Crosswatch: a System for Providing Guidance to Visually Impaired Travelers at Traffic Intersections

James M. Coughlan and Huiying Shen

The Smith-Kettlewell Eye Research Institute
San Francisco, CA
coughlan@ski.org, hshen@ski.org

Structured Abstract.

Purpose – This paper describes recent progress on the "Crosswatch" project, a smartphone-based system developed for providing guidance to blind and visually impaired travelers at traffic intersections. Building on past work on Crosswatch functionality to help the user achieve proper alignment with the crosswalk and read the status of walk lights to know when it is time to cross, we outline the directions Crosswatch is now taking to help realize its potential for becoming a practical system: namely, augmenting computer vision with other information sources, including geographic information systems (GIS) and sensor data, and inferring the user's location much more precisely than is possible through GPS alone, to provide a much larger range of information about traffic intersections to the pedestrian.

Design/methodology/approach – The paper summarizes past progress on Crosswatch and describes details about the development of new Crosswatch functionalities. One such functionality, which is required for determination of the user's precise location, is studied in detail, including the design of a suitable user interface to support this functionality and preliminary tests of this interface with visually impaired volunteer subjects.

Findings – The results of the tests of the new Crosswatch functionality demonstrate that the functionality is feasible in that it is usable by visually impaired persons.

Research limitations/implications – While the tests that were conducted of the new Crosswatch functionality are preliminary, the results of the tests have suggested several possible improvements, to be explored in the future.

Practical implications – The results described in this paper suggest that the necessary technologies used by the Crosswatch system are rapidly maturing, implying that the system has an excellent chance of becoming practical in the near future.

Accepted for publication in Journal of Assistive Technologies, 2013.

Originality/value – The paper addresses an innovative solution to a key problem faced by blind and visually impaired travelers, which has the potential to greatly improve independent travel for these individuals.

Keywords: Visual Impairment, Blindness, Assistive Technology, Traffic Intersection, Pedestrian Safety

Article Type: Research Paper

1 Introduction and Related Work

Crossing an urban traffic intersection is one of the most dangerous activities of a blind or visually impaired person's travel. Several types of technologies have been developed to assist blind and visually impaired individuals in crossing traffic intersections, which is crucial for facilitating independent travel for these individuals. Most prevalent among them are Accessible Pedestrian Signals, which generate sounds signaling the duration of the walk interval to blind and visually impaired pedestrians (Barlow *et al,*. 2003). In addition, Talking Signs® (Crandall *et al,*. 2001) allow blind travelers to locate and identify landmarks, signs, and facilities of interest at intersections and other locations, using signals from installed infrared transmitters that are converted to speech by a receiver carried by the traveler.

However, the adoption of both Accessible Pedestrian Signals and Talking Signs® is very sparse, and they are completely absent in most cities. Bluetooth beacons have also been proposed (Bohonos *et al.*, 2008) to provide real-time information at intersections that is accessible to any user with a standard mobile phone, but like Talking Signs® this solution requires special infrastructure to be installed at each intersection.

More recently, several smartphone apps have been devised to provide information about traffic intersections to blind and visually impaired travelers, including Intersection Explorer and Nearby Explorer for Android and Intersection and Sendero GPS LookAround for iPhone. Such apps are promising but are incapable of providing detailed information about the user’s precise location and orientation relative to a crosswalk or the current status of signal lights. A very ambitious and thoroughly tested project, the Mobile Accessible Pedestrian Signal (MAPS) system, has been created by Mr. Liao of the Univ. of Minnesota’s Traffic Observatory (Liao, 2012) to provide signal and intersection geometry information to smartphone users at signalized intersections. Again, however, special hardware infrastructure needs to be installed at each (signalized) intersection to provide information about the user’s precise location and the status of signal lights.

A relatively small amount of work has been done on computer vision algorithms for detecting crosswalks and traffic lights for the blind and visually impaired (Utcke, 1998; Se, 2000; Chung *et al.* 2002; Se and Brady, 2003; Aranda and Mares, 2004;

Uddin and Shioyama, 2005). The great advantage promised by the computer vision approach is that it interprets *existing* visual cues, such as crosswalk marking patterns and walk signal lights, without requiring any additional infrastructure to be installed at each intersection. More recently, the Crosswatch system (Coughlan and Shen, 2012) and a similar system (Ahmetovic *et al.*, 2011) have implemented such computer vision functionality on a smartphone platform, both of which provide real-time feedback to a blind or visually impaired user about their orientation and location relative to a crosswalk; experiments with both systems used by blind subjects have been reported, demonstrating the feasibility of the computer vision-enabled smartphone approach.

We note that the Crosswatch smartphone platform has been extended to include detection of two types of crosswalk markings (Fitzpatrick *et al.*, 2010), the continental (striped "zebra") crosswalk (Ivanchenko *et al.*, 2008; Fig. 2b) and the two transverse lines (two-stripe) crosswalk (Ivanchenko *et al.*, 2009; Fig. 2a), the second of which is more common and less highly visible than the continental (and which has not been tackled by other computer vision systems for blind and visually impaired persons). The same Crosswatch platform has also been configured to detect the presence of walk signal lights in real time (Ivanchenko *et al.*, 2009). Experience with Crosswatch suggests that it can be extended to incorporate a variety of functions to provide useful guidance to visually impaired travelers at traffic intersections.

2 Proposed Approach

We propose extending the Crosswatch system to obtain a broader range of information about traffic intersections, which may be categorized as "what", "where" or "when" information:

- "What" information includes not only the presence of crosswalks in an intersection and the type of intersection (e.g., four-way or T-junction) but also the presence of any signal lights (which may include traffic lights), important signs such as stop signs, walk buttons, raised medians and a variety of other important features.
- "Where" information includes the location of any crosswalks or other features listed above, which can be obtained from smartphone sensors in absolute geographic terms (i.e., latitude/longitude coordinates and bearing relative to north). To be useful to the traveler, it must be translated in terms relative to the user's location and bearing at each moment (e.g., to guide him/her to the entrance of the crosswalk).
- "When" information specifies the real-time status of walk lights or other traffic lights.

Previous work on Crosswatch has attempted to answer some of the "what" and "when" questions on a smartphone platform using computer vision algorithms. However, many "what"-type features are extremely challenging to determine solely through computer vision alone. For instance, walk buttons appear in a great variety of forms: some are small, recessed buttons while others are large and protruding; the signs labeling them appear in different colors and may contain text, graphics or both. Similarly, a raised median (Fig. 1) is hard to discern without detailed knowledge of the three-dimensional surface geometry (since the median is elevated relative to the road surface but otherwise looks similar to the road surface), and failure to detect the median can cause gross confusion about the length of the crosswalk. (The term “median” is used in North America, and is referred to elsewhere by other names, such as “central reservation”). We note that complex intersections pose the biggest challenge to visually impaired pedestrians – and are the intersections where assistive technology such as Crosswatch is most needed. Unfortunately, it is precisely these intersections that also pose the biggest challenge to computer vision algorithms!

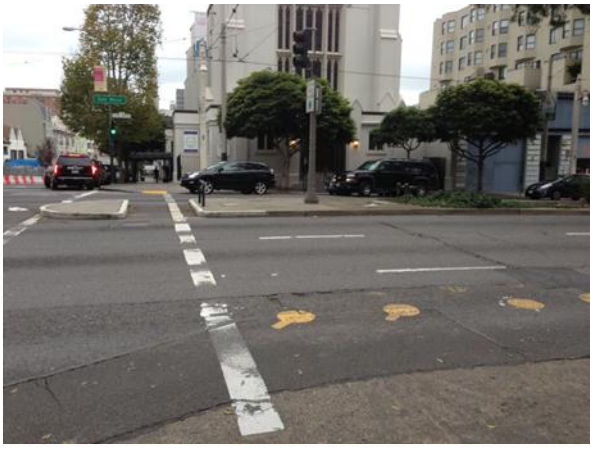

Figure 1. Some important intersection features are difficult to detect reliably using computer vision at typical viewing distances. One such example is the raised median (shown in picture), which is an elevated portion of a street that often serves as a place for pedestrians to pause while traversing a crosswalk. We propose that it is more practical to determine the presence of the median and a variety of other intersection features using data from a GIS.

Accordingly, we plan to focus on augmenting computer vision with other information sources, especially geographic information systems (GIS), which associate data with a given geographic location, and sensor data. For instance, given the pedestrian’s current location (GPS specifies location with enough accuracy to determine the nearest intersection) and bearing (indicated by the smartphone compass), a GIS can look up a host of information associated with that specific intersection, such as the intersection layout (including crosswalk lengths and directions), the presence and location of signs, crosswalks, signals, walk buttons and medians (or other specific features). We are currently researching the types of GIS data already available about traffic intersections (e.g., through municipal/transit data sources, OpenStreetMap, Google Maps, and other commercial sources). Crowd-sourcing approaches (see Sec. 5 for more details) may be the most practical way of adding to this data in the future, which would allow volunteers to contribute information about the intersections they are familiar with (and to focus on the intersections that are the most challenging to navigate).

Computer vision is still indispensable for certain information provided by the system, specifically, the pedestrian’s orientation relative to the crosswalk (i.e., detailed location information which GPS resolution is insufficient to determine), and the status of a walk (or traffic) light, for which no reliable non-visual cues exist. The detailed location information provided by computer vision can also be combined with GIS and sensor information to deduce information such as where the user is standing relative to the walk button, and thereby help the user find the button.

3 User Interface Design

An intersection analysis system must be designed with a user interface that allows a blind or visually impaired person to acquire and access whatever “what,” “where” and “when” information about the intersection is desired. In order for this information to be acquired and accessed easily and safely, the user must interact with the system in an appropriate manner. For instance, while some information about the intersection becomes known to the system anytime the user is standing close to it (e.g., the cross streets of the intersection, which can be determined using GPS), any information that must be acquired using the smartphone camera requires the user to aim the camera in an appropriate direction. Such aiming is an active process that requires real-time feedback from the system to help guide the user, which we describe in more detail in Sec. 4.

In this section we outline the user interface scheme that is planned for the entire Crosswatch system. The system will be tested repeatedly with blind and visually impaired users, whose experience with and feedback about the system will be used to implement modifications and improvements. The overall scheme is as follows:

1) As the user approaches an intersection, the approximate user location, estimated by GPS, determines the intersection and this is announced to her using text-to-speech.

2) GIS information about this intersection is looked up and reported, including all global “what” information that is available without knowing the precise location of the user at the intersection. If the user holds the smartphone in a known direction (e.g., with the camera pointed in the general direction of the intersection), then the smartphone compass also determines the approximate direction the user is pointing. We note that the GPS information can be acquired even while she is walking (in addition to the compass information for those users who are able to hold the smartphone in a consistent direction to permit a stable reading).

3) If the user wants more detailed information, including “when” or “where” information or more specific “what” information, she requests it using voice commands and/or touch screen menu commands.

4) To acquire this more detailed information, the user must approach the corner of the intersection, stop walking and “sweep” it with the smartphone camera. This will allow the system to estimate her location relative to the intersection. Details about this procedure are provided in Sec. 4.

5) The system will tell the user if she is well aligned to the desired crosswalk, and if not will guide her accordingly.

6) If a walk button is present at the current intersection corner and the user wants guidance towards it, she can request this and obtain help from the system in approaching it.

7) If the user needs “when” information, she will request this, and the system will guide her to point the camera to the appropriate signal light. It will then report the status of the light in real time (either the walk signal if this is present, or else a traffic signal if that is visible instead). Naturally, the onset of a walk signal or green traffic light provides important timing information about the traffic cycle but must not be interpreted as advice that it is now *safe* to enter and cross the intersection. The user must decide for herself when to enter an intersection based on the information provided by the system and all acoustic, tactile and other sources of information available to her.

8) Before entering the crosswalk, the user will put the smartphone away. It is vitally important that the user not be distracted by any device while walking in the crosswalk so that she can pay full attention to ambient sounds for maximum safety (Barlow *et al.* 2010).

9) If the user has not traversed the entire crosswalk but has reached a median where she can safely wait, she may wish to return to step (4) to perform a new “sweep” of the area, re-align herself to the crosswalk if necessary and/or obtain signal light timing information.

Earlier work on Crosswatch has explored several aspects of the above user interface scheme and demonstrated feasibility with blind and visually impaired users. In (Ivanchenko *et al.*, 2008; Ivanchenko *et al.*, 2009), blind users were able to use an earlier version of Crosswatch (implemented on a Symbian smartphone) to locate crosswalk patterns of two different types (the continental and two transverse lines) that they were standing near. Alignment information was also obtained by the system, indicating to the user whether he/she was standing in the crosswalk “corridor” (the rectangular path defined by the crosswalk pattern borders, extended onto the sidewalk it adjoins), to the left of it or the right.

Several blind and visually impaired users also successfully used another Crosswatch function to detect the presence of an illuminated walk signal within less than a second (Ivanchenko *et al.*, 2010). One novel user interface feature that was added to improve a visually impaired user's chances of aiming the smartphone camera in the direction of a walk signal light was a tilt feedback feature. This feature activates the smartphone vibrator any time the tilt sensor (accelerometer) indicates that the camera orientation deviates significantly from the horizontal. The purpose of this feature is to prevent users from inadvertently pointing the camera too far downwards (in a viewing direction that would not capture the walk signal or other important traffic intersection features), which was very useful for those who found it difficult to maintain a horizontal camera orientation based solely on their sense of proprioception. We describe a similar user interface feature in the next section.

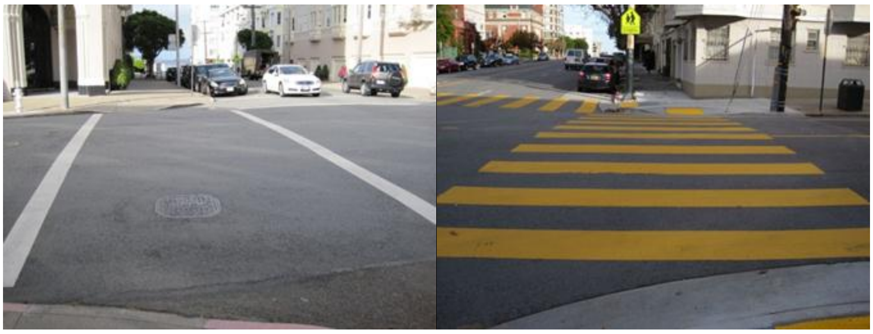

Figure 2. Two common kinds of crosswalk marking patterns. Left: two transverse lines ("two-stripe"). Right: continental ("zebra").

Many improvements are being investigated and developed for Crosswatch. Perhaps the most important enhancement we are currently developing is the ability to precisely locate the user in the intersection (relative to important features such as crosswalks), to provide more detailed "where" information (steps 3-6 above). Such functionality requires computer vision algorithms and a user interface that are substantially different from what was used in earlier versions of Crosswatch, and which are described in the next section.

Other possible improvements and variations to consider in the near future will include the following: (1) the use of bone conduction headphones as an alternative to the built-in smartphone speakers, which may be difficult to hear in noisy urban environments and which may draw unwanted attention to the user, but which interfere as little as possible with normal hearing of ambient environmental sounds; (2) making greater use of tactile feedback, which is currently only used to provide tilt feedback, but which might also substitute for some forms of audio feedback; and (3) different ways for the user to enter commands, including voice recognition, menu selection using the smartphone touchscreen and movement-based gestures (such as tilting or shaking the smartphone).

4 Precise “Where” Information: Overall Approach and User Interface Experiments

This section describes our recent work in developing and testing a preliminary version of a new Crosswatch function to locate the user within the traffic intersection with precision far greater than that provided by GPS, which is usually only accurate to within about 10 meters in an urban environment (Brabyn *et al.*, 2002). The spatial precision we hope to attain is a meter or better, which would be more than sufficient to determine whether the user is centered within a crosswalk corridor, and can be useful in helping him or her locate an important feature such as a walk button.

4.1 Overall Approach

The ability to localize the user with such high precision requires analysis of multiple images of the environment from the user’s vantage point. (By contrast, earlier versions of the Crosswatch system analyzed one video frame of the intersection at a time, independently of other video frames.) In our proposed approach, the user “sweeps” the environment by standing in one place and rotating the camera from one side to another while holding it horizontal. While some users might only be interested in approaching and traversing a specific crosswalk in the intersection, it is useful to incorporate imagery from a wider range of directions – including views of multiple crosswalks in the intersection, if possible. Multiple, overlapping images are stitched together into a horizontal *panorama* or image mosaic (Szeliski, 2010), as shown in Fig. 3.

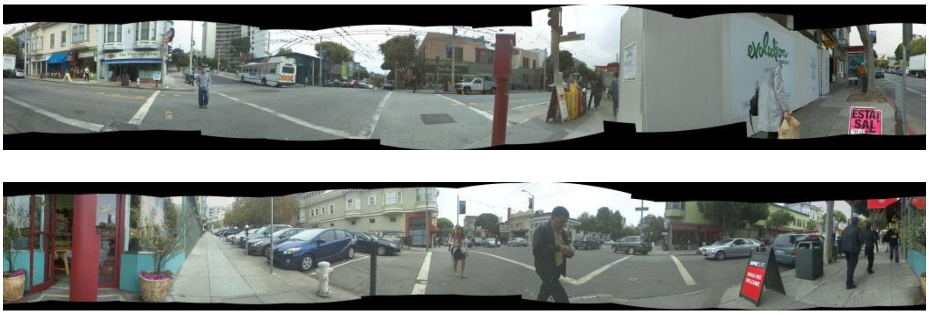

Figure 3. Image panoramas taken by blind volunteer subjects standing at the corners of four-way intersections using the Android app we developed for acquiring 360° imagery. Note that two crosswalks are visible in each panorama, which are approximately perpendicular to each other in both intersections.

The reason for acquiring this panorama is to detect all features of interest in the intersection and to locate the user’s position accurately in both the x- and y-dimensions along the ground. For instance, imagine a four-way crosswalk that is aligned perfectly to the north/east/south/west directions, with the positive x-axis pointing east and the positive y-axis pointing north. If the desired crosswalk direction runs east along the x-axis, then an image of this crosswalk alone provides accurate information about the user’s y coordinate (which determines whether or not the user is centered in the crosswalk corridor) but not the x coordinate. This is because of a basic property of parallax: shifts of the user along the y-axis manifest in the image as significant shifts in the apparent angles and locations of the crosswalk borders, whereas shifts along the x-axis produce only subtle changes in the apparent scale of the crosswalk that are harder to gauge. Conversely, a view of the other crosswalk running north along the y-axis provides accurate information about the x coordinate but not the y coordinate. Thus, views of both crosswalks taken from the same location are needed to infer accurate information about both the x and y coordinates. Moreover, given an aerial image (e.g., a Google Maps satellite image) of the intersection, whose features have known location, scale and orientation, the user’s (x,y) coordinates can be determined relative to the entire intersection.

An additional benefit of acquiring a panorama is that in many circumstances the user may not know the approximate direction to the crosswalk he/she wishes to cross (especially if the intersection is complicated or unfamiliar to the user), and so sweeping over a large range of directions ensures that the crosswalk(s) of interest will be captured by the camera, as well as other important features (such as walk or traffic signals).

We are exploring the possibility of augmenting the crosswalk features with other features in the scene to determine the (x,y) coordinates, such as corner or line features belonging to buildings and other structures visible across the intersection, as is done in a variety of self-localization techniques in computer vision (Arth *et al.*, 2011). Compared with crosswalk features on the street, these features have the benefit of being less often occluded by vehicles and pedestrians in the street, but their greater distance from the camera implies that they may provide less accurate localization information.

While a variety of smartphone apps are available to create image panoramas (indeed, an image panorama feature is now included with the iPhone), these apps are intended for users with normal vision, who can inspect the scene and decide which regions of the scene to include in the panorama. However, a blind or visually impaired user may not have sufficient vision to follow this approach to creating a usable panorama; in particular, some users with insufficient vision may inadvertently hold the camera far enough from horizontal that much or all of the intersection is not visible.

To address this problem, we devised the following simple user interface to facilitate the acquisition of a usable panorama (building on an earlier version of Crosswatch). Using the smartphone tilt sensor (accelerometer) to sense the camera orientation, the system activates the smartphone vibrator any time the orientation is sufficiently far from horizontal. Specifically, the roll (rotation about the camera’s line of sight, i.e., the angle between the horizontal frame of the image and the horizon defined by gravity) and pitch (the angle between the line of sight and the horizon defined by gravity) are constrained to be close to zero, so that only the yaw (i.e., azimuth, or bearing relative to north) can vary freely. The user is instructed to stand in place and turn slowly in a clockwise direction, making a complete circle. Using the smartphone compass (magnetometer), the system continuously estimates the user's bearing and acquires a new image every time the bearing increases by approximately 20°. Each time a new image is acquired, the system issues a brief audio tone; after the last image is acquired, the system issues an additional audio tone of lower frequency to signal that the panorama acquisition is complete.

4.2 User Experiments

The panorama acquisition function was implemented as an app on the Android LG-P990 smartphone, and was tested in a preliminary experiment with two blind volunteer subjects (both of whom have no light perception). After obtaining each subject's consent in accordance with an IRB protocol, the subjects were given brief indoors training sessions to acquaint them with the experiment and with the purpose and operation of the app. In each training session, the subject was taken to a quiet room, where he/she was shown how to hold the smartphone horizontally, move it slowly and avoid covering the lens with fingers. The experimenter demonstrated the acquisition of an image panorama, drawing attention to the audio tones issued by the system at roughly 20° intervals. The subject then acquired a few image panoramas in the room, with guidance from the experimenter, until the subject was able to perform this task independently.

Next, each subject was taken to two nearby traffic intersections, both of which are four-way intersections, one with traffic signals and one without. (The choice of intersections was fixed in advance of the experiment.) One panorama was acquired at each of the four corners of each intersection, for a total of eight panoramas per subject. During the outdoor experiments, the subject was accompanied by two normally sighted experimenters, one of whom had the sole job of ensuring the subject's safety (and intervening at any time if necessary). The approximate height of the camera above the ground, as held by the user to acquire panoramas, was measured for each subject, since this is an important geometric parameter that will be needed in future experiments. Both subjects used a white cane, which they took with them on the outdoor experiments, and which they were able to hold or put away while acquiring the panoramas.

All of the panoramas acquired in the experiment contained imagery of the traffic intersections that (a) sampled the entire 360° horizontal field from multiple, overlapping views and (b) consisted of images that were taken with the camera held roughly horizontal. These results (see Fig. 3 for sample panoramas) demonstrate the feasibility of the overall panorama approach, and more specifically of the user interface. However, the experiment did suggest the following possible improvements to be explored:

(1) In some cases the subject's finger obscured a small portion of the camera's field of view, so more attention needs to be focused on this issue during the training session.

(2) While the audio tones issues by the system at 20° intervals were easy for the subjects to hear indoors in a quiet room, they were often difficult to hear on a busy street corner. In particular, this meant that the subjects sometimes failed to hear the final lower-pitch tone signaling that the panorama was complete. Instead of using audio cues, perhaps a specific vibration pattern (using the smartphone vibrator) could be used to signal the end of the panorama image acquisition process. To avoid confusion, this vibration pattern would need to be distinct from the pattern used to signal when the camera deviates from the horizontal.

(3) One subject found it difficult to use the vibrator feedback to hold the camera horizontal, because when the vibration feedback was activated it was unclear how to right the camera to make it level. In hindsight it is now apparent that the feedback mechanism could be simplified to reflect pitch alone, without regard to roll. This simpler feedback mechanism should be easier to use because only one dimension (pitch) needs to be varied to make the vibration cease, rather than having to vary two dimensions (roll and pitch) simultaneously; the mechanism is sufficient because an image roll can always be eliminated (in software) simply by estimating the amount of roll using the tilt sensor (accelerometer) and un-rotating the image accordingly.

(4) Rather than taking a full 360° panorama, the compass could be used to estimate the approximate left and right limits needed to sample the entire traffic intersection. A user interface would then be devised to guide the user to sweep out a panorama between these two limits, which would typically span 180° or less, and which would take less time (and would be less cumbersome and disorienting) than sweeping out a full 360°. Multiple sweeps within these limits could also be taken in rapid succession, which would allow the system multiple opportunities to view portions of the intersection that are temporarily obscured (e.g., by a moving pedestrian or vehicle); a portion that was obscured by a moving obstacle in one sweep would be likely to be visible in a second sweep, thereby increasing the fraction of the intersection that is visible to the system.

In the future we will develop and test the necessary computer vision algorithms to estimate the user’s (x,y) location in the intersection from the panoramas. Once this function has been implemented, an appropriate user interface must be devised to convey to the user how to turn and/or move to align him/herself with a crosswalk, or to approach a nearby destination such as a walk button. We have had some experience with this kind of user interface in an earlier version of Crosswatch used to establish proper alignment with a desired crosswalk: the user had to first undergo a translation (shift to one side or another) so as to center him/herself in the crosswalk corridor, followed by a rotation of the torso, if necessary, to attain the proper bearing (parallel to the crosswalk corridor). Simple audio feedback indicated when each of the two stages of alignment were achieved. We plan to explore the use of this type of mechanism, as well as more elaborate alternatives, including explicit instructions (using synthetic speech, such as “shift to the left 1 meter and then turn east by 30 degrees”) and 3D spatialized sound (which requires the use of stereo headphones).

5 Discussion

In this section we discuss a variety of social, human factors and technical issues that will have an important influence on the development of Crosswatch, and on the chances of its adoption in the future.

First, we would like to emphasize that Crosswatch is grounded in the actual needs of blind and visually impaired travelers, as the conception and development of Crosswatch has been driven by ongoing consultations with members of the blind and visually impaired community, as well as with rehabilitation engineers and orientation and mobility instructors. However, we are acutely aware that this group of people is extremely heterogeneous, with widely varying needs and abilities. For instance, some blind or visually impaired travelers may prefer to use Crosswatch only for addressing complicated and unfamiliar intersections (perhaps while traveling to a new destination); other travelers may prefer to ask for help from a companion or passer-by at a challenging traffic intersection rather than rely on any high-tech tool for assistance. A traveler with partial vision may only require access to the “what” information provided by Crosswatch (or another GPS-enabled travel aid) because he/she is able to see crosswalk markings and traffic signals, while another traveler may be unable to use the system to capture usable images because of a persistent hand tremor or shake. We seek to make Crosswatch accessible and useful to as many persons with blindness or visual impairment as possible, while acknowledging that the fit between Crosswatch and each user will be unique, and that it will not fulfill every blind or visually impaired traveler’s needs.

In addition to the enormous variability among potential users, we expect to encounter a wide range of traffic intersection layouts, crosswalk markings and signal lights in different countries and regions. GIS information will need to be tailored accordingly to the characteristics of each locale. For instance, the shape of a Walk light icon may vary significantly in different regions, as will the shapes, sizes and colors of different crosswalk marking patterns; the defining shapes and colors of such features can be specified for each individual intersection in the GIS. While current Crosswatch research is conducted on intersections specifically in the San Francisco area, the overall Crosswatch concept extends to any place with paved roads and standardized crosswalk markings, and we expect that many important computer vision/sensor and user interface issues will be revealed and solved in this initial test ground. Once these issues are addressed locally, the challenges in translating Crosswatch functionality to other countries can begin to be tackled.

Another related and particularly important issue is the variable quality of GIS and mapping data available in different regions. This data will need to be combined from various sources, including Google, OpenStreetMap and other municipal/transit and commercial sources. The quality of the data is a function of its availability, completeness, format and accuracy, and even top-quality data about an intersection is rendered out of date any time the intersection is changed. Important information such as the presence and location of walk buttons may also be unavailable from some existing data sources.

For these reasons we propose a crowdsourcing approach for collecting additional data from volunteers. This is the approach taken by OpenStreetMap; a successful example of crowdsourcing applied specifically to the needs of blind and visually impaired persons is Benetech's Bookshare project, which is an online free library of accessible educational e-books created by volunteers scanning and uploading paper books (Bigham *et al.*, 2011). We plan to develop accessible software tools to allow volunteers to enter data about specific traffic intersections from a web browser. It might also be fruitful to consider integrating Crosswatch with route planning services to help travelers choose the most convenient route to a desired destination. Such integration could draw on crowdsourced data labeling the traversability of each intersection (and possibly even the separate traversability of each crosswalk within the intersection) in terms of the traveler’s specific needs – for instance, a crosswalk that is well laid out for a wheelchair user may be hazardous for a slower-walking blind person because of a short traffic cycle. Another useful aspect of crowdsourcing is that sometimes such data tends to update faster than municipal databases; for example, a crosswalk that is inaccessible due to construction may not be labeled as such in municipal or other databases, but could be marked in the crowdsourced database accordingly by a fellow traveler.

Various technical issues relating to the computer vision and sensor components of Crosswatch will affect the system’s performance and need to be considered as well. For instance, limitations of camera technology and computer vision algorithms mean that some of the Crosswatch “where” functionality may be unavailable in challenging lighting conditions such as dusk or nighttime, or in the presence of rain, snow, or when the paint marking a crosswalk is peeling. The error of GPS readings, which is usually less than about 10 meters in an urban environment, can sometimes be even more inaccurate due to obstacles and high buildings, causing confusion about which traffic intersection the traveler is standing at (Brabyn *et al.*, 2002). One way to cope with these limitations is to provide confidence measures of each reading provided by Crosswatch, so that the user is informed when the system is less confident of its readings.

We will also investigate a variety of human factors issues that affect the usability and convenience of Crosswatch. It may be challenging or impractical for some users to sweep the smartphone in a circle to acquire an image panorama, especially when holding a white cane and/or encumbered by shopping bags, etc. For this reason some users may prefer a different camera platform, such as a camera worn on eyeglasses, an approach that has been popularized with the advent of head-mounted displays such as

Google Glasses (more recently renamed Project Glass). Another approach is an inexpensive lens that attaches to a smartphone to provide either a wide-angle view or even a full 360° image panorama in a single video frame (such as the Kogeto for the iPhone: http://www.kogeto.com/), which obviates the need to sweep the smartphone in a circle or large arc to acquire a panorama. A 3D spatialized sound interface (which requires the use of stereo headphones) may be preferred by some users to get accurate, intuitive 3D orientation information, as mentioned in the previous section. Finally, different users will likely use Crosswatch in very different ways. For some, Crosswatch may be useful only for traversing unfamiliar intersections, or even as a learning tool to become better acquainted with an intersection (after which time the system may no longer be needed). An option to "bookmark" familiar but challenging intersections may be useful for streamlining interactions with them: for instance, the only information that a frequent visitor to such an intersection may desire is the status of the walk light.

There are many other practical measures to be considered in improving the accessibility of Crosswatch and increasing the chances that it becomes widely adopted. First, it is crucial that training materials be prepared to acquaint new users with the system, and it will be important to get orientation and mobility instructors and other rehabilitation professionals involved in the creation and dissemination of these materials. Ongoing feedback from users and the visually impaired population overall, including formal surveys, will be necessary to ensure that the system provides the functionality that users need, in a way that is as accessible and useful as possible. Second, we plan to make the Crosswatch software free and open source (FOSS), so that anyone can download and use it for free. This model will also maximize the chances that interested 3rd party software developers can build upon and improve Crosswatch.

6 Conclusion

We propose to extend the functionality of Crosswatch, a smartphone-based system for providing guidance to blind and visually impaired travelers at traffic intersections, to encompass a wide range of "what," "where" and "when" information about traffic intersections. This information can be obtained by augmenting computer vision with other information sources, including GIS and smartphone sensor data. We report recent work on a user interface for acquiring a 360° image panorama, needed to estimate the user's precise (x,y) location in a traffic intersection, which has been tested with blind subject volunteers, demonstrating the feasibility of a key component of the Crosswatch approach. This new component is absent in past work on Crosswatch or related computer vision approaches (nor is it available through GPS, which lacks sufficient localization accuracy), but we feel it is an essential part of the "where" information that many users need for guidance at traffic intersections. Such information is needed, for instance, to inform a blind traveler that he/she is about to walk into an intersection outside of the crosswalk corridor, or to help him/her locate a walk button (which may be difficult to find even if the user knows that the button is present).

Other possible improvements and variations to consider in the near future will include the following: (1) the use of bone conduction headphones as an alternative to the built-in smartphone speakers, which may be difficult to hear in noisy urban environments and which may draw unwanted attention to the user, but which interfere as little as possible with normal hearing of ambient environmental sounds; (2) making greater use of tactile feedback, which is currently only used to provide tilt feedback, but which might also substitute for some forms of audio feedback; and (3) different ways for the user to enter commands, including voice recognition, menu selection using the smartphone touchscreen and movement-based gestures (such as tilting or shaking the smartphone).

Finally, we discuss a variety of social, human factors and technical issues that will have an important influence on the development of Crosswatch, and on the chances of its adoption in the future. Such issues include the heterogeneity of the community of potential users and of the regions in which they travel, the need to combine GIS and mapping data from a variety of sources (and to augment this data with crowdsourced data supplied by volunteers), the possible advantages of different camera and audio feedback configurations, the necessity of supplying training materials and the availability of Crosswatch software as free and open source (FOSS) software.

7 Acknowledgments

The authors acknowledge support by the National Institutes of Health from grant No. 2 R01EY018345-04 and by the Department of Education, NIDRR grant number H133E110004. We would like to thank Dr. Megan Lawrence (Smith-Kettlewell) for several helpful conversations about GIS, which were instrumental in helping the authors formulate the current research approach. Dr. Ender Tekin (Smith-Kettlewell) contributed valuable feedback on the manuscript, and Mr. David Vásquez (Smith-Kettlewell) provided assistance in conducting the subject experiments.

References

1. Ahmetovic, D., Bernareggi, C. and Mascetti, S. (2011), “Zebralocalizer: identification and localization of pedestrian crossings”, in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. ACM, New York, NY, USA.
2. Aranda, J., Mares, P. (2004), “Visual System to Help Blind People to Cross the Street”, in *Proceedings of International Conference on Computers Helping People with Special Needs (ICCHP 2004).* Paris, France. LNCS 3118, pp. 454–461.

3. Arth, C., Klopschitz, M., Reitmayr, G. and Schmalstieg, D. (2011), "Real-Time Self-Localization from Panoramic Images on Mobile Devices", in *Science and Technology Proceedings of IEEE International Symposium on Mixed and Augmented Reality 2011*. 26 -29 October 2011, Basel, Switzerland.
4. Barlow, J.M., Bentzen, B.L. and Tabor, L. (2003), "Accessible pedestrian signals: Synthesis and guide to best practice" in *Proceedings of National Cooperative Highway Research Program 2003*.
5. Barlow, J.M., Bentzen, B.L., Sauerburger, D., and Franck, L. (2010), "Teaching Travel at Complex Intersections." In Wiener, W.R., Welsch, R.L. & Blasch, B.B., eds, *Foundations of orientation and mobility*, AFB Press.
6. Bigham, J., Ladner, R., and Borodin, Y. (2011), "The design of human-powered access technology." Proceedings ASSETS 2011.
7. Bohonos, S., Lee, A., Malik, A., Thai, C. and Manduchi, R. (2008), "Cellphone Accessible Information via Bluetooth Beaconing for the Visually Impaired", in *Proceedings of ICCHP 2008*, LNCS 5105, pp. 1117-1121.
8. Brabyn, J.A., Alden, A., Haegerstrom-Portnoy, G. and Schneck, M. (2002), "GPS Performance for Blind Navigation in Urban Pedestrian Settings", in *Proceedings of Vision 2002*, Goteborg, Sweden, July 2002.
9. Chung, Y-C, Wang, J-M and Chen, S-W. (2002), "A Vision-Based Traffic Light Detection System at Intersections", *Journal of Taiwan Normal University: Mathematics, Science and Technology*, 47(1), 67-86 (2002).
10. Coughlan, J. and Shen, H. (2012), "The Crosswatch Traffic Intersection Analyzer: A Roadmap for the Future", in *Proceedings of 13th International Conference on Computers Helping People with Special Needs (ICCHP '12)*. Linz, Austria. July 2012.
11. Crandall, W., Bentzen, B.L., Myers, L. and Brabyn, J. (2001), "New orientation and accessibility option for persons with visual impairment: transportation applications for remote infrared audible signage." *Clinical and Experimental Optometry*, May, 84(3): 120-131 (2001).
12. Fitzpatrick, K., Chrysler, S. T., Iragavarapu, V., & Park, E. S. (2010), "Crosswalk Marking Field Visibility Study." No. FHWA-HRT-10-068.
13. Ivanchenko, V., Coughlan, J. and Shen, H. (2008), "Crosswatch: a Camera Phone System for Orienting Visually Impaired Pedestrians at Traffic Intersections", in *Proceedings of 11th International Conference on Computers Helping People with Special Needs (ICCHP '08)*. Linz, Austria. July 2008.
14. Ivanchenko, V., Coughlan, J. and Shen, H. (2009), "Staying in the Crosswalk: A System for Guiding Visually Impaired Pedestrians at Traffic Intersections." In *Proceedings of Association for the Advancement of Assistive Technology in Europe (AAATE 2009)*. Florence, Italy. Sept. 2009.
15. Ivanchenko, V., Coughlan, J. and Shen, H. (2010), "Real-Time Walk Light Detection with a Mobile Phone." In *Proceedings of 12th International Conference on Computers Helping People with Special Needs (ICCHP '10)*. Vienna, Austria. July 2010.
16. Liao, C.F. (2012), "Using a Smartphone App to Assist the Visually Impaired at Signalized Intersections." Report no. CTS 12-25, Minnesota Traffic Observatory Laboratory, Department of Civil Engineering, Univ. of Minnesota. Aug. 2012.
17. Se, S. (2000), "Zebra-crossing Detection for the Partially Sighted." In *Proceedings of Computer Vision and Pattern Recognition (CVPR 2000)*, South Carolina, June 2000.
18. Se, S. and Brady, M. (2003), "Road Feature Detection and Estimation." *Machine Vision and Applications Journal*, 14(3), July 2003, 157–165.
19. Szeliski, R. (2010), *Computer Vision: Algorithms and Applications.* Springer, New York.

20. Uddin, M.S. and Shioyama, T. (2005), "Bipolarity- and Projective Invariant-Based Zebra-Crossing Detection for the Visually Impaired." In *Proceedings of Workshop on Computer Vision Applications for the Visually Impaired in Computer Vision and Pattern Recognition (CVPR 2005)*, San Francisco, June 2005.
21. Utcke, S. (1998), "Grouping based on Projective Geometry Constraints and Uncertainty." In *Proceedings of International Conference on Computer Vision (ICCV 1998)*, Bombay, India, Jan. 1998.