Mobile Crowd Assisted Navigation for the Visually Impaired

Greg Olmschenk[1] Christopher Yang[2] Zhigang Zhu[3] Hanghang Tong[4] William H. Seiple[5]
[1,3]The CUNY Graduate Center [1,2,3]The CUNY City College
[4]Arizona State University [5]Arlene R Gordon Research Institute - Lighthouse Guild
{[1]golmschenk@gradcenter| [2]cyang001@citymail | [3]zzhu@ccny}.cuny.edu
[4]hanghang.tong@asu.edu [5]wseiple@lighthouseguild.org

***Abstract*—The World Health Organization estimates that 285 million people are visually impaired worldwide: 39 million are blind and 246 million have low vision. In order to improve the overall situation without having the user feel encumbered, our Crowd Assisted Navigation app is designed for smartphones (including both iPhones and Android phones), which are by far the most commonly used mobile devices among those with low vision. Many individuals would rather forget their wallets at home than their phones. A smartphone is easily accessible and its use does not attract undue attention towards the user's need of aid for his/her disability. The app's primary objective is to assist a visually impaired or blind user in navigating from point A to point B through reliable directions given from an online community. The phone is able to stream live video to a crowd of sighted users through our website. The crowd is then able to give directions from the website with the push of one of the four arrow keys, indicating either left, right, forward, or stop. The aggregation of these directions will be relayed back to the user by audio.**

***Index Terms*—crowd sourcing; mobile; visually impaired; assistive technology; navigation**

I. Introduction

The global visually impaired population is over 285 million people according to the World Health Organization [1], and rapidly growing. In the United States alone, the visually impaired population is 6.6 million people [2] and expected to double by 2030 (from 2010 figures) [3] due to people living longer and thus prolonging chronic diseases, of which blindness or diminished sight are serious complications. With the technology advances in sensors and mobile computing, more and more research and development efforts have been directed at assisting in navigation for visually impaired people. However there is still a long way to go to achieving a wearable vision system comparable to the Google cars with advanced sensors and high computing capacities. Our user study found that the only new navigational technology to be widely adopted by the community has been the talking GPS that provides verbal walking directions. Recently, crowdsourcing assistance has been studied for various applications for the blind and visually impaired, elderly, and people in need, such as video annotation [4], label reading [5], and assisted navigation [6], which offer promise for real applications. In order for the above to be useful for individuals who are blind or who have low vision to perform realtime navigation tasks, several limitations must be overcome in order for it to reach its potential. First, feedback must as quick as possible, ideally in real-time. Second, numerous back and forth queries when the views of the user's picture are not usable would be time-consuming and inefficient for both the user and the volunteers. Third, still images would not provide the dynamic information of the scene such as traffic, pedestrians, etc. For a totally blind person, turn-by-turn guidance may be needed, particularly in an unfamiliar situation, thus video streaming is necessary.

In this paper we present the design of our Crowd Assisted Navigation platform, and describe an app for a smartphone. The app's primary objective is to assist a visually impaired or blind user in navigating from point A to point B through reliable directions given from an online community. The phone is able to stream live video to a crowd of sighted volunteers through our website. The crowd is then able to give directions from the website as simple as the push of one of the four arrow keys, indicating either left, right, forward, or stop. The aggregate of these directions will be relayed back to the user by audio in real time.

The organization of the paper is as follows. In Section II, we will provide a survey of some closely related work. In Section III, some key design considerations will be discussed. Current tests of the system and the app are described in Section IV. Finally in Section V we conclude the paper.

II. Related Work

A number of groups have explored the use of crowdsourcing for assisting blind user in various applications. BlindSquare [7] is MIPsoft's GPS navigation software for iPhone and iPad. It differs from other navigation applications by using crowd sourced data; it uses Foursquare for points of interest and OpenStreetMap for street information. Jeff Bigham of CMU pioneered the work on label reading using crowdsourcing in his Interactive Crowd Support system, VizWiz [5]. His team has collected thousands of images sent by visually impaired people. Researchers at Smith-Kettlewell Eye Research Institute use crowdsourcing for movie annotation for the blind [4]. However, the current research and services focus on static image queries which are inadequate for real-time assisted navigation.

So far the technologies in vision recognition are not reliable enough for a successful applications of automated navigation. On the other hand, smartphone video stream over Internet has been a mature technology, and humans are far more reliable than machines in recognizing situations for daily navigation and reading. Such real time video streaming services include Skype, Google Hangouts, WebEx, etc.

Crowdsourcing has been proved to be an effective way to collect large-amount of labels for many machine learning tasks [8, 9]. A key element in crowdsourcing is how to aggregate the noisy labels. Popular choices include average aggregation, majority voting, and minimax entropy based approaches, etc. [10, 11, 12]. We plan to further tailor these existing techniques to address the unique challenges in crowd-assisted navigation, such as the smoothness of label reliability of the volunteers, and the contextual information of video frames.

III. System Designs

We propose a crowdsourcing approach to multimedia data sharing and services to the navigation of visually impaired. An informal survey at the Lighthouse Guild of individuals who are blind found a high level of interest in such a device. Some respondents likened the proposed technology to having a "seeing-eye person". These potential users suggested that this technology would be useful for navigating unfamiliar areas, finding entrances and exits, identifying transportation options (finding the correct bus, navigating train stations and airports), and assistance with navigating sidewalks and street crossings. Figure 1 illustrates the data flows and possible processing of the services. The goal of the work is to provide crowd services that are user accessible (especially for visually impaired), flexible (with friendly HCI and APIs for the ease of plugging in new apps to motivate online volunteers for their services), and efficient (near real time response, and a balanced workload between mobile phone, the back end system, and the different types of users). In our research, we use the onboard sensors of a COTS smartphone (iPhone or Android Phone), such as camera, compass, GPS, and accelerometer, to assist the navigation of a blind user. The basic function of the mobile computing is to stream the video and other sensory information to the crowd server so that volunteers can use the information to provide service. In assisted navigation for the blind, volunteers send back their feedback via voice or typing and the crowd program combines the results to provide the final feedback to the blind user, through voice, vibration, or the combination of them, depending on what the tasks are.

In cases where there are more than one volunteer, each of them might possibly give a different instruction to the blind user. Some of the instruction might also be from machine vision algorithms (Figure 1) that provide direction information. Here, we want to aggregate all the available instructions into a single one that will be returned to the blind user.

In addition to the on-line process and data aggregation in the above and service evaluation with users, an offline analysis

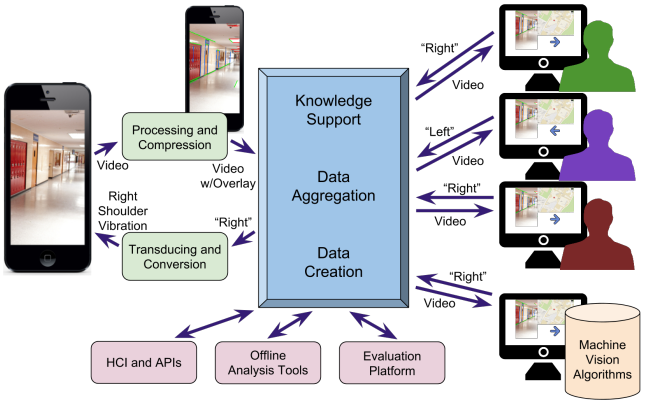

Figure 1. Dataflow in the application

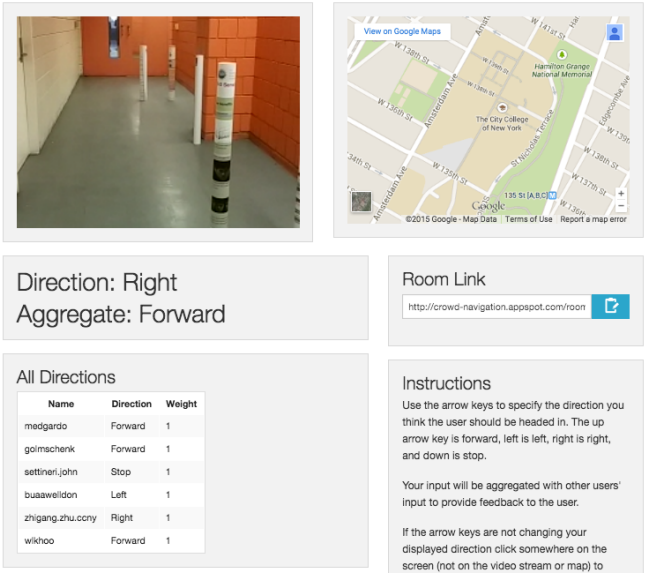

Figure 2. A screenshot of the webapp

will in turn help better tailor our context-aware human-computer interfaces and further improve the online analysis tasks; synergy among users, volunteers and the system.

We have set up a simple crowd navigation testing site (http://crowd-navigation.appspot.com/) using the Google App Engine platform in conjunction with a media server, showing that it is possible to use real-time video streaming (we use TokBox [http://tokbox.com/]) to assist a blind user to navigate by real-time feedback from volunteers online. In the following, we will discuss two main components: wedapp design and data aggregation.

A. Webapp Design

To effectively utilize the information from the crowd, a system had to be developed which would allow the instructions from the crowd to be examined, aggregated, and returned to the user in a timely manner. Furthermore, the system had to be easy and interesting to use for both the volunteers and

the users. To accomplish this, we have developed a webapp through Google App Engine. Users are able to log into the webapp using a regular Google account. The visually impaired users then have the option to create a uniquely identified video stream which we refer to as a "room". When sighted volunteers log onto the service, they can enter any of the existing rooms and provide instructions. For any given room, all the instructions from all the users are collected and aggregated via various methods. Each of the users can then be "graded" on their input and given "points" for doing a good job or docked them for malicious behavior. These points can be used in later instances to give more or less weight to each user's feedback. A screenshot of the website can be seen in Figure 2.

B. Aggregation

One of the major concerns and areas of focus for our study is how the user feedback is aggregated. With a crowd of users providing instruction, we have to be careful how this information is relayed to the visually impaired user.

The naive approach would just be to simply relay every instruction given from the crowd directly back to the visually impaired user. This of course would lead to an overwhelming amount of feedback, possibly conflicting with each other. Many of the crowd members may have different plans as to how the user should proceed and the constant changing of the instruction will be no help at all.

A more reasonable choice would be to take the aggregation of the instructions given from the crowd and send that back to the user. This way, only the primary opinion comes through to the user. There is of course the issue of how a single feedback message should be calculated from the alternatives provided by multiple volunteers. One option is to assume that all crowd members' instructions are valid. However, this may not always be true as some instructions may have been submitted after a time delay that makes them no longer relevant to the current situation. Instead, the average over a given time interval relative to the time of the user's request should be considered. Every piece of feedback will be based on the directions given in the relevant time interval. Of course, this raises its own issues as the time length and delay now plays a major role. Too short of a time interval will result in many cases where there is either no input during the interval or a single user's input is the only one considered. This leads to a problem that is almost identical to the naive approach above. On the other hand, too long of an interval means that the visually impaired user will not receive feedback until significantly after it is requested, and likely needed.

Another alternative for the aggregation is the use of a legion leader [13]. This approach still gathers all the instructions over an interval of time, but does not to send back a voted instruction to the visually impaired user. Instead, all the given instructions are considered and the crowd member who mostly closely matched the overall opinion of the crowd is chosen as the "leader" for the next time interval. The leader is given complete control during that time interval and only the leader's instructions are returned to the visually impaired user. This approach has several advantages. First, the feedback is immediately sent to the user. When the leader enters a command, there is no need to wait for the end of a time interval to send it to the user. Also, there is no problem with conflicting plans for how to proceed. One crowd member is given complete control for a short period of time, so it doesn't matter if about half the crowd think the person should go right around a pole while the other half think the user should go left. Only the person who is currently the leader picks and by the time they are no longer in control, the best choice for the route around the obstacle will probably be decided.

Table I
AGGREGATION TEST RESULTS

Maze #	1	2	3	4
Simple sum time (s)	221.64	180.27	292.86	322.79
Legion leader time (s)	219.65	182.41	263.50	228.89

IV. PRELIMINARY TESTS

Our first experiment was designed to test the aggregation method. This experiment consisted of a crowd of volunteers directing an avatar through virtual mazes in a virtual environment, whose video was streamed to the crowd. The application's feedback was used as direct commands for which direction the avatar should proceed in. The duration of the completion of the mazes were compared based on which aggregation method was used.

During the virtual reality experiment, 11 volunteers participated. The test consisted of 8 maze runs, where 4 mazes were tested once with the simple sum aggregation approach and once with the legion leader approach. While this is a small sample size, these preliminary results show that the difference in aggregation choice may significantly improve completion times. The results of this experiment can be seen in Table I. Overall, it seems that aggregation using the legion leader approach produced better performance than using the simple sum.

The following two experiments consisted of real humans being directed by the application. The first of these two experiments was a qualitative proof-of-concept. It consisted of users walking with their eyes closed from one room, around a U-turn shaped hallway, and entering another room. A single, experienced crowd member was the only individual giving feedback. Each of the eight participants (users) completed two trial runs. This was a simple test of the application's feasibility. The users were timed and the number of wall contacts was recorded. The results of this test can be seen in Table II. By comparison, this experiment showed that a "training" would be helpful for improving the navigation performance: overall the times for reaching the goal were shortened in the second trial for most of the participants. On average, it was a 20.5% improvement, even though more contacts with walls were made for some participants when moving too fast.

In the second of these real-world tests, an obstacle course was designed for the users to walk through and blackout

Table II
USERS BEING DIRECTED FROM ONE ROOM TO ANOTHER.

User	Trial 1 Time(s)	Trial 1 Contacts	Trial 2 Time(s)	Trial 2 Contacts	Improve
1	160	1	105	1	34.4%
2	129	0	102	0	20.9%
3	115	1	70	5	39.1%
4	85	0	92	1	7.6%
5	130	0	132	0	1.5%
6	103	1	103	0	0%
7	197	3	114	1	42.1%
8	120	0	108	2	10.0%
Average	129.9	0.75	103.3	1.25	20.5%

Table III
BLINDFOLDED USERS WALKING THROUGH AN OBSTACLE COURSE (I.T.: INDIVIDUAL TRIAL; C.T.: CROWD TRIAL).

user	I.T. Time(s)	I.T. Contacts	I.T. Course	C.T. Time(s)	C.T. Contacts	C.T. Course
1	75	1	1	71	0	6
2	79	0	7	112	0	2
3	46	0	6	40	2	1
4	69	0	5	79	1	5
5	69	0	4	130	3	3
6	69	0	3	96	0	8
7	77	1	8	60	2	7
8	53	0	5	73	0	4
Ave.	67.1	0.25	-	82.6	1	-

blindfolds were used. Eight participants ran two trial runs through the obstacle course with one of the runs being guided by an individual, experienced crowd member (assumed to be giving near perfect directions) and another run being directed by a crowd of 6 "untrained" volunteers. During each run a random obstacle course configuration was chosen (one as shown in Figure 2 , which the user being directed was not allowed to inspect before the navigation.

All of the crowd trials directing people through the obstacle course were run using the simple sum aggregation approach. This aggregation method was chosen because it's the simpler of the two aggregation methods and will provide a baseline for future experiments. The completion time as well as the number of wall contacts (both against the fake and real walls) is shown in Table III. In the table, the trial time, numbers of contacts and the trial courses for each participant are listed, under both the individual guidance and the crowd guidance. A wall contact only required that the user touch the wall to be counted. Also note that the two courses for each participant was different. Even though the test sample size is small, it seems that a small crowd of only six members generated slightly less reliable results than a single experienced individual (which is close to the legion leader approach), in terms of both the finishing times and numbers of wall contacts. Further study is needed to evaluate the crowd navigation performance with various changing factors of the crowd, as well as aggregation methods.

V. CONCLUSION AND DISCUSSION

Mobile devices have become ubiquitous, including among the visually impaired. Our Crowd-Assisted Navigation app is designed for smartphones, including both iPhones and Android phones, which are by far the most commonly used mobile device among those with visual impairment.

While the ability of an app to help guide visually impaired users is a valuable goal in its own right, there are various areas of research which this app will help advance. The app alone requires further study, experimentation, and evaluation of how to best aggregate large quantities of data from many users and how to most appropriately feed this information back to the visually impaired user.

VI. ACKNOWLEDGMENTS

This work is supported in part by NSF EFRI # 1137172, Dept of Veterans Affairs Rehabilitation R&D (WS), and a PSC-CUNY Award. We also thank Mr. Wai Khoo for providing the virtual environment design and test, and other CUNY students in helping the crowd navigation tests.

REFERENCES

[1] World Health Organization, "Visual impairment and blindness," http://www.who.int/mediacentre/factsheets/fs282/en/, accessed: March 2015.

[2] National Federation of the Blind, "Blindness statistics," https://nfb.org/factsaboutblindnessintheus, accessed: March 2015.

[3] National Eye Insititute, "Blindness, statistics, and data," http://www.nei.nih.gov/eyedata/blind.asp, accessed: March 2015.

[4] J. Miele, "Dvx: The descriptive video exchange," *The 27th Annual International Technology and Persons with Disabilities*, 2013.

[5] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, and T. Yeh, "Vizwiz: Nearly real-time answers to visual questions," in *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '10. New York, NY, USA: ACM, 2010, pp. 333–342. [Online]. Available: http://doi.acm.org/10.1145/1866029.1866080

[6] C. Asakawa, "Creating geo-voice-tags for and by the blind," *Workshop on Environmental Sensing Technologies for Visual Impairment*, 2013. [Online]. Available: https://nfb.org/factsaboutblindnessintheus

[7] BlindSquare, http://blindsquare.com, accessed: March 2015.

[8] R. Snow, B. O'Connor, D. Jurafsky, and A. Y. Ng, "Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 254–263. [Online]. Available: http://dl.acm.org/citation.cfm?id=1613715.1613751

[9] P. Welinder, S. Branson, S. Belongie, and P. Perona, "The multidimensional wisdom of crowds," in *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Eds., 2010, pp. 2424–2432.

[10] V. S. Sheng, F. Provost, and P. G. Ipeirotis, "Get another label? improving data quality and data mining using multiple, noisy labelers," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '08. New York, NY, USA: ACM, 2008, pp. 614–622. [Online]. Available: http://doi.acm.org/10.1145/1401890.1401965

[11] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds," *J. Mach. Learn. Res.*, vol. 11, pp. 1297–1322, Aug. 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1756006.1859894

[12] D. Zhou, J. Platt, S. Basu, and Y. Mao, "Learning from the wisdom of crowds by minimax entropy," in *Advances in Neural Information Processing Systems (NIPS)*, December 2012. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=175659

[13] W. S. Lasecki, K. I. Murray, S. White, R. C. Miller, and J. P. Bigham, "Real-time crowd control of existing interfaces," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '11. New York, NY, USA: ACM, 2011, pp. 23–32. [Online]. Available: http://doi.acm.org/10.1145/2047196.2047200