VR Training to Facilitate Blind Photography for Navigation

Jonggi Hong, James M. Coughlan

Smith-Kettlewell Eye Research Institute, California, USA

jonggi.hong@ski.org, coughlan@ski.org

Abstract

Smartphone-based navigation apps allow blind and visually impaired (BVI) people to take images or videos to complete various tasks such as determining a user's location, recognizing objects, and detecting obstacles. The quality of the images and videos significantly affects the performance of these systems, but manipulating a camera to capture clear images with proper framing is a challenging task for BVI users. This research explores the interactions between a camera and BVI users in assistive navigation systems through interviews with BVI participants. We identified the form factors, applications, and challenges in using camera-based navigation systems and designed an interactive training app to improve BVI users' skills in using a camera for navigation. In this paper, we describe a novel virtual environment of the training app and report the preliminary results of a user study with BVI participants.

Keywords

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

Introduction

Acquiring skills to take photos or videos with good quality has become a necessity for blind and low vision (BVI) people as they actively use cameras for various tasks such as saving memories and performing text recognition (Jayant *et al.* 2011). Using assistive mobile apps is also an essential reason for BVI individuals to use a camera. Smartphone-based navigation apps such as NaviLens and Clew enable BVI people to navigate indoors and outdoors independently by providing directional guidance based on images or videos of BVI users' surroundings. OrCam and Envision Glasses, which read a text and recognize objects, provide functions that detect traffic signs or landmarks for BVI users to navigate outdoors independently. Prior studies in computer vision and assistive technology developed camera-based navigation systems for BVI people to detect landmarks (Serrão *et al.*, 2012), objects (Afif *et al.*, 2020), and obstacles (Tapu *et al.*, 2013) or communication systems between BVI users and sighted people who interact with the BVI users through images or videos (Bigham *et al.* 2010).

However, taking photos or videos with good quality is challenging for BVI people due to the difficulty of image framing and focusing (Bigham *et al.* 2010). Prior studies employed non-visual guidance to help BVI individuals manipulate a camera better (Jayant *et al.* 2011, Vazquez *et al.* 2012, Manduchi *et al.* 2014, Lee *et al.* 2019). These interfaces detect an object of interest in users' images with computer vision techniques and provide guidance based on the object's position for better image framing. However, computer vision techniques are still error-prone for various reasons such as mismatches between training and real-world data, quality of images, and challenging light conditions. The audio and haptic guidance can be hard to perceive, distracting, or may raise privacy concerns (Easwara *et al.* 2015). To overcome these limitations, we aim to enable BVI users to learn how to take clear, properly framed images without audio and haptic guidance through interactive training in virtual reality (VR).

As VR holds a great opportunity to allow BVI individuals to explore simulated areas safely with augmented haptic and audio feedback, prior studies leveraged VR to enable BVI people to explore unfamiliar places in a virtual space (*e.g.*, Kunz *et al.*, 2018). To simulate navigation with VR effectively, Kreimeier *et al.* (Kreimeier *et al.*, 2020) evaluated different form factors for virtual navigation such as treadmills and trackers on the ankles. While these approaches employed VR to enable BVI users to have the experience of navigating unfamiliar places, we aim to improve BVI users' skills to interact with a camera for navigation. Based on BVI people's interactions with a camera for navigation identified through interviews, we designed an interactive training system using VR where BVI individuals can try out a simulated walk-light detector in a safe indoor environment with feedback to correct their gestures that may cause poor-quality images. The user study showed that interactive training allows BVI people to manipulate a camera properly when they scan their environments to find a walk light.

Discussion

As a preliminary study, we interviewed BVI participants to draw from their experiences with a camera for navigation when designing an interactive training tool. We report on the intermediate results from a user study that includes a training session followed by a task of using a walk light detector in a real environment through a user study.

Method

We conducted interviews with BVI individuals and used thematic coding to identify the main themes in their responses (Barun and Clarke, 2006). We recruited 10 participants (6 female and 4 male) from our email lists, whose ages ranged from 30 to 75 (*M*=46.4, *SD*=18.3). Seven participants reported being totally blind, two reported having light perception, and one reported being legally blind. Only two participants ever had full vision. Seven participants were congenitally blind. They have had the current level of vision for 35.5 years on average (*SD*=21.4). Nine participants had smartphones and used them for 10.7 years on average (*SD*=4.5). One participant never used a smartphone. All participants have used a camera. Seven of them reported using a camera several times a week or more (Figure 1a).

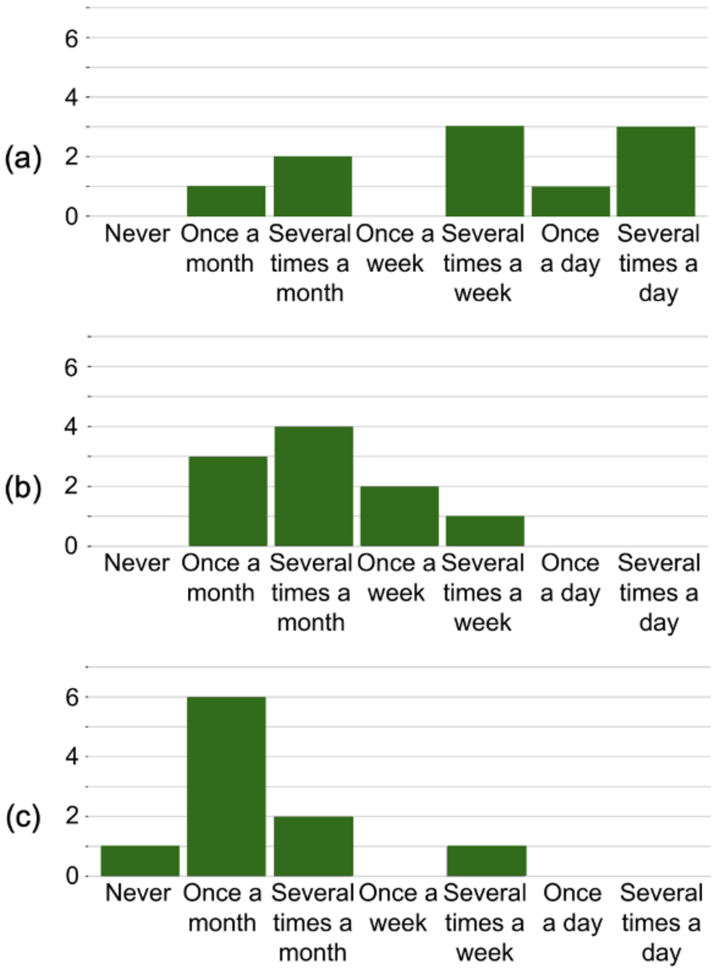


Fig. 1. Blind participants' responses to questions about (a) the frequency of using a camera, (b) the frequency of using navigation systems, and (c) the frequency of using a camera for navigation systems.

The interview includes topics relevant to BVI people's experience in using a camera and camera-based navigation systems: the methods to learn to take photos or videos; frequency of using assistive navigation systems; gestures to use a camera for assistive navigation systems; frequency of using a camera for navigation; apps or tools participants used with a camera for navigation; challenges of using a camera for navigation.

Findings

The most common tasks where the participants used a camera were reading text (N=8), keeping or sharing memories (N=6), video calling (N=5), and identifying objects (N=3). We found only four participants learned or practiced using a camera for these tasks. Two of the participants practiced with the feedback from assistive apps (*e.g.*, sound feedback to indicate the position of a document in the camera frame, whether the app recognizes something or not). The other two participants got feedback from sighted people to learn to take good-quality photos.

All participants reported using navigation systems (both with and without a camera) at least once a month (Figure 1b). They used sensor-based (GPS, compass) navigation systems such as Google Maps and BlindSquare (N=10), apps for video calls with sighted people such as Aira and BeMyEyes (N=7), and computer-vision systems such as OneStep Reader and Google Lookout (N=2). Most participants reported using a camera for navigation once a month (N=6) as shown in Figure 1c. While using a camera, the participants aimed a camera at a target object by following directions from a sighted person in a video call (N=7), maintaining a certain camera height or orientation (N=4), moving the camera around to find a target object (N=1), and touching a target object (N=1). When using a camera for navigation, participants had challenges in image framing (N=7), controlling network connection (N=4), adjusting light conditions (N=3), focusing (N=2), and holding the camera steadily (N=1).

When asked what participants wanted to capture with a camera for navigation, most of them (N=7) wanted to capture landmarks (*e.g.*, stores, restaurants). Other responses were door (*i.e.*, room, N=5), person (N=5), sign (N=5), and walk light or traffic light (N=3). The preferred form factors of a camera for navigation were smart glasses (N=8), a smartphone (N=3), and a shoulder-mounted camera (N=1). The difficulty of aiming a camera affects participants' preference in the form factors. P4 who preferred smart glasses mentioned *“No need to point the camera (on glasses) to a specific place.”* On the other hand, P8 preferred a smartphone because the participant did not want to carry multiple devices, saying *“[...] I don't like glasses because it is cumbersome to have glasses and masks together.”*

Interactive Training App for Using a Walk-Light Detector

Based on the findings in the interview, we designed and implemented an interactive training app that helps BVI individuals get used to interactions with a camera for navigation. While the interview revealed that BVI people are interested in recognizing various objects using their cameras (e.g., door, person, sign, walk light), in this study, we focused on detecting a walk light as a case of using a camera for navigation. Our BVI participants also mentioned aiming a camera at a target object and taking clear photos as the main challenges in using camera-based assistive technology. Therefore, we sought to improve BVI users’ ability to aim a camera at a walk light and take clear photos. We chose the smartphone as a form factor because it is a preferred form factor based on our interview results and because it is the most commonly used device to use a camera with assistive technology.

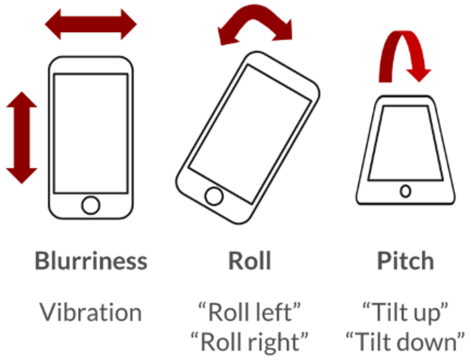
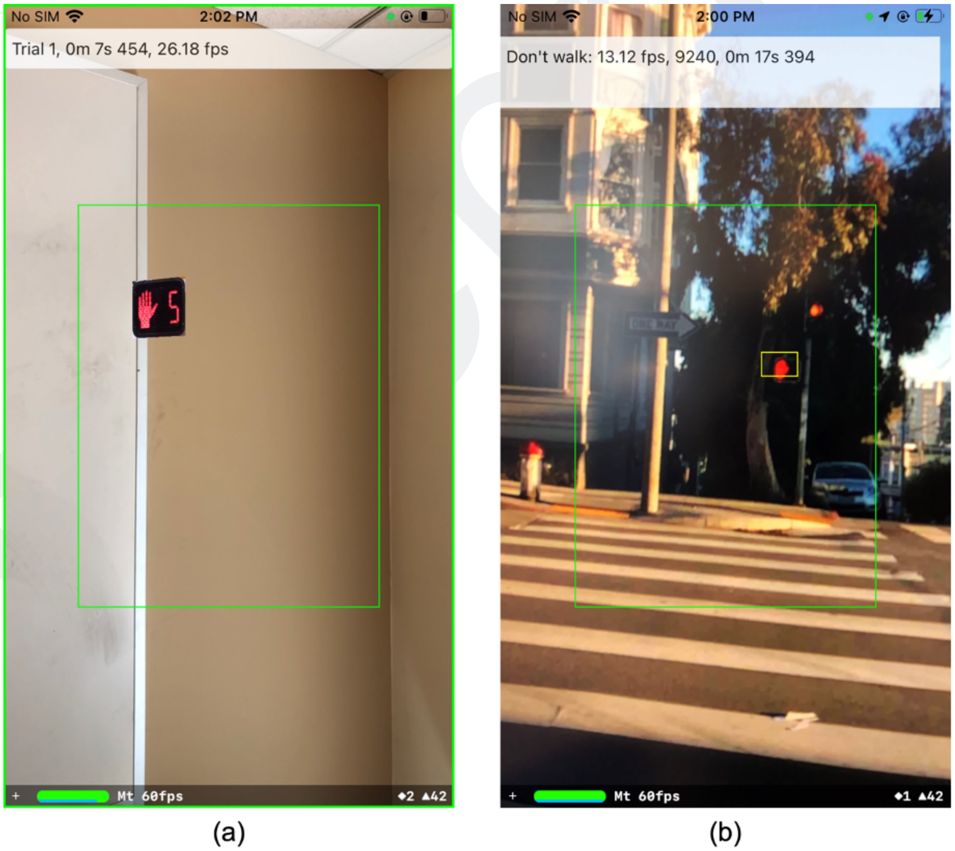


Fig. 2. Descriptions of audio and haptic feedback in the interactive training app.

We built an interactive training app where BVI individuals practice using a camera for a walk-light detector at a virtual traffic intersection. We implemented the virtual environment using ARKit in iOS on an iPhone 8 as shown in Figure 3a. BVI users can interact with the virtual environment with smartphones without additional devices such as head-mounted displays or wearable devices. The app provides audio and haptic feedback to BVI users related to two interactions, image framing and focusing, while they use a walk light detector that indicates the status of the walk light (Walk, Don't walk, Countdown) in real-time. The virtual walk lights have the same size, distance, and height as walk lights in the real world under typical viewing conditions. The size and heights of the virtual walk lights were chosen based on the design guidelines for pedestrian control features from the US Department of Transportation (https://mutcd.fhwa.dot.gov/htm/2009/part4/part4e.htm). The distances to the virtual walk lights were based on the widths of roads around our institution, ranging from 12.47 to 31.24 meters.

This training app is designed to simulate some of the experience of using a real walk light detector app (which we created to function at real traffic intersections), except that the training app's feedback on proper camera orientation is something that is missing from the real walk light detector app. The main function of the training app is to indicate when the virtual walk light is visible to the camera and what state the walk light is in. To allow BVI users to practice taking a clear photo or video for a walk light detector, the app provides audio and haptic feedback as shown in Figure 2. The feedback includes vibration to indicate overly fast camera movement that may cause blurriness of images. The app vibrates when a walk light moves farther than five pixels between two consecutive frames in the camera video stream as we observed that the performance of the walk light detector degrades under this condition due to image blur. It also provides verbal warnings for BVI users to avoid tilting the phone up or down too far from the horizon and rolling the phone to the left or right.



Trial 1, 0m 7s 454, 26.18 fps

Don't walk: 13.12 fps, 9240, 0m 17s 394

(a)

(b)

Fig. 3. Screenshots of (a) the interactive training app and (b) the walk light detector app. The green rectangles in the apps indicate the center area.

Why a virtual environment? Virtual reality has characteristics that enable the creation of an accessible training environment. It allows BVI users to try a navigation system in a safe environment (*e.g.*, a quiet room) before trying it in the real world (*e.g.*, a real traffic intersection). It allows them to have fewer distractions that are likely to occur in the real world (*e.g.*, pedestrians, noises). It also provides our training app with full control of the environment as the app can track the positions of virtual objects and the camera. Last, this approach allows BVI people to access the training environment independently with little help from sighted people because the training environment works in an empty space without installing any physical materials (*e.g.*, markers, beacons) in their environment.

Walk Light Detector for a User Study

To evaluate the effect of using our interactive training app, we will conduct a user study where BVI people will first have a practice session with our interactive training app and will then use the walk light detector in a real environment. For the user study, we implemented a walk light detector using an off-the-shelf object detector, YOLOv2 (Redmon *et al*., 2016) object detection model. We fine-tuned the object detection model using transfer learning with photos of traffic intersections near our institution. A researcher in our team labeled the photos with bounding boxes around walk lights with three different classes: Walk, Count down, and Don't walk. Examples of the photos are shown in Figure 4.

A challenge of detecting a walk light is that the target object appears small under typical viewing conditions (the walk light is about 12-32 meters away in our study), which makes it hard for the object detection model to detect walk lights. To address this challenge, we designed the

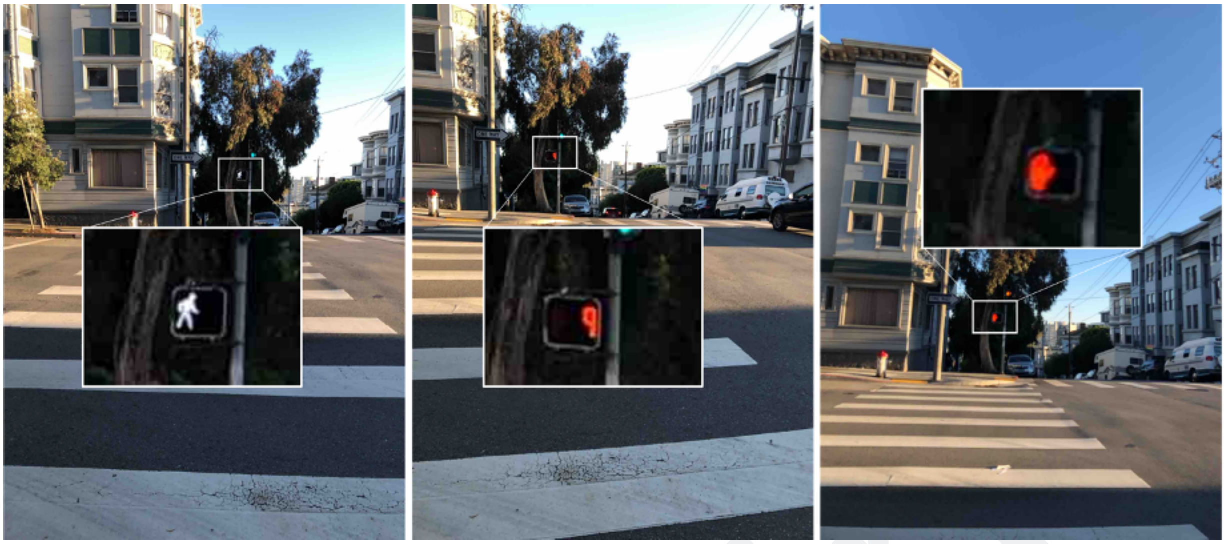


Fig. 4. Samples of the photos used to train a walk light detector model. Photos with "Walk", "Count down", and "Don't walk" signals from left to right.

app to crop the center area of the image as input for the walk light detector and provide feedback about the state of a walk light when it is in the area as shown in Figure 3b.

Results from a Pilot Study with BVI Participants

Before conducting a user study, we piloted a preliminary version of it with two BVI participants. The participants first completed a training session (one used the interactive VR training, and the other listened to verbal instructions with a demo video of the walk light detector). Afterward, both participants then used the walk light detector app outdoors. The main findings of this pilot study were to confirm that (a) the participant who received VR training was able, in multiple trials, to successfully locate the desired walk light (which was located at a 90-degree angle relative to a "distractor" walk light located in a different direction) and to indicate when the light entered the Walk state; and (b) both participants were able to complete the real-world version of this task in an outdoor setting, in which they walked with the experimenters along a path with eight walk lights. Based on our experience with the pilot study, we made a few changes to the user study: (a) since some walk lights are accompanied by accessible pedestrian signals (APS), which provide real-time audio cues about the walk light status, we included different types of intersections, including those with and without walk lights, and walk lights with and without APS; we asked the participants to ascertain if there is a walk light in each trial before providing information about it if it is present; (b) we improved the performance of the walk light detector to decrease the incidence of misrecognition errors.

Evaluating VR Training with a User Study

We increased the number of training images in order to increase the test accuracy of the walk light detector to 98.1%. During the pilot studies, we observed that BVI participants assumed that the intersections have walk lights, though determining whether a walk light exists or not at an intersection is a challenge for BVI users in practice. To reflect the practical usage scenario, we included intersections with no walk light. With these modifications, we conducted a user study with seven BVI participants (4 female and 3 male). Six of the participants ranged in age from 39 to 73 (*M*=38, *SD*=12.7); a participant in her 30s declined to provide her exact age. We randomly assigned the three and four participants to two groups: VR training (VR) and verbal instructions (VI) groups, respectively. All but one participant in each group completed all trials successfully (i.e., detected the Walk state of the walk lights or distinguished intersections with no walk light). A participant (P5, VR) could not capture a walk light at one intersection. The other participant (P7, VI) failed to finish a trial within the time limit (3 min.) due to the misrecognitions of the walk light detector. We observed that the participants in the VR group had proper orientation and camera movement speed while scanning the environment for 22.6% of the scanning time. On the other hand, the participants in the VI group did so only for 15.1% of the scanning time. These results suggest that the VR training method effectively allows BVI individuals to learn how to use a camera for a walk light detector.

Conclusions

In this study, we developed an interactive training app for BVI people where they practice interactions with a camera for a camera-based navigation system. To design the app, we interviewed BVI individuals to understand their challenges in using a camera for navigation. The interviews revealed that BVI people are interested in detecting landmarks and objects such as walk lights for navigation. They have challenges in image framing and taking clear photos with a camera. With these findings, we implemented an interactive training app with VR where BVI individuals can practice using a walk light detector with virtual walk lights. Through a pilot study, we found that interactive training is usable for BVI people, and we also encountered some user study design issues that we resolved for a user study. The intermediate results from the user study revealed that interactive training potentially enables BVI individuals to manipulate a camera better while they scan their environments to find a walk light.

Acknowledgments

JH was supported by Smith-Kettlewell's C.V. Starr Postdoctoral Fellowship. JMC was supported by NIDILRR grant number 90 REGE0018.

Works Cited

Afif, Mouna, et al. "An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation." Neural Processing Letters 51.3 (2020): 2265-2279.

Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, et al. 2010. *Vizwiz: nearly real-time answers to visual questions.* Proc. 23nd annual ACM symposium on User interface software and technology. 333–342.

Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. Qualitative research in psychology 3, 2 (2006), 77–101.

Aarthi Easwara Moorthy and Kim-Phuong L Vu. 2015. Privacy concerns for use of voice activated personal assistant in the public space. International Journal of Human-Computer Interaction 31, 4 (2015), 307–335.

Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P Bigham. 2011. *Supporting blind photography*. Proc. 13th international ACM SIGACCESS conference on Computers and accessibility. 203–210.

Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Kyungjun Lee, Jonggi Hong, Simone Pimento, Ebrima Jarjue, and Hernisa Kacorri. 2019. *Revisiting blind photography in the context of teachable object recognizers.* 21st International ACM SIGACCESS Conference on Computers and Accessibility. 83–95.

Kreimeier, Julian, Pascal Karg, and Timo Götzelmann. "BlindWalkVR: formative insights into blind and visually impaired people's VR locomotion using commercially available approaches." Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2020.

Kunz, Andreas, et al. "Virtual navigation environment for blind and low vision people." International Conference on Computers Helping People with Special Needs. Springer, Cham, 2018.

Roberto Manduchi and James M Coughlan. 2014. *The last meter: blind visual guidance to a target.* Proc. SIGCHI Conference on Human Factors in Computing Systems. 3113–3122.

Serrão, Miguel, et al. "Indoor localization and navigation for blind persons using visual landmarks and a GIS." Procedia Computer Science 14 (2012): 65-73.

Tapu, Ruxandra, Bogdan Mocanu, and Titus Zaharia. "A computer vision system that ensure the autonomous navigation of blind people." 2013 E-Health and Bioengineering Conference (EHB). IEEE, 2013.

Lida Theodorou, Daniela Massiceti, Luisa Zintgraf, Simone Stumpf, Cecily Morrison, Edward Cutrell, Matthew Tobias Harris, and Katja Hofmann. 2021. Disability-First Dataset Creation: Lessons from Constructing a Dataset for Teachable Object Recognition with Blind and Low Vision Data Collectors. 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, USA) (ASSETS '21). Association for Computing Machinery, New York, NY, USA, Article 27, 12 pages.

Marynel Vázquez and Aaron Steinfeld. 2012. *Helping visually impaired users properly aim a camera.* Proc. 14th international ACM SIGACCESS conference on Computers and accessibility. 95–102.