Staying in the Crosswalk: A System for Guiding Visually Impaired Pedestrians at Traffic Intersections

V. Ivanchenko, J. Coughlan and H. Shen
Smith-Kettlewell Eye Research Institute
San Francisco, CA USA 94115

Abstract. Traffic intersections are among the most dangerous parts of a blind or visually impaired person's travel. Our "Crosswatch" device [4] is a handheld (mobile phone) computer vision system for orienting visually impaired pedestrians to crosswalks, to help users avoid entering the crosswalk in the wrong direction and straying outside of it. This paper describes two new developments in the Crosswatch project: (a) a new computer vision algorithm to locate the more common – but less highly visible – standard "two-stripe" crosswalk pattern marked by two narrow stripes along the borders of the crosswalk; and (b) 3D analysis to estimate crosswalk *location* relative to the user, to help him/her stay inside the crosswalk (not merely pointing in the correct direction). Experiments with blind subjects using the system demonstrate the feasibility of the approach.

1. Introduction

Many urban traffic and pedestrian accidents occur at intersections, which are especially dangerous for blind or visually impaired pedestrians. Several types of Audible Pedestrian Signals (APS) have been developed to assist blind and visually impaired individuals in knowing when to cross intersections [1]. However, while widespread in some countries, their adoption is very sparse in others. Technology such as Talking Signs [2] allows blind travelers to locate and identify landmarks, signs, and facilities of interest, at intersections and other locations using infrared signals from installed transmitters, and has been found to enhance safety, efficiency and knowledge about the intersection. A number of related technologies have been proposed, and such systems are spreading, but are still only available in very few places.

The alternative approach that we have devised is embodied in our "Crosswatch" system [4], which uses computer vision software running on a mobile phone to identify important features in an intersection. With this system, the user takes an image of the intersection with a standard mobile camera phone, which is analyzed in real time by software run on the phone, and the output of the software is communicated to the user with synthesized speech or acoustic cues.

This paper describes two new developments in the Crosswatch project. First, we have devised a new computer vision algorithm to locate the more common – but less

highly visible – standard "two-stripe" crosswalk pattern marked by two narrow stripes along the borders of the crosswalk (see Fig. 1a). Second, we have added 3D analysis to estimate crosswalk *location* in addition to orientation, to help the user correct for two different forms of misalignment, translation error and direction error (Fig. 1b). A translation error occurs when the user is standing outside of the crosswalk borders, and may occur even if he/she is facing the correct direction; conversely, a direction error occurs when the user is facing a direction that will cause him/her to veer out of the crosswalk, even if he/she is currently inside its borders. Our new system provides audio feedback that allows users to correct for both translation errors and direction errors.
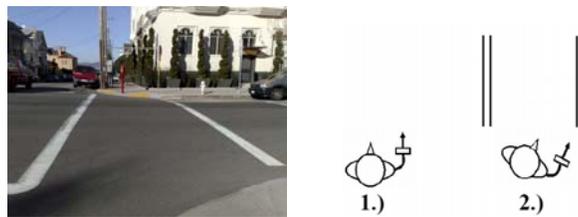

Fig. 1. (a) Two-stripe crosswalk. (b) Overhead view of users (holding mobile phones) with different kinds of alignment errors: (1) translation error and (2) direction error.

Finally, we describe a preliminary experiment with visually impaired subjects, demonstrating the feasibility of the system.

2. Finding Two-Stripe Crosswalks

Two-stripe crosswalks are more common than zebra (striped) crosswalks but are much less visible because the two-stripe pattern only demarcates the borders of the crosswalk, whereas the zebra crosswalk pattern is an alternating, high-contrast texture that fills the entire crosswalk area. This reduced visibility is especially problematic since vehicles or pedestrians in the crosswalk often block substantial parts of one or both stripes from view. Moreover, the limited field of view of the camera in a typical mobile phone means that an image is unlikely to contain *both* stripes unless the camera is very well aligned to the crosswalk.

To make it easier to detect two-stripe crosswalks we augment our analysis of image data with a non-visual cue that is available on an increasing number of mobile phones: the direction of gravity, as measured by the built-in accelerometer. When the phone is at rest (or moved steadily), the accelerometer vector indicates the direction perpendicular to the horizontal ground plane containing the crosswalk, which we denote by n. (Even if a street is on a slope, the street intersection containing the crosswalk is likely to be horizontal, and so the accelerometer still estimates n.)

Knowledge of n allows us to determine the location and orientation of the horizon line, which is determined by the angle the camera is held at (Fig. 2b). In addition, if we also know the camera focal length (which is fixed for each mobile phone model) and the approximate height the camera is held above the ground (about 1.5 meters for most adults), then we can reconstruct the geometry of everything on the ground plane. The significance of this ground plane reconstruction (Fig. 2c) is that it allows us to measure locations and distances on the ground plane in meters, and in particular *determines the location of the user's feet (more precisely, the point directly below the camera) relative*

to the crosswalk. This location estimate allows us to detect translation errors (as in Fig. 1b).

(a)   (b)

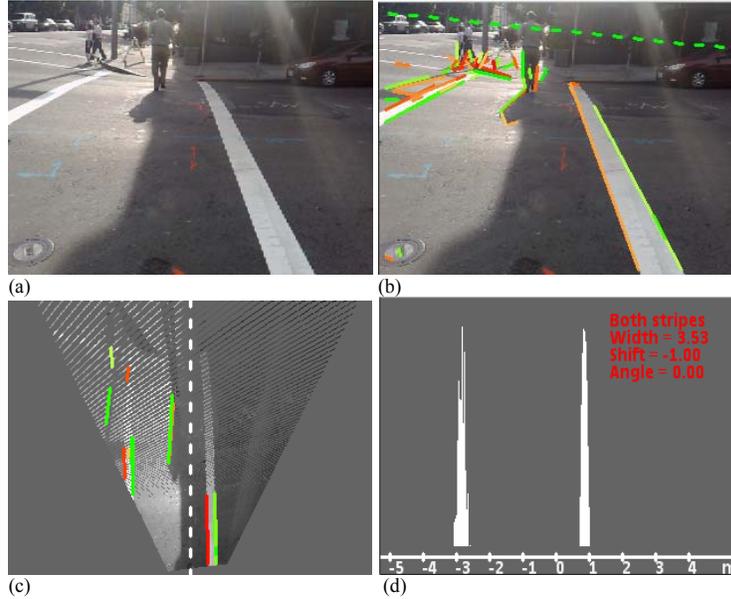(c)   (d)

Fig. 2. Main steps of algorithm. (a) Input image. (b) Colored lines indicate straight segments; dashed green line is horizon. (c) Ground plane geometry shown in aerial view; dashed white line shows dominant direction. (d) Two peaks along x-axis (units in meters) indicate two stripes.

Our algorithm proceeds as follows. First, for each camera frame (Fig. 2a) the accelerometer is read, and the corresponding value of n is calculated. Next, straight-line edge segments are extracted in the image (Fig. 2b) using a technique similar to that of [4], and those that are above the horizon line are discarded. The locations of the remaining segments are calculated assuming they lie on the ground plane, which typically yields a number of roughly parallel segments belonging to the crosswalk stripes, as well as stray background segments at random orientations. The dominant orientation of the segments is determined, and all segments with non-dominant orientations are removed.

To ascertain the presence of one or two stripes in the image, we first estimate which pixels in the image are likely to lie inside crosswalk stripes based on their brightness (since stripes are typically among the brightest parts of the scene), and on their proximity to the extracted segments. Then we construct an x-axis (in units of meters) on the ground plane that is perpendicular to the dominant direction, and count how many bright pixels near a segment in the image have the same x-coordinate, yielding a one-dimensional plot of pixels as a function of x (Fig. 2d). We expect one large peak in this plot for each stripe in the image, and so the number of significant peaks is an estimate of the number of stripes visible in the image (0, 1 or 2). If the peaks have an appropriate width (about 0.3 meters) and separation (at least 1.3 meters), then the algorithm declares the presence of a two-stripe crosswalk.

If a crosswalk is detected, then the dominant orientation on the ground plane defines the crosswalk bearing (i.e. 0° means the camera is pointed parallel to the

crosswalk direction). We also calculate the location of the user's feet relative to the corridor defined by the crosswalk (based on the x-coordinates of the two stripes), thus determining if the feet are inside the crosswalk corridor, or outside of it (to the left or right).

We ported our algorithm to the Nokia N95 mobile phone in Symbian C++. The algorithm was run in video mode, processing about three frames per second. We designed a user interface that allows a visually impaired user to quickly find a crosswalk, despite the camera's narrow field of view. For each frame, the presence of one or two crosswalk stripes was signaled with a brief low-pitched or high-pitched tone, respectively (no sound was generated if no stripes were detected). A user locates a nearby crosswalk by panning the phone left and right until low-pitched tones are repeatedly emitted, and then panning more finely until high-pitched tones are consistently produced, indicating that both stripes are currently in view. This interface exploits the fact that, while the algorithm may misinterpret individual camera frames, a consensus emerging from its analysis of several frames is very likely to be correct.

For the purposes of the experiment described in the next section, an additional interface component was added to indicate whether the user's feet are inside the crosswalk corridor, or outside of it (to the left or right): if two or more high-pitched tones are issued over the course of five consecutive frames, then the algorithm calculates the location of the user's feet, categorizes it as "inside", "left" or "right" of the corridor, and issues the appropriate speech signal. If the user holds the camera steady then this process repeats itself indefinitely, and he/she can decide if the system converges to a consistent output over time.

3. Experiments with Blind Subjects

We devised a simple experiment to test our system and demonstrate its feasibility. The experiment was based on objective information that was easily measured by the (sighted) experimenter: whether the user's feet lay inside the crosswalk corridor, left of it or right of it. (These three categories, or "zones," are denoted by I, L and R.) We avoided borderline cases which would have required precise measurements, and instead chose locations that were clearly in one of the three categories.

One outdoor crosswalk was chosen in advance for all the experiments, and a sequence of eight zones were chosen at random (with equal probability for I, L or R) for each of two blind subjects. A brief training period was first conducted indoors, using a model of a two-stripe crosswalk on the floor, to familiarize the subjects with the system and the experiment. One subject completed the trials for all eight zones, followed by the second subject. Each subject was led by the experimenter to stand in the appropriate zone for each trial, but the subject was not told which zone he was standing in. The subject was told to find the crosswalk using the mobile phone system, and to use the system to determine whether he stood in zone I, L or R.

In order to minimize the chances that the subject could ascertain each zone category from dead reckoning, the experimenter led the subject from one zone to the next in an indirect (i.e. intentionally disorienting) path. Of course, this procedure could not eliminate the effects of other cues available to the subjects, such as traffic sounds, or texture/slope of the ground. However, the subject was told to base his decision solely according to the output of the mobile phone system.

The result of the experiment was that both subjects indicated the correct zone category for all 8 trials. An exact binomial test shows that each subject responded significantly above chance, with $p = 1.5 \cdot 10^{-4}$ (i.e. $(1/3)^8$, the probability of responding correctly to all 8 trials by chance). In all but two trials the output of the system was unambiguous. However, there were two "borderline" trials in which the system incorrectly estimated that the subject was on the border between two adjacent zones, and at different times indicated one zone or the other. In these cases the subject was forced to guess which zone was correct, and may have drawn on other cues (mentioned above) not supplied by the mobile phone system. We emphasize that this experiment was preliminary, designed to demonstrate that blind users were able to extract reliable information about their location relative to the crosswalk; future experiments will need to be undertaken to further probe the operation of the system.

4. Conclusion

We have demonstrated a prototype mobile phone system that uses computer vision to detect two-stripe crosswalks in real time, extract 3-D information about crosswalk location, and convey this information to a visually impaired user with audio feedback. Simple experiments with blind subjects demonstrate the feasibility of the system. Future work will focus on user interface development and more extensive subject testing, and improving the system's ability to find crosswalks under more difficult conditions (e.g. large missing patches of paint in the stripes). Eventually we will integrate this functionality into a full traffic intersection analyzer, which will detect crosswalks of multiple types (zebra, two-stripe and others), analyze intersection layout (e.g. four-way or three-way), and locate and read signal lights (e.g. Walk/Don't Walk) to provide timing information.

References

[1] J.M. Barlow, B.L. Bentzen and L. Tabor, Accessible pedestrian signals: Synthesis and guide to best practice, National Cooperative Highway Research Program, 2003.

[2] W. Crandall, B.L. Bentzen, L. Myers and J. Brabyn, New orientation and accessibility option for persons with visual impairment: transportation applications for remote infrared audible signage, Clinical and Experimental Optometry **84**(3) (May 2001), 120–131.

[3] V. Ivanchenko, J. Coughlan and H. Shen. "Crosswatch: a Camera Phone System for Orienting Visually Impaired Pedestrians at Traffic Intersections." 11th Intl. Conference on Computers Helping People with Special Needs (ICCHP '08). Linz, Austria. July 2008.

[4] H. Shen, K.Y. Chan, J. Coughlan and J. Brabyn. "A Mobile Phone System to Find Crosswalks for Visually Impaired Pedestrians." Technology and Disability, Vol. 20, Number 3, pp. 217-224. 2008.