Enhancing Walk-Light Detector Usage for the Visually Impaired: A Comparison of VR Exploration and Verbal Instructions

Jonggi Hong
jhong8@stevens.edu
Stevens Institute of Technology
Hoboken, New Jersey, USA

James M. Coughlan
coughlan@ski.org
Smith-Kettlewell Eye Research Institute
San Francisco, California, USA, USA

ABSTRACT

People with visual impairments (PVI) increasingly rely on camera-enabled smartphone apps for tasks like photography, navigation, and text recognition. Despite the growing use of these applications, precise camera aiming remains a significant challenge. This study explores the impact of virtual reality (VR) exploration compared to traditional text/audio (TA) instructions in the context of learning to use a walk-light detector app at traffic intersections. We developed a VR exploration tool based on insights gathered from interviews with PVI. A user study was conducted, involving 13 PVI participants divided into two groups: VR exploration and TA instructions. Following indoor training using the respective approaches, participants from both groups used the walk light detector app outdoors. According to the participants' subjective feedback, a higher proportion of participants in the TA group found the training easier, potentially due to shortcomings in our VR protocol and differences between the real world and VR. However, more VR participants gained insights into walk light detection and felt unable to use the detector without VR training, compared to the TA group.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Empirical studies in interaction design**; **Ubiquitous and mobile devices**.

KEYWORDS

Blind photography; virtual reality; walk light; navigation; blindness and low vision

ACM Reference Format:
Jonggi Hong and James M. Coughlan. 2023. Enhancing Walk-Light Detector Usage for the Visually Impaired: A Comparison of VR Exploration and Verbal Instructions. In *The 21st International Web for All Conference, May 13–14, 2024, Singapore.* ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

W4A '24, May 13–14, 2024, Singapore

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The growing importance of smartphones as tools used to complete a wide range of everyday tasks means that the ability to use a camera has become an essential skill for nearly all smartphone users. The smartphone camera is used to take photographs and videos of friends and family and to document special places and occasions, scan QR codes and product barcodes to access menus and other information, scan paper receipts and credit cards, perform Optical Character Recognition (OCR) to read text, and even to acquire imagery of nearby landmarks to enhance GPS localization accuracy in navigation apps such as Apple Maps[1]. To better understand their surroundings and specific objects in it, PVI smartphone users often use scene/object recognition apps such as Seeing AI [3], TapTapSee [38], and Envision [2] and apps to obtain assistance from sighted online agents such as Be My Eyes [15] and Aira [4].

However, aiming the smartphone camera properly to frame an object or scene of interest, and capturing clear, high-quality images, is challenging for many PVI users [21, 25, 43]. While most sighted users can monitor the content and quality of the images they acquire in the camera viewfinder, this feedback is inaccessible to many PVI users, who may inadvertently aim the camera at the wrong target or acquire blurred images because of poor lighting or excessive camera motion. Previous research has explored the use of non-visual feedback to enhance the quality of photos taken by PVI [6, 25, 39, 44, 47]. However, not all camera-enabled apps provide audio and haptic feedback to help the user aim the camera properly (*e.g.*, [16]). Worse still, apps that use computer vision/AI to recognize visual targets often make mis-recognition mistakes (*e.g.*, false positive and false negative recognitions) even when the imagery is clear, so the feedback provided by these apps may cause additional confusion for the user. While omnidirectional cameras have also been utilized to simplify camera aiming [20, 46], they may not provide an ideal solution in scenarios where users must capture a specific target, such as detecting a walk light in a specific direction at an intersection or capturing an object of interest in a cluttered environment. Therefore, it is important that PVI users learn to use the camera as effectively as possible to be able to use camera-enabled apps whether or not they provide such feedback.

PVI has traditionally acquired knowledge of assistive technology through verbal instructions, videos, and sighted assistance. Notably, there has been a recent shift towards leveraging virtual reality (VR) as a tool for novice users to familiarize themselves with new technologies, as highlighted in the study by Zhu *et al.* [48]. PVI has been increasingly using VR to explore unfamiliar environments virtually [17, 18, 24, 32, 34, 40]. Previous research has demonstrated that immersing PVI in a virtual simulation of a real space through

[1]https://developer.apple.com/documentation/arkit/argeotrackingconfiguration

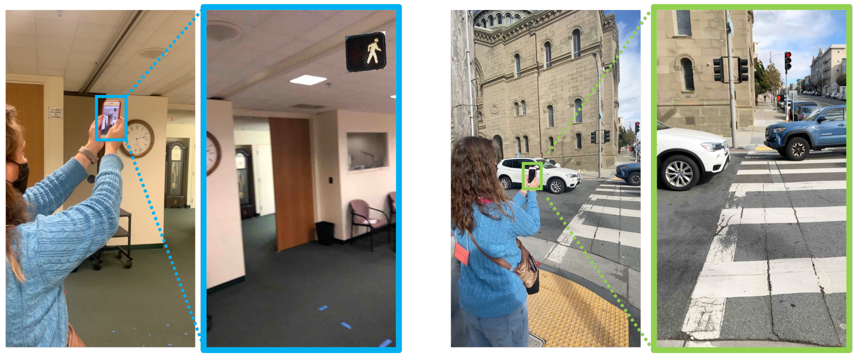

Figure 1: A participant using a walk light detector in a virtual space (left) and at a real intersection (right).

VR can enhance their spatial knowledge [10, 12, 17]. Furthermore, VR emerges as a promising tool for PVI to learn assistive technology specifically tailored for navigation, as emphasized by Theodorou *et al.* [40]. Despite the positive impact of VR on real-world navigation with assistive technology, there remains a scarcity of studies that effectively distinguish between the outcomes of VR-based learning and traditional text/audio instructions. This study aims to bridge this gap by investigating and uncovering the distinctive characteristics of these two instructional methods in the context of learning how to use a walk light detector at a traffic intersection.

To this end, we conducted a user study aimed at comparing the efficacy of a VR training tool against text/audio (TA) instructions in equipping users with camera manipulation skills, focusing particularly on their ability to detect pedestrian walk lights at traffic intersections [2]. Our study involved 13 participants with visual impairments, who were divided into two groups: one receiving VR training and the other receiving TA instructions. Following either an indoor VR session or TA instruction session, participants proceeded outdoors to utilize a designated app across various traffic intersections. Before the study, we designed the VR training tool specifically tailored for PVI, providing audio feedback to the user (in contrast with the visual feedback that most VR applications emphasize for sighted users). Our tool concentrated on helping the user refine their camera orientation techniques and take good-quality images (e.g., with minimal blur). Our VR application incorporated virtual visual targets within the user's environment, challenging participants to capture them within the camera's field of view. The tool provided guidance for camera orientation and speed control through non-visual feedback, informed by preliminary experiments assessing the impact of camera parameters on the accuracy of walk light detection by an object detection model.

The analysis of results, encompassing both quantitative and qualitative dimensions, highlights distinct outcomes arising from VR training and TA instructions. Despite the quantitative assessment indicating no significant difference in the quality of photos captured by participants or the efficacy of walk light detection between the two groups, a nuanced perspective emerges through qualitative analysis: a greater number of participants in the VR group expressed acquiring skills for manipulating a camera and conveyed a perceived inability to use the walk light detector proficiently without VR training, in contrast to the TA group. Conversely, a higher proportion of participants in the TA group found the learning process easy compared to their counterparts in the VR group.

[2]Project webpage: https://www.ski.org/projects/using-vr-help-train-visually-impaired-users-aim-camera

2 RELATED WORK

Prior research on camera-based assistive technologies has emphasized the need to develop skills for capturing high-quality images or videos. In this section, we explore previous research on virtual reality applications for training and blind photography. Additionally, we focus on camera-based systems aimed at aiding PVI in navigating their surroundings, which is directly aligned with the research scenario we emphasized.

2.1 Virtual Reality for Training

Prior studies have consistently demonstrated the effectiveness of VR in elevating student participation, confidence, and enthusiasm [11, 19, 33]. Beyond these benefits, VR offers distinct benefits that set it apart from traditional learning methods. It excels in simulating scenarios that would be challenging or impossible to recreate in real-life settings. Notable instances include the development of VR learning environments tailored for the internet of things [48], chemistry experiments [27], surgical training [26], and aviation training [30]. Moreover, VR contributes to a heightened understanding of spatial context by immersing students in the same virtual environments [22, 29, 41]. This immersive approach proves especially beneficial in teaching abstract concepts rich in spatial information, such as molecular structures [28, 45], astronomical objects [7], and sorting algorithms [35].

Due to these advantages, VR has proven effective in assisting people with disabilities to acquire spatial knowledge and learn how to use assistive technology. Theodorou *et al.* [40] have demonstrated high satisfaction with PVI learning new technology for navigation using VR. They highlighted that VR training can effectively empower PVI users to utilize assistive technologies and enhance the acceptance rate. Moreover, VR has been actively employed to facilitate the learning and exploration of unfamiliar environments for people with disabilities. Previous studies indicate that VR enables people with disabilities to gather information about unfamiliar environments, including identifying comfort areas [32] and getting general spatial knowledge [5, 10, 17, 23, 24], and assessing accessibility [18, 34] through exploration of simulated environments before encountering real-world settings.

2.2 Blind Photography

As camera-based assistive applications become popular, PVI have been actively using a camera for daily tasks. A prior study has revealed that blind people regularly take photos for various purposes such as sharing memories, reading texts, and identifying colors [21]. However, taking clear and well-framed photos remains a challenge for camera-based assistive applications. To address this issue, prior studies have presented real-time guidance for image framing in various applications, including text or barcode readers [39, 44], face detectors [6, 47], object recognizers [25], and blind navigation systems [42, 43]. These assistive systems for blind photography employ different types of non-visual feedback to provide guidance, such as verbal instructions (*e.g.*"left", "right", "up", "down") [6, 8, 21, 43, 44], sonification [8, 25], and haptic feedback [21, 25].

Although prior studies have shown that real-time non-visual feedback can improve the quality of images taken by PVI, they have some limitations. Firstly, many of these approaches rely on computer vision techniques to locate the object of interest in an image frame, which can be unreliable in practice due to blurry images, low light conditions, and cluttered backgrounds. Secondly, when the target object is outside the image frame, users must rely on their own skills without any feedback to adjust the camera's position and orientation. Lastly, there is no agreement on what types of feedback are suitable for blind photography across various camera-based assistive systems, meaning that PVI would need to learn new feedback for different applications. Due to these limitations, PVI still require further learning to capture photos for camera-based assistive applications, eliminating the need for external guidance.

3 DESIGNING THE VR TRAINING TOOL

To identify the challenges and skills in using a walk light detector effectively and navigating their surroundings with camera-based assistive applications, we interviewed PVI. We also carried out a series of experiments to assess the impact of users' proficiency in camera manipulation on the accuracy of walk light detection.

3.1 Understanding Blind Users' Challenges

We conducted interviews with PVI to discuss their experience using a camera and camera-based navigation systems through Zoom. We explored how participants learned to take photos or videos, how frequently they use assistive navigation systems, and the gestures they use to operate camera-based navigation systems. We also inquired about what types of objects they capture using a camera for navigation, as well as their preferred form factor for the camera. When assessing the frequencies of utilizing camera-based assistive apps and navigation systems, we employed an absolute 7-point scale adapted from Rosen *et al.* [37]. We identified the primary themes that emerged from their responses through thematic coding [9].

3.1.1 Participants. We recruited 10 participants from our email lists, including six females and four males, whose ages ranged from 30 to 75 ($M = 46.4, SD = 18.3$). Of the 10 participants, seven reported being totally blind, two reported having light perception, and one reported being legally blind. On average, participants maintained their current level of vision for 35.5 years ($SD = 21.4$). Nine participants had been using a smartphone for an average of 10.7 years ($SD = 4.5$), while one participant had never used one. Seven participants reported using a camera several times a week or more.

3.1.2 Findings. **How do PVI learn to take photos or videos?** Only four reported having experience in learning or practicing using a camera. Two of these participants learned through a process of trial and error, relying on the feedback from assistive apps (*e.g.*, sound feedback for document positioning in the camera frame). The other two received guidance from sighted individuals to capture high-quality photos. The most frequently performed tasks using a camera were reading text ($N = 8$), preserving or sharing memories ($N = 6$), video calling ($N = 5$), and object identification ($N = 3$).

How frequently do they use assistive navigation systems? Participants reported using navigation systems, with or without a camera, once a month ($N = 3$), several times a month ($N = 4$), once a week ($N = 2$), and several times a week ($N = 1$). The navigation systems they used included systems using sensors (*e.g.*, global positioning system, compasses) such as Google Maps and BlindSquare ($N = 10$), video calling apps for seeking assistance from sighted people such as Aira and Be My Eyes ($N = 7$), and computer-vision systems such as OneStep Reader and Google Lookout ($N = 2$).

What challenges do they encounter when using a camera for navigation? They faced several challenges while using a camera for navigation, including difficulties in image framing ($N = 7$), controlling internet connection ($N = 4$), adjusting light conditions ($N = 3$), focusing ($N = 2$), and holding the camera steady ($N = 1$). Drawing from these responses, we focused on assisting PVI in acquiring image framing and focusing skills through VR training.

What types of objects do PVI capture using a camera for navigation? Most participants captured landmarks such as stores and restaurants ($N = 7$). Participants also commonly reported capturing images of rooms in buildings ($N = 5$), people ($N = 5$), signs ($N = 5$), and pedestrian walk lights or traffic lights ($N = 3$). In this study, we focused on the scenario of detecting walk lights using a camera to show the feasibility of the VR training method. We discuss the potential of using VR training in other scenarios in the discussion section.

What are their preferred form factors for the camera? Participants expressed their preferences for different form factors for camera-based navigation systems. Smart glasses were the most preferred form factor ($N = 8$), followed by a smartphone ($N = 3$), and a shoulder-mounted camera ($N = 1$). The ease of aiming the camera influenced the participants' preference for a particular form factor. For instance, P4 who preferred smart glasses said that *"No need to point the camera (on glasses) to a specific place."* However, some participants preferred a smartphone because wearing a device is an inconvenience for them. P8 said, *"... I don't like glasses because it is cumbersome to have glasses and masks together."* Although smart glasses were the most preferred form factor of a camera for PVI, we chose the smartphone as the form factor because it is still the most commonly used device for camera-based assistive technology. The discussion section explores the feasibility of employing VR training for wearable devices.

3.2 Understanding the Walk-Light Detector

To understand the effect of a user's proficiency in manipulating a camera on the performance of a walk light detector, we built a walk light detector (*i.e.*, the outdoor app) and conducted experiments with it. Rather than relying on off-the-shelf applications like OKO [31], we developed a specialized outdoor app to conduct an in-depth analysis of the walk light detector's characteristics and explore the impact of VR training in our user study.

To create the outdoor app, we customized a YOLO v2 object detection model [36] pre-trained on ImageNet dataset [14] with transfer learning. We used a dataset comprised of images of the six walk lights along our route for the user study (shown in Figure 6). The dataset comprised 21,208 images, including 1,928 original images and 19,280 augmented images. Each image represented one of four states of the walk lights: "Walk", "Count down", "Don't walk", and 'Nothing' (*i.e.*, no walk light visible in the image), which were used as labels for the images. We annotated the images by marking bounding boxes around the walk lights and including labels denoting their respective states. To augment the dataset, we applied brightness variation and channel shift to each image [1], resulting in 10 new images per original image (*i.e.*, five from brightness variation and five from channel shift). During the development of the outdoor app, we observed that the walk lights captured in the images were too small to properly resolve visual features in a resized image (416x416) using the YOLO model. To address this, we modified the outdoor app and the VR training app to only process the central part of the image with half its original width and height. In this study, the output of the outdoor app consists of synthesized speech representing the walk light state, with no consideration for the locations of the bounding boxes.

3.2.1 Effect of Yaw on Walk Light Detection. When it comes to camera geometry, PVI need to control yaw, pitch, and roll to orient a camera correctly as shown in Figure 2. Out of these three rotations, pitch and roll affect the presence of the object in the frame, while yaw affects the orientation of the object captured in the image. Therefore, we conducted an experiment to investigate how yaw affects the accuracy of detecting objects. We captured 5650 images of walk lights in three states ("Walk", "Don't walk", "Countdown" as illustrated in Figure 3), taken at different yaw angles at intersections near our lab. Using the outdoor app, we measured the accuracy of detecting walk lights. We considered the output of the outdoor app for an image as accurate when the label of the bounding box

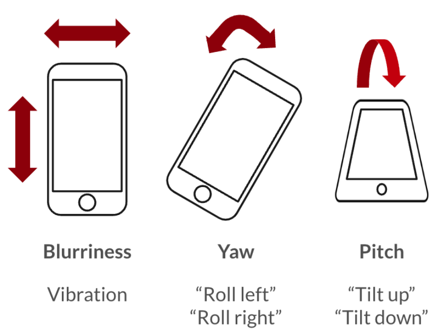

Figure 2: Feedback from the VR training app.

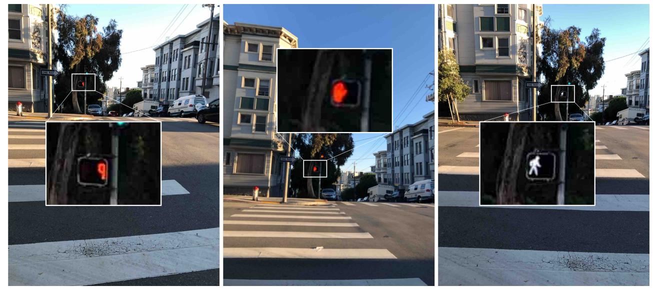

Figure 3: Images with three states of walk lights: Count down, Don't walk, and Walk.

with the highest confidence score matched the state of the walk light in the image, as annotated by a member of our research team, regardless of the bounding box's position. We analyzed the accuracy of detecting walk lights across images with different yaw angles, as shown in Figure 4. Our findings reveal a substantial decrease in accuracy, nearly 20 percent, when the yaw deviates from the straight-up position by more than 10 degrees. These results were employed to determine the appropriate timing for issuing a warning to a user based on the device's yaw orientation.

3.2.2 Effect of Camera Movement Speed on Walk Light Detection. We produced 500 images with varying levels of blur by applying horizontal and vertical motion blur filters to 25 original images of walk lights. These filters are represented by $k \times k$ matrices with ones at $(\lfloor k/2 \rfloor + 1, i)$ and $(i, \lfloor k/2 \rfloor + 1)$, respectively, while all other elements are zero. For each original image, we applied 10 horizontal and 10 vertical filters with varying sizes of 3, 5, 9, 17, 33, 65, 129, 257, 513, and 1025. The levels of blur were measured using the variance of pixels in the edge images generated through Laplacian edge detection [13]. A higher variance value indicates a less blurry image (since high variance is partly caused by the presence of crisp edges with strong image gradients). The edge images ranged from pixel values 0 to 255. The variance values of the 500 images fell between 1.8 and 354.2.

In Figure 5, we present the accuracy of walk light detection with images with different blur levels. We noticed that the accuracy begins to decline when the variance value falls below 180, which is the average variance value of images produced using a filter of size 5. Therefore, we established the threshold of a point movement in an exposure time to 5 pixels for the VR training app. Based on the pinhole camera model, the camera rotation speed corresponding to the threshold is $d\theta = du/f$ where $d\theta$ is the rotation threshold, du is the size of the filter (5 pixels), and f is the focal length (around 1200 pixels in our setup).

4 VR TRAINING APP PROTOTYPE

We developed an interactive training app that allows PVI to practice using a camera for a walk-light detector in a virtual traffic intersection. We built the virtual environment using ARKit in iOS on an iPhone 8, as illustrated in Figure 1. The app does not require additional devices such as head-mounted displays or wearable devices, allowing PVI to interact with the virtual environment using their smartphones. The app provides audio and haptic feedback related to two key interactions: image framing and focusing.

The users can practice using the walk-light detector, which indicates the real-time status of the walk light ("Walk", "Don't Walk", "Countdown"), at the virtual intersection. The virtual walk lights have the same size, distance, and height as those in the real world under typical viewing conditions. We chose the size (33.3 x 30.5 cm) and height (3.05 m) of the virtual walk lights based on design guidelines for pedestrian control features from the US Department of Transportation[3]. The training app simulates a range of distances between the user and the virtual walk light, with the distances based on the widths of roads in the vicinity of our lab, corresponding to the widths of roads near our laboratory. To account for the variability of real-world environments, including factors such as the slope of the road, the user's location, and height, which can affect the relative position of the walk light with respect to the user's camera, we introduced a random offset ranging from 0 to 2 meters to the x- and y- axis positions of the virtual walk light.

We designed the training app to replicate the experiences of using the outdoor app, except that the training app provides feedback on proper camera orientation, which is unavailable in the outdoor app. The training app, like its outdoor counterpart, detects the virtual walk light when it is fully included in the central portion of the image. The training app provides the following feedback (Figure 2):

[3]https://mutcd.fhwa.dot.gov/htm/2009/part4/part4e.htm

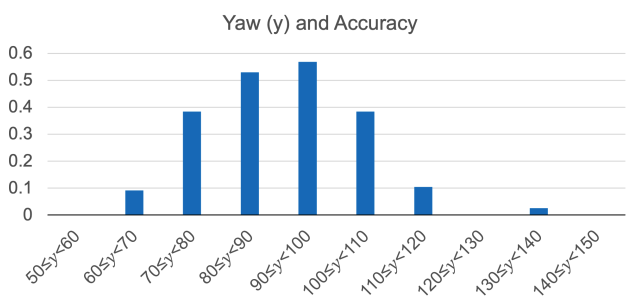

Figure 4: Yaw in degrees (x) and object detection accuracy (y). The smartphone is straight up when the yaw angle is 90 degrees.

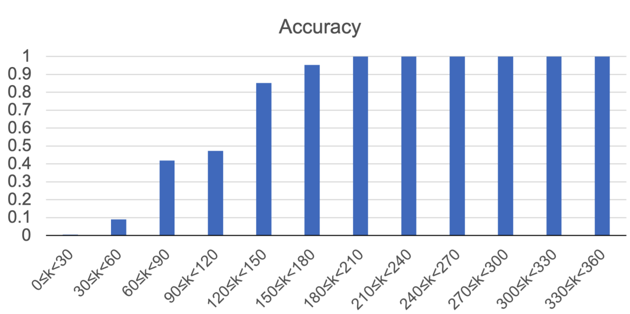

Figure 5: Variance of pixels in edge images (x) and object detection accuracy (y).

- If a walk light is detected, give verbal feedback about the status ("Walk", "Don't Walk", "Countdown")
- If no walk light is detected, say "Nothing"
- Say "Roll left"/"Roll right" if yaw is too far from zero
- Say "Tilt up"/"Tilt down" if pitch is too far from zero
- Issue vibration if the image is blurry
- In each video frame, with 5% probability, the app randomly acts as if a walk light has been detected to simulate the incidence of false positive detection.

Based on the experiments in Section 3.2, we have established the thresholds for the yaw angle of the device and the speed of rotating the device. The app provided a verbal warning whenever the yaw deviated from the upright position by 10 degrees. To address the pitch angle of the device, we established a threshold of $0.5fov$, where fov represents the field of view, to alert users when the virtual walk light is either above or below the central region of the image. The device vibrated when the pixels in the image shifted more than 5 pixels between consecutive frames.

5 USER STUDY

To compare the effect of VR training and TA instructions, we conducted a user study with two groups of PVI: the VR and TA groups. The TA instructions provided comprehensive detail, including guidance on using the walk light detector app and capturing clear, well-framed photos. We employed a between-subjects design to prevent the training effect between conditions. The participants went through the training session indoors and completed a walk-light detection task at the nearby intersections as shown in Figure 1.

5.1 Participants

We recruited 13 participants from our email lists and local organizations. Their ages ranged from 39 to 76 ($M = 53.1, SD = 12.8$) except VR4 who is in her 30s and did not want to reveal her exact age. We randomly assigned the participants to the two groups (*i.e.*, VR and TA). The participants' demographics and backgrounds are specified in Table 1. Nine participants reported that they had never had full vision before. All but one participant (VR5) owned and used iPhone devices. When it comes to participants' experience with camera-based assistive apps, participants used such apps to read texts ($N = 11$), recognize objects ($N = 6$), identify colors ($N = 3$), identify bills ($N = 2$), navigation ($N = 1$), and read barcodes ($N = 1$).

5.2 Procedure

The user study comprised four sections: a background questionnaire, a training session, a walk-light detection task, and a posthoc questionnaire. The participants completed the four sections in 75.1 minutes on average ($SD = 16.2$).

5.2.1 Background Questionnaire. At the beginning of the study, we asked questions about age, gender, and level of vision. To understand the participants' experience in using a camera and camera-based assistive apps, we asked 7-point Likert scale questions (*i.e.*, never, once a month, several times a month, once a week, several times a week, once a day, several times a day) and follow-up open questions regarding their responses; the participants' responses are shown in Table 1. Last, we asked participants to provide their

Table 1: Participants' demographics and experience with photo taking and camera-based assistive apps.

Group	ID	Age	Gender	Level of vision	Onset	Photo taking	Assistive apps
Virtual Reality (VR)	VR1	44	Woman	Legally blind	Birth	Several times a day	Once a day
	VR2	41	Woman	Totally blind	18	Several times a day	Several times a day
	VR3	73	Woman	Totally blind	26	Several times a month	Several times a month
	VR4	3x*	Woman	Legally blind	Since 2016*	Several times a day	Several times a week
	VR5	51	Woman	Totally blind	Birth	Once a week	Once a month
	VR6	61	Woman	Totally blind	1	Never	Never
	VR7	76	Man	Totally blind	13	Once a day	Once a day
Text/Audio (TA)	TA1	49	Man	Legally blind	Birth	Several times a week	Once a week
	TA2	42	Man	Totally blind	3	Several times a week	Several times a week
	TA3	39	Man	Light perception	5	Several times a week	Several times a week
	TA4	66	Man	Light perception	54	Several times a week	Several times a week
	TA5	50	Man	Totally blind	42	Once a week	Once a week
	TA6	45	Woman	Totally blind	1	Several times a day	Several times a week

*VR4 did not want to reveal the exact age.

familiarity with the area near our institute, which can be a factor in participants' performance in the walk-light detection task. Four participants in each group reported not being familiar at all. The remaining participants reported being somewhat familiar due to their previous participation in other user studies.

5.2.2 Training Session. The participants learned the procedure of the walk-light detection task and how to use a camera for walk-light detection during the training session in a conference room. The VR and TA groups obtained the information from VR training and TA instructions, respectively.

Virtual reality training. At the beginning of this session, the experimenter briefly explained that they needed to focus on holding the phone upright and moving the camera slowly as follows:

"Since the walk light detector works with the smartphone camera to recognize a walk light, it's important to hold and aim the camera properly so that it can see the walk light. First, hold the phone upright—this means it's held in portrait mode with the rear camera facing straight ahead. In other words, imagine that you are trying to balance the phone so it's standing up on a tabletop. This will make the walk light visible to the camera. Second, while keeping the phone upright, rotate the camera slowly from left to right when you need to scan the environment to find the walk light. If you move your camera too fast, your camera will be unable to recognize the images it captures."

During the training session, participants completed 15 trials of detecting virtual walk lights using the VR training app. Prior to the first trial, the experimenter provided detailed instructions regarding the virtual environment, including information on the size, height, and distance of the virtual walk light. The experimenter also explained the audio feedback related to the states of the virtual walk lights. During the second trial, the experimenter explained additional feedback related to camera orientation and movement speed. Following these initial trials, participants completed the remaining trials without instructions. Each trial, except for the first two, had a time limit of 2 minutes. Between each trial, participants simulated the scenario of navigating on the street by walking three steps and randomly turning left or right, as directed by the app.

Text/audio instructions. We designed the detailed text and audio instructions to deliver training information equivalent to the VR training. However, participants did not practice detecting virtual walk lights. The instructions were divided into two components: verbal and video. The verbal instructions provided the following information:

- Steps to start and finish a trial of detecting a walk light using the outdoor app
- Guidance on interpreting the verbal feedback from the outdoor app that indicates the states of a walk light
- The factors that may lead to incorrect predictions by the outdoor app, such as weather conditions or the presence of obstructions
- Step-by-step instructions on how to properly capture distant objects with a camera: holding the smartphone upright and scanning the environment slowly for optimal results

To allow participants to go through a complete trial step-by-step and to listen to the feedback from the outdoor app during the trial, we played a 3-minute video that demonstrated a trial step by step with corresponding feedback from the outdoor app.

5.2.3 Walk-Light Detection Task. For the task, participants used the outdoor app in Section 3.2. Following the training session, we provided the participants with an opportunity to practice detecting a walk light in images using the outdoor app indoors. The experimenter sequentially presented three printed images of the "Count down", "Don't Walk", and "Walk" states one at a time (as shown in Figure 3) in front of each participant to trigger the app to recognize and announce these states.

Following the practice session, the participant went outside with the experimenter and a sighted safety monitor to try the outdoor app. They followed a predetermined route near our lab, as shown in Figure 6, and completed six trials. The route was specifically designed to include two trials without a walk light, two trials with a walk light equipped with an accessible pedestrian signal (APS), and two trials with a walk light lacking an APS. In terms of the walking directions, the route consists of three trials where participants turn left or right and three trials where they continue straight ahead. Throughout the task, the experimenter and safety monitor accompanied the participants to ensure the participants' safety.

At the start of each trial, participants were instructed to approach the intersection corner without crossing the road for themselves. Of the 13 participants, 12 used a white cane during the task, while one was accompanied by a guide dog. The participants aimed their smartphones at the walk light when they believed they were close to the intersection corner. They then pressed either the volume up or down button to activate the detector. Once activated, the device indicated the trial number through synthesized speech and provided verbal feedback indicating the state of the walk light (*i.e.*, "Nothing", "Count down", "Don't walk", or "Walk"). Participants were instructed to end the trial when they believed the walk light was in the "Walk" state or when the detector repeatedly announced "Nothing", indicating that there was no walk light present. After each trial, the participants crossed the road with the experimenter and safety monitor to proceed to the next trial location.

5.2.4 Post-task Questionnaire. Once the task was completed, the participants were asked to return to our lab where we conducted a post-task interview to gather feedback on their experience. The interview focused on two themes: the effectiveness of the training session in teaching them how to use the outdoor app, and their overall experience with the training session. To assess their responses, we used a 5-point Likert scale, with response options ranging from "strongly agree" to "strongly disagree".

6 RESULTS

To assess the impact of VR training versus traditional TA instructions, we conducted a comparative analysis of trial completion time and the accuracy of walk light detection. We conducted the Mann-Whitney U test to compare the two groups statistically. To discern the behavioral disparities among participants in camera manipulation, we conducted a detailed analysis of group differences, including camera orientation, speed of camera movement, and the number of well-framed images. Additionally, we gleaned insights from participant feedback, shedding light on their learning experiences with VR training and TA instructions, thus providing valuable context for the observed disparities between the groups.

Prior to conducting the analysis, we manually filtered out image frames that were captured inadvertently by participants (*e.g.*, due to accidental button presses). Our analysis focused on two phases of each trial, scanning and waiting. Scanning refers to the initial phase of a trial, during which participants scan their environment until the walk-light detector detects the walk light for the first time. Specifically, scanning encompasses the period from the beginning of the trial until the first detection of the walk light. Waiting, on the other hand, encompasses the later phase of the trial, from the first detection of the walk light until the end of the trial. When participants fail to detect a walk light during a trial, the entire trial is considered a scanning phase.

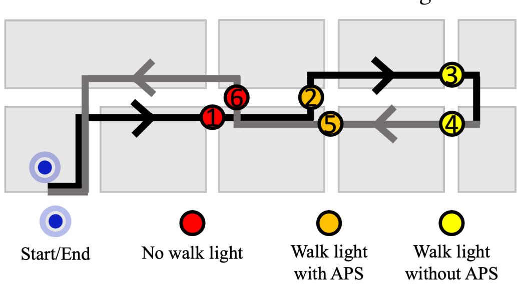

Figure 6: The route where participants completed trials of detecting walk lights. They used the outdoor app at locations marked by circles, with each circle's color indicating the type of walk light. The numbers in the circles are trial indices.

6.1 Accuracy of Walk-Light Detection

To evaluate the accuracy (*i.e.*, f1-score) of the walk-light detectors in our task, we compared manually annotated labels with detected labels from 24,286 images obtained from 13 participants ($M = 1868.15, SD = 969.43$). We assessed the correspondence of the labels without considering the bounding boxes, as both the VR training app and walk-light detector provide labels only and ignore bounding box locations. In this analysis, we removed the extreme outliers, defined as the f1-score less than $Q1 - 3 \cdot IQ$ or greater than $Q3 + 3 \cdot IQ$, where $Q1$ and $Q3$ are the lower and upper quartiles of data distribution, respectively, and IQ is the interquartile range.

Figure 7 presents the f1-score of the VR and TA groups across three phases. Overall, the VR and TA groups exhibited similar f1-scores at 0.99 ($SD = 0.02$) and 0.98 ($SD = 0.02$), respectively. We did not observe a statistically significant difference ($U = 518.0, Z = 0.71, p = .465, r = 0.55$). Specifically, in the scanning phase, the VR group achieved a f-score of 0.98 ($SD = 0.04$) and the TA group achieved 0.98 ($SD = 0.02$). In the waiting phase, the VR group had an accuracy of 0.98 ($SD = 0.02$) while the TA group had an accuracy of 0.96 ($SD = 0.03$). The f-scores in the scanning ($U = 96.0, Z = 1.05, p = .255, r = 0.63$) and waiting phases ($U = 191.5, Z = 1.80, p = .074, r = 0.68$) revealed no significant difference. This suggests that the effect of the VR training and the

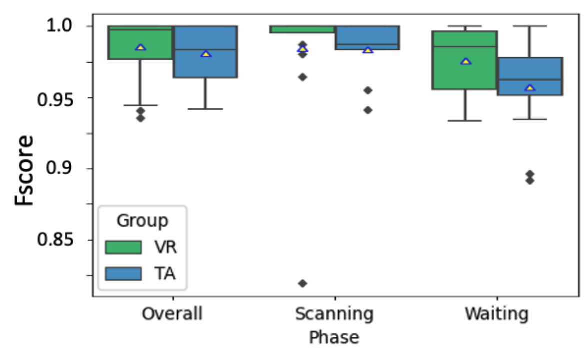

Figure 7: F1-score of the outdoor apps.

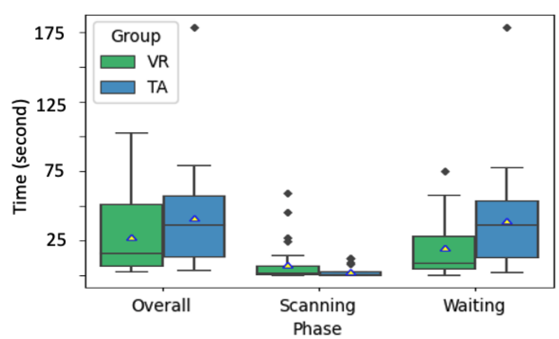

Figure 8: Trial completion time of Trial 2-5.

TA instructions were similar in terms of improving the accuracy of walk light detection.

6.2 Trial Completion Time

In Figure 8, we can observe the completion time of Trials 2-5, during which intersections had walk lights. The VR and TA group completed the trial in 27.3 seconds ($SD = 29.5$) and 41.2 seconds ($SD = 37.9$), respectively. The statistical analysis revealed no significant difference between the groups ($U = 238.0, Z = -1.80, p = .073, r = 0.35$). The average scanning time in the VR group is higher ($M = 7.6s, SD = 14.4$) than the TA group ($M = 2.2s, SD = 3.8$) although this difference did not reach statistical significance ($U = 425.5, Z = 1.64, p = .097, r = 0.63$). A plausible explanation for this difference is that the participants in the VR group moved the camera more slowly when scanning their surroundings as shown in Section 6.3. The VR group had a shorter waiting time ($M = 19.8s, SD = 22.0$) compared to the TA group ($M = 38.9s, SD = 38.1$) with statistically significant difference ($U = 211.0, Z = -2.29, p = .022, r = 0.31$). This implies that participants in the VR group exhibited better maintenance of camera orientation and retention of the walk signal compared to the TA group, following the initial capture of the walk light state.

6.3 Speed of Camera Movement

We used the threshold for the speed of camera movement from our VR training app to assess the percentage of image frames with a suitable camera speed, specifically below the established threshold. The VR group had a higher overall percentage of suitable camera speed at 49.1% ($SD = 32.5$) compared to the TA group at 46.0% ($SD = 30.6$). The difference was more pronounced during the scanning phase, where the VR group had 31.3% suitable camera speed images and the TA group only had 18.6%. In contrast, during the waiting phase, camera movement was limited, resulting in a higher percentage of slow camera speed. The VR group had 70.1% ($SD = 30.4$) and the TA group had 60.1% ($SD = 28.1$) in the waiting phase.

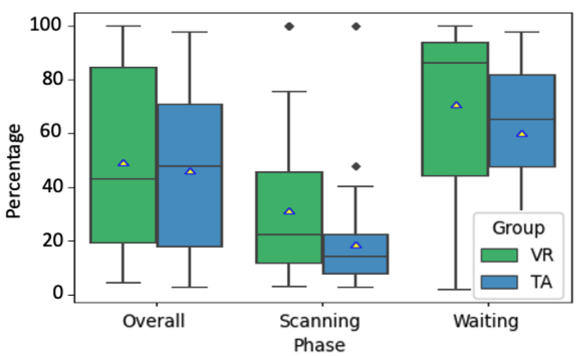

Figure 9: Images with camera speed below the threshold.

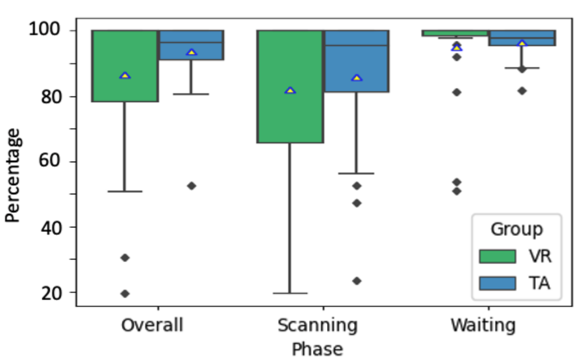

Figure 10: Images with appropriate camera orientation.

To assess the impact of camera movement on image quality, we measured image blurriness by analyzing the variance of pixels in the edge images generated through Laplacian edge detection [13]. Higher variance values indicate less blurriness. Pixel values in the edge images ranged from 0 to 255. We compared two groups, the VR group and the TA group, and found that the VR group had a higher variance of 183.5 ($SD = 25.9$) than the TA group with a variance of 163.5 ($SD = 20.0$). Additionally, when we analyzed images captured above and below a pre-defined threshold camera speed, we found that the variance was 183.5 ($SD = 74.8$) when the speed was above the threshold, and 173.9 ($SD = 87.3$) when the speed was below the threshold. Our results support our expectation that slower camera movement results in clearer images.

6.4 Camera Orientation

We investigated participants' ability to maintain proper camera orientation while capturing a walk light with a smartphone. We used the pitch and yaw thresholds in a VR training app to assess the percentage of image frames with correct camera orientation. The TA group had a higher percentage of properly oriented images (93.6%, $SD = 8.8$) than the VR group (86.5%, $SD = 22.0$). The difference in camera orientation performance between the VR and TA groups may have been due to the offset of the virtual walk light position in the VR training app, which could have led participants in the VR group to experiment with different camera orientations. Additionally, the TA group likely had better camera manipulation skills than the VR group. These issues will be further discussed in the Discussion section. During the scanning phase, both groups had lower percentages of proper orientation (VR : 82.0%, $SD = 26.2$; TA : 85.9%, $SD = 20.1$). However, participants in both groups performed better during the waiting phase (VR : 95.0%, $SD = 13.2$; TA : 96.4%, $SD = 4.7$).

We measured the proportion of images with or without a walk light in the frame by manually annotating the images to indicate whether they included a walk light. In the VR and TA groups, 69.5% ($SD = 31.5$) and 69.8% ($SD = 28.2$) of images contained walk lights, respectively. When the walk light was not visible, it was typically located outside the frame (VR : 29.9%, $SD = 31.3$; TA : 28.9%, $SD = 28.3$). Only a small percentage of images had no walk light due to occlusion: 0.5% ($SD = 1.3$) and 1.3% ($SD = 2.0$) in the VR and TA groups, respectively.

6.5 Subjective Feedback

The subjective responses shed light on both the advantages and disadvantages of VR training and traditional TA instructions. The results of the post-task questionnaire are visualized in Figure 11.

Five participants in the VR group (71.4%) and three participants (50%) in the TA group agreed or strongly agreed that they learned

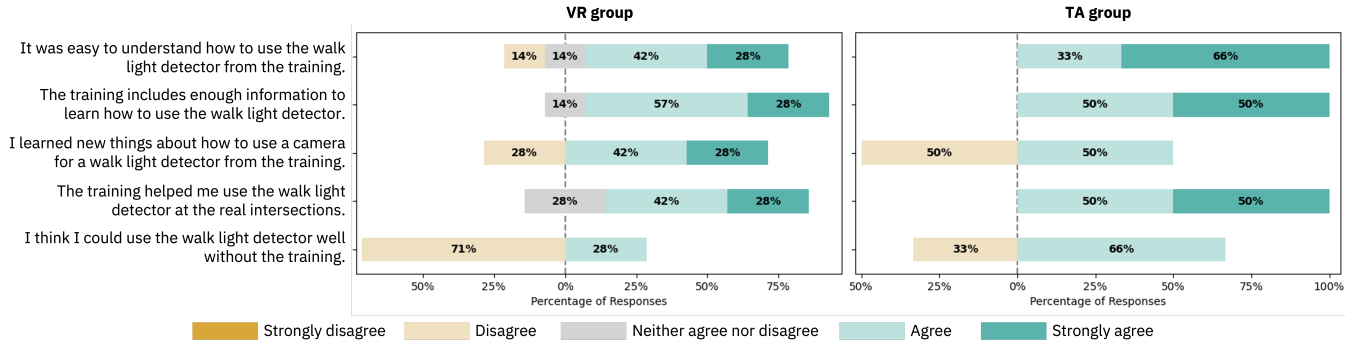

Figure 11: Participants' responses in the post-task questionnaire.

new things about how to use a camera for an outdoor app from the training. Participants in the TA group mentioned that the training reminded them of some concepts for taking good photos. For example, TA1 said *"...I don't think about that perfectly when I'm taking a picture or whatever. ...this helps me remember that I need to make sure that it's in focus, and stuff like that."* The participants in the VR group highlighted that the VR training app was useful for learning how to manipulate a camera. VR3 said *"I learned how to manipulate the camera so that it points in the correct direction."* VR6 also highlighted the opportunity to practice manipulating a camera independently, saying *"Because I have never done it. And so it was interesting to experience it. ...My teacher who was also blind, he was trying to get me to use it. And we had such a hard time, both of us."*

Five participants (71.4%) in the VR group and two in the TA group (33.3%) disagreed that they could use the outdoor app well without the training. VR3 thought that the practice in the training helped her to hold the smartphone in the right way. She stated, *"I wouldn't know how to hold it. ...Without training, I didn't know what that meant."* Meanwhile, VR5 thought that understanding the possibility of mistakes by the outdoor app was useful. She explained, *"I saw that the [training] app makes mistakes. And I saw the real app make mistakes, too. That worries me."* Some participants in the TA group emphasized the importance of hands-on experience, such as that provided by the VR training, as a part of the training process. TA3, who agreed, found the instructions helpful but believed that he would need to practice personally. He commented, *"I think instructions are absolutely necessary for some people. ...I think for me personally, I would have been able to figure it out by just experimenting, just standing there patiently...."* TA6 mentioned *"I know what to do with it. I think it'll be more helpful if there's more instruction on how you should aim it. Like what kind of angle? Like how far do you have to move? Which direction? ..."*

Most participants in both groups found the outdoor app easy to understand after completing the training. However, one participant (VR4) neither agreed nor disagreed and was somewhat confused about how to start and finish a trial using a volume button. Another participant (VR5) disagreed, stating that the task in the VR training was harder than the real walk light detection task. She said *"The training app made one thinks that this job was going to be a lot harder than it is actually."* She pointed out a lack of context in the VR training, elaborating *"When you're in the training, it's in virtual, but it's like you're in black space.... there's no context."* Nevertheless, all participants except VR4 agreed or strongly agreed that the training provided enough information to learn how to use the outdoor app. VR4, who was neutral, found it difficult to learn due to the limited space and obstacles in the training room, noting that *"I think it's a bit confusing because the area is small. ...Most of the times when the app would tell me turn right or left, there were chairs."*

Five in the VR group (71.4%) and six in the TA group (100%) agreed or strongly agreed that the training helped them use the outdoor app at real intersections. However, VR4 and VR5 were neutral due to the limited space of the training room and the lack of context. Meanwhile, VR6, who agreed, pointed out that the VR training app was picky, making the training harder than the actual walk light detection task. She noted that *"it made a lot more vibration when inside and it said a lot more nothing. It was easier [at the real intersections], a bit more difficult to find the light [indoors]."*

7 DISCUSSION

In this section, we will reflect on the key findings of our study and discuss the implications of designing an accessible virtual environment for learning photography. We will also examine the limitations of this work and provide suggestions for future research that can address these limitations and further enhance the impact of the VR training approach.

7.1 Implications

We have identified the advantages of using VR compared to text and audio instructions as well as the implications of designing VR environments for learning. Firstly, it is crucial to provide hands-on experience to resolve the limitations of verbal camera guidance instructions. Secondly, to make the virtual environment effective, it is essential to make it as close to reality as possible. Thirdly, when designing a virtual environment, it is important to consider both the user's background and the properties of the task.

Provide hands-on experience to resolve the limitations of verbal instructions. The quantitative and qualitative results of this study demonstrate the positive impact of VR training on the use of a walk light detector in real intersections. One of the primary benefits of VR training is that it allows participants to experience and practice abstract concepts mentioned in verbal instructions within a virtual environment. For instance, we instructed both the VR and TA groups to move their camera slowly to capture clear photos, and those in the VR group were able to follow the instructions better as they received haptic feedback on their camera movements during the training. Similarly, photography involves several ambiguous concepts that are challenging to describe verbally or require users to practice, such as "scan the area to your left" or "keep the camera still until you hear a sound" (which may lead to unintentional camera movement, even if the user tries to hold it steady). For other applications beyond walk light detection, instructions may include more complex and tricky descriptions like "turn left at 45 degrees" or "move the camera slightly to the right." We anticipate that users can better understand these instructions by experimenting with feedback in virtual environments, as the participants in our study did.

Provide clear contextual information within the virtual environment. Some participants in our user study pointed out that the virtual environment should be more realistic for learning photography and the task more easily. Although we tried to make the virtual environment as close to reality as possible, we did not replicate some real-world cues such as the textures or bumps on the ground and sounds from traffic and pedestrians. In the real world, these cues are important for PVI because they can indicate the distance to the cars as well as the moving directions of cars and people around the user. Due to the absence of these factors, it took time for some participants to understand the exact setup of the virtual environment during the training. Therefore, developers of the VR training app should describe the context clearly including what is missing in the virtual environment.

After the user study, we observed that certain parameters in the virtual environment, such as the probability of errors and offset of the position of walk lights, differed from those in the real-world task. As a result, some participants in the VR group found the training session more challenging than the actual task. Therefore, when building a virtual environment, it is crucial to determine the design parameters by accurately measuring the corresponding parameters in the real world. This ensures that the virtual environment reflects the actual task as closely as possible, enabling users to transfer the acquired skills to the real-world task effectively.

Take into account both the user's experience and the characteristics of the task. In this study, the participants in both groups achieved a high percentage of images with the proper camera orientation overall at around 90%. This can be because adjusting the orientation of the device is easy or because the participants are familiar with using a smartphone and taking photos with it, as many of them reported using a camera or camera-based assistive apps on a weekly basis (as shown in Table 1). When designing the VR training environment, considering the user's background would enable us to pinpoint which specific skills should be the focus of the training. The subjective feedback from the participants also suggested that the effect of VR training may vary depending on the user's background. For instance, VR4, who was taking photos several times a day, disagreed that she learned new things about how to use a camera. In contrast, VR6, who had never taken a photo before, strongly agreed that she learned new things, stating, *"because I have never done it. And so it was interesting to see it and experience it."* Furthermore, the level of difficulty may vary among different systems/tasks. If a walk light detector can leverage the full image from a camera with a wide field of view angle, aiming the camera would be easier with it compared to the outdoor app used in our study, which only utilized the central part of the image for better performance. However, aiming a camera could be more challenging with other form factors of a camera, such as head- or chest-mounted cameras, since users cannot control the orientation of the camera with their hands. Therefore, the effectiveness of VR training may vary depending on factors such as the user's skill level and the characteristics of the task.

7.2 Limitations

In this research, using a between-subject design for the user study served as a strategic choice. One of its notable advantages was its ability to mitigate the training effect, a phenomenon where participants who experience both VR training and traditional TA instruction methods may exhibit biased or confounded responses due to their exposure to multiple training approaches. By allocating distinct groups of participants to either VR training or TA instruction, we ensured that the assessment of each method's effectiveness remained independent and unaffected by carryover effects.

However, it is essential to acknowledge that this design choice also presents a limitation, primarily stemming from the small sample size within each group. With limited participants in each training modality, the findings may have limited applicability to a wider population and limited statistical power of data analysis. Also, the limited sample size may have influenced the balance of skill levels in camera manipulation between the two groups although we randomly assigned the participants to the groups. Future studies could explore larger sample sizes or employ other experimental designs to enhance the generalizability and robustness of our conclusions. Despite this limitation, our study highlights the promising potential of VR training as a valuable tool for enhancing skills in camera-based assistive technology for PVI based on the discernible trends of positive or equivalent effects observed in VR training compared to comprehensive TA instruction methods.

8 CONCLUSION

This study explored the efficacy of VR training compared to traditional text/audio instructions concerning the acquisition of skills in using a walk-light detector at traffic intersections. Through insights from interviews with PVI, we tailored a VR training tool to facilitate effective training. Subsequently, a user study involving 13 PVI participants was conducted, stratified into VR training and TA instruction groups. Analysis encompassing quantitative and qualitative measures revealed nuanced distinctions between the two instructional approaches. While there was no statistically significant difference in the photo quality between the groups, participants' subjective feedback illuminated divergent perceptions. Notably, participants tended to perceive VR training as instrumental in honing camera manipulation skills. Conversely, a larger proportion of TA participants found the training easier to comprehend compared to their VR counterparts. These findings underscore the multifaceted nature of instructional methodologies in facilitating skill acquisition among PVI.

ACKNOWLEDGMENTS

Jonggi Hong is supported by Smith-Kettlewell's C.V. Starr Postdoctoral Fellowship. James M. Coughlan is supported by NIDILRR (grant number: 90 REGE0018).

REFERENCES

[1] Vardan Agarwal. 2020. *Complete Image Augmentation in OpenCV.* https://towardsdatascience.com/complete-image-augmentation-in-opencv-31a6b02694f5
[2] Envision AI. 2023. *Enabling vision for the blind.* https://www.letsenvision.com
[3] Seeing AI. 2023. *Seeing AI App from Microsoft.* https://www.microsoft.com/en-us/ai/seeing-ai
[4] Aira. 2023. *Visual Interpreting – Get Live, On-demand Access to Visual Information.* https://aira.io
[5] Nida Aziz, Tony Stockman, and Rebecca Stewart. 2022. Planning your journey in audio: design and evaluation of auditory route overviews. *ACM Transactions on Accessible Computing* 15, 4 (2022), 1–48.
[6] Jan Balata, Zdenek Mikovec, and Lukas Neoproud. 2015. Blindcamera: Central and golden-ratio composition for blind photographers. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation.* 1–8.
[7] Michael Barnett. 2005. Using virtual reality computer models to support student understanding of astronomical concepts. *Journal of computers in Mathematics and Science Teaching* 24, 4 (2005), 333–356.
[8] Jeffrey P Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010. VizWiz:: LocateIt-enabling blind people to locate objects in their environment. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops.* IEEE, 65–72.
[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[10] Carmen Chai, Bee Theng Lau, and Zheng Pan. 2019. Hungry Cat—a serious game for conveying spatial information to the visually impaired. *Multimodal Technologies and Interaction* 3, 1 (2019), 12.
[11] Athanasios Christopoulos, Nikolaos Pellas, and Mikko-Jussi Laakso. 2020. A learning analytics theoretical framework for STEM education virtual reality applications. *Education Sciences* 10, 11 (2020), 317.
[12] Antonio Cobo, Nancy E Guerrón, Carlos Martín, Francisco del Pozo, and José Javier Serrano. 2017. Differences between blind people's cognitive maps after proximity and distant exploration of virtual environments. *Computers in Human Behavior* 77 (2017), 294–308.
[13] Kanjar De and V Masilamani. 2013. Image sharpness measure for blurred images in frequency domain. *Procedia Engineering* 64 (2013), 149–158.
[14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition.* Ieee, 248–255.
[15] Be My Eyes. 2023. *Be My Eyes - See the world together.* http://www.bemyeyes.org/
[16] Ricardo E Gonzalez Penuela, Paul Vermette, Zihan Yan, Cheng Zhang, Keith Vertanen, and Shiri Azenkot. 2022. Understanding How People with Visual Impairments Take Selfies: Experiences and Challenges. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility.* 1–4.
[17] João Guerreiro, Daisuke Sato, Dragan Ahmetovic, Eshed Ohn-Bar, Kris M Kitani, and Chieko Asakawa. 2020. Virtual navigation for blind people: Transferring route knowledge to the real-World. *International Journal of Human-Computer Studies* 135 (2020), 102369.
[18] C Harrison, PM Dall, PM Grant, MH Granat, TW Maver, and BA Conway. 2000. Development of a wheelchair virtual reality platform for use in evaluating wheelchair access. In *3rd International Conference on Disability, VR and Associated Technologies, Sardinia, Edited by P. Sharkey.*
[19] Hsiu-Mei Huang, Ulrich Rauch, and Shu-Sheng Liaw. 2010. Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education* 55, 3 (2010), 1171–1182.
[20] Masakazu Iwamura, Naoki Hirabayashi, Zheng Cheng, Kazunori Minatani, and Koichi Kise. 2020. VisPhoto: photography for people with visual impairment as post-production of omni-directional camera image. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–9.
[21] 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.* 203–210.
[22] Hannes Kaufmann, Dieter Schmalstieg, and Michael Wagner. 2000. Construct3D: a virtual reality application for mathematics and geometry education. *Education and information technologies* 5 (2000), 263–276.
[23] Orly Lahav. 2022. Virtual Reality Systems as an Orientation Aid for People Who Are Blind to Acquire New Spatial Information. *Sensors* 22, 4 (2022), 1307.
[24] Orly Lahav, Hadas Gedalevitz, Steven Battersby, David Brown, Lindsay Evett, and Patrick Merritt. 2018. Virtual environment navigation with look-around mode to explore new real spaces by people who are blind. *Disability and rehabilitation* 40, 9 (2018), 1072–1084.
[25] Kyungjun Lee, Jonggi Hong, Simone Pimento, Ebrima Jarjue, and Hernisa Kacorri. 2019. Revisiting Blind Photography in the Context of Teachable Object Recognizers. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19).* Association for Computing Machinery, New York, NY, USA, 83–95. https://doi.org/10.1145/3308561.3353799
[26] G Michael Lemole Jr, P Pat Banerjee, Cristian Luciano, Sergey Neckrysh, and Fady T Charbel. 2007. Virtual reality in neurosurgical education: part-task ventriculostomy simulation with dynamic visual and haptic feedback. *Neurosurgery* 61, 1 (2007), 142–149.
[27] Wei-Kai Liou and Chun-Yen Chang. 2018. Virtual reality classroom applied to science education. In *2018 23rd International Scientific-Professional Conference on Information Technology (IT).* IEEE, 1–4.
[28] Eric Nersesian, Adam Spryszynski, and Michael J Lee. 2019. Integration of virtual reality in secondary STEM education. In *2019 IEEE Integrated STEM Education Conference (ISEC).* IEEE, 83–90.
[29] Sebastian Oberdörfer, David Heidrich, and Marc Erich Latoschik. 2019. Usability of gamified knowledge learning in VR and desktop-3D. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–13.
[30] Matthias Oberhauser and Daniel Dreyer. 2017. A virtual reality flight simulator for human factors engineering. *Cognition, Technology & Work* 19 (2017), 263–277.
[31] OKO. 2023. *The AI App for the Blind.* https://www.ayes.ai/
[32] Shanmugam Muruga Palaniappan, Ting Zhang, and Bradley S Duerstock. 2019. Identifying comfort areas in 3D space for persons with upper extremity mobility impairments using virtual reality. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility.* 495–499.
[33] Jocelyn Parong and Richard E Mayer. 2021. Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning* 37, 1 (2021), 226–241.
[34] Siyou Pei, Alexander Chen, Chen Chen, Franklin Mingzhe Li, Megan Fozzard, Hao-Yun Chi, Nadir Weibel, Patrick Carrington, and Yang Zhang. 2023. Embodied Exploration: Facilitating Remote Accessibility Assessment for Wheelchair Users with Virtual Reality. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility.* 1–17.
[35] Johanna Pirker, Johannes Kopf, Alexander Kainz, Andreas Dengel, and Benjamin Buchbauer. 2021. The potential of virtual reality for computer science education-engaging students through immersive visualizations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* IEEE, 297–302.
[36] Joseph Redmon and Ali Farhadi. 2017. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 7263–7271.
[37] Larry D Rosen, Kelly Whaling, L Mark Carrier, Nancy A Cheever, and Jeffrey Rokkum. 2013. The media and technology usage and attitudes scale: An empirical investigation. *Computers in human behavior* 29, 6 (2013), 2501–2511.
[38] TapTapSee. 2023. *Mobile camera application designed specifically for the blind and visually impaired iOS users.* http://www.taptapseeapp.com/
[39] Ender Tekin and James M Coughlan. 2010. A mobile phone application enabling visually impaired users to find and read product barcodes. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010, Vienna, Austria, July14-16, 2010, Proceedings, Part II 12.* Springer, 290–295.
[40] Paraskevi Theodorou, Kleomenis Tsiligkos, Apostolos Meliones, and Costas Filios. 2022. A training smartphone application for the simulation of outdoor blind pedestrian navigation: usability, UX evaluation, sentiment analysis. *Sensors* 23, 1 (2022), 367.
[41] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology.* 161–174.
[42] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility.* 95–102.
[43] Marynel Vázquez and Aaron Steinfeld. 2014. An assisted photography framework to help visually impaired users properly aim a camera. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 5 (2014), 1–29.
[44] Samuel White, Hanjie Ji, and Jeffrey P Bigham. 2010. EasySnap: real-time audio feedback for blind photography. In *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology.* 409–410.
[45] Mihye Won, Mauro Mocerino, Kok-Sing Tang, David F Treagust, and Roy Tasker. 2019. Interactive immersive virtual reality to enhance students' visualisation of complex molecules. In *Research and Practice in Chemistry Education: Advances from the 25th IUPAC International Conference on Chemistry Education 2018.* Springer, 51–64.
[46] 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 *International Conference on Mobile and Ubiquitous Systems: Computing, Networking,*

and Services. Springer, 150–168.
[47] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A face recognition application for people with visual impairments: Understanding use beyond the lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.
[48] Zhengzhe Zhu, Ziyi Liu, Youyou Zhang, Lijun Zhu, Joey Huang, Ana M Villanueva, Xun Qian, Kylie Peppler, and Karthik Ramani. 2023. LearnIoTVR: An End-to-End Virtual Reality Environment Providing Authentic Learning Experiences for Internet of Things. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.