

# Determining a Blind Pedestrian's Location and Orientation at Traffic Intersections

Giovanni Fusco, Huiying Shen, Vidya Murali and James M. Coughlan

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

**Abstract.** This paper describes recent progress on Crosswatch, a smartphone-based computer vision system developed by the authors for providing guidance to blind and visually impaired pedestrians at traffic intersections. One of Crosswatch's key capabilities is determining the user's location (with precision much better than what is obtainable by GPS) and orientation relative to the crosswalk markings in the intersection that he/she is currently standing at; this capability will be used to help him/her find important features in the intersection, such as walk lights, pushbuttons and crosswalks, and achieve proper alignment to these features. We report on two new contributions to Crosswatch: (a) experiments with a modified user interface, tested by blind volunteer participants, that makes it easier to acquire intersection images than with previous versions of Crosswatch; and (b) a demonstration of the system's ability to localize the user with precision better than what is obtainable by GPS, as well as an example of its ability to estimate the user's orientation.

**Keywords:** visual impairment, blindness, assistive technology, smartphone, traffic intersection.

## 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. Most prevalent among them are Accessible Pedestrian Signals, which generate sounds signaling the duration of the walk interval to blind and visually impaired pedestrians [3]. However, the adoption of Accessible Pedestrian Signals is very sparse, and they are completely absent at the vast majority of intersections. More recently, Bluetooth beacons have been proposed [4] to provide real-time information at intersections that is accessible to any user with a standard mobile phone, but this solution requires special infrastructure to be installed at each intersection.

Computer vision is another technology that has been applied to interpret existing visual cues in intersections, including crosswalk patterns [6] and walk signal lights [2,9]. Compared with other technologies, it has the advantage of not requiring any additional infrastructure to be installed at each intersection. While its application to

the analysis of street intersections is not yet mature enough for deployment to actual users, experiments with blind participants have been reported in work on Crosswatch [7] and a similar computer vision-based project [1], demonstrating the feasibility of the approach.

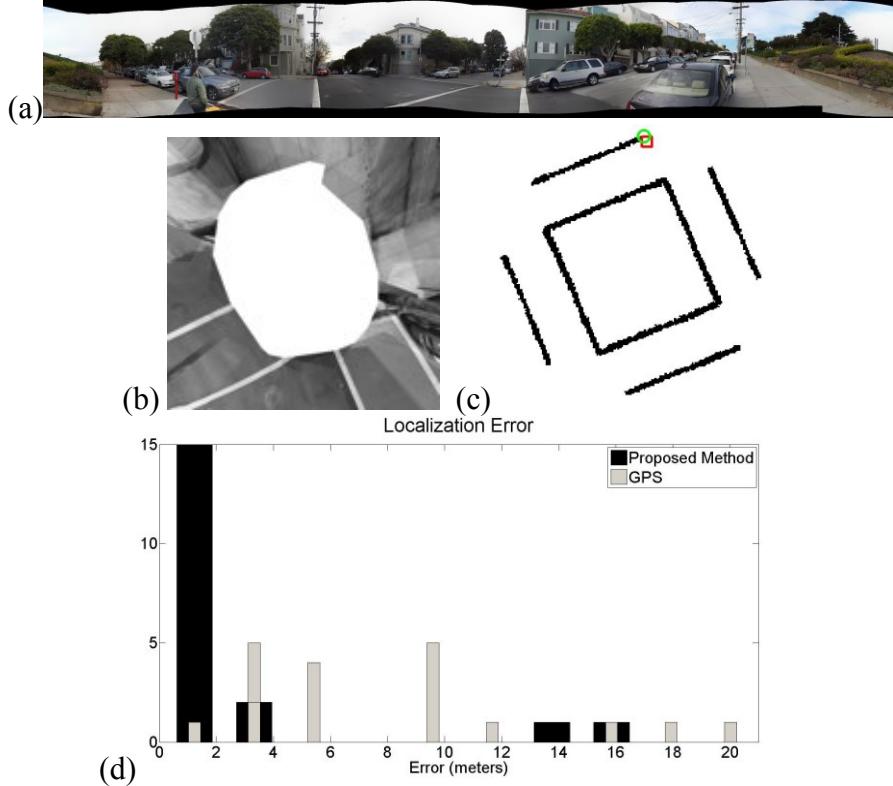
## 2 Overall Approach

Crosswatch uses a combination of information obtained from images acquired by the smartphone camera and from onboard sensors and offline data to determine the user's current location and orientation relative to the traffic intersection he/she is standing at. The goal is to ascertain a range of information about the intersection, including "what" (e.g., what type of intersection?), "where" (the user's precise location and orientation relative to the intersection) and "when" information (i.e., the real-time status of walk and other signal lights).

We briefly describe how the image, sensor and offline data are combined to determine this information. The GPS sensor determines which traffic intersection the user is standing at; note that GPS resolution, which is roughly 10 meters in urban settings [5], is sufficient to determine the current intersection but not necessarily which corner the user is standing at, let alone his/her precise location relative to crosswalks in the intersection. Given knowledge from GPS of which intersection the user is standing at, a GIS (geographic information system, stored either on the smartphone or offline in the cloud) is used to look up detailed information about the intersection, including the type (e.g., four-way, three-way) and a detailed map of the intersection, including features such as crosswalks, median strips, walk lights or other signals, push buttons, etc.

The IMU (inertial measurement unit, which encompasses an accelerometer, magnetometer and gyroscope) sensor estimates the direction the smartphone is oriented in space relative to gravity and magnetic north. Finally, information from panoramic images acquired by the user of the intersection, combined with IMU and GIS data, allows Crosswatch to estimate the user's precise location and orientation relative to the intersection, and specifically relative to any features of interest (such as a crosswalk, walk light or push button); given the pose, the system can direct the user to aim the camera towards the walk light, whose status can be monitored in real time and read aloud.

In [7] we reported a procedure that allows Crosswatch to capture images in all directions from the user's current location: the user stands in place and turns slowly in a clockwise direction (i.e., panning left to right) while the Crosswatch app periodically acquires images (every 20° of bearing relative to north), making a complete circle. The system activates the smartphone vibrator any time the camera orientation is sufficiently far from horizontal, to help the user acquire images that contain as much of the visible crosswalks as possible (and prevent inadvertently pointing too far up or down). The system stitches the images thus acquired into a full 360° panorama (see Fig. 1a), which was analyzed in [8] to determine the user's (x,y) location (Fig. 1b,c).



**Fig. 1.** Crosswatch user localization. (a) 360° panorama. (b) Aerial reconstruction of intersection based on (a) shows crosswalk stripes as they would appear from above. (c) Intersection template (i.e., map of crosswalk stripes). Matching features from (b) to (c) yields (x,y) localization estimate (shown in red, near ground truth in green). (d) Localization error of our algorithm (black) and of GPS (gray), shown as histograms with x-axis in meters and the number of counts on the y-axis, demonstrating superior performance of our algorithm.

However, capturing the full 360° panorama forces the user to turn on his/her feet in a complete circle, which is both awkward and disorienting. In this paper we describe our initial experiments with an approach that requires only a 180° (half-circle) panorama, which spans enough visual angle to encompass multiple crosswalk features in the intersection; this approach has the benefit that it can be executed by the user by moving arms, shoulders and/or hips without having to move his/her feet. The assumption underlying this approach is that the user can orient him/herself accurately enough using standard Orientation & Mobility skills, so that the 180° panorama encompasses enough of the intersection features.

We tested this new user interface in informal experiments with two blind users. The user interface was modified in two additional ways relative to the earlier (360°) version of Crosswatch: (a) in order to signal the beginning and end of the 180° sweep, the system not only issues loud audio beeps but also augments them with special vi-

bration patterns (three short vibration bursts in rapid succession), since past experiments revealed that street noises sometimes obscured the audio tones; and (b) in order to capture more of the crosswalk stripes in the images, the camera is designed to be held with a slightly negative pitch, i.e., pointing slightly down such that the horizon line appears horizontal but somewhat above the middle of the image (see Fig. 2a); when this pitch requirement is violated, or the camera is held too far from horizontal, the smartphone vibration is turned on to warn the user.

We trained the first user to use the system indoors, before going outdoors to a single intersection, where we acquired a panorama to check that the system was working correctly. The second user was also trained indoors, and was then led outside to a total of 18 locations at a total of five distinct traffic intersections. The main challenge for this user was to orient himself sufficiently accurately so that the 180° sweep encompassed all the nearby crosswalk stripes; he tended to begin his sweeps farther to the right than was optimal for the system, even with repeated hints given by the experimenters who accompanied him outdoors.

In these informal experiments with the second user we recorded which corner he was standing at for all sweeps. The localization algorithm estimated the correct corner in 16 out of the 18 cases. While these experiments were informal, we note the following observations: (a) users were able to acquire panoramas without having to move their feet during image acquisition, by using a combination of hip, shoulder and arm movements; (b) better training is required to help users orient themselves accurately enough to encompass all important intersection features in the 180° sweep. In the future we will experiment with improving the training procedure, as well as expanding the sweep range (e.g., to 225°) in a way that still permits the user to acquire the entire sweep without moving his/her feet.

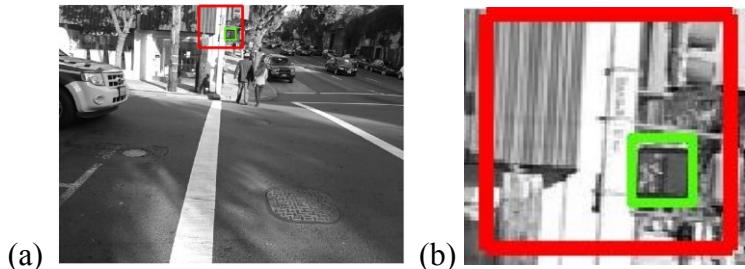
### 3 Evaluating Localization Estimates

In a separate experiment, we evaluated the accuracy of the Crosswatch localization estimates relative to ground truth estimates. In this experiment, one of the authors (who is sighted) photographed a total of 19 panorama sweeps (seventeen 180° sweeps and two 360° sweeps), each in a distinct location. The locations were distributed among three intersections, two of which were four-way intersections and one of which was a T-junction intersection. Another experimenter estimated the photographer's "ground truth" location for each panorama, making reference to curbs, crosswalk stripes and other features visible in satellite imagery; such ground truth is not perfect, but we estimate that it is accurate to about 1-2 meters. Using this ground truth, we evaluated the localization error of our algorithm and compared it with the localization error of GPS alone (Fig. 1d). The histograms show that our algorithm is more accurate than GPS. The typical localization error of our algorithm is roughly 1 meter or less, except when gross errors occur (i.e., failure to detect one or more crosswalk stripes, or mistaking a non-crosswalk feature in the image as a crosswalk stripe). In ongoing work we are creating more accurate ground truth and decreasing the rate of crosswalk stripe detection errors.

## 4 Demonstration: Where is the Walk Light?

The intersection template in the GIS includes the 3D locations of features such as crosswalks and walk lights. Once the user's (x,y) location is determined, the system has enough information to help the user find a specific feature of interest, such as the walk light. We choose the walk light as a sample target of interest because a blind user will need to be given guidance to aim the camera towards it, so that its status can be monitored and reported (e.g., using synthesized speech) in real time. Fig. 2 shows an example of how Crosswatch can predict the location of the walk light even before detecting it: in Fig. 2a, a photo acquired by the second blind user in our experiment (during a panorama sweep) is marked with a red rectangle indicating the system's prediction (calculated offline in this experiment) of the approximate walk light location; the green box shows the precise location of the walk light inside (see Fig. 2b for zoomed-in portion of image), as detected by a simple walk light detection algorithm (similar to that used in [9]).

This ability to predict the location of a feature in an intersection before detecting it in an image is important for two reasons. First, using high-speed orientation feedback from the IMU, the prediction provides a way for the system to rapidly guide the user to aim the camera towards the feature, without relying on computer vision, which could greatly slow down the feedback loop. Second, the prediction may be used either to simplify the computer vision detection/recognition process (by ruling out regions of images that are far from the prediction), and is available if computer vision is unable to perform reliable detection/recognition (for instance, for features such as median strips, which are extremely difficult to detect based on computer vision alone).



**Fig. 2.** Example of Crosswatch's prediction of the approximate walk light location (see text for explanation).

## 5 Conclusion

We have demonstrated new modifications to the Crosswatch user interface for acquiring panoramic imagery intended to make the procedure simpler and less disorienting for the user. A basic performance analysis of the Crosswatch localization algorithm that this user interface facilitates shows that the algorithm performs much better than GPS.

We are planning more rigorous experimental tests with more blind users, which will suggest further improvements to the system and to the user training process. Meanwhile, we are improving the localization algorithm and creating more accurate ground truth data to better evaluate its performance. Future work on Crosswatch will include integrating past work on another common crosswalk pattern (the zebra) and walk light detection into a single app, which may be accomplished by offloading some of the computations required by our algorithms to a remote server. Finally, we plan to investigate ways of creating an extensive GIS of intersections throughout San Francisco (and eventually other cities worldwide) through a combination of municipal data sources and crowdsourcing techniques, and to make the GIS data freely available (e.g., via OpenStreetMap).

**Acknowledgments.** The authors acknowledge support by the National Institutes of Health from grant No. 2 R01EY018345-06 and by the Department of Education, NIDRR grant number H133E110004.

## References

1. Ahmetovic, D., Bernareggi, C. and Mascetti, S. “ZebraLocalizer: identification and localization of pedestrian crossings.” In Proc. 13<sup>th</sup> Int’l Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI ’11). ACM, 2011.
2. Aranda, J., Mares, P. “Visual System to Help Blind People to Cross the Street.” ICCHP 2004. Paris, France. LNCS 3118, pp. 454–461 (2004).
3. Barlow, J.M., Bentzen, B.L. and Tabor, L. “Accessible pedestrian signals: Synthesis and guide to best practice.” National Cooperative Highway Research Program 2003.
4. Bohonos, S., Lee, A., Malik, A., Thai, C. and Manduchi, R. “Cellphone Accessible Information via Bluetooth Beacons for the Visually Impaired.” ICCHP 2008, LNCS 5105, pp. 1117-1121, 2008.
5. 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.
6. Coughlan, J. and Shen, H. “The Crosswatch Traffic Intersection Analyzer: A Roadmap for the Future.” 13th International Conference on Computers Helping People with Special Needs (ICCHP ’12). Linz, Austria. July 2012.
7. Coughlan, J. and Shen, H. “Crosswatch: a System for Providing Guidance to Visually Impaired Travelers at Traffic Intersections.” Special Issue of the Journal of Assistive Technologies (JAT). Vol. 7, Issue 2, pp. 131-142. 2013.
8. Fusco, G., Shen, H. and Coughlan, J. “Self-Localization at Street Intersections.” 11<sup>th</sup> Conference on Computer and Robot Vision (CRV 2014). Montréal, Canada. May 2014.
9. Ivanchenko, V., Coughlan, J. and Shen, H. “Real-Time Walk Light Detection with a Mobile Phone.” 12<sup>th</sup> International Conference on Computers Helping People with Special Needs (ICCHP ’10). Vienna, Austria. July 2010.