Echtzeitfähigkeit erhalten wird.
Ein weiterer Bestandteil des Systems ist die Relokalisierung in bereits bekannten Umgebungen. Befindet sich der Anwender in einem zuvor erfassten Bereich, kann die aktuelle Kameraposition innerhalb des bestehenden dreidimensionalen Modells bestimmt werden. Darüber hinaus wurde die Möglichkeit geschaffen, die lokale Rekonstruktion mit georeferenzierten Umgebungsmodellen, beispielsweise auf Basis von Satellitenbildern, in Beziehung zu setzen. Dadurch kann die Position des Anwenders nicht nur relativ zum lokal aufgebauten Umfeldmodell, sondern auch innerhalb einer globalen Umgebung bestimmt werden. Ein Beispiel der Relokalisierung ist in Abbildung 9 dargestellt.

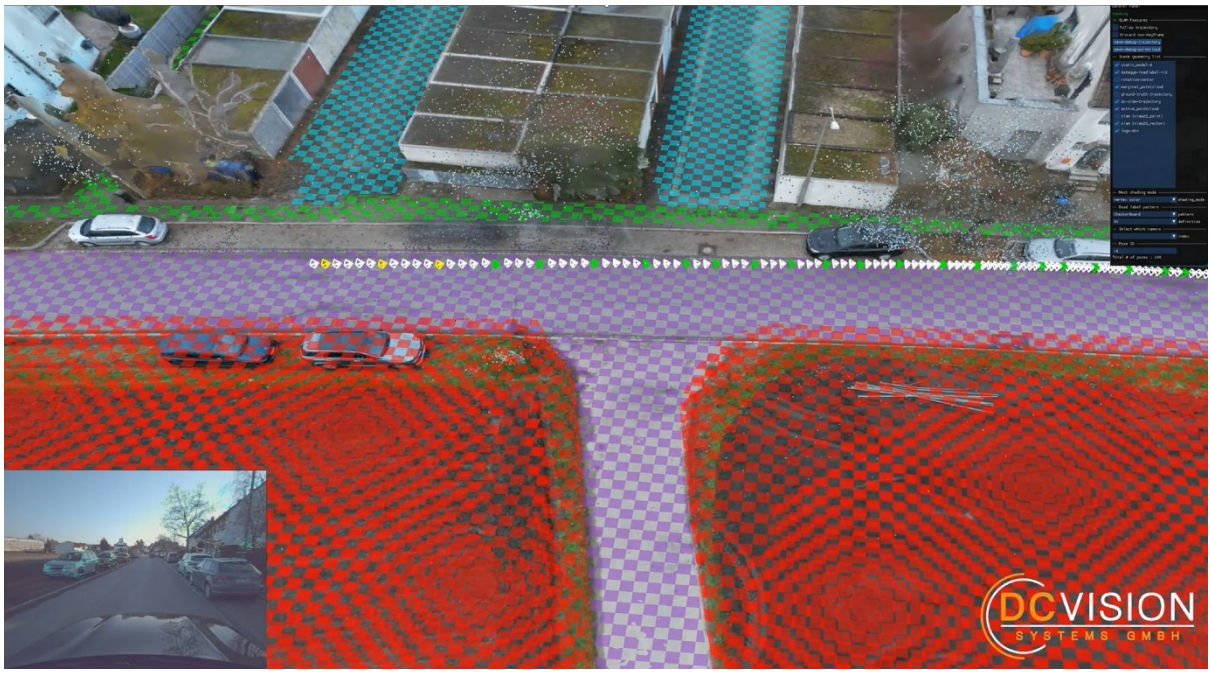

Abbildung 9: Relokalisierung in rekonstruierter Umgebung

2.4.3 3D Rekonstruktion

Aufbauend auf der im vorherigen Kapitel beschriebenen Lokalisierung wurde eine echtzeitfähige 3D-Rekonstruktion entwickelt, welche die vom Kamerasystem erfasste Umgebung fortlaufend zu einem konsistenten dreidimensionalen Modell zusammenführt. Während das SLAM-Verfahren die räumliche Referenz bereitstellt, erzeugt die 3D-Rekonstruktion die geometrische Beschreibung der Umgebung und bildet damit eine wesentliche Grundlage für das spätere Umfeldmodell.

Abbildung 10: Live 3D Rekonstruktion und Visualisierung

Die Tiefeninformation wird ausschließlich aus den Bildern des passiven Stereo-Kamerasystems gewonnen. Hierzu werden sowohl klassische Stereo-Korrespondenzverfahren als auch Mehrbildverfahren (Multi View Stereo, MVS) eingesetzt, wodurch auch in komplexen Szenen eine robuste Tiefenschätzung erreicht wird. Die

gewonnenen Tiefenkarten werden anschließend unter Verwendung einer Truncated Signed Distance Function (TSDF) in ein gemeinsames dreidimensionales Volumen integriert [4]. Zur effizienten Speicherung großer Umgebungen kommt ein Sparse-Voxel-Ansatz auf Basis von Voxel Hashing [5] zum Einsatz, wodurch ausschließlich tatsächlich beobachtete Bereiche der Szene im Speicher abgelegt werden.
Die entwickelte Rekonstruktion arbeitet vollständig echtzeitfähig und ermöglicht den kontinuierlichen Aufbau eines dreidimensionalen Modells bereits während der Bewegung des Anwenders. Ein Beispiel der Live-Rekonstruktion mit der zugehörigen Visualisierung ist in Abbildung 10 dargestellt.

Im Rahmen des Projekts konnten die Robustheit und Qualität der Rekonstruktion insbesondere für große Innen- und Außenbereiche deutlich verbessert werden. Hierzu wurden sowohl die Tiefenschätzung als auch die Integration der Messdaten und die Behandlung dynamischer Szenen deutlich weiterentwickelt. Das Verfahren liefert dadurch auch unter anspruchsvollen Einsatzbedingungen eine stabile geometrische Repräsentation der Umgebung.

2.4.4 Semantische Klassifikation

Die semantische Klassifikation dient der Zuordnung von Objektklassen zu Bereichen des 3D-Umfeldmodells. Dazu wird zunächst eine semantische Klassifizierung auf Basis der 2D-Bilddaten pro Pixel durchgeführt (siehe Abbildung 11). Die Segmentierung lässt sich durch Prompts steuern [6], d. h. je nach textueller Vorgabe können unterschiedliche Objektklassen erkannt werden. Je nach Use-case werden dementsprechend unterschiedliche Objektklassen konfiguriert. In der oberen Reihe sind z. B. Objekte für den Use-case „Greifen und Interaktion“ selektiert. In der unteren Reihe Objekte für das Verkehrs- und Innenraum-Szenario.

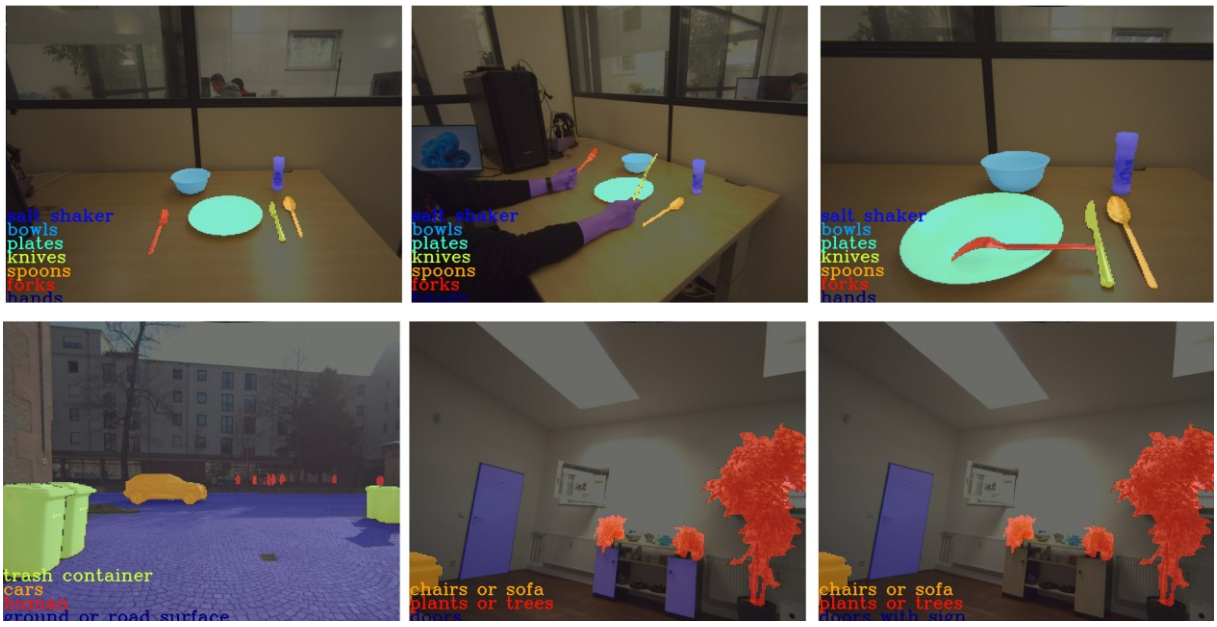

Abbildung 11: Semantische Klassifikation basierend auf Prompts

Auf dieser Grundlage erfolgt die Annotation der Objektklassen für die einzelnen Voxel (siehe Abbildung 12). Die durch den Klassifikator bestimmten, bildbasierten Annotationen werden dabei durch Projektion aus den 2D-Bildern in den 3D-Raum übertragen und dort weiterverarbeitet sowie gefiltert.
Dadurch kann das 3D Umfeldmodell pro Voxel um die Zugehörigkeit zu Objektklassen angereichert werden. Die Möglichkeit der Projektion der 2D Bild-Informationen in das 3D Umfeldmodell ist ein weiteres Beispiel für die Bedeutung des SLAM-Verfahrens für die gesamte Verarbeitungskette.

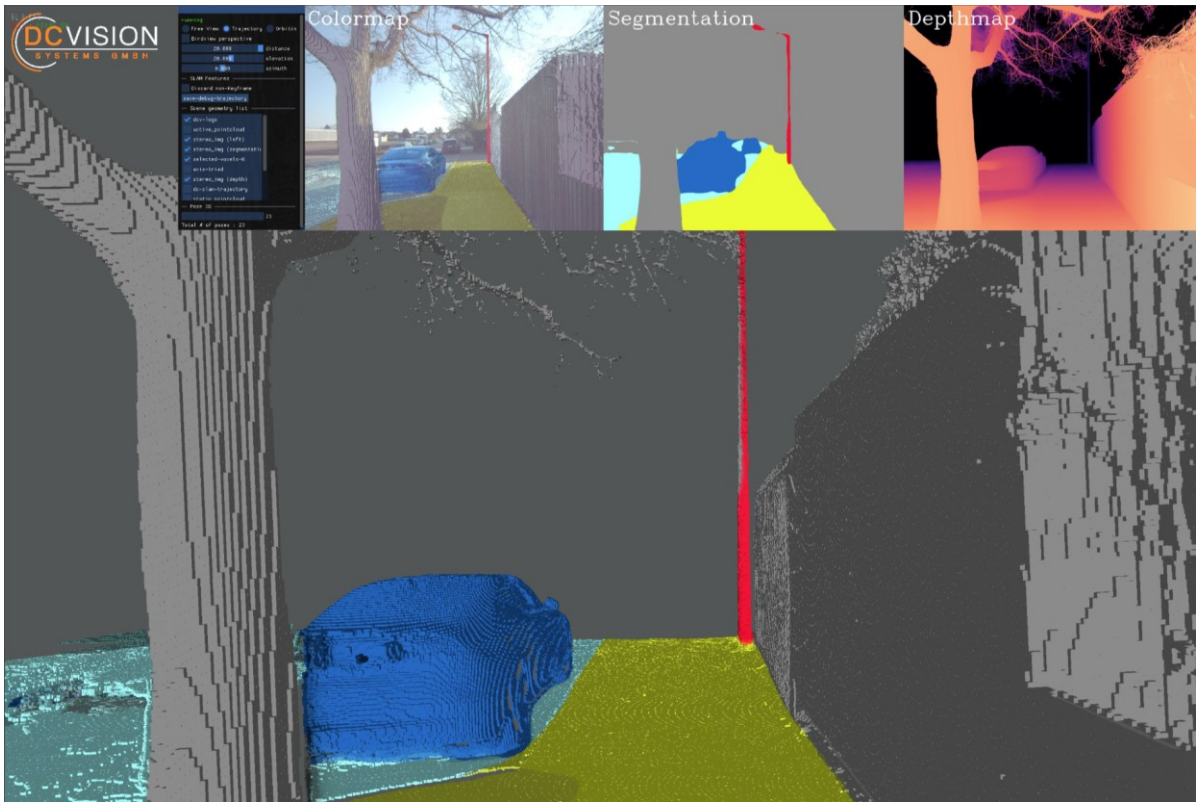

Abbildung 12: Semantische Klassifikation auf Voxel-Ebene

2.4.5 Erkennung und Tracken von Objekten

Aufbauend auf der 3D-Rekonstruktion und dem Umfeldmodell wurde eine Infrastruktur zur Erkennung und Verfolgung statischer und dynamischer Objekte entwickelt. Hierzu wurden moderne KI-basierte Klassifikationsverfahren in die Bildverarbeitung integriert und um Verfahren zur dreidimensionalen Lokalisierung und Verfolgung der erkannten Objekte erweitert.
Die Objekterkennung erfolgt zunächst im Kamerabild, wobei verschiedene 2D- und 3D-Klassifikatoren untersucht, integriert und hinsichtlich ihrer Robustheit und Laufzeit optimiert wurden. Durch die Verknüpfung mit der Tiefeninformation sowie der kontinuierlich bestimmten Kameraposition können die erkannten Objekte als dreidimensionale Bounding-Boxes im Umfeldmodell verortet und über mehrere Bildfolgen hinweg verfolgt werden. Ein Beispiel der 3D-Objekterkennung und des Trackings ist in Abbildung 13 dargestellt.

Abbildung 13: 3D Objekterkennung und Tracking. Links: Projektion ins Kamera-Bild. Rechts: Lokalisierung im 3D Umfeldmodell

Im Gegensatz zu einer rein zweidimensionalen Bildauswertung stehen dadurch die räumliche Position, Orientierung und Ausdehnung der Objekte im gemeinsamen Umfeldmodell zur Verfügung. Dies ermöglicht die Berechnung von Sicherheitsabständen sowie die Vorhersage möglicher Kollisionen mit bewegten Objekten.

2.4.6 Umfeldmodell

Das Umfeldmodell bildet die zentrale Wissensbasis des gesamten Assistenzsystems. Während die vorhergehenden Verarbeitungsschritte jeweils Teilinformationen über die Umgebung liefern, werden diese im Umfeldmodell zu einer konsistenten dreidimensionalen Beschreibung der aktuellen Situation zusammengeführt. Sämtliche Assistenzfunktionen greifen auf dieses Modell zu und erhalten dadurch eine einheitliche Sicht auf die Umgebung sowie den aktuellen Zustand des Anwenders.
Grundlage des Umfeldmodells ist die kontinuierliche Lokalisierung des Kamerasystems mittels visueller SLAM-Verfahren. Diese bestimmen fortlaufend die Position und Orientierung des Anwenders innerhalb der Umgebung und ermöglichen es gleichzeitig, neu aufgenommene Sensordaten in ein gemeinsames dreidimensionales Koordinatensystem einzubringen. Dadurch entsteht nicht lediglich eine Momentaufnahme der Szene, sondern ein über die Zeit wachsendes räumliches Modell, das Informationen aus verschiedenen Blickrichtungen miteinander verbindet.
Ein wesentlicher Vorteil dieses Ansatzes besteht darin, dass Informationen auch dann erhalten bleiben, wenn sie sich nicht mehr unmittelbar im Sichtfeld der Kameras befinden. Bereits rekonstruierte Bereiche können jederzeit wieder genutzt werden, sodass das Assistenzsystem auf eine deutlich vollständigere Beschreibung der Umgebung zugreifen kann, als dies allein auf Basis des aktuellen Kamerabildes möglich wäre. Gleichzeitig ermöglicht die kontinuierliche Relokalisierung eine robuste Bestimmung der eigenen Position innerhalb des bereits aufgebauten Umfeldmodells.
Die Kollisionsvermeidung nutzt die dreidimensionale Beschreibung freier und belegter Bereiche zur Erkennung potenzieller Gefährdungen und zur Bestimmung sicherer Bewegungsräume. Der Use-Case „Greifen und Interaktion“ verwendet die räumliche Lage von Objekten und der Hand des Anwenders zur Berechnung von Bewegungshinweisen. Die Navigationsfunktionen greifen auf das Modell zur lokalen Pfadplanung sowie zur kontinuierlichen Positionsbestimmung des Anwenders zurück. Darüber hinaus dient das Umfeldmodell als gemeinsame Datengrundlage für den zentralen Zustandsautomaten und die KI-basierte Aufgabenplanung des Gesamtsystems.
Die Kombination aus visueller Lokalisierung, Echtzeit-3D-Rekonstruktion, semantischer Szenenanalyse und Objekttracking stellt eine wesentliche Innovation des Teilvorhabens dar.

2.4.7 Visualisierung des Umfeldmodells

Zur Überwachung, Analyse und Demonstration des Gesamtsystems wurde eine leistungsfähige Visualisierung des Umfeldmodells entwickelt. Die Visualisierung basiert auf OpenGL und stellt sämtliche Informationen des Umfeldmodells dreidimensional und in Echtzeit dar. Sie greift dabei direkt auf die Datenströme des Assistenzsystems zu, sodass Änderungen der Umgebung sowie die Ergebnisse der Bildverarbeitung und KI ohne zusätzliche Nachbearbeitung unmittelbar visualisiert werden können.
Die Visualisierung ist modular aufgebaut und umfangreich konfigurierbar. Je nach Anwendungsfall können unterschiedliche Informationen einzeln oder kombiniert dargestellt werden. Hierzu zählen unter anderem die rekonstruierte 3D-Geometrie der Umgebung, semantische Objektklassen, erkannte und verfolgte Objekte einschließlich ihrer Trajektorien, Kamerapositionen, Freiräume, Kollisionsbereiche sowie weitere Zustandsinformationen des Gesamtsystems. Dadurch kann die Verarbeitungskette vom Sensor bis zu den Assistenzfunktionen vollständig nachvollzogen werden.
Ein wesentliches Merkmal der entwickelten Architektur ist die Trennung zwischen Datenverarbeitung und Darstellung. Die Visualisierung kann sowohl lokal auf dem tragbaren Rechner als auch remote auf einem separaten Rechner betrieben werden. Hierfür werden ausschließlich die Echtzeit-Datenströme des Assistenzsystems übertragen, wodurch eine

Überwachung und Analyse des Systems ohne Eingriff in die eigentliche Verarbeitung möglich ist. Dies ermöglicht beispielsweise die Beobachtung des Systems durch einen Operator oder Entwickler während des laufenden Betriebs sowie die Analyse aufgezeichneter oder live übertragener Szenarien.
Die Visualisierung wurde während des gesamten Projekts zur Integration, Validierung und Demonstration der entwickelten Algorithmen eingesetzt und stellt ein wichtiges Werkzeug zur Bewertung des Systemverhaltens dar. Sämtliche in diesem Bericht dargestellten 3D-Visualisierungen des Umfeldmodells sowie der Ergebnisse der Bildverarbeitung wurden mit der entwickelten OpenGL-basierten Visualisierung erzeugt.

2.5 Systemsoftware und Systemintegration

2.5.1 Integration des Basis-Systems und Inbetriebnahme der Gesamt-Applikation

Im Rahmen dieses Arbeitspakets wurde die Systemsoftware des tragbaren Computers entwickelt und die Grundlage für die Integration sämtlicher im Projekt entstandener Hard- und Softwarekomponenten geschaffen. Hierzu wurde das Linux-Betriebssystem auf die Embedded-Plattform portiert und an die entwickelte Hardware angepasst. Neben den Standardkomponenten wurden die erforderlichen Treiber für die projektspezifische Peripherie, insbesondere für die Bildsensorik, entwickelt und integriert. Weitere Komponenten wie Audio, Mobilfunk, GNSS und die Energieverwaltung wurden über die entsprechenden Systemdienste und Schnittstellen eingebunden.
Aufbauend auf dem Betriebssystem wurde eine modulare Basissoftware entwickelt, welche die zentrale Laufzeitumgebung des Assistenzsystems bildet. Sie übernimmt die Initialisierung und Konfiguration der Hardware, stellt einheitliche Schnittstellen zwischen den Softwarekomponenten bereit und koordiniert deren Ausführung. Dadurch können sowohl eigenständige Dienste als auch direkt in die Verarbeitungspipeline integrierte Algorithmen gemeinsam betrieben und zur Laufzeit flexibel konfiguriert werden.
Ergänzend wurden Diagnose-, Verwaltungs- und Remote-Funktionen implementiert, welche die Überwachung des Gesamtsystems sowie die Analyse und Wartung während der Entwicklung und im späteren Betrieb unterstützen.

2.5.2 Operator Schnittstelle

Zur Unterstützung des Anwenders sowie für Entwicklung, Inbetriebnahme und Validierung des Systems wurde eine Operator-Schnittstelle entwickelt. Diese ermöglicht den Fernzugriff auf das Assistenzsystem und stellt dem Operator die relevanten Sensor-, Bild- und Zustandsinformationen in Echtzeit zur Verfügung. Gleichzeitig können Systemfunktionen überwacht, konfiguriert und gesteuert, sowie gezielt Rückmeldungen an den Anwender übermittelt werden. Das grundlegende Kommunikationskonzept der Operator-Schnittstelle ist in Abbildung 14 dargestellt.
Ergänzend wurde ein Dashboard entwickelt, das die wesentlichen System- und Sensordaten übersichtlich visualisiert und den aktuellen Betriebszustand des Assistenzsystems darstellt (siehe Abbildung 15). Die Visualisierung wird kontinuierlich durch die Systemsoftware erzeugt, komprimiert und an den Operator übertragen. Die Kommunikation zwischen Assistenzsystem und Operator erfolgt über eine MQTT-basierte Nachrichtenarchitektur, wodurch sowohl die Übertragung von Sensordaten als auch die Fernsteuerung einzelner Systemfunktionen in Echtzeit ermöglicht wird.

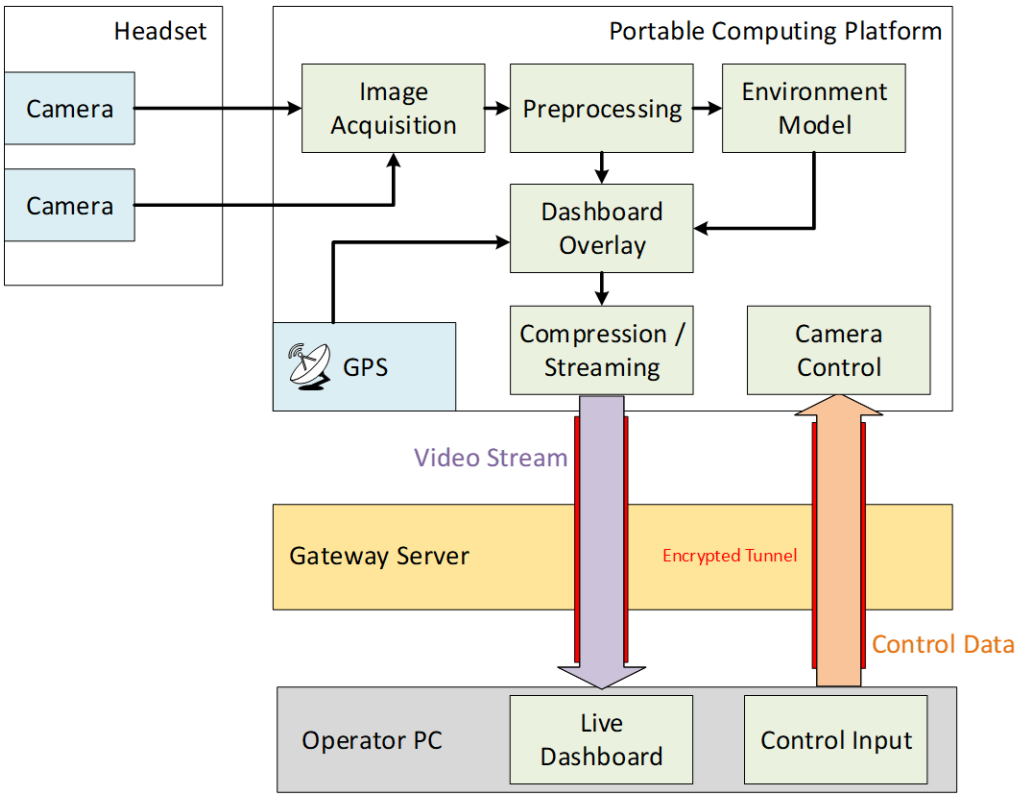

Abbildung 14: Architektur der Verbindung zum Operator

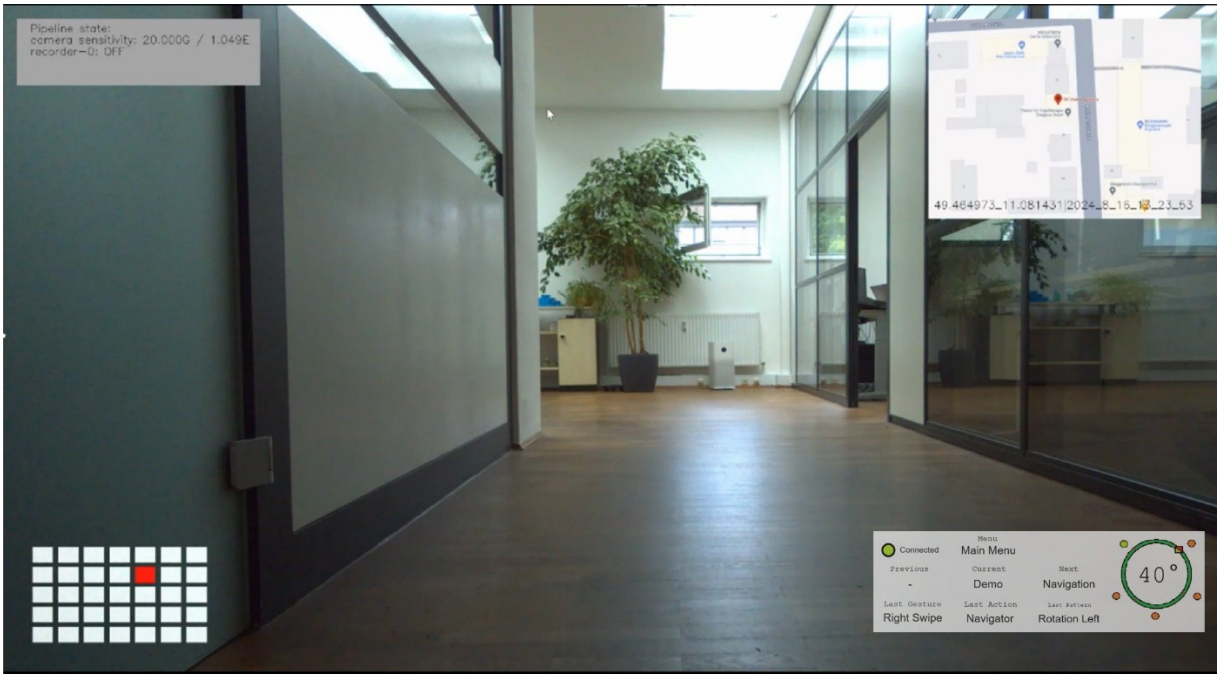

Abbildung 15: Dashboard zur Anzeige der Sensordaten für den Operator

2.5.3 Integration des taktilen Armbandcontrollers

Der Armbandcontroller dient als Steuerelement für den Anwender und benötigt deshalb eine Anbindung an die Steuersoftware der Hauptanwendung. Der Armbandcontroller wird von einem Projektpartner entwickelt und von DC Vision Systems in Zusammenarbeit mit dem Projektpartner in das Gesamtsystem integriert. Die Anbindung erfolgt über Bluetooth und der Controller stellt Host-Seitig eine Runtime zur Ansteuerung bereit. Mit dem Armbandcontroller konnte der Use-case „Greifen und Interaktion“ realisiert werden.

2.6 Realisierung von Use-cases

2.6.1 Greifen und Interaktion

Der Use-Case „Greifen und Interaktion“ dient der Unterstützung des Anwenders bei der gezielten Interaktion mit Objekten in seiner unmittelbaren Umgebung. Ziel ist es, Objekte zuverlässig zu lokalisieren und den Anwender während der Greifbewegung so zu unterstützen, dass diese sicher und präzise ausgeführt werden kann. Der Anwendungsfall bildet eine wichtige Grundlage für zahlreiche Alltagssituationen, beispielsweise das Auffinden und Greifen von Getränken, Lebensmitteln oder anderen Gebrauchsgegenständen.
Der Ablauf des Use-Cases beginnt mit der kontinuierlichen Erfassung der Umgebung durch das Kamerasystem. Es werden relevante Objekte erkannt, klassifiziert und innerhalb des Umfeldmodells räumlich lokalisiert. Dadurch stehen ihre Position und räumliche Ausdehnung unabhängig vom aktuellen Kamerablick kontinuierlich zur Verfügung.
Wählt der Anwender ein Objekt zur Interaktion aus, wird dessen Position fortlaufend mit der aktuellen Bewegung des Anwenders in Beziehung gesetzt. Hierzu werden neben der Position des Zielobjekts auch die Bewegung des Kamerasystems sowie die Lage der Hand berücksichtigt. Auf dieser Grundlage werden kontinuierlich situationsabhängige Hinweise erzeugt, welche den Anwender bei der Annäherung an das Objekt unterstützen. Gleichzeitig können potenzielle Kollisionen der Hand oder des Arms mit anderen Gegenständen erkannt werden, wodurch die Sicherheit während der Interaktion erhöht wird.
Die Rückmeldung an den Anwender erfolgt über den entwickelten Armbandcontroller sowie ergänzend über die weiteren Ausgabekomponenten des Assistenzsystems. Die Hinweise werden fortlaufend an die aktuelle Situation angepasst und unterstützen den Anwender während der gesamten Greifbewegung bis zum Erreichen des Zielobjekts. Die Verarbeitungsschritte vom Kamerabild über die Bildverarbeitung bis zur Ausgabe der Assistenzinformationen sind in Abbildung 16 dargestellt.
Die entwickelten Verfahren konnten im Projekt erfolgreich in das Gesamtsystem integriert und anhand verschiedener Demonstrationsszenarien validiert werden. Die Bildverarbeitung, die Erzeugung des Umfeldmodells sowie die Integration des Armbandcontrollers arbeiteten dabei in Echtzeit zusammen und ermöglichten eine kontinuierliche Unterstützung des Anwenders während des Greifvorgangs.

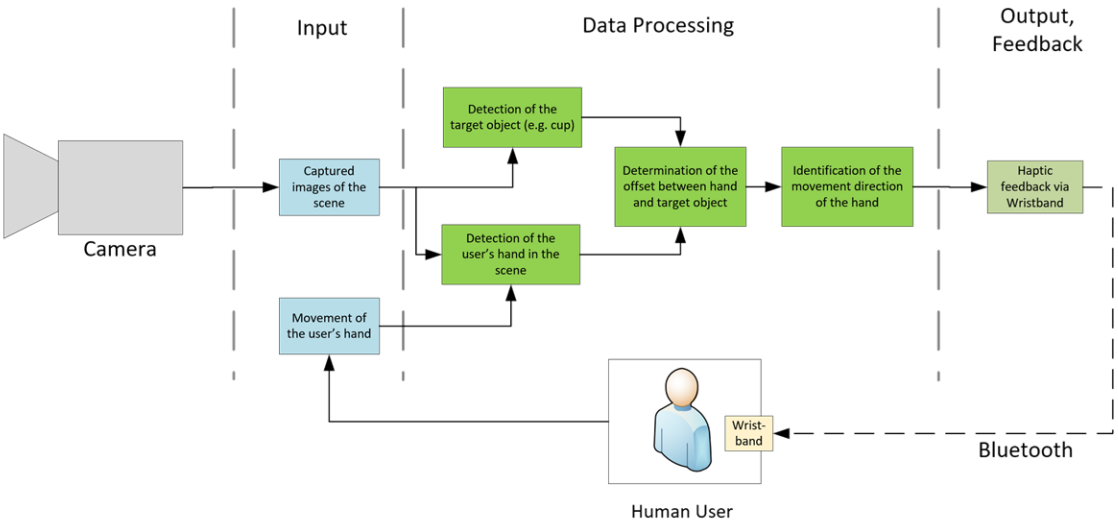

Abbildung 16: Verarbeitungskette beim Use-case “Greifen und Interaktion”

Abbildung 17 links zeigt den Versuchsaufbau zur Validierung des Use-Cases „Greifen und Interaktion“. Der Proband trägt das vorläufige Kamerasystem am Kopf sowie den Armbandcontroller zur Ausgabe der Assistenzinformationen. Das zu greifende Objekt befindet sich innerhalb des Erfassungsbereichs des Stereo-Kamerasystems. Die Umgebung, die Hand

des Anwenders sowie das Zielobjekt werden kontinuierlich erfasst und in Echtzeit verarbeitet, um die für den Greifvorgang erforderlichen Bewegungsinformationen zu berechnen.

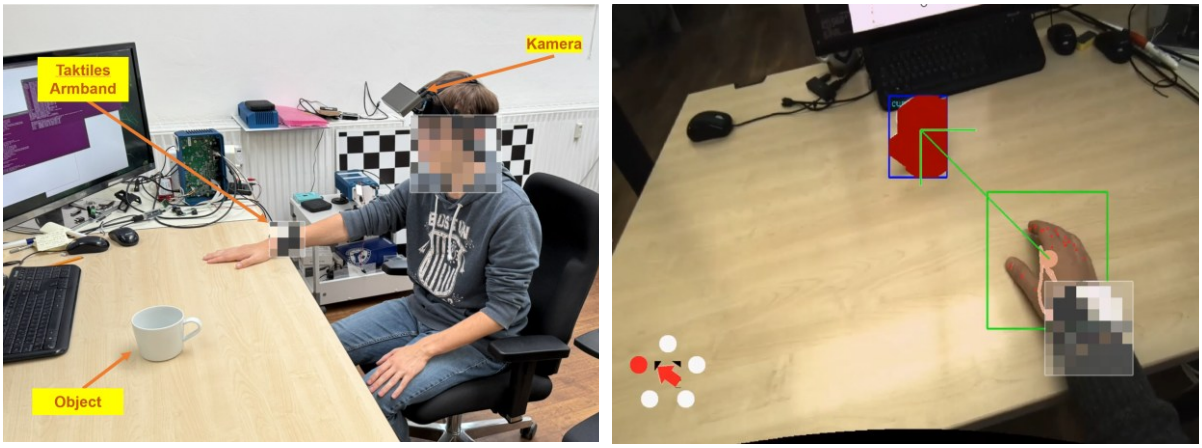

Abbildung 17: Links: Test-setup bei „Greifen und Interaktion“, Rechts: Visualisierung des Umfeldmodells/Dashboard

Abbildung 17 rechts zeigt die Visualisierung der Bildverarbeitung während des Greifvorgangs. Das Zielobjekt wird durch die Objekterkennung lokalisiert und markiert, während gleichzeitig die Hand des Anwenders erkannt und verfolgt wird. Die räumliche Beziehung zwischen Hand und Zielobjekt bildet die Basis für die Berechnung der Assistenzhinweise, welche dem Anwender über den Armbandcontroller bereitgestellt werden.

2.6.2 Kollisionsvermeidung und 3D-Freiraumerkennung

Ziel ist es, potenzielle Kollisionen frühzeitig zu erkennen und den Anwender bei einer sicheren Fortbewegung zu unterstützen. Hierzu werden die dreidimensionale Struktur der Umgebung, begehbare Bereiche sowie statische und dynamische Hindernisse kontinuierlich erfasst und bewertet.
Abbildung 18 zeigt beispielhafte Alltagssituationen mit erhöhtem Kollisionspotenzial für Menschen mit Seheinschränkungen. In vielen dieser Situationen stößt der Blindenlangstock an seine systembedingten Grenzen, da beispielsweise seitlich in den Gehweg hineinragende oder oberhalb des Bodens befindliche Hindernisse sowie Engstellen nur eingeschränkt oder erst unmittelbar vor einer Kollision erkannt werden können. Die dargestellten Szenarien verdeutlichen die Motivation für die im Projekt entwickelten Verfahren zur 3D-Freiraumerkennung und Kollisionsvermeidung.

Abbildung 18: Schwierige Szenarien mit Potential für Kollisionen (Quelle: BFW Düren)

Aufbauend auf dem Umfeldmodell wurde eine 3D-Freiraumerkennung entwickelt. Diese ermittelt kontinuierlich freie Bewegungsräume sowie potenzielle Hindernisse und leitet daraus mögliche Kollisionssituationen ab. Das Prinzip der Kollisionserkennung ist in Abbildung 19 dargestellt.

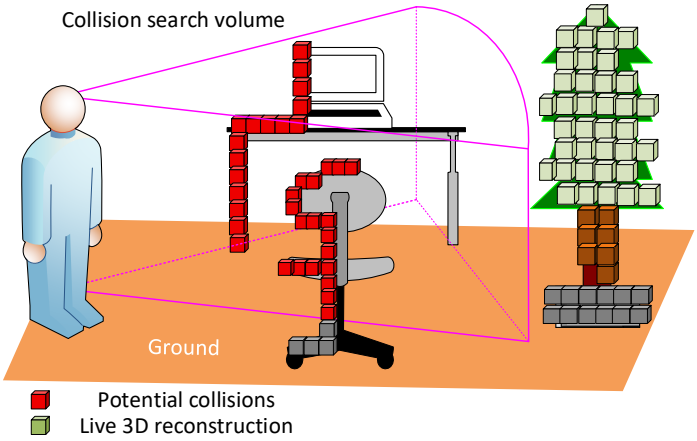

Abbildung 19: Prinzip der Kollisionserkennung

Abbildung 20 zeigt die Anwendung der entwickelten 3D-Freiraumerkennung und Kollisionsbewertung. Dargestellt ist die Live-3D-Rekonstruktion der Umgebung, in der das für die Kollisionsbewertung verwendete Prüfvolumen in Form einer Kegelgeometrie visualisiert ist. Sämtliche Voxel, die innerhalb dieses Volumens liegen und eine potenzielle Kollision darstellen, werden hervorgehoben und rot markiert. Auf diese Weise lassen sich kritische Bereiche der Umgebung unmittelbar identifizieren und kontinuierlich hinsichtlich möglicher Gefährdungen bewerten.

Die entwickelten Verfahren wurden in die bestehende Softwareplattform integriert und sowohl in Innen- als auch in Außenbereichen validiert. Aus Sicht der DC Vision Systems GmbH stellt die entwickelte Lösung einen erheblichen Mehrwert für Menschen mit Seheinschränkungen dar, da sie eine frühzeitige Erkennung komplexer Kollisionssituationen ermöglicht, die mit einem Blindenlangstock nur eingeschränkt oder erst unmittelbar vor einer Kollision erkannt werden können.

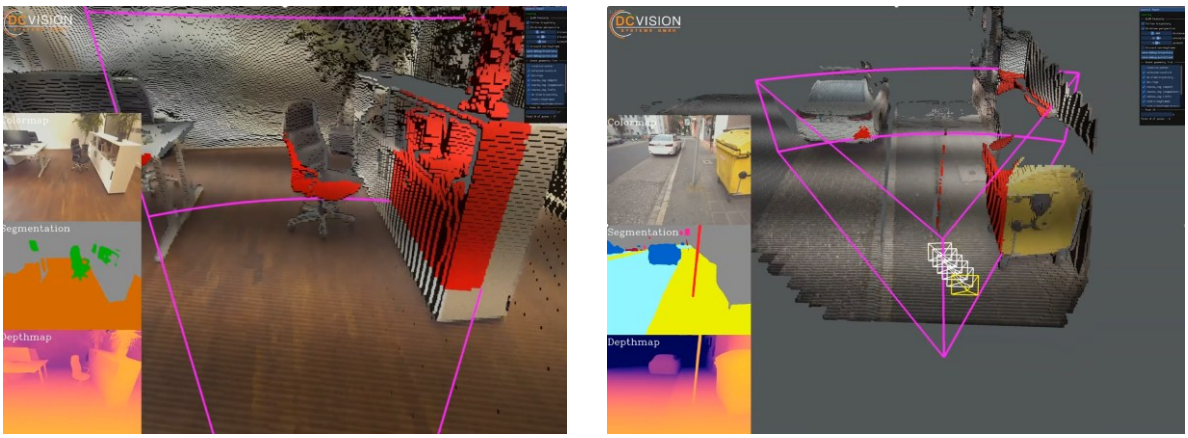

Abbildung 20: Kollisionserkennung auf realen Daten im Innenraum (links) und Außenbereich (rechts)

2.6.3 Outdoor-Navigation

Für die Realisierung der Outdoor-Navigation wurde eine abstrahierte Repräsentation der Umgebung entwickelt, welche die für die Navigation relevanten Szenenbestandteile in einem gemeinsamen Modell zusammenführt. Im Gegensatz zur dreidimensionalen Nahbereichsrekonstruktion stehen hierbei nicht die exakte 3D-Geometrie, sondern die semantische Beschreibung der Umgebung und deren Navigierbarkeit im Vordergrund. Erfasst werden insbesondere begehbare Bereiche, Fahrbahnen, Radwege, Fußgängerüberwege, Hindernisse sowie statische und dynamische Objekte.

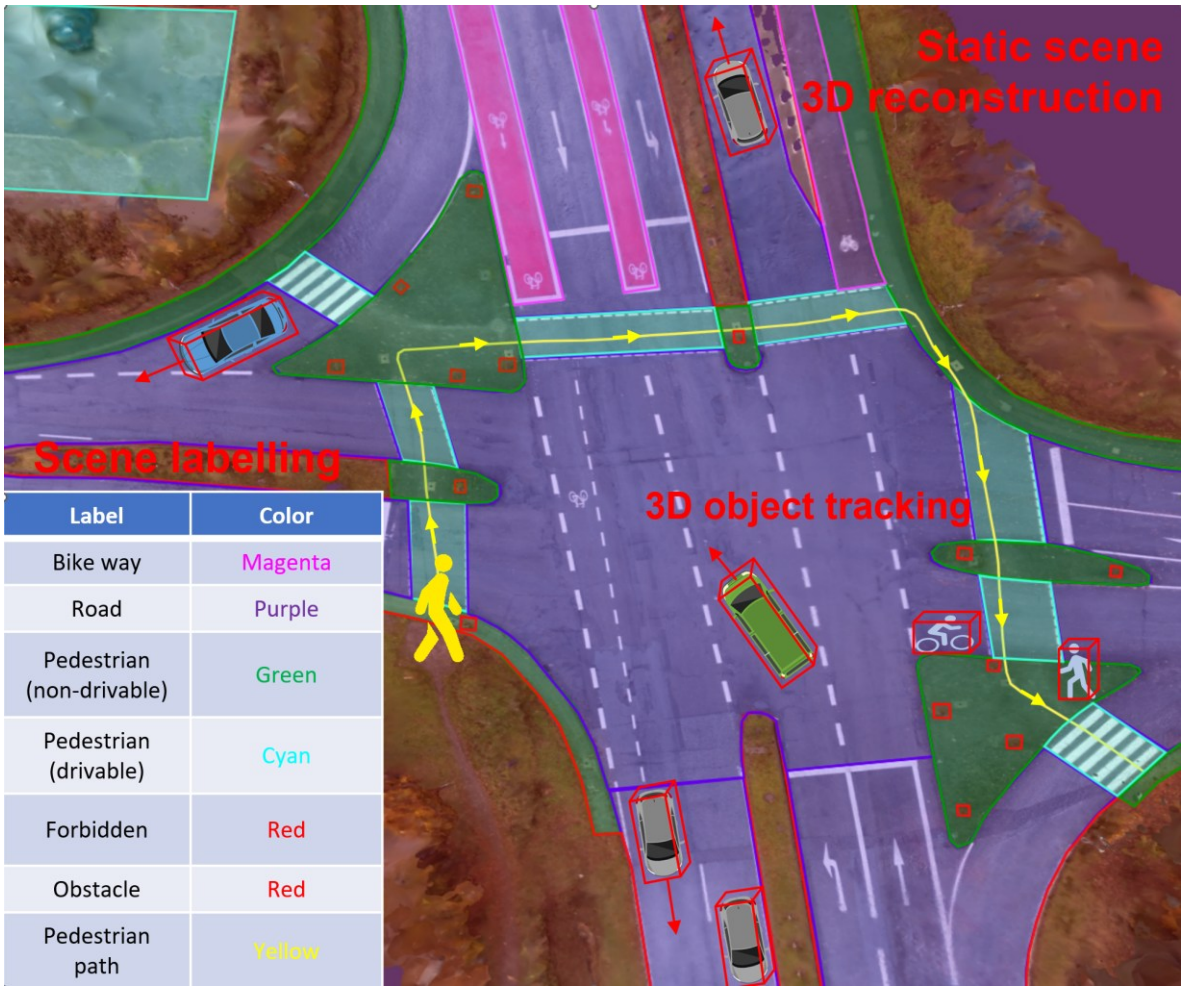

Abbildung 21: Repräsentation der Oudoor-Szene im Umeldmodell

Abbildung 21 zeigt beispielhaft diese Repräsentation einer Outdoor-Szene. Die statischen Bestandteile der Umgebung können dabei entweder manuell auf Grundlage von Luft- oder Satellitenbildern annotiert oder automatisiert erkannt werden. Ergänzend werden dynamische Objekte, beispielsweise Fahrzeuge oder Fußgänger kontinuierlich erfasst und in das Szenenmodell integriert. Die resultierende Szenenrepräsentation bildet die Grundlage für die anschließende Routen- und Pfadplanung.
Für die Navigation werden die aktuelle Position des Anwenders sowie eine Zielposition vorgegeben. Auf Basis der semantischen Szenenrepräsentation wird ein sicherer Pfad berechnet und in einzelne Wegpunkte unterteilt. Wie in Abbildung 22 dargestellt, orientiert sich die Navigation am jeweils nächsten Wegpunkt. Hierzu werden die aktuelle Position sowie die Blickrichtung des Anwenders fortlaufend mit dem geplanten Pfad abgeglichen, sodass jederzeit die optimale Bewegungsrichtung bestimmt werden kann. Parallel hierzu überwacht die im vorherigen Kapitel beschriebene Kollisionsvermeidung die unmittelbare Umgebung und berücksichtigt statische sowie dynamische Hindernisse bei der Navigation.

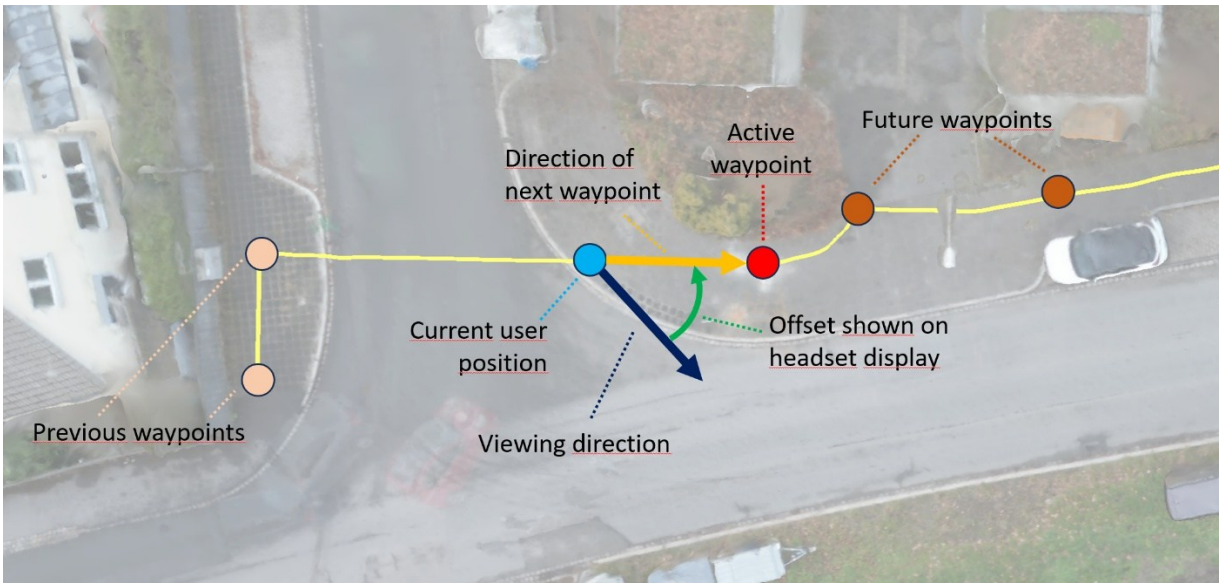

Abbildung 22: Aufteilung eines Pfades in Wegpunkte

Die berechneten Navigationsinformationen werden anschließend an die verschiedenen Ausgabemodule des Assistenzsystems übergeben. Je nach Anwendung erfolgt die Ausgabe der Richtungsinformationen über das taktile Display, das Armband oder eine akustische Benutzerschnittstelle. Abbildung 23 zeigt beispielhaft die Projektion der relevanten Informationen des Umfeldmodells einschließlich der berechneten Pfadvorgabe in das Kamerabild. Diese Visualisierung dient der Entwicklung, Validierung und Analyse der Navigationsverfahren und verdeutlicht das Zusammenwirken von Szenenrepräsentation, Pfadplanung und lokaler Kollisionsvermeidung.

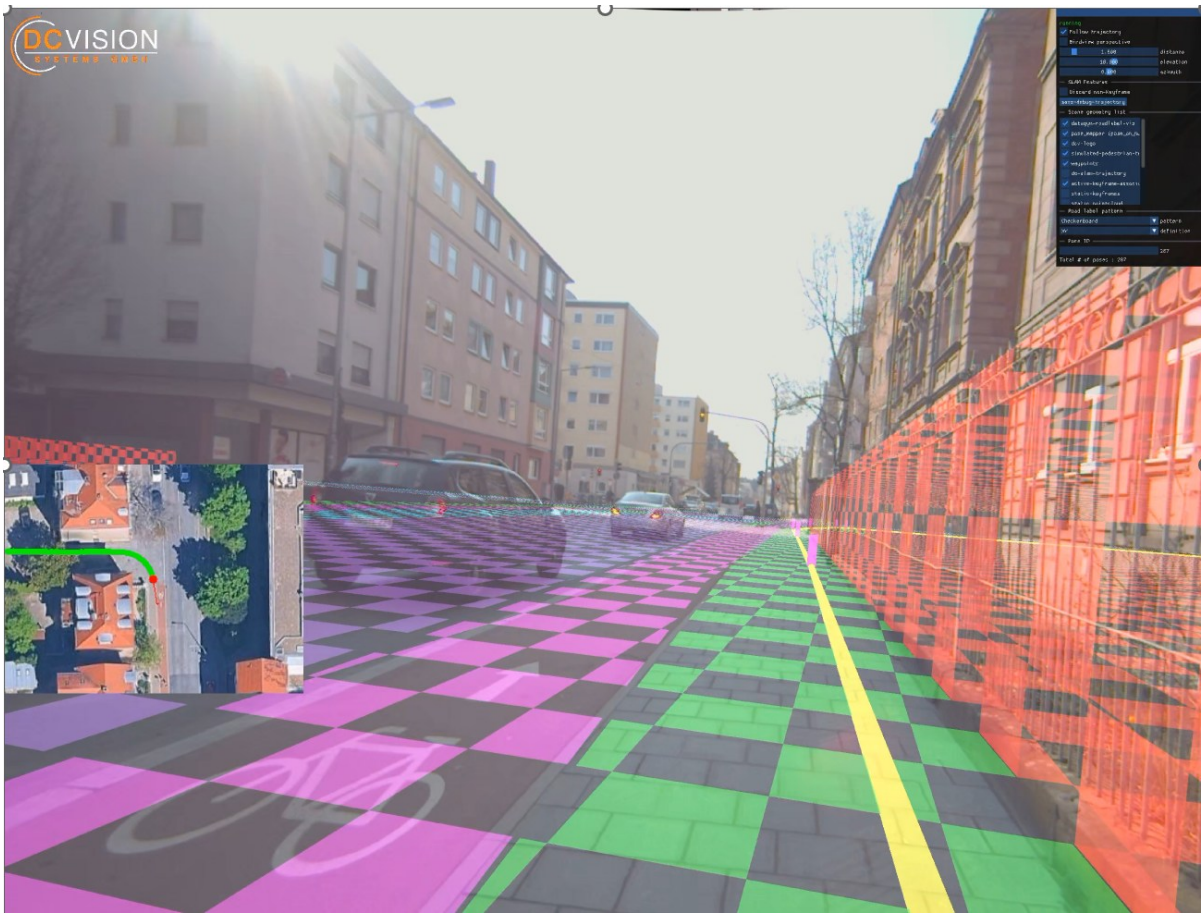

Abbildung 23: Projektion des Umfeldmodells mit Pfadvorgabe in das Kamerabild

3 Verwendung der Zuwendung

3.1 Darstellung und Begründung der wichtigsten Positionen des zahlenmäßigen Nachweises

Die größte Ausgabenposition entfällt auf die Personalkosten. Dies ist auf den Forschungscharakter des Vorhabens zurückzuführen, dessen Schwerpunkt auf der Entwicklung neuer Verfahren der Computer Vision, der Echtzeit-3D-Rekonstruktion, der künstlichen Intelligenz sowie der Hard- und Softwareintegration lag.
Die Materialkosten entstanden überwiegend für den Aufbau der Hardwareprototypen und der Kamerasysteme. Die Reisekosten dienten der Durchführung von Konsortial- und Arbeitstreffen mit den Projektpartnern.

3.2 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Arbeiten wurden entsprechend der im Vorhaben definierten Arbeitspakete durchgeführt und umfassten die Konzeption, Weiterentwicklung, Integration und Validierung der Hard- und Softwarekomponenten des Assistenzsystems. Dabei konnte auf bereits vorhandene Vorentwicklungen von DC Vision Systems, insbesondere im Bereich der Recheneinheit sowie grundlegender Verfahren zur visuellen Lokalisierung (SLAM) und 3D-Rekonstruktion, aufgebaut werden. Diese Vorarbeiten wurden im Rahmen des Vorhabens jedoch wesentlich erweitert und hinsichtlich Funktionalität, Leistungsfähigkeit und Robustheit deutlich weiterentwickelt. Die in Kapitel 2 dargestellten Ergebnisse dokumentieren diesen Ausbau der Technologien sowie deren Integration zu einer leistungsfähigen Gesamtlösung. Darüber hinaus wurden geeignete Open-Source-Komponenten und aktuelle KI-Modelle genutzt und an die Anforderungen des Vorhabens angepasst, wodurch der Entwicklungsaufwand auf die projektrelevanten Erweiterungen und die Systemintegration konzentriert werden konnte.
Die Angemessenheit der geleisteten Arbeiten ergibt sich aus der Komplexität der Aufgabenstellung sowie der Notwendigkeit, die bestehenden Technologien zu einer robusten, echtzeitfähigen Assistenzlösung für den mobilen Einsatz weiterzuentwickeln und in das Gesamtsystem zu integrieren.

3.3 Darstellung des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit der Ergebnisse im Sinne des fortgeschriebenen Verwertungsplans

Die Verwertung des im Verbundvorhaben entwickelten Gesamtsystems wird im Schlussbericht des Gesamtverbundes dargestellt. Im Folgenden wird daher ausschließlich die Verwertung der im Teilvorhaben von DC Vision Systems entwickelten Ergebnisse beschrieben.
Die im Vorhaben entwickelten Verfahren und Softwarekomponenten werden unmittelbar in das Vision SDK von DC Vision Systems übernommen und bilden dort wesentliche Kernfunktionen. Hierzu zählen insbesondere die Verfahren zur Echtzeit-Lokalisierung (SLAM), 3D-Rekonstruktion, Umfeldmodellierung sowie die darauf aufbauenden Funktionen zur Szenenanalyse. Gleichzeitig fließen die Ergebnisse in die von DC Vision Systems entwickelte KI-Kamera ein, auf der das Vision SDK als integrierte Softwareplattform betrieben wird. Dadurch kann DC Vision Systems sowohl einzelne Softwarekomponenten und Lizenzen als auch vollständige Hard- und Softwarelösungen anbieten.
Mit Abschluss des Vorhabens beginnt die Vermarktung des Vision SDK sowie der darauf aufbauenden KI-Kamerasysteme. Die entwickelten Technologien erweitern das bestehende Produktportfolio des Unternehmens erheblich und können in einer Vielzahl unterschiedlicher Anwendungen eingesetzt werden. Hierzu zählen insbesondere mobile Robotik, autonome Systeme, industrielle Automatisierung sowie intelligente kamerabasierte Sensorsysteme.
Die Verwertung erfolgt über die Lizenzierung des Vision SDK, den Vertrieb der eigenen KI-Kamerasysteme einschließlich der integrierten Software sowie über Entwicklungsdienstleistungen für kundenspezifische Anwendungen. Darüber hinaus bilden die Projektergebnisse eine wesentliche Grundlage für die Weiterentwicklung der Technologieplattform und zukünftige Forschungs- und Entwicklungsprojekte. Die Verwertung

der Ergebnisse erfolgt am Unternehmensstandort Nürnberg und damit innerhalb Deutschlands.

3.4 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Durchführung des Vorhabens wurden verschiedene nationale und internationale Entwicklungen im Bereich KI-gestützter Assistenzsysteme für Menschen mit Sehbeeinträchtigungen beobachtet. Hierzu zählen insbesondere kommerzielle Systeme wie Envision, das KI-basierte Szenenbeschreibung und Texterkennung auf Smart Glasses bereitstellt [7], sowie 7Sense (SuperBrain), das mittels haptischer Rückmeldung die Umgebung für den Nutzer erfassbar macht [8] und die sichere Mobilität unterstützen soll. Ein weiterer vergleichbarer Ansatz wird mit den .lumen Glasses verfolgt. Das System erfasst die Umgebung mittels Sensorik und KI, erkennt sichere Bewegungsräume und Hindernisse und übermittelt dem Nutzer in Echtzeit haptische Richtungshinweise zur Unterstützung einer selbstständigen Navigation [9]. Zu den während der Projektlaufzeit bekannt gewordenen Entwicklungen zählt außerdem das Assistenzsystem NOA von Biped AI, das als auf der Schulter getragenes KI-Assistenzsystem mittels räumlicher Sensorik die Orientierung und Interaktion von Menschen mit Sehbeeinträchtigungen unterstützen soll [10].
Darüber hinaus war während der Projektlaufzeit ein erheblicher technologischer Fortschritt in den Bereichen Computer Vision, Deep Learning und Foundation Models zu beobachten. Insbesondere Verfahren zur Objekterkennung, semantischen Segmentierung, Tiefenschätzung, 3D-Rekonstruktion, visuellen Lokalisierung sowie Vision-Language-Modelle haben sich deutlich weiterentwickelt und sind zunehmend auch als leistungsfähige Open-Source-Lösungen verfügbar geworden. Diese Entwicklungen erweitern die Möglichkeiten intelligenter Assistenzsysteme erheblich und eröffnen insbesondere neue Funktionen zur semantischen Szenenanalyse, natürlichen Sprachinteraktion und kontextbezogenen Unterstützung.
Die im Vorhaben entwickelte modulare Systemarchitektur konnte von diesen technologischen Fortschritten unmittelbar profitieren.

3.5 Veröffentlichungen der Ergebnisse

Im Rahmen des Statusseminars „Photonik in den Lebenswissenschaften 2024“ wurden Projektergebnisse in Form eines Posters sowie eines Fachvortrags präsentiert. Darüber hinaus sind ausgewählte, aus dem Vorhaben hervorgegangene Technologien und Funktionen in die Darstellung des Vision SDK auf der Unternehmenswebsite der DC Vision Systems GmbH eingeflossen und werden dort im Rahmen der Produkt- und Technologiekommunikation vorgestellt.
Geplant ist insbesondere die Präsentation der aus dem Vorhaben hervorgegangenen Technologien und Produkte auf der VISION 2026 in Stuttgart. Ergänzend sind weitere Marketing- und Vertriebsmaßnahmen vorgesehen, um die entwickelten Technologien und daraus abgeleiteten Produkte am Markt zu etablieren. Darüber hinaus sind weitere Veröffentlichungen im Rahmen des Verbundprojekts geplant, um die erzielten wissenschaftlich-technischen Ergebnisse sowie die Verwertung des Gesamtsystems kontinuierlich sichtbar zu machen.

4 Anhang

Referenzen

[1] The Directions API overview | Google Developers

[2] bemyeyes.com/language/german

[3] https://www.microsoft.com/de-de/ai/seeing-ai

[4] B. Curless und M. Levoy, „A Volumetric Method for Building Complex Models from Range Images“, Proceedings of SIGGRAPH, 1996.

[5] M. Nießner, M. Zollhöfer, S. Izadi und M. Stamminger, „Real-time 3D Reconstruction at Scale using Voxel Hashing“, ACM Transactions on Graphics, Bd. 32, Nr. 6, 2013.

[6] A. Kirillov et al., „Segment Anything“, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023.

[7] Envision Technologies B.V., „Envision Glasses – Smart Glasses for People who are Blind or Low Vision“, Unternehmenswebsite, abgerufen 2026.

[8] 7Sense OÜ, „SuperBrain 1“, Unternehmenswebsite, abgerufen 2026.

[9] .lumen, „Glasses for the Blind“, Produkt- und Unternehmenswebsite, https://www.dotlumen.com/glasses, abgerufen am 04.08.2026.

[10] Biped AI, *NOA – AI Companion for Blind and Low Vision Users*, https://www.biped.ai/noa, abgerufen am 04.08.2026.