Leveraging Augmented Reality to Create Orientation and Mobility Apps for People with Visual Disabilities

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

ABSTRACT

With the introduction of augmented reality technology to the iOS and Android platforms, mainstream smartphones now have the ability to determine their own motion in 3D space with high accuracy. Here, we present our work leveraging these new capabilities to create two smartphone apps for people with visual disabilities: (1) an app that provides automatic navigation guidance when backtracking along a route and (2) an app that leverages crowdsourcing to provide easy to follow, automatic guidance to objects of interest. Along with a discussion of the design of the apps themselves, we present a preliminary usability study that supports their utility. We conclude with a discussion of the promises and limitations of augmented reality smartphone technology to create assistive apps for people with visual disabilities.

need a connection to orientation and mobility

Probably need to up the claims on novelty here as well.

ACM Classification Keywords

K.4.2. Assistive technology for persons with disabilities: Orientation and mobility tools for people who are blind or visually impaired

Author Keywords

visual disability; orientation and mobility; augmented reality; assistive technology; smartphone apps

INTRODUCTION

For people who are blind or visually-impaired (B/VI), improvements in orientation and mobility (O&M) have been shown to increase economic opportunity as well as psychological well-being. For instance, while only 30% of working-age Americans who are B/VI are employed [3, 27] (compared with 65% of the general population), individuals with better O&M skills have a higher likelihood of being employed [12, 11, 28, 34]. Similarly, while the link between visual-impairment and depression has been well-documented [36, 35, 22, 23], several studies have suggested that it is disability rather than blindness itself that is at the root of this linkage [35, 38]. It was found that one's ability to perform daily-life tasks — such as shopping for basic necessities, a task that is facilitated greatly for individuals with good O&M skills — was *more* predictive of overall life satisfaction than was degree of vision loss [38].

This paragraph feels a bit disconnected. There probably needs to be something to make it more cohesive with the rest of the paper.

Due to the high degree of importance of O&M, there is a long history of assistive technology designed to bolster the O&M skills of people who are B/VI (for an overview of this work refer to the *Background and Related Work* section). Here, we explore new technological developments and trends that have the potential to be utilized to create widely accessible assistive technology for O&M. Firstly, the high rate of ownership of smartphones by people who are B/VI [33]. The availability of these devices has dramatically improved accessibility for people who are B/VI (e.g., through GPS-based directions and OCR software). Secondly, the introduction of high accuracy 3D-tracking capabilities into mainstream smartphones. While the primary purpose of these modules, e.g., Apple's ARKit [25], Google's ARCore [31], is to enable augmented reality applications — whereby virtual and real world content are mingled, e.g., by overlaying images on a smartphone video feed — these modules can be repurposed to create powerful assistive technology for O&M.

need the explicit link to the failure of these technologies

In this document we present our work on leveraging the augmented reality modules in modern smartphones to create assistive technologies to help users who are blind with everyday O&M tasks. In order to design maximally impactful technologies, we employ user-centered design principles throughout our research and development process. Specifically, we worked longitudinally with co-designers who are B/VI and two of the study authors, who are visually impaired themselves, contributed to all aspects of the project and provided design guidance based on their personal experiences.

Our first app, *Clew*, enables users to backtrack along previously traveled routes. This app is designed to alleviate various pain points experienced by non-visual travelers (e.g., finding one's way independently after being led to a location by a sighted guide). The second app, *ViewShare*, utilizes crowdsourcing to enable a user to find objects and receive automated guidance to objects in cluttered environments. In the remainder of the paper we present background and related work on O&M assistive technology, discuss some of the algorithms that underlie the AR technology on modern smartphones, present the design of *Clew* and *ViewShare*, provide preliminary usability study for each app, and finally conclude with a discussion of future challenges and promising directions for smartphone-based AR technology for people who are B/VI.

Need to sell harder in the intro once the story of the pa-

ASSETS'18, October 22–24, 2018, Galway, Ireland

ISBN 123-4567-24-567/08/06...$15.00

DOI: http://dx.doi.org/10.475/123_4

BACKGROUND AND RELATED WORK

Engineers have long sought to use technology to improve the O&M capabilities of people who are B/VI. Most of the early work in this area has focused on the usage of technology to help with obstacle detection and avoidance. For example, researchers have created obstacle-detecting versions of the long cane by instrumenting it with sensors such as lasers [5] and sonar [9]. These sorts of devices have not achieved high adoption in the B/VI community for a number of reasons, including their cost, lack of reliability, and limited new functionality over the white cane [37].

Recently, spurred, in part, by the high ownership rates of smartphones among people who are B/VI, the focus of this work has shifted from obstacle avoidance to problems such as navigation and spatial awareness. For instance, GPS-based navigation apps, such as BlindSquare [32] and Google Maps, allow users who are B/VI to navigate to destinations of interest and learn about nearby landmarks and businesses. The combination of location awareness, as provided by GPS, along with databases of businesses, streets, and points of interest, allow for high-quality assistance to people who are B/VI.

While incredibly useful, GPS-based apps for O&M have a number of shortcomings. Most notably, GPSes in modern smartphones are only accurate to about 10m in ideal conditions (they are even worse in challenging environments such as cities) and do not work indoors. To circumvent these challenges, researchers have pursued roughly two types of approaches. The first is to utilize crowdsourcing, whereby people who are B/VI connect over the internet with a sighted person for on-demand assistance. Examples of this approach include the pioneering VizWiz project [6]), the free BeMyEyes app [24], and the subscription-based Aira system [10] (both BeMyEyes and Aira connect a user who is B/VI with a sighted volunteer through a video chat interface). The second class of approaches combine rudimentary motion estimates derived from inertial sensors (gyroscopes and accelerometers) with sensing of environmental infrastructure (e.g., Bluetooth-enabled location beacons or Wifi access points) to enable relatively accurate indoor positioning. For instance, [18, 17, 19] developed a system for navigating indoor environments using smartphone-detectable RFID tags. Dias and her collaborators utilized WiFi fingerprinting (calibrated using a highly-precise indoor mapping robot) and dead-reckoning (using the accelerometers and gyroscopes on a standard smartphone) for indoor navigation [14]. A similar system uses low-energy Bluetooth beacons instead of WiFi access points as landmarks [26, 1, 2]. A third approach utilizes a calibrated database containing images at known locations, which are then matched to camera images at run time [4].

figure out a place to integrate some of Aaron Steinfeld's work

AUGMENTED REALITY

Recently, both Apple and Google have released support on their smartphone platforms for highly sophisticated augmented reality (AR) experiences. In an AR app, virtual and real content are combined, most commonly by overlaying virtual objects or characters on a smartphone's video feed. For instance, Figure 1 shows a virtual cat projected into a real world scene. As the user moves around in space, the phone senses the user's motion with high accuracy and rerenders the image of the cat at an appropriate distance and viewing angle, providing the illusion that the cat exists in the physical world.

Figure 1. A virtual cat overlaid on a smartphone camera feed. The phone is able to accurately sense its movement in order to render the cat at the appropriate viewing angle and depth.

These AR systems are made possible by 3D spatial-processing algorithms that are vastly more accurate than the inertial-based systems utilized in pervious assistive O&M apps. The high accuracy of these systems has been driven by two key trends: the development of sophisticated algorithms for visual-inertial odometry (VIO) [30, 29, 8, 16] (which combine optical tracking with inertial sensing to perform motion estimation) and the development of special purpose hardware that allows these computationally intensive algorithms to run on a smartphone with minimal heat generation and power consumption. The high-degree of motion estimation accuracy enabled by these systems, unlocks many new possible O&M apps for people who are B/VI, which we will described later.

Algorithms for Visual Inertial Odometry

In order to understand the potential of VIO for creating assistive O&M applications, it helps to understand a bit about how these algorithms work. A full explanation of VIO is beyond the scope of this document (for a more comprehensive treatment consult [20]). Approaches for VIO have been developed for both the stereo and the monocular (or single camera) settings. The monocular setting is directly applicable to most smartphones, which rarely have cameras configured as a stereo pair. For the remainder of this paper VIO should be understood to be referring to the monocular setting.

VIO algorithms utilize sensor fusion to blend motion estimates generated by a camera and an IMU. An estimate of the motion of a camera can be made by tracking salient visual features (for instance, corners or other highly textured portions of the image) over the course of multiple frames. Utilizing the mathematics of perspective geometry, one can estimate the rotation and translation of the camera based on the global pattern of movement of these visual features [21]. Of particular interest to the creation of assistive apps-based on this technology, is that the accuracy of these motion estimates is highly dependent on tracking a large number of visual features. Further, for optimal performance these visual features should correspond to points at a range of depths from the camera and be distributed uniformly over the image. The flip side of this is that VIO is susceptible to inaccurate motion tracking when few visual features are tracked or when the tracked visual features are impoverished (e.g., all at the same depth or in the same region of the image). While some environments are more difficult for visual tracking, in some cases users may hold their phones in a suboptimal orientation (e.g., with the camera facing the ground).

Assuming a suitable set of visual features is tracked, the motion estimates of the translation of the camera are only determined up to an arbitrary scale factor. This problem is known as scale indeterminacy. This indeterminacy arises due to the fact that the depths, defined as the perpendicular distance of a point from the image plane, of the 3D points corresponding to the tracked visual features are unknown [21]. For example, for any estimate of the translation of the phone, it is equally valid to say that the phone moved twice as far and the depths of the tracked points were all twice as great. The shortcomings of motion estimates from optical tracking, scale-indeterminacy and inaccurate performance in feature-poor environments, can be overcome (to some degree) by the fusion of inertial sensing data (gyroscopes and accelerometers). Gyroscopes, which provide accurate estimates of angular velocity over short timescales, can be used to refine the estimate of rotation generated by visual tracking and accelerometer data can be integrated to obtain an estimate of linear velocity to overcome the scale-indeterminacy problem. Since inertial sensors don't function optimally when subjected to extremely fast rotations or high accelerations, users need to hold their phones relatively stable to get the full benefits of VIO.

VIO in Mass-Market Smartphones

Both Apple and Google have released AR modules based on VIO. Unfortunately, the precise details of the algorithms employed by each platform are not publicly available, however, there are important high-level distinctions between these frameworks that assistive technology researchers and application developers should keep in mind.

Google Tango

The first devices based on the Google Tango platform were released by Google's ATAP (Advanced Technology and Projects) division in late 2014. The Tango platform utilizes a wide-angle (fisheye) lens equipped and a global image shutter to enable maximally accurate visual-feature tracking. Further, the platform includes a PrimeSense depth-sensing camera. Two commercial smartphones have been released based on the Google Tango platform: the Lenovo Phab2 Pro and the Asus Zenfone AR. While the tracking capabilities of Tango devices are superior to both ARKit and ARCore (discussed next), the reliance of the platform on special-purpose hardware has severely limited the adoption of the technology. As a result, Google suspended the project in early 2018 [13].

ARKit and ARCore

Apple's ARKit [25] and Google's ARCore [31], both released in 2017, provide AR-capabilities that do not require special purpose cameras. Since these platforms utilize conventional cameras, the richness of visual features available for tracking is not as great as phones based on Google Tango. Based on our anecdotal observation, the narrower field-of-view of the cameras used in these frameworks result in less accurate tracking than Google Tango. Further, since neither of these platforms have depth sensing cameras, the availability of 3D information is limited to objects with special structure (e.g., planar horizontal and vertical surfaces). Despite their drawbacks, importantly these frameworks are capable of running on a wider share of phones than Google Tango. Further, given the high degree of preference for iOS devices among people who are B/VI [33], ARKit is the primary platform of interest for researchers seeking to develop assistive apps based on smartphone AR technology.

Although quite specific and certainly at the implementation detail level, a challenge presented when developing assistive apps for ARKit is that the ARSCNView class provided by Apple does not allow for the presentation of subviews. This limitation makes it particular hard to achieve optimal integration with Apple's VoiceOver. In particular, it is very difficult to make an app that consistently announces the identity of activated buttons, while also properly announcing new controls that appears on the screen.

USER-CENTERED DESIGN PROCESS

Our lab has been working to develop O&M assistive technology for the last several years. We take a user-centered approach in which we deeply engage with people who are B/VI to understand their needs, wants, and values as well as any pain points and areas of opportunity that exist within their daily lives. As such, we've used a mixture of quantitative and qualitative methods to arrive at problem spaces along with ideas for assistive technologies to help with O&M.

might want to put this later. The paper could run the risk of dragging at this point

Identified Areas of Opportunity

Based on background research, interviews, and observation we identified several areas of opportunity for the development of assistive apps for people who are B/VI. The theme of high precision indoor navigation came up repeatedly during this phase of our design process. Further, the notion of being able to browse, e.g. in a grocery store or other public space, was often identified by our participants. Finally, users expressed a desire to more quickly build facility in navigating, independently, through unfamiliar environments.

Usability Factors of VIO Technology

Based on the identified areas of opportunity, along with the availability of VIO on modern smartphones, it was natural to investigate the suitability of these approaches. Since VIO is a sensor-fusion algorithm that relies, in part, on optical tracking, any assistive app based on VIO must requires the user leave their smartphone's camera is unoccluded. In the designs explored in this paper, we assume that the user would hold the phone in one hand while holding their long cane in the other. In some cases the user may require handsfree operation of their smartphone (handsfree operation is an area we are actively researching).

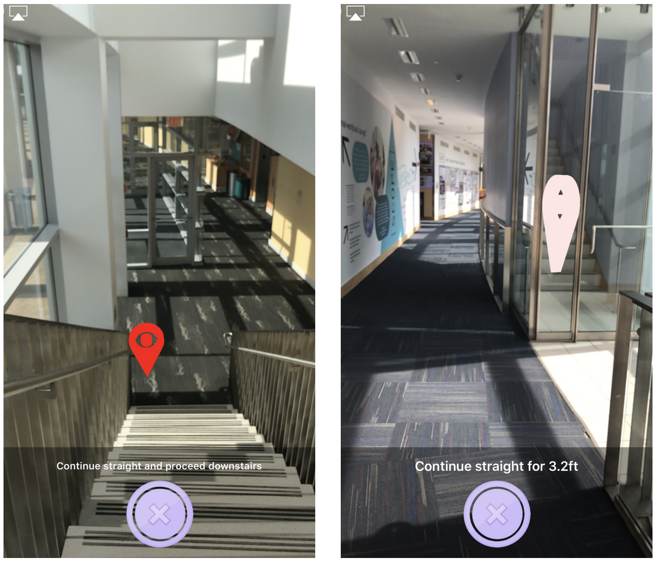

Figure 2. Two screenshots from our app "Clew." Both images show the app in navigation mode where a user is using the app to retrace a route they have previously traveled. The text for the left image says "Continue straight and proceed downstairs" and the text on the right says "Continue straight for 3.2 feet." TODO: the distance seems to be in the wrong units. Should take new screenshots.

Participatory Design

Throughout the development of the apps described in this document, we utilized a participatory approach to design. This manifested itself in two ways. First, we worked with a college student with Retinitis Pigmentosa (a vision disorder that leads to the loss of peripheral vision to the point where someone eventually has little to no functional vision) over the course of five, three-hour long co-design sessions to develop the basic concept and test various prototypes of our *Clew* app. Second, two of the authors of this paper are themselves visually impaired. In addition to their contributions to the design and implementation of the apps, their personal experiences helped guide the design process.

CLEW: AUTOMATIC GUIDANCE ALONG PREVIOUSLY TRAVELED ROUTES

We present our iPhone app, *Clew*, based on the Apple's ARKit platform (Clew is available on the app store, link redacted for anonymous review). Clew provides automatic, turn-by-turn directions to guide users who are B/VI along previously traveled routes.This use case arises, primarily, in the following two scenarios. Firstly, when a user is led to a location in an unfamiliar environment by a sighted guide, they may desire to return to their previous location at a later time — without the need for further assistance from the sighted guide. Secondly, there are examples of places that are easier for someone who is B/VI to navigate out of than to navigate back in to. For instance, it is easier to leave a conference room than it is to find your way back to the particular seat. For both of these uses cases, *Clew* is able to provide high-accuracy, easy-to-follow, navigational guidance to enable people who are B/VI to travel independently indoors. Importantly, Clew requires no environmental modification in order to work (e.g., the introduction of beacons or special signage).

The app works by dropping a trail of virtual breadcrumbs (representing timestamped positions), which can be followed in reverse order at a later time. We call the phase of laying down these virtual breadcrumbs, *path recording* mode. In *path recording* mode, the user holds their smartphone with the screen facing towards them and the camera facing roughly parallel to the ground (recall that ARKit is based on VIO, which relies on optical tracking). Once path recording mode is active, the user travels to a new location (either via their traditional O&M process or via assistance from a sighted guide). Figure 3 shows a sample path along with virtual breadcrumbs.

When the user has recorded their path, they can either pause (described later) or navigate back along their route. When navigating back along the route, the trail of breadcrumbs is first processed by the Douglas-Ramer-Peucker (DPR) algorithm [15] for path simplification. This algorithm winnows down the path, represented by a discrete number of breadcrumbs, by removing sequences of breadcrumbs that can be faithfully represented by a single straight line. Figure 3 shows the breadcrumbs selected for by the DPR algorithm and the resultant straight line path obtained by navigating between the selected breadcrumbs (we refer to the selected breadcrumbs as *waypoints*). In *navigation mode*, the app synthesizes directions to the next waypoint using one of three mechanisms: (1) speech (e.g., "continue straight for 10 feet", (2) haptic feedback or (3) an audible beep when the phone is pointing, approximately, towards the next waypoint. The latency of the update of the iPhone's position is so low that when using feedback mechanism (2) or (3), the user can rapidly sweep their phone back and forth until they feel the haptic feedback or hear the auditory cue. Even with one sweep it is possible to sense the direction to the next waypoint with high accuracy. Screenshots of the app in navigation mode are show in Figure 2.

As the user navigates, the app continuously checks to see if the user has reached the next waypoint. We define the condition of "reached the next waypoint" as the user entering a *waypoint checkoff zone* (see Figure 3). Instead of making these checkoff zones spherical, we chose to make them rectangular prisms where the side length for the dimensions perpendicular to the direction of travel from the previous waypoint are greater than for dimensions parallel to the direction of travel. This choice of checkoff zone shape enables the user to deviate laterally from the intended path without missing a waypoint. The app also handles the user going up or down stairs by specially marking waypoints as belonging to a flight of stairs if the vertical angle of the segment connecting two waypoints exceeds a threshold.

Pause feature

In order for an ARKit tracking session to function properly, the user must not switch to a different app, lock the phone, or block the camera. This can be a hindrance in situations where the user wants to wait a significant amount of time between traveling to a destination and navigating back, when the user would like to use some other app on their phone, or when the user needs both of their hands to perform some task. ARKit does support a pause feature for a tracking session,

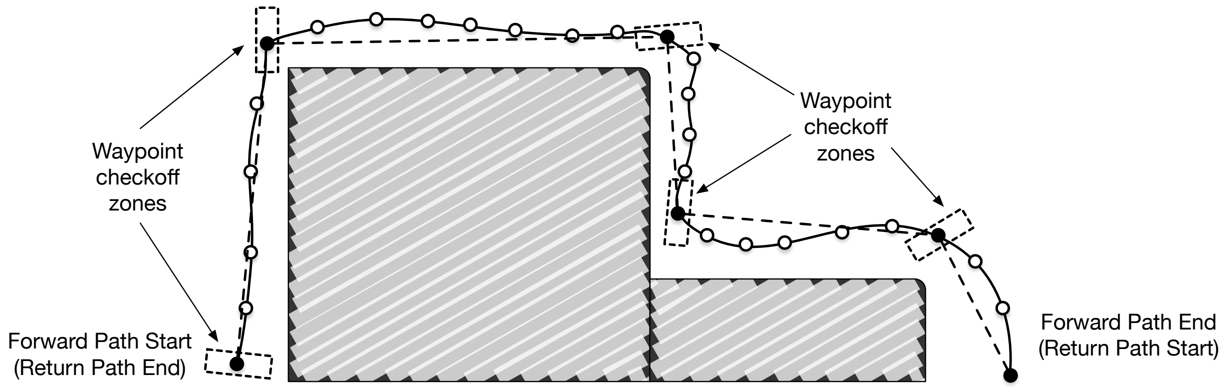

Figure 3. A topdown view of a sample path generated by the *Clew* app. The raw path shown as a solid line is sampled, in time, as a trail of virtual breadcrumbs (white circles with black outlines). The Ramer-Douglas-Peucker algorithm is applied to reduce the breadcrumbs to a more manageable number of waypoints (solid black circles). The resultant path consists of straight lines connecting these waypoints in reverse time order (since the user will be navigating the path in reverse). Waypoint checkoff zones (dashed boxes) are create to be wide in dimensions perpendicular to the direction of travel and narrow in directions parallel to the direction of travel. The light gray area filled with diagonal lines denotes a region of the space that is impassible.

however, when the session is unpaused the phone's position, as reported by ARKit, will remain the same as before it was paused. Given this limitation, we designed our pause feature to prompt the user to place the phone in a position and orientation before pausing that is easy to return to when the user wants to unpause the app. Features such as walls or doors work well for this purpose. It is not as crucial that the user is able to exactly recreate the position and orientation at pause time. Specifically, the most important aspect of the pose to recreate is its yaw (see Figure TODO). The reason for this is that the phone's accelerometer can easily determine the other two degrees of freedom of rotation (as these are not perpendicular to gravity) and deviations in the phone's position cause a fixed amount of error, whereas an error in the phone's yaw will be magnified the farther the user travels (see Figure TODO).

Usability testing

TBD.

Future Work on the Clew App

As stated earlier the basic version of the Clew app is already in the app store. Beyond engaging with users deeply to understand how they experience the current version of the app and improving the usability based on their feedback, there are two major areas of future work we intend to explore. The first is to improve the ability for the app to detect and communicate tracking failures. When ARKit throws an error message, it is straightforward to communicate this information to the user, however, if the tracking drifts or becomes inaccurate without entering an error condition the user will be in a much more difficult situation. In future versions of Clew we would like to be able to detect degraded tracking and alert the user. A second area of future work would be to enable sharing of previously traveled routes with other Clew users. This feature could work very similarly to the pause functionality where the user puts their phone in a known starting pose (e.g., against a door as they enter a building) which allows the phone to register its current location with the previously saved route.

VIEWSHARE: OBJECT FINDING IN CHALLENGING ENVIRONMENTS

Object-finding in unfamiliar environments can be very challenging for people who are B/VI . Researchers have attempted to build systems to enable people who are B/VI to more easily find objects whose locations are unknown (either because the location was never known to begin with or because the objects were moved by a third party). Inspired by previous work [7] that utilized a crowdsourcing model, whereby sighted online volunteers labeled an object of interest and a mobile phone automatically provided sonic cues to guide a user to the object, we developed the *ViewShare* app.

citation needed

The ViewShare app employs the following sequence of steps to provide high-precision, automatic guidance to objects of interest. First, the user, who is B/VI, launches the ViewShare app, which starts an ARKit tracking session. Once the user is ready to find an object, they announce it, the phone recognizes their voice command using Apple's Siri kit, and a localization job is created. This localization job is assigned to a number of sighted volunteers who are alerted via push notifications on their smartphones. On the user's device, every two seconds an image is captured and added to the localization job. Once the volunteer clicks on the relevant push notification they see the collected images. Once the locate the object of interest, the volunteer taps its position on their screen. The identified location, represented as a 2D pixel coordinates, is relayed back to the user's app, which then attempts to convert this 2D pixel coordinate into a 3D position using the procedures outlined in the sub section below on *2D to 3D Transformation*. If a 3D location can be determined, the user is provided with automatic guidance to the object in the form of computer-generated speech and haptic feedback.

Object Localization Feedback

The ViewShare app generates a haptic cue when the phone is oriented towards a localized object. As mentioned in our discussion of Clew, the low latency of this mechanism allows the user to sweep the phone to easily find objects. Further,

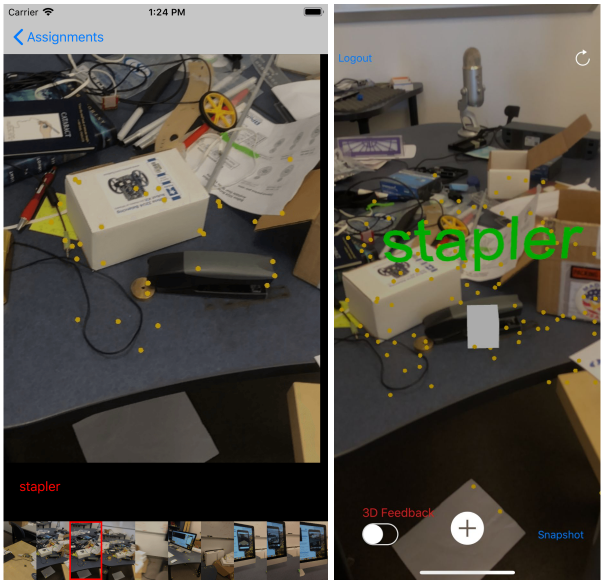

Figure 4. Left image: the ViewShare interface for the crowdworker. The crowdworker is instructed to locate the stapler in the series of images collected from the B/VI user's phone. Right image: the interface for the B/VI user. This interface shows a visualization of the location of the stapler as selected by the crowdworker. While the visualization is not useful for user's who are blind, it is potentially useful for people who with low vision. All elements of the interface are accessible via Apple's VoiceOver technology.

the app synthesizes speech feedback to announce both the name of the object and its distance from the user. The app can provide feedback in two modes. The first mode is 2D position, whereby notions of distance and orientation towards an object are determined by projecting the 3D spatial information into the floor plane. This mode is useful for navigating to objects that are far away (e.g., farther than 10 feet) where the height of the object is irrelevant until the user gets close. The second mode is 3D position, whereby notions of distance as well as whether or not the phone is oriented towards an object are determined by using the unmodified 3D positions and orientations of the phone and localized objects. This mode is appropriate for finding objects that are close by and are relatively small in terms of their vertical extant. For example, it might be useful to use 3D mode to find a pen, whereas finding a door would be best approached using 2D.

2D to 3D Transformation

In order to map a 2D pixel coordinate to a 3D point in space we utilize ARKit's built-in functionality for plane fitting. We test the ray corresponding to the 2D pixel coordinate to see if it intersects a 3D plane (as determined by ARKit). TODO a figure would be good here. ARKit version 1.5, which we utilize for this study, is capable of finding both vertical and horizontal planes. If the ray intersects a plane, we mark the object as localized and compute its 3D position using straightforward geometric formulas. If the ray *does not* intersect a known plane (either because the object located on a plane or because the plane has not been found by ARKit) we return an error condition to the volunteer. In this situation, the volunteer must click on the object again from a second viewpoint. Once the 2D pixel coordinate of the object is provided in two views, we use triangulation to compute its 3D position. The optimal 3D triangulation of two rays with endpoints x_1 and x_2 and unit directions u_1 and u_2 is given by the following equation.

$$\text{triangulate}(x_1,u_1,x_2,u_2) = \frac{x_1+x_2}{2} + \frac{u_1u_1^\top u_2u_2^\top + u_1u_1^\top + u_2u_1^\top u_2u_1^\top - u_2u_2^\top}{2\left(1-(u_1^\top u_2)^2\right)}(x_2-x_1)$$

Usability Study

TBD

Future Work on the ViewShare App

While we designed this app primarily for finding objects, it can also potentially be utilized for other tasks. One area of opportunity is finding things that are not objects (e.g., the precise location of a bus stop or a doorway in outdoor settings where GPS is not sufficiently accurate). A second area of opportunity is automatically providing the location of relevant features in the environment (e.g., the locations of doors, windows, chairs) so that the user can make a mental map of the space around them. A third area of exploration is quality control. Aggregating judgments from multiple crowdworkers is a very active area of research in the machine learning communityand applying some of these techniques to figure out whom to trust and how to filter out inaccurate localizations from the crowd would be immensely helpful.

need citation

DISCUSSION AND FUTURE WORK

The development of *Clew* and *ViewShare* provide useful information as to the potential of AR-based assistive apps. The results demonstrate that for object finding — within a room or on a tabletop — and navigation along relatively short routes — Xm — the 3D motion estimation abilities of Apple's ARKit are sufficient for good performance. As the complexity of the environment increase (either because it is larger or has few easily trackable features, which are essential for VIO to function well) other sources of information must be brought to bear on the problem. For instance, in such cases external landmarks (e.g., signs, such as AprilTags, that are easily detectable in a phone's camera feed or Bluetooth Beacons) could be used to reduce tracking drift and provide anchors in the space. A second method for adding robustness to AR-based apps would be to employ crowdsourcing to correct drift in the phone's motion estimates. For example, a crowdworker might select the location of a particular landmarks in images collected at two points in time so that drift in the system could be reduced.

CONCLUSION

We have presented two smartphone apps that each allow people who are B/VI to perform significant, new tasks with their smartphones. In contrast to the typical use cases of indoor navigation and image process where smartphones have provided significant value for users who are B/VI, our apps provide some of the first apps that can be used for navigation and object finding in arbitrary indoor environments. While the initial results of our usability test are promising, much more work is required to refine the developed apps to be maximally useful to the B/VI community.

Further, we have outlined several promising areas of opportunity for the development of new augmented reality-enabled apps to support people who are B/VI. We have also provided a discussion of the limitations of this technology. With a combination of the development of new algorithms, careful co-design with users who are B/VI, and the improvement of the underlying AR capabilities from smartphone vendors these limitations can hopefully be overcome to create impactful new assistive smartphone technology for people who are B/VI.

ACKNOWLEDGMENTS

Removed for anonymous review.

REFERENCES

1. Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris M Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: a navigational cognitive assistant for the blind.. In *MobileHCI*. 90–99.

2. Dragan Ahmetovic, Masayuki Murata, Cole Gleason, Erin Brady, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. 2017. Achieving Practical and Accurate Indoor Navigation for People with Visual Impairments. *Web for All* (2017).

3. American Federation for the Blind. 2017. Interpreting Bureau of Labor Statistics Employment Data. (January 2017). http://www.afb.org/info/blindness-statistics/interpreting-bls-employment-data/24

4. Yicheng Bai. 2014. *A wearable indoor navigation system for blind and visually impaired individuals*. Ph.D. Dissertation. University of Pittsburgh.

5. J Malvern Benjamin. 1973. The new C-5 laser cane for the blind. In *Proc. Carnahan Conf. on Electronic Prosthetics*. 77–82.

6. Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and others. 2010a. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 333–342.

7. Jeffrey P Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010b. VizWiz:: LocateIt-enabling blind people to locate objects in their environment. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 65–72.

8. Michael Bloesch, Sammy Omari, Marco Hutter, and Roland Siegwart. 2015. Robust visual inertial odometry using a direct EKF-based approach. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 298–304.

9. Johann Borenstein and Iwan Ulrich. 1997. The guidecane-a computerized travel aid for the active guidance of blind pedestrians. In *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, Vol. 2. IEEE, 1283–1288.

10. Aira Tech Corp. 2018. aira: your life, your schedule, right now. https://aira.io/. (2018).

11. Adele Crudden and Lynn W McBroom. 1999. Barriers to employment: A survey of employed persons who are visually impaired. *Journal of Visual Impairment and Blindness* 93 (1999), 341–350.

12. Adele Crudden, Lynn W McBroom, Amy L Skinner, and J Elton Moore. 1998. Comprehensive Examination of Barriers to Employment among Persons Who Are Blind or Visually Impaired. *Mississippi State: Rehabilitation Research and Training Center on Blindness and Low Vision, University of Mississippi.* (1998).

13. Tech Crunch. 2017. Google retires the Tango brand as its smartphone AR ambitions move wider. https://techcrunch.com/2017/08/29/google-retires-the-tango-brand-as-its-smartphone-ar-ambitions-mov (August 2017).

14. M Bernardine Dias . 2014. *NavPal: Technology Solutions for Enhancing Urban Navigation*. Technical Report CMU-RI-TR-21. Robotics Institute, Pittsburgh, PA.

15. David H Douglas and Thomas K Peucker. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization* 10, 2 (1973), 112–122.

16. Christian Forster, Matia Pizzoli, and Davide Scaramuzza. 2014. SVO: Fast semi-direct monocular visual odometry. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 15–22.

17. Aura Ganz, Siddhesh Rajan Gandhi, James Schafer, Tushar Singh, Elaine Puleo, Gary Mullett, and Carole Wilson. 2011. PERCEPT: Indoor navigation for the blind and visually impaired. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 856–859.

18. Aura Ganz, James M Schafer, Yang Tao, Larry Haile, Charlene Sanderson, Carole Wilson, and Meg Robertson. 2015. PERCEPT based interactive wayfinding for visually impaired users in subways. *Journal on Technology & Persons with Disabilities* 3, 22 (2015).

19. Aura Ganz, James M Schafer, Yang Tao, Carole Wilson, and Meg Robertson. 2014. PERCEPT-II: Smartphone based indoor navigation system for the blind. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 3662–3665.

20. Jianjun Gui, Dongbing Gu, Sen Wang, and Huosheng Hu. 2015. A review of visual inertial odometry from filtering and optimisation perspectives. *Advanced Robotics* 29, 20 (2015), 1289–1301.

21. R. I. Hartley and A. Zisserman. 2004. *Multiple View Geometry in Computer Vision* (second ed.). Cambridge University Press, ISBN: 0521540518.

22. Karen J Hayman, Ngaire M Kerse, Steven J La Grow, Trecia Wouldes, M Clare Robertson, and A John Campbell. 2007. Depression in older people: visual impairment and subjective ratings of health. *Optometry & Vision Science* 84, 11 (2007), 1024–1030.

23. Vera Heyl and Hans-Werner Wahl. 2001. Psychosocial adaptation to age-related vision loss: A six-year perspective. *Journal of Visual Impairment & Blindness* 95, 12 (2001).

24. Bill Holton. 2015. A Review of the Be My Eyes Remote Sighted Helper App for Apple iOS. *Access World Magazine* 16, 2 (2015).

25. Apple Inc. 2018. ARKit Apple Developer. `https://developer.apple.com/arkit/`. (2018).

26. Tatsuya Ishihara, Jayakorn Vongkulbhisal, Kris M Kitani, and Chieko Asakawa. 2017. Beacon-Guided Structure from Motion for Smartphone-Based Navigation. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*. IEEE, 769–777.

27. Corinne Kirchner, Emilie Schmeidler, and Alexander Todorov. 1999. Looking at Employment through a Lifespan Telescope: Age, Health, and Employment Status of People with Serious Visual Impairment. *Mississippi State, MS: Rehabilitation Research and Training Center on Blindness and Low Vision.* (1999).

28. Robin Leonard, Tana D'Allura, and Amy Horowitz. 1999. Factors associated with employment among persons who have a vision impairment: A follow-up of vocational placement referrals. *Journal of Vocational Rehabilitation* 12, 1 (1999), 33–43.

29. Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. 2015. Keyframe-based visual–inertial odometry using nonlinear optimization. *The International Journal of Robotics Research* 34, 3 (2015), 314–334.

30. Mingyang Li and Anastasios I Mourikis. 2013. High-precision, consistent EKF-based visual-inertial odometry. *The International Journal of Robotics Research* 32, 6 (2013), 690–711.

31. Google LLC. 2018. ARCore Overview. `https://developers.google.com/ar/discover/`. (2018).

32. MIPsoft. 2018. BlindSquare: Pioneering accessible navigation. `http://www.blindsquare.com/`. (2018).

33. John Morris and James Mueller. 2014. Blind and deaf consumer preferences for android and iOS smartphones. In *Inclusive designing*. Springer, 69–79.

34. Bonnie O'Day. 1999. Employment Barriers for People with Visual Impairments. *Journal of Visual Impairment & Blindness* 93, 10 (1999).

35. Barry W Rovner, Pamela M Zisselman, and Yochi Shmuely-Dulitzki. 1996. Depression and disability in older people with impaired vision: a follow-up study. *Journal of the American Geriatrics Society* 44, 2 (1996), 181–184.

36. Gary S Rubin, Karen Bandeen Roche, Patty Prasada-Rao, and Linda P Fried. 1994. Visual impairment and disability in older adults. *Optometry & Vision Science* 71, 12 (1994), 750–760.

37. William R Wiener, Richard L Welsh, and Bruce B Blasch. 2010. *Foundations of orientation and mobility*. Vol. 1. American Foundation for the Blind.

38. Rebecca A Williams, Barbara L Brody, Ronald G Thomas, Robert M Kaplan, and Stuart I Brown. 1998. The psychosocial impact of macular degeneration. *Archives of Ophthalmology* 116, 4 (1998), 514–520.