Mobile Vision as Assistive Technology for the Blind: An Experimental Study

Roberto Manduchi

Department of Computer Engineering
University of California, Santa Cruz
manduchi@soe.ucsc.edu

Abstract. Mobile computer vision is often advocated as a promising technology to support blind people in their daily activities. However, there is as yet very little experience with mobile vision systems operated by blind users. This contribution provides an experimental analysis of a sign-based wayfinding system that uses a camera cell phone to detect specific color markers. The results of our experiments may be used to inform the design of technology that facilitates environment exploration without sight[1].

1 Introduction

There is increasing interest in the use of computer vision (in particular, mobile vision, implemented for example on smartphones) as assistive technology for persons with visual impairments [7]. Advances in algorithms and systems are opening the way to new and exciting applications. However, there is still a lack of understanding of how exactly a blind person can operate a camera-based system. This understanding is necessary not only to design good user interfaces (certainly one of the most pressing and as yet unsolved problems in assistive technology for the blind), but also to correctly dimension, design and benchmark a mobile vision system for this type of applications.

This paper is concerned with a specific mobile vision task: the detection of "landmarks" in the environment, along with mechanisms to guide a person towards a detected landmark without sight. Specifically, we consider both the *discovery* and *guidance* components of this task. Consider for example the case of a blind person visiting an office building, looking to find the office of Dr. A.B. He or she may walk through the corridor (while using a white cane or a dog guide for mobility), using the camera phone to explore the walls in search of a tag with the desired information. This is the *discovery* phase of sign-based wayfinding. Suppose that each office door has a tag with the room number and

[1] The project described was supported in part by Grant Number 1R21EY021643-01 from NEI/NIH and in part by Grant Number IIS-0835645 from NSF. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the NEI, NIH, or NSF.

the occupant's name. The vision algorithm implemented in the camera phone can be programmed to identify tags and read the text in each tag.

Once a target has been detected, and the user has been informed (via acoustic or tactile signal) of its presence, the user may decide to move towards the target guided by the vision system. In some cases, only the general direction to the target is needed (e.g., the entrance door to a building). In other cases, more precise guidance is required, for example to reach a specific item on a supermarket's shelf, a button on the elevator panel, or to get closer to a a bulletin board in order to take a well-resolved picture of a posted message which can then be read by OCR. This *guidance* task calls for the user to maintain appropriate orientation of the camera towards the target as he or she moves forward, ensuring that the target remains within the camera's field of view so that its presence and relative location can be constantly monitored. Although trivial for a sighted person, this operation may be surprisingly challenging for a blind person.

This paper presents a user study with eight blind volunteers who performed discovery and guidance tasks using a computer vision algorithm implemented in a cell phone. Specific "markers", designed so as to be easily detectable by a specialized algorithm, were used as targets. Since the goal of this investigation is to study how a blind person interacts with a mobile vision system for discovery and guidance tasks, the choice of the target to be detected is immaterial: similar results would be obtained with a system designed to find other features of interest, such as an informational sign, an office tag, or an elevator button. The experiments described in this paper were inspired by a previous user study that was run on a smaller scale [8]. These previous tests turned out to be inconclusive, due to the small sample size and poor experimental design, which resulted in experiments that were too challenging for the participants to complete. The new user study presented here was more carefully designed: all participants were able to complete all tasks, yet the tasks were challenging enough that informative observations were obtained.

We note here that vision-based landmark detection is not the only technology available for blind wayfinding. Other approaches considered include the use of

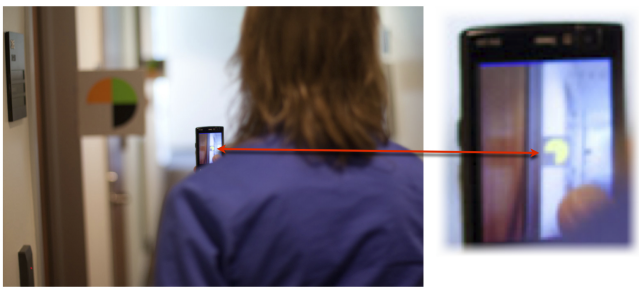

Fig. 1. Our color marker system, tested in the Env3 environment. The right image shows the view from the viewfinder; the pie-shaped color marker is displayed in yellow, signaling that detection has occurred.

active light beacons such as TalkingSigns [4], GPS [2], indoor positioning systems (e.g. Wi-Fi triangulation), inertial navigation [6] and RFID [5].

2 Experiments: Design and Outcomes

2.1 System Design

For our wayfinding tests, we selected the system proposed in [3] that uses carefully designed, pie-shaped color markers, easily detected by a camera phone with a minimal amount of computation. We used the implementation of the color marker detector for the Nokia N95 by Bagherinia and Manduchi [1]. Equipped with an ARM 11 332 MHz processor, the Nokia N95 is certainly not a state of the art platform; however, its processing rate (about 8 frames per second on VGA-resolution images) was fast enough for our experiments. The system can reliably detect markers with diameter of 16 cm at a distance of about 3.5 meters, and is insensitive to rotations of the cell phone around the camera's optical axis by up to $\pm 40^\circ$. We observed virtually no false alarms during our tests. By printing the same color marker with spatially permuted colors, we were able to obtain a variety of markers with ID embedded in the color permutation.

The feedback provided by the detection system to the user is in the form of an acoustic signal (a sequence of "beeps"). There are two distinct beeping rates: a slower rate (about 2 beeps per second) at distances beyond 1 meter, and a faster rate (about 5 beeps per second) at closer distances. The pitch of the beep is kept constant, while its volume depends on whether the target is within the central third of the image (higher volume) or in the left or right third of the image (lower volume). This allows the user to figure out the approximate bearing angle to a detected marker. In previous preliminary tests we experimented with richer types of interface (e.g., multiple sound pitch), but the feedback we received from blind users led us to select the simple and "minimalistic" interface described above. Indeed, all participants to this user study commented positively on the chosen interface. Finally, the cell phone vibrates if the user rotates the cell phone by more than 30° around its optical axis. This provides a discreet warning and reminds the user to keep the cell phone straight up.

2.2 Experiment Design

We considered three environments that were representative of a variety of realistic indoor situations. The first environment (Env1) was a wide (4 meters by 5 meters) hall opening onto a corridor with the markers attached to two opposing walls. The starting position for each trial was in the middle of the open side of the hall. Note that some of the markers could be detected (if aiming in the correct direction) already from the starting point. The second environment (Env2) was a fairly wide corridor (about 2 meters in width), with markers placed flat on just one wall at several meters of distance from each other (some but not all located near office doors). The participants were informed of which wall the

Fig. 2. Bind volunteers during our experiments in the Env1 (left) and Env2 (right) environments.

markers were attached to. The starting position was at either end (alternating) of a stretch about 20 meters long. Two markings ("fiducials") were taped to the floor at a distance of 3.25 meters from each other at one end of the test stretch, defining a "probe" segment. The walking speed of each participant during the tests was measured by recording the time at which he or she crossed each fiducial. The third environment (Env3) was a narrower corridor (1.6 meters wide), with markers attached by velcro strips so that they would jut out orthogonally from the wall (see Fig. 1). Copies of the same marker were attached to both faces of a piece of cardboard, allowing it to be seen in fronto-parallel view from either side of the corridor. The participants were instructed to start walking from either end (alternating) of a 15 meters long stretch. As in the previous case, fiducials were placed on the floor to measure the participant's walking speed.

Eight blind volunteers (two men and six women, aged 50 to 83) participated in our experiments. Only one of them was congenitally blind; the others lost their sight at various stages in life. All but one of the participants had only at most some light perception; the remaining participant had enough sight to recognize a marker at no more than a few centimeters of distance. Two participants were already familiar with the system, having tried it one year earlier, while the other six were new users. Three participants used a guide dog during the experiments; everyone else used a white cane, except for one who elected to walk without any assistance.

Each participant was read the IRB-approved informed consent form and was given the opportunity to ask questions afterwards. The participants demographic details were filled in by the investigators, and the participant signed the consent form (which included permission to use pictures taken of them in scientific publications). After these preliminary instructions, the participant was taken to each one of the chosen environments in turn (the order of the environments in the test was chosen randomly for each participant). The participant was explained the correct usage of the system and was given ample time to experiment with a test marker at a known location within each environment. The participant then completed a "dry run" sequence of at least eight trials. During the dry run, the participant was allowed to ask questions; some general recommendations were

also offered by the investigator supervising the experiment. Each participant was allowed to continue the dry run trials until he or she felt comfortable with the system.

After the dry run phase, the official test began. Each test comprised eight trials. In each trial, the participant was led by hand to the starting location for the current environment and asked to search, using the cell phone, the camera pointing forward, for one specific marker (whose location was unknown to the participant). Rather than manually changing the location of the marker at each trial, we placed five markers in different position of the wall at the beginning of the experiment, and programmed the cell phone so as to only detect one specific marker at each trial. The sequence of marker IDs to be detected during the trials was chosen randomly for each environment (the same sequence was used for all participants at that environment).

At each trial, the investigator used a stopwatch to record the time at which the phone first beeped after detecting a marker, and the time at which the participant touched the marker (which concluded the trial). The walking speed of the participant during the trial was also measured in Env2 and Env3 using the fiducials on the floor, as explained earlier. If a participant was not able to complete a trial in Env1 within a period if five minutes, or walked past the designated marker without finding it in Env2 or Env3, the trial was declared unsuccessful. The total experiment (including initial training) took between three and four hours per participant.

2.3 Results

Results from the tests are shown in Fig. 3 for the three different environments considered. Each figure reports the median *guidance time*, defined as the time between the first beep (when the system first detected the marker) and the time at which the participant touched the marker. The number of unsuccessful trials (if any) is also reported in each figure. For Env2 and Env3, we also reported the average *probe time*, that is, the time it took to each participant to walk trough the 3.25 meter probe.

One thing that results apparent from the plots is that the median probe time was, in general, quite smaller than the median guidance time. Considering that the target was detected at no more than 3.5 meters of distance, and often at a shorter range[2], it results clear that the participants walked faster during the "discovery" phase (as measured by the probe time) than during the "guidance" phase. In fact, guidance often proved to be a long and painstaking process.

Different environments called for different search strategies. In the case of Env1, a few participants methodically explored all walls in the hall (keeping at an approximate constant distance from the wall) until they came upon the marker. Others participants would move towards the center of the room, slowly rotating the camera to obtain a panoramic view of the space. One participant

[2] This was especially the case for the case of markers placed flat on a corridor's wall (Env2), and thus seen from a slanted angle.

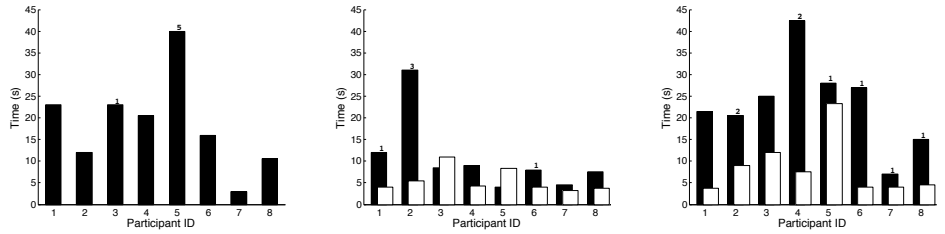

Fig. 3. The median guidance time (black bars) and median probe time (white bars) for the environments considered. The number of unsuccessful trials (if any) for each participant is shown by a number on top of the bar.

(who did not use a cane or guide dog during the trials) experienced serious difficulty in this environment, as she would soon get disoriented. Indeed, self-location awareness is important for successful exploration and discover (as noted in the post-test interview by two other participants). This seems to be less of a concern during the guidance phase, in which case only one's relative location with respect to the marker needs to be controlled.

Env2 had markers only on one side of the corridor, placed flat on the wall. Successful detection required walking at a specific distance to the wall, holding the cell phone at an approximately constant angle. Due to the geometry of this environment, the target was typically detected at a closer distance than in Env1 or Env3 (in which case the marker could be found facing the participant). This explains why guidance time was in most cases smaller for Env2 than for the other environments. One participant using a guide dog pointed out that maintaining the desired location in the corridor was challenging, as the dog was trained to walk at a specific distance from the wall. At the same time, this participant remarked that the guide dog helped her maintaining a straight directions, often a difficult task (even in a corridor) without sight.

Env3 was considered the most challenging by five participants (while five participants considered Env2 to be the easiest one). This came somewhat as a surprise, as we originally thought that the fronto-parallel geometry of the marker placement would simplify both detection and guidance. In fact, median guidance times in Env3 were almost always higher than for the other environments (with a median value of 23 seconds, versus 18.25 seconds for Env 3 and 8 seconds for Env2). The probe times in Env3 were also higher than in Env2 (with a median value of 6 seconds, compared to 4.12 seconds for Env2), showing that the participants preferred to walk slower in this environment. Part of the difficulty was that in Env3 the markers were found on both walls, and participants were asked to explore both walls as they proceeded. This required methodically scanning the scene by rotating the cell phone around its vertical axis, an operation that several participants found challenging. Once the cell phone beeped signaling a detection, some participants had trouble understanding whether they should

keep searching on the left or on the right wall. This is not surprising, given the relatively large field of view of the camera compared width the small width of the corridor.

From the answers to the post-test questionnaire, several common themes emerged. Participants seemed to think that the system worked well for what it was supposed to do: on a scale from 0 to 5 (where 0 meant "did not work" and 5 meant "worked perfectly"), the average score was 4.25. This was certainly very encouraging. Two participants commented that they would prefer a wider field of view. The need for rotating the cell phone to search for a target was also commented negatively by some. These two aspects are clearly related: a larger field of view would require less active interaction, since the marker could be found without having to constantly rotate the phone. Several participants commented positively on the fact that the markers were all at the same height. Indeed, earlier experiments [8] with markers at different heights resulted in very poor performance.

Several participants commented on the importance of keeping the phone at the correct height and orientation. Indeed, we observed that at least two participants had serious difficulty with holding the phone properly. For example, Participant 5 found that she had to hold the cell phone "locked" onto her shoulder (see Fig. 2), and thus would rotate her whole body when looking for a target. Other participants, however, showed good wrist control, which enabled effective discovery and guidance. Understanding the correct direction to a detected target was also challenging for some participants. For example, one participant observed that she was constantly misestimating the location of the marker by two feet or so.

Finally, almost all participants commented positively on the chosen interface, but noted that it may be impractical to use it if other people were nearby (who could be annoyed by the beeping). Most participants appreciated the information provided by the interface: approximate distance to the target (through the beeping rate) and bearing angle (through beeping loudness). Indeed, when asked to describe their guidance strategy in words, most participants said that they tried to always aim the cell phone so that the beeping was loud (signaling that the marker was seen straight ahead). However, at least two participants seemed to confuse the role of two features (beeping rate and volume). This confirms our previous observations that rich interfaces may easily become too complex, especially when one is already concentrated in other mobility tasks (e.g., avoiding obstacles).

3 Conclusions

Our experiments have resulted in a number of interesting (and at times unexpected) observations, which may inform the design of future wayfinding systems mediated by computer vision. We summarize our main conclusions below.

Field of view: The limited field of view of typical camera phones forces the user to actively explore the environment in search of a target, an operation that may

be challenging for some people. A natural solution would seem the use of shorter focal lengths (and thus wider field of view). It should be noted, however, that a wider field of view reduces the angular resolution of each pixel and thus the distance at which a target of given size can be found.

Camera placement: Several participants found the use of a hand-held camera to explore the environment difficult, and some observed that they would prefer the camera to be attached to their body or their garment. Further investigation is necessary to establish whether a wearable camera could be used for effective exploration

Target location: Our experiments have shown that, even with an "ideal" system with carefully designed targets, detection and guidance can be difficult and time-consuming in some environments. This suggests that the environment layout and target location have an important role in the success of vision-based systems for information access and wayfinding without sight.

References

1. H. Bagherinia and R. Manduchi. Robust real-time detection of multi-color markers on a cell phone. *Journal of Real-Time Image Processing*, June 2011.
2. J. Brabyn, A. Alden, H.-P. G., and M. Schneck. GPS performance for blind navigation in urban pedestrian settings. In *Proc. Vision 2002*, 2002.
3. J. Coughlan and R. Manduchi. Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. *International Journal on Artificial Intelligence Tool*, 18(3):379–397, 2009.
4. W. Crandall, B. L. Bentzen, and L. Meyers. Talking signs®: Remote infrared auditory signage for transit, intersections and ATMs. In *Proceedings of the CSUN*, Los Angeles, CA, 1998.
5. V. Kulyukin, C. Gharpure, J. Nicholson, and S. Pavithran. RFID in robot-assisted indoor navigation for the visually impaired. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS '04, 2004.
6. Q. Ladetto and B. Merminod. An alternative approach to vision techniques - pedestrian navigation system based on digital magnetic compass and gyroscope integration. In *Proc. WMSCI*, 2002.
7. R. Manduchi and J. Coughlan. (Computer) vision without sight. *Commun. ACM*, 55(1), 2012.
8. R. Manduchi, S. Kurniawan, and H. Bagherinia. Blind guidance using mobile computer vision: A usability study. In *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, 2010.