Leveraging Augmented Reality Technology for 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 track the motion of a user's phone 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 misplaced objects. 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.

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. While only 30% of working-age Americans who are B/VI are employed [3, 26] (compared with 65% of the general population), individuals with better O&M skills have a higher likelihood of being employed [11, 10, 27, 33]. Similarly, while the link between visual-impairment and depression has been well-documented [35, 34, 21, 22], several studies have suggested that it is disability rather than blindness itself that is at the root of this linkage [34, 37]. For example, it was found that one's ability to perform daily-life tasks, such as shopping for basic necessities, was *more* predictive of overall life satisfaction than was degree of vision loss [37].

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

Due to its high degree of importance, there is a long history of engineers creating assistive technology to empower people are B/VI to perform O&M tasks more easily (see the *Background and Related Work* section for a discussion of this). Here, we explore new technological developments and trends, which are enabling the development of widely accessible assistive technology for O&M. The first enabling trend is the high rate of ownership of smartphones by people who are B/VI [32]. The availability of these devices has already had revolutionary impacts on accessibility (e.g., through GPS-based directions, OCR software). The second driver of this trend is the introduction of high accuracy 3D-tracking capabilities into mainstream smartphones. While the primary purpose of these modules, e.g., Apple's ARKit [24], Google's ARCore [30], 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 very powerful assistive technology for O&M.

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 service of our goal of designing impactful technologies, we employed user-centered design principles throughout the research and development lifecycle, including working longitudinally with a co-designers who are blind. Further, two of the study authors, who have severe visual impairments, contributed to all aspects of the project and additionally provided much needed design guidance to the project based on their personal experiences.

Our first app is called Clew, and it enables users to backtrack along previously traveled routes. This capability is designed to help with various pain points experienced by non-visual travelers, including finding one's way independently after being brought to a location by a sighted guide. The second app is called View Share, which 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 provide some relevant background and related work on O&M assistive technology, discuss some of the algorithms that underlie the AR technology of modern smartphones, discuss our two developed apps in detail, provide preliminary usability, and finally conclude with a discussion of future challenges and promising directions of smartphone-based AR technology for people who are B/VI.

Need to sell harder in the intro once the story of the paper be-

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 things such as 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 [8]. These sorts of devices based around obstacle detection have not achieved high adoption by the B/VI community for a number of reasons, including their cost, lack of reliability, and limited new functionality over the white cane (see [36] for a discussion of this issue).

Recently, spurred by the high ownership rates of smartphones among people who are B/VI, the focus of this work has shifted from obstacle avoidance problems such as navigation and spatial awareness. For instance, GPS-based navigation apps, such as BlindSquare [31] 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 extensive databases of businesses, streets, and points of interest have allowed these apps to provide high-quality assistance to people who are B/VI.

GPS-based apps for O&M have a number of limitations, including limited accuracy (in good conditions GPS on smartphones is accurate to about $10m$, and can be much less accurate in urban environments) and lack of function inside. To circumvent these challenges, researchers are pursuing largely two major approaches. The first approach is to utilize crowdsroucing, whereby people who are B/VI connect with a sighted person for on-demand assistance. Crowdsourcing has a successful track record in the area of assistive tech (e.g., the highly successful VizWiz project [6]) and this same model has been used by the BeMyEyes app [23] and the Aira system [9], which each connect a user who is B/VI with a sighted volunteer through a video chat interface. The second approach is to utilize along with inertial sensing (e.g., accelerometers and gyroscopes) to enable accurate positioning (both indoors and outdoors). The second class of apps combine rudimentary (and somewhat inaccurate) motion estimation via inertial sensors (gyroscopes and accelerometers) along with special environmental infrastructure (e.g., Bluetooth-enabled location beacons) that provide sporadic location cues to enable relatively accurate indoor positioning. For instance, [17, 16, 18] 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 [13]. A similar system uses low-energy Bluetooth beacons instead of WiFi access points as landmarks [25, 1, 2]. A third approach is to utilize 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 begun to rollout significant new capabilities that enable users to enjoy highly sophisticated augmented reality (AR) experiences on their smartphones. The key characteristic of AR-enabled apps is that virtual content is seamlessly combined with real world content. The most common instantiation of this idea is to overlay virtual objects or characters on top of the video feed form a smartphone. For instance, Figure 1 shows a virtual cat that is projected into a real world scene. As the user moves around in the space, the phone will detect the user's motion with high accuracy and automatically render the cat from an appropriate angle and distance, 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 new AR systems are made possible by spatial processing algorithms that are vastly more accurate than the simple 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) [29, 28, 7, 15] (which combine optical tracking using a smartphone's camera with inertial sensing for motion estimation) and the development of special purpose hardware that allows these highly computationally intensive algorithms to run on a user's 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 function. A full explanation of VIO is beyond the scope of this document. For a more comprehensive treatment consult [19]. Approaches for VIO have been developed for both the stereo setting and the monocular (or single camera) setting. The monocular setting is directly applicable to most modern smartphones, which either only have one rear-facing camera or have a second camera that is unsuitable for use in a stereo pair. For the remainder of this section, we'll focus on the monocular case only.

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 [20]. Of particular interest to the creation of assistive apps-based on this technology, the accuracy of these motion estimates is highly dependent on being able to track a large number of these visual features that ideally correspond to points at a range of distances from the camera and are distributed uniformly over the image. This dependency means that VIO is susceptible to inaccurate motion tracking when few visual features are able to be tracked frame-to-frame or when the visual features that are tracked represent are impoverished (e.g., all at the same depth or in the same region of the image). While some environments are simply more difficult for visual tracking, in some cases users may hold their phones in a suboptimal position (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 (perpendicular distance from the image plane) of the visual features are unknown [20]. For example, for any particular estimate of the translation of the camera, it is equally valid that the camera moved twice as far and the depths of the visual features 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, their are important high-level differences in these implementations for application developers to keep in mind.

Google Tango

The Google Tango platform was first made publicly available by Google's ATAP (Advanced Technology and Projects) division in late 2014. Google's platform utilized a special wide angle (fisheye) camera equipped with a global (more precise) shutter that enabled maximally accurate visual feature tracking. Further, the platform included a PrimeSense depth-sensing camera, although the depth sensor was not utilized to estimate the motion of the phone. Two commercial products have been released based on the Google Tango technology: the Lenovo Phab2 Pro and the Asus Zenfone AR. While the tracking capabilities of Tango devices is superior to other platforms (discussed next), the reliance of the platform on special-purpose hardware severely limited the adoption of the technology, leading Google to suspend the project in March of 2018 [12].

ARKit and ARCore

Apple's ARKit [24] and Google's ARCore [30], 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 high as those that can be harvested from fisheye lens on Tango platforms. Based on our anecdotal observation, this results less accurate tracking performance than the Tango. Further, since neither of these platforms have built-in depth sensing cameras, the availability of 3D information about the landmarks in the environment is limited to objects with special structure (e.g., horizontal and vertical planes). Of primary importance is that these frameworks are capable of running on a much wider share of phones than Google Tango. Further, given the much higher preference for iOS devices by people who are B/VI [32], ARKit is the primary platform of interest for researchers seeking to develop assistive apps based on smartphone-based AR technology.

USER-CENTERED DESIGN PROCESS

My 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 both our background research along with our interviews and observation of individuals who are B/VI, we identified several areas of opportunity for the development of assistive apps. The themes of high precision and indoor navigation came up again and again in our interactions with users. Further, the notion of being able to browse (for instance in a grocery store or other public space) was also something that was repeatedly 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 emerging implementation 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 visual tracking, any assistive app based on VIO must maintain the condition that the user's 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. The development of handsfree methods for utilizing VIO is an area we are actively researching.

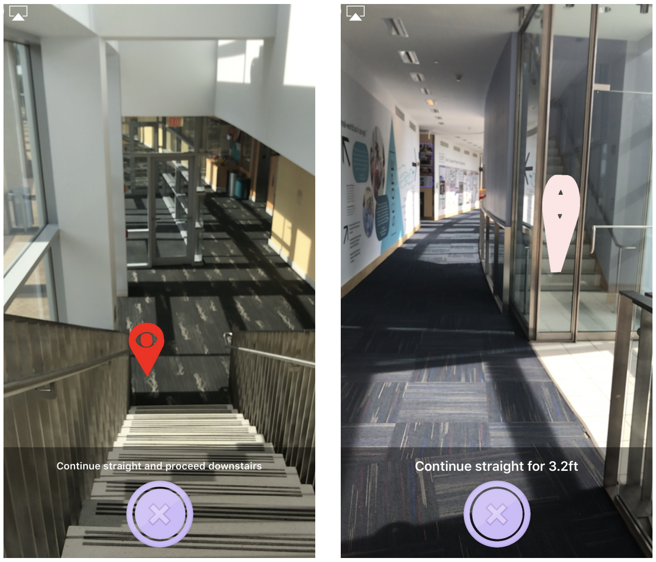

Figure 2. Two screenshots from our app "Clew." Both images show the app in navigation mode where a user is using the app to retrace a route they have previously traveled. The text for the left image says "Continue straight and proceed downstairs" and the text on the right says "Continue straight for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

Participatory Design

Throughout the development of the apps described in this document, we utilized a participatory approach to design. This manifested itself in two specific ways. First, worked with a college student who is blind over the course of five, three-hour co-design sessions to develop the basic concept and test various prototypes for our app *Clew*. 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 experience served as a valuable guiding light throughout the design process.

CLEW: AN APP FOR AUTOMATIC GUIDANCE ALONG PREVIOUSLY TRAVELLED ROUTES

1. Basic idea of the app
2. Co-design elements
3. Lack of suitability for guide dog travel
4. Path recording (including Douglas-Ramer-Peucker algorithm [14] (cool page for generating examples of the algorithm running http://karthaus.nl/rdp/))
5. Path following (including feedback mechanisms)
6. Pause feature
7. Usability test

VIEWSHARE: AN APP FOR OBJECT FINDING IN CHALLENGING ENVIRONMENTS

Todo: connect to idea of sense of space and building mental maps.

1. Basic idea of the app
2. Co-design elements
3. Overview of the 3D location mechanism
4. App for the searcher
   (a) Speech interface
   (b) Automatic snapshotting
   (c) Guidance to object (3D versus 2D feedback)
5. App for the finder
   (a) Description of general interface
   (b) Special interface for triangulation
6. Usability test

DISCUSSION AND FUTURE WORK

1. Interpretation of the results
2. Future work for each of the apps
   (a) New feature development
   (b) Robust usability testing
3. Unresolved Issues
4. Promises of AR technology
5. Limitations of AR technology
   (a) Accuracy
   (b) Accessibility to voice over (minor problem, but worth mentioning)

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. 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.
7. 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.
8. 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.
9. Aira Tech Corp. 2018. aira: your life, your schedule, right now. `https://aira.io/`. (2018).
10. 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.
11. 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).
12. 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-move-wider/`. (August 2017).
13. M Bernardine Dias . 2014. *NavPal: Technology Solutions for Enhancing Urban Navigation*. Technical Report CMU-RI-TR-21. Robotics Institute, Pittsburgh, PA.
14. 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.
15. 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.
16. 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.
17. 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).
18. 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.
19. 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.
20. R. I. Hartley and A. Zisserman. 2004. *Multiple View Geometry in Computer Vision* (second ed.). Cambridge University Press, ISBN: 0521540518.
21. 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.
22. 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).
23. Bill Holton. 2015. A Review of the Be My Eyes Remote Sighted Helper App for Apple iOS. *Access World Magazine* 16, 2 (2015).
24. Apple Inc. 2018. ARKit Apple Developer. `https://developer.apple.com/arkit/`. (2018).

25. 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.

26. 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).

27. 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.

28. 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.

29. 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.

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

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

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

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

37. 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.