Snap&Nav: Smartphone-based Indoor Navigation System For Blind People via Floor Map Analysis and Intersection Detection

MASAYA KUBOTA*, Waseda University, Japan
MASAKI KURIBAYASHI*, Waseda University, Japan
SEITA KAYUKAWA, IBM Research - Tokyo, Japan
HIRONOBU TAKAGI, IBM Research - Tokyo, Japan
CHIEKO ASAKAWA, Miraikan - The National Museum of Emerging Science and Innovation, Japan and IBM Research, United States
SHIGEO MORISHIMA, Waseda Research Institute for Science and Engineering, Japan

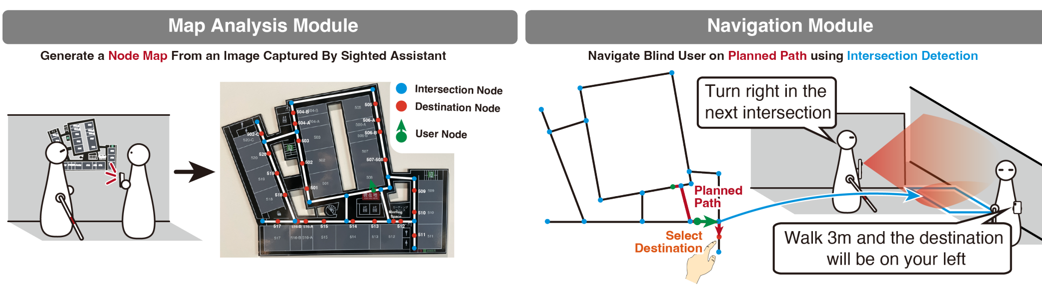

Fig. 1. Our system first requires a sighted assistant to capture an image of a floor map that is commonly available at buildings. The system extracts a node map from the image by applying a map analysis algorithm. Then, the system plans a path to the selected destination by a blind user and navigates them to the destination by using an intersection detection algorithm.

We present Snap&Nav, a navigation system for blind people in unfamiliar buildings, without prebuilt digital maps. Instead, the system utilizes the floor map as its primary information source for route guidance. The system requires a sighted assistant to capture an image of the floor map, which is analyzed to create a node map containing intersections, destinations, and current positions on the floor. The system provides turn-by-turn navigation instructions while tracking users' positions on the node map by detecting intersections. Additionally, the system estimates the scale difference of the node map to provide distance information. Our system was validated through two user studies with 20 sighted and 12 blind participants. Results showed that sighted participants processed floor map images without being accustomed to the system, while blind

*Both authors contributed equally to this research.

Authors' addresses: Masaya Kubota, Waseda University, Tokyo, Japan; Masaki Kuribayashi, Waseda University, Tokyo, Japan; Seita Kayukawa, IBM Research - Tokyo, Tokyo, Japan; Hironobu Takagi, IBM Research - Tokyo, Tokyo, Japan; Chieko Asakawa, Miraikan - The National Museum of Emerging Science and Innovation, Tokyo, Japan and IBM Research, Yorktown Heights, New York, United States; Shigeo Morishima, Waseda Research Institute for Science and Engineering, Tokyo, Japan.

Author Version

participants navigated with increased confidence and lower cognitive load compared to the condition using only cane, appreciating the system's potential for use in various buildings.

CCS Concepts: • **Human-centered computing → Accessibility systems and tools**; • **Social and professional topics → People with disabilities**.

Additional Key Words and Phrases: visual impairment; orientation and mobility; intersection detection; map-less navigation

ACM Reference Format:
Masaya Kubota, Masaki Kuribayashi, Seita Kayukawa, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. 2024. Snap&Nav: Smartphone-based Indoor Navigation System For Blind People via Floor Map Analysis and Intersection Detection. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 275 (September 2024), 22 pages. https://doi.org/10.1145/3676522

1 INTRODUCTION

A major challenge in daily life for blind people is to navigate to their destinations in unfamiliar buildings. As they need a long-term familiarization of buildings for independent navigation with canes and guide dogs, they usually need to rely on sighted assistance by asking to accompany them to their destinations or asking them for the route [10, 36]. The reliance particularly becomes repetitive when they need to travel to multiple destinations, significantly impeding their independence. Nonetheless, many prefer not to rely on sighted people but rather to travel independently by themselves [10, 36]. Thus, to assist blind people in navigating to their destinations independently, many navigation systems have been proposed in past years. The majority of these systems use prebuilt digital maps, *i.e.*, maps that contain locations and descriptions about points of interest (POIs), and localization methods to provide users with turn-by-turn navigation instructions [18, 29, 32, 46]. While these systems are promising solutions for navigating blind people in unfamiliar buildings, these systems are often unavailable in public, as prebuilt digital maps are costly to deploy [28, 46]. In response to this issue, navigation systems that do not require prebuilt digital maps have been proposed in recent years [27, 28, 30, 31, 43]. These systems assist users by leveraging real-time sensing results from various sensors to convey information about the environment and have users decide their decisions at each intersection. As these systems don't possess any pre-existing knowledge of the route users need to walk, users or systems must rely on external route information to be used. Previously, this route information was sourced from sighted people who provided descriptions of routes [27, 28, 43]. However, the descriptions of routes provided by sighted passersby can be an inaccurate source of information [39]. Also, it can be cumbersome for blind users to constantly ask for routes when they need to go to a new destination using these navigation systems.

In this research, we propose Snap&Nav, a navigation system that utilizes an image of a floor map in a building, which contains a walkable path along with names of possible destinations, as an information source of the environment, thereby eliminating the need for prebuilt digital maps (Fig. 1). Firstly, to use floor maps for navigation, it is necessary to obtain the image of the floor maps for the system. As blind people may have difficulty in capturing an image of the target object by themselves [20, 25], we designed the system such that sighted people capture floor maps instead of blind users. Also, from the floor map image, it is necessary to extract information such as intersections, possible destinations, connection relationships, and the user's initial position and orientation for guidance to the destination. While intersections, destinations, and their connections can be extracted by utilizing image processing or computer vision algorithms on a floor map image [51], the users' position and orientation are not always apparent in all floor maps. Thus, we designed the system so that sighted assistants annotate the blind user's initial position and orientation. After a sighted assistant captures an image of a floor map, the system analyzes the image and creates a node map, which is a map involving the aforementioned information represented by nodes and connections. Secondly, by using the extracted node map, the system navigates blind people who hold the smartphone in their hands. To use the node map for navigation, it is necessary to continuously localize the user's position on the node map to provide turn-by-turn navigation instructions. In our approach, users scan surroundings with the smartphone, and the system detects an intersection in the real world and localizes the user's position by comparing its shape to that of the node map.

We implemented Snap&Nav by prototyping two functionalities, namely the map analysis and navigation, and conducted two user studies to assess the validity of the design of Snap&Nav. The first user study was conducted with 20 sighted participants to evaluate the usability wherein a sighted person captures an image of a floor map and annotates the position and orientation of a blind user. The study revealed that most participants were able to use the system without being accustomed to the system, but also revealed improvements such as the need for specification of how to verify whether the generated map is correct. Then, the main study was conducted with 12 blind participants. We prepared two conditions: a *system-aided* condition where they navigated using the proposed system, and a *cane-only* condition where they navigated with a description of routes by sighted people. Participants were asked to navigate to multiple destinations in sequence. Throughout the study, we revealed that usage of our system enabled participants to navigate with increased confidence and reduced cognitive load, without affecting the task completion time. The participants generally appreciated the fact that they did not have to ask for route descriptions multiple times when using the system, which allowed them to gain more independence by relying less on sighted people. Additionally, ten of the blind participants expressed that they find it acceptable to involve sighted assistants, given the potential benefits they would receive.

The code will be publicly available on the following link https://github.com/chestnutforestlabo/Snap-and-Nav.

2 RELATED WORK

2.1 Assistance Systems Using Prebuilt Digital Maps

To provide blind people with an advanced understanding of the environment and routes, systems have been proposed that leverage tactile maps [38, 45, 50] or virtual environments [17, 23]. In particular, one common solution for navigating blind people to a destination in buildings was to use a prebuilt digital map of the building with infrastructures for localization. For example, such systems would utilize localization methods that would rely on Bluetooth Low Energy (BLE) beacons [4, 16, 26, 37, 46], visual features [29, 55], magnetic signatures [15] or radio frequency identifier (RFID) tags [41]. However, they are not available in every building, as they usually require additional prebuilt digital maps and infrastructure for localization, which needs cumbersome preparation and maintenance. In contrast, our proposal is to analyze the layout of the building from floor maps commonly found at entrances, offering a more cost-effective solution.

2.2 Navigation Systems without Maps

Systems that aim to navigate blind users to their destinations without needing prebuilt digital maps have also been proposed. Since these systems do not have prebuilt digital maps, they primarily depend on real-time sensing and route information sourced externally, such as prior route knowledge from blind people [30, 31] or route descriptions by sighted people [27, 28, 43]. In particular, route descriptions have been utilized to navigate blind users in unfamiliar environments, such as ones described by experimenter [27, 43] or sighted people [28]. However, the usage of routes described by the sighted people leads to three major limitations. Firstly, route descriptions provided by sighted people may be inaccurate, which may lead blind users to reach the wrong destinations. Secondly, blind users need to ask for descriptions of routes multiple times when they have to travel to multiple destinations, leading to less independence for blind people. Finally, users need to remember described routes, which can potentially cause navigational errors, particularly in lengthy routes, due to mistakes in memorization. Thus, we aim to utilize an image of floor maps instead of route descriptions from sighted people as an external knowledge of the system without prebuilt digital maps. This approach could potentially overcome three limitations by providing more accurate information than routes described by sighted passersby, navigating to multiple destinations by referring to the analyzed map, and eliminating the need for users to memorize route descriptions.

2.3 System Using Indoor Floor Map Analysis

Researchers have proposed various methods for creating navigation routes with edges and nodes from floorplans of buildings [6, 14, 33, 40, 47, 54]. While floorplans accurately represent a floor's structure, these are not often available for public usage, which impedes assistance systems from using them for navigation purposes. Prior research proposed analysis systems for floor maps (*i.e.*, ones found on information boards in shopping malls or at entrances of buildings) to extract walkable areas [22] or localize the user's position in shopping malls [51]. Following previous research, we prototype a method to analyze floor maps, but for navigation purposes for blind people. To provide turn-by-turn navigation instructions, the system extracts information such as intersections, destinations, and their connections and generates a node map.

3 SYSTEM DESIGN

The proposed system, Snap&Nav, has a map analysis module and a navigation module (Fig. 1). The map analysis module is aimed to be used by a sighted assistant, by having a blind user ask sighted people to use this module. Then, based on the analyzed map, blind users could select the destination and navigate using the navigation module. The system is designed to acquire the route to destinations from a floor map, which is commonly available in buildings. Thus, the advantage of this design is that it has the potential to be used in various buildings that have floor maps, without any preparation. To realize the design described above, we implemented the system on the iPhone 12 Pro, which is a smartphone equipped with a LiDAR sensor.

3.1 Map Analysis Module

To provide blind users with turn-by-turn navigation instructions, this module in the system creates a node map consisting of information of intersections, destinations, and the position and orientation of a blind user. Users of this module are sighted assistants (Fig. 1). As they are expected to be asked to use the system on the first view, the interface of the system must be used without any prior tutorial. Thus, we design the map analysis module to provide voice instructions to sighted assistants. Firstly, the system instructs sighted assistants to capture the floor map image. Then, sighted assistants are asked to annotate blind users' position and orientation in the captured image by interacting with the system, as it is not always apparent in all floor maps. Then, the system processes the captured image to extract a node map consisting of the positions of intersections, destinations, and the blind user's position with their connections. Finally, the system asks sighted assistants to determine whether the connections in the node map match the information in the floor map. They can compare the node map displayed on the smartphone screen with the floor map. If they determine that the generated node map is not sufficient, they capture and annotate it again.

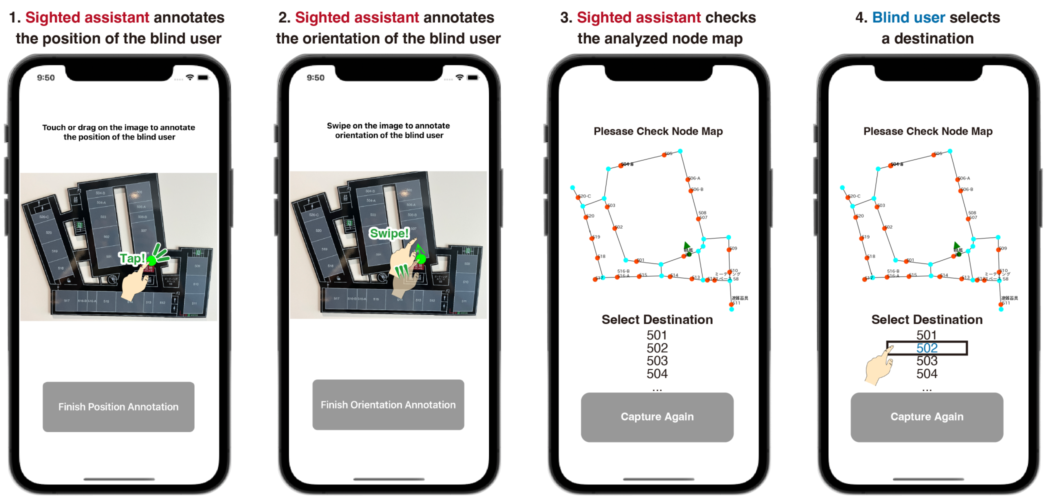

Fig. 2. The figure showing how sighted assistants and blind users use the system.

3.2 Navigation Module

Firstly, blind users can select the desired destination from the list of extracted destination nodes. Once the destination is selected, the system plans a path from the current user's node (Fig. 1). To provide turn-by-turn instructions, the system tracks the user's position on the node map (*i.e.*, which nodes or connections users are in) by using the intersection detection algorithm to verify the shape of their current intersection and match it with the intersection nodes on the node map. To guide users to their destination, it's essential for the system to convey the distance they need to proceed from the final intersection. Therefore, the system calculates the scale difference between the node map and the real world by comparing the distance between two nodes in pixel space to the distance between two real-world intersections in meters. Once the scale difference has been obtained, it can offer instructions accompanied by the distance covered after that intersection.

4 IMPLEMENTATION: MAP ANALYSIS MODULE

4.1 Interface for Sighted Assistants

Fig. 2 shows the interface of Snap&Nav. When sighted assistants press the "capture floor map" button on the initial screen, the system activates the camera, and with voice feedback, the system instructs assistants to capture an image of a floor map so that it is placed in the center with minimum lighting (Fig. 2-1). Then, assistants are instructed to annotate the position of blind users. Assistants can either tap or drag on the image to specify the position of blind users, at which point a green dot appears to indicate their location (Fig. 2-1). When the annotation is completed, assistants can tap a button placed on the bottom of the screen to complete the annotation process of the position. Then, the system provides voice feedback to instruct assistants in setting the orientation of the blind user at any angle by using a swipe gesture on the image. When assistants swipe, a green arrow pointing from the green dot will be displayed in the swiped direction (Fig. 2-2). Finally, assistants can tap a button on the bottom to complete the whole annotation process.

The system sends the image to a remote server to apply a map analysis algorithm (Sec. 4.2). After the analysis, the system receives and displays the image of the analyzed node map from the server.

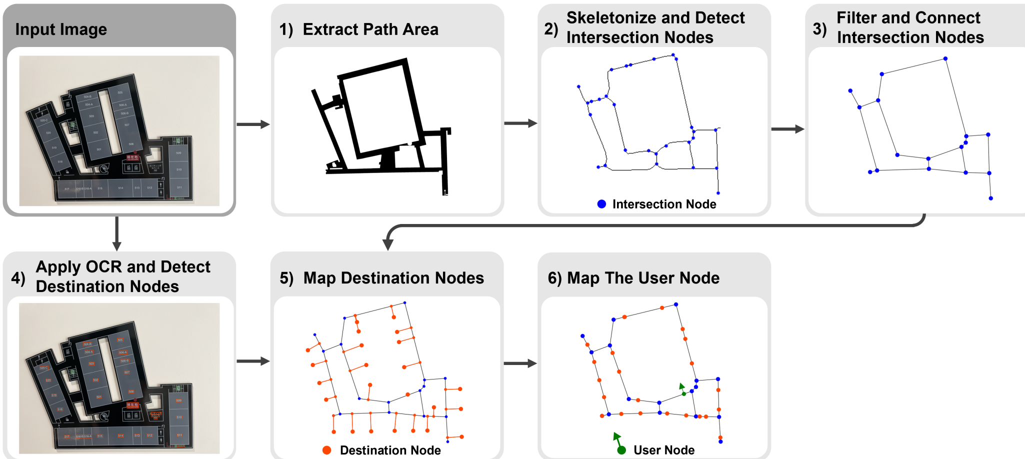

Fig. 3. The figure showing floor map analysis algorithm

Assistants can verify if the image is accurately analyzed (Fig. 2-3). If not, assistants can repeat the process by pressing the capture image button.

4.2 Floor Map Analysis Algorithm

We prototyped a map analysis algorithm. Our algorithm creates a node map, which is a representation of a floor where each node corresponds to an intersection or a destination, and the connection between each node represents walkable pathways. The node map consists of three types of nodes: (1) a user node that represents the initial position and orientation of the blind user, (2) an intersection node, and (3) a destination node. Specifically, the node map will be obtained by applying the algorithm described below to an RGB image (resolution of 4032 × 3024) obtained by a smartphone camera on a remote processing server with RTX-3060 GPU with 8GB memory capacity.

The red circular and rectangular icon in Fig. 3 represents the user's location. To prevent the icon from affecting the following image processing algorithm, we first mask out the icon, which indicates the user's position. To do so, we assume that the red color indicates the user's location in this study and mask out colors close to red in the image. Then, the image is binarized, and the connected component algorithm is applied to identify connected regions. The largest area, which is expected to represent the path or corridor, is extracted (Fig. 3-1). Then, the skeletonization algorithm [56] is applied to the extracted path area. After that, Harris corner detection is applied to the skeletonized image to identify potential intersection nodes (Fig. 3-2). The connections between these nodes are ascertained based on the connections presented in the skeletonized image. To eliminate extra intersection nodes that were accidentally detected, we filter out nodes with only two connections where the connection angle exceeds 140°, and get the intersection-only node map (Fig. 3-3). Next, to find destination nodes in the floor map, optical character recognition (OCR) [24] is applied to the original RGB image. As a result, multiple bounding boxes with destinations' names are obtained (Fig. 3-4). The center point of each bounding box is extracted to represent the location of each destination, which is then mapped to the nearest connection between nodes in the node map. At the same time, the system determines on which side of the path the destinations are located (Fig. 3-5). Finally, we map the user node, whose location and direction were annotated by sighted assistants, to the closest connection (Fig. 3-6).

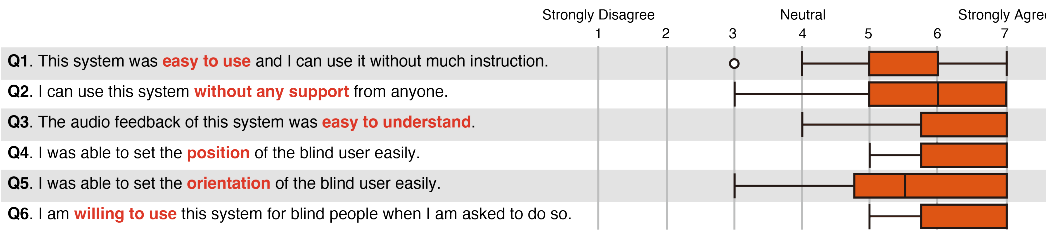

Fig. 4. Questions and responses from our study with 20 sighted participants. Responses are rated on a seven-point Likert scale.

5 USER STUDY FOR MAP ANALYSIS MODULE WITH SIGHTED PARTICIPANTS

We conducted a user study to evaluate the map analysis module with 20 sighted participants (16 male and four female), aged 22 to 31 years old (mean=23.8 and standard deviation(SD)=2.3). Eighteen participants were familiar with the experimental location, and two were unfamiliar. One aim of this study was to investigate if sighted assistants could use the system without being accustomed to it. This user study was approved by an Institutional Review Board (IRB), and informed consent was obtained from every participant before the study.

5.1 Tasks and Procedure

We took participants in front of five different floor maps, which are already available in the building. Then, we asked them to use the system, assuming they were asked to do so by a blind person. The experimenter, who acted as a blind person, stood within three meters of a floor map and faced towards it. Before handing the system, we gave concise instructions to capture the floor map image and annotate it by following the system's voice instructions. They were allowed to recapture and repeat the annotation until they felt confident with the generated node map. When participants finished the task, we asked them to return the system to the experimenter.

The floor map participants first capture using the system is the most important, as the map analysis module is designed for scenarios where sighted assistants are asked to use it by blind people. To ensure that each floor map is captured an equal number of times, we randomized the order of capturing each floor map, with each floor map being captured first by different participants precisely four times. Finally, we interviewed participants with questions on Fig. 4. The whole study was recorded and took 30 minutes in total. Participants were compensated with 7$. Below, we refer to *first trial* as the first task of floor map capturing and *overall trial* as the all task of floor map capturing.

5.2 Metrics

5.2.1 APLS. To evaluate the performance of the map analysis algorithm, we used Average Path Length Similarity (APLS) [11], which is a standard metric for evaluating node maps such as ones generated from satellite images. The metric assesses the similarity between two node maps by comparing differences in their path lengths. This process involves identifying corresponding nodes between the predicted node map and the GT node map. The algorithm calculates the shortest path distances between all pairs of corresponding nodes using the Dijkstra algorithm [9], and records these path lengths. Subsequently, it computes the ratio of the length differences between these paths. If a corresponding node is absent in the predicted node map, resulting in a missing path, a maximum penalty of 1.0 is assigned. The final step calculates the sum of the differences between the predicted node map and the GT node map. This sum is then averaged across all nodes and subtracted

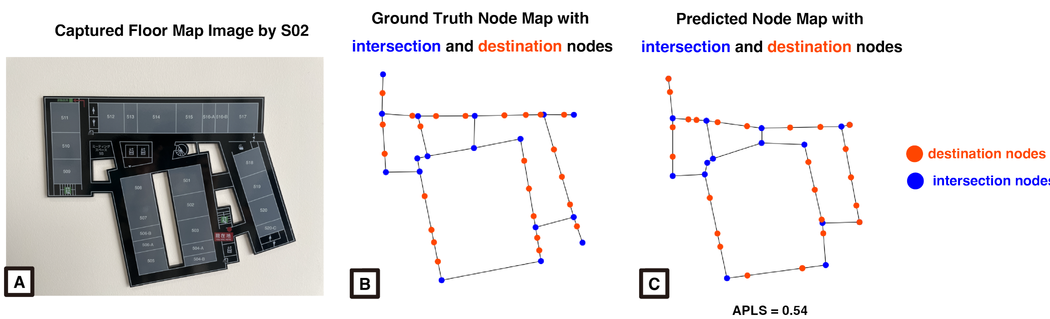

Fig. 5. Figure showing (A) a floor map image captured by S02, (B) its ground truth node map, and (C) generated node maps that contain both intersection and destination nodes.

from 1 to derive the APLS score. The APLS score ranges from 0 (indicates poor similarity) to 1 (indicates high similarity), and is defined as follows, $APLS = 1 - \frac{1}{N}\sum \min\left\{1, \frac{|L(a,b)-L(a',b')|}{L(a,b)}\right\}$ where $L(a, b)$ represents the path length between nodes a and b as computed by the Dijkstra algorithm in the node map. This metric is influenced by the nodes' topological connectivity and geographical positioning, as it is highly penalized when there is an absent connection and primarily measures the differences in path length.

5.2.2 Task Completion Time (TCT). We measured the task completion time, which is the time to capture an image of a floor map and edit the user's position and orientation. We also define the system process time (SPT), which is the time it takes to send the captured image to the server and generate the node map, and confirmation time (CT), which is the time it takes for sighted participants to verify if the received node map is correct.

5.2.3 User Node Accuracy. We measured how accurately participants annotated the position and orientation of a blind user. We considered the position of a point to be correctly annotated if it was within 227 pixels of the ground truth (GT) location in the original image coordinate space. The value of 227 pixels was determined by imitating the minimum button radius of 22 points on the screen coordinate space for iOS devices [8]. As the screen width is 390 points for the iPhone 12 Pro and the image width is 4032 pixels, we calculated the value with the following equation, $4032 \times \frac{22}{390} \simeq 227$. We also defined the orientation as the correct orientation if the orientation is within 45 degrees from the GT orientation in the captured map. In a building where the system may be used, such as our experimental environment, floor maps may be installed on either side of the corridor wall. Thus, the orientation of a user standing in front of a floor map can be one of the two possible orientations. The system classifies the input user orientation into two categories based on which wall of the path the user node is orientated. If the error between the input and the annotation is within 90 degrees, the system determines the correct orientation. In this study, 45 degrees was used as a strict condition. We defined the GT position as the position of the floor map and the GT orientation as the side of the wall on which the floor map exists. The experimenter manually annotated the GT position and the orientation for the evaluation.

5.2.4 Subjective Ratings. We asked seven-point Likert-scale questions as shown in Fig. 4. The questions were designed based on the system usability scale (SUS) [3] questionnaire. To fit within the time constraints of the study, we selected relevant questions from the SUS questionnaire, minimizing the total number of questions.

Table 1. Evaluation of the map analysis algorithm and participants' performance: Average Path Length Similarity APLS [11], task completion time (TCT), user node accuracy, and average number of recaptures per trial. For TCT, we report for overall time, system process time (SPT), and confirmation time (CT). We evaluated each metric for the first Trial and the overall trial.

	APLS [11] Intersection & Destination	Task Completion Time (Mean±SD [Seconds])			User Node Accuracy		Number of Recaptures per trial
		Overall	SPT	CT	Location	Direction	
First Trial	0.57	88.62±35.41	6.54±4.20	20.27±14.93	0.95	0.85	0.35±1.11
Overall Trials	0.56	62.92±28.40	6.42±3.06	14.91±12.68	0.99	0.95	0.21±0.60

5.3 Result

5.3.1 APLS. Tab. 1 reports the average APLS of the generated node maps. The values did not differ between their first and overall trials, indicating how sighted participants captured images did not differ before and after getting accustomed to the system. In Fig. 5, we provide an example of a captured image and its corresponding generated node map with its APLS. The average APLS value was 0.57 in the First Trial and 0.56 in the Overall Trial. The APLS values, which are close to 0.5, can be attributed to the misdetections of some nodes and deviations in node mappings, as illustrated in Fig. 3-5. For instance, some nodes at the ends of corridors were missing because our algorithm mainly extracts intersection nodes with corner detection, and thus, corridor ends were not detected. Furthermore, while the algorithm successfully captured the overall structure of the node map, slight deviations in node placement from their actual positions were noted. These deviations also led to a decrease in APLS values.

5.3.2 Task Completion Time (TCT). Tab. 1 shows the mean and SD of TCT. The mean value of the overall trial decreased compared to that of the overall TCT and CT of their first trial. Tab. 1 also shows the results of system process time (SPT) and confirmation time (CT) involved in overall TCT. The maximum CT was 63.94 seconds by S06.

5.3.3 User Node Accuracy. Tab. 1 shows the ratio of this metric. Generally, all participants were able to set the user's position correctly. On the other hand, three participants mistakenly set the orientation as they thought they were asked to set the orientation the blind person would be heading.

5.3.4 Number of Recaptures. Tab. 1 reports the mean and SD of the number of recaptures per trial. Seventeen participants finished the task without recapturing the floor map in their first trial. In overall trials, ten participants recaptured floor map images. Seven recaptures occurred in the first trials, and 21 occurred in the overall trials. Out of the ten participants, S18 showed the most confusion in their first trial. Firstly, S18 captured a floor map from a distance because S18 thought it was important to remove the light reflection. It caused the floor map in the image to be small. Thus, the system was not able to generate the appropriate map. As S18 thought that the cause of the failure was light reflections, S18 repeatedly captured the floor map in the same manner.

5.3.5 Subjective Ratings. Fig. 4 shows the results of seven-point Likert-scale questions. For all questions (Q1–Q6), more than 17 participants gave positive scores (greater than 5). While all of them felt that they were able to set the position of the blind user (Q4), S03 and S19 felt that they were unable to set the orientation easily (they scored 3 points in Q5).

5.3.6 Qualitative Feedback. Aligned with the result in Q6, all participants stated that they are willing to use the system when asked by blind people, as it offers reliable assistance than explaining

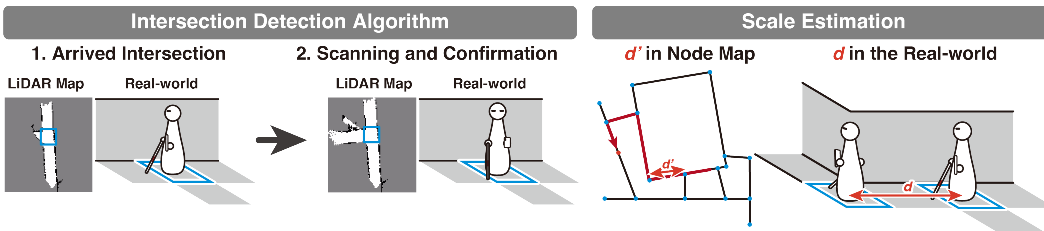

Fig. 6. Figure showing how the system detects intersections and how the system estimates the scale difference of the node map with the real world. The displayed 2D occupancy grid map on the left panel is the actual grid map captured by the system.

routes or guiding them: **C1**: "*It isn't easy to know if the route description is conveyed correctly. If the destination is far away, it takes time to convey the information, and I am also concerned about whether the information I provided is accurate. On the other hand, the system offers easy assistance just by capturing floor map images. And it doesn't take much time.*" (S06) Also, S20 commented, assuming the scenario guiding a blind person. **C2**: "*I found it relatively easier than walking with them and guiding them to their destination. For example, when a blind person's destination is the exact opposite of the direction I want to go, or when the distance is very long, or when there is not much time available, I feel that this kind of application can reduce the burden on the person providing guidance.*" (S20)

We also received negative feedback. Three out of twenty participants made errors in annotating the orientations of the blind user in their first trial. In this regard, six participants pointed out the ambiguity of the explanation for the orientation. **C3**: "*As for the orientation, I wasn't sure if I should input the direction the blind person was facing or the direction of the pathway.*" (S14) Eight participants also commented about the difficulty of checking the node maps generated by the system. **C4**: "*In the case of a complex map, it would take a lot of time and effort to check if it is accurate.* " (S01) and, **C5**: "*The definition of whether the node map is good or not was not clearly stated. ... I thought it would be easier to check if the items to be checked were clearly indicated, for example, whether the room nodes are properly taken and whether the intersections are in place.*" (S10)

6 IMPLEMTATION: NAVIGATION MODULE

6.1 Destination Selection and Path Planning

Firstly, the system lets a blind user select a destination on the node map from a list of destinations extracted in the floor map analysis (Fig. 2-4) by using VoiceOver, the built-in screen reader on iOS. Then, the system employs Dykstra's algorithm [9] to plan a path from the current position, which is initially set to the user node, to the selected destination.

6.2 Tracking Users Position Using Intersection Detection and Node Map

6.2.1 Intersection Detection and Confirmation. We use the intersection detection and intersection confirmation algorithm, which utilizes a method used in the previous work [30]. The system generates a 2D occupancy grid map (*i.e.*, LiDAR Map) of the surrounding environment using the smartphone's LiDAR sensor and employs the YOLOv7 object detection model [49] to identify intersections, where the position of the bounding box indicates the location of the intersection and the label specifies its shape. Nine distinct intersection shapes, composed of combinations of the words Left, Right, Front, and Back, can be recognized (*e.g.*, Intersection on Fig. 6–1 indicate "Left, Front, Back").

When an intersection has an uneven structure like an alcove, the system may mislabel it. Thus, the system checks the LiDAR map to confirm the shape of intersections by verifying whether each side bounding box contains a sufficient amount of walkable area [30]. If the criteria for a specific direction are met, the system confirms that the intersection leads in that direction. Users are instructed to scan specific directions in intersections to ensure the necessary features appear in the LiDAR map (Fig. 6–2).

6.2.2 Tracking Users Position. The system navigates the blind user to the destination by tracking their position on the node map. Every time the user reaches a detected intersection, the system compares the shape of the detected intersection with the shape of the next intersection in the planned path. The system calculates the direction of the paths, *i.e.*, the angle relative to the heading direction, in the node map and that of the detected intersection and compares them. If the error of these angles is within 40 degrees, the system determines that the intersections are matched, and the system updates their position on the node map and announces the next instruction.

6.3 Scale Estimation of Node Map

The system calculates the scale difference between the node map and the real world to convey users the distance they have to proceed once they have reached the first intersection. Every time the system passes an intersection, it calculates the distance d between the previous intersection in real-world scale and calculates the difference scale by $Scale = \frac{d}{d'}$, where d' denotes the distance between two intersections in the node map in pixel space coordinates. The system could calculate the distance to walk by multiplying the calculated scale by the length of sides in the node map.

6.4 Voice Feedback while Navigation

We designed our voice feedback that provides instructions on which direction to turn and the distance to proceed. While there are various types of voice feedback, including clock position instructions for tasks of lining up in a queue [29], simplified instruction to turn right or turn left for navigation[12, 13, 46, 55], and slight turn instructions (between 30 and 60 degrees) [19], we refer to the work by Kuribayashi *et al.* [28, 30], as they also convey intersection information in their task.

First, the system instructs users which way to proceed, along with the direction to proceed to the next intersection (*e.g.*, *"Face right, proceed forward, and turn left in the next intersection"*). Note that the initial direction for blind users to face (*e.g.*, *"Face right"*) is computed from the annotated orientation of a sighted assistant. When users have arrived at an intersection, the system instructs users to scan the intersection (*e.g.*, *"You have arrived at an intersection. Scan left and right for confirmation."*). If the scanned shape of the intersection is the same as the one users have to be at, the system instructs users to turn (*e.g.*, *"You are at an intersection to turn. Turn right"*). When the users have turned, the system provides users with the distance to proceed, along with the next direction turn. (*e.g.*, *"Proceed for 11 meters and face left.""*) Note at this point, as the system has already calculated the scale difference between the node map and real world, it could convey how much distance users should proceed. Finally, when users have arrived at the destination, the system notifies them of the way the destination is. (*e.g.*, *"Face left. You have arrived at the destination."*) The system could provide instructions in real time as the whole process in the navigation module operates ten times per second.

7 USER STUDY FOR NAVIGATION MODULE WITH BLIND PARTICIPANTS

To evaluate the navigation module of the Snap&Nav, we performed a user study in our university building with 12 blind participants. Specifically, the user study was conducted to compare the system and their navigation scenario where blind people use a white cane and walk based on the

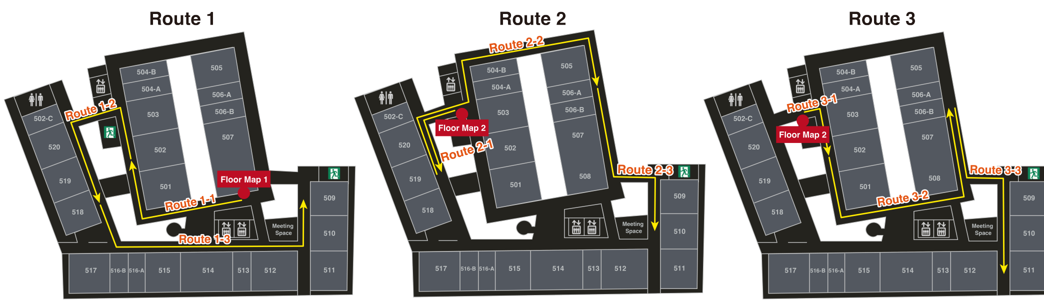

Fig. 7. Routes used in the user study with blind people.

route described by sighted people. Tab. 2 shows the demographic information of the participants. All participants are totally blind and use a white cane in their daily lives. While three participants had visited the experimental environment, they had no experience navigating the routes used in this user study. This user study was approved by the university's IRB, and informed consent was obtained from every participant.

7.1 Tasks and Conditions

The task of the study was to navigate several routes, each with three predefined destinations. Specifically, we prepared three routes, Route 1, Route 2, and Route 3, whose lengths were 122 m, 116 m, and 106 m, respectively (Fig. 7). To mimic the scenario of moving to multiple destinations in an unfamiliar building, each route had three sub-routes, for example, Route 1 consists of Route 1-1, Route 1-2, and Route 1-3. For each route, they were asked to navigate sub-routes one by one and speak out to the experimenter when they reached the destination.

For comparison, we prepared two conditions, *system-aided* condition, and *cane-only* condition. Participants walked each route under two conditions, completing a total of six walks. In the system-aided condition, participants held their cane in their right hand and the system in their left hand. The experimenter, who acted as a sighted assistant, handed the smartphone equipped with the system in front of the floor map, assuming the situation of seeking an assistant to the sighted person and capturing the floor map has already been completed. Then, participants walked three sub-routes within a single route independently using the system. This assumption was explained to the participants prior to the task. To focus on the evaluation navigation aspects, the system used two node maps with a high value of APLS obtained in the user study with sighted participants. The APLS value of the node map used in Route 1 was 0.63, and the value of the node map used in Route 2 and 3 was 0.59. In the cane-only condition, participants only had their cane. The experimenter provided a description of each sub-route at each starting point of sub-route. (*i.e.*, descriptions were given three times per route) When participants believed they had arrived at their destination, they verbally indicated their arrival to the experimenter. The descriptions of the route consisted of an accurate number of turns and distances they had to walk. They were allowed to ask the experimenter for the route. In such a case, the experimenter would explain the route from their current position to destinations.

7.2 Procedure

We first explained the purpose of the study and conducted 20 minutes interview asking about their demographic information and daily experience when navigating unfamiliar buildings. We then introduced the proposed system and participants practiced using the system in a test area for 30 minutes to get familiar with the system. For the training session, participants navigated through five pre-determined routes using the system. Some participants navigated an additional route if they needed to familiarize themselves more with the system. Then, participants were asked to conduct the main task. In order to counterbalance potential order effects, participants systematically rotated through the experimental conditions. For the first half of the participants, the progression began with Route 1 with the system-aided condition, and subsequently alternated conditions with each successive route (*e.g.*, B01 walked Route 1 with system-aided, then Route 2 with cane-only, Route 3 with system-aided, Route 1 with cane-only, Route 2 with system-aided, Route 3 with cane-only). The second participant in this group started with Route 2 with the cane-only condition, maintaining the alternating route and condition order. The translation of the route and condition was maintained until the sixth participant. For the latter half, this order was reversed (*e.g.*, Route 3 with cane-only, Route 2 with system-aided, Route 1 with cane-only, and so on). Finally, after the main task, we conducted a post-interview, asking them questions regarding the usability of the system with both open-ended questions and questions using Likert scores [3]. The whole study was recorded, and the study took approximately 135 minutes. Participants were compensated with 25$ per hour.

Table 2. Participants' demographic information and corresponding values for system usability scale (SUS) score.

ID	Age	Gender	Total Blindness	Smartphone Usage	SUS
B01	58	Male	18 years	8 years	92.5
B02	37	FeMale	23 years	9 years	95
B03	31	Male	21 years	10 years	90
B04	50	Female	13 years	13 years	92.5
B05	28	Male	14 years	9 years	95
B06	55	Female	51 years	12 years	92.5
B07	48	Male	8 years	7 years	77.5
B08	50	Female	5 years	5 years	87.5
B09	38	Female	37 years	6 years	75
B10	48	Female	45 years	10 years	100
B11	50	Female	15 years	10 years	70
B12	57	Male	3 years	20 years	72.5

7.3 Metrics

7.3.1 Task Completion Time (TCT). We measured task completion time, which is the time to complete routes. Task completion time was recorded both for the time they travel the whole route and the time they travel from one destination to another. A timer was started when participants started navigating and was stopped when they stated their arrival at the destination. We also measured the time participants scanned at each intersection, by observing the recorded video.

7.3.2 Distance to Destination Area. Our system is designed to offer turn-by-turn navigation instructions to reach a destination, which in this study's context, is a specific area. It is not our goal to provide last-few-meters guidance, such as guiding users to the exact entrance of the room [13, 44]. Thus, to evaluate the accuracy of our navigation system, we measured the distance between the point where participants stated their arrival and the destination area where the rooms are located. If participants stopped within the width of the rooms, we considered this metric as zero. However, if they walked past the destination area, we measured the distance from the end of the room to where they stopped.

7.3.3 Subjective Rating. We conducted subjective ratings to quantitatively assess the usability of the system (Fig. 9). As illustrated in Fig. 9, we evaluated confidence and cognitive load for each functionality by comparing system-aided and cane-only conditions. Additionally, we assessed ease of understanding, usefulness, and appropriateness for the system-aided conditions. To design the questionnaire, we referred to the question presented in the previous research [18, 26, 28, 34, 35]. We note that while some questions for the cognitive load can be measured using the NASA-TLX questionnaire [21], we adopted the aforementioned design method to minimize the total number of questions and fit within the time constraints of the study.

8 RESULTS

8.1 Overall Performance

8.1.1 Task Completion Time. Fig. 8–A shows the mean and 95% confidence interval (CI) for task completion time for each sub-route. We conducted the Shapiro-Wilk test for the normality of TCT for nine sub-routes. Out of nine sub-routes, normality was not confirmed for three sub-routes. Thus, we used the nonparametric test, the Wilcoxon signed-rank test, to compare the metric between system-aided and cane-only conditions. We compared TCT for two conditions using the Wilcoxon signed-rank test and revealed that the system-aided condition significantly took a longer time for Route 2-1 and Route 3-3 and a shorter time for Route 2-3 ($p < .05$ for all Route 2-1, 2-3, and 3-3), compared to the cane-only condition. In the cane-only condition, some participants had difficulty finding intersections and destinations and sometimes got lost. B04 got lost in Route 1-3, and B09 and B10 got lost in Route 2-2, resulting in the cane-only condition having larger confidence intervals for task completion time for these routes. Also, Tab. 3 reports the average scanning time in each sub-route.

8.1.2 Distance to Destination Area. Fig. 8–B shows the result of this metric. We conducted the Shapiro-Wilk test on this metric for nine sub-routes. Out of nine sub-routes, normality was not confirmed for eight sub-routes. Thus, we used the nonparametric test, the Wilcoxon signed-rank test, to compare the metric between system-aided and cane-only conditions. There were no significant differences for all routes except Route 3-2 ($p < .05$). This is because some participants had no difference in this value, zero, regardless of a condition. Still, the Figure shows that the system-aided condition produced generally smaller mean values and confidence intervals than the cane-only condition. This can be attributed to the system's ability to guide participants within the destination area and prevent them from making significant navigation errors. In the cane-only condition, while we provided accurate distances in the route description, seven participants made more than three meters of navigation errors. For example, B04 arrived 9.6 meters, and B08 arrived 7.6 meters away from the destination.

8.1.3 Number of Times Asked for Route Description and Subjective Rating. Tab. 3 shows the average number of times participants asked for route descriptions. Participants did not ask for the route descriptions in the system-aided condition.

Fig. 9 shows the results of subjective ratings. We performed the statistical analysis using the Wilcoxon signed-rank test for Q1–Q6 and observed significance for all questions, indicating that the system-aided condition was rated higher than the cane-only condition ($p < 0.05$ for all Q1–Q6). Moreover, all participants gave positive scores (greater than 5) to our system for Q7, Q9–Q11. For the ability to tell the shape of intersections (Q8), most participants except B03 and B11 gave positive scores. Also, Tab. 2 shows the SUS scores.

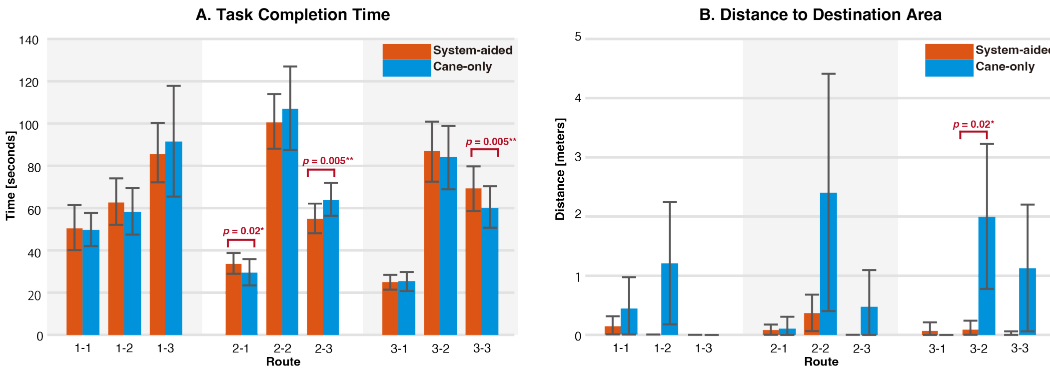

Fig. 8. Bar graph showing the distribution of task completion time and distance to destination area for each sub-routes.

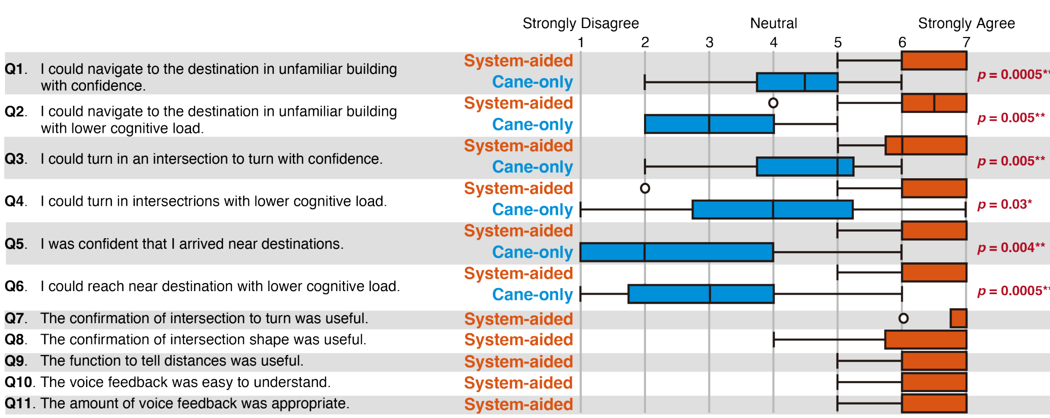

Fig. 9. Questions and seven-point Likert scale responses from our study with blind participants. Responses marked with * indicate $p < .05$ and ** indicate $p < .01$ significant difference between the systems when applying the Wilcoxon sign rank test .

8.2 Qualitative Feedback

Compared to their usual strategy of navigating unfamiliar buildings, ten participants appreciated the system's design, which involves the capturing process of a floor map image by sighted assistants, as it may enable them to navigate to multiple destinations independently: **C6**: "*(Route description by sighted people is) fine if all you have to do is go to the room. However, you may need to leave the room and move around the building. In such cases, if I have a picture of a floor map taken by a sighted assistant, I may be able to move around independently. I think it is a good idea because we can reduce various costs just by having the photos taken.*" (B10) and, **C7**: "*Asking where I want to go in the first place is, of course, hard and stressful, but in the end, I catch people again and ask, 'Where is the entrance?' is also stressful. This system does not require that, so it is good that I don't have to ask for help all the time.*" (B03) In addition, regarding their cognitive load, participants appreciated that they did not have to remember the route to their destination: **C8**: "*For example, I can remember an explanation of just going straight and turning right, but I can't remember if there is some further*

Table 3. Route features and results of Average Times Asked for Route Description and Average Scanning Time at each intersection

		Number of Intersecitons	Length of Route [Meters]	Average Times Asked per Route	Average Scan Time Per Intersection [Seconds]
Route 1	1-1	1	32	0.75	4.56
	1-2	2	35	0.67	5.38
	1-3	2	55	1.00	6.14
Route 2	2-1	1	16	0.67	6.84
	2-2	4	64	2.33	4.47
	2-3	2	36	1.00	3.68
Route 3	3-1	1	12	0.25	5.50
	3-2	2	54	1.00	4.90
	3-3	2	40	0.67	4.17

explanation. The system was very good because I didn't have to remember, and I could leave it to the system to guide me." (B08)

On the other hand, B06 and B09 disagreed with the design that required them to ask the sighted people to capture the floor map. B06 and B09, as well as B07, expressed concern about handing their smartphones to others: **C9**: *"My iPhone is really precious to me, so I honestly don't like asking someone I don't know to take a picture with it."* (B06) B09 and four other participants (B01, B07, B08, and B11) also expressed concern about relying on sighted people who did not know when using the system: **C10**: *"I think getting assistance from sighted strangers easily is a challenge in this system. (When I talk to strangers) I don't know who we are talking to or what kind of people we are talking to."* (B09)

Many participants appreciated the ability to tell distance. **C11**: *"The system told me when I arrived, "Please face the right" or "Please face left." So, I felt the distance was right. With a white cane, thinking about how many meters I have gone, I walk 1m, 2m, 3m, 4m...and so on. I can only walk with the feeling that this would be about 8 meters. Thinking about meters is extremely tiring because I'm using my nerves to walk."* (B04) Related to Q3, Q4, Q7, and Q8, ten participants provided a score above 4, commenting positively on the function of notifying them of the information on intersections. **C12**: *" Usually, I am unsure where intersections are, so I have to follow the wall to find the edge of the wall. Intersections are scary for those who cannot see. So I was impressed that the system told me the exact location of the intersection."* (B08) In particular, ten participants provided positive feedback towards systems feedback, conveying the shape and directions of intersections. An example of the feedback to this feedback was: **C13**: *" Using a cane, I cannot determine if this is truly an intersection or just a hollow part of the wall. If the system knows that we are at an intersection, it tells me to "go to left", and then I can understand that I need to turn left without verifying with my cane."* (B12)

9 DISCUSSION

9.1 Acceptance of Snap&Nav

Throughout the study with sighted participants, most of them were able to use the system without being accustomed to the system, and appreciated that the system eliminates the need to explain routes (C1) and the need to guide blind people to their destination (C2). Crucially, all participants answered they were willing to use the system when asked by blind people (Q6 in Fig. 4), implying that our design involving sighted assistants may be appreciated.

Throughout the study with blind people, participants appreciated that they were able to navigate independently in the unfamiliar building compared to their daily experiences. Usually, blind people have to ask others multiple times to navigate multiple destinations (C7) and memorize the description of the route by sighted people (C8). In contrast, the design of our system allows them to independently navigate to multiple destinations within a floor once a floor map image has been obtained without memorizing the description of the route. Therefore, with the proposed system, participants were able to complete all tasks without asking for a route (Sec. 8.1.3). Furthermore, although the system required users to scan at every intersection, adding roughly five seconds to the process, they were satisfied with the overall experience the system provided (Fig. 9). Most importantly, ten participants expressed that the total benefits of the system outweigh the inconvenience of asking for assistance with image capturing once before the navigation (C6), indicating the potential acceptance of the design the system, which involves sighted people in the intermediate step, by blind people.

While recent research has primarily aimed at reducing user effort in navigation systems through automation [4, 16, 26, 37, 46], our approach emphasizes interaction with sighted assistants and the system. We achieve map and sensor infrastructure (*e.g.*, BLE beacons) less solutions by engaging sighted people in acquiring floor map images and blind users in scanning environments at intersections. Although this approach might initially cause reluctance among both sighted assistants and blind users, our experimental results demonstrate its potential acceptance by both groups. However, we note that future solutions may adopt an approach from Teng *et al.* [48], which would seamlessly connect blind users to nearby sighted assistance. Overall, by employing a design that involves assistance from sighted people and scanning interaction from blind users, we achieve the first step towards scalable and map-less navigation for blind people in potentially diverse buildings.

9.2 User Experience of Map Analysis Module

Sighted participants were able to use the map analysis module without being accustomed to it. Following the system's voice instructions, participants were able to capture floor maps that appeared wide in the image with minimum lighting, resulting in a node map with minimum errors, as illustrated in Fig. 5. On their first trials, they were able to complete the task with an average time of 88.62 seconds and felt they were able to use the system easily (Q1–2 in Fig. 4). Out of 100 overall trials, most participants successfully annotated position and orientation, with only one position error and five orientation errors. The sighted participants agreed that they were willing to use the system when asked by blind people (Q6 in Fig. 4). Meanwhile, we observed concerns from participants. One major concern was the difficulty in determining whether node maps can be used for navigation (C4 and C5). After generating a node map, the system instructs the user to check the node map, but it did not instruct what specifically should be checked. Therefore, the future system should display the criteria to be checked, such as the correctness of the locations of nodes and their connections.

9.3 User Experience of Navigation Module

Using the system, blind participants were able to arrive closer to their destinations with increased confidence and reduced cognitive load, without affecting the task completion time. While in the system-aided condition, it took time to scan at intersections (Tab. 3), in the cane-only condition, it also took time for participants to complete the task for reasons such as the difficulty in finding intersections and destinations (Sec. 8.1.1). As a result, we did not observe significant differences for all routes between the TCT of the system-aided condition and the cane-only condition (Fig. 8). This indicates that the participants mostly maintained their usual walking speed while using the system for navigation. The distance to the destination area of the system-aided condition was less than one meter on average, and its confidence intervals were smaller compared to the cane-only condition (Fig. 8). It indicated the system navigated participants to the destination with small errors. In addition, we gave accurate route descriptions with precise numbers for distances and intersections in the user study. However, the route description given by sighted people is often inaccurate, for example, containing wrong distances [28]. In such cases, participants may not be able to navigate to destinations accurately. In the future, we believe the system can be made more practical by introducing mechanisms to solve the last-one-meter problem (*e.g.*, sign detection [1, 5, 44]).

Furthermore, the system enabled participants to navigate with more confidence and lower cognitive load (Q1–6 of Fig. 9), which is due to two functionalities: announcement of the existence of an intersection, and announcement of distances. All participants provided feedback that the system's announcement of distances via scale estimation (Sec. 6.3) and announcement of the entrance of an intersection via intersection detection (Sec. 6.2.1) helped reduce their cognitive load, as they no longer had to keep track of how far they had walked (C11) and struggle finding intersections (C12). While we designed simple voice feedback to convey the shape of intersections or directions to turn, ten participants appreciated it (C13). Considering the above, we conclude that the system enhanced their navigation experience to the destination.

9.4 Concern of Dependence on Sighted Assistants

While the system design was appreciated by blind users, several considerations need to be made for the design of the system. Two participants disagreed with the design of the system, which incorporates sighted assistants' help in the system usage flow. They were concerned that they had to ask sighted people to capture a floor map image. One of them mentioned that they would not like to hand their smartphone to others as it is expensive and precious to them (C9), and the other expressed concern about relying on strangers (C10). To address the first concern, we may adopt a design to have sighted assistants capture floor map image with their own smartphone and send it to blind users' devices (*e.g.*, via Airdrop[2]). To address the second concern, an alternative strategy could involve blind users scanning the environment while the system attempts to localize itself from the information acquired through it. The system may analyze features such as the shapes of intersections or the names of stores on signage [52]. Such information could serve as landmarks, allowing us to refine and apply localization techniques previously established in research [7].

9.5 Limitation and Future Work

This user study design and the system had several limitations. Firstly, the study was conducted only in our university building. In real world settings, there are obstacles, which the current system does not handle. For more practical usage, we aim to incorporate obstacle avoidance methods proposed in the previous work for smartphones [30, 42] and evaluate the usability. Also, the experimental location contained only simple 90-degree turns. However, in real world environments, there may be complex-shaped intersections, such as those with large open spaces or Y-shapes, which the current system is unable to detect. To achieve this, we can modify the intersection detection method used in more practical settings with robots [28, 53], and integrate it into our smartphone-based system. The optimal feedback method for complex-shaped intersections may differ from the straightforward instructions we used in this study (*e.g.*, "left" and "right") Therefore, we aim to explore these feedback methods further by conducting a user study in real-world buildings. The floor map analysis module also does not handle users' orientation in large open spaces. While detailed orientations are required in such spaces, the system only classifies the input user orientation into two orientations, which in this study used the threshold of 45 degrees to evaluate. The system, however, can still incorporate orientations annotated by sighted assistants, as it allows sighted assistants to annotate precisely. Moreover, the system could utilize the localization method described in Section 9.4 for localizing and ensuring precise orientation.

The performance of floor map analysis is the core element of this system. We prototyped the floor map analysis algorithm and evaluated it with five floor maps in the experimental location, as the main focus of our study was on validating the system design, not the accuracy of general floor maps. The floor map analysis algorithm involved certain assumptions. First, the red region was assumed to represent the current location. Second, the largest area identified by the connected component algorithm was assumed to represent the path area of the floor map. However, when considering the practical usage in the real world, the floor map analysis algorithm needs to be more generalized and evaluated on other floor maps. For future work, we aim to develop a more generalized algorithm. In addition, the system assumed that the scale of the captured floor map was the same over the entire image. The scale of floor maps may not always be accurately presented. Thus, the algorithm also needs to take into account scale differences for real world deployment.

Finally, there were limitations in the participant recruitment. In the first user study with sighted participants, the average age of the participants was 23.8, and the maximum age was 31. In addition, 18 of the 20 sighted assistants were familiar with the building. Therefore, it is unclear how other generations and people unfamiliar with the building would evaluate the system's usability. Thus, we aim to conduct the study with a broader age range in the future. In the second user study with blind participants, three participants had previously visited the experimental location for the previous study. Thus, the three participants may have had a bias, such as positive impressions of the study.

10 CONCLUSION

This paper proposes Snap&Nav, a system that navigates blind people based on the analyzed floor map images captured by sighted people. Our study with 20 sighted participants revealed that they could use the system without being accustomed to the system, and all of them were generally willing to use the system when asked by blind people. The second user study was 12 blind people navigating in an unfamiliar building. The results indicated that Snap&Nav increased confidence and reduced cognitive load, without affecting the task completion time. Additionally, ten blind participants felt that the system design, which involves sighted assistants, was acceptable, as they felt that the benefit it would provide would outweigh the inconvenience of asking for assistance. For future work, we aim to further generalize and evaluate the floor map analysis algorithm.

ACKNOWLEDGMENTS

We would like to thank all the participants who took part in our user study. We would also thank Miraikan - The National Museum of Emerging Science and Innovation for their support. This work was supported by JSPS KAKENHI (JP23KJ2048).

REFERENCES

[1] Mrouj Almuhajri and Ching Y. Suen. 2022. A Complete Framework for Shop Signboards Detection and Classification. In *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, Piscataway, NJ, USA, 4671–4677. https://doi.org/10.1109/ICPR56361.2022.9956399

[2] Apple. 2023. How to use AirDrop on your iPhone or iPad. Retrieved in July, 2023 from https://support.apple.com/en-us/HT204144.

[3] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7. https://doi.org/10.1201/9781498710411-35

[4] Hsuan-Eng Chen, Yi-Ying Lin, Chien-Hsing Chen, and I-Fang Wang. 2015. BlindNavi: A navigation app for the visually impaired smartphone user. In *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems*. ACM, New York, NY, USA, 19–24. https://doi.org/10.1145/2702613.2726953

[5] Xiaoxue Chen, Lianwen Jin, Yuanzhi Zhu, Canjie Luo, and Tianwei Wang. 2020. Text Recognition in the Wild: A Survey. arXiv:2005.03492 [cs.CV]

[6] Eliseo Clementini. and Andrea Pagliaro. 2020. The Construction of a Network for Indoor Navigation. In *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management - GISTAM*. SciTePress, Setúbal, Portugal, 254–261. https://doi.org/10.5220/0009488902540261

[7] Ryan Crabb, Seyed Ali Cheraghi, and James M Coughlan. 2023. A Lightweight Approach to Localization for Blind and Visually Impaired Travelers. *Sensors* 23, 5 (2023), 2701. https://doi.org/10.3390/s23052701

[8] Apple Developer. 2023. Apple Human Interface Guidelines. Retrieved in September, 2023 from https://developer.apple.com/design/human-interface-guidelines/buttons.

[9] Edsger W Dijkstra. 2022. A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: His Life, Work, and Legacy*. ACM, New York, NY, USA, 287–290. https://doi.org/10.1145/3544585.3544600

[10] Christin Engel, Karin Müller, Angela Constantinescu, Claudia Loitsch, Vanessa Petrausch, Gerhard Weber, and Rainer Stiefelhagen. 2020. Travelling More Independently: A Requirements Analysis for Accessible Journeys to Unknown Buildings for People with Visual Impairments. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA. https://doi.org/10.1145/3373625.3417022

[11] Adam Van Etten, Dave Lindenbaum, and Todd M. Bacastow. 2019. SpaceNet: A Remote Sensing Dataset and Challenge Series. arXiv:1807.01232 [cs.CV]

[12] Navid Fallah, Ilias Apostolopoulos, Kostas Bekris, and Eelke Folmer. 2012. The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 425–432. https://doi.org/10.1145/2207676.2207735

[13] Alexander Fiannaca, Ilias Apostolopoulous, and Eelke Folmer. 2014. Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. ACM, New York, NY, USA, 19–26. https://doi.org/10.1145/2661334.2661453

[14] Meiqing Fu, Rui Liu, Bing Qi, and Raja R. Issa. 2020. Generating straight skeleton-based navigation networks with Industry Foundation Classes for indoor way-finding. *Automation in Construction* 112 (2020), 103057. https://doi.org/10.1016/j.autcon.2019.103057

[15] Nicholas Giudice, William Whalen, Tim Riehle, Shane Anderson, and Stacy Doore. 2019. Evaluation of an Accessible, Real-Time, and Infrastructure-Free Indoor Navigation System by Users Who Are Blind in the Mall of America. *Journal of Visual Impairment & Blindness* 113 (04 2019), 0145482X1984091. https://doi.org/10.1177/0145482X19840918

[16] Nicholas A. Giudice, Benjamin A. Guenther, Toni M. Kaplan, Shane M. Anderson, Robert J. Knuesel, and Joseph F. Cioffi. 2020. Use of an Indoor Navigation System by Sighted and Blind Travelers: Performance Similarities across Visual Status and Age. *ACM Transactions on Accessible Computing* 13, 3, Article 11 (2020), 27 pages. https://doi.org/10.1145/3407191

[17] João Guerreiro, Dragan Ahmetovic, Kris M Kitani, and Chieko Asakawa. 2017. Virtual navigation for blind people: Building sequential representations of the real-world. In *Proceedings of the 19th International ACM SIGACCESS Conference on computers and accessibility*. ACM, New York, NY, USA, 280–289. https://doi.org/10.1145/3132525.3132545

[18] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M Kitani, and Chieko Asakawa. 2019. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 68–82. https://doi.org/10.1145/3308561.3353771

[19] João Guerreiro, Eshed Ohn-Bar, Dragan Ahmetovic, Kris Kitani, and Chieko Asakawa. 2018. How Context and User Behavior Affect Indoor Navigation Assistance for Blind People. In *Proceedings of the 15th International Web for All Conference* (Lyon, France) *(W4A '18)*. Association for Computing Machinery, New York, NY, USA, Article 2, 4 pages. https://doi.org/10.1145/3192714.3192829

[20] Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham. 2018. VizWiz Grand Challenge: Answering Visual Questions From Blind People. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Piscataway, NJ, USA, 3608–3617.

[21] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Radarweg 29, Amsterfdam, Netherlands, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[22] Tomoya Honto, Yoshihiro Sugaya, Tomo Miyazaki, and Shinichiro Omachi. 2017. Analysis of floor map image in information board for indoor navigation. In *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Piscataway, NJ, USA, 1–7. https://doi.org/10.1109/IPIN.2017.8115896

[23] Gesu India, Mohit Jain, Pallav Karya, Nirmalendu Diwakar, and Manohar Swaminathan. 2021. VStroll: An audio-based virtual exploration to encourage walking among people with vision impairments. In *Proceedings of the 23rd International ACM SIGACCESS Conference on computers and accessibility*. ACM, New York, NY, USA, 1–13. https:

//doi.org/10.1145/3441852.3471206
[24] JaidedAI. 2023. JaidedAI/EasyOCR. Retrieved in July, 2023 from https://github.com/JaidedAI/EasyOCR.
[25] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting Blind Photography. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. Association for Computing Machinery, New York, NY, USA, 203–210. https://doi.org/10.1145/2049536.2049573
[26] Jee-Eun Kim, Masahiro Bessho, Shinsuke Kobayashi, Noboru Koshizuka, and Ken Sakamura. 2016. Navigating Visually Impaired Travelers in a Large Train Station Using Smartphone and Bluetooth Low Energy. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. ACM, New York, NY, USA, 604–611. https://doi.org/10.1145/2851613.2851716
[27] J. Taery Kim, Wenhao Yu, Yash Kothari, Jie Tan, Greg Turk, and Sehoon Ha. 2023. Transforming a Quadruped into a Guide Robot for the Visually Impaired: Formalizing Wayfinding, Interaction Modeling, and Safety Mechanism. arXiv:2306.14055 [cs.RO]
[28] Masaki Kuribayashi, Tatsuya Ishihara, Daisuke Sato, Jayakorn Vongkulbhisal, Karnik Ram, Seita Kayukawa, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2023. PathFinder: Designing a Map-less Navigation System for Blind People in Unfamiliar Buildings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–16. https://doi.org/10.1145/3544548.3580687
[29] Masaki Kuribayashi, Seita Kayukawa, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. 2021. LineChaser: A Smartphone-Based Navigation System for Blind People to Stand in Lines. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445451
[30] Masaki Kuribayashi, Seita Kayukawa, Jayakorn Vongkulbhisal, Chieko Asakawa, Daisuke Sato, Hironobu Takagi, and Shigeo Morishima. 2022. Corridor-Walker: Mobile Indoor Walking Assistance for Blind People to Avoid Obstacles and Recognize Intersections. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022), 1–22. https://doi.org/10.1145/3546714
[31] Gerard Lacey and Shane MacNamara. 2000. Context-aware shared control of a robot mobility aid for the elderly blind. *The International Journal of Robotics Research* 19, 11 (2000), 1054–1065. https://doi.org/10.1177/02783640022067968
[32] Bing Li, J Pablo Munoz, Xuejian Rong, Jizhong Xiao, Yingli Tian, and Aries Arditi. 2016. ISANA: wearable context-aware indoor assistive navigation with obstacle avoidance for the blind. In *European Conference on Computer Vision*. Springer, New York, NY, USA, 448–462. https://doi.org/10.1007/978-3-319-48881-3_31
[33] Jun Li, Chee Leong Chan, Jian Le Chan, Zhengguo Li, Kong Wah Wan, and Wei Yun Yau. 2021. Cognitive Navigation for Indoor Environment Using Floorplan. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Piscataway, NJ, USA, 9030–9037. https://doi.org/10.1109/IROS51168.2021.9635850
[34] Guanhong Liu, Tianyu Yu, Chun Yu, Haiqing Xu, Shuchang Xu, Ciyuan Yang, Feng Wang, Haipeng Mi, and Yuanchun Shi. 2021. Tactile Compass: Enabling Visually Impaired People to Follow a Path with Continuous Directional Feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (, Yokohama, Japan,) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 28, 13 pages. https://doi.org/10.1145/3411764.3445644
[35] Chen-Lung Lu, Zi-Yan Liu, Jui-Te Huang, Ching-I Huang, Bo-Hui Wang, Yi Chen, Nien-Hsin Wu, Hsueh-Cheng Wang, Laura Giarré, and Pei-Yi Kuo. 2021. Assistive Navigation Using Deep Reinforcement Learning Guiding Robot With UWB/Voice Beacons and Semantic Feedbacks for Blind and Visually Impaired People. *Frontiers in Robotics and AI* 8 (2021), 1–23 pages. https://doi.org/10.3389/frobt.2021.654132
[36] Karin Müller, Christin Engel, Claudia Loitsch, Rainer Stiefelhagen, and Gerhard Weber. 2022. Traveling More Independently: A Study on the Diverse Needs and Challenges of People with Visual or Mobility Impairments in Unfamiliar Indoor Environments. *ACM Transactions on Accessible Computing* 15, 2 (2022), 1–44. https://doi.org/10.1145/3514255
[37] Masayuki Murata, Dragan Ahmetovic, Daisuke Sato, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. 2018. Smartphone-based indoor localization for blind navigation across building complexes. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, Piscataway, NJ, USA, 1–10. https://doi.org/10.1109/PERCOM.2018.8444593
[38] Ruth G Nagassa, Matthew Butler, Leona Holloway, Cagatay Goncu, and Kim Marriott. 2023. 3D Building Plans: Supporting Navigation by People who are Blind or have Low Vision in Multi-Storey Buildings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–19. https://doi.org/10.1145/3544548.3581304
[39] John Nicholson. 2010. Generation and Analysis of Verbal Route Directions for Blind Navigation. https://digitalcommons.usu.edu/etd/672
[40] Yueyong Pang, Liangchen Zhou, Bingxian Lin, Guonian Lv, and Chi Zhang. 2020. Generation of navigation networks for corridor spaces based on indoor visibility map. *International Journal of Geographical Information Science* 34, 1 (2020), 177–201. https://doi.org/10.1080/13658816.2019.1664741

[41] Manoj Penmetcha, Arabinda Samantaray, and Byung-Cheol Min. 2017. Smartresponse: Emergency and non-emergency response for smartphone based indoor localization applications. In *International Conference on Human-Computer Interaction*. Springer, New York, NY, USA, 398–404. https://doi.org/10.1007/978-3-319-58753-0_57

[42] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2019. WatchOut: Obstacle sonification for people with visual impairment or blindness. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 402–413. https://doi.org/10.1145/3308561.3353779

[43] Vinitha Ranganeni, Mike Sinclair, Eyal Ofek, Amos Miller, Jonathan Campbell, Andrey Kolobov, and Edward Cutrell. 2023. Exploring Levels of Control for a Navigation Assistant for Blind Travelers. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, NY, USA, 4–12. https://doi.org/10.1145/3568162.3578630

[44] Manaswi Saha, Alexander J Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. 2019. Closing the gap: Designing for the last-few-meters wayfinding problem for people with visual impairments. In *The 21st international acm sigaccess conference on computers and accessibility*. ACM, New York, NY, USA, 222–235. https://doi.org/10.1145/3308561.3353776

[45] Elen Sargsyan, Bernard Oriola, Marc JM Macé, Marcos Serrano, and Christophe Jouffrais. 2023. 3D Printed Interactive Multi-Storey Model for People with Visual Impairments. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3544548.3581304

[46] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. 2019. NavCog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *ACM Transactions on Accessible Computing* 12, 3 (2019), 14. https://doi.org/10.1145/3340319

[47] Simon Schmitt, Larissa Zech, Katinka Wolter, Thomas Willemsen, Harald Sternberg, and Marcel Kyas. 2017. Fast routing graph extraction from floor plans. In *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Piscataway, NJ, USA, 1–8. https://doi.org/10.1109/IPIN.2017.8115868

[48] Yuanyang Teng, Connor Courtien, David Angel Rios, Yves M Tseng, Jacqueline Gibson, Maryam Aziz, Avery Reyna, Rajan Vaish, and Brian A Smith. 2024. Help Supporters: Exploring the Design Space of Assistive Technologies to Support Face-to-Face Help Between Blind and Sighted Strangers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–24.

[49] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, San Francisco, CA, USA, 7464–7475.

[50] Xiyue Wang, Seita Kayukawa, Hironobu Takagi, and Chieko Asakawa. 2022. BentoMuseum: 3D and Layered Interactive Museum Map for Blind Visitors. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3517428.3544811

[51] Ziwei Xu, Haitian Zheng, Minjian Pang, Yangchun Zhu, Xiongfei Su, Guyue Zhou, and Lu Fang. 2017. Utilizing High-level Visual Feature for Indoor Shopping Mall Navigation. arXiv:1610.01906 [cs.CV]

[52] Yutaro Yamanaka, Seita Kayukawa, Hironobu Takagi, Yuichi Nagaoka, Yoshimune Hiratsuka, and Satoshi Kurihara. 2021. One-Shot Wayfinding Method for Blind People via OCR and Arrow Analysis with a 360-degree Smartphone Camera. In *Proceedings of the 18th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (Beppu, Japan) *(MobiQuitous '21)*. Springer, New York, NY, USA, 19 pages. https://doi.org/10.1007/978-3-030-94822-1_9

[53] Fan Yang, Dung-Han Lee, John Keller, and Sebastian Scherer. 2021. Graph-based topological exploration planning in large-scale 3d environments. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Piscataway, NJ, USA, 12730–12736. https://doi.org/10.1109/ICRA48506.2021.9561830

[54] Liping Yang and Michael Worboys. 2015. Generation of navigation graphs for indoor space. *International Journal of Geographical Information Science* 29, 10 (2015), 1737–1756. https://doi.org/10.1080/13658816.2015.1041141

[55] Chris Yoon, Ryan Louie, Jeremy Ryan, MinhKhang Vu, Hyegi Bang, William Derksen, and Paul Ruvolo. 2019. Leveraging Augmented Reality to Create Apps for People with Visual Disabilities: A Case Study in Indoor Navigation. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 210–221. https://doi.org/10.1145/3308561.3353788

[56] T. Y. Zhang and C. Y. Suen. 1984. A Fast Parallel Algorithm for Thinning Digital Patterns. *Commun. ACM* 27, 3 (mar 1984), 236–239. https://doi.org/10.1145/357994.358023

Received February 2024; revised May 2024; accepted June 2024