One-Shot Wayfinding Method for Blind People via OCR and Arrow Analysis with a 360-degree Smartphone Camera

Yutaro Yamanaka[1], Seita Kayukawa[2], Hironobu Takagi[3], Yuichi Nagaoka[4], Yoshimune Hiratsuka[5], and Satoshi Kurihara[1]

[1] Graduate School of Science and Technology, Keio University, Kanagawa, Japan
[2] Waseda University, Tokyo, Japan
[3] IBM Research - Tokyo, Tokyo, Japan
[4] Tokyo Independent Living Support Center for the Visually Impaired, Tokyo, Japan
[5] Department of Ophthalmology, Juntendo University School of Medicine, Tokyo, Japan

Abstract. We present a wayfinding method that assists blind people in determining the correct direction to a destination by taking a *one-shot* image. Signage is standard in public buildings and used to help visitors, but has little benefit for blind people. Our *one-shot wayfinding method* recognizes surrounding signage in all directions from an equirectangular image captured using a 360-degree smartphone camera. The method analyzes the relationship between detected text and arrows on signage and estimates the correct direction toward the user's destination. In other words, the method enables wayfinding for the blind without requiring either environmental modifications (*e.g.* Bluetooth beacons) or preparation of map data. In a user study, we compared our method with a baseline method: a signage reader using a smartphone camera with a standard field of view. We found that our method enabled the participants to decide directions more efficiently than with the baseline method.

Keywords: Visual impairment · signage · OCR · arrow detection

1 Introduction

Signage is standard in public buildings and shows directions toward points of interest to help visitors find their way [14], but it has little benefit for blind people. Recent studies have proposed assistive technologies that can recognize signage information (*e.g.* text or pictograms on signage) by combining a smartphone camera and computer vision technologies such as optical character recognition (OCR) [2, 29]. One difficulty for blind people in using such signage recognition systems is taking pictures with the appropriate framing and aiming the camera toward a sign quickly and accurately [15, 24]. Thus, blind users sometimes cannot obtain required information from these systems. In this situation, it can be difficult for them to distinguish whether the reason is a lack of signage in the environment or incorrect camera framing.

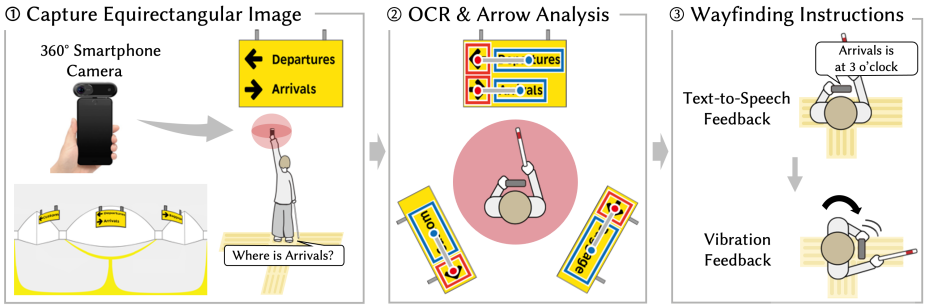

Fig. 1. Overview of our *one-shot wayfinding* method for blind people. 1) When approaching an intersection in a public building, a blind user takes a picture with a 360-degree camera attached to a smartphone. 2) The method detects text and arrows on surrounding signage and links them to estimate the direction to a destination. 3) The method provides wayfinding instructions for the estimated direction via text-to-speech and vibration feedback.

To overcome this limitation, we developed a *one-shot wayfinding* method, which uses a 360-degree smartphone camera that captures all signage around a user in only one shot without having to adjust the camera aim. To decide the correct direction to a destination, blind people need to understand the directions of arrows on signs. For example, as shown in Fig. 1, a sign with a right arrow indicates to turn right at the intersection. Therefore, our wayfinding method was developed to recognize not only text but also arrows that point in the direction to a destination. The method recognizes text, arrows, and text-arrow associations on surrounding signage, and it converts the direction of each arrow into an egocentric direction (*i.e.* a direction relative to the user's body). Then, it verbalizes the egocentric direction in terms of a clock position, which is a standard way of presenting directions to blind people. In other words, our approach enables a wayfinding method for blind people that does not require either environmental modifications, such as markers, Bluetooth Low Energy (BLE) beacons, and Wi-Fi beacons, or preparation of data, such as maps and points-of-interest datasets.

The proposed method first detects text and arrows from a captured equirectangular image by using an OCR system and a convolutional neural network (CNN) object detector. It then links detected text to each detected arrow via a minimum spanning tree (MST). We set the edge weights to link text and arrows by considering the relationship between them (*e.g.* text above an arrow tends to have a weaker correspondence with the arrow than text below the arrow). For example, in Fig. 1 (2), the method links "Departures" and "Arrivals" to the left and right arrows, respectively. It then estimates the directions to the destinations in egocentric coordinates relative to the user's current orientation. For instance, when a sign in front of the user shows "Arrivals" and this is linked to a right arrow, the smartphone says, *"'Arrivals' is at 3 o'clock."* To further convey the estimated direction, the smartphone gives vibration alerts when the user faces the correct direction [Fig. 1 (3)].

To evaluate the usability of our method, we performed a user study with eight blind people. To provide a baseline system, we implemented a simple signage reader system that uses the RGB camera built into a smartphone (not a 360-degree camera). We asked the participants to find the correct direction to a destination by using either the proposed system or the baseline system. To evaluate the effectiveness of each system's interface, we designed a Wizard-of-Oz-style [17] study using images for which our algorithm worked successfully. Because of the COVID-19 outbreak, we conducted the user study in a laboratory space that reproduced wayfinding decision-making situations in public buildings by using pre-captured images from places such as an international airport and a railway station. We observed that the proposed system enabled the participants to determine directions with a smaller amount of rotation than with the baseline system. The participants' feedback also supported our hypothesis that the proposed system is useful for wayfinding tasks in public buildings. On the basis of our findings, we discuss future directions to develop a more flexible and comfortable wayfinding system for public buildings.

2 Related Work

2.1 Indoor Navigation and Wayfinding System for Blind People

Navigating large and unfamiliar public buildings (*e.g.* international airports, railway stations, and shopping centers) is a challenging task for blind people [12, 13]. Thus, researchers have proposed various types of indoor navigation systems for blind people. Most of these systems provide turn-by-turn navigation instructions by using localization technologies (*e.g.* Bluetooth Low Energy (BLE) beacons [3, 30], ultra-wide band (UWB) [4], and Wi-Fi [10]) or environment databases (*e.g.* images [32] and maps [22]). While these navigation or wayfinding systems can provide accurate wayfinding instructions, they require installing sensors or code in the environment or constructing a database of the environment.

To implement a wayfinding or navigation system that does not require additional sensors or databases, recent research using computer vision has enabled systems that can recognize useful information for wayfinding (*e.g.* doors [8], flat floors [11], pictograms and text on signage [29, 35]). However, sign locations do not always correspond to the route to a destination. For example, a sign with a right arrow indicates that the destination is to the right, not at the sign's location. To overcome this limitation, we propose a wayfinding method that can recognize not only text but also arrows that show the direction to a destination. By analyzing the relationship between detected text and arrows, the method gives blind users egocentric directions toward their destinations.

2.2 Environment Recognition via Smartphone Camera

With the expansion of smartphone usage in the blind people community [25], various smartphone camera-based recognition systems have been proposed to

help blind users obtain information on their surroundings (*e.g.* object [1, 2, 6, 18, 29, 38], text [1, 2, 6, 38], and signage [29, 31]). However, it is still challenging for blind users to capture an entire target object with a smartphone camera [15, 24]. While capture-assistance systems using audio [15, 21, 33, 37] or vibration [21] have been proposed, standard smartphone cameras require blind users to rotate them and face them toward objects. Having blind people change their orientation may cause them to lose their way and become disoriented [16]. We thus use a 360-degree smartphone camera, which can capture all surrounding signage in one shot, for wayfinding tasks; we call this *one-shot wayfinding*.

3 Design: One-Shot Wayfinding Method

Here, we describe our wayfinding method design specifically for the following typical situation: Blind pedestrians walk through a public building such as an airport, railway station, or shopping center. They walk along the tactile pavings in the building but is unfamiliar with the route. Thus, when they approache a tactile paving intersection, they cannot decide which direction to take.

3.1 One-Shot Wayfinding method with 360-degree Camera

While there are smartphone-based assistive technologies that can recognize information on the surroundings via a smartphone camera [2, 29, 18, 6, 38, 1, 31, 37], it is challenging for blind users to point a camera toward a target and capture its entirety [15, 24]. When blind users cannot obtain required information with such technologies, it can be difficult for them to distinguish whether the reason is a lack of signage in the environment or incorrect camera framing.

Therefore, we attach a 360-degree camera to a smartphone. Compared with built-in smartphone cameras with a standard field of view (FoV), 360-degree cameras have three advantages: (1) they can capture all surrounding signage (including directly behind) in one shot, (2) they can capture the whole of each sign (*i.e.* no text is cut off), and (3) they do not require aiming. This is why we call our method a *one-shot wayfinding*. In other words, it can distinguish whether there is signage around a user with only one camera shot.

3.2 Wayfinding Instructions via OCR and Arrow Analysis

The combination of a 360-degree camera and OCR can recognize text appearing around a user, including non-signage text (*e.g.* posters and signboards). However, reading out all text can cognitively overwhelm the user [26]. In addition, sign locations do not always correspond to the route to a destination. For example, when a user approaches an intersection and a sign with a right arrow is in front of the user, it indicates that the destination is to the right, not at the sign location [Fig. 1]. In this situation, the system should tell the user to turn right.

To overcome these limitations, we designed our wayfinding method to detect not only text but also arrows on signage. The method then links detected text to

each detected arrow by considering their spatial relationship, through a process we call *arrow analysis*. By using the linking results, the method recognizes only signage text and estimates the egocentric direction to each destination [Sec. 4]. It then instructs the user on the correct direction to the destination.

4 Implementation

Our one-shot wayfinding method consists of two components: (1) a **web API** that performs equirectangular image preprocessing, arrow detection, OCR, and arrow analysis; and (2) a **smartphone interface** that estimates egocentric directions to destinations and provides wayfinding instructions. For our user study with blind people, we attached an Insta360 ONE[6], which can capture 7K (6912×3456 pixels) equirectangular images, to an iPhone6 [7]. Captured images are horizontally corrected by the camera's built-in gyroscope. As a result, blind users can capture equirectangular images horizontally without concern for the smartphone's angle and rotation. After capturing an image, the method sends it to the web API on our server.

4.1 Equirectangular Image Preprocessing

Because equirectangular images are spatially distorted and unsuitable for arrow detection, the method first converts a captured image into cubemap images (1728×1728 pixels). The method converts the equirectangular image into five cubemap images having 18-degree horizontal overlaps. The method uses the five cubemap images for arrow detection, and the original equirectangular image and the back cubemap image for OCR.

4.2 Arrow Detection and OCR

The method detects arrows by using the YOLOv3 object detector [28]. To train the arrow detection model, we collected 1140 arrow images taken in public spaces from Open Images Dataset [20] and Flicker API[8] (only Creative-Commons-licensed images). We annotated the collected images with bounding boxes and four types of arrow labels (straight, down, right, or left). The method detects bounding boxes of arrows from the five cubemap images and obtains their positions in the equirectangular image coordinate system. Because of the cubemap images' overlaps, the method may detect the same arrow twice from different cubemap images. In that case, it picks the bounding box with the higher confidence value.

The method detects text on the captured equirectangular image by using an OCR package from Google Cloud Vision API[9]. As each end of the equirectangular image may contain separated text from behind the user, the method also

[6] https://www.insta360.com/product/insta360-one/
[7] https://support.apple.com/kb/sp705
[8] https://www.flickr.com/services/api/
[9] https://cloud.google.com/vision/docs/ocr/

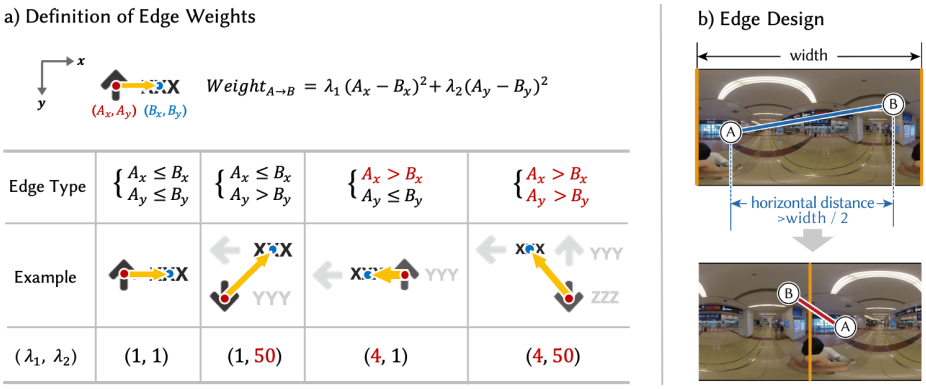

Fig. 2. a) Edge weights are defined according to the related positions of two nodes. b) If the horizontal distance between two nodes is more than half the width of the equirectangular image, the distance is recalculated by shifting the image 180 degrees.

uses the back cubemap image for OCR. Then, the method obtains the center positions of detected text in the equirectangular image coordinate system.

4.3 Arrow Analysis

Next, the method connects detected text to each detected arrow by using a minimum spanning tree (MST) [9]. It constructs a directed graph with two types of nodes: (1) arrow nodes representing the center positions of the detected arrow bounding boxes, and (2) text nodes representing the center positions of the detected text. In the graph, edges connect among text nodes and between text and arrow nodes, but not among arrow nodes.

As shown in Fig. 2a, the method defines edge weights according to the related positions of two nodes and the lengths of edges. A signage design guideline [14] reported that left-aligned signage (text to the right of arrows) makes recognition more comfortable for those whose language is read from left to right. Thus, the method sets the edge weight higher when a node connects to the left or above nodes. Following our observations, we set the edge weight values (λ_1, λ_2) as illustrated in Fig. 2a. The horizontal distance of an edge may be more than half the width of the equirectangular image (blue edge in Fig. 2b). In that case, the method horizontally shifts the equirectangular image 180 degrees and then calculates the edge weight (red edge in Fig. 2b).

The method adds a new node that links to each arrow node with a zero-weight edge and applies the MST algorithm [9] from the new node. Then, it removes edges whose weight is zero or more than 5000. As a result, it obtains trees with an arrow node as the root node and text nodes as the child nodes [Fig. 4a]. We assumed that each root (arrow) node's label (*i.e.* right, left, straight, or down) indicates the direction toward a destination provided by the child (text) nodes.

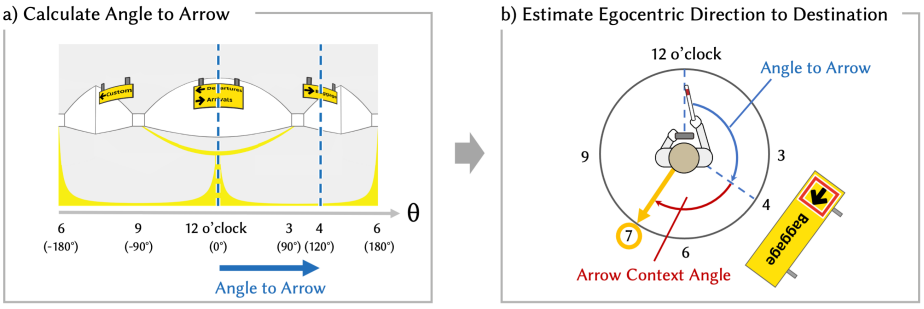

Fig. 3. Direction estimation. a) The method calculates the *angle to arrow*, which is the angle between the center positions of an arrow and the equirectangular image. b) The method estimates the egocentric direction to a destination by adding the *arrow context angle* to the *angle to arrow*.

4.4 Direction Estimation

From the arrow analysis results, the method estimates the angle to the destination indicated by the text on a sign. As illustrated in Fig. 3a, it first calculates the angle between an arrow's center position and the center of the equirectangular image (*angle to arrow*). It then estimates the angle to the destination on the basis of the predicted arrow label [Fig. 3b]. We assumed that a straight or down arrow indicates that the destination is in the same directions as a sign and that a right or left arrow indicates that the destination is 90 degrees to the right or left of a sign. Concretely, we define an *arrow context angle* on the basis of four types of arrow labels: straight/down arrows are 0 degrees, left is -90 degrees, and right is 90 degrees. The method obtains the egocentric direction to the destination by adding the *arrow context angle* to the *angle to arrow* [Fig. 3b].

4.5 Process Evaluation

We evaluated the OCR and arrow analysis processes by using equirectangular images captured in public buildings. As there are no open datasets of equirectangular images capturing signage in public buildings, we constructed our own dataset. It consists of 104 images captured at tactile paving intersections in an international airport (43 images) and a railway station (61 images). We also used these images in our user evaluation [Sec. 5]. For the process evaluation, we annotated 255 arrows and 330 text instances on signage that are useful for wayfinding at intersections, the area of a sign covered by each arrow, and the directions of the tactile paving branches indicated by each arrow on a sign.

Tab. 1 shows each process's accuracy. We obtained the arrow analysis accuracy by calculating the algorithm's success rate in linking the annotated arrows and text. For the direction estimation accuracy, we calculated the success rate in recognizing the correct tactile paving branch as the closest branch to the direction estimated by the method for each annotated arrow. Two diagonal arrows

	Arrow detection (%)	OCR (%)	Arrow analysis (%)	Direction estimation (%)	Overall performance (%)
International airport	88.0	55.2	90.0	83.6	45.7
Railway station	85.7	58.4	60.0	85.0	41.1
Total	86.7	57.3	70.9	84.5	42.7

Table 1. Summary of process evaluation results.

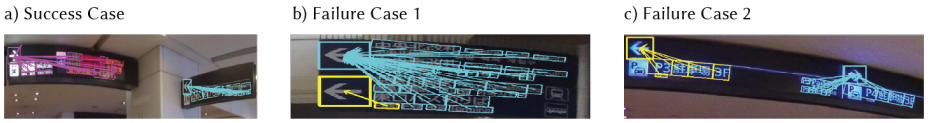

Fig. 4. a) Success case: all text linked to the appropriate arrow. b) Failure case 1: narrow space between arrows, 9 of 104 images. c) Failure case 2: wide space separating text, 33 of 104 images.

were also included in the failed cases of the arrow detection. The table also lists the method's overall performance as defined by the success rate in linking the text and the correct tactile paving branch. Fig. 4 shows examples of the OCR and arrow analysis results. In Fig. 4a, the method linked text on sign to the appropriate arrow. On the other hand, we observed many failure cases. One cause of failure was incorrect arrow and text detection. The detection performance, especially for OCR, was worse for small text and arrows. Another cause was signage design: the arrow analysis accuracy decreased when there was a narrow space between arrows (Fig. 4b) or a wide space separating text (Fig. 4c). We will discuss possible solutions to improve each process's accuracy in Sec. 7.2.

4.6 User Interface

Wayfinding Instructions The method provides wayfinding instructions to the user via text-to-speech and vibration feedback. After estimating the direction of text, the smartphone first reads out instructions in terms of clock positions. For example, suppose the user inputs "Arrivals" to the smartphone, and the estimated direction to "Arrivals" is 120 degrees to the right of the user's current orientation. In this case, the smartphone calculates the clock position of the estimated direction and says, *"Arrivals is at 4 o'clock.''* Loomis *et al.* showed that instructions given with clock positions can help blind people navigate to a specific destinations [23].

While text-to-speech feedback can provide clear information for blind people, it is difficult for them to slightly adjust their orientation [30]. Accordingly, we designed an interface combining audio and vibration alerts. The smartphone gives vibration alerts when the user is facing the expected direction. The current orientation is obtained with the smartphone's built-in gyroscope.

Smartphone Interface The smartphone interface has three buttons: (1) Record button located at the top pf the smartphone screen: Used to register destination-related keywords via speech input. (2) Capture button located in the bottom left: Used to capture an equirectangular image while holding the camera overhead. (3) All button located in the bottom right: Used to hear readout of all text linked to arrows. The user can push this button when the smartphone does not read out any audio instructions related to the registered keywords. On the basis of audio feedback, the user can register new keywords or conclude that there is no useful signage around the user.

5 User Study

To evaluate the effectiveness of our wayfinding method interface, we performed a user study with eight blind participants: five legally blind people and three totally blind people (P3. P5, and P6), as listed in Tab. 2. They all considered themselves to have good orientation and mobility skills. Seven participants (P1–P4, P6–P8) regularly used a white cane, and P5 as their navigation aids and P5 owned a guide dog. In this study, we compared our one-shot wayfinding system against a baseline system: a signage reader using a smartphone camera with a standard field of view (FoV).

5.1 Experimental Setup

User Study in Laboratory Space We performed our study in a laboratory space rather than public buildings given the restriction under the COVID-19 pandemic situation. We used equirectangular images pre-captured at tactile paving intersections in public buildings [Sec. 4.5]. For each captured image, we laid tactile paving on the laboratory floor to reproduce the real intersections captured at those points. To focus on evaluating the effectiveness of each interface for signage-recognition-based wayfinding, we designed a Wizard-of-Oz-style study [17] using equirectangular images for which our algorithm worked successfully [Sec. 5.1].

Proposed System When a user pushed the "capture button" [Sec. 4.6], the system obtained the smartphone's orientation relative to the tactile paving via the smartphone's gyroscope. Using this orientation, the system shifted the equirectangular image to match the direction and the center of the image. We argue that this process reproduced the scenario of a user capturing an equirectangular image at a tactile paving intersection in a public building. Next, the system sent the shifted image to the web API to get wayfinding instructions.

Baseline System Inspired by smartphone camera-based recognition applications for visually impaired people [2, 18, 29, 31], we implemented a simple signage reader system using a smartphone camera with a standard FoV as the baseline system. To operate in the laboratory space, the baseline system used

Demographic information			Task accuracy (%)		SUS score	
ID	Eyesight	Age	Proposed	Baseline	Proposed	Baseline
P1	Legally blind	43	100	50	60	**80**
P2	Legally blind	46	100	50	**90**	65
P3	Totally blind	48	87.5	87.5	**72.5**	62.5
P4	Legally blind	52	100	75	62.5	**85**
P5	Totally blind	41	87.5	62.5	**75**	30
P6	Totally blind	47	87.5	87.5	**95**	82.5
P7	Legally blind	55	100	87.5	**90**	82.5
P8	Legally blind	39	87.5	100	**82.5**	20
	Mean	46.4	93.8	75.0	78.4	63.4
	SD	5.4	6.7	18.9	13.1	25.3

Table 2. Demographic information of our participants, task accuracy, and SUS score for each system.

pre-captured equirectangular images and the pre-obtained results of arrow detection and the OCR for these images. During the wayfinding task, the system obtained the smartphone's current orientation via its gyroscope sensor. It then read out the registered text and all arrow labels within the pre-defined FoV (horizontal FoV: 100 degrees; vertical FoV: 80 degrees) around the smartphone's direction. When more than one arrow label was within the FoV, the system read out the labels in order from top left to bottom right.

Dataset For the Wizard-of-Oz-style study, we picked eight pairs of equirectangular images for which our algorithm worked successfully from the dataset used for the process evaluation [Sec. 4.5] and asked participants to perform wayfinding tasks with either the proposed or baseline system for each pair of images. We chose pairs with (1) the same building (an airport or a station), (2) the same direction on the target sign (right, left, forward, or backward from the participant's orientation during image capture), and (3) the same number of tactile paving branches at the intersection (three or four). The participants were divided into two groups, X and Y, and the dataset was divided into two groups of images, A and B. Group X completed the wayfinding tasks for image group A with the proposed system and image group B with the baseline, while group Y used the opposite system for each of A and B. The order of the systems and images was randomized for each participant.

5.2 Task

We asked the participants to choose the pre-defined correct direction from the tactile paving branches on the floor with either the proposed or baseline system. The participants held the smartphone with one hand and their white cane

with the other. At the beginning of each task, the experimenter registered the destination and gave the phone to the participants. We asked them to find the correct tactile paving branch to the destination (*e.g.* "Please select the tactile paving branch to 'Arrivals.' "). The participants used each system to decide the correct direction and reported verbally that they had completed the task if they found the correct branch.

To provide egocentric clock-position-based instructions correctly, we instructed the participants to capture images while keeping the camera's horizontal direction with the user's face direction. The proposed system could continuously obtain the smartphone's current angle relative to the initial angle when the user captured the image by using the smartphone's gyroscope. Therefore, the vibration feedback can correctly convey the estimated direction to users even if the smartphone's direction and the face direction are misaligned.

5.3 Procedure

After obtaining informed consent (IRB approved) from the participants, we first administered a questionnaire on demographics and navigation habits. Next, we described the two systems (proposed and baseline) and conducted a short training session (15–20 minutes) to familiarize the participants with each system. Then, we asked the participants to perform wayfinding tasks with either the proposed or baseline system. As they performed the tasks, the interfaces and the dataset (dataset A and B described in Sec. 5.1) were changed in a counterbalanced order. After all tasks were completed, we interviewed the participants. The task process took around 20 minutes, while the whole experiment took approximately 90 minutes per participant.

5.4 Metrics

Task Accuracy and Task Completion Time We defined the rate of success in deciding the correct tactile paving branch as the *task accuracy*. In addition, during the main session, we measured the *task completion time* for each task. Note that the proposed system estimated the correct direction by using the web API for every task, but the baseline system used pre-obtained results for arrow detection and OCR and thus required no processing time to read signage. One of the study's main goals was to evaluate the effectiveness of each system's interface. Therefore, in measuring the task completion time with the proposed system, we both included and excluded the processing time.

Rotation Efficiency We measured the amount of rotation (yaw angle) of each system during each task with the smartphone's gyroscope. We then obtained the *rotation efficiency* by calculating the absolute difference between the participant's rotation and the angle between the participant's initial orientation and the correct tactile paving branch. Lower rotation efficiency values mean that participants could decide the correct direction without extra rotation. Because

the rotation efficiency became too large when the participants chose the wrong direction, we only calculated it for cases of success.

Interview After completing all the tasks, we asked participants to rate four sentences by using a 7-point Likert scale ranging from "1: strongly disagree" to "7: strongly agree", with 4 denoting "neutral."

Q1: *"I decided the direction confidently with the proposed/baseline/no system."*[10]
Q2: *"The proposed/baseline system helped me in wayfinding."*
Q3: *"The proposed/baseline system was easy to use."*
Q4: *"I felt comfortable with the proposed/baseline system."*

We also asked the participants to rate each item on the system usability scale (SUS) [7]. Finally, we asked open-ended questions about the advantages and disadvantages of each system, and we asked for suggestions to improve each system.

6 Results

6.1 Overall Performance

Task Accuracy Tab. 2 lists the task accuracy of each interface for each participant. Five of the eight participants had a higher task accuracy with the proposed system than with the baseline system. Though the average task accuracy of the proposed system (93.8%; 60/64) was higher than that of the baseline (75%; 48/64), we found no significant differences between them ($p = 0.057$).

Task Completion Time Fig. 5 shows the average task completion time of each system for each condition (target sign direction: left, front, right, and back) and for all conditions. Fig. 5 also shows the average task completion time for the proposed system excluding the processing time Here, we report the mean and SD of the processing time: the communication time was 1.01 ± 0.62 seconds, the web API processing time was 6.91 ± 0.78 seconds (Intel Xeon E5-2698 v4, 2.20 GHz, NVIDIA GTX Station), and the total processing time was 7.92 ± 0.88 seconds. Our statistical analysis by using a Wilcoxon signed-rank test revealed that the proposed system (**excluding** the processing time) enabled the participants to complete the tasks significantly quicker than with the baseline system ($p < 0.0001$). This was also the case when the participants tried to read signage behind them ($p = 0.004$ for the task completion time **including** the processing time).

[10] All communication with the participants was in their native language. In this paper, we describe any translated content in the form of *"translated content"*.

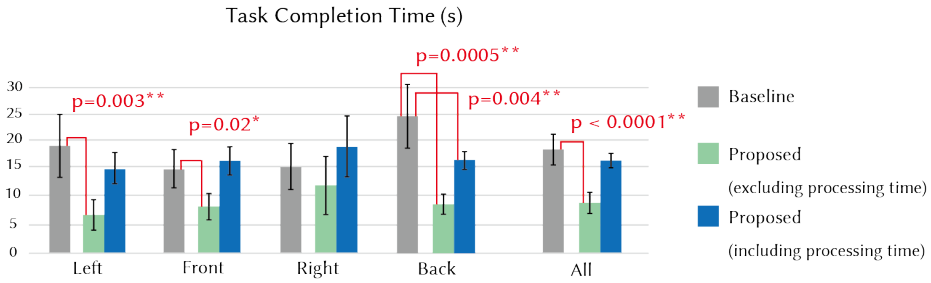

Fig. 5. Task completion time, with bars showing 95% confidence intervals and p-values for a Wilcoxon signed-rank test on the task completion time. ** and * indicate 0.005 and 0.05 levels of significance, respectively.

Rotation Efficiency The mean, SD, and 95% confidence interval of the rotation efficiency for each system were as follows: mean = 66.6, SD = 94.3, and 95% confidence interval = 40.1˜93.1 for the proposed system; mean = 110.1, SD = 98.3, and 95% confidence interval = 81.6˜138.7 for the baseline system. When we compared each system's results by using a Mann-Whitney test, we observed a significant difference ($p = 0.00095$) in the rotation efficiency. This result showed that participants using the proposed system found the correct direction without extra rotation as compared with the baseline system.

Video Observation The video recordings enabled us to analyze the participants' behavior when they chose the wrong tactile paving branch. Four participants using the proposed system (P3, P5, P6, and P8) sometimes selected the wrong branch when the system estimated that the correct direction was between two branches (4 failure cases /64 total trials). On the other hand, six participants using the baseline system (P2–P7) sometimes chose the wrong branch when the system read out multiple arrow labels.

System Ratings Tab. 2 lists the SUS scores for each participant. Six of the eight participants gave a higher SUS score to the proposed system than to the baseline system. Fig. 6 summarizes the results for the Likert scale questions. For all questions, the proposed system received positive ratings (*i.e.* the median rating was more than four). Participants who valued the baseline system more on Q4 (comfort) mainly pointed out the weight of the proposed system. We describe the detailed comments on the usability of each system in a later section.

6.2 Qualitative System Feedback

Six of the eight participants (P3–P8) agreed that our signage reader systems (both proposed and baseline) can be useful in wayfinding decision-making situations: A1: *"I am not confident in walking alone in public spaces, so I would like*

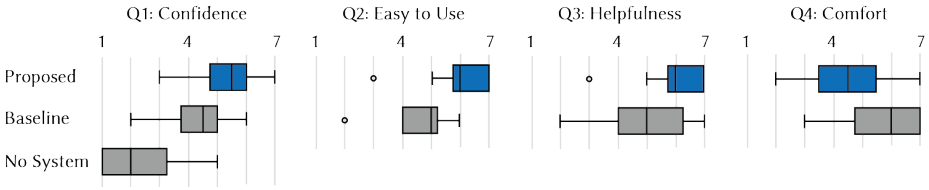

Fig. 6. Summary of Likert scale responses (1: strongly disagree to 7: strongly agree).

to use a system that reads signage to tell me the direction to my destination." (P3); and A2: *"It would be easier to move with confidence if the system read the signage in unfamiliar places."* (P4)

Six participants (P2, P3, P5–P8) gave positive feedback on our one-shot wayfinding system because it does not require users to change their orientation and scan surrounding signage: A3: *"Rotating on the spot with the second (baseline) system destroyed my mental map. I appreciated the first (proposed) system because it did not require me to turn my body in various directions."* (P5); and A4: *"The 360-degree camera made it possible to determine directions without moving and rotating, and I thought it could be used in buildings I have never visited."* (P8) However, P4 preferred to face his camera toward surrounding signage: A5: *"I found the first (baseline) system natural and easy to use because it read the signage in the direction I was facing."* (P4)

Six participants (P2, P3, P5–P8) appreciated the proposed system's wayfinding instructions with integrated clock-position-based audio feedback and vibration feedback: A6: *"The second (proposed) system allowed me to intuitively and accurately understand the direction. On the other hand, with the first (baseline) system, I took more time to think about the direction to the destination after facing the sign direction."* (P2); A7: *"The vibration feedback of the first (proposed) system gave me confidence that I was facing the correct direction. The directional feedback of the second (baseline) system was vague, and I was not confident in my direction after rotating."* (P7); and A8: *"The directional instructions using the clock position instantly gave me a clear vision of the direction I needed to face. I thought the vibration feedback would be useful for determining the direction even where there are no landmarks such as tactile paving."* (P6)

Five participants (P3–P7) gave negative feedback on the baseline because it sometimes read out multiple arrow labels: A9: *"When the first (baseline) system read out multiple arrow labels, I had to figure out which one was the true direction to the destination."* (P4) In contrast, P1 preferred the baseline system's simple audio instructions ("turn right/left") to the proposed system's clock-position instructions: A10: *"It was difficult for me to understand the clock-position directions of the first (proposed) system, while the second (baseline) system's directional instructions were easy to understand intuitively."* (P1)

Regarding suggestions to improve our systems for use in public buildings, we obtained the following comments: A11: *"I want to check the direction when I*

walk along a road and reach an intersection." (P3); and A12: *"I want to check my direction when I lose confidence while walking on tactile paving."* (P6)

Half the participants (P2, P4–P6) mentioned that the proposed system with the 360-degree camera was heavy: A13: *"The camera was so heavy that walking with it all the time was a burden."* (P4) Finally, one participant mentioned the limitation of the user study in a laboratory space as compared with real-world use: A14: *"It was difficult to get a true feel for the usability of both systems without using them in a real environment."* (P3)

7 Discussion

7.1 Effectiveness of One-Shot Wayfinding Method

The results showed that the task accuracy of the proposed system (93.8%) was higher than that of the baseline (75%), with $p = 0.057$. In addition, the proposed system significantly reduced the extra rotation as compared with the baseline system [Sec. 6.1]. The participants also agreed that the proposed system had an advantage in not requiring users to change their orientation and scan surrounding signage [A3–A4]. Participants' feedback on the interface also supported the proposed system's effectiveness [A6–A8].

Regarding the rotation effectiveness, users with the baseline system had to decide correct positions on the basis of arrow reading results, but the proposed system directly provides the correct direction from arrow analysis results. We argue that this accounted for the difference in the task completion time. Excluding the processing time, the proposed system's task completion time [Fig. 5] was significantly shorter than that of the baseline system. Moreover, when the target sign was located behind the participant's initial position, the proposed system's task completion time, including the processing time, was significantly shorter than that of the baseline [Fig. 5].

7.2 Toward More Accurate Wayfinding Systems

For the Wizard-of-Oz-style study, we used equirectangular images for which our algorithm worked successfully from the dataset. However, through the process evaluation [Sec. 4.5], We observed that our algorithm can be improved for real-world usage, and further evaluation in various environments is needed. The arrow detection and OCR performance was a bottleneck for the wayfinding method's accuracy. Specifically, the accuracy, particularly for OCR, decreased when signage was far away from the user. To improve the detection accuracy, we will consider designing an interface that guides the user close to a sign by using arrow or signage detection [34, 35] results and then allows them to retake the equirectangular image at a place close to the target sign.

As listed in Tab. 1, while our graph-theory-based arrow analysis method achieved relatively high accuracy at an international airport (90.0%), its accuracy at an railway station was only 60.0%. The method analyzes arrows by

considering only the relative positions of arrows and text. To improve the arrow analysis, we will consider using signage boundaries or CNN-based computer vision techniques such as document layout analysis [5].

While our direction estimation process achieved 84.5% task accuracy, the estimation accuracy decreased when a sign did not directly face the 360-degree camera. To increase direction estimation accuracy, one possible solution would be to detect surrounding cues such as the directions of tactile paving branches from the captured images [36] and use these results for direction estimation.

7.3 Future System Design

Processing Time Reduction We found no significant differences in the overall task completion time between the proposed system (**including** the processing time) and the baseline system [Fig. 5]. Regarding the processing time, we expect to exploit the ever-improving processing power of CPUs and GPUs and the ever-increasing communication bandwidth. We will also design faster algorithms and find a better edge-cloud balance to reduce the processing time.

User Interface Design While the proposed system's overall performance was positive, we also found opportunities to improve the user interface. Six participants preferred the proposed system, which automatically estimates the correct direction [A3–A4], but P4 gave the baseline system a higher SUS score because it allows users to estimate the correct direction from arrow detection results [A5]. While many participants commented postiively on the proposed system's clock-position-based instructions [A6 and A8], P1 found them difficult to understand [A10]. We argue that the requirements for a wayfinding system depend on the user's orientation and mobility (O&M) skills, familiarity with the target public building, and individual preferences. We aim to further explore various types of interface options, including sonification method [27], 3D spatialized audio [23], vibration patterns [16, 19], and shape-changing devices [16], to provide more suitable wayfinding instructions to users.

7.4 Limitation of Laboratory-Based User Study

As P3 commented [A14], we agree that there are many differences between our study in a laboratory space and a real-world study in public buildings. First, in a public building, the user would have to stop at a tactile paving intersection to capture images. The lab-based study, which used pre-captured images, did not reproduce this procedure. Second, in complex buildings that repeatedly require wayfinding tasks, users may choose the wrong direction. The lab-based study missed an opportunity to understand how users recover from errors in wayfinding tasks. Third, the acoustic environment in public buildings is hardly reproducible in a laboratory study, which prevents the use of techniques such as echolocation (*e.g.* the direction of wide-open spaces) and sound landmarks (*e.g.* escalators). Thus, we need to confirm the system's practical usability in public buildings

that facilitate echolocation and other senses. To explore more suitable interfaces and algorithms in a real-world setting, we will conduct a study in which blind participants are asked to approach a specific goal turn-by-turn in a public space by recognizing signage with our method.

8 Conclusion

We proposed a *one-shot wayfinding* method that uses a 360-degree smartphone camera to recognize all signage around a blind user. The method analyzes the relationship between detected text and arrows on signage and estimates the egocentric direction to a destination. It provides text-to-speech feedback of the estimated direction on the basis of clock positions and gives vibration alerts when the user faces the indicated direction. A user study with eight blind participants in a laboratory revealed that the proposed system enabled them to choose the correct tactile paving branch to a destination more efficiently than with a baseline system. The proposed system significantly reduced the extra rotation, and the task completion time excluding the processing time was significantly shorter than that of the baseline system. While the participants' feedback supported our hypothesis that the proposed method is useful for wayfinding tasks, we also recognized the need for a real-world user study in public buildings. The proposed method has the possibility of assisting users in unknown places without requiring either environmental modifications like distributed beacons or preparation of maps or points-of-interest datasets. We hope to explore this possibility further and make the technology practical to help blind people with daily activities.

ACKNOWLEDGMENTS

We would like to thank all participants who took part in our user study. We would also thank Japan Airport Terminal Co., Ltd. and East Japan Railway Company. This work was supported by AMED (JP20dk0310108, JP21dk0310108h0002), JSPS KAKENHI (JP20J23018), and Grant-in-Aid for Young Scientists (Early Bird, Waseda Research Institute for Science and Engineering, BD070Z003100).

References

1. Bespecular. (2016), https://www.bespecular.com
2. Seeing ai. (2017), https://www.microsoft.com/en-us/seeing-ai
3. Ahmetovic, D., Gleason, C., Ruan, C., Kitani, K., Takagi, H., Asakawa, C.: Navcog: A navigational cognitive assistant for the blind. In: MobileHCI (2016)
4. Alnafessah, A., Al-Ammar, M.A., Alhadhrami, S., Al-Salman, A., Al-Khalifa, H.S.: Developing an ultra wideband indoor navigation system for visually impaired people. IJDSN **12**, 403–416 (2016)
5. Augusto Borges Oliveira, D., Palhares Viana, M.: Fast cnn-based document layout analysis. In: ICCVW (2017)

6. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., Yeh, T.: Vizwiz: Nearly real-time answers to visual questions. In: UIST (2010)
7. Brooke, J.: Sus: a "quick and dirty'usability. Usability evaluation in industry p. 189 (1996)
8. Fiannaca, A., Apostolopoulous, I., Folmer, E.: Headlock: A wearable navigation aid that helps blind cane users traverse large open spaces. In: ASSETS (2014)
9. Gabow, H.N., Galil, Z., Spencer, T., Tarjan, R.E.: Efficient algorithms for finding minimum spanning trees in undirected and directed graphs. Combinatorica **6**(2) (1986)
10. Gallagher, T., Wise, E., Li, B., Dempster, A.G., Rizos, C., Ramsey-Stewart, E.: Indoor positioning system based on sensor fusion for the blind and visually impaired. In: IPIN (2012)
11. Garcia, G., Nahapetian, A.: Wearable computing for image-based indoor navigation of the visually impaired. In: WH. A (2015)
12. Guentert, M.: Improving public transit accessibility for blind riders: A train station navigation assistant. In: ASSETS (2011)
13. Guerreiro, J.a., Ahmetovic, D., Sato, D., Kitani, K., Asakawa, C.: Airport accessibility and navigation assistance for people with visual impairments. In: CHI (2019)
14. Guidelines, I.H.F.: Wayfinding guidelines international health facility guidelines (2016), http://www.healthfacilityguidelines.com/GuidelineIndex/Index/Wayfinding-Guidelines
15. Jayant, C., Ji, H., White, S., Bigham, J.P.: Supporting blind photography. In: ASSETS (2011)
16. Kayukawa, S., Tatsuya, I., Takagi, H., Morishima, S., Asakawa, C.: Guiding blind pedestrians in public spaces by understanding walking behavior of nearby pedestrians. IMWUT **4**(3) (2020)
17. Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications. TOIS **2**(1), 26–41 (1984)
18. Ko, E., Ju, J.S., Kim, E.Y.: Situation-based indoor wayfinding system for the visually impaired. In: ASSETS (2011)
19. Kuribayashi, M., Kayukawa, S., Takagi, H., Asakawa, C., Morishima, S.: Linechaser: A smartphone-based navigation system for blind people to stand in line. In: CHI (2021)
20. Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., Ferrari, V.: The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. IJCV **128**(7), 1956–1981 (2020)
21. Lee, K., Hong, J., Pimento, S., Jarjue, E., Kacorri, H.: Revisiting blind photography in the context of teachable object recognizers. In: ASSETS (2019)
22. Li, B., Muñoz, J.P., Rong, X., Xiao, J., Tian, Y., Arditi, A.: Isana: Wearable context-aware indoor assistive navigation with obstacle avoidance for the blind. In: ECCVW (2016)
23. Loomis, J.M., Lippa, Y., Klatzky, R.L., Golledge, R.G.: Spatial updating of locations specified by 3-d sound and spatial language. JEP:LMC **28**(2), 335 (2002)
24. Manduchi, R., Coughlan, J.M.: The last meter: Blind visual guidance to a target. In: CHI (2014)
25. Pal, J., Viswanathan, A., Song, J.H.: Smartphone adoption drivers and challenges in urban living: Cases from seoul and bangalore. In: IHCI (2016)
26. Panëels, S.A., Olmos, A., Blum, J.R., Cooperstock, J.R.: Listen to it yourself! evaluating usability of what's around me? for the blind. In: CHI (2013)

27. Presti, G., Ahmetovic, D., Ducci, M., Bernareggi, C., Ludovico, L., Baratè, A., Avanzini, F., Mascetti, S.: Watchout: Obstacle sonification for people with visual impairment or blindness. In: ASSETS (2019)
28. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv (2018)
29. Saha, M., Fiannaca, A.J., Kneisel, M., Cutrell, E., Morris, M.R.: Closing the gap: Designing for the last-few-meters wayfinding problem for people with visual impairments. In: ASSETS (2019)
30. Sato, D., Oh, U., Naito, K., Takagi, H., Kitani, K., Asakawa, C.: Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. In: ASSETS (2017)
31. Shen, H., Coughlan, J.M.: Towards a real-time system for finding and reading signs for visually impaired users. In: ICCHP (2012)
32. Treuillet, S., Royer, E.: Outdoor/indoor vision based localization for blind pedestrian navigation assistance. IJIG **10**, 481–496 (2010)
33. Vázquez, M., Steinfeld, A.: Helping visually impaired users properly aim a camera. In: ASSETS (2012)
34. Wang, S., Tian, Y.: Indoor signage detection based on saliency map and bipartite graph matching. In: ICBBW (2011)
35. Wang, S., Tian, Y.: Camera-based signage detection and recognition for blind persons. In: ICCHP (2012)
36. Yamanaka, Y., Takaya, E., Kurihara, S.: Tactile tile detection integrated with ground detection using an rgb-depth sensor. In: ICAART (2020)
37. Zhao, Y., Wu, S., Reynolds, L., Azenkot, S.: A face recognition application for people with visual impairments: Understanding use beyond the lab. In: CHI (2018)
38. Zhong, Y., Lasecki, W.S., Brady, E., Bigham, J.P.: Regionspeak: Quick comprehensive spatial descriptions of complex images for blind users. In: CHI (2015)