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

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 erforscht. 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 stehen 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.

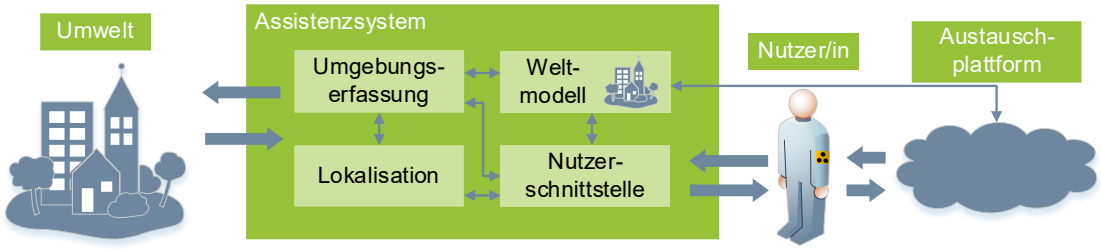

Abbildung 1: Darstellung der Interaktion des Assistenzsystems mit Umwelt und sehbehinderter Person zur Unterstützung von Orientierung und Mobilität

Das LOMOBI-System (Abbildung 1) ist als interaktives Gesamtsystem konzipiert. Es erfasst die "Umwelt" mittels "Umgebungserfassung" und "Lokalisation". Die "Nutzer/in" interagiert direkt mit dem System sowie mit einer "Austauschplattform". Diese Plattform ermöglicht die Vernetzung der Nutzenden sowie den Austausch von Daten, Karten und spezialisierten Modellen.

Die Forschungsschwerpunkte lagen auf der Entwicklung spezialisierter Methoden des maschinellen Lernens (ML) zur effizienten Umgebungsinterpretation und der adaptiven Gestaltung der Mensch-System-Interaktion. Die Navigationsinformationen werden über interaktive Schnittstellen kommuniziert, die sich flexibel an die Fähigkeiten und Präferenzen der Nutzenden anpassen lassen.

Die Entwicklung folgte einem strukturierten Arbeitsplan, der sich am User-Centered-Design orientierte. Zu Beginn stand die Anforderungsanalyse, für die ein ELSI-Workshop (Ethische, Legale und Soziale Implikationen) mit 12 Personen, darunter drei Blinde, durchgeführt wurde. Als wichtigste identifizierte Anforderungen galten unter anderem das nutzerbestimmte Ausschalten, die Implementierung von Kartenmaterial öffentlicher Gebäude, das Erkennen von Hindernissen auf Kopfhöhe und die Mitteilung an Nutzer bei einem Systemausfall.

Darauf aufbauend erfolgten der Systementwurf und die Integration. Es wurde eine modulare Softwarearchitektur im Robot Operating System (ROS) implementiert, die vier Hauptkomponenten umfasst: Umgebungserfassung, Lokalisierung/Kartierung, Pfadplanung und Feedbackgenerierung. Parallel dazu lag der Fokus auf der Umgebungssegmentierung, um eine robuste Segmentierung von Wegen und Hindernissen mittels maschinellem Lernen (ML) zu erreichen. Ein weiteres Kernstück war

die Implementierung der Nutzerinteraktion. Basierend auf der im ELSI-Workshop identifizierten Notwendigkeit einer hohen Individualisierung wurden dafür multimodale Kommunikationsschnittstellen entwickelt.

Die Lokalisierung wurde spezifisch für unterschiedliche Szenarien konzipiert. Für die Indoor-Lokalisierung, bei der kein GNSS verfügbar ist, basiert das System auf einer Stereokamera (D435) und einer Inertial Measurement Unit (IMU). Mittels RTAB-Map, eines quelloffenen SLAM-Frameworks (Simultaneous Localization and Mapping), erstellt das System Karten der Umgebung und bestimmt die eigene Position darin. Diese Karten können gespeichert und über die Austauschplattform geteilt werden. Im Gegensatz dazu basiert die grobe Outdoor-Lokalisierung im Freien auf GNSS (z. B. GPS). Um die Genauigkeit zu verbessern, werden diese Daten mittels eines erweiterten Kalman-Filters (EKF) mit den Daten der IMU und der visuellen Odometrie (VO) der Kamera fusioniert. Zusätzlich wurde OrienterNet evaluiert, eine KI-gestützte visuelle Lokalisierung, die die Position anhand von Kamerabildern in OpenStreetMap (OSM)-Karten schätzt und den Lokalisierungsfehler reduziert.

Im Bereich der Umgebungsinterpretation und Pfadplanung wurde zunächst eine Umgebungsklassifikation entwickelt. Zur Unterscheidung verschiedener Navigationsumgebungen, wie Innenräume oder städtische Straßen, wurden vier neuronale Netzarchitekturen getestet, wobei SE-ResNeXt-50 die besten Ergebnisse erzielte. Zur Verbesserung der Umgebungsinterpretation wurde eine automatisierte Trainingsdatengenerierung entwickelt. Mithilfe des CARLA-Simulators konnten über 10.000 synthetische, fußgängerzentrierte Szenen erzeugt und mit realen Aufnahmen kombiniert werden, was die Modellrobustheit erhöhte. Ein neuartiger Auto-Labeling-Ansatz nutzte den tatsächlich gegangenen Pfad zur automatischen Segmentierung begehbarer Flächen und erzielte eine mittlere IoU von 93,9 %. Ergänzend kam eine Zero-/Single-Shot-Segmentierungspipeline (Grounded SAM 2, SegGPT) zum Einsatz, um Daten mit hoher Unsicherheit automatisiert zu annotieren. Für die Pfadplanung nutzt LOMOBI einen dreistufigen hierarchischen Ansatz. Die oberste semantische Ebene nutzt globale Daten wie OpenStreetMap (OSM) sowie "Funktionale Landmarken" (z.B. Türen, Treppen, Ampeln), an denen eine Nutzeraktion erforderlich ist. Darunter berechnet die mittlere, globale Ebene im Nav2-Framework den Pfad zur nächsten Landmarke. Die unterste, lokale Ebene nutzt Echtzeit-Sensordaten, um den Pfad zu verfolgen und dynamischen Hindernissen auszuweichen. Um komplexe Handlungsabfolgen an diesen funktionalen Landmarken zu steuern, wie das Warten an einer Ampel, werden Behavior Trees (Verhaltensbäume) eingesetzt. Diese gelten als modularer und reaktiver als traditionelle Zustandsautomaten. Zur Erkennung des Zustands der Landmarken, etwa ob eine Ampel rot oder grün ist oder eine Tür geöffnet ist, wird ein trainiertes YOLO-Netzwerk verwendet.

Die Kommunikation der Navigationsinformationen an den Nutzer erfolgt multimodal. Für das taktile Feedback wurde als zentrales Element ein Navigationsgürtel integriert; dieser vibrotaktile Gürtel übermittelt Richtungsanweisungen und Warnungen vor Hindernissen. Ergänzend können optional Knochenkopfhörer für 3D akustisches Feedback genutzt werden. Diese geben räumliche Signale, ohne dabei die wichtige auditive Wahrnehmung der Umgebung zu blockieren.

Durchgeführte Probandenstudien validierten zentrale Designentscheidungen. Die vibrotaktile Gürtellösung wurde im Vergleich zur vorherigen Lösung der integrierten Vibrationsmotoren in den Schulterträger des Rucksacksystem als deutlich intuitiver und besser für Richtungsanweisungen bewertet. Zudem reduzierte die Hindernisaugmentation des Gürtels die Anzahl der Kollisionen.