

# Detecting Stairs and Pedestrian Crosswalks for the Blind by RGBD Camera

Shuihua Wang and Yingli Tian

*Department of Electrical Engineering,  
The City College of New York, NY, 10031, USA  
{swang15, ytian}@ccny.cuny.edu*

## Abstract

*A computer vision-based wayfinding and navigation aid can improve the mobility of blind and visually impaired people to travel independently. In this paper, we develop a new framework to detect and recognize stairs and pedestrian crosswalks using a RGBD camera. Since both stairs and pedestrian crosswalks are featured by a group of parallel lines, we first apply Hough transform to extract the concurrent parallel lines based on the RGB channels. Then, the Depth channel is employed to further recognize pedestrian crosswalks, upstairs, and downstairs using support vector machine (SVM) classifiers. Furthermore, we estimate the distance between the camera and stairs for the blind users. The detection and recognition results on our collected dataset demonstrate that the effectiveness and efficiency of our proposed framework.*

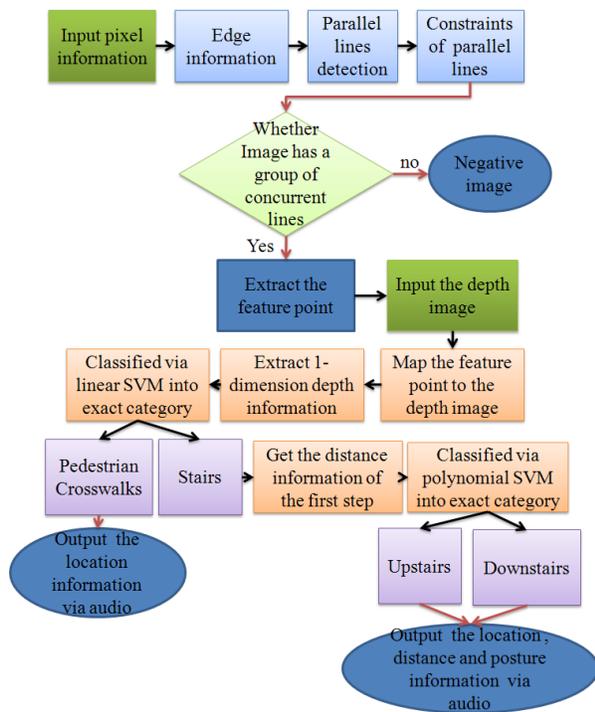
## 1. Introduction

Independent travel and active interactions with the dynamic surrounding environment are well known to present significant challenges for individuals with severe vision impairment, thereby reducing quality of life and compromising safety. In order to improve the ability of people who are blind or have significant visual impairments to access, understand, and explore surrounding environments, many assistant technologies and devices have been developed to accomplish specific navigation goals, obstacle detection, or wayfinding tasks.

The most popular electronic mobility assistant systems are those based on conversion sonar information into an audible signal for the visually impaired persons to interpret [2, 6, 11]. However, they are not so commonly used and only provide limited information. Recently, researchers focus on interpreting the visual information into high level

representation before sending to the visually impaired persons. Coughlan *et al.* [3] developed a method of finding crosswalks based on figure-ground segmentation, which they casted in a graphical model framework for grouping geometric features into a coherent structure. Ivanchenko *et al.* [5] further extended the algorithm to detect the location and orientation of pedestrian crosswalks for a blind or visually impaired person using a cell phone camera. The prototype of the system can run in real time on an off-the-shelf Nokia N95 camera phone. The cell phone automatically took several images per second, analyzed each image in a fraction of a second and sounded an audio tone when it detected a pedestrian crosswalk. Advanyi *et al.* [1] employed the Bionic eyeglasses to provide the blind or visually impaired individuals the navigation and orientation information based on an enhanced color preprocessing through mean shift segmentation. Then detection of pedestrian crosswalks was carried out via a partially adaptive Cellular Nanoscale Networks algorithm. Se *et al.* [13] proposed a method to detect zebra crosswalks. They first detected the crossing lines by looking for groups of concurrent lines. Edges were then partitioned using intensity variation information. Se *et al.* [14] also developed a Gabor filter based texture detection method to detect distant stair cases. When the stairs are close enough, stair cases were then detected by looking for groups of concurrent lines, where convex and concave edges were portioned using intensity variation information. The pose of stairs was also estimated by homograph search model. The “vOICe” system [16] is a commercially available vision-based travel aid that transfers image information to sound. The system contains a head-mounted camera, stereo headphones and a laptop. Uddin *et al.* [16, 17] proposed a bipolarity-based segmentation and projective invariant-based method to detect zebra crosswalks. They first segmented the image on the basis of bipolarity and selected the candidates on the basis of area, then extracted feature points on the candidate area based on

the fisher criterion. The authors recognized zebra crosswalks based on the projective invariants. Everingham *et al.* [4] developed a wearable mobility aid for people with low vision using scene classification in a Markov random field model framework. They segmented an outdoor scene based on color information and then classified the regions of sky, road, buildings etc. Lausser *et al.* [9] introduced a visual zebra crossing detector based on the Viola-Jones approach. Shoal *et al.* [12] discussed the use of mobile robotics technology in the Guide-Cane device, a wheeled device pushed ahead of the user via an attached cane for the blind to avoid obstacles. When the Guide-Cane detects an obstacle, it steers around it. The user immediately feels this steering action and can follow the Guide-Cane's new path. Tian *et al.* [15] developed a proof-of-concept computer vision-based wayfinding aid for blind people to independently access unfamiliar indoor environments. They mainly focus on indoor object detection and context information extraction and recognition.



**Figure 1. Flow chart of the proposed algorithm for stair and pedestrian crosswalk detection and recognition.**

In this paper, we propose a computer vision-based method to detect stair-cases and pedestrian crosswalks by using a commodity RGBD camera. The recent introduction of the cost-effective RGBD cameras eases the task by providing both RGB information and depth information of the scene. As shown in Figure 1, our

method consists of three main steps. First, a group of parallel lines are detected via Hough transform and line fitting with geometric constraints from RGB information (see details in Section 2.1). In order to distinguish stairs and pedestrian crosswalks, we extract the feature of one dimension depth information according to the direction of the detected longest line from the depth image. Then the feature of one dimension depth information is employed as the input of a SVM-based classifier to recognize stairs and pedestrian crosswalks. For stairs, a further detection of upstairs and downstairs is conducted. Furthermore, we estimate the distance between the camera and stairs for the blind users.

The paper is organized as following: Section 2 describes the methodology of our proposed algorithm including 1) detection whether the scene image contains stair-cases or pedestrian crosswalks based on RGB image analysis; 2) since both stairs and pedestrian crosswalks are featured by a group of parallel lines in RGB images, we further employ depth information to distinguish stairs from pedestrian crosswalks, then stairs will be further recognized as upstairs and downstairs. Section 3 displays the evaluation effectiveness and efficiency of proposed method and summarizes the experiment results. Section 4 concludes the paper and our future work.

## 2. Methodology of RGBD Camera based Stair and Pedestrian Crosswalk Detection

### 2.1. Detecting Candidates of Pedestrian Crosswalks and Stairs from RGB images

There are various kinds of stair-cases and pedestrian crosswalks. In this paper, we focus on stair cases with uniform trend and steps, and pedestrian crosswalks of the most regular zebra crosswalks with alternating white bands. In our application of blind navigation and wayfinding, we focus on detecting stairs or pedestrian crosswalks in a close distance.

Stairs consists of a sequence of steps which can be regarded as a group of consecutive curb edges, and pedestrian crosswalks can be characterized as an alternating pattern of black and white stripes. To extract these features, we start with an edge detection to obtain the edge map from RGB image of the scene and then perform a Hough transform to extract the lines in the extracted edge map image. These lines for are parallel for both stairs and pedestrian crosswalks. Therefore, a group of concurrent parallel lines will most likely represent the structure of stairs and pedestrian crosswalks. In order to eliminate the noise

from unrelated lines, we add constraints including the number of concurrent lines, line length, etc.

**Extracting Parallel Lines based on Hough Transform:** We apply Hough transform to detect straight lines based on the edge points. A number of edge points  $(x_i, y_i)$  in an image that form a line can be expressed in the slope-intercept form:  $y=ax+b$ , where  $a$  is the slope of the line and  $b$  is the  $y$ -intercept. The main idea here is to consider the characteristics of a straight line not as image points  $(x_1, y_1), (x_2, y_2)$ , etc., but instead, in terms of its parameters. Based on that fact, the straight line  $y=ax+b$  can be represented as a point  $(a, b)$  in the parameter space. However, we face the problem that vertical lines give rise to unbounded values of the parameters  $a$  and  $b$ . Considering the unbounded values of the parameters  $a$  and  $b$ , it is better to use the Polar coordinates, denoted  $r$  and  $\theta$ , for the lines in the Hough transform (as shown in Figure 2).

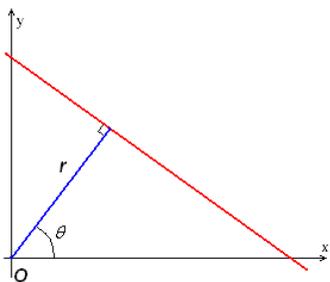


Figure 2. Illustration of polar coordinates of a line.

The parameter  $r$  represents the distance between the line and the origin, while  $\theta$  is the angle of the vector from the origin to the closest point, then the equation of a line can be represented as:

$$r = y \sin \theta + x \cos \theta \quad (1)$$

The algorithm of Hough transform line fitting is summarized as following:

*Step1:* Detect edge maps from the RGB image by edge detection.

*Step2:* Compute the Hough transform of the RGB image to obtain  $r$  and  $\theta$ .

*Step3:* Calculate the peaks in the Hough transform matrix.

*Step4:* Extract lines in the RGB image.

*Step5:* Detect a group of parallel lines based on constraints such as the length and total number of detected lines of stairs and pedestrian crosswalks.

As shown in Figures 3(c), 4(c), and 5(c), the detected parallel lines of stairs and pedestrian crosswalks are marked as green, while yellow dots and

red dots represent the beginning and the end of the lines respectively. However, these lines are often separated with small gaps caused by noises, so we group the line fragments as the same line if the gap less than a threshold. In general, stairs and pedestrian crosswalks contain multiple parallel lines with a reasonable length. If the length of a line  $\leq \phi$ , then the line is not belong to the line group. And if the number of parallel lines less than  $\beta$ , the scene image is a negative image which does not contain stairs and pedestrian crosswalks. In our experiment, we set the line length  $\phi$  as 60 pixels in the acquired images and the number parallel lines  $\beta$  as 5.

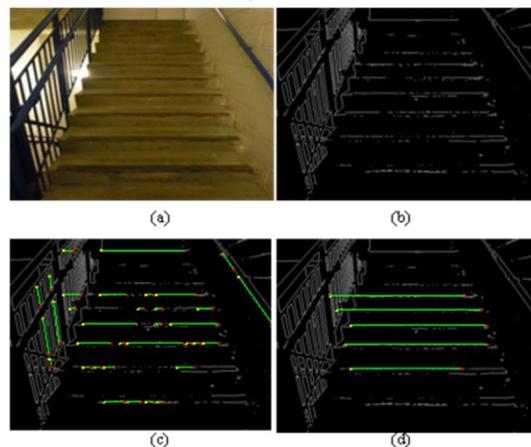


Figure 3. An example of upstairs. (a) Original image; (b) edge detection; (c) line detection; (d) concurrent parallel lines detection (yellow dots represent the beginnings, red dots represent the ends of the lines, and green lines represent the detected lines.)

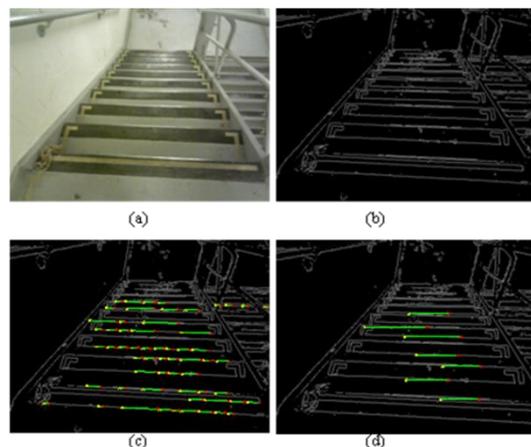


Figure 4. An example of downstairs. (a) Original image; (b) edge detection; (c) line detection; (d) concurrent parallel lines detection (yellow dots represent the beginnings, red dots represent the ends of the lines, and green lines represent the detected lines.)

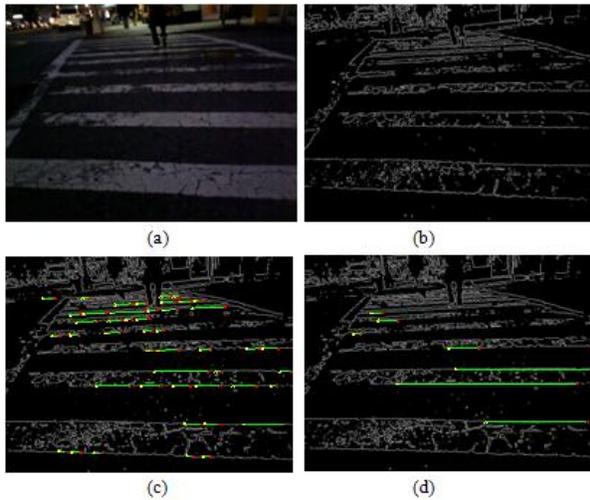


Figure 5. An example of Pedestrian crosswalks. (a) Original image; (b) edge detection; (c) line detection; (d) concurrent parallel lines detection (yellow dots represent the beginnings, red dots represent the ends of the lines, and green lines represent the detected lines.)

## 2.2. Recognizing Pedestrian Crosswalks and Stairs from Depth Images

Based on the above algorithm, we can detect the candidates of stairs or pedestrian crosswalks by detecting parallel lines with constraint condition in a scene image captured by a RGBD camera. From the depth images, we observe that upstairs have rising steps and downstairs have decreasing step, and pedestrian crosswalks are flat with smooth depth change as shown in Figure 6. Considering the safeness for the visually impaired people, and the further application for the robotic, it is necessary to classify the different stairs and pedestrian crosswalks into the correct categories.

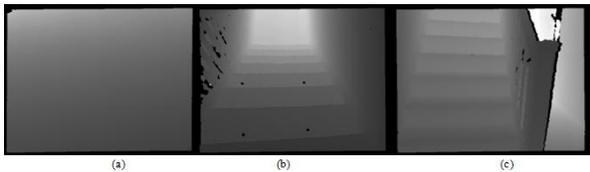


Figure 6. Depth images of (a) pedestrian crosswalks, (b) downstairs, and (c) upstairs.

In order to distinguish stairs and pedestrian crosswalks, we first calculate the orientation and position for extract the one-dimensional based feature from depth information. As shown in Figure 7, the orientation is perpendicular to the parallel lines detected from RGB images. The position will be determined by the middle point of the longest line of

the parallel lines. In Figure 7, the blue square indicates the middle point of the longest line and the red line shows the orientation to calculate the one-dimensional depth features. The typical one-dimensional depth feature for upstairs, downstairs, and pedestrian crosswalks are demonstrated in Figure 8.

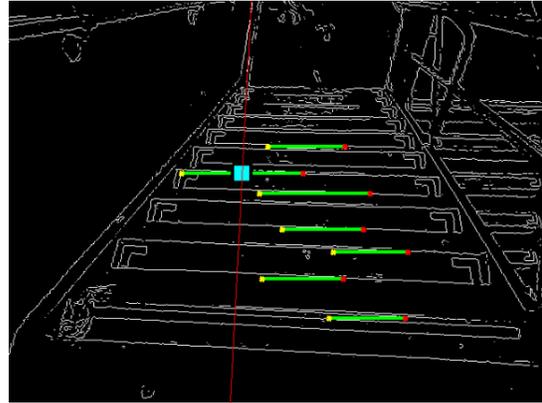


Figure 7. Orientation and position to calculate one-dimensional depth features from edge image. The blue square indicates the middle point of the longest line and the red line shows the orientation which is perpendicular to the detected parallel lines.

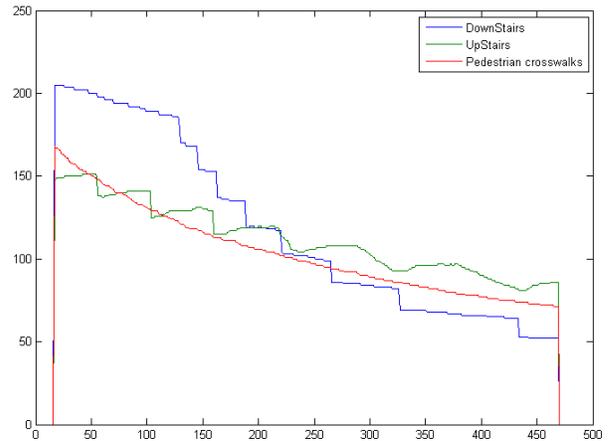


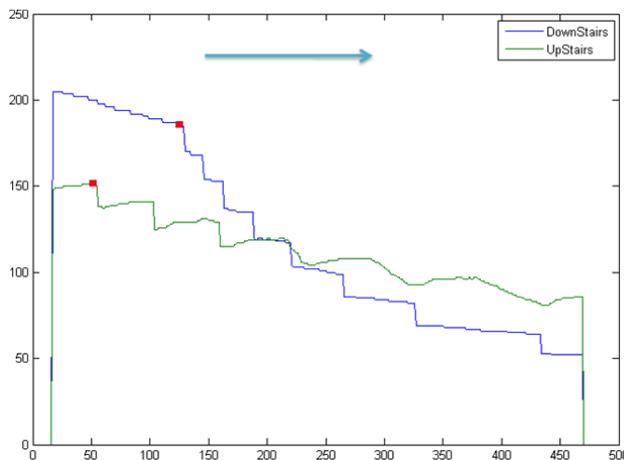
Figure 8. One-dimensional depth feature for upstairs (green), downstairs (blue), and pedestrian crosswalks (red). The horizontal axis indicates the distance from the camera in centimeters. The vertical axis represents the intensity of the depth image.

As shown in Figure 6, the resolution of depth images is 480\*640 pixels. The effective depth range of the RGBD camera is about 0.15 to 4.0 meters. The intensity value range of the depth images is [0, 255]. Therefore, as shown in Figure 8, the intensities of all the curves of the one dimension depth features are between 50 and 220 but are 0 if the distance is out of the depth range of a RGBD camera.

In order to classify upstairs, downstairs, and pedestrian crosswalks, we propose a hierarchical SVM structure by using the extracted one-dimensional depth features. The SVM builds a set of hyper-planes in an infinite-dimensional space, which can be used for classification, regression, or other tasks. The high classification accuracy can be achieved by the hyper-plane that has the largest distance to the nearest training data point of any class. In the classification section, we will have two steps, first classify pedestrian crosswalks from stairs, and then we further classify upstairs and downstairs.

### 2.3 Estimating Distance between Stairs and the Camera

When walking on stairs, we should adjust our foot height as the stairs has a steep rising or decreasing. For blind users, stairs, in particular downstairs, may cause injury if they fall. Therefore, it is essential to provide the distance information to the blind or visually impaired individuals how far is the first step of the stairs away from the camera position to remind them when they should adjust their foot height. In our method, the distance information between the first step of the stairs and the camera position will be calculated by detecting the first turning point from the one-dimension depth information as shown in Figure 9 marked as the red dots.



**Figure 9. Detecting the first turning points (red points) of the one-dimensional depth features of upstairs and downstairs.**

From the near distance to far distance (e.g., from left side to the right side as the blue line with arrow shown in Figure 9) along the one-dimensional depth features, a point  $x$  satisfies the following two conditions is considered as a turning point:

$$f(x) - f(x-1) > \lambda \quad (1)$$

or

$$f'(x) - f'(x-1) > \varepsilon \quad (2)$$

where  $f(x)$  is the intensity value of the depth information,  $\lambda$  and  $\varepsilon$  are the thresholds. In our experiment, we set  $\lambda=8$  and  $\varepsilon=50$ .

After we obtain the position of the turning point which indicates the first step of the stairs, the distance information from the camera and the first step of the stairs can be read from the original RGBD depth data. This distance will be provided to the blind traveler by speech.

## 3. Experiments and Discussion

### 3.1 Database

To evaluate the effectiveness and efficiency of the proposed method, we collect two databases: a testing database and a training database. The testing database contains 106 stairs including 56 upstairs and 50 downstairs, 52 pedestrian crosswalks, and 70 negative images which contain neither stairs nor pedestrian crosswalks. Some of negative images contain objects structured with a group of parallel lines such as bookshelves. The training database contains 30 images for each category to train the SVM classifiers. The images in the databases include small changes of camera view angles  $[-30^\circ, 30^\circ]$  because the visually impaired people pay more attention to the area in front of them. The experiment example used in our algorithm is shown in Figure 10. The first row displays examples of upstairs with different camera angles and the second row shows the corresponding depth images. Similarly, the third and fourth rows are the RGB depth images for examples of downstairs, and the fifth and sixth rows are the examples of pedestrian crosswalks.

### 3.2 Experimental Results

We evaluate the accuracy of the detection and the classification of our proposed method. The proposed algorithm achieves an accuracy of detection rate at 91.14% among the positive image samples and 0% false positive rate as shown in Table 1. For the detection step, we correctly detect 103 stairs from 106 images, and 41 pedestrian crosswalks from 52 images of pedestrian crosswalks. Here, positive image samples indicate images containing either stairs or pedestrian crosswalks, and negative image samples indicates images containing neither stairs nor pedestrian crosswalks.

The negative samples include some objects such as bookshelves, which are constructed similar edges as stairs and pedestrian crosswalks as shown in Figure 11.

With the current camera configuration, in general, only one to two shelves can be captured. The detected parallel lines will not meet the constraint conditions as described in Section 2.1. Therefore, the bookshelves will not be detected as candidates of stairs and pedestrian crosswalks.

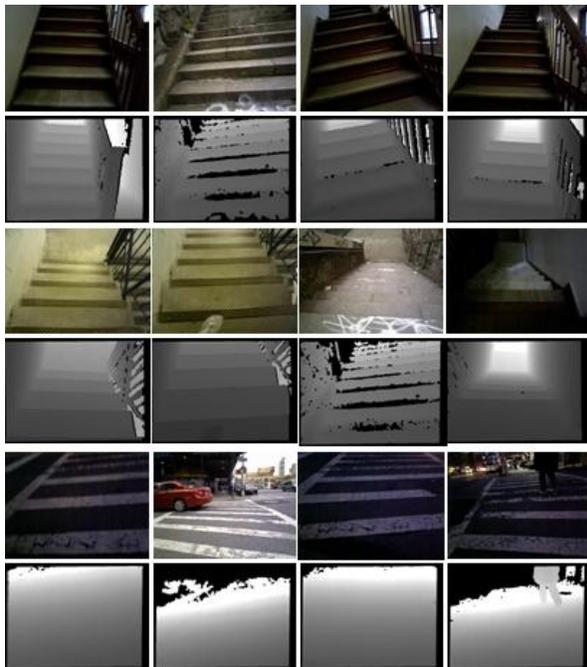


Figure 10. Examples of RGB and depth images for upstairs (1<sup>st</sup> and 2<sup>nd</sup> rows), downstairs (3<sup>rd</sup> and 4<sup>th</sup> rows), and pedestrian crosswalks (5<sup>th</sup> and 6<sup>th</sup> rows) in our database.

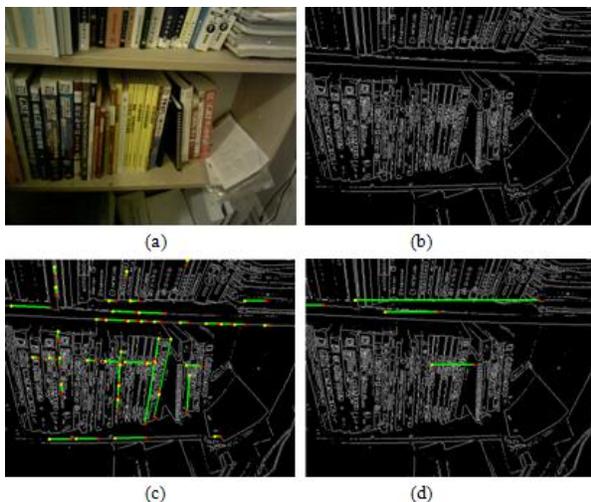


Figure 11. Negative examples of a bookshelf which has similar edge lines to stairs and pedestrian crosswalks.

In order to classify stairs and pedestrian crosswalks, the detected positive images are input into a SVM-based classifier. As shown in Table 2, our method achieves a classification rate for the stairs and pedestrian crosswalks at 95.8% which correctly classified 138 images from 144 detected candidates. Total of 6 images of stairs are wrongly classified as pedestrian crosswalks.

For stairs, we further classify they are upstairs or downstairs by inputting the one-dimensional depth features into a different SVM classifier. We achieve an accuracy rate of 90.2%. More details of the classification of upstairs and downstairs are listed in Table 3.

Table 1. Detection accuracy of stairs and pedestrian crosswalks

Classes	No. of Samples	Correctly Detected	Detection Accuracy
Stairs	106	103	97.2%
Crosswalks	52	41	78.9%
Negative samples	70	70	100%
<b>Average</b>	<b>228</b>	<b>214</b>	<b>93.9%</b>

Table 2. Accuracy of classification between stairs and pedestrian crosswalks

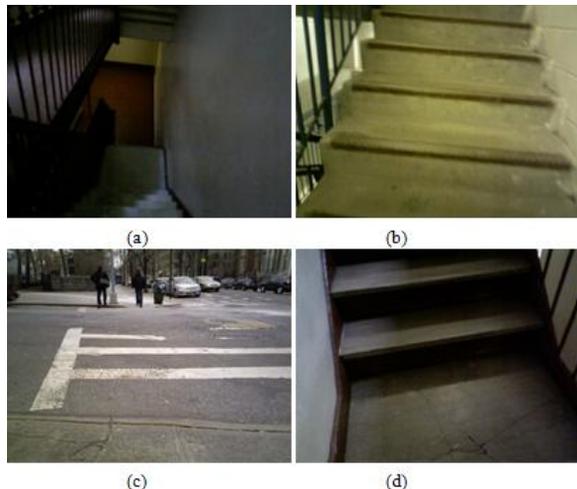
	Stairs	Crosswalks
Stairs	97	0
Crosswalks	6	41

Table 3. Accuracy of classification between upstairs and downstairs

	Upstairs	Downstairs
Upstairs	48	5
Downstairs	5	45

In database capture, we observe that it is hard to capture good quality depth images of pedestrian crosswalks compared to capture images of stairs. The main reason is the current RGBD cameras cannot obtain good depth information for outdoor scenes if the sunshine is too bright. Therefore, the field of view of the obtained depth maps is restricted compared to the RGB images. Some of the images our method cannot handle are shown in Figure 12. For example, the depth information of some parts of the images (see the 2<sup>nd</sup> and 4<sup>th</sup> columns of the 6<sup>th</sup> row of Figure 12) is missing. Furthermore, the white band patterns of pedestrian crosswalks are often disappeared caused by the long time exposure and no well maintained as shown in

Figure 11(c). In this case, it will be hard to extract parallel lines to satisfy the candidate detection constraints we described in Section 2.1. In our method, stairs with less than 3 steps will not be able to be detected, as shown in Figure 12 (a) and (d).



**Figure 12. Examples of our proposed method fails. (a) Downstairs with poor illumination; (b) Upstairs with less detected lines caused by noise; (c) Pedestrian crosswalks with missing white patterns; and (d) Stairs with less steps.**

#### 4. Conclusion and Future Work

We have developed a novel method for automatic detection of pedestrian crosswalks, upstairs, and downstairs by using a RGBD camera to improve the travel safeness of the blind and visually impaired people. The proposed method can run in real time. Our method has been evaluated on the database of stairs and pedestrian crosswalks, and achieved accuracy rates of 91.1% for detection stairs and pedestrian crosswalks from scene images, 95.8% for classification of stairs and pedestrian crosswalks, and 90.3% for classification of upstairs and downstairs,

Our further research will focus on enhancing our algorithm to handle stairs and pedestrian crosswalks with large perspective projections, more types of objects, user interface study with evaluation by blind subjects.

#### Acknowledgment

This work was supported by NIH 1R21EY020990, NSF IIS-0957016 and EFRI-1137172, DTFH61-12-H-00002, ARO W911NF-09-1-0565, Microsoft Research, and CITY SEEDs grant.

#### References

- [1] R. Advanyi, B. Varga, and K. Karacs, "Advanced crosswalk detection for the Bionic Eyeglass," 12th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA), pp.1-5, 2010
- [2] M. Bousbia-Salah, A. Redjati, M. Fezari, and M. Bettayeb, "An Ultrasonic Navigation System for Blind People," IEEE International Conference on Signal Processing and Communications (ICSPC), pp.1003-1006, 2007.
- [3] J. Coughlan and H. Shen, A fast algorithm for finding crosswalks using figure-ground segmentation. The 2nd Workshop on Applications of Computer Vision, in conjunction with ECCV, 2006.
- [4] M. Everingham, B. Thomas, and T. Troscianko, Wearable Mobility Aid for Low Vision Using Scene Classification in a Markov Random Field Model Framework, International Journal of Human Computer Interaction, vol. 15, pp.231-244, 2003.
- [5] V. Ivanchenko, J. Coughlan, and H. Shen, Detecting and Locating Crosswalks using a Camera Phone, Computers Helping People with Special Needs Lecture Notes in Computer Science, Vol. 5105, PP. 1122-1128, 2008.
- [6] G. Kao, "FM sonar modeling for navigation" Technical Report, Department of Engineering Science, University of Oxford. 1996.
- [7] R. Kuc, "A sonar aid to enhance spatial perception of the blind: engineering design and evaluation", IEEE Transactions on Biomedical Engineering Vol. 49 (10), pp. 1173-1180, 2002.
- [8] B. Laurent and T. Christian, "A sonar system modeled after spatial hearing and echolocating bats for blind mobility aid", International Journal of Physical Sciences Vol. 2 (4), pp. 104-111, April, 2007
- [9] L. Lausser, F. Schwenker, and G. Palm, "Detecting zebra crossings utilizing AdaBoost", European Symposium on Artificial Neural networks, Advances in computational intelligence and learning, 2008.
- [10] J. Liu, J. B. Liu, L. Q. Xu, and W. Jin, "Electronic travel aids for the blind based on sensory substitution", The 5<sup>th</sup> International conference on Computer Science and Education (ICCSE), pp. 328-1331, 2010.
- [11] C. Morland, D. Mountain "Design of a sonar system for visually impaired humans" , Proceedings of the 14th International Conference on Auditory Display, Paris, France June 24 - 27,

2008

- [12] S. Shoval, I. Ulrich, and J. Borenstein, "Computerized Obstacle Avoidance Systems for the Blind and Visually Impaired," Invited chapter in "Intelligent Systems and Technologies in Rehabilitation Engineering." Editors: Teodorescu, H.N.L. and Jain, L.C., CRC Press, ISBN/ISSN: 0849301408, pp. 414-448.
- [13] S. Se, "Zebra-crossing Detection for the Partially Sighted", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 211-217, 2000.
- [14] S. Se and M. Brady, "Vision-based Detection of Stair-cases", Proceedings of Fourth Asian Conference on Computer Vision (ACCV), pp. 535-540, 2000.
- [15] Y. Tian, X. Yang, C. Yi, and A. Ardit, "Toward a Computer Vision-based Wayfinding Aid for Blind Persons to Access Unfamiliar Indoor Environments," Machine Vision and Applications, 2012, DOI: 10.1007/s00138-012-0431-7.
- [16] Seeing with Sound – The vOICe: <http://www.seeingwithdound.com/>
- [17] Uddin and T. Shioyama, "Bipolarity and Projective Invariant-Based Zebra-Crossing Detection for the Visually Impaired," 1st IEEE Workshop on Computer Vision Applications for the Visually Impaired, 2005.