Crosswalk Guidance System for the Blind

Hojun Son, Divya Krishnagiri, V. Swetha Jeganathan, and James Weiland, *Member, IEEE*

***Abstract*— Street crossing can be a significant challenge for visually impaired people, limiting their mobility especially in urban environments. To date, there are few solutions for this significant problem. Current approaches for guiding blind pedestrians in crosswalks have mainly focused on detection of crosswalks and crosswalk signals. Few studies have taken into consideration the mobility of a visually impaired person while street crossing. We programmed a commercially available, wearable goggle system to detect crosswalk signals, to plan a path across the street, and to provide verbal guiding cues with real-time semantic features to keep the user on the correct path. During verification testing, we found crosswalk signal detections were typically reliable but depended on hyper-parameters to reduce false positive errors in the crosswalk signs in a small number of cases. Testing with visually impaired subjects resulted in successful guidance at an outdoor crosswalk.**

***Clinical Relevance*— Independent and safe mobility is a significant problem for people with visual impairment. Our work shows a way to improve the safety of blind travelers by guiding them at street crossings.**

I. Introduction

People with visual impairment (VI) report challenges such as the ability to travel independently which further limited their mobility and led to loss of independence, depression, reduced quality of life and an overall decline in health [1]. To address this important issue, we propose a wearable assistive technology that could guide orientation and mobility (O&M). for individuals with VI. The specific module we report here aids street crossing at signalized crosswalks. Using the system that we describe in this paper, algorithms interpret the environment, plan a path, and provide simple cues to the user about when and how to proceed. We use crosswalk navigation as a specific scenario that requires several visual tasks: locating the crosswalk signal, determining the state of the signal, and walking from one end of the crosswalk to the other. We tested our device in blind users and demonstrated successful crosswalk guidance and important design considerations for future systems.

Several prototype devices for crosswalk navigation have been reported. CrossNavi detects the large white stripes found in many crosswalks using a smartphone camera and app. The system maintains subject heading within the detected crosswalks with a customized cane [2], but does not detect the state of the crosswalk signal. Cross-safe uses a commercial stereo vision device to help detect the state of crosswalk signs, using neural networks[3]. It did not provide feedback guidance for mobility. Cross-safe was not tested on visually impaired subjects [4].

Our research is unique in that we created a completely self-contained detection and guidance system on a commercial, wearable HMD. No special infrastructure (e.g. beacons), bulky hardware and sensors, or wireless connectivity was required. Our system has been tested with visually impaired subjects in both simulated and real crosswalks.

II. Method

Our navigation assistive device utilized a commercially available smart goggles, programmed with a custom set of algorithms. The algorithms had two main tasks: 1) detection and classification of crosswalk signals (see II.A) and 2) guiding users to their destination (across the target crosswalk) (see II.B). These two modules were implemented on the ODG R7 smart glasses (Osterhout Design Group see Fig. 1*)*. Fig. 2 shows the block diagram of a full software pipeline. The pipeline operated in near real-time at 10 frames/second, which is adequate for this application.

Fig. 1: ODG R7 smart glasses. The smart glasses had a Qualcomm Snapdragon 805 2.7 GHz quad-core processor, a monocular camera, diagonal 30-degrees field of view, 10-degree vertical field of view, a Bluetooth link, and an Inertial Measurement Unit (IMU) sensor.

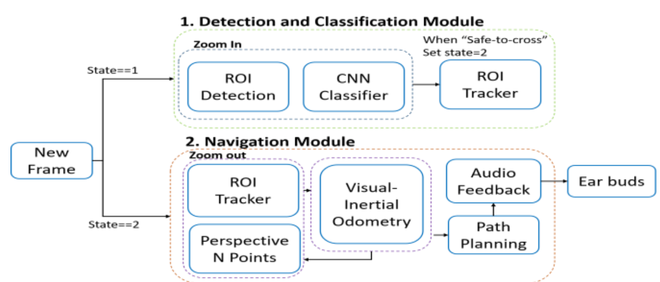

Fig. 2: Block diagram of full pipeline of the system. The new frame was the input frame from the smart glasses. Each module ran dependent on its current state.

A. Detection and Classification Module

The detection and classification module contained a cascade classifier [5] using LBP features [6] to set a region of interest (ROI) for the crosswalk signals, which were classified as "safe-to-cross" or "do-not-cross". Both ROI detection and signal classification required high accuracy and real-time performance. This was important because wrong classification could cause the user to move forward despite a "do-not-cross" signal. Furthermore, real-time performance

* Supported by the University of Michigan and a Research to Prevent Blindness Departmental Grant to the University of Michigan
Hojun Son is with the Biomedical Engineering, University of Michigan (e-mail: hojunson@umich.edu).
Divya is with the Biomedical Engineering, Ohio State University (e-mail: krishnagiri.1@buckeyemail.osu.edu).
V. Swetha Jeganathan is with the Dept. of Biomedical Engineering, University of Michigan (e-mail: jvswetha@med.umich.edu).
James D. Weiland is with the Dept. of Biomedical Engineering, University of Michigan (e-mail: weiland@umich.edu).

was needed since delays could result in the detection of the "safe-to-cross" sign being missed.

1) Classification of ROI

A simple convolutional neuron network (CNN) was used to classify the ROI. The network's structure had four blocks consisting of two convolution layers and a max-pooling layer. A batch normalization layer [7] was inserted after every convolution layer and a dropout layer attached after each max pooling layer. There were two fully connected layers to classify the inputs. Three different classes were used: "do-not-cross", "safe-to-cross", and "background", utilizing transfer learning with VGG16 parameters [8]. Any input not classified as "do-not-cross" or "safe-to-cross" was classified as "background". The background training data avoided overfitting as our target images did not have many features and the network was liable to overfitting. More sophisticated CNN are available (e.g. Yolov3 [9]), but could not be used due to computational resource limits on the R7 glasses. For CNN training and testing, ten video sets were recorded, each video set had 8different crosswalks. 12,450 images were used for training and 791 sets were used for testing.

2) Tracking

The ROI tracker combined the Median Flow tracker [10] and a cascade classifier with classification by CNN to adjust for the different zoom levels between the two task modules and to set the crosswalk signal as a landmark. The tracker was used for both tasks: 1) detection and classification and 2) navigation (Fig. 4). If the cascade classifier detected the ROI, then this location was used for localization. In cases when the cascade classifier did not detect the ROI in a given frame, then the ROI was tracked only by the Median flow tracker.

B. Navigation Module

The navigation module localized the user relative to the safe path and provided real-time audio feedback to keep the user within the safe path boundary. Red texture plates at each end of the crosswalk were used as start and end points.

1) Visual-Inertial Odometry with Geometry Constraints.

Visual- Inertial Odometry was used to localize the user's current position, which was mapped onto a grid for path planning. Geometry constraints were determined by detected crosswalk signs as a landmark. To update the user's position as they walked, a Visual-Inertial Odometry estimator was used, based on the quaternion kinematics error-state extended Kalman filter [11]. Updating user position involved a prediction step and a correction step. IMU pre-integration with sensor fusion of accelerometer and gyroscope was utilized in the prediction step of the Kalman Filter. Results from solving the perspective-n-point problem with matched 3D virtual coordinate of the detected crosswalk sign and 2D image pixel points were applied in the correction step of the Kalman filter.

2) Path Planning

The goal of path planning was to keep the user on a safe path. The local area was divided into a 31 x 50 grid for the outdoor crosswalk. Each grid was 50 cm x 50 cm. For every pose estimation, an A* algorithm [12] was performed iteratively. The A* search was used to find a path with the least nodes using path cost and heuristics. A difference between the current grid and desired grid on the preferred path prompted cues such as "veer left" or "veer right" to direct the user back to the safe path.

3) Audio Feedback

Audio feedback was incorporated as system guiding cues. The following five verbal commands were electronically synthesized to guide the user while street crossing: standby, veer left, veer right, forward and arrived. The commercially available smart glasses included ear buds that magnetically attached to the headset to provide audio to the user.

III. RESULTS

A. Verification with sighted subjects

Two sighted persons were tested as controls. They verified the systems correct operation in both indoor (Fig. 3) and outdoor (Fig. 4) settings, in different seasons and weather conditions and at multiple crosswalks.

B. Accuracy of region of interest classification

Table 1 shows the confusion matrix results for the ROI classification accuracy for two test sets: 1) a collection of crosswalk signal images including purposefully distorted or badly cropped images and 2) frames from video recorded during outdoor testing in sighted persons on three different days. The outdoor crosswalk signal results show that the classification algorithm worked correctly in a variety of weather, time, brightness, and scale of the ROI.

Table 1**:** The plain text showed the results when purposefully distorted test images were included. The **bold** text shows the classification results at 15 different outdoor crosswalks at 3 different intersections on 3 different days.

labels	Results of Classification		
	safe-to-cross	**do-not-cross**	**Background**
"safe-to-cross"	197/217 (90.7%) **60/60 (100%)**	5/217 (2.3%) **0/60 (0%)**	15/217 (6.9%) **0/60 (0%)**
"do-not-cross"	1/304 (0.003%) **0/975 (0%)**	257/304 (0.845%) **975/975 (100%)**	46/304 (0.151%) **0/975 (0%)**
Background	4/270 (0.148%) **0/5 (0%)**	1/270 (0.003%) **0/5 (0%)**	265/270 (0.981%) **5/5 (100%)**

C. Experiment with visually impaired subjects

The University of Michigan Institutional Review Board approved the study. After informed consent, three visually impaired subjects participated in the study. All had self-reported difficulty with street crossing. Two of the subjects used guide dogs and one used a cane; their testing was performed both in indoor and outdoor settings, at a single crosswalk. Indoor testing (Fig. 3) was a 8.87m simulated crosswalk in a conference room. A video of the crosswalk signals was displayed to mimic an outdoor crosswalk. A real crosswalk was used for outdoor testing and was 14.87 meters in length. For safety, several experimenters flanked the user while they were in the crosswalk and were ready to intervene if any hazardous situation arose. Successful completion of a trial required the subjects to start at the correct time and finish at the specific endpoint. Hardware failures (e.g. dead battery) were not included as trials.

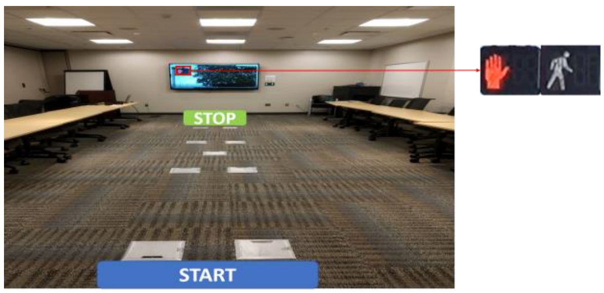

Fig. 3: Indoor crosswalk testing environment used to train the subjects. The monitor displayed a video of the crosswalk changes from “do-not-cross” to “safe-to-cross”

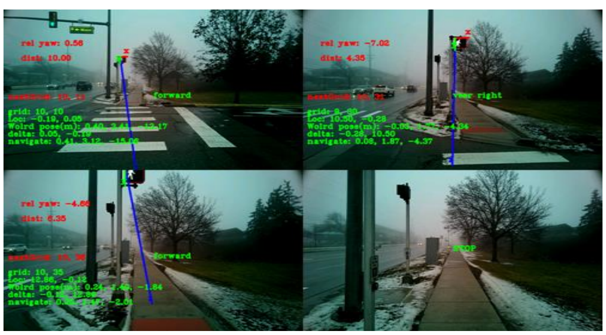

Fig. 4: System verification test carried out in winter, with sighted subjects. (Top Left) – The crosswalk signal has been located and the safe-to-cross symbol detected. The system provides the verbal cue “forward”, which is also displayed at the center of the screen. (Top Right) The subject has drifted to the left of the preferred path. The system detects the position and provides the verbal cue “veer right”. (Bottom Left) The user has responded to the cue and is now on the correct heading, which is confirmed by the verbal cue “forward”. (Bottom Right) The user has reached the destination and is provided the verbal cue “stop”.

Subject S001 had retinitis pigmentosa with visual acuity of 20/800. He uses a guide dog but did not use the animal during testing. He could locate the indoor monitor but could not see details of what was displayed. During outdoor crosswalk testing, he used the sound of the traffic as a guidance cue. S001 completed 7/11 indoor trials and 4/4 outdoor trials. Battery loss limited this subject’s outdoor trials to 4.

Subject S002 had severe retinitis pigmentosa with bare light perception vision. He had an Argus 2 retinal prosthesis (which was not powered during the testing). He navigated with the aid of a guide dog but chose not to use the animal during testing. The guide dog required handling which led to 1) diverted attention 2) sudden movement which resulted in loss of crosswalk signal tracking and 3) the dog being uncomfortable indoors. Initially, S002 walked slowly and cautiously and “side-stepped” during his experiment but adapted in later trials. Comparing the number of steps taken on the indoor course, S002 needed more steps than either S001 or S003 (see Fig. 5). He completed 8/11 indoor trials. His outdoor crosswalk testing was unsuccessful. He had difficulty when attempting to position the crosswalk signal in the camera’s FOV due to his extreme blindness. Also, weather conditions reduced the ROI detection rate (bright sunlight on the crosswalk signal which reduced contrast).

Subject S003 had no measurable visual acuity (best vision was detection of motion) from combined retinopathy of prematurity and secondary glaucoma. He could not see the crosswalk sign but could see the stripes in the crosswalk, which he used as orientation cues, in addition to a white cane to navigate. He completed 8/11 indoor trials and 7/10 outdoor trials. In the three failed outdoor trials, the subject received a stop command prior to reaching the endpoint, due to inaccurate position estimation. The investigators walking with him instructed him to take a few more steps, which he did to reach the endpoint.

Both S001 and S003 were able to complete the outdoor crosswalk in the time allowed by the crosswalk, meaning the signal was either “safe-to-cross” or blinking “do-not-cross” while they were navigating.

IV. Discussion

A crosswalk navigation system was demonstrated in visually impaired patients. The users however needed assistance for proper alignment with the crosswalk since the magnified camera setting needed for signal classification resulted in a narrow camera field of view. During training, the subjects learned how to adjust their mobility based on audio feedback received. For example, when the subjects’ made sharp turns, tracking of the crosswalk sign was lost, which led either to lost time when attempting to recover to the crosswalk path or to experiment failure due to lost tracking. For example, S003 initially interpreted the “veer left” command as a 45 degree turn. To train them on how to respond, we used a sighted guide approach, where the user held the arm of an investigator, who spoke commands and demonstrated the degree of turn they should use. This helped the users adjust their responses. After training, simple verbal cues were effective in guiding the subjects, consistent with our earlier studies [13]. Use of bone conduction headphones is recommended so that the ear canal, and natural hearing, are available for the users since they often rely on environmental cues for navigation.

We demonstrated some key requirements for a crosswalk navigation system, which included performing verification testing at the crosswalks in different conditions. Test data the included distorted images resulted in 91% accuracy, a number which was too low for a safety critical function. However, improved ROI detection allowed 100% accurate classification of “safe-to-cross” and “do-not-cross” states under a variety of conditions and novel crosswalks not used in classifier training. In spite of our effort to gather training data under a wide range of conditions, a novel condition was encountered (bright sunshine on the crosswalk sign) that decreased the ROI detection rate for S002 outdoor testing. Methods for improving ROI detection are needed for future systems.

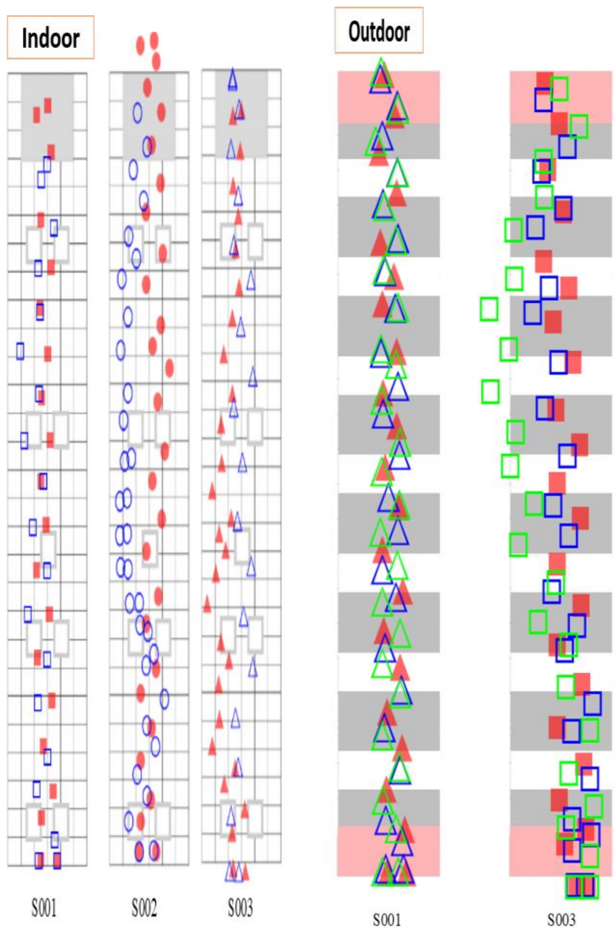

Fig. 5: Footstep maps from indoor (left) and outdoor (right). Top view of each course. Each shape represents a step taken by the user while completing the trial. Representative data shown. **Indoor Maps.** The filled shapes represent trial 1 and the open shapes represent trial 2 (from 11 trials indoors). The light grey, open rectangles indicate floor plates, which were used as reference points for analysis of video. The light grey, filled rectangle on the top represents the end point. In trial 1, Subject S002's steps were closer together and he required more steps vs. S001 and S003. S002 adapted quickly, as shown in trial 2. **Outdoor Maps.** S001 and S003 maps. Three trials are shown with colors indicating different trials. The pink rectangles represent the start and end points and the white rectangles represent the crosswalk stripes. During 1 trial (green boxes), S003 walked out of crosswalk but recovered his path using guiding cues.

During testing in sighted persons, there was no loss of crosswalk signals because the subjects' movements were stable. On the other hand, visually impaired users lacked visual feedback required to point the camera towards the crosswalk signal. This required the investigators to guide the subjects' heads until the crosswalk signal was in view of the camera. In the case of S002, this was not possible. Cameras with both higher resolution and wider FOV are needed to address this issue, so that magnification is not required to find the ROI. However, the larger image size will increase computational demand. The use of prior maps and algorithms for localization of the user within these maps, regardless of head position, is the ultimate solution to the localization problem. Collision avoidance was not considered, that is we assumed the crosswalk would be a free path. Other pedestrians and cars encroaching on the crosswalk are potential hazards. Although it is well known that sighted pedestrians typically will yield to users of a white cane or a guide dog, real-time obstacles detection must be added to our system to account for dynamic changes in the environment and path correction based on these changes.

V. Conclusion

In summary, our crosswalk navigation system was able to guide blind users safely at a crosswalk. Signal state classification was excellent but depended highly on successful ROI detection. In the controlled indoor setting, detection was more consistent and such a setting may be an important first step for training people to use such a device. Navigation utilizing simple verbal cues was effective, once the users were trained. However, we uncovered some challenges and limitations to the current approach. Proper aiming of the camera was difficult due to magnified camera settings and lack of visual feedback in the VI test subjects. An extensive training set is needed to ensure robustness under a wide variety of conditions which can affect image contrast. These findings will incorporated into the redesign of our system, followed by testing in a wider pool of individuals.

Acknowledgment

Yumou Wei, Edgar Cheung, and Cynthia Lin contributed to this project.

References

[1] Owsley, Cynthia, and Gerald McGwin Jr. "Depression and the 25-item national eye institute visual function questionnaire in older adults." Ophthalmology 111.12 (2004): 2259-2264.

[2] Shangguan, Longfei, et al. "Crossnavi: enabling real-time crossroad navigation for the blind with commodity phones." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

[3] Freund, Y., Schapire, R.E., et al.: Experiments with a new boosting algorithm. In: ICML, vol. 96, pp. 148–156, Bari, Italy (1996)

[4] Li, Xiang, et al. "Cross-Safe: A computer vision-based approach to make all intersection-related pedestrian signals accessible for the visually impaired." Science and Information Conference. Springer, Cham, 2019.

[5] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." CVPR (1) 1 (2001): 511-518.

[6] Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." IEEE Transactions on Pattern Analysis & Machine Intelligence 12 (2006): 2037-2041.

[7] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." arXiv preprint arXiv:1502.03167 (2015).

[8] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[9] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[10] Kalal, Zdenek, Krystian Mikolajczyk, and Jiri Matas. "Forward-backward error: Automatic detection of tracking failures." 2010 20th International Conference on Pattern Recognition. IEEE, 2010.

[11] Sola, J. Quaternion kinematics for the error-state Kalman filter. arXiv 2017, arXiv:1711.02508.

[12] Dechter R, Pearl J. Generalized best-first search strategies and the optimality of A. Journal of the ACM (JACM). 1985;32(3):505-536.

[13] Adebiyi, Aminat, et al. "Assessment of feedback modalities for wearable visual aids in blind mobility." PloS one 12.2 (2017): e0170531.