eLabrador: A Wearable Navigation System for Visually Impaired Individuals

Meina Kan, *Member, IEEE*, Lixuan Zhang, *Student Member, IEEE*, Hao Liang, Boyuan Zhang, Minxue Fang, Dongyang Liu, Shiguang Shan, *Fellow, IEEE*, and Xilin Chen, *Fellow, IEEE*

***Abstract*— Visually impaired individuals encounter significant challenges when walking and acting in unfamiliar environments, particularly in outdoor scenarios. The complexity of outdoor environments, characterized by diverse obstacles, traffic signals, and societal norms, poses substantial barriers to mobility of visually impaired individuals and makes long-distance walking especially arduous. Although GPS-based navigation systems can facilitate long-distance travel, they often suffer from location inaccuracies in urban areas and even completely fail indoors. Moreover, these systems lack the capability to provide detailed information about walkways and immediate surroundings, which are crucial for safe and efficient walking. To address these limitations, we introduce a proof-of-concept wearable navigation system named eLabrador, designed to assist visually impaired individuals in long-distance walking in unfamiliar outdoor environments. The eLabrador integrates public maps (e.g. Amap or Google Maps) and GPS for global route planning, while leveraging computational visual perception to provide precise and safe local guidance. This hybrid approach enables accurate and safe navigation for visually impaired individuals in outdoor scenarios. Specifically, the eLabrador utilizes a head-mounted RGB-D camera to capture environmental geometric terrain and objects in outdoor urban environments. These inputs are processed into a 3D semantic map, offering a detailed representation of the surrounding environment. The planning module then integrates this 3D semantic map with route information from the global map (i.e. Amap) to generate an optimized walking path. Finally, the interaction module utilizes the audio-haptic dual-channel to relay navigation instructions to visually impaired user. Together, these three modules work seamlessly to facilitate long-distance navigation for visually impaired individuals in outdoor environments. The eLabrador is evaluated with two real-world outdoor scenarios, involving 10 visually impaired and visually masked participants. The experiments show that eLabrador successfully guides visually impaired participants to their destinations in outdoor environments. Additionally, the eLabrador provides descriptive information about landmarks and other navigation cues, helping visually impaired users better understand their surroundings. Subjective evaluations further indicate that most participants felt a sense of safety and reported an acceptable cognitive load during navigation, indicating its usability and effectiveness.**

***Note to Practitioners*—Visually impaired individuals almost cannot walk long distance in unfamiliar outdoor environments. Without proper assistance, their mobility and quality of life can be severely impacted. To address this issue, this article presents a wearable navigation system eLabrador to assist visually impaired individuals in walking outdoors, such as traveling from a residential entrance to a nearby park. Experimental results from real-world scenarios involving 10 participants demonstrate that eLabrador safely guides visually impaired users to their destination, significantly enhancing their mobility and independence.**

***Index Terms*— Wearable navigation system, navigation for visually impaired, outdoor navigation, eLabrador.**

I. Introduction

ACCORDING to the Vision Loss Expert Group, more than 338 million people worldwide experience moderate to severe visual impairment, with 43 million suffering from significant vision loss [1]. Visual perception is essential for daily activities, and visual impairment severely limits an individual's ability to perceive the environment. This impairment adversely affects various aspects of life, including mobility, behavior, travel, and social interactions [2]. Previous research [3] shows that severe vision loss significantly reduces the mobility of visually impaired individuals, diminishing their confidence in participating social activities. Consequently, over 30% of visually impaired individuals rarely engage in independent outdoor activities [4]. Moreover, visual impairment increases the risk of suffering physical and mental health issues [5]. Given these issues, the development of an intelligent navigation system for visually impaired individuals holds substantial social and technological importance.

In recent years, researchers have proposed numerous assistive systems for visually impaired individuals, with a particular focus on indoor environments [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23].

Received 26 July 2024; revised 28 October 2024; accepted 28 December 2024. Date of publication 11 February 2025; date of current version 16 April 2025. This article was recommended for publication by Associate Editor X. Li and Editor L. Zhang upon evaluation of the reviewers' comments. This work was supported in part by the National Science and Technology Major Project under Grant 2021ZD0111901 and in part by the Natural Science Foundation of China under Grant U2336213 and Grant 62122074. *(Corresponding author: Xilin Chen.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of the Institute of Computing Technology, Chinese Academy of Science under Application No. JLS2023.

Meina Kan, Lixuan Zhang, Hao Liang, Boyuan Zhang, Minxue Fang, Dongyang Liu, and Xilin Chen are with the State Key Laboratory of AI Safety, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: kanmeina@ict.ac.cn; lixuan.zhang@vipl.ict.ac.cn; hao.liang@vipl.ict.ac.cn; zhangboyuan17@mails.ucas.ac.cn; fangminxue17@mails.ucas.ac.cn; liudongyang21s@ict.ac.cn; xlchen@ict.ac.cn).

Shiguang Shan is with the State Key Laboratory of AI Safety, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, also with the University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Peng Cheng National Laboratory, Shenzhen 518055, China (e-mail: sgshan@ict.ac.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TASE.2025.3541055, provided by the authors.

Digital Object Identifier 10.1109/TASE.2025.3541055

For example, Lee et al. [11] proposed a wearable navigation system based on RGB-D camera for indoor wayfinding and obstacle avoidance. Guerreiro1 et al. [15] developed "Cabot," a suitcase-shaped navigation robot, to enhance indoor mobility. Despite significant advancements in indoor navigation, research on outdoor scenarios remains relatively limited in both depth and breadth [24], [25], [26], [27], [28], [29], [30]. For example, Kammoun et al. [24] adopted the GPS to assist the visually impaired individuals in outdoor navigation, however the system provided only coarse directional cues, lacking detailed guidance. Ni et al. [25] developed a cart-shaped robot to improve the wayfinding abilities of visually impaired individuals in simple campus environments. Similarly, Duh et al. [28] introduced V-Eye, a vision-based wearable navigation system designed for outdoor use, however it requires predefined landmarks, restricting its usability primarily to university environments. The Augmented Cane [29] was also designed for outdoor navigation, however its reliance on a 2D LiDAR sensor limited its ability to detect obstacles at head height.

In general, these navigation systems either rely on coarse GPS-based navigation [24], [29] which lack precision for detailed guidance, or are limited to small, confined areas such as indoor spaces or school campuses [11], [25]. Consequently, both approaches face significant limitations in supporting long-distance navigation for visually impaired individuals in unfamiliar outdoor environments.

To guide visually impaired individuals in long-distance walking through unfamiliar outdoor environments, we follow the common framework shown in Fig. 2 and elaborately refine each module to find feasible local walking path based on a GPS route. The core innovation of our work lies at the system level, particularly in *Global-Local Collaborative Navigation*, which simultaneously leverages both global route map (e.g. Amap[1] or Google Maps) and local semantic map. Although the adjustments within each module may appear minor in isolation, together they form a cohesive system that enables visually impaired individuals to successfully walk in unfamiliar outdoor environments. A summary of each module's design is given below:

- **In the perception module**, we utilize the existing mask2former [31], [32] to obtain semantic information about the environment and and employ VINS-MONO [33] for odometry estimation, thereby generating a 3D semantic map of the user's local surroundings.
- **In the planning module**, we propose a newly designed Direction A* algorithm (Dir-A*), which extends the original A* method to find a local walking path using coarse guidance from route planner. As the route provided by Amap (comprising a set of GPS waypoints) typically lies on driving road, directly following those waypoints would be dangerous for pedestrians. To address this issue, Dir-A* refines the walking path from driving road to safe sidewalk, and ensure safer navigation.
- **In the interaction module**, we design an audio-haptic dual-channel interaction mechanism to deliver navigation instructions and key object information, helping visually impaired users follow the planned path properly. The haptic channel provides direction and timely prompts with low cognitive load, while the audio channel conveys richer semantic details such as object names, road signs, and system notifications.

[1]https://ditu.amap.com/, Amap is also known as Gaode in China.

II. Related Work

Over the past years, several navigation systems have been developed to assist visually impaired individuals in both indoor and outdoor environments, thereby improving their mobility, confidence, and self-esteem. Among these efforts, smart canes, navigation robots, and wearable devices have emerged as the most popular ones. Each of these systems offers unique advantage that improve the independence and quality of life for visually impaired individuals.

A. Smart Cane

The white cane is one of the most widely used assistant tools within the visually impaired community. Through frequent physical contact with their surroundings, visually impaired individuals can sense the terrain and avoid obstacles effectively. Despite its benefits, the white cane alone offers limited assistance and may not always provide sufficient guidance, even in familiar environments.

To overcome the white cane's limitation in exploration range, researchers proposed various smart cane designs, which integrate multiple sensors (including sonar [34], infrared [35], and ultrasonic [36]) to gather more environmental information. These additional sensors allow smart canes to detect obstacles of varying heights and at longer distance, thus enhancing their overall utility. Among these solutions, Agrawal et al. [21] mounted an RGB-D camera on a white cane to search nearby seats. Similarly, the GuideCane [37], [38] employed multiple ultrasonic sensors for nearby obstacle detection. These sensors enable the perception of the surroundings at a distance without requiring direct physical contact with the environment, and thus improve the capability of mobility, obstacle avoidance, and indoor wayfinding for visually impaired individuals.

Moreover, some cane-based navigation systems enhance the white cane with robotic mechanisms to provide smart assistance. For instance, Aigner and McCarragher [39] developed a Robotic Cane featuring a wheel at its end, which steers the user in the correct direction upon detecting obstacles or deviations from the intended path. This system also designs a shared-control framework, allowing the user to override or compromise with the autonomous commands and ensuring that the device helps rather than hinders. The Co-Robotic Cane (CRC) [10] and the Robotic Navigation Aid (RNA) [16] incorporate RGB-D cameras for pose estimation, object recognition, and path planning. More recently, Slade et al. [29] proposed a promising Augmented Cane, which combines multiple sensors (e.g. 2D LiDAR, camera, and GPS) to offer comprehensive assistance for visually impaired individuals, including obstacle avoidance, indoor and outdoor wayfinding, and key-object detection. In addition, Ranganeni et al. [23] and Zhang et al. [40] investigated varying levels of shared control

between humans and robots, highlighting how different control schemes affect users' sense of agency.

By incorporating additional sensors and robotic mechanisms, these enhanced white cane–based systems offer substantial support for visually impaired individuals. However, as noted by users in [29], over installed sensors will increase the cane's weight, diminishing its portability and suitability for daily use. Moreover, a heavier cane may reduce the frequency of environmental contact, thereby discouraging active exploration and interaction.

B. Navigation Robot

The guide dog is another widely used partner for visually impaired individuals. However, training a guide dog typically requires about six months and costs around $50,000, while it can only serve for six to seven years [23]. To address these limitations, several navigation robots have been developed. Kayukawa et al. [17] proposed BlindPilot, a cart-like robotic platform [6], [14], [25] designed to guide visually impaired users to landmark objects. This system detects and estimates target objects (e.g. an empty seat) using an RGB-D camera and constructs a 2D map of the surroundings with a LiDAR sensor. Once the target is recognized, the robot plans an obstacle-free path and navigates the user reach the destination through the handle. Guerreiro et al. [15] implemented the Cabot, a suitcase-shaped navigation system that allows users to walk alongside it while holding the handle for added confidence and safety. This design has since been adopted to guide visually impaired individuals in public spaces [18] and unfamiliar buildings [22]. Meanwhile, Xiao et al. [41] implemented the Robotic Guide Dog, which is connected to the user via a leash, enabling navigation through narrow and cluttered spaces. Chen et al. [42] further developed a controllable traction device with adjustable length and force between the user and the robot to ensure user comfort. Additionally, Balatti et al. [43] proposed a guidance planner that accounts for both the robot and the user, ensuring a collision-free path.

These navigation robots enhance mobility and confidence for visually impaired users, particularly on flat ground. However, such systems often underperform on the uneven terrain commonly found in outdoor environments. Although quadrupedal robots can mitigate this issue, their downtime can still pose problems for users. Furthermore, while the combination of white canes and robotic systems (e.g. Robotic Cane [39] and Augmented Cane [29]) has significantly benefited the visually impaired individuals, the added weight and the constant need to occupy one hand remain barriers to active exploration. To address these fundamental challenges, our research focuses on wearable assistive technology that achieves seamless integration with daily activities while maintaining unimpeded mobility.

C. Wearable Devices

In recent years, a few wearable devices have emerged, such as goggles, helmets, smartphones, and backpacks. They naturally blend with both users and their environments. These devices present a promising choice for navigation systems that are designed to assist visually impaired individuals. Previous work [44] also highlights the importance of providing comfortable assistance while mitigating social stigma concerns.

Smartphones equipped with integrated cameras, GPS, and Inertial Measurement Units (IMUs) provide a variety of convenient aids for visually impaired users, including obstacle avoidance [45], indoor self-localization [46], [47], wayfinding [48], and pedestrian/intersection detection [20], [49]. Chen et al. [26] designed an extensible Android-based system offering features such as messaging, street-view descriptions, and navigation to specific destinations. Furthermore, Duh et al. [28] proposed V-Eye, which integrates a global localization method (VB-GPS) and image-segmentation techniques to improve scene understanding using a single camera. V-Eye provides location and orientation information, detects unexpected obstacles, and supports navigation in both indoor and outdoor environment. Although these outdoor navigation systems enhance mobility for visually impaired individuals, they have primarily been tested on university campuses rather than urban roads. This limitation highlights the need for systems tailored to the complexities of urban environments. Accordingly, one of our primary contributions is designing a wearable navigation system specifically for urban environments.

Head-, wrist-, or chest-mounted sensors represent another common design in wearable navigation systems. For instance, Blessenohl et al. [50] developed a helmet-mounted, camera-based solution to measure the depth of surrounding floors and walls. Lee et al. [11] designed an RGB-D camera–based indoor navigation system that relies on a camera and an IMU for user position estimation. Wang et al. [13] attached a depth camera to the chest as the main perception sensor, enabling the system to identify walkable areas, plan step-by-step movements, and recognize key objects in indoor environments. Similarly, Ma et al. [30] used a shoulder-mounted camera to detect important objects in the surroundings.

In addition, smart goggles equipped with cameras provide another solution for assisting visually impaired individuals. Al-Khalifa [51] implemented an Arabic navigation system called Ebsar, which uses Google goggles to support indoor navigation. In this system, a sighted individual pre-builds the building map, and the visually impaired user's location is determined by scanning Quick Response (QR) codes. Meanwhile, Katzschmann et al. [52] proposed a contactless, intuitive, hands-free, and discreet wearable device featuring a sensor belt fitted with an array of LiDARs to measure distances to nearby obstacles and walls, thus facilitating obstacle avoidance. Liu et al. [53] further developed an RGB-D goggles to assist visually impaired individuals in jogging.

To summarize, smartphone is an excellent wearable navigation option attributing to its lightweight and equipped multiple sensors. However, using a smartphone's camera requires hanging or holding it around chest level, bringing an extra burden on hands or neck [20], [49]. Consequently, hands-free, head-mounted cameras or smart goggles serve as more convenient alternatives. In this work, we adopt a head-mounted camera as it offers an effective proof-of-concept solution and can be seamlessly adapted to lightweight smart goggles in future.

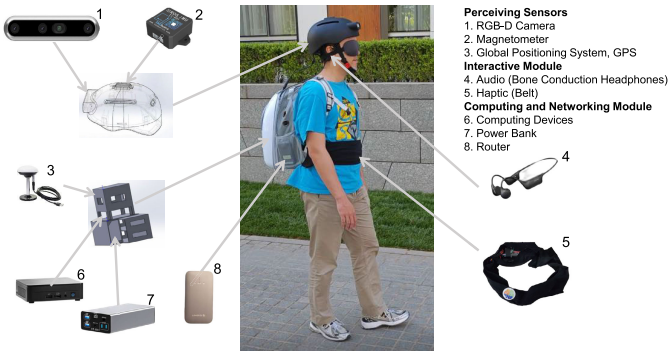


Fig. 1. Hardware configuration of the eLabrador.

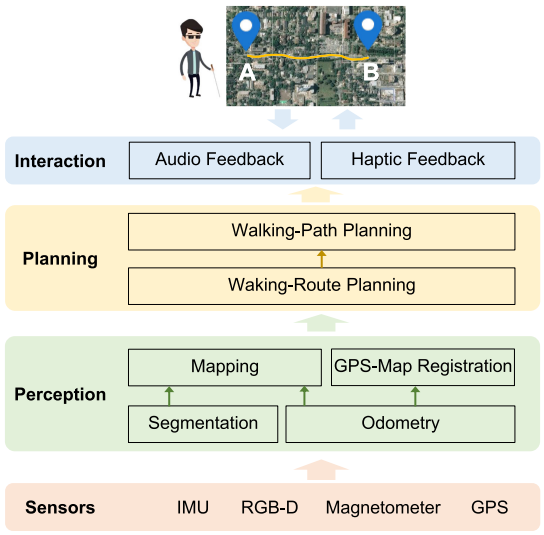


Fig. 2. Information processing procedure of the eLabrador.

III. System Design

Although some efforts on assistive systems have been proposed to aid the visually impaired individuals, the goal of achieving complete independence in their daily lives remains a big challenge. In this work, we contribute to the field by proposing and implementing the eLabrador, a wearable navigation system designed to assist visually impaired individuals in long-distance walking in unfamiliar outdoor environments. The system integrates hardware, software, and network connectivity, working in unison to capture environmental data, process it, and deliver actionable instruction to the user.

The system operates through a coordinated workflow. First, multiple sensing inputs, from an RGB-D camera with an Inertial Measurement Unit (IMU), a magnetometer, and a Global Positioning System (GPS) unit, are used to capture environmental information. Then, perception and planning modules analyze these inputs to generate concise and informative guidance instructions, such as optimized walking path and direction. Finally, these guiding instructions are delivered to visually impaired users through audio and haptic channels, ensuring a safe and efficient navigation. The details about the whole system are described in the following sub-sections.

A. Hardware Configuration

As illustrated in Fig. 1, the system hardware includes three primary components: perceiving sensors, computing and networking module, and interactive module (including audio and haptic vibrators).

The perceiving sensors consist of an RGB-D camera, a magnetometer, and a Global Positioning System (GPS) unit. The RGB-D camera, a RealSense D455, captures RGB-D images of the surroundings with a wide field of view [54] and is integrated with an Inertial Measurement Unit (IMU) to measure the head/body's pose, including position and orientation. The magnetometer determines the user's orientation within the east-north-up coordinate system, while the GPS provides the user's position in terms of latitude and longitude based on the World Geodetic System-1984 Coordinate System (WGS84). All sensors, except for the GPS, are mounted within a customized helmet. The GPS is relocated to the accompanying backpack due to its larger size.

The portable computer is an Intel NUC 12 mini PC equipped with Intel Core i7-1260P processors, chosen for its compact size and processing power. Power is supplied by a portable power bank with a capacity of 260 watt-hours, ensuring extended operational time. For network connectivity, a wireless router with 4G LTE-FDD is included, offering a maximum bandwidth of approximately 150 Mbps. The portable computer, wireless router, power bank, and the GPS unit are housed within a backpack, as illustrated in Fig. 1.

The interactive module (including audio and haptic vibrators) consists of bone-conduction headphones and an elastic belt, designed to deliver audio and haptic guidance instructions to visually impaired users. Bone-conduction headset is chosen since it does not obstruct the user's sense of hearing, allowing them to remain aware of environmental sounds [55]. The elastic belt delivers navigation instructions with haptic feedback using five vibrators evenly distributed across the front of the belt. These vibrators are controlled by an Arduino Mega 2560 embedded board. Additionally, the belt is made of elastic material to ensure it can comfortably accommodate individuals with varying waist sizes.

Overall, the weight of the whole navigation system (excluding the helmet) is approximately 1.76 kg, with the 260 watt-hours power bank contributing an additional 1.58 kg. The power bank capacity can be adjusted as needed to reduce weight. Under full load, the portable power bank provides a minimum of 7 hours of continuous operation.

B. Procedure of Information Processing

From sensoring to action instruction, three modules, perception, planning, and interaction, are involved for information processing, as illustrated in Fig. 2. These modules collaborate to generate navigation instruction that guides visually impaired individuals to walk independently from a starting point (location A) to a destination (location B).

Among these modules, *the perception module* aims to perceive the user's position and surroundings. Specifically, this module processes data from the sensors, constructs a 3D semantic map of the surrounding environment, and estimates the self-position of the visually impaired user. Additionally,

it performs GPS-Map Registration to align the east-north-up coordinate of the GPS with the coordinate of the local surrounding map. Then, *the planning module* is responsible for optimizing a practical walking path by considering both the global GPS-based route and local sidewalk conditions, such as obstacles, traffic signals, and pedestrians. Leveraging the aligned GPS data, self-position, 3D semantic map, and destination, this module performs walking-route planning followed by walking-path planning to determine the specific walking path. Finally, *the interaction module* is responsible for delivering navigation instruction from the planning module to visually impaired users through audio and haptic channels. Additionally, this module accepts user voice commands, such as destination changes.

C. Network Connectivity

To balance the local computing power and portability of eLabrador, we adopt an edge-cloud collaborative computing design. The modules of information processing in eLabrador are strategically divided between the cloud and edge devices based on their importance and time complexity. This design allows the system to leverage the computational power of the cloud for faster processing while ensuring critical safety and essential functions remain operational even if the cloud connection fails. Specifically, tasks such as segmentation, mapping, and path search step of path planning are deployed on the cloud (a Lenovo Y9000K Laptop with the RTX 3080), while other modules are deployed on the edge (an Intel NUC 12 mini PC with Intel Core i7-1260P Processors). This design ensures that the onboard computing device remains lightweight and does not impose a significant burden on the user.

Additionally, based on our experiments, the bandwidth of the eLabrador ranges from 8 to 16 Mbps, which is stably provided by major telecommunication providers. However, in cases where Internet or mobile connections are poor, GPU-based modules in the cloud, such as segmentation, may fail to function. To address this, our system periodically monitors network connectivity, and the system will switch to navigation without those modules in the cloud when a network failure is detected.

In eLabrador, all sensors are connected to the mini PC via USB 2.0 (for the GPS and magnetometer) or USB 3.0 (for the camera). Haptic vibrators is transmitted to the belt using an HC-05 Bluetooth 2.0 module, while audio feedback is delivered to the bone-conduction headphones via a Bluetooth 5.3 module.

IV. Details of Software Modules

The problem that our eLabrador aims to address is formulated as a prompt-driven navigation problem. This problem takes continuous sensor observations and destinations as input, and generates actionable instruction as output, as illustrated in Eq. (1) and Fig. 3.

$$f(O, P) \rightarrow A, \tag{1}$$

where O represents the sensor observations, consisting of RGB-D image, IMU, and Magnetometer data; P represents the destinations; A represents the action instruction, including audio and haptic ones. To address the prompt-driven navigation problem, the software in our system contains several functional modules to enable visually impaired individuals to walk long distance in outdoor environment. These modules include perception, global-local collaborative planning, and dual-channel interaction. These modules form a pipeline as in Fig. 2, with each module meticulously designed to ensure optimal functionality. The novelty of eLabrador lies at system level, particularly in the aspect of *Global-Local Collaborative Navigation*, which integrates both global route map (i.e. Amap) and local semantic map to provide accurate navigation guidance.

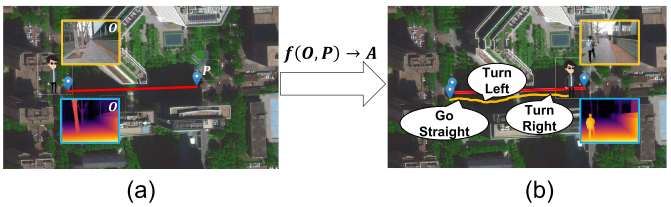


Fig. 3. Illustration of the prompt-driven navigation problem. (a) shows the inputs, including the sensor observations O and prompt P, (b) shows the output, i.e. the action instruction A, which is delivered to the visually impaired user through audio and haptic channels. 'yellow line': the actual trajectory the visually impaired user walks.

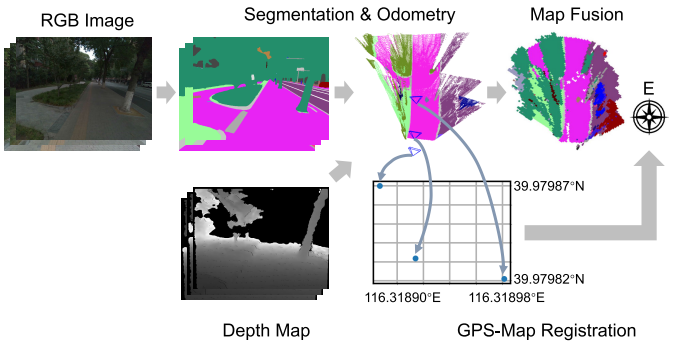


Fig. 4. Illustration of the workflow in the perception module.

A. Visual Perception Module

Visual perception module is designed to extract both geometric and semantic information of the user's local surrounding environment. This includes localizing the sidewalk region, detecting obstacles, and recognizing key objects, all of which are essential for enabling safe and efficient navigation. As illustrated in Fig. 4, the perception module is composed of four sequential sub-modules: image segmentation, odometry, map fusion, and GPS-Map registration. Together, these sub-modules work to generate a 3D semantic map, which is described in detail below.

1) Semantic Segmentation: The first sub-module is image segmentation, which takes synchronized RGB and depth images as inputs and generates a point cloud enriched with semantic information. Specifically, the state-of-the-art image semantic segmentation method, Mask2Former [31], is employed to produce semantic categories and classification probabilities for each pixel in the image. Mask2Former is

pre-trained on the Mapillary Dataset [56], a street-level dataset containing 103 categories. Subsequently, using the depth map D and camera intrinsic matrix $\mathbf{K}$, each pixel in the image is accurately mapped to a 3D coordinate in the camera coordinate system, resulting in semantically annotated point clouds. The intrinsic matrix $\mathbf{K}$ of the camera is defined as:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

where f_x and f_y represent the focal lengths in the x and y directions, respectively, and c_x and c_y represent the coordinates of the principal point. For a given pixel (u, v) in the image plane with a depth value d, obtained from depth camera the corresponding 3D point (X, Y, Z) in the camera coordinate system can be computed as follows:

$$\begin{aligned} X &= (u - c_x) \cdot \frac{Z}{f_x}, \\ Y &= (v - c_y) \cdot \frac{Z}{f_y}, \\ Z &= d. \end{aligned} \tag{3}$$

This transformation is applied to every pixel in the depth map to generate the 3D point cloud.

2) Odometry Estimation: The second sub-module is odometry estimation, which calculates the camera's pose in the local coordinate system at each timestamp by leveraging data from the IMU and the captured images. In the eLabrador, we utilize VINS-MONO [33] as the SLAM backend. VINS-MONO estimates odometry from RGB images and IMU data, providing a robust and versatile monocular visual-inertial state estimation. As the map generated by VINS-MONO is sparse and unsuitable for direct use in obstacle avoidance and path planning, the derived odometry is utilized in the subsequent map-fusion sub-module to integrate semantic point clouds from different locations. Each odometry estimation $\mathbf{F}_{wc}(t)$ at time t represents the transformation from the camera coordinate system to the local coordinate system and is typically expressed as a 4×4 homogeneous transformation matrix:

$$\mathbf{F}_{wc}(t) = \begin{bmatrix} \mathbf{R}_{wc}(t) & \mathbf{T}_{wc}(t) \\ 0 & 1 \end{bmatrix}, \tag{4}$$

where $\mathbf{R}_{wc}(t)$ is a 3×3 rotation matrix and $\mathbf{T}_{wc}(t)$ is a 3×1 translation vector.

3) Map Fusion: The Map Fusion sub-module fuses point clouds from multiple time steps to construct a 3D map enriched with semantic information. Inspired by Semantic SLAM [57], we utilize a voxel map to represent this 3D environment. Each voxel in the map stores three key pieces of information: the occupancy probability, the semantic category, and the classification probability associated with that voxel.

Firstly, the semantic point clouds are transformed from the camera coordinate system to the local coordinate system based on the estimated camera poses from the submodule of odometry estimation. For a 3D point $\mathbf{P}_c(t) = [X_c, Y_c, Z_c, 1]^T$ in the camera coordinate system at time t, its corresponding point $\mathbf{P}_w(t)$ in the local coordinate system is given by:

$$\mathbf{P}_w(t) = \mathbf{F}_{wc}(t) \cdot \mathbf{P}_c(t). \tag{5}$$

By applying this transformation to each point in the point cloud captured at time t, the entire point cloud is transformed from the camera coordinate system to the local coordinate system. Repeating this process for point clouds captured at different timestamps enables the alignment of all point clouds to a common reference frame.

Then, the transformed point clouds are inserted into the semantic map in chronological order. For each point cloud, the points within it are inserted into the map one by one. When inserting a point into its corresponding voxel, we adopt the Max fusion in Semantic SLAM [57] to integrate its semantic information. This procedure produces a 10m×10m octomap, which serves as a detailed and semantically enriched representation of the local environment.

4) GPS-Map Registration: To align the global east-north-up (ENU) coordinate system of the GPS with the local coordinate system of the 3D semantic map, GPS-Map Registration is introduced to estimate the transformation matrix between them. Given the sequence of GPS points $\{\mathbf{p}_i^{enu}(lat, lon)\}$ in global ENU coordinate system and the odometry sequence $\{\mathbf{p}_i^{w}(x, y)\}$ in local coordinate system, we align the $\mathbf{p}^{enu}$ and $\mathbf{p}^{w}$ through their time stamps and get the aligned point pair sequence $\{(\mathbf{p}_i^{enu}, \mathbf{p}_i^{w})\}$. Specifically, the GPS points are transformed into the Cartesian coordinate system as follows:

$$\mathbf{p}_i^{car} = \begin{cases} (0, 0) & i = 0 \\ \mathbf{p}_{i-1}^{car} + (D_{i-1,i} \cos \alpha_i, D_{i-1,i} \sin \alpha_i) & i \geq 1 \end{cases}, \tag{6}$$

where the $D_{i-1,i}$ is the horizontal distance between GPS points $\mathbf{p}_{i-1}^{enu}$ and $\mathbf{p}_i^{enu}$, and α_i is azimuth of line between $\mathbf{p}_{i-1}^{enu}$ and $\mathbf{p}_i^{enu}$ at point $\mathbf{p}_{i-1}^{enu}$. Here, the algorithm for the inverse problem in [58] is adopted to compute the horizontal distance and azimuth. Finally, same as that in [59], the transformation is computed by solving the following equation:

$$(\mathbf{R}, \mathbf{t}) = \underset{\mathbf{R} \in SO(2), \mathbf{t} \in \mathbb{R}^2}{\arg\min} \sum_i \left\| \mathbf{p}_i^{w} - (\mathbf{R}\mathbf{p}_i^{car} + \mathbf{t}) \right\|^2, \tag{7}$$

where $\mathbf{R}$ is a 2×2 rotation matrix, and $\mathbf{t}$ is a 2×1 translation vector, representing the transformation from the GPS Cartesian coordinate system to the local coordinate system of the 3D semantic map respectively. Since the z-axis of map coordinate system has been aligned with the anti-gravity direction, we only need to estimate the transformation of the x- and y-axis here, which means $\mathbf{p}_i^{car} \in \mathbb{R}^2$, $\mathbf{p}_i^{w} \in \mathbb{R}^2$.

B. Global-Local Collaborative Planning

The visual perception module described above generates a 3D semantic local map aligned with the global East-North-Up (ENU) coordinate system. Consequently, the planning results (i.e. GPS waypoints) from Amap can be projected onto this 3D semantic map. However, simply following these waypoints is not feasible for visually impaired users because an Amap route is located on driving roads rather than sidewalks, posing significant safety risks for pedestrians. Moreover, the 3D semantic map contains only semantic and geometric information without explicitly conveying traversability, making it incompatible with typical planning algorithms.

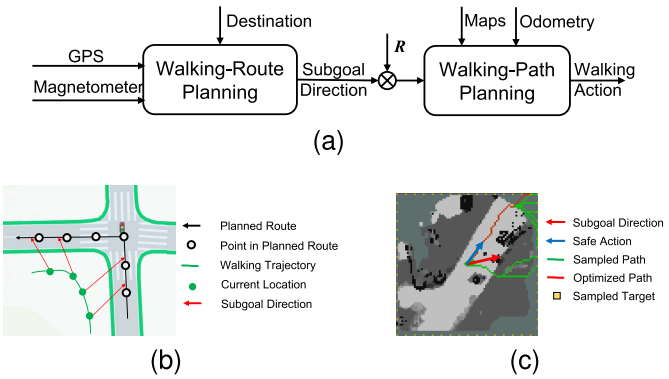


Fig. 5. Illustration of (a) the workflow in planning module, (b) outputs (red lines) of the walking-route planning sub-module, and (c) optimized walking path from current position (red line), obtained by the walking-path planning sub-module.

To address these issues, the planning module is designed to optimize a practical walking path by considering both the global GPS route and local sidewalk conditions (e.g. obstacles, traffic signals, pedestrians), enabling visually impaired individuals to reach their destinations safely. As shown in Fig. 5a, the planning module consists of two sub-modules: walking-route planning and walking-path planning. The walking-route planning sub-module rectifies the waypoints from driving roads to sidewalks by determining a directional subgoal. Given this subgoal, the walking-path planning sub-module then routes a safe walking path by considering local traversability.

1) Walking-Route Planning: The walking-route planning sub-module aims to obtain the walking route on the sidewalk with the Amap and GPS location, as shown in Fig. 5b. Given the user-assigned destination and current (start point) GPS location and an option of sidewalk, the Amap returns a planned route from start to the destination. This route is represented by a series of waypoints, which typically lie on roads rather than sidewalks. Each pair of successive waypoints defines a subroute, with an approximate length of 10 meters. The visually impaired user's current GPS location is then matched to the nearest subroute, and the azimuth between them is identified as the directional subgoal, which begins from the sidewalks.

2) Walking-Path Planning: The directional subgoal derived from the walking-route planning sub-module may not always be feasible due to potential obstructions on the pavement, such as pedestrians or vehicles. To address this issue, the walking-path planning sub-module dynamically generates a safe and walkable path by taking into account the user's immediate surroundings and the intended subgoal direction. This ensures that the user can walk safely and avoid obstacles in real time.

The walking-path planning sub-module comprises three steps: cost map construction (detailed in Appendix A), path searching (detailed in Appendix B), and path following (detailed in Appendix C). These three steps run asynchronously to enable real-time obstacle avoidance and ensure user safety.

First, a cost map represented by a 2D grid map $\mathcal{M}_c$ is constructed from the 3D semantic map, considering both the collision risk and terrain undulation. In $\mathcal{M}_c$, each grid stores a cost value between 0 and 1, where 0 corresponds to a region with no cost (e.g. sidewalk) and 1 indicates a region with high cost (e.g. vehicle). For more details on the construction of the cost map, please refer to Appendix A.

The path searching submodule generates a walkable path based on the directional subgoal and the 2D cost map. Since the subgoal from the walking-route planning is directional, it is not applicable for conventional path planner. Therefore, we developed a Direction A* (Dir-A*) algorithm to compute a walkable path given a directional subgoal. In this algorithm, a set of target points is uniformly sampled along the boundary of the cost map $\mathcal{M}_c$. For each target point, a path is generated using the A* algorithm. The cost of each path is calculated by considering the accumulation of map costs from the start point to the target point, the distance from the start point to the target point, and the angle relative to the subgoal direction. Among all searched paths, the path with the minimum path cost is selected as the optimal path. For more details on the Dir-A* algorithm, please refer to Appendix B.

Although the path from the path search step is theoretically optimized and safe, visually impaired individuals may not follow it exactly. Therefore, the path following (PF) algorithm is designed to determine safe walking actions. Using the searched path and cost map, the PF algorithm first looks ahead by two meters to identify the next waypoint on the planned path, similar to the pure pursuit algorithm [60]. A collision-checking step then determines whether any obstacles lie along the direct path from the current location to next waypoint. If no obstacles are found, the direction from the current location to the endpoint is deemed safe. Otherwise, the foresight distance is reduced to the distance of the first detected obstacle, and the process repeats until a safe behavior is found. For more information, please refer to Appendix C.

C. Audio-Haptic Dual-Channel Interaction

To effectively guide visually impaired individuals, we developed an audio-haptic interaction module that leverages the complementary strengths of both channels to convey navigation instruction and environmental information. Specifically, audio feedback can communicate high-density semantic content via short sentences, albeit with some delay and requiring more cognitive effort. By contrast, haptic feedback offers higher control frequency with minimal delay and lower cognitive load, making it ideal for conveying real-time movement commands necessary to avoid collisions.

As illustrated in Fig. 1, audio feedback is conveyed through bone-conduction headphones, enabling the delivery of rich semantic information such as object names, road signs, and system messages. Meanwhile, haptic feedback is generated by five vibrators mounted uniformly across the front of a belt, offering intuitive and concise cues—e.g., veering left or right. The specific navigation messages conveyed via audio feedback are summarized as follows:

- Route Information: The route information includes the route name, distance, and direction, for example, "walk 49 meters on current road and turn left."
- Action Information: The action information refers to the specific actions users need to perform. For instance, the

system will announce “prepare to turn left” when a turn is approaching. If no safe path can be found or an obstacle is encountered, it will issue a “Stop!” warning.

- System Information: The system information messages convey the status of the system, such as “system initialization completes,” or “the camera module exception.”

The haptic feedback is designed to deliver fine-grained navigation information at a rate of 1 Hz. Five vibrators are mounted on a belt at angles of −90° (Left), −45° (Front-left), 0° (Front), 45° (Front-right), 90° (Right), with 0° facing forward. Their respective feedback modes are summarized below:

- vibrates @ −90°/−45°: When the next action direction lies to the left of the user’s current orientation by more than 30°/10°, the vibrator positioned at −90°/−45° will vibrate continuously, indicating the user should walk to the left.
- vibrates @ 90°/45°: When the next action direction lies to the right of the user’s current orientation by more than 30°/10°, the vibrator positioned at 90°/45° will vibrate continuously, indicating the user should walk to the right.
- vibrates @ 0°: When the difference between the next action direction and the user’s current orientation is at most 10°, the vibrator @ 0° vibrates intermittently, indicating the user should continue walking forward.

During operation, we offer binary (on/off) feedback to guide users effectively. To accommodate varying preferences and clothing thicknesses, two vibration intensity modes–low and high—have been implemented. Users can select the mode that ensures vibrators are both perceptible and comfortable.

V. Experimental Setting

As the eLabrador is designed for outdoor navigation, the evaluation was carried out in typical real-world outdoor environments frequently encountered by visually impaired individuals. The following sections describe the evaluation routes, systems, participants, testing procedures, and the metrics used.

A. Evaluation Routes

We selected two routes commonly traversed in the daily life of visually impaired individuals to test the eLabrador. At each route, the subject starts from a designated location and follows the system’s guidance until stopping near the destination and turning to face it. Fig. 6 shows the layouts of Route 1 and Route 2, which are detailed below.

- Route 1: The first route represents a daily stroll from the residential entrance to a nearby wayside park. For this route, each participant starts from the residential entrance, walks south for about 30 meters, passes a turn, then goes east for 240 meters, and finally arrives at the entrance of the wayside park. At the entrance, the participant turns around according to the system’s guidance and finally faces the landmark of this park. The landmark of this park is a slogan “Honoring the Pioneers of Science” (Fig. 6-C). On this route, the main challenges lie in obstacles including walking persons, trees (Fig. 6-B) as well as passing a turn (Fig. 6-A). Although tactile paving provides directional cues, it is sometimes absent or obstructed, making it unreliable as the sole guidance for passing the turn.
- Route 2: The second route simulates the scenario of coming home after attending a social activity held in the Cultural Activities Center about 300 meters nearby. Though this route is straight, it has several challenges. The biggest challenges of this route are two T-junctions with the crosswalk (Fig. 6-D, E) where the tactile paving is missing so no directional information can be utilized. Meanwhile, the low steps at the end of each sidewalk can potentially trip the visually impaired participants. In addition, pedestrians and vehicles passing through the T-junctions could also temporally block the crosswalk. Besides, the pavilion of the bus stop with a narrow passageway (Fig. 6-F) also complicates this route. The residential entrance, marked by the community name “Kexing Unit” (Fig. 6-G), serves as the final landmark.

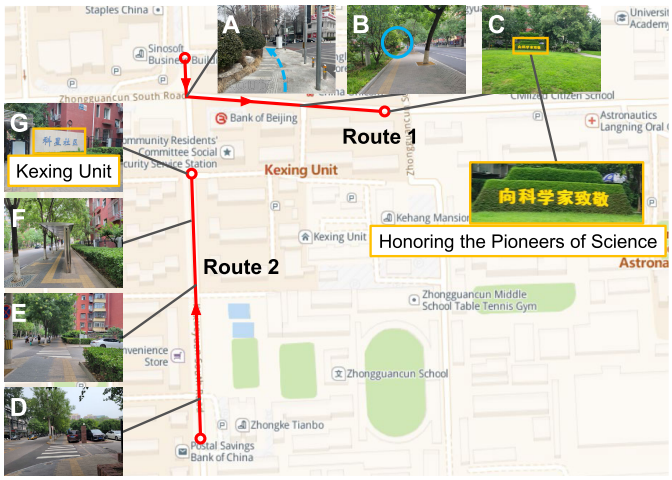


Fig. 6. The layout of Route 1 and Route 2. The map is cropped from Amap. In each route, several representative street-views (A: left turn, B: low-hanging branches, C: the landmark of the destination in Route 1; D: the first crosswalk, E: the second crosswalk; F: the pavilion of the bus stop, G: the landmark of the destination in Route 2) are presented. The landmarks are marked by the orange rectangles. The videos to illustrate Route 1 and Route 2 can be found in supplementary materials.

B. Evaluation Systems

To compare our eLabrador to other tools, three navigation approaches are tested with selected routes:

- White Cane + Amap app: Through communication with visually impaired individuals, we found that using a white cane in combination with Amap’s navigation application is a commonly adopted navigation method. Therefore, this approach is used as a baseline for comparison. In this baseline, participants followed the directional instructions from the Amap app while using any preferred white cane technique, such as touching or constant contact, two-point touch, shorelining, and trailing.
- eLabrador Only: With eLabrador, participants are navigated using real-time guidance, including haptic feedback for walking direction and audio instructions for system information and destination orientation. In this setting,

only the strategies from the system are used; other daily strategies, such as shorelining, are not allowed.

- eLabrador + White Cane: Our system is wearable and hands-free, allowing users to utilize their hands for other tasks, such as using a white cane to detect low steps. In this setting, participants utilize both the white cane and our eLabrador, seamlessly integrating the system's strategies with their daily navigation habits.

C. Subjects

Ten subjects were recruited to test our system, including seven visually impaired participants (V1–V7) and three sighted participants with eye mask (M1–M3). The group consisted of seven males and three females, from 16 to 63 years old. All visually impaired participants have lived with their impairment for many years. In contrast, the sighted participants with their eye mask simulate individuals who have recently become visually impaired and are still adapting to their new circumstances. This setup allows us to evaluate the effectiveness of our system for both experienced and newly impaired users.

D. Testing Procedure

The testing procedure consists of three steps: a tutorial, walking tests, and a questionnaire. At the beginning of each test session, subjects receive a 15-minute tutorial that explains the system's components, outlines IRB and safety protocols, and instructs them on how to respond to the system's feedback. During the test, subjects complete walking tests on both Route 1 and Route 2 sequentially. For each route, participants navigate using three different approaches in the following order: White Cane + Amap, eLabrador, and eLabrador + White Cane. After the walking tests, participants complete a questionnaire on their experience using the system during the walking tests, details about their daily travel experiences, and additional suggestions.

E. Evaluation Metrics

The performance of eLabrador is evaluated in terms of three dimensions: *safety*, *function*, and *friendliness*:

- Safety: Safety refers to the system's ability to detect and prompt the user to avoid obstacles, thereby ensuring the user's safety. This dimension is evaluated by the number of *contacts with obstacles*, such as traffic poles, low steps, and low-hanging branches.
- Functionality: Functionality refers to the system's ability to guide users successfully to their destination (i.e. wayfinding ability), which is the primary objective of the navigation system for the visually impaired individuals. This dimension is assessed using two indicators: *success times* and *correction number*. A trial is considered successful if the subject arrives at the destination and correctly faces the destination landmark. The *success* indicator is recorded as '1' for successful trials and '0' otherwise. The *correction number* counts the number of times the participant had to be intervened by the experimenter to change his / her current action when they encountered obstacles such as wrong way or temporary closed path. Additionally, we report the average walking *speed* for reference.
- Subjective Assessment: Subjective assessment involves the sense of security, cognitive load, and helpfulness while using the system. For this dimension, we adopt the questionnaire to gather participants' subjective feedback on their experiences with the system.

VI. Experimental Results and Analysis

A. Safety Evaluation

As mentioned, the safety of the navigation system is quantified by the number of contacts between participants and obstacles, where a lower number indicates better obstacle avoidance and greater safety. The experimental results are presented in Fig. 7 and Fig. 8. For Route 1 (Fig. 7), participants using only a white cane collided with obstacles such as trees an average of once. When guided by our eLabrador, the average number of collisions decreased to 0.7, achieving a 30% reduction. Furthermore, when using both our eLabrador and a white cane simultaneously, collisions were further reduced to an average of 0.5. These results demonstrate that our system is highly compatible with the white cane, yielding superior safety.

In Route 2, the results for visually impaired participants, as illustrated in Fig. 8, reinforce the same conclusion observed in Route 1: our eLabrador significantly reduces the frequency of contact with obstacles, demonstrating its ability to enhance safety for visually impaired individuals during outdoor navigation. However, the trend differs for participants with eye mask. When guided by a white cane, subjects with eye mask experience fewer collisions with obstacles compared to using our eLabrador. This discrepancy arises from difference in walking behavior between the two groups.

Eye masked participants, aware that obstacles are rarely present at the edges of the route, tend to walk directly along the edge of the sidewalk while using the white cane, thereby minimizing the likelihood of collisions. In contrast, our eLabrador is designed to guide users roughly through the middle of the walkable area rather than along the edge of the sidewalk. While the eLabrador ensures broader coverage of the walkable space, it may increase the chances of encountering obstacles for visually impaired users. These findings highlight the importance of tailoring navigation strategies to the specific needs and behaviors of visually impaired individuals, as well as the potential for further optimizing our system to better align with their natural walking patterns.

Through careful observation of collision with obstacles during the experiment, we find that the white cane can assist the visually impaired participants in avoiding stationary obstacles placed vertically on the ground by using ground contact. However, it has limitations in perceiving and avoiding non-vertical obstacles (e.g. low-hanging branches) and obstacles in motion (e.g. pedestrians). Differently, the eLabrador relies on visual perception and thus can handle situations that the white cane struggles with. The above analysis highlights the distinct strengths of our eLabrador and the white cane in

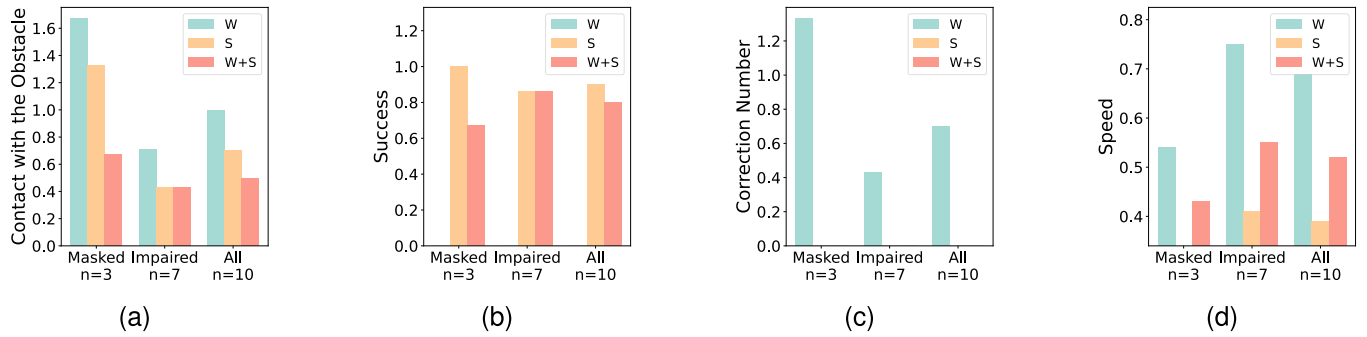


Fig. 7. Quantitative results on Route 1 in terms of "(a) number of contact with the obstacle↓, (b) success times↑, (c) correction number↓, (d) walking speed↑." In total, ten subjects, including visually impaired and eye-masked individuals, are tested. 'W' means 'White Cane', 'S' means our proposed wearable system eLabrador, and 'S+W' means our eLabrador and white cane.

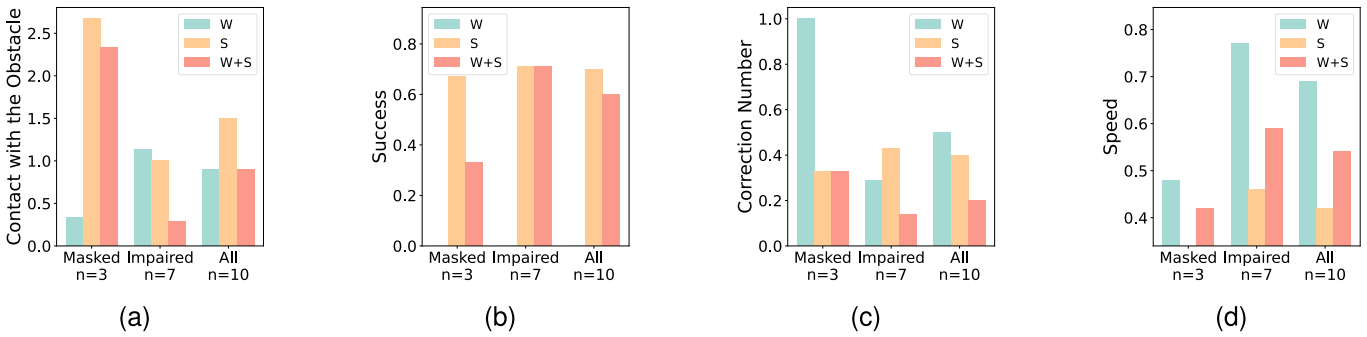


Fig. 8. Quantitative results on Route 2 in terms of "(a) number of contact with the obstacle↓, (b) success times↑, (c) correction number↓, (d) walking speed↑." In total, ten subjects, including visually impaired and eye-masked individuals, are tested. 'W' means 'White Cane', 'S' means our proposed wearable system eLabrador, and 'S+W' means our eLabrador and white cane.

obstacle avoidance scenarios, establishing them as effectively complementary. These results demonstrate that the eLabrador can be either used alone to successfully guide the visually impaired individuals walking outdoors or combined with the white cane for better navigation.

B. Functionality Evaluation

In our study, the functionality of the navigation system is evaluated in terms of the *success times* and *correction number* as shown in Fig. 7 and 8. As seen from Fig. 7, the participants with the white cane achieve no success in reaching the designated landmark. While the white cane can guide visually impaired participants to the general destination, but fails to provide precise guidance toward the specific landmark, resulting in a score of '0' for all participants. In contrast, participants using our system perform significantly better than those relying on the white cane, though occasional failures still occur. These failures primarily stem from GPS location inaccuracies, which arise from multi-path effects due to the tall buildings and numerous metallic objects, such as bike-sharing, along Route 1 and Route 2.

As shown by the result of *correction number* in Fig. 7, the participants with our eLabrador pass Route 1 without any correction, while the participants with white cane need 0.37 corrections on average, most of which occur at the turns along the route. These results confirm that passing the turn is particularly challenging for white cane users, while the eLabrador effectively addresses this issue by leveraging the intelligent visual perception of the environment. At Route 2, the participants with the white cane and our system both need correction as shown in Fig. 8. The corrections occur in the crosswalk, where the absence of directional information makes it difficult to guide visually impaired individuals accurately. Compared to the white cane, our eLabrador requires fewer corrections, demonstrating its superior effectiveness. Furthermore, when the participants use eLabrador and the white cane simultaneously, the correction decreases significantly, exhibiting the compatibility of our system with the white cane.

The eLabrador works effectively mainly because it delivers more informative guidance to visually impaired individuals. However, this brings additional cognitive load, resulting in a reduction in walking speed as shown in Fig. 7 and Fig. 8. Moving forward, we plan to focus on optimizing the system to minimize this cognitive burden.

C. Performance of Visually Impaired and Eye-Masked Participants

For the eye masked participants, eLabrador significantly reduces the number of corrections but causes more collisions on Route 2. This outcome is likely attributed to the distinct behavioral patterns exhibited by eye masked participants, particularly their cautious and deliberate movement.

For the visually impaired participants, eLabrador significantly reduces the number of collisions, however results in a higher number of corrections. Through a detailed analysis of the trial records, we find that this issue arises from

TABLE I
SUBJECTIVE RESULTS FOR ELABRADOR. EACH SUBJECTIVE RATING SCORE IS BETWEEN 0 AND 10, WHERE 0 INDICATES "POOR", 5 INDICATES "MEDIOCRE", AND 10 INDICATES "GOOD"

	V1	V2	V3	V4	V5	V6	V7	avg±std
Security↑	9	6	9	8	4	8	4	6.9±2.19
Cognitive load↓	10	3	6	8	8	9	4	6.9±2.61
Helpfulness↑	8	5	9	8	9	7	5	7.3±1.70
Expectation↑	9	10	10	10	9	7	8	9.0±1.15

an area on the route lacking clear boundaries, where even minor deviations in direction easily trigger manual corrections. Participants accustomed to independent walking were better able to maintain a straight trajectory, which helped them avoid misdirection when relying solely on a white cane. In contrast, the eye masked participants, who are less skilled at walking in a straight line, were more susceptible to misdirection under the same conditions. Additionally, due to limitations in GPS accuracy, eLabrador occasionally provides incorrect directional guidance, further contributing to the need for manual corrections. These findings underscore the complex interplay between user behavior, environmental factors, and technological constraints in shaping navigation outcomes.

D. Subjective Assessment

In addition to the objective evaluation presented in the previous section, we further conducted a subjective assessment on eLabrador through a structured questionnaire. This questionnaire was designed to gather feedback from visually impaired participants regarding their sense of security, perceived cognitive load, the system's helpfulness, and their expectations for its performance. These aspects reflect: (1) how safe users feel when using the system (Security); (2) the level of attention required to follow the system's guidance (Cognitive Load); (3) the system's ability to assist in daily life navigation (Helpfulness); and (4) users' enthusiasm about the system's potential for future development (Expectation). In addition to these metrics, we also collected suggestions from visually impaired participants on how the system could be further improved to better meet their needs in the future.

Each indicator in the questionnaire was rated on a 0-10 scale, with detailed explanations provided for each score to ensure clarity and consistency in responses. For instance, a score of 0 represents a very poor experience (e.g. no sense of security, excessively high cognitive load), while a score of 10 indicates an excellent experience (e.g. a strong sense of security, minimal effort required). Considering that the eye masked participants do not have enough experience living with visual impairments, they are excluded from this part. The results of visually impaired participants are shown in Table I.

As shown in Table I, the average scores for the three perspectives—security, cognitive load, and helpfulness—are all close to 7, indicating that most visually impaired users hold a positive attitude toward our system. For example, V1 mentioned *"As a real-time navigation system, it stays stable and sensitive and also follows my pace without any bug or stuck. This is more than I've expected since I walk rather fast compared to other visually impaired. To me, haptic feedback is vital as it not only tells me which way to go but also confirms me when I'm in the correct direction."*; V3 said *"A problem with the white cane is the areas outside of reach, you can't detect every obstacle on the road, nor tree branches around your head. On the other hand, the system is sensitive enough to warn you about even the tinniest danger so that you can avoid it. That is very helpful to me."*

Several participants consider the current system not helpful enough for their daily life. However, they also believe in the system's potential to help the visually impaired. For example, V2 said *"I think this system could bring great help to those visually impaired individuals who are afraid of going out, but for us who goes out very often and follow a few fixed routes? Maybe not that helpful since we are too familiar with these routes*;" V7 said *"It's all about time. The longer I use this system in the future, the more familiar I would become with it. And familiarity determines how helpful the system is to me. For a morning's training, I could give it a score of 6, but I believe future practice would bring that score higher."*

In summary, all 7 visually impaired users consider our system helpful for their walking and make them feel safer. In the meantime, some users (V2, V3, V7) need to pay their major attention to receiving guidance from the system, which can be further improved. Despite these limitations, the system offers hope for visually impaired users to walk independently in outdoor environment. For example, V3 says that *"I think this system is great. It provides me with much safety with voice feedback about exactly how far the destination is frequently. This feedback, along with feedback about exactly when to turn left or right, helped me form a general idea of where I am. I look forward to walking in the park independently with this system someday."*

Additionally, regarding the weight of the eLabrador, most users believe they can use it for up to one hour. One user estimates a usage time of 40 minutes, while two users feel that the system's weight does not pose a significant burden.

E. Users' Comments and Suggestions

Below are some comments and suggestions from visually impaired participants about how we can improve our system in the future. A general suggestion from all visually impaired participants is that the system could adopt more voice instructions in the future. For example, a more detailed description of what the obstacle is, confirmation of walking in the right direction, and the complexity of the way ahead. All participants confirmed the importance of haptic feedback except V5 and V6, who argued haptic feedback could cause confusion and thus should be deleted. Some other suggestions include recognizing which bus line is coming to the station, reducing the total weight of the system, and adding warnings about dogs' poop or puddles on the road and so on.

Particularly, V4, being enthusiastic about attending all kinds of activities for years and having communicated with countless visually impaired people, gives his valuable opinion: *"Visually impaired individuals, though far fewer in number than sighted individuals, are no less complex. There are those with visual*

impairments, congenital visual impairments, acquired visual impairments, and so on. Some of them learn new things fast while others may not. It would be a huge loss if we abandon either voice or haptic feedback. What you researchers should do is embrace this diversity and provide visually impaired people with various choices and let us decide whatever we like, just like sighted people."

VII. Conclusion

In this work, we have conceptualized and developed a wearable navigation system eLabrador, tailored explicitly for enabling independent long-distance walking in outdoor environments for the visually impaired individuals. Both the hardware and software are cohesively integrated to facilitate environmental perception, global-local collaborative path planning, and audio-haptic dual-channel interactive engagement with visually impaired users. The eLabrador is tested on two real-world routes in the visually impaired individuals' daily life, with each route being about 300 meters. Both objective metrics and subjective feedbacks affirm that eLabrador not only empowers the visually impaired individuals with independent mobility but also instills a heightened sense of security.

While the current system is promising, there exists substantial scope for enhancement, including refining environmental perception accuracy, accelerating user walking speed, incorporating comprehensive voice instructions, and diversifying system functionalities.

Appendix A
Cost Map Construction

Cost map construction is the first step of the walking-path planning sub-module as illustrated in Section IV-B2. The cost map is represented by a 2D grid $\mathcal{M}_c$. In $\mathcal{M}_c$, each grid cell is assigned a value ranging from '0' to '1,' which indicates the cost of traversing that cell. A value of '0' represents a region with no traversing cost (e.g. a sidewalk) while a value of '1' represents a region with high traversing cost (e.g. a vehicle-occupied region). The cost map $\mathcal{M}_c$ is constructed by fusing the collision map $\mathcal{M}_t$ and the undulation map $\mathcal{M}_h$ as follows:

$$\mathcal{M}_c = \mathcal{M}_t(1 - \mathcal{M}_h) + \mathcal{M}_h. \tag{8}$$

Eq. (8) specifies that a grid cell in the cost map $\mathcal{M}_c$ is unwalkable if the corresponding grid is marked as untraversable in collision map $\mathcal{M}_t$, or is blocked by the objects with a large undulation in undulation map $\mathcal{M}_h$.

A. Collision Map $\mathcal{M}_t$

The collision map $\mathcal{M}_t$ is designed as a 2D grid map where the value in each grid reflects its likelihood of collision or traversability. The value of each grid in the collision map $\mathcal{M}_t$ also ranges between 0 and 1. A value of '1' indicates a high probability of collision, rendering the grid non-traversable, and a value of '0' signifies that the grid is fully traversable with no risk of collision.

To construct this collision map $\mathcal{M}_t$, we compress the 3D semantic map into a 2D representation. Let the index of any voxel in the 3D semantic map from the perception module be denoted as (g, h), where g refers to the index of the grid in the 2D grid plane and h means the height index of the voxel. Each voxel in the 3D semantic map is characterized by two attributes: semantic class and occupancy probability. To simplify the computation of the collision probability for each grid in the collision map $\mathcal{M}_t$, we suppose that

1) The collision probability of a grid g in $\mathcal{M}_t$ is the accumulation of the collision probability of all voxels in the 3D semantic map that share the same plane coordinate g but locate at different height h.
2) The collision probability of a voxel is determined by its semantic and geometric properties (e.g. a high tree branch does not affect walking while a low branch may hurt the subject). The contributions of these two dimensions are independent.

TABLE II
The Collision Probability of Different Semantic Categories

category	probability	category	probability
Crosswalk Plain	0.00	Other Rider	1.00
Curb Cut	0.00	Mountain	1.00
Parking	0.00	Banner	1.00
Pedestrian Area	0.00	Bench	1.00
Sidewalk	0.00	Bike Rack	1.00
Crosswalk Lane	0.00	Billboard	1.00
General Lane	0.00	Catch Basin	1.00
Sky	0.00	CCTV Camera	1.00
Car Mount	0.00	Fire Hydrant	1.00
Ego Vehicle	0.00	Junction Box	1.00
Sand	0.10	Mailbox	1.00
Snow	0.10	Manhole	1.00
Bike Lane	0.30	Phone Booth	1.00
Service Lane	0.40	Pothole	1.00
Unlabeled	0.50	Street Light	1.00
Road	0.60	Pole	1.00
Terrain	0.60	Traffic Sign Frame	1.00
Vegetation	0.60	Utility Pole	1.00
Water	0.80	Traffic Light	1.00
Rail Track	0.99	Traffic Sign (Back)	1.00
Bird	1.00	Traffic Sign (Front)	1.00
Ground Animal	1.00	Trash Can	1.00
Curb	1.00	Bicycle	1.00
Fence	1.00	Boat	1.00
Guard Rail	1.00	Bus	1.00
Barrier	1.00	Car	1.00
Wall	1.00	Caravan	1.00
Bridge	1.00	Motorcycle	1.00
Building	1.00	On Rails	1.00
Tunnel	1.00	Other Vehicle	1.00
Person	1.00	Trailer	1.00
Bicyclist	1.00	Truck	1.00
Motorcyclist	1.00	Wheeled Slow	1.00

TABLE III

QUANTITATIVE RESULTS ON ROUTE 1 IN TERMS OF "NUMBER OF CONTACT WITH OBSTACLES, SUCCESS TIMES, CORRECTION NUMBER, WALKING SPEED." IN TOTAL, TEN SUBJECTS, INCLUDING EYE-MASKED AND VISUALLY IMPAIRED INDIVIDUALS, ARE TESTED. 'M' MEANS EYE-MASKED AND 'V' MEANS VISUALLY IMPAIRED. 'W' STANDS FOR 'WHITE CANE', 'S' FOR OUR PROPOSED WEARABLE SYSTEM ELABRADOR, AND 'S+W' FOR BOTH

Subject		M1	M2	M3	Avg.	V1	V2	V3	V4	V5	V6	V7	Avg.	Avg.
Type		Eye-masked				Acquired				Congenital				
Age/Gender		25/F	23/M	22/M		18/F	34/M	56/F	63/M	16/M	44/M	48/M		
Contact with the Obstacle ↓	W	1.00	3.00	1.00	1.67	2.00	1.00	2.00	0.00	0.00	0.00	0.00	0.71	1.00
	S	1.00	3.00	0.00	1.33	0.00	0.00	1.00	0.00	2.00	0.00	0.00	**0.43**	0.70
	S+W	0.00	2.00	0.00	**0.67**	0.00	0.00	2.00	1.00	0.00	0.00	0.00	**0.43**	**0.50**
Success Times↑	W	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	S	1.00	1.00	1.00	**1.00**	1.00	0.00	1.00	1.00	1.00	1.00	1.00	**0.86**	**0.90**
	S+W	1.00	0.00	1.00	0.67	1.00	0.00	1.00	1.00	1.00	1.00	1.00	**0.86**	0.80
Correction Number ↓	W	3.00	0.00	1.00	1.33	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.43	0.70
	S	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	**0.00**
	S+W	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	**0.00**
Speed ↑	W	0.31	0.86	0.45	**0.54**	0.74	0.86	0.58	0.43	0.64	1.05	0.94	**0.75**	**0.69**
	S	0.39	0.43	0.21	0.34	0.47	0.43	0.18	0.29	0.27	0.62	0.60	0.41	0.39
	S+W	0.40	0.42	0.47	0.43	0.59	0.58	0.39	0.40	0.59	0.69	0.62	0.55	0.52

TABLE IV

QUANTITATIVE RESULTS ON ROUTE 2 IN TERMS OF "NUMBER OF CONTACT WITH OBSTACLES, SUCCESS TIMES, CORRECTION NUMBER, WALKING SPEED." IN TOTAL, TEN SUBJECTS, INCLUDING EYE-MASKED AND VISUALLY IMPAIRED INDIVIDUALS, ARE TESTED. 'M' MEANS EYE-MASKED AND 'V' MEANS VISUALLY IMPAIRED. 'W' STANDS FOR 'WHITE CANE', 'S' FOR OUR PROPOSED WEARABLE SYSTEM ELABRADOR, AND 'S+W' FOR BOTH

Subject		M1	M2	M3	Avg.	V1	V2	V3	V4	V5	V6	V7	Avg.	Avg.
Type		Eye-masked				Acquired				Congenital				
Age/Gender		25/F	23/M	22/M		18/F	34/M	56/F	63/M	16/M	44/M	48/M		
Contact with the Obstacle ↓	W	0.00	0.00	1.00	**0.33**	1.00	2.00	3.00	1.00	1.00	0.00	0.00	1.14	**0.90**
	S	6.00	0.00	2.00	2.67	2.00	0.00	2.00	2.00	0.00	1.00	0.00	1.00	1.50
	S+W	2.00	0.00	5.00	2.33	0.00	2.00	0.00	0.00	0.00	0.00	0.00	**0.29**	**0.90**
Success Times↑	W	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	S	1.00	0.00	1.00	**0.67**	0.00	1.00	1.00	1.00	0.00	1.00	1.00	**0.71**	**0.70**
	S+W	1.00	0.00	0.00	0.33	0.00	0.00	1.00	1.00	1.00	1.00	1.00	**0.71**	0.60
Correction Number ↓	W	2.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.29	0.50
	S	1.00	0.00	0.00	**0.33**	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.43	0.40
	S+W	1.00	0.00	0.00	**0.33**	0.00	0.00	1.00	0.00	0.00	0.00	0.00	**0.14**	**0.20**
Speed ↑	W	0.35	0.60	0.50	**0.48**	0.71	0.88	0.91	0.40	0.69	1.08	0.74	**0.77**	**0.69**
	S	0.32	0.40	0.31	0.34	0.53	0.31	0.34	0.28	0.49	0.82	0.43	0.46	0.42
	S+W	0.38	0.60	0.28	0.42	0.46	0.40	0.62	0.39	1.02	0.78	0.44	0.59	0.54

3) The collision probability of the semantics is position-free, which means it is not related to the specific position of the voxel, but only related to semantics property itself.

Based on the above assumptions, the collision probability of a grid g is computed as

$$p(g) = \frac{1}{\sum_h p(g,h)} \sum_h p(s)p(m)p(g,h), \tag{9}$$

where $p(g)$ is the collision probability of the grid at g; $p(g,h)$ is the occupancy probability of the voxel at (g,h); $p(s)$ and $p(m)$ represent the collision probability of semantic s and geometric m respectively; s and m are the semantic and geometric properties of the voxel at (g,h). Moreover, we ignore the mathematical proof of Eq. (9) for brevity and provide an intuitive explanation to understand it: Eq. (9) is the discrete form of the expectation of $p(s)p(m)$ whose probability distribution is $p(g,h)$. The $p(s)p(m)$ means that a voxel has a high collision-probability only if the semantic and geometric properties of this voxel are both not traversable (high $p(s)$ and $p(m)$), otherwise has a low collision-probability.

To compute Eq. (9), the occupancy probability $p(g,h)$ is directly queried from the 3D semantic map. For the collision probability of semantic s, each semantic class is assigned a value ranging from '0' to '1' according to our daily experience

Algorithm 1 Dir-A* for Path Searching With Goal Direction

function DIR-A*(start point: p_{start}, goal direction: q_{goal}, cost map: $\mathcal{M}_c$)
  uniformly sample n_s target points $\mathcal{T}$ on the boundary of the cost map $\mathcal{M}_c$
  $P \leftarrow \varnothing$
  $C \leftarrow \varnothing$
  for each target point p_{target} in $\mathcal{T}$ **do**
    (path, cost) $\leftarrow$ ASTAR(p_{start},p_{target},$\mathcal{M}_c$)
    cost $\leftarrow$ cost + α path_length + $\beta(\cos\theta + 1)$
    $P \leftarrow P \cup \{\text{path}\}$
    $C \leftarrow C \cup \{\text{cost}\}$
  end for
  idx = $\arg\min_i C[i]$
  return P[idx]
end function
function ASTAR(start point: p_{start}, target point:p_{target}, cost map: $\mathcal{M}_c$)
  classical A* algorithm with cost map $\mathcal{M}_c$.
  return (path, cost) when the feasible path is found. Otherwise, **return** ($\varnothing$, ∞). ▷ ∞: infinite path cost.
end function

Algorithm 2 Path Following Algorithm

function PF(current location: $p_{current}$, searched path: path, cost map: $\mathcal{M}_c$, foresight distance: d_{fs})
  $\text{idx}_{near} \leftarrow \arg\min_i \|p_{current} - \text{path}[i]\|$
  $d \leftarrow d_{fs}$
  while True **do**
    $p_{next} \leftarrow \text{path}[\text{idx}_{near} + d/\text{map\_resolution}]$
    (is_collision, d_{obs})
    $\leftarrow$ CHECKCOLLISION($p_{current}$,p_{next})
    if not is_collision **then**
      break
    end if
    $d \leftarrow d_{obs}$
  end while
  return $p_{next} - p_{current}$
end function
function CHECKCOLLISION(p_{start},p_{end})
  check whether the obstacle between the straight path of p_{start} and p_{end} exists.
  return (True, d_{obs}) when detecting the obstacle in the straight path of p_{start} and p_{end}. Otherwise, **return** (False, -1). ▷ -1: collision-free identifier.
end function

as shown in Table II. Moreover, the geometric property is defined as the height difference between the voxel and the camera (i.e. $m = h_{voxel} - h_{camera}$). The collision probability of m is set as '1' if its value satisfies -2 meters $\leq m \leq 0$ meter, otherwise m is set as '0'.

B. Undulation Map $\mathcal{M}_h$

The undulation map $\mathcal{M}_h$ is also designed as a 2D grid map but with binary value. The grids that are taller in height than their neighboring grids (e.g. the step and curb) are marked as '1,' and all others (e.g. flat road) as '0.' To create the undulation map, we first project the 3D semantic map onto the horizontal plane, transforming it into a 2D grid map. Each cell in this 2D map represents the height of the highest occupied voxel within the corresponding vertical pillar, as below:

$$\mathcal{H}(g) = \arg\max_h \{h; \mathcal{M}_{3D}(g, h) = 1\}, \tag{10}$$

where $\mathcal{M}_{3D}(g, h) = 1$ means the voxel at (g, h) is occupied. Next, Sobel filter is applied to determine the degree to which a grid's height value exceeds that of its neighboring grids, similar to edge detection. Finally, grids with edge degrees beyond a certain threshold are marked as '1,' while the rest are marked as '0,' resulting in the undulation map $\mathcal{M}_h$.

APPENDIX B
THE PATH SEARCHING ALGORITHM: DIR-A*

Path searching is the second step of the walking-path planning sub-module as illustrated in Section IV-B2. In this appendix, we describe the algorithm for path searching.

In unfamiliar outdoor environments, only a coarse route can be obtained from global maps such as Amap or Google Maps. However, these routes typically lie on driving roads rather than sidewalks, making them unsafe for pedestrian walking. To address this, we transform the coarse global route into a series of directional subgoals and design the Direction A* algorithm (Dir-A*) to do path searching based on these subgoals.

As described in Algorithm 1, the Dir-A* first samples a set of the target points along the boundary of the cost map $\mathcal{M}_c$. For each target point, the proposed algorithm searches an optimal path on the cost map $\mathcal{M}_c$ by using the classical A* algorithm. After identifying potential paths, the cost of each path is evaluated based on three criteria: the map cost, the path length, and the angle relative to the subgoal direction, calculated as follows:

$$\text{cost}_{\text{path}} = f_c(p_{target}) + \alpha d(p_{target}) + \beta(\cos\theta + 1), \tag{11}$$

where p_{target} is the sampled target point; $f_c(p_{target})$ means the accumulation of the map cost from the start point to the target point; $d(p_{target})$ represents the distance from the start point to the target point; θ represents the angle between the goal direction and the direction from the center of the map to sampled point; α and β are the hyperparameters to balance the magnitude of above items. Finally, the planned path is selected as the path with minimum cost.

APPENDIX C
THE PATH FOLLOWING ALGORITHM: PF

Path following is the last step of the path walking-planning sub-module as illustrated in Section IV-B2. The Path Following (PF) algorithm is designed to acquire actionable walking direction based on the planned path, the cost map, and the user's current location. Its primary goal is to guide the user walking along the planned path.

Fig. 9. Illustration of light conditions at different locations along the two routes. The light conditions are relatively diverse, enabling a comprehensive assessment of the effects of different systems.

Fig. 10. Illustration of light conditions when different subjects walked in the same position while using different systems. The light conditions will vary somewhat as the experiment was conducted on different days for different subjects.

As shown in Algorithm 2, the Path Following (PF) algorithm begins by identifying the nearest point p_{near} to the current location $p_{current}$ on the optimized path. Starting from p_{near}, the algorithm looks ahead along the path by a foresight distance of 2 meters to determine the next target point p_{next}. A collision-checking mechanism is then employed to detect obstacles along the straight-line path between p_{near} and p_{next}. An obstacle is defined as any grid with a cost value greater than 0.9. If no collision is detected, the walking direction is set as the vector from $p_{current}$ to p_{next}. If a collision is found, the foresight distance is reduced to d_{obs}, which represents the distance from p_{near} to the nearest obstacle along the straight-line path.

APPENDIX D
EXPERIMENTAL DETAILS

The detailed experimental results on Route 1 and Route 2 are presented in Table III and Table IV, corresponding to Fig. 7 and Fig. 8 in the main body. The experimental setup, including road conditions and testing procedures, has been described in the main text. Additionally, factors such as the time of day and lighting conditions during the experiments are also important, especially for the visual perception module. Most experiments were conducted between 8 a.m. and 11 a.m. over a span of 23 days. Each subject completed all routes on a single day, ensuring consistent lighting conditions for the same subject. However, lighting conditions varied across different subjects due to changes in weather. To illustrate these variations, we have included representative images from several subjects in Fig. 9 and Fig. 10.

REFERENCES

[1] R. Bourne et al., "Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the global burden of disease study," *Lancet Global Health*, vol. 9, no. 2, pp. e130–e143, Dec. 2021.

[2] C. Y. Wong, R. A. Ananto, T. Akiyama, J. P. Nemargut, and A. J. Moon, "Perspectives on robotic systems for the visually impaired," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Boulder, CO, USA, Mar. 2024, pp. 1134–1138.

[3] H. T. V. Vu, J. E. Keeffe, C. A. McCarty, and H. R. Taylor, "Impact of unilateral and bilateral vision loss on quality of life," *Brit. J. Ophthalmol.*, vol. 89, no. 3, pp. 360–363, 2005.

[4] D. D. Clark-Carter, A. D. Heyes, and C. I. Howarth, "The efficiency and walking speed of visually impaired people," *Ergonomics*, vol. 29, no. 6, pp. 779–789, Jun. 1986.

[5] S. Resnikoff et al., "Global data on visual impairment in the year 2002," *Bulletin World Health Org.*, vol. 82, no. 11, pp. 844–851, 2004.

[6] G. Lacey and S. MacNamara, "Context-aware shared control of a robot mobility aid for the elderly blind," *Int. J. Robot. Res.*, vol. 19, no. 11, pp. 1054–1065, Nov. 2000.

[7] N. Kawarazaki and T. Yoshidome, "Remote control system of home electrical appliances using speech recognition," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Seoul, South Korea, Aug. 2012, pp. 761–764.

[8] X. Qian and C. Ye, "NCC-RANSAC: A fast plane extraction method for navigating a smart cane for the visually impaired," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Madison, WI, USA, Aug. 2013, pp. 261–267.

[9] H. He, Y. Li, Y. Guan, and J. Tan, "Wearable ego-motion tracking for blind navigation in indoor environments," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1181–1190, Oct. 2015.

[10] C. Ye, S. Hong, X. Qian, and W. Wu, "Co-robotic cane: A new robotic navigation aid for the visually impaired," *IEEE Syst., Man, Cybern. Syst. Mag.*, vol. 2, no. 2, pp. 33–42, Apr. 2016.

[11] Y. H. Lee and G. Medioni, "RGB-D camera based wearable navigation system for the visually impaired," *Comput. Vis. Image Understand.*, vol. 149, pp. 3–20, Aug. 2016.

[12] Q. Chen et al., "CCNY smart cane," in *Proc. IEEE 7th Annu. Int. Conf. CYBER Technol. Autom., Control, Intell. Syst. (CYBER)*, Honolulu, HI, USA, Jul. 2017, pp. 1246–1251.

[13] H. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarré, and D. Rus, "Enabling independent navigation for visually impaired people through a wearable vision-based feedback system," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Singapore, May 2017, pp. 6533–6540.

[14] K. Tobita, K. Sagayama, M. Mori, and A. Tabuchi, "Structure and examination of the guidance robot LIGHBOT for visually impaired and elderly people," *J. Robot. Mechatronics*, vol. 30, no. 1, pp. 86–92, Feb. 2018.

[15] J. Guerreiro, D. Sato, S. Asakawa, H. Dong, K. M. Kitani, and C. Asakawa, "CaBot: Designing and evaluating an autonomous navigation robot for blind people," in *Proc. 21st Int. ACM SIGACCESS Conf. Comput. Accessibility*, New York, NY, USA, Oct. 2019, pp. 68–82.

[16] H. Zhang and C. Ye, "Human–robot interaction for assisted wayfinding of a robotic navigation aid for the blind," in *Proc. 12th Int. Conf. Human Syst. Interact. (HSI)*, Richmond, VA, USA, Jun. 2019, pp. 137–142.

[17] S. Kayukawa, T. Ishihara, H. Takagi, S. Morishima, and C. Asakawa, "BlindPilot: A robotic local navigation system that leads blind people to a landmark object," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, Honolulu, HI, USA, Apr. 2020, pp. 1–9.

[18] S. Kayukawa, T. Ishihara, H. Takagi, S. Morishima, and C. Asakawa, "Guiding blind pedestrians in public spaces by understanding walking behavior of nearby pedestrians," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–22, 2020.

[19] L. Wang, J. Zhao, and L. Zhang, "NavDog: Robotic navigation guide dog via model predictive control and Human–Robot modeling," in *Proc. 36th Annu. ACM Symp. Appl. Comput.*, Mar. 2021, pp. 815–818.

[20] M. Kuribayashi et al., "Corridor-Walker: Mobile indoor walking assistance for blind people to avoid obstacles and recognize intersections," in *Proc. ACM Hum.-Comput. Interact. (PACMHCI)*, vol. 6, Sep. 2022, pp. 1–22.

[21] S. Agrawal, M. E. West, and B. Hayes, "A novel perceptive robotic cane with haptic navigation for enabling vision-independent participation in the social dynamics of seat choice," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 9156–9163.

[22] M. Kuribayashi et al., "PathFinder: Designing a map-less navigation system for blind people in unfamiliar buildings," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Hamburg, Germany, Apr. 2023, pp. 1–16.

[23] V. Ranganeni et al., "Exploring levels of control for a navigation assistant for blind travelers," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, Stockholm, Sweden, Mar. 2023, pp. 4–12.

[24] S. Kammoun, C. Jouffrais, T. Guerreiro, H. Nicolau, and J. Jorge, "Guiding blind people with haptic feedback," in *Proc. Pervasive Workshop Frontiers Accessibility Pervasive Comput.*, vol. 3, New Castle, U.K., Jun. 2012, pp. 1–3.

[25] D. Ni, L. Wang, Y. Ding, J. Zhang, A. Song, and J. Wu, "The design and implementation of a walking assistant system with vibrotactile indication and voice prompt for the visually impaired," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Shenzhen, China, Mar. 2013, pp. 2721–2726.

[26] R. Chen, Z. Tian, H. Liu, F. Zhao, S. Zhang, and H. Liu, "Construction of a voice driven life assistant system for visually impaired people," in *Proc. Int. Conf. Artif. Intell. Big Data (ICAIBD)*, Chengdu, China, May 2018, pp. 87–92.

[27] H. Liu, D. Guo, X. Zhang, W. Zhu, B. Fang, and F. Sun, "Toward image-to-tactile cross-modal perception for visually impaired people," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 521–529, Apr. 2021.

[28] P.-J. Duh, Y.-C. Sung, L. F. Chiang, Y.-J. Chang, and K.-W. Chen, "V-eye: A vision-based navigation system for the visually impaired," *IEEE Trans. Multimedia*, vol. 23, pp. 1567–1580, 2021.

[29] P. Slade, A. Tambe, and M. J. Kochenderfer, "Multimodal sensing and intuitive steering assistance improve navigation and mobility for people with impaired vision," *Sci. Robot.*, vol. 6, no. 59, Oct. 2021, Art. no. eabg6594.

[30] Y. Ma, Y. Shi, M. Zhang, W. Li, C. Ma, and Y. Guo, "Design and implementation of an intelligent assistive cane for visually impaired people based on an edge-cloud collaboration scheme," *Electronics*, vol. 11, no. 14, p. 2266, Jul. 2022.

[31] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 1290–1299.

[32] B. Cheng, A. Schwing, and A. Kirillov, "Per-pixel classification is not all you need for semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 34, 2021, pp. 17864–17875.

[33] T. Qin, P. Li, and S. Shen, "VINS-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.

[34] N. Mahmud, R. Saha, R. Zafar, M. Bhuian, and S. Sarwar, "Vibration and voice operated navigation system for visually impaired person," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, Dhaka, Bangladesh, May 2014, pp. 1–5.

[35] S. Gallo et al., "Augmented white cane with multimodal haptic feedback," in *Proc. 3rd IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics*, Tokyo, Japan, Sep. 2010, pp. 149–155.

[36] R. Pyun, Y. Kim, P. Wespe, R. Gassert, and S. Schneller, "Advanced augmented white cane with obstacle height and distance feedback," in *Proc. IEEE 13th Int. Conf. Rehabil. Robot. (ICORR)*, Seattle, WA, USA, Jun. 2013, pp. 1–6.

[37] I. Ulrich and J. Borenstein, "The guidecane-applying mobile robot technologies to assist the visually impaired," *IEEE Trans. Syst., Man, Cybern. A, Syst. Hum.*, vol. 31, no. 2, pp. 131–136, Mar. 2001.

[38] S. Shoval, I. Ulrich, and J. Borenstein, "NavBelt and the guide-cane [obstacle-avoidance systems for the blind and visually impaired]," *IEEE Robot. Autom. Mag.*, vol. 10, no. 1, pp. 9–20, Mar. 2003.

[39] P. Aigner and B. McCarragher, "Shared control framework applied to a robotic aid for the blind," *IEEE Control Syst. Mag.*, vol. 19, no. 2, pp. 40–46, Apr. 1999.

[40] Y. Zhang et al., "I am the follower, also the boss: Exploring different levels of autonomy and machine forms of guiding robots for the visually impaired," in *Proc. ACM Conf. Hum. Factors Comput. Syst. (CHI)*, 2023, pp. 1–22.

[41] A. Xiao, W. Tong, L. Yang, J. Zeng, Z. Li, and K. Sreenath, "Robotic guide dog: Leading a human with leash-guided hybrid physical interaction," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Xi'an, China, May 2021, pp. 11470–11476.

[42] Y. Chen et al., "Quadruped guidance robot for the visually impaired: A comfort-based approach," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, London, United Kingdom, May 2023, pp. 12078–12084.

[43] P. Balatti et al., "Robot-assisted navigation for visually impaired through adaptive impedance and path planning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 2310–2316.

[44] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and technologies for blind and visually impaired navigation support: A review," *IETE Tech. Rev.*, vol. 39, no. 1, pp. 3–18, Jan. 2022.

[45] A. R. See, B. G. Sasing, and W. D. Advincula, "A smartphone-based mobility assistant using depth imaging for visually impaired and blind," *Appl. Sci.*, vol. 12, no. 6, p. 2802, Mar. 2022.

[46] D. Croce et al., "An indoor and outdoor navigation system for visually impaired people," *IEEE Access*, vol. 7, pp. 170406–170418, 2019.

[47] M. Murata, D. Ahmetovic, D. Sato, H. Takagi, K. M. Kitani, and C. Asakawa, "Smartphone-based localization for blind navigation in building-scale indoor environments," *Pervas. Mobile Comput.*, vol. 57, pp. 14–32, Jul. 2019.

[48] P. Du and N. Bulusu, "An automated AR-based annotation tool for indoor navigation for visually impaired people," in *Proc. Int. ACM SIGACCESS Conf. Comput. Accessibility (ASSETS)*, New York, NY, USA, Oct. 2021, pp. 1–4, doi: 10.1145/3441852.3476561. Accessed: Aug. 16, 2022.

[49] M. Kuribayashi, S. Kayukawa, H. Takagi, C. Asakawa, and S. Morishima, "LineChaser: A smartphone-based navigation system for blind people to stand in lines," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Yokohama, Japan, May 2021, pp. 1–13.

[50] S. Blessenohl, C. Morrison, A. Criminisi, and J. Shotton, "Improving indoor mobility of the visually impaired with depth-based spatial sound," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Santiago, Chile, Dec. 2015, pp. 418–426.

[51] S. Al-Khalifa and M. Al-Razgan, "Ebsar: Indoor guidance for the visually impaired," *Comput. Electr. Eng.*, vol. 54, pp. 26–39, May 2016.

[52] R. K. Katzschmann, B. Araki, and D. Rus, "Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 583–593, Mar. 2018.

[53] X. Liu, B. Wang, and Z. Li, "Vision-based wearable steering assistance for people with impaired vision in jogging," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 15270–15275.

[54] W. W. Mayol-Cuevas, B. J. Tordoff, and D. W. Murray, "On the choice and placement of wearable vision sensors," *IEEE Trans. Syst., Man, Cybern. A, Syst. Hum.*, vol. 39, no. 2, pp. 414–425, Mar. 2009.

[55] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognit. Lett.*, vol. 137, pp. 37–52, Sep. 2020.

[56] G. Neuhold, T. Ollmann, S. R. Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4990–4999.

[57] Z. Xuan and F. David. (2018). *Real-time Voxel Based 3D Semantic Mapping With a Hand Held Rgb-d Camera*. [Online]. Available: https://github.com/floatlazer/semantic_slam

[58] C. F. F. Karney, "Algorithms for geodesics," *J. Geodesy*, vol. 87, no. 1, pp. 43–55, 2013.

[59] O. Sorkine-Hornung and M. Rabinovich, "Least-squares rigid motion using svd," *Computing*, vol. 1, no. 1, pp. 1–5, 2017.

[60] R. C. Coulter, "Implementation of the pure pursuit path tracking algorithm," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-92-01, Jan. 1992.

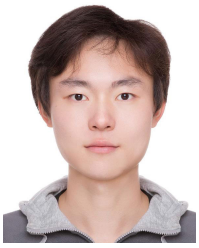
Meina Kan (Member, IEEE) received the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China. She is currently a Professor with the Institute of Computing Technology, Chinese Academy of Sciences (CAS). Her research interests include machine vision especially embodied AI, transfer learning, and deep learning.

Lixuan Zhang (Student Member, IEEE) received the B.Eng. degree in automation from Xi'an Jiaotong University, Xi'an, China, in 2021. He is currently pursuing the Ph.D. degree with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China. His research interests include visual navigation, decision-making, and planning.

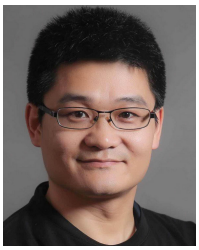
Hao Liang received the B.S. degree in computer science and technology from Peking University, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing. His research interests include object detection and active vision.

Boyuan Zhang received the bachelor's degree in computer science and technology from the University of Chinese Academy of Sciences in 2021. He is currently pursuing the master's degree with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS). His research interests include 3D vision, SLAM, and structure from motion.

Minxue Fang received the bachelor's degree in computer science and technology from the University of Chinese Academy of Sciences in 2021. He is currently pursuing the master's degree with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS). His research interests include human–computer interaction.



Dongyang Liu received the B.E. degree from Tongji University, Shanghai, China, in 2021. He is currently pursuing the master's degree with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China. His research interests include knowledge distillation and model compression.



Shiguang Shan (Fellow, IEEE) is currently a Professor with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), and the Director of the Key Laboratory of Intelligent Information Processing, CAS. His research interests include image processing, computer vision, pattern recognition, and machine learning. He has published more than 350 articles in related areas. He served as the General Co-Chair for the IEEE Face and Gesture Recognition 2023 and Asian Conference on Computer Vision (ACCV) 2022 and the Area Chair for tens of international conferences, including CVPR, ICCV, ECCV, NeurIPS, ICML, AAAI, IJCAI, ACCV, ICPR, FG, and WACV. He was/is an Associate Editor of several journals, including IEEE TRANSACTIONS ON IMAGE PROCESSING, *Neurocomputing*, *Computer Vision and Image Understanding*, *Transactions on Machine Learning Research*, and *Pattern Recognition Letters*.

Xilin Chen (Fellow, IEEE) is currently a Professor with the Institute of Computing Technology, Chinese Academy of Sciences (CAS). He has authored one book and more than 400 articles in refereed journals and proceedings in the areas of computer vision, pattern recognition, image processing, and multimodal interfaces. He is a fellow of ACM, IAPR, and CCF. He is also an Information Sciences Editorial Board Member of Fundamental Research, an Editorial Board Member of Research, a Senior Editor of the *Journal of Visual Communication and Image Representation*, and the Associate Editor-in-Chief of *Chinese Journal of Computers* and *Pattern Recognition and Artificial Intelligence*. He served as an Organizing Committee Member for multiple conferences, including the General Co-Chair for FG 2013/FG 2018 and VCIP 2022, the Program Co-Chair for ICMI 2010/FG 2024, and the Area Chair for ICCV/CVPR/ECCV/NeurIPS for more than ten times.