Leveraging Augmented Reality to Create Apps for People with Visual Disabilities: A Case Study in Indoor Navigation

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
The introduction of augmented reality technology to iOS and Android enables, for the first time, mainstream smartphones to estimate their own motion in 3D space with high accuracy. For assistive technology researchers, this development presents a potential opportunity. In this spirit, we present our work leveraging these technologies to create a smartphone app to empower people who are visually disabled to more easily navigate through indoor environments. Our app, *Clew*, allows users to record routes and then load them, at any time, providing automatic guidance (using haptic, speech, and sound cues) along the route. We present our user-centered design process, Clew's system architecture and technical details, as well as both small and large-scale usability studies. Our work expands the capabilities of technology for orientation and mobility that can be distributed cheaply and at massive scale. We also discuss opportunities, pitfalls, design guidelines, and limitations of augmented reality for orientation and mobility apps.

ACM Classification Keywords
K.4.2. Assistive technology for persons with disabilities: Orientation and mobility tools for persons with visual disabilities

Author Keywords
orientation and mobility; augmented reality; assistive tech

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 [2, 25], individuals with better O&M skills fare better at finding employment [10, 9, 26, 33]. Similarly, while the link between blindness and depression is well documented [35, 34, 20, 21], studies suggest that it is disability rather than blindness itself that is at the root of this linkage [34, 41]. For example, it was found that the ability to perform activities of daily living was *more* predictive of overall life satisfaction than was degree of vision loss [41]. Due to the importance of O&M, there is a long history of assistive technologies designed to bolster these skills [3, 6]. Despite considerable effort, historically, few of these technologies have achieved much impact beyond the lab [39]. This lack of impact has been driven, largely, by the fact that the technologies were either expensive, unreliable, cumbersome, did not provide significant benefits over simpler solutions, or were hard to distribute at mass scale.

One notable exception to this disappointing track record are GPS navigation apps for smartphones. Since these apps are mainstream technologies that happen to be universally accessible — as opposed to special purpose assistive technologies — they are highly robust, powerful, and extremely useful to people who are B/VI. Further, the fact that a majority of people who are B/VI own smartphones [31] makes them distributable at either no additional cost (e.g., Google Maps) or at modest cost (e.g., BlindSquare which is designed for people who are B/VI and costs about $30).

While being incredibly useful, GPS apps for O&M are not applicable to all mobility tasks. Most notably, mass market GPSes are only accurate to about 5*m* under open sky (and are even worse in challenging environments such as cities) and do not work indoors. Researchers are working to build systems that overcome these challenges (see *Related Work*).

Recently, smartphone manufacturers have introduced augmented reality modules (AR), which support high-accuracy 3D-tracking. While the primary purpose of these modules is to enable AR applications — whereby virtual and physical content are mingled, e.g., by overlaying virtual characters on a smartphone's camera feed — these modules have the potential to be repurposed to create assistive technology for O&M that is robust, accurate, usable, widely deployable, and free.

With the significant potential of smartphone-based AR technology comes critical research questions. Are the motion estimates provided robust enough to use for O&M? If so, which O&M tasks might be facilitated? What usability challenges does AR technology bring, and how can we, as designers, best support users in harnessing such technology? Here, we take preliminary steps towards answering these questions by presenting our work utilizing user-centered design to leverage the AR modules in modern smartphones to create, and to release to a large global audience, an application to assist with indoor navigation. Our app, *Clew*, enables users to record routes using their smartphone so that they can navigate these

ASSETS'19, October 28–30, 2019, Pittsburgh, Pennsylvania

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

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

routes later. *Clew* 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 or practicing a new route in an unfamiliar environment).

In the remainder of the paper we present related work on O&M assistive technology, discuss the algorithms that underlie AR technology on modern smartphones, present the design of *Clew*, provide small and largescale usability data for *Clew*, and finally conclude with a discussion of future challenges and promising directions for smartphone-based AR technology for people who are B/VI.

RELATED WORK

A number of researchers have worked to overcome the limitations (e.g., less-than-ideal accuracy, lack of availability indoors) of GPS technology for assisting people who are B/VI with O&M. Researchers have pursued roughly two approaches. The first is to utilize crowdsourcing, whereby people who are B/VI connect over the internet with a sighted person for real-time assistance. Examples of this approach include the VizWiz project [4], BeMyEyes [22], and Aira [8] (both BeMyEyes and Aira support a video chat interface). A second approach combines rudimentary motion estimates derived from inertial sensors (gyroscopes and accelerometers) with detection of fixed environmental infrastructure (e.g., Bluetooth beacons, Wifi access points). For instance, [16, 15, 17] developed a system for navigation using smartphone-detectable RFID tags. Dias and her collaborators utilized WiFi fingerprinting and dead-reckoning for indoor navigation [12], and similar systems based on Bluetooth beacons have also been developed [24, 1]. Others have explored the use of robots, with sophisticated sensors, to act as guides for people who are blind [32].

Generally need to expand on this given that we are not trying to hit the 5-page limit. More detail can be added as well as incorporating some of the references from my 2018 CAREER submission.

AUGMENTED REALITY

Both Apple and Google have released support for sophisticated, smartphone-based AR experiences, whereby virtual and real content are combined. For instance, an app might show a virtual cat projected into a real world scene. As the user moves, the phone senses the user's motion and renders the cat at an appropriate distance and angle, providing the illusion that the cat exists in the physical world.

These AR systems utilize 3D motion-tracking algorithms that, combine data from inertial sensors (gyroscopes and accelerometers) with visual information (obtained from the phone's camera), to generate motion estimates that are far more accurate than what could be obtained using only inertial sensing. The high accuracy of these systems is driven by two key trends: the development of sophisticated algorithms for visual-inertial odometry (VIO) [28, 27, 5, 14] (which are algorithms that enable the combination of optical and inertial data for motion estimation) and the development of special-purpose hardware to allow computationally intensive algorithms to run with minimal heat generation and power consumption. The combination of high accuracy and low power consumption provides the motivation for the exploration of the potential of AR technology for creating O&M apps for people who are B/VI.

Algorithms for Visual Inertial Odometry

While a full explanation of VIO [18] is beyond the scope of this paper, it helps to have a conceptual understanding of VIO. VIO algorithms are designed for either the monocular (single camera) or stereo setting. Since the monocular setting is the one applicable to mass-market smartphones, here we use the term VIO to refer to monocular VIO specifically.

VIO algorithms utilize sensor fusion to blend optical and inertial motion estimates. Optical motion estimates are made by tracking salient visual features — e.g., corners or other highly textured portions of an image — through multiple video frames. Utilizing the mathematics of perspective geometry, one can estimate the rotation and translation of the phone [19]. Importantly, the accuracy of these estimates is dependent on tracking a large number of visual features that should, ideally, correspond to points at a range of depths from the camera.

Further complicating matters, the translation estimated using optical tracking is only determined up to an arbitrary scale factor. This indeterminacy arises due to the fact that the depths of the tracked visual features are unknown [19]. For example, given an estimate of the translation of the phone, it is possible that the phone moved twice as far and the depths of the tracked points were twice as great. The shortcomings of optical tracking, scale-indeterminacy and inaccurate performance in feature-poor environments, can be overcome by the fusion of inertial measurements from gyroscopes and accelerometers. Gyroscopes, which provide accurate estimates of angular velocity, can refine estimates of rotation while accelerometer data can be integrated over time to obtain estimates of linear velocity, overcoming the scale-indeterminacy problem.

VIO in Mass-Market Smartphones

Both Apple and Google have released AR modules based on VIO. While, the details of their VIO algorithms are not publicly available, there are distinctions between these frameworks that researchers should keep in mind.

Google Tango

Release in late 2014 by Google's ATAP (Advanced Technology and Projects) division, the Tango utilizes a wide-angle lens and a global image shutter to enable accurate visual-feature tracking. The platform also includes a PrimeSense depth-sensing camera. Two commercial smartphones have been released based on the 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 on special-purpose hardware severely limited the adoption of the technology. As a result, Google suspended the project in early 2018 [11].

ARKit and ARCore

Apple's ARKit [23], released in 2017, and Google's ARCore [29], released in 2018, do not require special purpose hardware. Since these platforms utilize conventional cameras, which have comparatively limited field-of-view when compared to a fisheye camera, the richness of visual features available for tracking is not as great as with Tango, and consequently, their motion estimates are less accurate. Further, since neither of these platforms have depth sensing cameras, the availability of

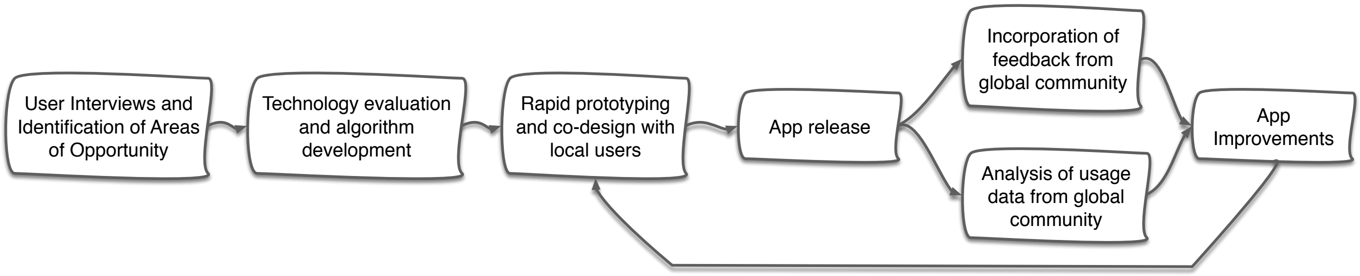

Figure 1. Our design process used to investigate the usage of AR technology for creating assistive technologies for people who are B/VI. We are using this process to design our app *Clew*.

3D information is limited to objects with special structure (e.g., planar horizontal and vertical surfaces). Despite their drawbacks, these frameworks can run on a wider array of phones than Google Tango; however, given the large preference for iOS among people who are B/VI [31], ARKit is the primary platform of interest for researchers seeking to develop O&M apps for people who are B/VI.

DESIGN PROCESS OVERVIEW

We employed a user-centered design process that involved people who were B/VI in all phases. Our process is summarized in Figure 1. In the initial phase, we focused on in-person interactions with local members of the B/VI community to identify and explore areas of opportunity for improving access to physical spaces (see *User Interviews and Areas of Opportunity*). Based on these initial interactions, we came up with concepts for O&M apps that fed into an intensive co-design process where we created system designs, selected and developed algorithms, and ultimately produced prototype apps, which our co-designers evaluated (see *User Interviews and Areas of Opportunity*). Based on these co-design sessions, we generated suggestions for making these prototypes better.

Once we had an app that we felt was sufficiently polished as well as useful to the B/VI community, we released the app on the iOS App Store. The release of the app generated feedback in two forms. First, users from all over the world gave us their impressions of the app and how they would want to see it improved. Second, users who opted-in to sharing usage data, provided a rich dataset to understand, in a quantitative manner, how the app was being utilized and what its limitations are (see *Large-scale, Data-Driven User Study*).

User Interviews and Areas of Opportunity

We employed semi-structured interviews and participatory design approaches [7, 37] with members of the local B/VI community. We engaged in a series of co-designs with Joe (pseudonym), a college student who has no functional vision due to Retinitis Pigmentosa (a degenerative vision disorder that leads to the breakdown of cells in the retina); he is a non-visual traveler, mostly navigating with the help of his guide dog. We also learned from an O&M trainer for primary and middle school students who shared stories drawn from his own experiences. Furthermore, three members of our research team, two of whom are low-vision and the other of whom is completely blind, contributed both to the design and implementation of the apps and leveraged their personal experiences and knowledge of the B/VI community to inform our design process. Our team identified a number of pain points and areas of opportunity related to orientation and mobility and access to physical spaces.

Navigating in unfamiliar indoor environments is difficult. Joe expressed that navigating in buildings that he has never visited is an especially difficult task. In these situations Joe will often need to either call a sighted person to assist him once he arrives at the building or bring along a sighted friend or familiar member to help him find his way. If he will be navigating this environment over a long period of time, for instance if he is starting a new job, then he will often work with a mobility instructor in order to learn how to effectively get around in this new environment. He identified his need to rely on others for assistance in these scenarios as a major impediment to independence.

Navigating newly traveled routes towards previously visited locations is difficult. A specific subset of the difficulties encountered in navigating new indoor environments is navigating back to a starting point after traveling to a new location. As an example, Joe expressed that finding his seat (e.g., in a classroom) after going to the restroom was challenging, particularly when traveling without his dog. This story resonated for one of our visually impaired team members who found it difficult to find his seat on a dimly lit airplane while returning from the restroom.

The identification of these pain points motivated the creation of *Clew* to support precise indoor navigation in unfamiliar environments. In *Clew App Overview*, we present the app's features and technical architecture. In *Usability Testing and Technology Evaluation*, we present both small and large-scale user studies that provide insight into the app's capabilities, suggest improvements to its design, and provide guidelines for developers of similar apps.

CLEW APP OVERVIEW

Clew is based on ARKit (link to Clew on the app store redacted for anonymous review). Clew provides turn-by-turn directions to guide users who are B/VI along recorded routes. The app is designed for two primary use cases.

The first use case is to allow the user to record a path through an indoor environment and then navigate the path back to the starting location when they are ready. This use case can arise, e.g., when a user is led somewhere by a sighted guide and

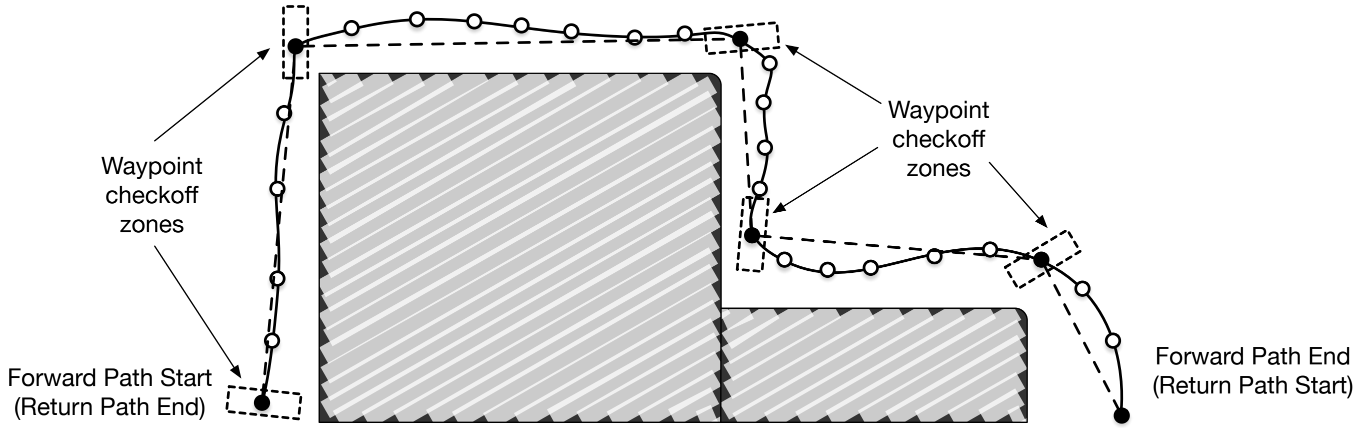

Figure 2. A topdown view of a path generated by the *Clew*. The figure shows the raw path (solid line) and breadcrumbs (white circles with black outlines). The Ramer-Douglas-Peucker reduces the breadcrumbs to a more manageable number of waypoints (solid black circles). The resultant path consists of straight lines connecting these waypoints in reverse order. Waypoint checkoff zones are shown as dashed boxes).

they want to return to their previous location — without being guided back. As a second example, sometimes it is easier to navigate from a location than back to it. For instance, it is easier for someone who is B/VI to leave a conference room than it is to find their way back to their particular seat.

The second use case is to allow the user to record a path through an indoor environment, save this path, and then navigate the path either in the forward or reverse direction at a later point in time. This function is useful when a user is either learning to navigate a new route and could use guidance when practicing the route or when the user finds themselves in a situation where they will be navigating a new route repeatedly for a brief period of time (e.g., during a hotel stay a user may want to be able to navigate to their room or to the hotel pool).

In order to serve each of these use cases, *Clew* provides high-accuracy, easy-to-follow, navigational guidance to enable people who are B/VI to travel independently indoors, without the need for modifications to the environment (e.g., the introduction of beacons or special signage).

Path Recording

In path recording mode, the app creates a trail of virtual breadcrumbs (representing timestamped, 3D positions and orientations). The estimates of position and orientation are generated by Apple's ARKit library, which provides these at a rate of 60 Hz. Recall that these position and orientation estimates are based on fusing visual and inertial measurements. We record the phone's position every 300 milliseconds (this sampling interval was chosen to tradeoff computational constraints with faithfully representing the travelled route).

In our *technology evaluation and algorithm development* design phase (see Figure 1), we discovered that the motion tracking estimates of ARKit are most accurate when the user holds their smartphone with the screen facing towards them and the optical axis of the camera facing roughly parallel to the ground. This anecdotal finding is supported by the details of the underlying VIO algorithms, whose geometrical calculations are best conditioned when visual features are tracked at a range of depths from the camera. If the user's phone is pointed down at the floor or to the side at a wall, the phone will track features that primarily lie on a single plane, which will result in less reliable estimates of motion.

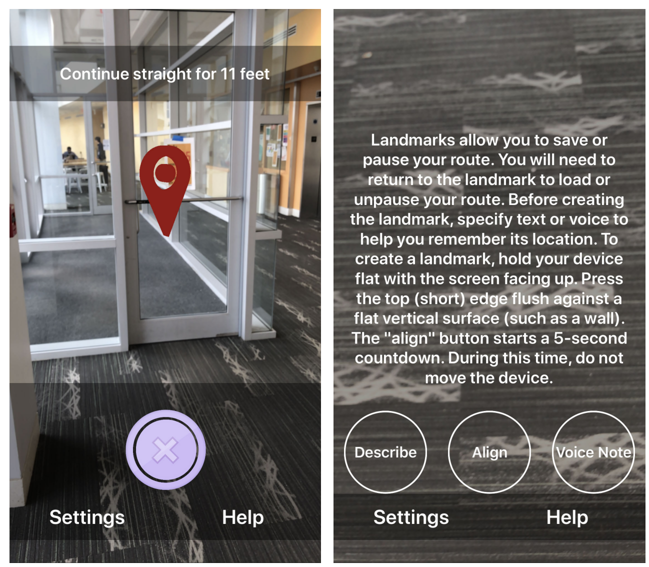

Figure 3. Two screenshots from our app "Clew" in navigation mode (where a user is retracing a previously traveled route). The text for the left image says "Continue straight and proceed downstairs" and the text on the right says "Continue straight for 3.2 feet."

Once the user starts path recording, they travel to a new location (either via their traditional O&M process or via assistance from a sighted guide) and then stop the recording. Figure 2 shows a sample path along with virtual breadcrumbs.

Route Pausing and Saving through Landmark Creation

ARKit tracks motion in a coordinate system whose origin coincides with the device's position when the app starts. Thus the positions of any breadcrumbs dropped during the path recording phase are only meaningful *within* the context of the current ARKit tracking session. Once a path is recorded, if the user doesn't switch to a different app, lock the phone, or occlude the camera, then they can navigate back to their starting location using the procedures described in *Path Navigation Mode*. However, if the user wishes to wait a significant amount of time before navigating back, would like to use another app, needs both hands to perform some task, or if the user would like to save the route for use at a later time, these limitations can be prohibitive. In order to support such cases, Clew provides support for route pausing and saving through coordinate system registration.

Registration through Visual Alignmnent

With the release of iOS 12.0, Apple's ARKit supports a relocalization feature whereby a visual representation of landmarks and their associated 3D positions can be stored to form a sparse map of an ARKit session. Given a new ARKit session (e.g., the user has restarted the app), ARKit can load the previous 3D map and attempt to match the phone's current image features to the map. If a match is found, the current position and orientation of the phone is updated to be relative to the coordinate system in the 3D map (thereby registering the two coordinate systems). In order to support route saving and pausing, *Clew* will save the current 3D map of the tracking session and attempt to use visual matching to relocalize the user when the user wants to navigate the route at a later time.

The main limitation of visual alignment is that it doesn't always succeed. This can occur for a multitude of reasons. First, if the environment has changed significantly, for instance when lighting conditions have changed drastically, the visual environment may not look similar enough to the saved 3D map. Alternatively, even if the environment is relatively static the user's viewing angle on the environment might be significantly different. In particular, this is likely to happen when navigating routes in the reverse direction as visual features in the environment often appear very differently when viewed from the opposite direction (thus causing matching problems).

Registration through Physical Alignment

In order to allow users to reload or pause routes when visual alignment is impossible, Clew supports physical alignment through landmarks. A user can create a landmark by placing their phone in a position and orientation (collectively called a "pose") that is easy to return to, unassisted, when the user wants to resume or reload the route. It is not crucial that the user be able to perfectly recreate this pose. Specifically, the most important aspect of the pose to recreate is its yaw. The reason for this is that the phone's accelerometer can determine the other two degrees of freedom of rotation, pitch and roll, as these are not perpendicular to gravity and deviations in the phone's position contribute a fixed amount of error, whereas an error in the phone's yaw will be magnified over long routes.

The suggested procedure for landmark creation is for the user to place their phone's top (short-edge) flush against a flat vertical surface (such as a door or wall). In this configuration the user's camera will be facing down, which will allow it to track visual features on the floor, and the screen will be facing up, which allows the user to see what's on the screen (if they have usable vision) or interact with the phone using VoiceOver. In order to make the alignment process more robust, we artificially "level" the phone by undoing the effects of roll and pitch to create a virtual landmark pose in which the phone is perfectly flat. Further, Clew allows the user to enter text or record a voice message to help them remember the details of how they positioned their phone when creating the landmark (e.g., "Office front door, right above the handle").

We tested several other procedures for landmark creation (e.g., with the screen flat against the wall) and we found that this mode was both easy for the user to perform and accurate in terms of its ability to align to paused or saved routes.

Requiring Landmark Creation

In order to load a route and navigate it at a later time or to pause a route so the user can switch to another app or lock their phone, they must create a route landmark. While in theory we could support registration with visual alignment only, we chose not to allow this as an option since it isn't guaranteed to succeed in relocalizing. When reloading a route, the user can choose to navigate the route in either the forward or reverse direction in comparison to the direction it was originally recorded. In order to navigate in a particular direction, the user must have created a landmark at either the beginning of the route (if navigating int he forward direction) or at the end of the route (if navigating in the reverse direction).

Path Navigation Mode

When the user is ready to navigate using the app, the trail of breadcrumbs is processed by the Douglas-Ramer-Peucker (DPR) algorithm [13] for path simplification. This algorithm winnows down the path by removing sequences of breadcrumbs that are well represented by a straight line. Figure 2 shows the breadcrumbs selected by the DPR algorithm, called *waypoints*, and the resultant piecewise straight navigation route obtained by connecting the 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 or (3) audible feedback when the phone is pointing towards the next waypoint. When using (2) or (3), the low latency of the update of the phone's position allows the user to sweep their phone back and forth until they sense a haptic or auditory cue, providing an accurate sense of the direction to the next waypoint.

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 2). Instead of making these checkoff zones spherical, we made them rectangular prisms with sides perpendicular to the direction of travel longer than for dimensions parallel to the direction of travel. This choice of shape enables the user to deviate laterally from the intended path without missing a waypoint. The app also announces flights of stairs by detecting if the vertical angle of the segment connecting two waypoints exceeds a threshold.

Novelty in Relationship to Previous Work

Our work on the design of *Clew* represents the first research into automatically guiding a user back along a route, without the need for special modifications to the environment.

USABILITY TESTING AND TECHNOLOGY EVALUATION

Previously in this document we discussed our initial design work that we undertook to inform the concept of Clew. Here, we discuss the results of co-designs as well as data analysis experiments that informed the creation of specific aspects of Clew. A number of our results have implications for researchers who would like to use AR technology to create assistive tech for people who are B/VI.

Longitudinal Design with Local Co-Design Partners

In addition to our work with Joe, whose insights helped inform the concept of the Clew app. We engaged with three other local co-design partners to further refine the design of Clew. We worked with each co-designer for five, two-hour sessions spaced a week apart.

Insight 1: Maintaining Optimal Phone Position

For apps that require the user to hold the phone in their hand, recall that VIO algorithms work best when they are able to detect a large number of visual features at a range of depths. This condition is best achieved when the user holds their phone upright with the camera facing approximately parallel to the ground. We found that Joe had difficulty maintaining the phone in this configuration, perhaps due to the lack of visual feedback about the phone's orientation. As a result, tracking performance suffered. Developers of AR-powered assistive apps should consider adding feedback mechanisms to help the user to maintain their phone in an optimal orientation. In our large-scale evaluation, we were able to demonstrate this quantitatively.

Insight 2: Alignment Between Body and Phone

Through interactions with a co-designer from the local area named Jim (pseudonym), we discovered that some users have a difficult time understanding the orientation of their phone relative to the orientation of their body. Jim is congenitally blind, whereas the other three local co-designers, none of whom had this difficulty, lost their vision in their teens. That the age at which sight was lost would have this effect was not wholly unexpected (see, e.g., [30, 39, 36, 38, 40] for discussions of how spatial processing is affected). We found that Jim could easily rotate his body in an effort to elicit haptic or auditory guidance from the app, which would indicate that he was facing the correct direction.

In order to provide feedback relative to his body direction rather than the phone's direction, we developed a feature to constantly update an estimate of the phone's offset relative to the user's body. The calculation of this offset can only be performed when the user is moving forward (moving laterally will throw this calculation off). We found that this modification enabled Jim to use the app effectively. Developers of AR-powered assistive apps should be cognizant of the differing abilities of users to sense spatial relationships between various parts of their body and their phone.

Technology Evaluation with College Students

As a preliminary usability test, four college students used Clew to navigate six routes. For each route, the participant was led by one of the authors — mimicking the role of a sighted guide — and then navigated back using Clew. Participants were instructed to follow the guidance of the app, avoiding the use of their memory to correct for the app's errors. In order to become accustomed to the app, the participants were sighted for the first two trial routes. A trial was unsuccessful if the user couldn't follow the app's guidance (e.g., the app's estimation of a waypoint's location was inside a wall due to a failure of the ARKit's tracking algorithm). For successful trials (the user navigated to all waypoints), we recorded the distance from the user's final position to the ground truth position. A summary of each of the routes and the results are shown in Table 1. These results demonstrate the robustness of Clew for navigating fairly complex routes, including routes that navigate stairs and/or through places with poor visual features. The current study, while promising, is limited by its use of participants with occluded vision, rather than people who are B/VI. While our co-designer who is blind tested Clew extensively, we are currently working on a full evaluation of Clew with multiple users who are B/VI.

Length	*Stairs*	*Visuals*	*Success*	*Mean Error*	*Median Error*
21.3m	no	good	100%	0.15m	0.15m
40.2m	yes	mixed	100%	0.99m	0.91m
11.9m	no	poor	100%	0.91m	0.61m
22.9m	no	good	100%	1.45m	1.37m
47.2m	yes	mixed	100%	1.45m	1.37m
43.4m	yes	good	75%	2.03m	1.83m

Table 1. Usability results for Clew. The *Visuals* column indicates the availability of trackable visual features, with "mixed" indicating that a route had segments with both good and poor quality visual features.

Feedback from Global User Community

write this

The Importance of Route Saving

Supporting older iOS versions

write this

Designing for Low-vision Users

write this

Large-scale, Data-Driven User Study

write this

Adoption and Usage of Clew

write this

Device Pose During Navigation

Include the histogram of device verticality

Communicating Tracking Failures to the User

this is a good story linking data analysis to creating an app feature.

write this

Regression Analysis

We now have a result that shows that when the phone is not vertical, users are more likely to give the route a thumbs down.

write this

FUTURE WORK FOR CLEW

First, in order to make *Clew* successful, we must continue to engage deeply with users to understand how they experience the current version of the app and improve it based on their feedback. Second, we are working to improve the ability of the app to detect and communicate tracking failures. Specifically, when ARKit throws an error, it is straightforward to communicate this information to the user; however, if tracking drifts or becomes inaccurate without throwing an error, detection is much more difficult. Third, we are working to enable the sharing of previously traveled routes with other users. This feature will work similarly to the pause feature in that the user will place their phone in a known starting pose (e.g., against a door as they enter a building), allowing their phone to register its current pose with the starting pose of a saved route.

SUMMARY OF CONSIDERATIONS FOR RESEARCHERS WHO WANT TO USE AR FOR ASSISTIVE TECHNOLOGY

Robustness of AR motion estimates for precise O&M

Our study demonstrates that for navigation routes of $\sim 45m$ the motion estimates of ARKit are sufficiently accurate. Additionally, for object finding within a room or on a tabletop we have found, anecdotally, that ARKit's performance is sufficient.

As the complexity of the environment increases other sources of information (e.g., the detection of landmarks such as optical tags or Bluetooth beacons) could reduce tracking drift, providing spatial anchors. Alternatively, crowdsourcing could correct drift in the phone's motion estimates (e.g., a crowdworker might identify a landmark at two points in time).

Usability Concerns of O&M technology using AR

Since AR relies on VIO, which leverages optical tracking, apps based on AR require the camera to be unoccluded. Currently, our apps assume that the user holds their phone in one hand while holding their cane in the other. Future research should consider whether a method of handsfree operation can be developed that allows the user to continue to use their phone as an interface. This could work by using an attachment (e.g., a neck lanyard) for holding the phone coupled with a speech-based user interface. We are currently exploring this idea.

FUTURE WORK AND AREAS OF OPPORTUNITY

Finding objects in unfamiliar environments was difficult. While Joe mentioned he had no difficulty in finding objects in spaces he has control to organize himself, he admitted that locating objects of interest in unfamiliar places or when others disturb his organization can be very challenging. Additionally, finding precise locations in outdoor environments (e.g., the correct part of a train platform or button to activate a crosswalk) is very difficult. This set of insights helped support the creation of our app *ViewShare* for object finding.

CONCLUSION

We have presented two smartphone apps that allow people who are B/VI to perform difficult O&M tasks with their smartphones. Our apps are among the first mass-distributable apps for navigation and object finding in unmodified indoor environments. While the results of our usability study are promising, more work, particularly in concert with users who are B/VI, is required to further refine the developed apps.

Additionally, we have outlined several promising areas of opportunity for the development of AR apps for O&M and provided a discussion of the usability factors of this technology. Through the development of new algorithms, co-design with users, and the improvement of AR technology itself, researchers can hopefully continue to leverage AR to create impactful 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. 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`
3. J Malvern Benjamin. 1973. The new C-5 laser cane for the blind. In *Proc. Carnahan Conf. on Electronic Prosthetics*. 77–82.
4. Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and others. 2010. 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.
5. 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.
6. 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.
7. Christian Bühler. 2001. Empowered participation of users with disabilities in R&D projects. *International Journal of Human-Computer Studies* 55, 4 (2001), 645–659.
8. Aira Tech Corp. 2018. aira: your life, your schedule, right now. `https://aira.io/`. (2018).
9. 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.
10. 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).

11. Tech Crunch. 2017. Android Guys. `http://www.androidguys.com/2017/08/30/google-rebrands-tango-as-arcore/`. (August 2017).
12. M Bernardine Dias . 2014. *NavPal: Technology Solutions for Enhancing Urban Navigation*. Technical Report CMU-RI-TR-21. Robotics Institute, Pittsburgh, PA.
13. 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.
14. 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.
15. 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.
16. 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).
17. 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.
18. 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.
19. R. I. Hartley and A. Zisserman. 2004. *Multiple View Geometry in Computer Vision* (second ed.). Cambridge University Press, ISBN: 0521540518.
20. 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.
21. 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).
22. Bill Holton. 2015. A Review of the Be My Eyes Remote Sighted Helper App for Apple iOS. *Access World Magazine* 16, 2 (2015).
23. Apple Inc. 2018. ARKit Apple Developer. `https://developer.apple.com/arkit/`. (2018).
24. 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.
25. 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).
26. 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.
27. 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.
28. 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.
29. Google LLC. 2018. ARCore Overview. `https://developers.google.com/ar/discover/`. (2018).
30. Richard G Long and EW Hill. 1997. Establishing and maintaining orientation for mobility. *Foundations of orientation and mobility* 1 (1997).
31. John Morris and James Mueller. 2014. Blind and deaf consumer preferences for android and iOS smartphones. In *Inclusive designing*. Springer, 69–79.
32. Amal Nanavati, Xiang Zhi Tan, and Aaron Steinfeld. 2018. Coupled Indoor Navigation for People Who Are Blind. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 201–202. DOI: `http://dx.doi.org/10.1145/3173386.3176976`
33. Bonnie O'Day. 1999. Employment Barriers for People with Visual Impairments. *Journal of Visual Impairment & Blindness* 93, 10 (1999).
34. 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.
35. 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.

36. Victor R Schinazi, Tyler Thrash, and Daniel-Robert Chebat. 2016. Spatial navigation by congenitally blind individuals. *Wiley Interdisciplinary Reviews: Cognitive Science* 7, 1 (2016), 37–58.

37. Douglas Schuler and Aki Namioka. 1993. *Participatory design: Principles and practices*. CRC Press.

38. Catherine Thinus-Blanc and Florence Gaunet. 1997. Representation of space in blind persons: vision as a spatial sense? *Psychological bulletin* 121, 1 (1997), 20.

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

40. Michele A Williams, Caroline Galbraith, Shaun K Kane, and Amy Hurst. 2014. just let the cane hit it: how the blind and sighted see navigation differently. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. ACM, 217–224.

41. 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.