Abschlussbericht zum Forschungsprojekt LOMOBI - „Lernfähiges Assistenzsystem für Orientierung und Mobilität von sehbeeinträchtigten Personen im Alltag“.

Zuwendungsempfänger:	*Friedrich-Alexander-Universität Erlangen-Nürnberg*
Projektleitung:	*Prof. Dr.-Ing. Jörg Franke*
Berichtsautor:	*Helmut Engelhardt, Matthias Kalenberg*
Förderkennzeichen:	*16SV9081*

Abschlussbericht zum Forschungsprojekt LOMOBI - „Lernfähiges Assistenzsystem für Orientierung und Mobilität von sehbeeinträchtigten Personen im Alltag“. 1

1 Aufgabenstellung 1

2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde 2

3 Planung und Ablauf des Vorhabens 3

4 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde 5

5 Angabe bekannter Konstruktionen, Verfahren und Schutzrechte, die für die Durchführung des Vorhabens benutzt wurden 6

6 Angabe der verwendeten Fachliteratur sowie der benutzten Informations- und Dokumentationsdienste 6

Literaturverzeichnis ..... Fehler! Textmarke nicht definiert.

7 Zusammenarbeit mit anderen Stellen 9

8 Ergebnis 9

8.1 Anforderungsanalyse 9
8.2 Soft- und hardwareseitiger Systementwurf und Navigationskonzept 12
8.3 Umgebungsklassifikation durch Machine Learning Modelle 13
8.4 Nutzerinteraktion 14
8.4.1 Smartphone-App 14
8.4.2 Vibrotaktile Nutzerinteraktion 15
8.4.3 3D Audio 16
8.5 Automatische Segmentierung und Trainingsdatengenerierung 17
8.5.1 Carla Simulation für Trainingsdatengenerierung 17
8.5.2 Automatische Segmentierung anhand des gegangen Weg 18
8.5.3 Zero- und Single Shot Segmentierungs Pipeline 19
8.6 Indoor- und Outdoor-Lokalisierung für ETAs 20
8.6.1 Indoor-Lokalisierung 21
8.6.2 Outdoor-Lokalisierung 22
8.7 Pfad- und Ablaufplanung 23
8.7.1 Hierarchisches Pfadplanungskonzept 23
8.7.2 Ablaufplanung mittels Behavior Trees 25
8.8 Erkenntnisse aus den Probandenstudien zur multimodalen Nutzerinteraktion 27
8.8.1 Smartphone App 27
8.8.2 Vibrotaktile Nutzerinteraktion 27
8.8.3 3D Audio 28

9 Voraussichtlicher Nutzen 29

10 Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf diesem Gebiet bei anderen Stellen 29

11 Erfolgte oder geplante Veröffentlichung des Ergebnisses 30

Am Lehrstuhl FAPS der FAU wird im Rahmen des Forschungsprojekts LOMOBI ein neuartiges, interaktives Assistenzsystem zur Förderung von Orientierung und Mobilität (O&M) für sehbeeinträchtigte Personen entwickelt. Ziel ist es, eine intuitive und adaptive Unterstützung im Alltag bereitzustellen, welche über die Möglichkeiten konventioneller Mobilitätshilfen hinausgeht. Die Kommunikation von Navigationsinformationen erfolgt dabei über interaktive Schnittstellen, die sich flexibel an die Fähigkeiten und Präferenzen der Nutzenden anpassen lassen. Im Fokus des Projekts steht die Entwicklung spezialisierter Methoden des Maschinellen Lernens zur effizienten Umgebungsinterpretation sowie die adaptive Gestaltung der Mensch-System-Interaktion zur Steigerung der Selbstständigkeit und Lebensqualität der Betroffenen.

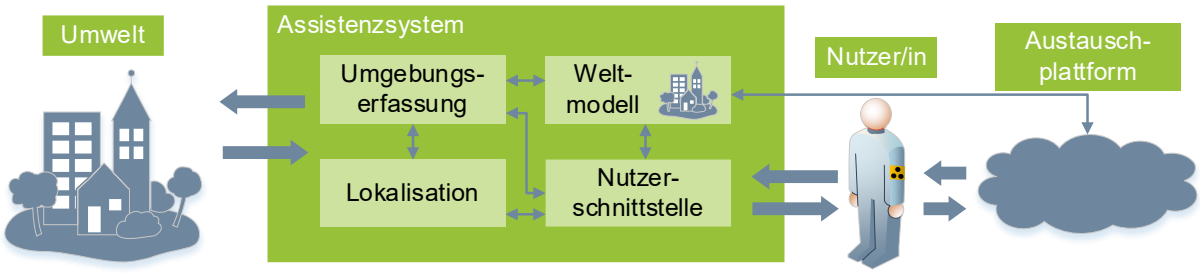


Abbildung 1: Darstellung der Interaktion des Assistenzsystems mit Umwelt und sehbehinderter Person zur Unterstützung von Orientierung und Mobilität.

1 Aufgabenstellung

Weltweit waren 2020 43,2 Mio. Menschen von Blindheit betroffen und 295,3 Mio. Menschen litten unter milden bis schwerwiegenden Sehbeeinträchtigungen. In Europa wurden 2,6 Mio. blinde und 26,5 Mio. mittel bis schwerwiegend sehbeeinträchtigte Personen geschätzt. Bis zum Jahr 2050 wird weltweit ein Anstieg auf 895,5 Mio. Menschen mit Sehbeeinträchtigungen prognostiziert. [1]

Die Beeinträchtigung der optischen Wahrnehmung stellt Betroffene im Alltag vor vielfältige Herausforderungen, die sich oftmals durch eine eingeschränkte Teilhabe am alltäglichen Leben und einhergehend einer verminderten Lebensqualität äußern [2]. Orientierung und Mobilität (O&M) ist eine entscheidende Fähigkeit zur Förderung der Selbständigkeit im Alltag und wird entsprechend intensiv in strukturierten Mobilitätsschulungen trainiert. Hierbei wird primär der Umgang mit Mobilitätshilfen wie einem Langstock oder Blindenhund durch geschulte RehabilitationslehrerInnen im alltäglichen Umfeld erlernt. Dennoch weisen die Methoden gewisse Einschränkungen auf. So können Hindernisse mit dem Langstock nur in unmittelbarer Nähe erfasst werden, was die O&M vor allem in unbekannten und veränderten Umgebungen zu einer großen Herausforderung werden lässt. Hier sind blinde und sehbeeinträchtigte Personen häufig auf die Unterstützung sehender Personen angewiesen. [3]

Neben den klassischen Mobilitätshilfen können sich Betroffene bei Bedarf auch ergänzende Hilfsmittel von einem Arzt verschreiben und die Kosten durch die Krankenkasse übernehmen lassen. Mögliche Systeme sind hier im Hilfsmittelkatalog des Spitzenverbandes Bund der Krankenkassen (GKV) eingetragen. Die verfügbaren Systeme dienen jedoch lediglich als situationsabhängige Ergänzung konventioneller Hilfsmittel und ermöglichen keine vollständig selbstständige Mobilität. Das folgende Kapitel liefert einen detaillierten Überblick über kommerziell verfügbare Lösungen des deutschen

Hilfsmittelverzeichnisses und den Stand der Forschung. Es ist bislang kein Assistenzsystem bekannt, das eine umfassende Unterstützung von O&M ermöglicht. [4]

Das Forschungsvorhaben LOMOBI adressiert diese Herausforderung, indem ein interaktives Assistenzsystem zur umfassenden Unterstützung von O&M im Alltag sehbeeinträchtigter Personen entwickelt wird. Dabei liegen die Forschungsschwerpunkte auf der Gestaltung eines interaktiven Assistenzsystems, zur Kommunikation von Navigationsinformationen über omnidirektionale interaktive Kommunikationsschnittstellen zwischen Nutzenden, dem System und der Umwelt. Die Navigationsinformationen können dabei an unterschiedlichen Stellen der Verarbeitungskette entsprechend der Fähigkeiten und Präferenzen der Nutzenden kommuniziert werden. Die Teilaufgaben der Navigation, wie Lokalisation, Umgebungsinterpretation und Pfadplanung werden spezifisch für unterschiedliche Szenarien (z. B. öffentliche Gebäude, urbane Umgebungen) konzipiert und je nach aktueller Umgebung genutzt. Der für die sichere Navigation essentielle und rechenintensive Aspekt der Umgebungsinterpretation wird mittels spezialisierter und dadurch möglichst effizienter Methoden des Maschinellen Lernens (Künstliche Neuronale Netze) umgesetzt. Das System wird so einerseits lernfähig konzeptioniert, um an unterschiedliche Gegebenheiten angepasst werden zu können und andererseits adaptiv gestaltet, um an die Präferenzen des Nutzenden angepasst zu werden und eine intuitive Nutzung zu ermöglichen.

Eine Austauschplattform ermöglich die Vernetzung und Interaktion der sehbeeinträchtigten Personen und den Austausch von Daten und Modellen. So können z. B. je nach Umgebung spezialisierte Modelle zur Umgebungssegmentierung kommuniziert oder spezifische Karten zur Lokalisation genutzt werden. Die transferierten Daten können auf leistungsfähigen Cloudservern automatisiert segmentiert oder durch sehende Nutzer manuell segmentiert und zur Optimierung der spezialisierten Modelle genutzt werden. Zudem können sehende Personen bei einem Systemausfall zur Unterstützung als Fall-Back Strategie herangezogen werden.

2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Das Forschungsvorhaben LOMOBI - Lernfähiges Assistenzsystem für Orientierung und Mobilität von sehbeeinträchtigten Personen im Alltag wurde im Zeitraum vom 01.02.2023 bis zum 30.04.2025 durch die beiden wissenschaftlichen Mitarbeitenden Matthias Kalenberg und Helmut Engelhardt (E13) unter Leitung von Prof. Dr.-Ing. Jörg Franke am Lehrstuhl für Fertigungsautomatisierung und Produktionssystematik (FAPS) der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) durchgeführt.

3 Planung und Ablauf des Vorhabens

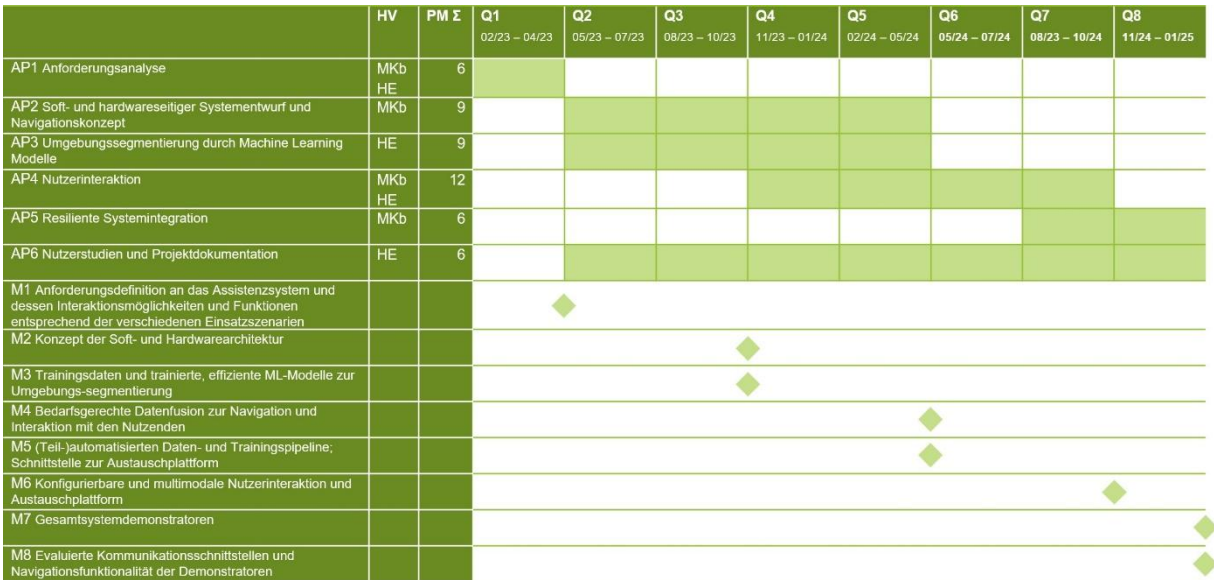


Abbildung 2: Darstellung des Arbeitsplans für das Projekt LOMOBI.

Die Anforderungsanalyse bildete im Rahmen des Gesamtvorhabens LOMOBI das erste Arbeitspaket (AP 1) und wurde zu Beginn des Projekts durchgeführt. Dieser Prozess orientierte sich am User-Centered-Design und dem Prinzip der Lebensweltorientierung. Ein zentraler Bestandteil dieser Phase war die Durchführung eines ELSI-Workshops (Ethische, Legale und Soziale Implikationen). Dabei kam das etablierte MEESTAR-Modell zum Einsatz, um ethische, rechtliche und soziale Problemstellungen sowie deren Implikationen frühzeitig zu identifizieren und zu diskutieren. Ziel des Workshops war es, durch den Austausch zwischen Entwicklern und Betroffenen ein ganzheitliches und nutzerorientiertes Bild der Mensch-Technik-Interaktion zu erhalten. Für den Workshop haben wir drei sehbeeinträchtigte Personen gewinnen können, die zusätzlich das Zentrum für selbstbestimmtes Leben Behinderter e. V. als auch den Bayerischen Blinden- und Sehbehindertenbund e.V. vertreten. Damit konnten Einblicke sowohl in die individuelle Perspektive als auch gesellschaftliche und organisatorische gewonnen werden. Die Diskussionen gliederten sich in die Ebenen der individuellen, gesellschaftlichen und organisatorischen Sichtweise, um differenzierte Problemprofile zu erhalten. Parallel dazu wurden auf Basis der gewonnenen Erkenntnisse Personas abgeleitet, welche spezifische Personen der sehbeeinträchtigten Nutzergruppe charakterisieren. Diese Personas dienten den Projektbeteiligten als Grundlage für die Ableitung spezifischer technischer und interaktiver Funktionen und Anforderungen an das Assistenzsystem. Das Ergebnis dieser initialen Analyse war die Anforderungsdefinition für das Assistenzsystem sowie dessen Interaktionsmöglichkeiten und Funktionen.

Die darauffolgenden Arbeitspakete AP 2 („Soft- und hardwareseitiger Systementwurf und Navigationskonzept“) und AP 3 („Umgebungssegmentierung durch Machine Learning Modelle“) wurden parallel in den Quartalen Q2 bis Q5 durchgeführt und bildeten die Grundlage für die technische Umsetzung des interaktiven Assistenzsystems. Im Rahmen des AP 2 wurde eine modulare Systemarchitektur entworfen, welche die bedarfsgerechte Fusionierung der Datenverarbeitungsprozesse ermöglicht. Das Navigationskonzept adaptierte Prinzipien autonomer, mobiler Systeme und umfasste die Lokalisierung, Pfadplanung und die Generierung von Feedback-

Nachrichten. Ein zentrales, aus dem Workshop abgeleitetes Erfordernis war die Berücksichtigung von externem Kartenmaterial in der globalen Pfadplanung. Dementsprechend wurden die Integration von GPS und eine Schnittstelle zu Open Street Map (OSM) realisiert, um sowohl die Lokalisierung zu unterstützen als auch die Möglichkeit für alternative Routen bei der Pfadplanung zu berücksichtigen, was als essenzielle Nutzeranforderung identifiziert wurde. Parallel dazu konzentrierte sich AP 3 auf die robuste und effiziente Segmentierung des Wegverlaufs und von Hindernissen. Die Umgebungserfassung erfolgte mittels spezialisierter Machine-Learning-Modelle, wobei die Anforderungen an die Detektion von Objekten oberhalb des Blindenstocks (d. h. auf Kopf- und Oberkörperhöhe) sowie die Erkennung von Löchern oder Abstufungen im Bodenbereich adressiert wurden, um die Hindernisvermeidung über die gesamte Körperfläche zu gewährleisten. Die Anpassung an vielfältige Umgebungen (wie Indoor, Outdoor Trails, Urban Streets) wurde durch Umgebungsklassifizierung und die Optimierung der ML-Modelle für eingebettete GPUs sichergestellt. Schließlich wurde die multimodale Kommunikation der Navigations- und Feedback-Signale modular konzipiert, um variable Feedback-Modalitäten (wie haptisches Feedback über Vibrationssignale oder akustische Signale) situations- und bedarfsgerecht an die Nutzenden zu vermitteln.

Die Implementierung der Nutzerinteraktion stellte das Hauptziel des Arbeitspakets 4 dar. Ein wesentliches Ziel dieses APs war die Integration der interaktiven, adaptiven Kommunikationsschnittstellen in ein modulares Navigationsframework. Ein zentrales Designkriterium, das aus dem ELSI-Workshop hervorging, ist die Notwendigkeit einer möglichst variablen Schnittstelle und hohen Individualisierung. Dies resultierte aus der Erkenntnis, dass unterschiedliche Personen unterschiedliche Arten der Informationsübermittlung bevorzugen. Um dieser Anforderung gerecht zu werden, wurden verschiedene Kanäle adressiert, wobei die vielschichtigen Schnittstellen eine bedarfsgerechte und flexible Kommunikation ermöglichen sollten. Die taktile Komponente wurde durch die Integration eines vibrotaktilen Gürtels zur sensorischen Augmentation für die Navigation von sehbehinderten Personen fokussiert. Dadurch können Virbationssignale zu Richtungsanweisungen und zur Hindernisvermittlung implementiert werden. Ergänzend dazu wurde die Umgebungswahrnehmung durch auditive Signale substituiert. Die Umfrageergebnisse des ELSI-Workshops priorisierten die Integration von Knochenkopfhörern. Diese optionalen Kopfhörer ermöglichen die Ausgabe von akustischen Signalen und Sprachbefehlen, während die wichtige auditive Wahrnehmung der Umgebung weiterhin gewährleistet bleibt. Die resultierenden Schnittstellen umfassen somit taktiles und akustisches Feedback und sind auf die Präferenzen der Nutzenden adaptierbar.

Die resiliente Systemintegration (AP 5) bildete den Abschluss der Entwicklungsphase. Hierbei wurden die interaktiven, adaptiven Kommunikationsschnittstellen und die Navigationsfähigkeiten in ein modulares Navigationsframework zusammengeführt. Die Evaluation des Gesamtsystems erfolgte parallel dazu im AP 6, dessen Ziel die Evaluierung der Kommunikationsschnittstellen und Navigationsfunktionalität war. Die Probandenstudien wurden sowohl in einer kontrollierten Simulationsumgebung als auch in einer Laborumgebung durchgeführt, in der die Teilnehmer entweder blind einen Roboter steuerten oder durchs Labor navigiert wurden.

4 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Das Projekt LOMOBI knüpfte in seiner Entwicklung an einen wissenschaftlichen und technischen Stand an, der die dringende Notwendigkeit eines umfassenden, lernfähigen Assistenzsystems für sehbeeinträchtigte Personen im Alltag verdeutlichte. Die steigende Zahl sehbeeinträchtigter Menschen korreliert mit einem wachsenden Bedarf an geeigneten Alternativen zur herkömmlichen Navigation. Bislang dienen kommerziell verfügbare Lösungen, die im Hilfsmittelkatalog eingetragen sind, oft nur als situationsabhängige Ergänzung zu konventionellen Hilfsmitteln wie dem Langstock oder dem Blindenführhund, ohne eine vollständig selbstständige Mobilität zu ermöglichen (z. B. Ultra Body Guard [5], Laser Langstock [6]). Weitere Möglichkeiten bieten Smartphones mit Sprachein- und Ausgabe zur Kommunikation [7]. Entsprechend sind auch spezielle Apps zur Unterstützung bestimmter O&M-Aspekte außerhalb [8], [9] oder innerhalb [10] geschlossener Räume verfügbar. Auch kann durch Smartphones die Unterstützung sehender Personen genutzt werden [11]. Des Weiteren sind Systeme verfügbar, die bei der Umgebungserfassung unterstützen. Z. B. Texte lesen oder Objekte lokalisieren (ORCam [12]). Das System BipedAI unterstützt sehbeeinträchtigte Personen durch eine Hindernisvermeidung basierend auf Tiefenkameras und einem Audio Feedback [13]. Diese Hilfsmittel sind in ihrer Anwendung jeweils begrenzt und es zeigt sich, dass kein umfassendes System gibt, das eine vollständige Navigation sowohl im Indoor als auch im Outdoor-Bereich bietet und dabei die verschiedenen Wege segmentiert und die verschiedenen Lösungen bedarfsgerecht kombiniert.

Die Lokalisierung außerhalb geschlossener Räume basiert primär auf der Satellitennavigation (GNSS), welche Systeme wie GPS, GLONASS, Galileo und BeiDou umfasst. Die Genauigkeit schwankt je nach Örtlichkeit und kann in dicht bebauten Gebieten oder bewaldeten Regionen Abweichungen von fünf bis zu 50 Metern aufweisen. Um diese Ungenauigkeiten zu kompensieren und die Pose (Position und Orientierung) dynamisch zu erfassen, sind Kamerabilder, Visual Odometry (VO) und Inertial Measurement Units (IMUs) notwendig, die akkumulierten Drift entgegenwirken und Signalausfälle überbrücken. Innerhalb geschlossener Räume ist GNSS in der Regel nicht verfügbar. Hier sind entweder spezielle Infrastrukturen (wie WLAN oder BLE) oder die Kartierung und Lokalisierung mittels Simultaneous Localization and Mapping (SLAM) erforderlich. [14] Softwarepakete wie RTAB-Map SLAM ermöglichen dabei die gleichzeitige Kartenerstellung und Positionsbestimmung in Echtzeit [15]. Für eine situationsgerechte Navigation in beiden Szenarien ist die Sensordatenfusion (SDF) zentral, welche Daten von GNSS, VO und IMU (häufig mittels Kalman-Filter) kombiniert, um eine robuste und präzisere Positionsbestimmung zu erzielen. Diese hybriden Navigationslösungen nutzen GNSS und externe Karten (wie OpenStreetMap) für die globale Routenplanung und ergänzen diese durch SLAM und SDF für die detaillierte lokale Navigation.

Cheng et al. [16] nutzt ML-Verfahren zur detaillierten Klassifikation von Umgebungen, wobei die gewonnenen Informationen jedoch nicht zur Unterstützung der Navigation herangezogen werden. Zhang et al. [17] setzen eine Transformer-Architektur zur Detektion und Segmentierung von transparenten Objekten wie Glasscheiben ein, um sehbeeinträchtigte Personen zu unterstützen. Die hierbei erfassten Umgebungsinformationen werden der nutzenden Person allerdings nur über ein

rudimentäres Sprachfeedback übermittelt und nicht mit Lokalisations- oder Pfadplanungsmethoden verknüpft. Ein Framework zur Fusion von Lokalisations- und Umgebungsinformationen wurde von Elmannai et al. [18] konzipiert, um Orientierung und Mobilität (O&M) zu unterstützen. Bei diesem Ansatz erfolgt die Bildverarbeitung jedoch auf einem Onlineserver, was eine gewisse Abhängigkeit und Anfälligkeit schafft. Zudem werden dabei nur sehr große Hindernisse detektiert, und die Fusionierung ermöglicht lediglich eine sehr rudimentäre Pfadplanung. Bai et al. [19] verwenden ebenfalls ein Framework zur Fusionierung von GNSS und 3D-Kameras. Hierbei ist die Bildverarbeitung jedoch primär für Indoor-Szenarien konzipiert und kann nur bekannte Objekte (wie Stühle oder Personen) detektieren. Das Konzept zeigt Defizite in gering strukturierten Szenerien und bei unspezifischen Hindernissen auf.

Angesichts dieser Einschränkungen besteht eine Forschungslücke in der Entwicklung eines einzelnen, umfassenden Systems LOMOBI zur intuitiven Unterstützung von O&M sowohl in Indoor- als auch Outdoor-Bereichen mit einer adaptiven Wegsegmentierung für unterschiedliche Umgebungen.

5 Angabe bekannter Konstruktionen, Verfahren und Schutzrechte, die für die Durchführung des Vorhabens benutzt wurden

Der feelSpace naviGürtel wurde im LOMOBI-Projekt zur Implementierung von haptischem Feedback und zur sensorischen Augmentation in Form der Hinderniserkennung in Probandenstudien evaluiert, wobei die Steuerung über ROS2-Nodes unter Verwendung der PyBelt-Bibliothek erfolgte.

6 Angabe der verwendeten Fachliteratur sowie der benutzten Informations- und Dokumentationsdienste

Fachliteratur wurde über die Universitätsbibliothek der Friedrich-Alexander-Universität, sowie über Google Scholar, Scopus und PubMed recherchiert. Der Zugang erfolgte im Rahmen von Open Access oder über die Lizenzen der Friedrich-Alexander-Universität. Zur Erstellung von Dokumenten wurden im Rahmen des Forschungsvorhabens LOMOBI folgende Softwaretools genutzt:

- Office 2019 (Microsoft)
- Zotero (Corporation for Digital Scholarship)
- Citavi (Swiss Academic Software)
- Nextcloud (Nextcloud)
- Creo Paramateric (PTC Creo)

Verwendete Fachliteratur:

[1] BOURNE, R., J.D. STEINMETZ, S. FLAXMAN, P.S. BRIANT, H.R. TAYLOR, S. RESNIKOFF, R.J. CASSON, A. ABDOLI, E. ABU-GHARBIEH, A. AFSHIN, H. AHMADIEH, Y. AKALU, A.A. ALAMNEH, W. ALEMAYEHU, A.S. ALFAAR, V. ALIPOUR, E.W. ANBESU, S. ANDROUDI, J. ARABLOO, A. ARDITI, M. ASAAD, E. BAGLI, A.A. BAIG, T.W. BÄRNIGHAUSEN, M. BATTAGLIA PARODI, A.S. BHAGAVATHULA, N. BHARDWAJ, P. BHARDWAJ, K. BHATTACHARYYA, A. BIJANI, M. BIKBOV, M. BOTTONE, T. BRAITHWAITE, A.M. BRON, Z.A. BUTT, C.-Y. CHENG, D.-T. CHU, M.V. CICINELLI, J.M. COELHO, B. DAGNEW, X. DAI, R. DANA, L. DANDONA, R. DANDONA, M.A. DEL MONTE,

J.P. DEVA, D. DIAZ, S. DJALALINIA, L.E. DREER, J.R. EHRLICH, L.B. ELLWEIN, M.H. EMAMIAN, A.G. FERNANDES, F. FISCHER, D.S. FRIEDMAN, J.M. FURTADO, A.M. GAIDHANE, S. GAIDHANE, G. GAZZARD, B. GEBREMICHAEL, R. GEORGE, A. GHASHGHAEE, M. GOLECHHA, S. HAMIDI, B.R. HAMMOND, M.E.R. HARTNETT, R.K. HARTONO, S.I. HAY, G. HEIDARI, H.C. HO, C.L. HOANG, M. HOUSEH, S.E. IBITOYE, I.M. ILIC, M.D. ILIC, A.D. INGRAM, S.S.N. IRVANI, R.P. JHA, R. KAHLOUN, H. KANDEL, A.S. KASA, J.H. KEMPEN, M. KERAMATI, M. KHAIRALLAH, E.A. KHAN, R.C. KHANNA, M.N. KHATIB, J.E. KIM, Y.J. KIM, S. KISA, A. KISA, A. KOYANAGI, O.P. KURMI, C. VAN LANSINGH, J.L. LEASHER, N. LEVEZIEL, H. LIMBURG, M. MAJDAN, N. MANAFI, K. MANSOURI, C. MCALINDEN, S.F. MOHAMMADI, A. MOHAMMADIAN-HAFSHEJANI, R. MOHAMMADPOURHODKI, A.H. MOKDAD, D. MOOSAVI, A.R. MORSE, M. NADERI, K.S. NAIDOO, V. NANGIA, C.T. NGUYEN, H.L.T. NGUYEN, K. OGUNDIMU, A.T. OLAGUNJU, S.M. OSTROFF, S. PANDA-JONAS, K. PESUDOVS, T. PETO, Z. QUAZI SYED, M.H.U. RAHMAN, P.Y. RAMULU, S. RAWAF, D.L. RAWAF, N. REINIG, A.L. ROBIN, L. ROSSETTI, S. SAFI, A. SAHEBKAR, A.M. SAMY, D. SAXENA, J.B. SERLE, M.A. SHAIKH, T.T. SHEN, K. SHIBUYA, J.I. SHIN, J.C. SILVA, A. SILVESTER, J.A. SINGH, D. SINGHAL, R.S. SITORUS, E. SKIADARESI, V. SKIRBEKK, A. SOHEILI, R.A.R.C. SOUSA, E.E. SPURLOCK, D. STAMBOLIAN, B.W. TADDELE, E.G. TADESSE, N. TAHHAN, M.I. TAREQUE, F. TOPOUZIS, B.X. TRAN, R.S. TRAVILLIAN, M.K. TSILIMBARIS, R. VARMA, G. VIRGILI, Y.X. WANG, N. WANG, S.K. WEST, T.Y. WONG, Z. ZAIDI, K.A. ZEWDIE, J.B. JONAS und T. VOS. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study [online]. *The Lancet Global Health,* 2021, **9**(2), S. e130-e143. ISSN 2214109X. Verfügbar unter: doi:10.1016/S2214-109X(20)30425-3

[2] LANGELAAN, M., M.R. de BOER, R.M.A. VAN NISPEN, B. WOUTERS, A.C. MOLL und G.H.M.B. VAN RENS. Impact of visual impairment on quality of life: a comparison with quality of life in the general population and with other chronic conditions [online]. *Ophthalmic epidemiology,* 2007, **14**(3), S. 119-126. ISSN 0928-6586. Verfügbar unter: doi:10.1080/09286580601139212

[3] VIRGILI, G. und G. RUBIN. Orientation and mobility training for adults with low vision [online]. *The Cochrane database of systematic reviews,* 2010, (5), S. CD003925. Verfügbar unter: doi:10.1002/14651858.CD003925.pub3

[4] GKV SPITZENVERBAND. *Hilfsmittelverzeichnis des GKV-Spitzenverbandes - Blindenhilfsmittel* [online] [Zugriff am: 12. Januar 2020]. Verfügbar unter: https://hilfsmittel.gkv-spitzenverband.de/HimiWeb/produktgruppeAnzeigen_input.action?gruppeId=7

[5] HILFSMITTELVERZEICHNIS DES GKV-SPITZENVERBANDES. *Ultra Body Guard - Rehabilitations-Technik Broer GmbH & Co. KG* [online]. 16 Februar 2021 [Zugriff am: 16. Februar 2021]. Verfügbar unter: https://hilfsmittel.gkv-spitzenverband.de/HimiWeb/produktAnzeigen_input.action?produktId=3765

[6] HILFSMITTELVERZEICHNIS DES GKV-SPITZENVERBANDES. *Laser Langstock - VISTAC GmbH* [online]. 16 Februar 2021 [Zugriff am: 16. Februar 2021]. Verfügbar unter: https://hilfsmittel.gkv-spitzenverband.de/HimiWeb/produktAnzeigen_input.action?produktId=54444

[7] GOOGLE INC. *Android-Bedienungshilfen - Bedienungshilfen für Android-Hilfe* [online]. 23 Februar 2021 [Zugriff am: 23. Februar 2021]. Verfügbar unter: https://support.google.com/accessibility/android/answer/6006564?hl=de&ref_topic=6007234

[8] ROUTAGO. *Routago - Assist - Sichere Fußgängernavigation* [online]. 18 Februar 2021 [Zugriff am: 23. Februar 2021]. Verfügbar unter: https://www.routago.de/

[9] *Was ist BlindSquare? – BlindSquare* [online]. 16 Oktober 2025 [Zugriff am: 16. Oktober 2025]. Verfügbar unter: https://www.blindsquare.com/de/about/

[10] FRAUNHOFER FOKUS. *Indoor-Navigation* [online]. *Hochpräzise Indoor-Navigation auf dem Smartphone.* 23 Februar 2021 [Zugriff am: 23. Februar 2021]. Verfügbar unter: https://www.fokus.fraunhofer.de/go/indoor-navigation

[11] *Be My Eyes - See the world together* [online]. 12 Januar 2022 [Zugriff am: 12. Januar 2022]. Verfügbar unter: https://www.bemyeyes.com/

[12] ORCAM GMBH. *Hilfe für Blinde und Menschen mit eingeschränktem Sehvermögen* [online] [Zugriff am: 12. November 2019]. Verfügbar unter: https://www.orcam.com/de/

[13] *NOA by biped.ai, Live AI mobility companion* [online]. 24 September 2025 [Zugriff am: 16. Oktober 2025]. Verfügbar unter: https://www.biped.ai/

[14] FERNANDES, H., P. COSTA, V. FILIPE, H. PAREDES und J. BARROSO. A review of assistive spatial orientation and navigation technologies for the visually impaired [online]. *Universal Access in the Information Society,* 2019, **18**(1), S. 155-168. ISSN 1615-5289. Verfügbar unter: doi:10.1007/s10209-017-0570-8

[15] LABBÉ, M. und F. MICHAUD. Multi-Session Visual SLAM for Illumination-Invariant Re-Localization in Indoor Environments [online]. *Frontiers in robotics and AI,* 2022, **9**, S. 801886. Verfügbar unter: doi:10.3389/frobt.2022.801886

[16] CHENG, R., K. WANG, J. BAI und Z. XU. Unifying Visual Localization and Scene Recognition for People With Visual Impairment [online]. *IEEE Access,* 2020, **8**, S. 64284-64296. Verfügbar unter: doi:10.1109/ACCESS.2020.2984718

[17] ZHANG, J., K. YANG, A. CONSTANTINESCU, K. PENG, K. MULLER und R. STIEFELHAGEN. Trans4Trans: Efficient Transformer for Transparent Object and Semantic Scene Segmentation in Real-World Navigation Assistance [online]. *IEEE Transactions on Intelligent Transportation Systems,* 2022, S. 1-14. ISSN 1524-9050. Verfügbar unter: doi:10.1109/TITS.2022.3161141

[18] ELMANNAI, W.M. und K.M. ELLEITHY. A Highly Accurate and Reliable Data Fusion Framework for Guiding the Visually Impaired [online]. *IEEE Access,* 2018, **6**, S. 33029-33054. Verfügbar unter: doi:10.1109/ACCESS.2018.2817164

[19] BAI, J., Z. LIU, Y. LIN, Y. LI, S. LIAN und D. LIU. Wearable Travel Aid for Environment Perception and Navigation of Visually Impaired People [online]. *Electronics,* 2019, **8**(6), S. 697. Verfügbar unter: doi:10.3390/electronics8060697

[20] RONNEBERGER, O., P. FISCHER und T. BROX. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: N. NAVAB, J. HORNEGGER, W.M. WELLS und A.F. FRANGI, Hg. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* Cham: Springer International Publishing, 2015, S. 234-241. ISBN 978-3-319-24573-7

[21] ENGELHARDT, H., M. KALENBERG, J. FRANKE und S. MARTIN. Learn Where I Can Walk: Auto-labeling of Walked Areas Using Monocular Camera Trajectory. In: J. RÖNING und J. FILIPE, Hg. *Robotics, Computer Vision and Intelligent Systems.* Cham: Springer Nature Switzerland, 2026, S. 348-363. ISBN 978-3-032-00985-2

[22] RAVI, N., V. GABEUR, Y.-T. HU, R. HU, C. RYALI, T. MA, H. KHEDR, R. RÄDLE, C. ROLLAND, L. GUSTAFSON, E. MINTUN, J. PAN, K.V. ALWALA, N. CARION, C.-Y. WU, R. GIRSHICK, P. DOLLÁR und C. FEICHTENHOFER. *SAM 2: Segment Anything in Images and Videos:* arXiv, 2024

[23] REN, T., S. LIU, A. ZENG, J. LIN, K. LI, H. CAO, J. CHEN, X. HUANG, Y. CHEN, F. YAN, Z. ZENG, H. ZHANG, F. LI, J. YANG, H. LI, Q. JIANG und L. ZHANG. *Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks:* arXiv, 2024

[24] WANG, X., X. ZHANG, Y. CAO, W. WANG, C. SHEN und T. HUANG. *SegGPT: Segmenting Everything In Context:* arXiv, 2023

[25] SARLIN, P.-E., D. DETONE, T.-Y. YANG, A. AVETISYAN, J. STRAUB, T. MALISIEWICZ, S.R. BULO, R. NEWCOMBE, P. KONTSCHIEDER und V. BALNTAS. OrienterNet: Visual Localization in 2D Public Maps with Neural Matching. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR):* IEEE, 2023, S. 21632-21642. ISBN 979-8-3503-0129-8

7 Zusammenarbeit mit anderen Stellen

Die Zusammenarbeit mit anderen Stellen umfasste die Planung eines Workshops zu Chancen und Risiken der LOMOBI-Anwendung in Kooperation mit Verena Müller und Philipp Mehl von der Technischen Universität München (TUM). Ziel war es, mögliche spätere Nutzer und Betroffene einzuladen, um gemeinsam deren Hoffnungen, Sorgen und Fragen zur Technologie zu diskutieren. Die gesammelte Einschätzung sollte maßgeblich dazu beitragen, sichere und wirklich hilfreiche Neuerungen zu entwickeln, und die Ergebnisse sollen in einer wissenschaftlichen Veröffentlichung zusammengefasst werden.

8 Ergebnis

Im folgenden Abschnitt werden die Ergebnisse des Forschungsprojekts LOMOBI vorgestellt. Zur besseren Übersichtlichkeit erfolgt keine chronologische Darstellung, sondern entlang des gesamten Entwicklungszyklus, von der Anforderungsanalyse bis zur Evaluation. Aus ELSI-Workshop, Personas und Umgebungsanalyse wurden priorisierte Anforderungen abgeleitet und in einem ROS-basierten Assistenzsystems mit Umgebungsklassifikation und -segmentation, Sensorfusion (VO/IMU/GNSS), hierarchischer Pfad- und Ablaufplanung sowie barrierearmer multimedialer Interaktion (App, Vibration, 3D-Audio) umgesetzt. Probandenstudien bestätigen die Nutzbarkeit und Wirksamkeit der multimodalen Interaktion und zeigen gezielte Verbesserungsfelder.

8.1 Anforderungsanalyse

Für die Definition von ELSI-Aspekten ist zu Beginn des Projektes ein ELSI-Workshop nach dem MEESTAR Modell durchgeführt worden. Das Ziel ist es dadurch Anforderungen an das zu entwickelnde LOMOBI-System abzuleiten. An dem Workshop nahmen 12 Personen teil, darunter drei Blinde. Jeder der beeinträchtigten Teilnehmer und Teilnehmerinnen hat dabei eine für sehbeeinträchtigte Menschen wichtige gesellschaftliche Position bei der Stadt oder in einem Verein. Dadurch spiegelt das Teilnehmerfeld nicht nur die Sicht der Betroffenen wider, sondern bietet auch Expertise in gesellschaftlichen Belangen, was für die Identifizierung von ELSI-Aspekten sehr wichtig ist. Die Whiteboards mit den identifizierten ELSI-Aspekten der verschiedenen Ebenen nach dem MEESTAR-Modell sind in den Abbildungen 1-3 dargestellt.

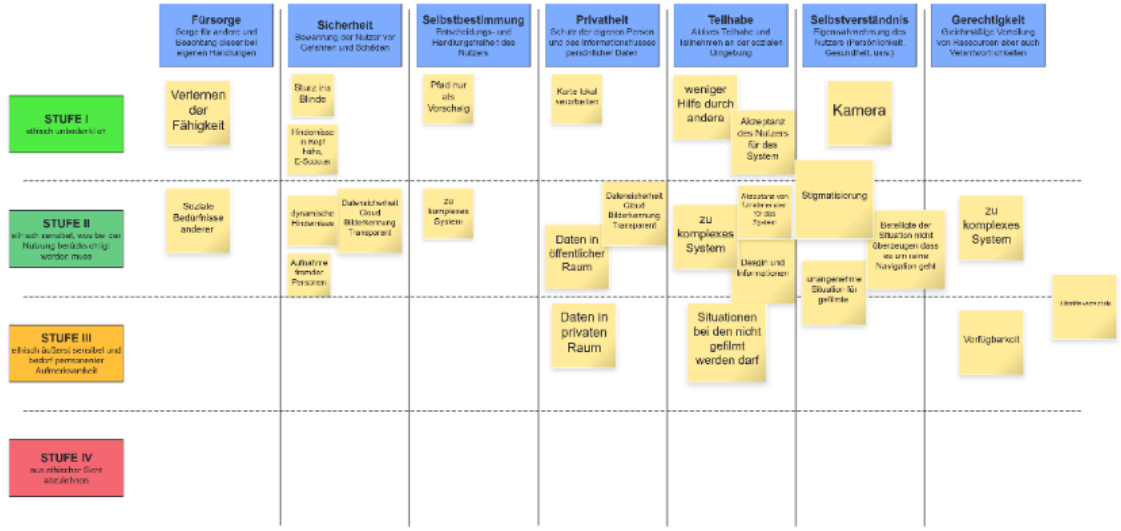


Abbildung 1: Individuelle Ebene des ELSI-Workshops basierend auf dem MEESTAR-Modell.

Abbildung 2: Gesellschaftliche Ebene des ELSI-Workshops basierend auf dem MEESTAR-Modell.

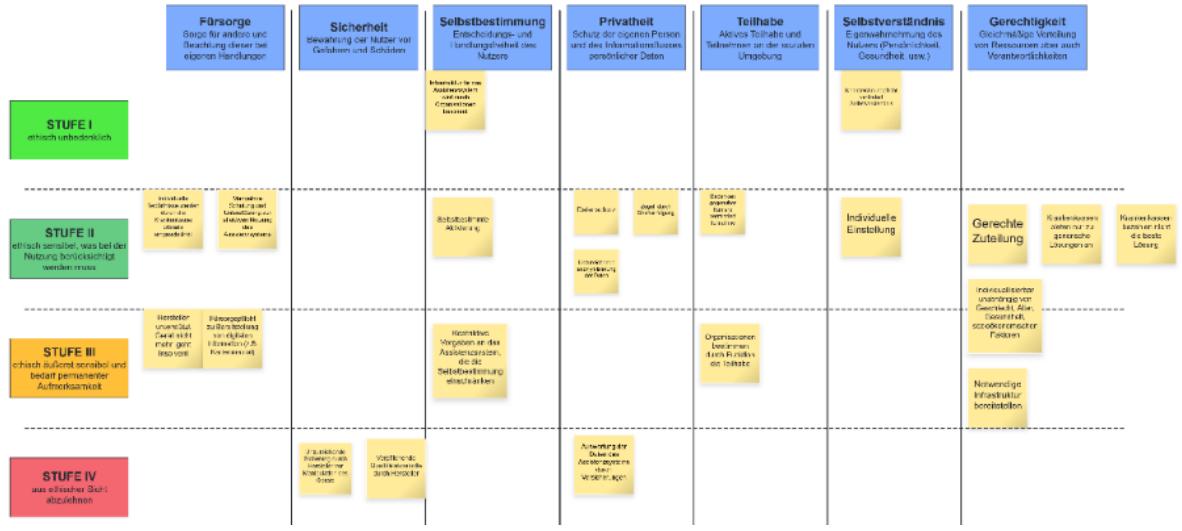


Abbildung 3: Organisatorische Ebene des ELSI-Workshops basierend auf dem MEESTAR-Modell.

Zusätzlich zu den Workshop Ergebnissen sind weitere Personas definiert worden, um weitere Perspektiven zur Eruierung potenzieller Einsatzszenarien des Assistenzsystems zu berücksichtigen. Dazu sind Personas nach den Eigenschaften Alter, Familienstand, Tätigkeit, Ausbildung, Körperliche Einschränkungen, Computer-Kenntnisse und Hobbys erstellt worden. Verschiedene potenzielle Navigationsumgebungen im Alltag für das LOMOBI-System sind definiert und nach dem Morphologischen Kasten in Tabelle 1 klassifiziert worden. Beispiele für Navigationsumgebungen sind:

- Fußgängerwege in bekannter, urbaner Umgebung
- Gering strukturierte Outdoor Umgebung
- Öffentliche, strukturierte Outdoor Umgebungen
- Outdoor Umgebungen mit Verkehr

Tabelle 1: Morphologischer Kasten zur Klassifizierung von Navigationsumgebungen

Eigenschaft	**Parameter 1**	**Parameter 2**	**Parameter 3**
Räumlichkeit	Indoor	Outdoor	
Bebauung	Urbane Bebauung	Keine Bebauung	
Bekanntheit	Bekannt	Unbekannt	
Wege	Autostraße	Verkehrsberuhigt	Gering strukturiert
Zugänglichkeit	Privat	Nicht öffentlich	Öffentlich
Personen	Sehr belebt	Normal belebt	Wenig belebt

Die abgeleiteten Anforderungen sind von den Workshop Teilnehmern nach ihrer Wichtigkeit bewertet worden. Die Liste der Anforderungen ist in Tabelle 2 dargestellt.

Tabelle 2: Liste der Anforderungen sortiert nach der Bewertung der Workshop Teilnehmer. Jede Funktion konnte von einem Teilnehmer 0-5 Punkte erhalten.

Funktion / Spezifikation	**Punktewert (von maximal 115 Punkten)**	**Punktewert sehbehinderter Befragter (von maximal 30 Punkten)**
Nutzerbestimmtes Ausschalten	100 Punkte	23 Punkte
Implementierung von Kartenmaterial öffentlicher Gebäude	97 Punkte	26 Punkte
*Erkennen von Hindernissen auf Kopfhöhe**	88 Punkte	23 Punkte
Modularer Aufbau der Funktionen	88 Punkte	20 Punkte
Mitteilung an Nutzer bei Ausfall des Systems	88 Punkte	24 Punkte

Bereitstellung von Alternativrouten	84 Punkte	28 Punkte
Warnung des Nutzers bei unsicherer Umgebungsanalyse	84 Punkte	20 Punkte
Einstellbare Scan-Distanz	83 Punkte	25 Punkte
Variable Stärke der Vibrationselemente	80 Punkte	16 Punkte
Automatisches Standby nach längerer Inaktivität	74 Punkte	26 Punkte
Bereitstellen des Quellcodes für Dritte	74 Punkte	18 Punkte
*Genaue Identifizierung von Hindernissen mit Informationsweitergabe an den Nutzer**	73 Punkte	18 Punkte
Integration von Knochenkopfhörern	72 Punkte	20 Punkte
Verschiedene wählbare Rücksackgrößen	68 Punkte	24 Punkte
Abschaltbare Vibrationselemente	52 Punkte	20 Punkte
Einschalten eines Fernassistenzsystems bei Ausfall des Systems	52 Punkte	8 Punkte
individuelle Platzierung der Vibrationselemente	52 Punkte	12 Punkte
Signalisierung an Außenstehende bei Ausfall des Systems	16 Punkte	0 Punkte

8.2 Soft- und hardwareseitiger Systementwurf und Navigationskonzept

Die Softwarearchitektur wird im Robot Operating System (ROS) implementiert, das die Integrierbarkeit und Wartbarkeit einzelner Softwaremodule gewährleistet. Die Architektur basiert auf den vier Hauptkomponenten Umgebungserfassung, Lokalisierung und Kartierung, Pfadplanung und Feedbackgenerierung, für die jeweils ein eigenes ROS-Paket implementiert wird.

Die Umgebungserfassung zur Weiterverarbeitung durch Bildverarbeitungsalgorithmen der künstlichen Intelligenz basiert auf Kamerabildern. Diese Verarbeitung wird in Kapitel 8.3 ausführlicher beschrieben. Zur Lokalisierung sind weitere Sensorprinzipien untersucht worden.

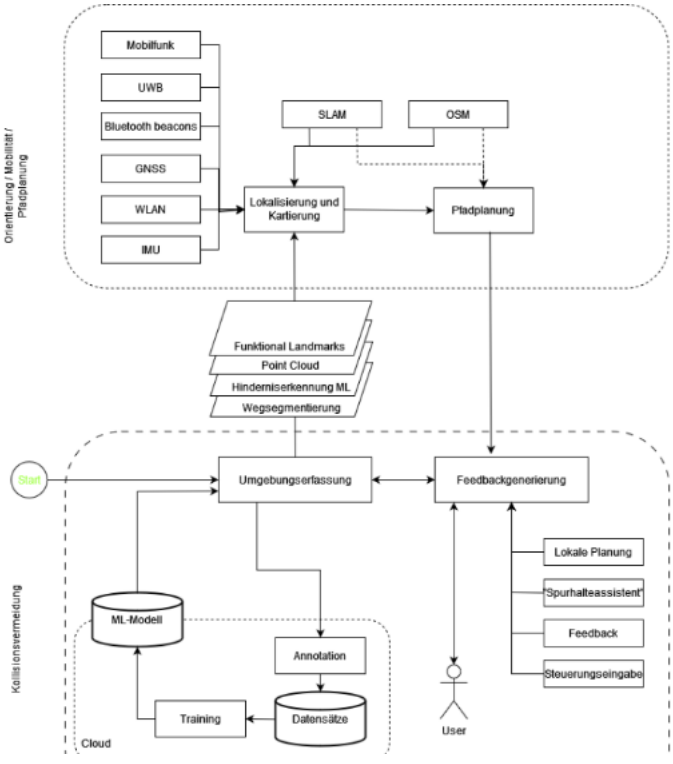


Abbildung 3: Übersicht des softseitigen Systementwurf und Navigationskonzept.

Grundlegend ist eine inertiale Messeinheit zur Bestimmung von Beschleunigungswerten. Darüber hinaus können im öffentlichen Raum die Mobilfunkdaten sowie WLAN- und GNSS-Daten zur globalen Lokalisierung genutzt werden. Für den privaten Indoor-Bereich sind noch Ultrabreitband oder Bluetooth-Technologien denkbar, würden aber eine Anpassung der Umgebung erfordern und werden deshalb zunächst nicht weiterverfolgt. Die Kameradaten können genutzt werden, um eine Karte von der Umgebung zu erstellen. Dafür ist ein Algorithmus zum simultanen Kartieren und Lokalisieren (SLAM) implementiert worden. Innerhalb dieser Karten kann sich das LOMOBI-System auf Basis der Kameradaten lokalisieren. Für eine möglichst robuste Lokalisierung werden die verschiedenen Sensoren im Laufe des Projektes mit Hilfe eines Kalman Filters fusioniert.

Für die in Kapitel 8.1 genannten Navigationsumgebungen wurden die verschiedenen Sensorprinzipien evaluiert. Insbesondere unterscheidet sich die Genauigkeit der Positionsdaten zwischen Indoor und Outdoor. Während im Indoor-Bereich keine zuverlässigen GNSS- und Magnetometriedaten erwartet werden können, wird der SLAM-Algorithmus für große Karten im öffentlichen Outdoor-Bereich ineffizient. ROS bietet standardisierte Schnittstellen zu den Sensoren, die je nach Bedarf mit der Umgebungsklassifikation aus Kapitel 8.3 fusioniert werden können. Innerhalb von ROS wurde eine Schnittstelle zu Open Streetmap implementiert, mit dem Ziel Wege in öffentlichen Umgebungen zu planen. Damit soll die Navigation auch in unbekannten öffentlichen Umgebungen ermöglicht werden.

8.3 Umgebungsklassifikation durch Machine Learning Modelle

Es erfolgte für die Klassifizierung der in Kapitel 8.1 identifizierten Navigationsumgebungen eine iterative Testung und Optimierung von vier verschiedenen künstlichen neuronalen Netzarchitekturen. Hierfür wurden verschiedene Sensor-Daten für die jeweiligen spezifischen Navigationsaufgaben aufgezeichnet und diese entsprechend klassifiziert und in Trainings- und Testdaten aufgeteilt. Zudem wurden die öffentlich verfügbaren Datensätze SUN und Places mit in die Trainings- und Testdaten integriert. Die ausgewählten Netzarchitekturen sind mit den Trainingsdaten trainiert und anhand der Testdaten auf deren Klassifizierungsgenauigkeit untersucht worden. Die besten Ergebnisse konnten hierbei mit SE-ResNeXt-50 erzielt werden, die Ergebnisse der Klassifizierungsgenauigkeiten sind in Tabelle 1 dargestellt.

Tabelle 1: Zusammenfassung der Genauigkeiten und F1-Scores für jede Klasse für die optimierten Algorithmen von EfficientNet, DenseNet, ResNeXt und SENet. Die besten Ergebnisse sind fett markiert.

	EfficientNet-B0		**DenseNet-121**		**ResNeXt-101**		**SE-ResNeXt-50**	
Klasse	Genauigkeit [%]	F1 Score [%]	Genauigkeit [%]	F1 Score [%]	Genauigkeit [%]	F1 Score [%]	Genauigkeit [%]	F1 Score [%]
Indoor	95,74	95,31	96,18	95,18	96,74	95,84	**96,84**	**95,95**
Privat	96,29	85,09	96,34	83,52	**96,80**	**85,62**	96,65	85,14
Outdoor Pfade	94,48	86,55	**95,37**	**87,94**	94,48	86,63	95,21	87,57

Städtische Straßen	93,96	87,72	94,13	87,97	94,11	87,45	**94,80**	**89,33**
Keine Straße	96,23	87,54	96,82	88,82	96,65	87,94	**96,96**	**89,64**
Viele Menschen	93,84	83,83	93,79	84,69	93,94	85,15	**94,36**	**85,21**
Wenig Menschen	91,23	83,51	91,78	84,37	92,02	84,91	**92,58**	**86,37**
Keine Menschen	93,30	93,79	94,44	94,8	94,86	95,16	**95,09**	**95,43**
Transportfahrzeug	99,30	88,62	99,31	86,46	99,31	88,06	**99,51**	**90,66**
Fahrzeuge	95,36	80,55	94,43	78,64	94,92	80,30	**95,70**	**82,53**
Durchschnitt	94,97	87,25	95,26	87,24	95,38	87,71	**95,77**	**88,78**

Für die beste Netzarchitekturen „SE-ResNeXt-50“ wurde ein ROS2 Knoten implementiert, welcher anhand der Kameradaten die Navigationsumgebung klassifiziert und diese über definierte ROS2 Nachrichtenschema, wie in Kapitel 8.2 beschrieben, an das Navigationsframework kommuniziert.

8.4 Nutzerinteraktion

Die Gestaltung einer intuitiven und adaptiven Nutzerinteraktion ist von zentraler Bedeutung, da das Assistenzsystem als kritische Schnittstelle zwischen der komplexen Umgebung (Umwelt), der internen Systemlogik und der nutzenden Person fungiert. Die Kommunikation der gewonnenen Navigationsinformationen, wie Wegverlauf, Hindernisse und Richtungsanweisungen, erfolgt über multimodale Kanäle. Hierbei kommen taktiles Feedback (über Vibrationsmodule) und akustische Signale (z. B. über Knochenschallkopfhörer) zum Einsatz.

Die Entwicklung der Kommunikationsschnittstellen und deren Interaktionsmöglichkeiten basiert auf einem nutzerzentrierten Ansatz (User-Centered-Design), bei dem die Bedürfnisse und Präferenzen der Zielgruppe frühzeitig durch Workshops und Studien ermittelt wurden. Dieser Ansatz gewährleistet die adaptive Gestaltung des Systems an die Fähigkeiten und Anforderungen der Nutzenden, was zur Steigerung der Nutzerfreundlichkeit und Akzeptanz beitragen soll.

8.4.1 Smartphone-App

Die für das LOMOBI-System entwickelte Smartphone-App dient als zentrale und barrierefreie Benutzerschnittstelle und wurde mithilfe des Cross-Plattform-Frameworks Flutter realisiert, was ihre Eignung für die Zielgruppe sehbeeinträchtigter Personen unterstreicht, da diese moderne Smartphones häufig nutzen [20]. Die App ermöglicht eine bidirektionale Kommunikation über WLAN mit dem ROS2-basierten LOMOBI-Assistenzsystem. Zu den Kernfunktionen, wie in Abbildung 4 dargestellt, gehören die Eingabe von Zielen, die Suche nach Orten sowie das Speichern von Favoriten. Darüber hinaus übermittelt die App kontinuierlich die aktuellen Standortdaten des Smartphones zur Verbesserung der Gesamtgenauigkeit der Lokalisierung an das LOMOBI-System. Zusätzliche

Konfigurationsmöglichkeiten umfassen die Anpassung der Lautstärke, der Sprechgeschwindigkeit der Sprachausgabe sowie die Wahl zwischen einem hellen und einem dunklen Farbschema der App.

Um eine hohe Zugänglichkeit zu gewährleisten, wurde die Benutzeroberfläche einfach gehalten und die Barrierefreiheit gemäß den „Web Content Accessibility Guidelines 2.1“ Prinzipien umgesetzt. Die App ist mit gängigen Screenreadern, wie dem TalkBack bei Android oder dem VoiceOver vom iPhone, kompatibel. Entscheidend für das Feedback ist ein akustischer Heartbeat-Mechanismus, der den Verbindungsstatus zum LOMOBI-System über Töne signalisiert, was essenziell ist, da die Nutzer das Smartphone nicht ständig bedienen. Ein separater Text-to-Speech (TTS)-Dienst der App liefert akustische Rückmeldungen über den Erfolg oder Misserfolg von Aktionen und liest zudem Status- oder Ereignisnachrichten vor, die das LOMOBI-System über ROS2 an die NutzerInnen sendet.

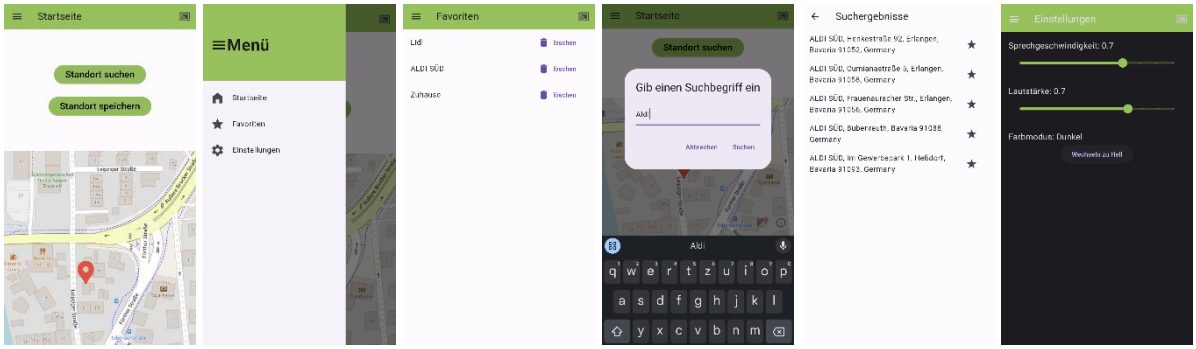


Abbildung 4: Bilder der Oberfläche der für das LOMOBI-System entwickelte Smartphone-App. Von links nach rechts: Startseite, Menüleiste, Suchzeile, Ergebnisliste der Suche, Favoritenliste und der Einstellungen in der LOMOBI-App.

8.4.2 Vibrotaktile Nutzerinteraktion

Das Assistenzsystem LOMOBI nutzt den naviGürtel von feelSpace als zentrale Komponente für die vibrotaktile Nutzerinteraktion. Dieser Gürtel verfügt über 16 Vibrationsmodule, die um die Taille angeordnet sind, um eine 360°-Abdeckung zu gewährleisten. Der Vibrationsgürtel wird hierbei zum einen für die sensorische Augmentation von Hindernissen über die vorderen 5 Vibrationsmotoren, sowie zur Anweisung der Navigation über die 10 hinteren Vibrationsmotoren verwendet. Die Kommunikation mit dem naviGürtel erfolgt via Bluetooth Low Energy.

Die sensorische Augmentation von Hindernissen zielt darauf ab, blinden und sehbeeinträchtigten Personen Informationen über ihre Umgebung zu vermitteln, die ihnen normalerweise nicht zur Verfügung stehen. Diese Augmentation übermittelt dem Nutzer mittels Vibrationssignalen die Position von Hindernissen, wobei fünf Vibrationsmotoren im Bauchbereich verwendet werden, um einen Detektionsbereich von 112,5° abzudecken (siehe Abbildung 5). Durch die Aktivierung dieser Motoren kann der Nutzer die räumliche Einordnung der Objekte wahrnehmen, zum Beispiel signalisiert eine Vibration am Bauchnabel ein direkt vorausliegendes Hindernis. Die Intensität der Vibration dient dabei als Kodierung der Entfernung zum Objekt: Bei nahen Hindernissen (≤ 0,5 m) wird die maximale Intensität ausgelöst, während bei weiter entfernten Objekten die Intensität linear abnimmt oder bei einer Entfernung über einem Meter ganz ausbleibt. Diese taktile Rückmeldung ermöglicht es dem Nutzer oder

der Nutzerin, mehrere Hindernisse gleichzeitig zu erfühlen, sie räumlich einzuordnen und ihnen gezielt auszuweichen, was das Risiko von Kollisionen erheblich reduziert.

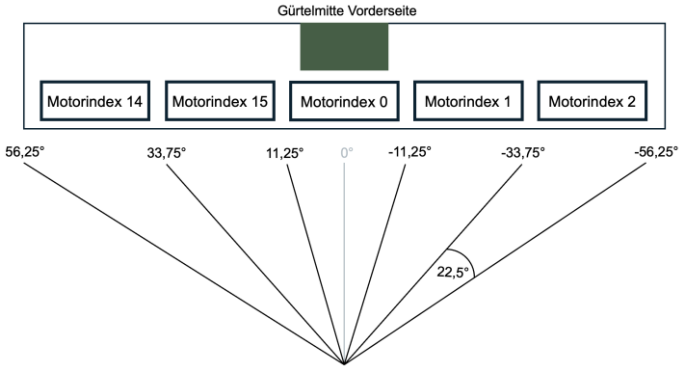


Abbildung 5: Die Hindernisaugmentation erfolgt über eine Abdeckung von 112,5° in fünf Segmenten mit jeweils 22,5° Abdeckung. Im Falle eines Hindernisses im jeweiligen Segment vibriert der zugehörige Vibrationsmotor in der Richtung des Segments mit einer Intensität in Abhängigkeit des Abstands zum nächsten Hindernis.

Die vibrotaktile Übermittlung der Navigationsanweisungen des Assistenzsystems erfolgt über die auf der Rückseite des Vibrationsgürtels angebrachten Motoren. Abhängig von der Richtung, in die sich die Nutzerin oder der Nutzer bewegen soll, vibriert der entsprechende Motor auf der Rückseite des Gürtels und vermittelt so den Eindruck eines sanften Schiebens in die gewünschte Richtung. Diese taktile Rückmeldung soll intuitiv die Bewegungsrichtung signalisieren und ist so gestaltet, dass der Hörsinn vollständig frei bleibt.

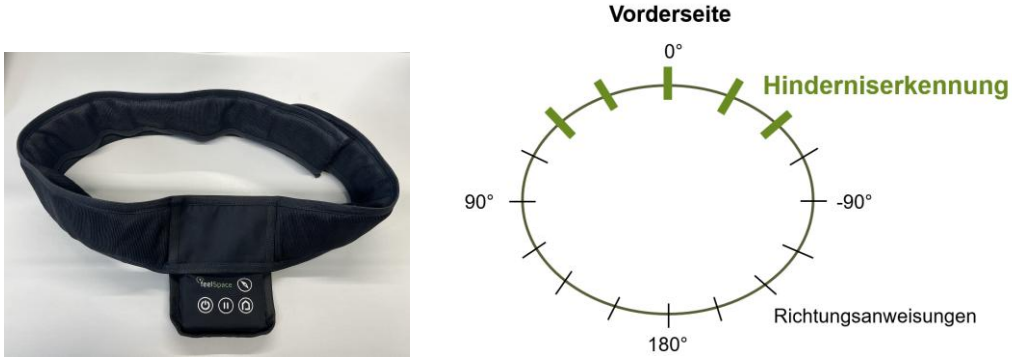


Abbildung 6: Links der NaviGürtel von feelSpace, rechts die Zuordnung der jeweiligen Funktionen zu den 16 Vibrationsmotoren. Fünf vorne positionierte Motoren (grün) sind der Hindernisaugmentation zugeordnet, während die übrigen Motoren der Richtungsanweisung dienen.

8.4.3 3D Audio

Als weitere Schnittstelle des multimodalen Hinkanal nutzt das LOMOBI Assistenzsystem zur Nutzerinteraktion die räumliche Audiowiedergabe (3D-Audio), um sehbehinderten Personen Navigationsanweisungen und Hindernisinformationen intuitiv zu übermitteln. Dieses akustische Feedback wird über Knochenschallkopfhörer an die Nutzenden übertragen. Der Einsatz von Knochenschallkopfhörern ist dabei ein wesentlicher Vorteil, da der Gehörgang der Nutzenden frei bleibt, was die gleichzeitige Wahrnehmung wichtiger Umgebungsgeräusche (wie Verkehrslärm oder Sprache) ermöglicht. Dies ist entscheidend, da der Hörsinn für sehbeeinträchtigte Personen eine der wichtigsten Sinneswahrnehmungen zur Orientierung darstellt. Die 3D-Audio-Technologie ermöglicht es, Klänge

unter Nutzung des psychoakustischen Modells, der Head-Related Transfer Functions (HRTF), räumlich präzise zu positionieren. Dadurch entsteht der Eindruck, dass Navigations- oder Warnsignale aus einer spezifischen Richtung stammen und somit intuitiv wahrgenommen werden können. Das System kann damit Informationen kontinuierlich, präzise und schneller vermitteln, als es mit diskreten Sprachbefehlen möglich wäre.

Die Funktionalität des Systems umfasst die gleichzeitige Vermittlung von Routenanweisungen und der Position von Hindernissen. Die akustischen Hinweise werden dabei dynamisch an die Bewegung und den Abstand zu Objekten angepasst; nähert sich ein Hindernis, erhöht sich beispielsweise die Frequenz des Warnsignals. Die gesamte Kommunikation und Steuerung des akustischen Feedbacks ist modular im Rahmen des Robot Operating System 2 (ROS2) implementiert.

8.5 Automatische Segmentierung und Trainingsdatengenerierung

Die Entwicklung KI-gestützter Umgebungswahrnehmungssysteme erfordert große Mengen qualitativ hochwertiger, annotierter Trainingsdaten. Deren manuelle Erstellung stellt jedoch einen erheblichen zeitlichen und finanziellen Aufwand dar und ist zugleich fehleranfällig. Um diese Hürde zu überwinden, wurden im Projekt verschiedene Strategien zur automatisierten und teilautomatisierten Generierung sowie Verbesserung von Trainingsdaten entwickelt und erprobt. Diese Ansätze reichen von der Nutzung synthetischer Daten aus der CARLA-Simulation über innovative Auto-Labeling-Verfahren auf Basis des tatsächlich gegangenen Weges bis hin zu aktiven Lernstrategien mit Zero- und Single-Shot-Segmentierung. Ziel war es, den manuellen Annotierungsaufwand drastisch zu reduzieren, die Datenqualität zu erhöhen und die Generalisierungsfähigkeit der Modelle in realen Einsatzszenarien nachhaltig zu verbessern.

8.5.1 Carla Simulation für Trainingsdatengenerierung

Für die Generierung von Trainingsdaten im Rahmen des Projekts wurde der Open-Source-Simulator CARLA (Car Learning to Act) eingesetzt. Mithilfe von CARLA lassen sich verschiedene städtische Umgebungen simulieren und entsprechende Trainingsdaten erzeugen. Die Erstellung großer Mengen realer annotierter Daten ist zeitaufwendig, kostspielig und arbeitsintensiv, weshalb synthetische Daten eine notwendige Ergänzung darstellen. CARLA bietet hierfür eine skalierbare und kontrollierte Simulationsumgebung, um die Generalisierungsfähigkeit des Deep-Learning-Modells zu verbessern und diverse Trainingsszenarien abzubilden, die in der realen Welt nur schwer zu erfassen wären.

Der Simulator ermöglicht die Nachbildung einer fußgängerzentrierten Perspektive, was ein wesentlicher Vorteil gegenüber anderen Tools ist, die sich oft auf fahrzeugzentrierte Ansichten konzentrieren. Zu diesem Zweck wurden virtuelle Kameras in CARLA in etwa Brusthöhe (1,6 Meter) platziert und mit einem Neigungswinkel von etwa 30 Grad nach unten ausgerichtet, um die reale Kamerakonfiguration des LOMOBI-Systems nachzuahmen. Ein Vorteil von CARLA ist die automatische Erzeugung von pixelgenauen Segmentierungsmasken mithilfe des semantischen Segmentierungsmodus, wie in Abbildung 7 dargestellt.

Abbildung 7: CARLA-Simulator-Ausgaben: (Links) RGB-Bild, (Mitte) semantische Segmentierung und (Rechts) binäre Maske, die die Segmentierung des Gehwegs zeigt.

Für die Steigerung der Szenenvielfalt wurden die synthetischen Daten aus vier unterschiedlichen virtuellen Städten (Town02, Town03, Town05 und Town10) gesammelt, die verschiedene architektonische Stile und städtische Layouts bieten. Es wurden 10.200 synthetische Bilder generiert, die zusammen mit realen Aufnahmen in gemischten Datensätzen verwendet wurden. Der reale Datensatz wurde dabei zum Vergleich händisch mit Unterstützung von dem Segment Anything Model gelabelt. Obwohl eine visuelle Diskrepanz (Domain Gap) zwischen den synthetischen und realen Bildern besteht, führt die Kombination beider Datenquellen zu einer verbesserten Generalisierung und Robustheit der Modelle (Tabelle 2).

Tabelle 2: Validierungsergebnisse für verschiedene Trainings- und Validierungsdatenkonfigurationen (Reale-Bilder oder Carla-Bilder) unter Verwendung des U-Net-Modells [21].

Trainingsdaten / Validierungdaten	*Pixel Acc. (%)*	*Dice Score*	*IoU*
Reale (7.000 Bilder) / Reale (3.200 Bilder)	97,86	0,9787	95,85
Carla (7.000 Bilder) / Carla (3.200 Bilder)	98,21	0,9824	96,57
Carla (7.000 Bilder) / Reale (3.200 Bilder)	88,42	0,9818	80,55
Gemischt (3.500 Carla + 3500 Reale) / Reale (3.200 Bilder)	**98,34**	**0,9842**	**96,89**

8.5.2 Automatische Segmentierung anhand des gegangen Weg

Die Automatische Segmentierung anhand des gegangenen Weges stellt einen neuartigen Auto-Labeling-Ansatz dar, der zur Generierung von Trainingsdaten für die Navigation sehbeeinträchtigter Personen entwickelt wurde [22]. Die präzise Identifizierung begehbarer Flächen ist essenziell für die sichere Navigation. Herkömmliche Datensätze sind jedoch oft aus der Perspektive autonomer Fahrzeuge aufgenommen und daher für die Fußgängernavigation unzureichend. Der Ansatz zielt darauf ab, den zeit- und kostenintensiven Prozess der manuellen Annotation zu umgehen, indem es den tatsächlich vom Kameraoperateur zurückgelegten Pfad zur automatischen Segmentierung der durchquerten Bereiche nutzt. Als Datenbasis dienen dabei sequenzielle monokulare Kamerabilder, die von kollaborativen Plattformen wie Mapillary stammen.

Der Ansatz beruht auf einem mehrstufigen Prozess. Zuerst wird die Trajektorie des Kameraoperators aus den geschätzten Kameraposen extrahiert. Anschließend werden Teile des zukünftigen Pfades, die

durch Hindernisse verdeckt sind, herausgefiltert. Dies erfolgt mithilfe des Modells Depth Anything V2 zur metrischen Tiefenschätzung, wodurch die Notwendigkeit eines zusätzlichen Tiefensensors entfällt. Die so gewonnenen, gefilterten Punkte dienen als Prompts für die eigentliche Segmentierung der begehbaren Fläche, welche mit dem Modell Segment Anything Model 2 (SAM 2) [23] durchgeführt wird. Die Segmentierungsgenauigkeit konnte durch die Verwendung einer inkrementellen Auswahlstrategie für die Prompts, bei der die Segmentierung für Punkte außerhalb der zuvor erstellten Maske wiederholt wird, optimiert werden.

Die Validierung des Ansatzes erfolgte anhand eines manuell gelabelten Datensatzes von Mapillary, der Gehwege, gepflasterte Fußwege und gering strukturierte Umgebungen umfasste. Im direkten Vergleich mit dem hochmodernen Zero-Shot-Segmentierungsmodell Grounded SAM 2 [24] zeigte der Ansatz eine überlegene Leistung. Insgesamt wurde eine mittlere Intersection over Union (mIoU) von 93,9% und einen mittleren F1 Score (mF1) von 96,6% erreicht. Dies stellt eine Leistungssteigerung von +1,6 Prozentpunkten beim mIoU und +1,0 Prozentpunkten beim mF1 im Vergleich zu Grounded SAM 2 dar. Die Methode ist besonders effektiv in strukturierten Umgebungen und demonstriert Robustheit beim automatischen Labeln begehbarer Flächen ohne menschliche Aufsicht.

8.5.3 Zero- und Single Shot Segmentierungs Pipeline

Die Iterative Verbesserung der Umgebungssegmentierung stellt einen fortschrittlichen Ansatz dar, um die Genauigkeit und Skalierbarkeit von Navigationsassistenzsystemen für sehbeeinträchtigte Personen zu gewährleisten. Angesichts der hohen Sicherheitsanforderungen im Anwendungsbereich (Ziel-IoU > 0,8, angestrebt > 0,9) war ein effizienteres Verfahren erforderlich, um potenzielle Fehlerquellen im Datensatz schnell zu identifizieren und zu korrigieren, sowie potenziell neue hilfreiche Trainingsbilder während der Anwendung zu identifizieren.

Das zugrunde liegende Konzept baut auf Techniken des Active Learning auf, fokussiert jedoch auf die automatisierte Annotation von Bildern mit hoher Unsicherheit. Zunächst wird die Unsicherheit der Vorhersagen auf Pixelebene mithilfe der Monte-Carlo-Dropout-Methode berechnet (basierend auf der Varianz). Bilder mit der höchsten Varianz (den sogenannten unsichersten Bildern) werden gezielt identifiziert, da sie die größten potenziellen Schwachstellen in der Lerngrundlage des Modells repräsentieren. Diese kritischen Stichproben werden anschließend nicht manuell, sondern automatisiert durch den Einsatz moderner Zero- und Single-Shot-Segmentierungsmodelle gelabelt.

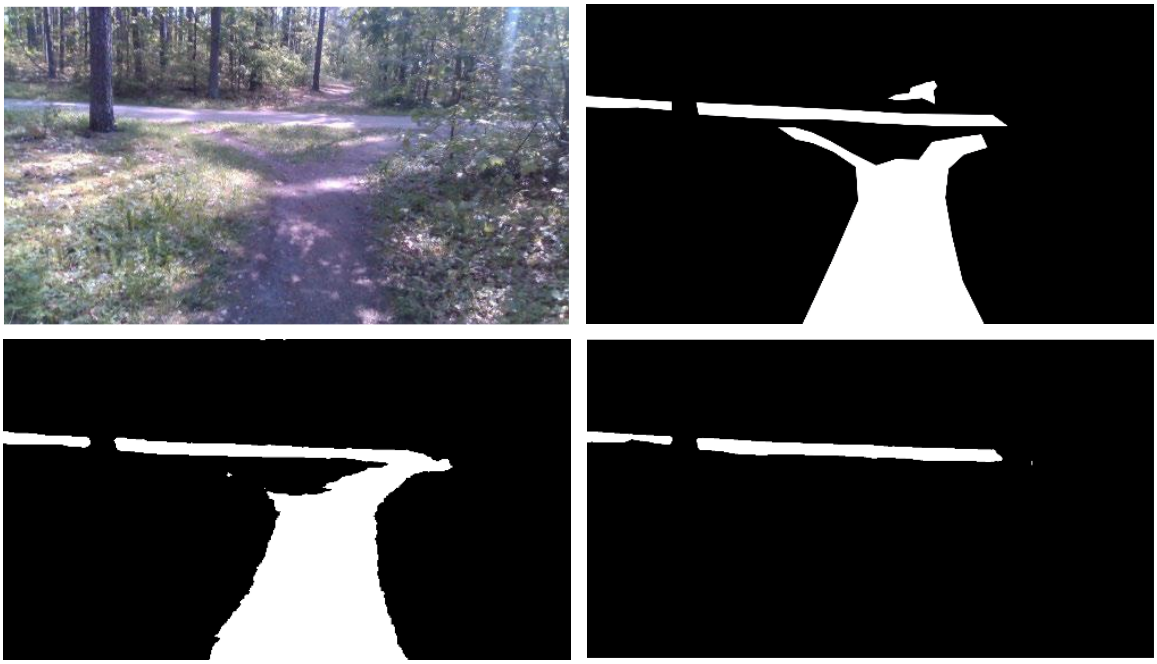


Abbildung 8: Original (oben links), Ground-Truth-Maske (oben rechts), SegGPT-Maske (unten links), Grounded SAM2-Maske (unten rechts).

Bei der Evaluierung dieser automatisierten Label-Pipeline zeigte das Single-Shot-Modell SegGPT [25] (basierend auf In-Context-Learning) eine leicht überlegene Leistung in der Straßensegmentierung gegenüber dem Zero-Shot-Modell Grounded SAM2 [24]. SegGPT erzielte einen durchschnittlichen Intersection over Union (IoU) von 0,9039 (Median 0,9578), während Grounded SAM2 einen durchschnittlichen IoU von 0,8982 erreichte (siehe Tabelle 3). Die Anwendung dieser Methodik auf die als am schwierigsten identifizierten Bilder (mit hoher Unsicherheit) konnte die Qualität des Trainingsdatensatzes gezielt stabilisieren und die Vielzahl an nötigen Trainingsdaten reduzieren.

Tabelle 3: Vergleich der Segmentierungsleistung von SegGPT und Grounded SAM 2 anhand ausgewählter Metriken

Metrik / Detail	***SegGPT***	***Grounded SAM 2***
Gesamtanzahl der Bilder im Datensatz	375	375
Anzahl erfolgreich erzeugter Masken	371	**375**
Bilder ohne Maskenerzeugung	4	**0**
Durchschnittlicher IoU-Wert	**0,9039**	0,8982
Median des IoU-Werts	**0,9578**	0,9547
Maximaler IoU-Wert	**0,9922**	0,9884
Minimaler IoU-Wert	0,0000	**0,0152**

Dieser innovative Ansatz ermöglicht langfristig ein kontinuierliches Lernverfahren für das Assistenzsystem, wodurch die Anpassungsfähigkeit und Verlässlichkeit der Mobilitätsunterstützung sehbeeinträchtigter Personen in wechselnden Umgebungen signifikant erhöht werden kann, ohne dass umfangreiche manuelle Nachbearbeitung erforderlich ist.

8.6 Indoor- und Outdoor-Lokalisierung für ETAs

Das Assistenzsystem LOMOBI wird entwickelt, um Nutzenden eine eigenständige Bewegung sowohl im Innen- als auch im Außenbereich zu ermöglichen. Umfassende Unterstützung bei Orientierung und Mobilität (O&M) im Alltag setzt voraus, dass das System die jeweilige Umgebung (urbane Umgebungen,

öffentliche Gebäude etc.) interpretieren und die eigene Position präzise bestimmen kann. Die Lokalisierung und die darauf aufbauende Pfadplanung unterscheiden sich dabei grundlegend, je nachdem, ob sich der Nutzer in geschlossenen Räumen oder im Freien befindet. Im Folgenden werden die unterschiedlichen untersuchten Lokalisierungsmöglichkeiten detaillierter beschrieben.

8.6.1 Indoor-Lokalisierung

Für die Navigation in Innenräumen kann das Global Positioning System (GPS/GNSS) aufgrund der starken Einschränkungen des Signals nicht eingesetzt werden. Daher stützt sich die Indoor-Lokalisierung und -Navigation des LOMOBI-Systems hauptsächlich auf Sensorik wie Stereokamera und IMU in Kombination mit Kartenmaterial.

Die relative Bewegungsänderung, auch genannt Odometrie, basiert auf einer Stereokamera und einer IMU. Im Rahmen von LOMOBI sind verschiedene Stereokameras untersucht worden. Eine Stereokamera liefert zum einen ein Bild als auch Tiefendaten über Triangulation. Anhand der Bilddaten kann eine visuelle Odometrie, die die relative Bewegung der Kamera beschreibt, geschätzt werden. Bewegt sich eine Person mit der Kamera einen Meter vorwärts, so sollte eine perfekte Odometrie eben diesen Meter ausgeben. Nachteile dieses Verfahrens sind, dass es abhängig von den visuellen Merkmalen der Umgebung ist. Das bedeutet, dass Bereiche mit wenigen Merkmalen sowie mit vielen dynamischen Objekten problematisch sind. Deshalb wird bei der Berechnung der Odometrie in der Regel eine IMU, die die Beschleunigungen misst, mit fusioniert, um eine stabilere Messung zu erhalten. In diesem Projekt kommt final die interne IMU der D435 zum Einsatz, aber auch eine externe IMU der Firma Olive Robots ist untersucht worden.

Zur Kartierung und Lokalisierung in unbekannten Innenräumen wird das Verfahren Simultaneous Localization and Mapping (SLAM) angewandt. SLAM erstellt gleichzeitig eine Umgebungskarte und bestimmt die eigene Position darin. Innerhalb des LOMOBI-Projekts wurde RTAB-Map (Real-Time Appearance-Based Mapping), ein quelloffenes Framework für Graphen-basiertes SLAM, verwendet und als geeignet für die Fortführung des Projekts erachtet. RTAB-Map kombiniert visuelle Odometrie mit Loop-Closure-Erkennung (Schleifenschließung), um Driftfehler zu korrigieren und die Präzision der aufgezeichneten Karte zu verbessern. Die Lokalisierung und Kartierung basieren auf den Kameradaten und ermöglichen die Erstellung einer Belegungskarte (Costmap), die für die Pfadplanung essenziell ist. Die Kartierung mittels SLAM auf Basis von Kameradaten wurde technisch als machbar befunden und deckt einen breiteren Anwendungsbereich ab als andere Technologien. Für die Lokalisierung in bekannten Innenräumen, wie dem FAPS-Lehrstuhlgebäude, wird eine statische Karte der Umgebung genutzt, die zuvor mittels LOMOBI aufgezeichnet wurde. Die Lokalisierung erfolgt dann durch den

Abgleich der aktuellen Sensordaten mit dieser statischen Karte. Auf diese Weise können Gebäude kartiert und auch anderen Nutzern des Systems über eine Austauschplattform bereitgestellt werden.

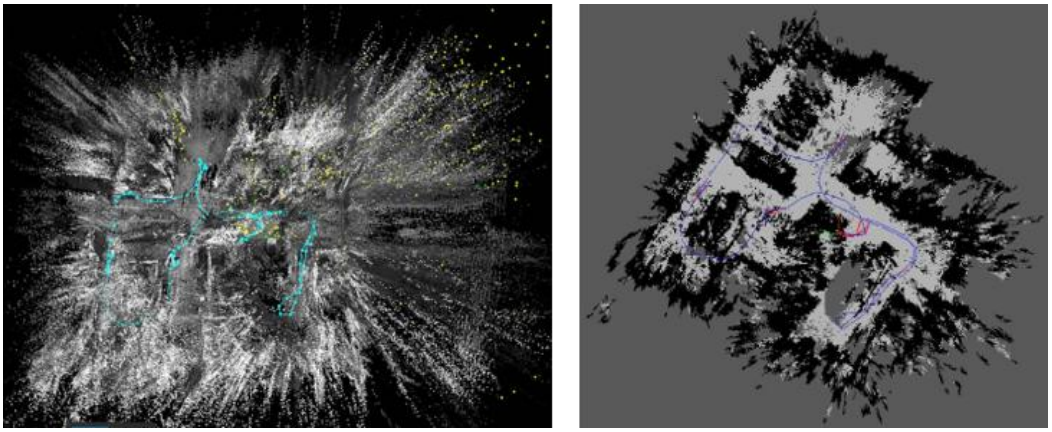


Abbildung 9: RTAB-Map Merkmalskarte der Indoor-Umgebung zur Lokalisierung (links) und die draus abgeleitete Rasterkarte zur Pfadplanung (rechts).

8.6.2 Outdoor-Lokalisierung

Die Outdoor-Navigation adressiert die Herausforderungen der Fortbewegung im Freien, insbesondere in urbanen, dynamischen Umgebungen. Im Gegensatz zum Innenbereich basiert die grobe Lokalisierung hier primär auf Satellitennavigation.

Für die Gewährleistung einer absoluten Positionsbestimmung im Außenbereich nutzt das LOMOBI-Assistenzsystem das GPS beziehungsweise das GNSS, wobei das System hierfür mit einem spezialisierten GNSS-Modul und einer externen Antenne erweitert wurde, um die Positionsbestimmung zu ermöglichen. Um die erforderliche Präzision der Lokalisierung im Freien zu erreichen, ist die Fusion verschiedener Sensordaten erforderlich, welche im LOMOBI-Projekt mittels eines Erweiterten Kalman Filters (EKF) bzw. eines Dual-EKF-Ansatzes umgesetzt wird. Dieser Ansatz kombiniert GPS-Daten mit Informationen der IMU und der VO. Die Struktur umfasst einen lokalen EKF, der IMU- und Odometrie-Daten für kurzfristige Positionsinformationen fusioniert. Ergänzend dazu integriert ein globaler EKF zusätzlich GPS-Daten sowie SLAM-basierte Lokalisierungsinformationen, um eine globale, langfristig stabile Odometrie im map-Frame zu generieren. Des Weiteren können SLAM-Verfahren (zum Beispiel RTAB-Map) auch im Freien zur Kartierung genutzt werden, obwohl die Anzahl der Merkmale geringer ausfallen kann, hat die Integration von GPS/GNSS in RTAB-Map zur Verbesserung der Lokalisierung und Orientierung beitragen.

Für Gebiete mit unzureichender GNSS Signal Verfügbarkeit, wie diese in komplexen städtischen Gebieten vorkommen können, wurde eine alternative Outdoor-Lokalisierung untersucht. Diese basiert auf OrienterNet [26], zur KI-gestützten visuellen Feinlokalisierung in OSM integriert und evaluiert. Ziel war es, die Lokalisierungsungenauigkeit, die durch ungenaue GNSS-Signale verursacht wird, in urbanen Umgebungen zu kompensieren und damit die Navigationssicherheit für sehbeeinträchtigte Personen zu erhöhen. Das Modell schätzt hierbei eine 3-DoF-Pose (Position x, y und Orientierung θ) aus einem einzigen Bild oder Bildsequenz. OrienterNet ahmt dabei die menschliche

Orientierungsfähigkeit nach, indem es ein neuronales Matching zwischen der Sicht der Umliegenden Struktur und der Draufsichtkarte von OSM durchführt.

Trotz der deutlichen Verbesserung des Fehlers durch den sequenziellen Ansatz wurde das OrienterNet-Modell in seiner aktuellen Form als nicht geeignet für den Einsatz in sicherheitskritischen Anwendungen im Außenbereich befunden. Abbildung 10 zeigt die Teststrecke und Fehlerverlauf des Versuchs. Eine verbesserte Lokalisierung gegenüber des GPS Signals konnte hierbei nicht realisiert werden. Der verbleibende Durchschnittsfehler von 4,69 m und der maximale Fehler von 13,19 m stellen im öffentlichen Raum, insbesondere im Straßenverkehr, ein inakzeptables Sicherheitsrisiko dar. Ein Lokalisierungsfehler von nur 5 m kann die Unterscheidung zwischen einem sicheren Gehweg und einer Fahrbahn bedeuten. Die Fehler traten konsistent in kritischen Szenarien auf, die durch visuelle Ambiguität gekennzeichnet waren, wie beispielsweise Bereiche mit dichter Vegetation, monotoner Architektur oder sich wiederholenden Fassaden. Diese Umgebungen führen zu einem Problem des perzeptuellen Aliasings (Verwechslungsgefahr), bei dem das System visuell ähnliche Orte auf der OSM-Karte nicht eindeutig unterscheiden kann.

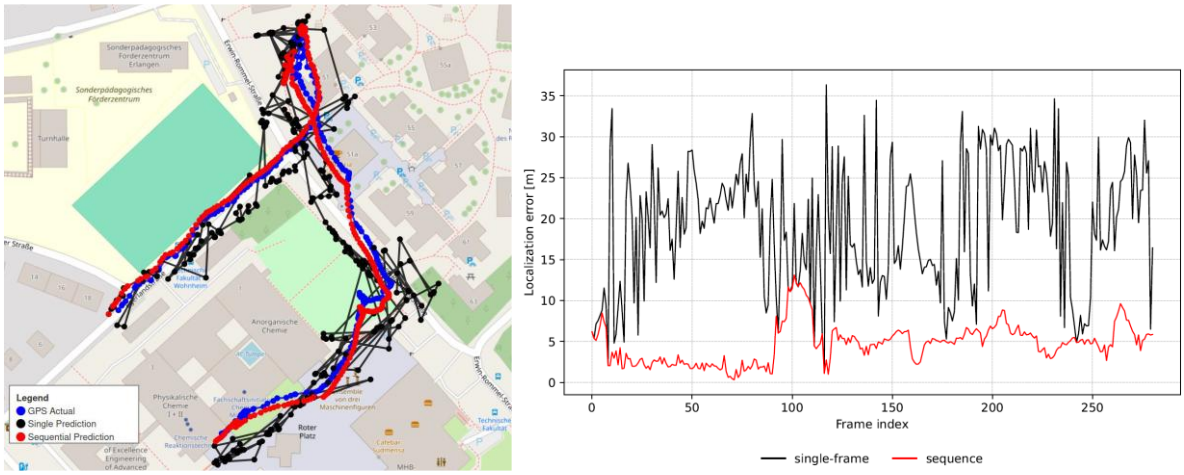


Abbildung 10: Vergleich von Einzelbild- und sequenzieller Prädiktion. Die Grafik links zeigt die GPS-Grundwahrheit (blau) im Vergleich zu den Schätzungen der Einzelbild-Lokalisierung (schwarz) und der sequenziellen Lokalisierung (rot). Das rechte Diagramm zeigt den Fehlerverlauf der Einzelbild-Lokalisierung (schwarz) und der sequenziellen Lokalisierung (rot). Die x-Achse gibt den Frame-Index an, während die y-Achse den Lokalisierungsfehler in Metern darstellt.

8.7 Pfad- und Ablaufplanung

Die Pfadplanung im LOMOBI-Assistenzsystem ist eine zentrale Komponente, um sehbeeinträchtigten Nutzenden eine optimale und kollisionsfreie Trajektorie von einer Start- zu einer Zielposition zu berechnen. Der gesamte Navigationsablauf erfordert das präzise Zusammenspiel von Lokalisierung, Umgebungsinterpretation, Pfadplanung und der Ausführung geplanter Aktionen an bestimmten Positionen im Raum. Um diese komplexen Anforderungen sowohl in Innenräumen als auch im Freien zu erfüllen, setzt LOMOBI auf einen hierarchischen Planungsansatz, der globale Routen mit lokaler Echtzeitanpassung und sequenzieller Verhaltenssteuerung kombiniert.

8.7.1 Hierarchisches Pfadplanungskonzept

Das hierarchische Pfadplanungskonzept des LOMOBI-Assistenzsystems ist darauf ausgelegt, eine nahtlose Navigation in komplexen Umgebungen zu gewährleisten, indem es die Steuerung und

Routenfindung in drei logische Ebenen unterteilt, die von globalen semantischen Informationen bis zu lokalen Bewegungsbefehlen reichen.

Die oberste Ebene basiert auf einem semantischen Graphen, der die Umgebungsinformationen abstrahiert. Für die globale Routenführung greift das System primär auf frei verfügbare geografische Daten wie OSM zurück. Dieses OSM-basierte Straßennetzwerk wird für den Innenbereich gezielt erweitert, um interne Straßennetzwerke und funktionale Landmarken abzubilden. Funktionale Landmarken sind Orte, an denen spezifische Verhaltensweisen des Nutzers erforderlich sind, wie zum Beispiel Türen, Treppen und Aufzüge in Gebäuden. Die Koordinaten dieser Landmarken werden in der Karte gespeichert (z. B. in einem YAML-Dokument). Der gesamte Pfad von Start zum Endziel wird in Teilpfade zerlegt, wobei jeder Teilpfad bis zur nächsten funktionalen Landmarke führt

Der globale Pfadplaner, der im Rahmen des Nav2-Frameworks implementiert ist, bildet die mittlere Ebene. Er empfängt den nächsten semantischen Wegpunkt (das Teilziel), beispielsweise die Position einer erkannten Tür, von der obersten Ebene. Der globale Pfadplaner berechnet dann einen niedrig auflösenden High-Level-Pfad vom aktuellen Standort zu diesem Teilziel, wobei der A*-Suchalgorithmus typischerweise für diese Berechnung des kürzesten Weges zum Einsatz kommt. Die Koordination dieser sequenziellen Teilaufgaben wird durch Behavior Trees (Verhaltensbäume) gesteuert, welche die Logik zur Abfrage der nächsten Landmarke und zur Initiierung der Navigation (mittels der Nav2-Aktion NavigateToPose) bereitstellen.

Der resultierende globale Pfad wird an den lokalen Pfadplaner auf der untersten Ebene übergeben. Der lokale Pfadplaner, der Teil des Navigations-Stacks, arbeitet in Echtzeit („online“). Er nutzt aktuelle Sensordaten, um den globalen Pfad exakt zu verfolgen und dabei dynamischen Hindernissen (wie entgegenkommenden Personen) auszuweichen. Schließlich berechnet der lokale Planer Geschwindigkeitsbefehle. Diese Befehle werden an die Interaktionsschnittstelle (das haptische Feedback-System, z. B. über Vibrationsmodule in den Schultergurten) übermittelt, um dem Nutzer Richtungsanweisungen oder, beim Erreichen einer Landmarke, ein Signal zur Ausführung einer notwendigen Aktion (wie das Öffnen einer Tür) zu geben.

Zusätzlich zu den Sensordaten wird in der Karte des lokalen Pfadplaners die Wegsegmentierung als Kosten integriert. Auf diese Weise werden die unbegehbaren Flächen als Hindernisse dargestellt und

der Pfadplaner meidet diese Regionen, wodurch sehbeeinträchtigte Personen auch bei ungenauer GPS-Lokalisierung lokal beispielsweise auf einem Bürgersteig navigieren können.

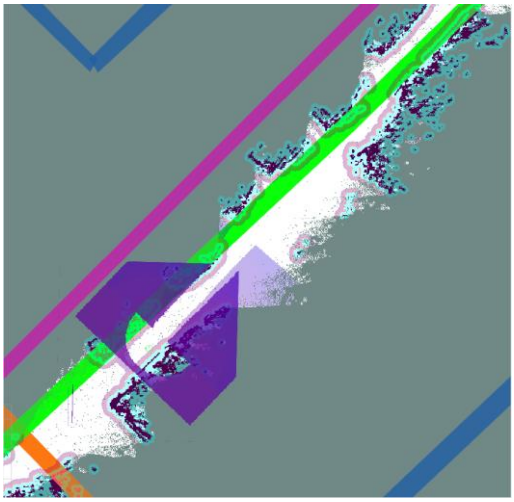


Abbildung 11 Fusionierte Pfadplanung aus semantischer Karte von OSM, globaler Karte von RTAB-Map und lokaler Karte mit integrierter Wegsegmentierung.

8.7.2 Ablaufplanung mittels Behavior Trees

In der herkömmlichen Pfadplanung robotischer Assistenzsysteme wird primär die kontinuierliche Fortbewegung des Roboters oder Nutzers entlang einer Trajektorie optimiert, wobei davon ausgegangen wird, dass Geschwindigkeitsbefehle kontinuierlich ausgeführt werden können. Diese Annahme ist jedoch in realen Alltags- und Navigationsszenarien für sehbeeinträchtigte Personen nicht haltbar, insbesondere an kritischen Übergängen und Orten im Raum. Das LOMOBI-Assistenzsystem muss die Fähigkeit zur Interaktion mit Hindernissen auf dem geplanten Weg besitzen. Diese Objekte, die als funktionale Landmarken bezeichnet werden, sind Orte, an denen eine Person eine Aktion ausführen muss oder warten muss, was die unterbrechungsfreie Ausführung von Bewegungsbefehlen blockiert. Funktionale Landmarken umfassen dabei im Indoor-Bereich Türen, Treppen und Aufzüge in Gebäuden, während im Outdoor-Bereich Ampeln und Zebrastreifen in urbanen Umgebungen betrachtet werden können. Die Planung des Verhaltens zur Erkennung und Bewältigung dieser Hindernisse ist daher von zentraler Bedeutung und ein essenzieller Bestandteil im Navigationsprozess. Das Ziel ist es, die Orte im Raum zu erkennen, an denen bestimmte Verhaltensweisen erforderlich sind, und die entsprechenden Aktionen auszuführen.

Zur Steuerung der Handlungsabfolge und zur Organisation komplexer Verhaltensmuster wurde der Ansatz der Behavior Trees (Verhaltensbäume) gewählt. Behavior Trees (BT) stellen eine Weiterentwicklung der traditionellen Zustandsautomaten (State Machines) dar und bieten signifikante Vorteile für die Ablaufsteuerung in der Robotik. Die Entscheidung für BTs begründet sich durch ihre Modularität, Reaktivität und Ausdruckskraft. Im Gegensatz zu sequenziellen Zustandsautomaten erlauben BTs die parallele Ausführung von Knoten und die Aufteilung komplexer Verhaltensmuster in einfache, wiederverwendbare Teilaufgaben. Diese Eigenschaften sind besonders vorteilhaft für die

Modellierung reaktiven Verhaltens in dynamischen Umgebungen und bei der Koordination komplexer Abläufe.

Abbildung 12: Aufforderung zum öffnen der Tür durch die Ablaufplanung mittels Behavior Tree.

Dem Nutzer wird während des Ablaufs mitgeteilt, wenn dieser eine Aktion wie zum Beispiel eine Tür zu öffnen oder warten muss, über eine Ausgabe mitgeteilt. Die Notwendigkeit zur Ausführung einer spezifischen Aktion an einem bestimmten Ort ist ein essenzieller Bestandteil des Navigationsprozesses. Die Entfernung zur Landmarke wird dabei über definierte Radien bestimmt, wobei im realen Assistenzsystem für Türen ein Radius von 1,5 Metern verwendet wird.

Um die Ablaufplanung präzise an die Umgebung anpassen und die Notwendigkeit zur Türöffnung prüfen zu können, ist die Erkennung des Zustands der Landmarke von zentraler Bedeutung. Als Erkennungsmethode für funktionale Landmarken wird das neuronale Netz YOLO (You Only Look Once) verwendet. Dieses Netz wurde spezifisch auf der Basis von Kameradaten trainiert. Im Rahmen der Türaufgabe ist das Netz auf Türen und deren Zustand geschlossen, geöffnet oder halbgeöffnet trainiert worden. Diese Echtzeit-Erfassung des Türzustands ist essenziell, da der Nutzer bei bereits geöffneter Tür nur informiert werden muss und keine zusätzliche Aktion ausführen muss. Zudem wurde das neuronale Netz für die Ampelaufgabe trainiert, um den Zustand (grün, gelb oder rot) zu erfassen. Diese Erkennung ermöglicht es, an Ampeln das erforderliche Verhalten (Warten oder Überqueren) präzise zu planen.

Abbildung 13 Die Erkennung verschiedener visueller Landmarken wurde mittels eines neuronalen Netztes erreicht. Links am Beispiel von Türen und rechts am Beispiel von Ampel.

8.8 Erkenntnisse aus den Probandenstudien zur multimodalen Nutzerinteraktion

Zur Bewertung der im Projekt entwickelten multimodalen Nutzerinteraktion wurden mehrere Probandenstudien durchgeführt, die die Praxistauglichkeit, Benutzerfreundlichkeit und Akzeptanz der unterschiedlichen Interaktionsmodalitäten untersuchten. Ziel war es, die Eignung der Systeme für sehbeeinträchtigte Nutzerinnen und Nutzer zu evaluieren und deren Beitrag zur intuitiven, sicheren und effizienten Navigation zu bewerten. Dabei wurden die Smartphone-App, der vibrotaktile Gürtel sowie das 3D-Audio-System in getrennten Untersuchungen hinsichtlich ihrer Effektivität, Verständlichkeit, Reaktionszeit und Nutzerzufriedenheit analysiert. Die Ergebnisse liefern wertvolle Erkenntnisse zur Gestaltung und Kombination multimodaler Schnittstellen, um eine möglichst natürliche und redundante Interaktion zwischen Mensch und Assistenzsystem zu ermöglichen.

8.8.1 Smartphone App

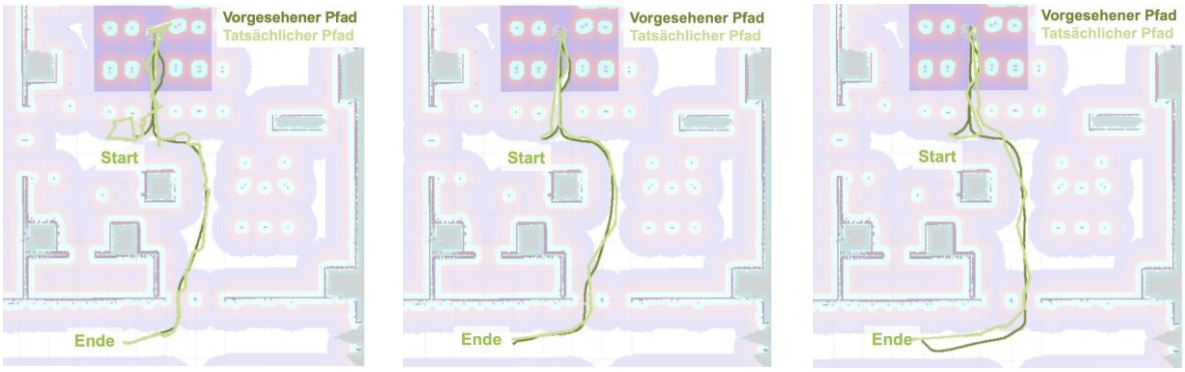
Eine Nutzerstudie bestätigte die Eignung der App als Schnittstelle: Sehbehinderte Teilnehmer bewerteten die Effizienz mit durchschnittlich 4,0 und die Effektivität mit 4,5 von 5 möglichen Punkten (5 = stimme voll zu); die Barrierefreiheit wurde mit 4,6 sehr positiv bewertet. Kritisch wurde lediglich die Überlagerung der Sprachausgaben bei gleichzeitiger Nutzung von TalkBack und dem App-eigenen TTS-Dienst angemerkt.

Die Nutzerinteraktion mit der Lomobi Smartphone-App wurde im Rahmen einer Nutzerstudie evaluiert, die darauf abzielte, die Eignung der in Flutter entwickelten und auf Barrierefreiheit ausgerichteten Anwendung als Schnittstelle für das Assistenzsystem zu bewerten. Die Evaluation der Benutzerfreundlichkeit basierte auf Metriken wie Effizienz, Effektivität und Verständlichkeit und umfasste zehn Teilnehmer, darunter drei Personen mit Sehbeeinträchtigung. Die Ergebnisse zeigten eine insgesamt gute bis sehr gute Benutzerfreundlichkeit (Usability), wobei die Gruppe der sehbehinderten Teilnehmer die Verständlichkeit, Robustheit und Bedienbarkeit mit der Note 5 bewertete. Im Vergleich zur Benchmark-App FeelSpace schnitt die LOMOBI-App in puncto Effizienz (MW 3,85 vs. 3,14) und Effektivität (MW 4,14 vs. 3,14) besser ab und wurde oft als einfacher empfunden, was auf den Fokus der App auf die Kernfunktionen zurückzuführen ist. Herausforderungen bestehen weiterhin in der Optimierung der nahtlosen Interoperabilität zwischen der App-eigenen Sprachausgabe (TTS-Dienst) und Screenreadern wie TalkBack. Die Nutzer wünschten sich zudem erweiterte Funktionen wie die Möglichkeit, häufig genutzte Ziele wie z. B. „Zuhause“ oder „Arbeit“ zu kennzeichnen und die Navigation dorthin direkt von der Startseite aus starten zu können.

8.8.2 Vibrotaktile Nutzerinteraktion

Die Nutzerstudie zur Interaktion mit dem vibrotaktilen Gürtel evaluierte dessen neu entwickelte Funktionen zur Hindernisinformationen (HI) und Richtungsanweisung im direkten Vergleich mit dem bestehenden Assistenzsystem in Form eines Rucksacks des FAPS mit verbauten Vibrationsmotoren in den Schultergurten. Zehn Probanden absolvierten die Navigationsaufgaben blind in einer simulierten Gazebo-Simulationsumgebung, wobei der Gürtel zur sensorischen Augmentation neben Navigationshinweisen auch Hindernisinformationen mittels fünf Vibrationsmotoren (Abdeckungsbereich 112,5°) im Bauchbereich vermittelte. Die quantitativen Ergebnisse zeigten eine Überlegenheit des

Gürtels: Das Zeitverhältnis (tatsächliche benötigte Zeit gegenüber geplanter Zeit) war mit einem Durchschnitt von 6,10 ± 2,50 (ohne HI) bzw. 6,41 ± 2,57 (mit HI) 2,3 Mal schneller gegenüber dem Rucksack (14,51 ± 4,16), was auf eine deutlich höhere Effizienz und schnellere Reaktionszeit hindeutet. Das Wegverhältnis des Gürtels lag in beiden Konfigurationen sehr nah am Idealpfad (0,99), was eine hohe Pfadtreue belegt. Subjektiv bewerteten die Probanden die Richtungsanweisungen des Gürtels als intuitiv (Mittelwert 5,0 von 5 möglichen Punkten) und eindeutiger als die des Rucksacks, und sie fühlten sich mit dem Gürtel sicherer. Die aktivierte Hinderniserkennung reduzierte die Anzahl der Kollisionen auf komplexen Pfaden signifikant um über 61%, obwohl die Aktivierung zu einer leichten Verzögerung bei dem Verständnis der Richtungsanweisungen führte. Verbesserungspotenzial besteht laut den Teilnehmern in der Anpassung der Vibrationsmuster auf pulsierende Signale für die Hinderniserkennung und in der Verbesserung der Passform des Gürtels, um die Vibrationswahrnehmung zu optimieren.



Rucksack | Gürtel ohne Hindernisinformationen | Gürtel mit Hindernisinformationen

Abbildung 14: Vergleich zwischen vorgesehenen und tatsächlichen Pfad der Probanden für das alte Rucksacksystem und den Vibrationsgürtel ohne und mit aktivierter Hindernisaugmentation.

8.8.3 3D Audio

Die Evaluation des entwickelten 3D-Audio-Navigationssystems fand im Rahmen einer systematischen Probandenstudie mit insgesamt acht Teilnehmern statt, wobei die Versuche zur Simulation der Sehbeeinträchtigung in einer kontrollierten Simulationsumgebung in Gazebo mit lichtundurchlässigen Brillen durchgeführt wurden. Das System übermittelte Navigations- und Hindernisinformationen akustisch über Knochenschallkopfhörer. Die quantitative Analyse der Navigationsleistung belegte die hohe Präzision und Effizienz des 3D-Audio-Systems. Bei längeren und komplexeren Routen lag die mittlere Abweichung vom Sollpfad (MF) lediglich bei 0,05 m, und die Aufgaben wurden schnell bewältigt, wobei die Ausführungszeit für die komplexeste Route 162 Sekunden betrug. Die subjektive Nutzerbewertung, basierend auf einer Likert-Skala (1 = stimme voll zu), zeigte eine hohe Akzeptanz: Die Probanden bewerteten das System als „insgesamt leicht zu verstehen“ (Mittelwert 1,63) und die akustischen Hinweise als „intuitiv nachvollziehbar“ (Mittelwert 1,63). Die räumliche Auflösung der 3D-Audio-Signale wurde ebenfalls positiv hervorgehoben, da die Teilnehmer „gut erkennen, aus welcher Richtung das Signal kam“ (Mittelwert 1,88). Darüber hinaus wurde die Audio-Navigation als „hilfreicher

für komplexe Umgebungen“ eingeschätzt (Mittelwert 1,88), und die verwendeten Knochenschallkopfhörer wurden als angenehm zu tragen empfunden (Mittelwert 1,88).

9 Voraussichtlicher Nutzen

Die gewonnenen Erkenntnisse sind für das Gründerteam von strategischer und technischer Bedeutung. Die Ergebnisse aus den Workshops und Umfragen bestätigen, dass die Aufnahme des Assistenzsystems in den Hilfsmittelkatalog der Krankenkassen als essenzielle Anforderung für eine gerechte Zugänglichkeit zu behandeln ist, was die wirtschaftliche Ausrichtung des Teams festigt. Technisch profitierte das Team von der Validierung zentraler Designentscheidungen: Die entwickelte vibrotaktile Gürtellösung wurde von Probanden bezüglich der Klarheit und Eindeutigkeit der Richtungsanweisungen als deutlich intuitiver und besser bewertet als das Rucksacksystem. Zudem konnte gezeigt werden, dass die aktivierte Hinderniserkennung des Gürtels die Anzahl der Kollisionen signifikant reduziert. Die Resultate liefern dem Team konkrete, validierte Bausteine für die weitere Produktentwicklung.

Die im Rahmen des LOMOBI-Projekts erzielten Ergebnisse liefern dem Lehrstuhl FAPS signifikante Erkenntnisgewinne. Dieser wissenschaftliche Nutzen basiert auf der echtzeitfähigen Anwendung von Machine Learning (ML)-Methoden insbesondere im Kontext der Umgebungssegmentierung, Lokalisation und autonomen Navigation in unterschiedlichsten Szenarien. Konkrete Lösungen wurden für spezifische Forschungsfragen entwickelt, darunter die Gestaltung effizienter und selbstlernender Segmentierungsmethoden von Umgebungsinformationen, die dynamische Fusion vielfältiger Sensordaten zur Gewährleistung sicherer Navigation, sowie die intuitive, interaktive und adaptive Kommunikation von Informationen des Nutzenden, der Umgebung und der Sensorik. Diese gewonnenen Erkenntnisse fließen in wissenschaftliche Publikationen, Konferenzen und Dissertationen ein und werden Studierenden über vielfältige Lehrveranstaltungen vermittelt.

10 Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf diesem Gebiet bei anderen Stellen

Nach dem Kenntnisstand des Projektteams existieren zum Zeitpunkt der Verfassung des Abschlussberichts keine Assistenzsysteme, die eine umfassende Unterstützung für sehbeeinträchtigte Personen im Alltag in dem angestrebten Umfang gewährleisten. Im erweiterten Themenfeld der Assistenztechnologien für sehbeeinträchtigte Menschen gab es Fortschritte, insbesondere im Bereich kamerabasierter Ansätze zur Hinderniserkennung und spezialisierter Navigationshilfen, wobei diese bestehenden Forschungsansätze stets nur die Navigation in einzelnen Szenarien adressieren. Mit den relevanten Akteuren in dem Feld, insbesondere betroffenen Personen, die in partizipativen Workshops und Umfragen eingebunden wurden (darunter Kooperationen mit regionalen Vereinen), ist die Forschungsgruppe eng vernetzt und in regelmäßigem Austausch.

11 Erfolgte oder geplante Veröffentlichung des Ergebnisses

Folgende Veröffentlichungen sind im Rahmen des Forschungsprojekts LOMOBI entstanden oder befinden sich zum Berichtszeitpunkt im Prozess der Veröffentlichung:

- Kalenberg, M., Lieret, M., Hofmann, C., & Franke, J. (2023, July). A multimodal a* algorithm to solve the two-dimensional optimization problem of accompanying a person for an intelligent wheelchair. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 1-6). IEEE.
- Kalenberg, M., Hofmann, C., Martin, S., & Franke, J. (2024, February). Human Comfort Factors in People Navigation: Literature Review, Taxonomy and Framework. In *International Conference on Robotics, Computer Vision and Intelligent Systems* (pp. 225-243). Cham: Springer Nature Switzerland.
- Engelhardt, H., Kalenberg, M., Franke, J., & Martin, S. (2025, February). Learn Where I Can Walk: Auto-labeling of Walked Areas Using Monocular Camera Trajectory. In *International Conference on Robotics, Computer Vision and Intelligent Systems* (pp. 348-363). Cham: Springer Nature Switzerland.
- Kalenberg, M., Gutzeit, M., Engelhardt, H., Franke, J., Martin, S., (2025, Dezember). Ethical, Legal, and Social Implications of Electronic Travel Aids for Visually Impaired People . In 2025 3nd International Conference on Integrated Systems in Medical Technologies (ISMT)