RGB-D Image-Based Detection of Stairs, Pedestrian Crosswalks and Traffic Signs

Shuihua Wang, Hangrong Pan, Chenyang Zhang and Yingli Tian*

Department of Electrical Engineering,
The City College of New York,
City University of New York, New York, NY 10016 USA
Email: { swang15, hpan05, czhang10, 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, pedestrian crosswalks, and traffic signals based on RGB-D (Red, Green, Blue, and Depth) images. 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 (Red, Green, and Blue) channels. Then, the Depth channel is employed to recognize pedestrian crosswalks and stairs. The detected stairs are further identified as stairs going up (upstairs) and stairs going down (downstairs). The distance between the camera and stairs is also estimated for blind users. Furthermore, the traffic signs of pedestrian crosswalks are recognized. The detection and recognition results on our collected datasets demonstrate the effectiveness and efficiency of our proposed framework.

Keywords — blind; visually impaired; wayfinding and navigation; RGB-D camera; object recognition

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.

Many electronic mobility assistant systems are developed based on converting sonar information into an audible signal for the visually impaired persons to interpret [3, 10, 12, 13, 18]. However, they only provide limited information. Recently, researchers have focused on interpreting the visual information into a high level representation before sending it to the visually impaired persons. Coughlan *et al.* [6] developed a method of finding crosswalks based on figure-ground segmentation, which they built in a graphical model framework for grouping geometric features into a coherent structure. Ivanchenko *et al.* [9] 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.* [24] 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*. [25] 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 a homograph search model. The "vOICe" system [29] is a commercially available vision-based travel aid that converts image information to sound. The system contains a head-mounted camera, stereo headphones and a laptop. Uddin *et al.* [30] 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. Omachi *et al.* [19] proposed an image-based traffic sign detection method using image shape and color information. They further improved their method by adding the traffic sign structure based on Hough transform as a critical factor [20]. Charette and Nashashibi implemented a real time image processing system for traffic sign based on the generic "adaptive templates" [5]. Alefs and Eschemann designed a computer vision-based system for reliable road sign detection [2]. The detection system is based on the method of feature selection and matching using edge orientation histograms [4]. Everingham *et al.* [8] 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.* [13] introduced a visual zebra crossing detector based on the Viola-Jones approach. Pallejà *et al.* developed an Electronic White Cane based on a LIDAR, a Tri-Axial Accelerometer and a Tactile Belt to help blind people [21]. Shoval *et al.* [23] 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's group has developed a series of computer vision-based methods for blind people to independently access and navigate unfamiliar environments [26-28, 31-33, 35, 36].

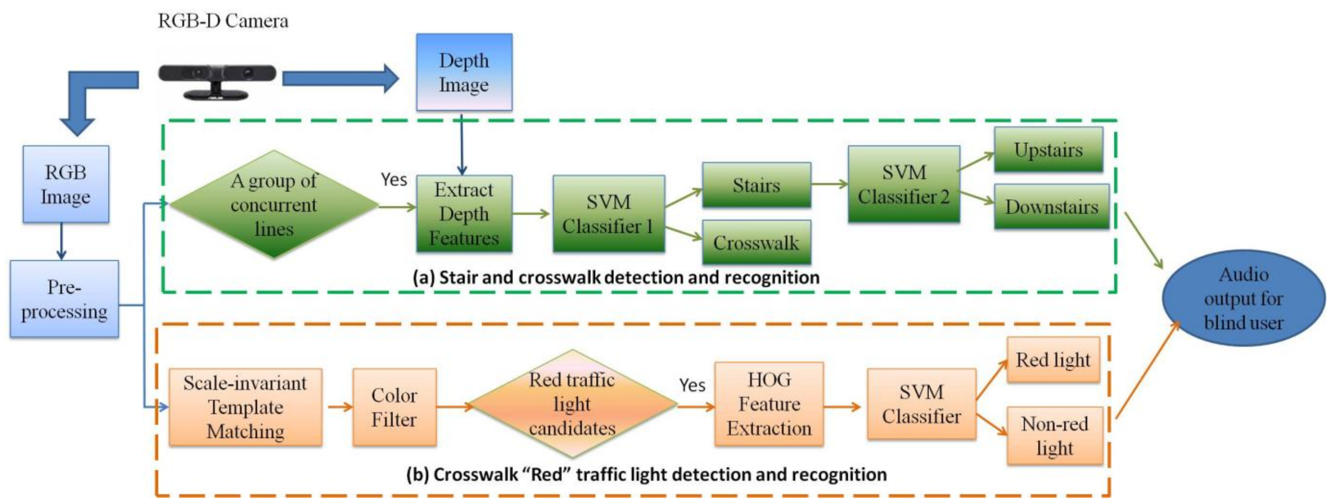


Figure 1. Flowchart of the proposed algorithm for stair, pedestrian crosswalk, and crosswalk "*Red*" traffic sign detection and recognition.

In this paper, we propose a computer vision-based framework to detect stair-cases, pedestrian crosswalks and traffic signs. The computer vision-based wayfinding and navigation aid for blind persons integrates an RGB-D camera [21], a microphone, a portable computer, and a speaker connected by Bluetooth for audio description of objects identified. In our prototype system, an HP mini laptop is employed to conduct image processing and data analysis. The RGB-D camera mounted on the user's belt is used to capture videos of the environment and connected to the HP mini laptop via a USB connection. The presence of environmental objects (stairs, crosswalks, traffic signs, etc.) is described to the blind user by a verbal display with minimal distraction to the user's sense of hearing. The user can control the system by speech input via microphone. The recent introduction of the cost-effective RGB-D cameras eases the task [17, 22]. We employ the RGB-D cameras based on the following advantages: a) RGB-D cameras contain both an RGB camera and a 3D depth camera which can provide more information of the scene; b) they work well in a low light environment; c) they are low-cost (about US$150); and d) they are efficient for real-time processing. The RGB-D camera captures RGB images at a resolution of 640x480 and depth points at 30 frames per second. The RGB-D cameras field of view is about 60 degrees. Compared to existing work [15, 16, 34] of staircase detection which only depend on RGB videos or stereo cameras, our proposed method is more robust and efficient. More importantly, our framework integrates multiple functions including detection of staircases, crosswalks, and traffic signs.

As shown in Figure 1, our whole framework consists of two main components: a) detection and recognition of stairs and pedestrian crosswalks; and b) detection and recognition for crosswalk "*red*" traffic signs. For stair and crosswalk detection and recognition, first, a group of parallel lines are detected via Hough transform and line fitting with geometric constraints from RGB information. In order to distinguish stairs and pedestrian crosswalks, we extract the feature of one dimensional depth information according to the direction of the longest detected line from the depth image. Then the feature of one dimensional depth information is employed as the input of a support vector machine (SVM) based classifier [4] to recognize stairs and pedestrian crosswalks. For stairs, a further detection of the upstairs and downstairs is conducted. Furthermore, we estimate the distance between the camera and stairs for the blind user. For crosswalk "*red*" traffic sign detection and recognition, first, a scale-invariant template matching method with a color filter is applied to RGB images to detect the candidates of the "*red*" traffic signs. Then the Histograms of Oriented Gradients (HOG) features [7] are extracted as the input of a SVM-based classifier to further distinguish the "red" traffic signs and the "non-red" traffic signs.

The reminder of paper is organized as follows. Section 2 describes the methodology of our proposed algorithm which includes the following main components: 1) detection of stair-cases or pedestrian crosswalks based on RGB image analysis; 2) classification between stairs from pedestrian crosswalks; 3) recognition of upstairs and downstairs. Section 3 introduces the proposed method for pedestrian traffic sign detection and recognition. Section 4 evaluates the effectiveness and efficiency of proposed method. Section 5 summarizes the paper.

2. Methodology of RGB-D Image-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 consist 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 by applying a Sobel operator 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 are parallel for both stairs and pedestrian crosswalks. Therefore, a group of concurrent parallel lines 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 transfer to the polar coordinates, denoted r and θ, for the lines in the Hough transform.

The parameter r represents the distance between starting point and the ending point of the line (i.e. the line length). The starting point of the line is used as the origin, while θ is the angle between the edge line and the horizontal direction, then the equation of a line can be represented as:

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

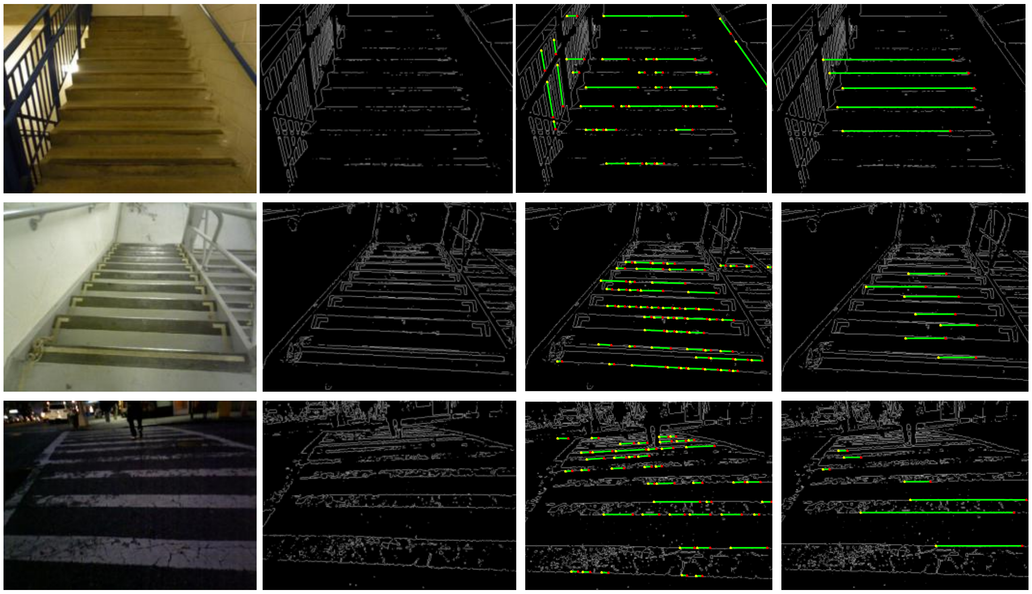


Figure 2. Example of upstairs (row 1), downstairs (row 2), and Pedestrian crosswalks (row 3). (a) Original image; (b) edge detection; (c) line detection; (d) concurrent parallel lines detection (yellow dots represent the starting points, red dots represent the ending points of the lines, and green lines represent the detected lines.)

Parallel lines have similar θ . The Hough transform based algorithm to detect parallel lines from RGB images is summarized as following:

<u>Step1</u>: Detect edge maps from the RGB image by edge detection.

<u>Step2</u>: Compute the Hough transform of the RGB image to obtain r and θ.

<u>Step3</u>: Calculate the peaks in the Hough transform matrix.

<u>Step4</u>: Extract lines in the RGB image.

<u>Step5</u>: 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 Figure 2(c), the detected parallel lines of stairs and pedestrian crosswalks are marked as green, while yellow dots and red dots represent the staring points and the ending points 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 is less than a threshold. In general, stairs and pedestrian crosswalks contain multiple parallel lines with a reasonable length. If the length of a line ≤ ϕ , then the line does not belong to the line group. And if the number of parallel lines less than β , the scene image is a negative image which does not contain stairs and pedestrian crosswalks. In our experiment, we set the line length ϕ as 60 pixels in the acquired images and the number parallel lines β as 5 based on the camera configurations to achieve the best performance.

2.2. Recognizing Pedestrian Crosswalks and Stairs from Depth Images

2.2.1. Extracting One-dimensional Depth Feature

Based on the above algorithm, we can detect the candidates of stairs and pedestrian crosswalks by detecting parallel lines under the constraints in a scene image captured by an RGBD camera. From the depth images, we observe that upstairs have rising steps and downstairs have descending steps, and pedestrian crosswalks are flat with smooth depth change as shown in Figure 3. Considering the safeness for the visually impaired people, and the further application of robotics, it is necessary to classify the different stairs and pedestrian crosswalks into the correct categories.

In order to distinguish stairs and pedestrian crosswalks, we first calculate the orientation and position of the feature line in the edge image to extract the one-dimensional feature from depth information. As shown in Figure 4(a), the orientation of the feature line is perpendicular to the parallel lines detected from RGB images. The position of the feature line is determined by the middle point of the longest line of the parallel lines. In Figure 4(a), the blue square indicates the middle point of the longest line and the red line is the feature line which indicates the orientation to calculate the one-dimensional depth feature. The typical one-dimensional depth feature for upstairs (green curve), downstairs (blue curve), and pedestrian crosswalks (red curve) are demonstrated in Figure 4(b).

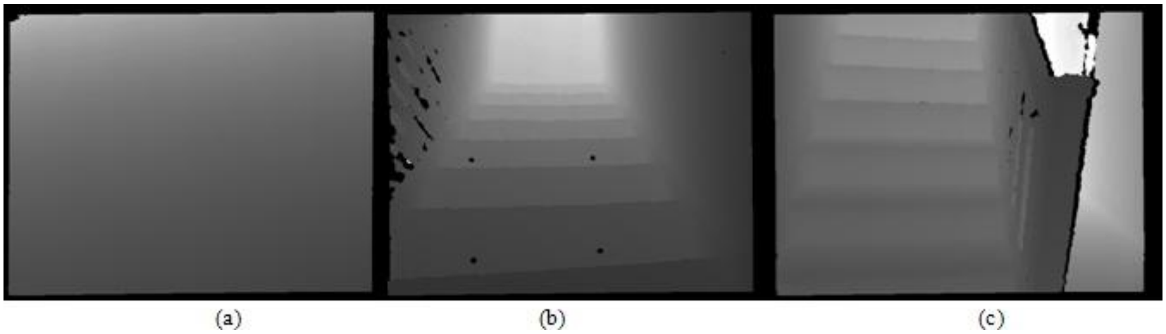


Figure 3. Depth images of (**a**) pedestrian crosswalks, (**b**) downstairs, and (**c**) upstairs.

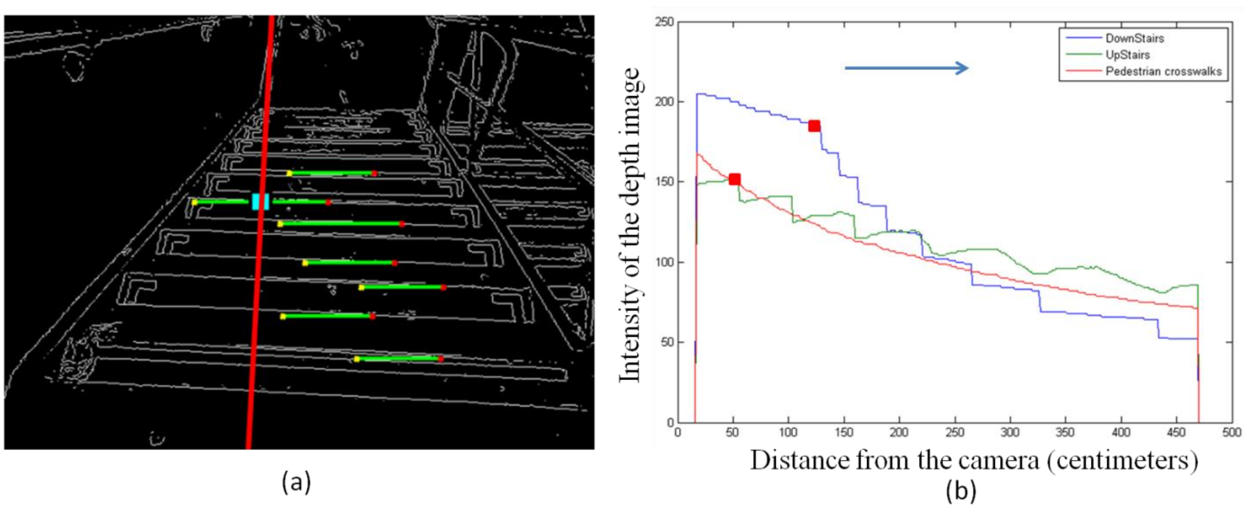


Figure 4. (a) The orientation and position of the feature line to extract one-dimensional depth features from the edge image. The blue square indicates the middle point of the longest line and the red line shows the orientation which is perpendicular to the detected parallel lines. (b) One-dimensional depth feature for upstairs (green curve), downstairs (blue curve), and pedestrian crosswalks (red curve). The red squares indicate the first turning points of the one-dimensional depth features of upstairs and downstairs.

The resolution of depth images in Figure 3 is *640*480* pixels. The effective depth range of the RGBD camera is about *0.15* to *4.7* meters. The intensity value range of the depth images is [0, 255]. Therefore, as shown in Figure 4(b), the intensities of the one dimensional depth feature for upstairs, downstairs, and crosswalks are between *50* and *220* (the vertical axis) but are *0* if the distance is out of the depth range of an RGBD camera. Therefore, the one-dimensional depth feature is a feature vector with 480 dimensions. We observe that the curve for crosswalks is very flat while the curves of upstairs and downstairs are with intensity changes of step shape which can be used to distinguish stairs and crosswalks.

2.2.2. Recognizing Crosswalks, Upstairs, and Downstairs

In order to classify upstairs, downstairs, and pedestrian crosswalks, we propose a hierarchical SVM structure by using the extracted one-dimensional depth feature vector as input. The SVM-based classifier 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. As shown in Figure 1(a), the classification processing includes two steps: 1) one SVM-based classifier to identify pedestrian crosswalks from stairs. 2) For those detected stairs, one more SVM-based classifier to further identify upstairs and downstairs. In our implementation, the Radial Basis Function (RBF) is employed. In the training procedure, the feature vectors (i.e. the one-dimensional depth feature vector with 480 dimensions) and the category labels (e.g. 0 indicates stairs and 1 indicates crosswalks) from all the training data are used as the input of the SVM. The SVM classification algorithm finds the maximum margin hyper-plane that classifies the data belonging to different categories.

2.3. Estimating Distance between Stairs and the Camera

When walking on stairs, we should adjust our walking speed and 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 of the first step from the camera position to the blind or visually impaired individuals to remind them when they should adjust their walking speed and foot height. In our method, the distance information between the first step of the stairs and the camera position is calculated by detecting the first turning point from the one-dimensional depth feature as shown in Figure 4(b) marked as the red squares.

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 4(b) along the one-dimensional depth feature, a point x satisfies the following two conditions is considered as a turning point:

$$\|f(x) - f(x-1)\| > \lambda \; and \; \|f'(x) - f'(x-1)\| > \varepsilon \qquad (2)$$

where *f(x)* is the intensity value of the depth information, λ and ε are the thresholds which are determined by the RGBD camera configuration. In our experiment, we observe that the best results with λ =8 and ε =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 and provided to the blind traveler by speech.

3. Methodology of RGB Image-Based Crosswalk Traffic Sign Detection and Recognition

3.1. Method Overview

Blind and visually impaired pedestrians face critical safety challenges while crossing street at intersections, especially in unfamiliar environments. In this paper, we focus on crosswalk traffic sign detection and recognition. As shown in Figure 1(b), our method of crosswalk traffic sign detection and recognition contains four main steps: 1) Preprocessing of the original input images which includes down-sampling the original images and Gaussian smoothing to remove noise. 2) Detection of the candidate traffic sign regions by applying scale-invariant template matching and color segmentation based filtering. 3) Extraction of the HOG features from the detected candidate image regions. 4) Recognition of the crosswalk "*red*" sign from the non-red sign using a SVM classifier.

3.2. Detecting Candidates of the Crosswalk Red Sign

In order to detect the candidates of the crosswalk red sign in images, we first convert the RGB image to grayscale, and then apply the scale-invariant template matching method [11] to find the largest value of the cross correlation coefficients between the template images and the target image. Cross correlation is a standard method of estimating the degrees to which two sets of data are correlated with each other. Given a template *t(i, j)* of size *mxn* and an image *f(x, y) of size WxH*, the correlation between the template and the image for the pixel *(x, y)* can be calculated as:

$$c(x,y) = \sum_{i=-m/2}^{m/2} \sum_{j=-n/2}^{n/2} f(x+i, y+j)t(i,j). \qquad (3)$$

The template size *mxn* is normally be an odd numbers and is smaller than the image size *WxH*. Then normalized cross correlation is calculated based on the correlation in order to remove the linear distortions among different images. The normalized correlation coefficient can be expressed as below:

$$r(x,y) = \frac{\sum_{i,j}[f(x+i,\ \ y+j)-\bar{f}][t(i,j)-\bar{t}\,]}{\{\sum_{i,j}[f(x+i,y+j)-\bar{f}]^2 \sum_{i,j}[t(i,j)-\bar{t}\,]^2\}^{0.5}} \qquad (4)$$

where summation is computed over the region shared by *f(x, y)* and *t(i, j)*, $\bar{t}$ is the average value of the template and $\bar{f}$ is the average value of the image intensities in the intersection area with the template *t*.

In order to make the system invariant to the scale changes and detect candidate regions with different sizes from the input images, a scale index *k* is introduced to enhance the performance of the template matching method. In our system, we first set a standard template of the crosswalk red sign at size 40x40 pixels. Then, the range of scale index *k* is defined as [0.5, 1.5] with an interval of 0.1. This allows our system to detect crosswalk red traffic sign with size between 20x20 pixels and 60x60 pixels. For each *k*, the size of the resized template can be calculated as:

$$\mathrm{row}' = \mathrm{row} \times \mathrm{k};\ \text{and}\ \mathrm{col}' = \mathrm{col} \times \mathrm{k}; \qquad (5)$$

where *row'* and *col'* is the number of rows and columns respectively for the resized template image. A cubic interpolation is applied during the process of resizing. Figure 5 displays the searching results by performing the scale-invariant template matching method on the target images, where the regions in the red boxes indicate the matched regions.

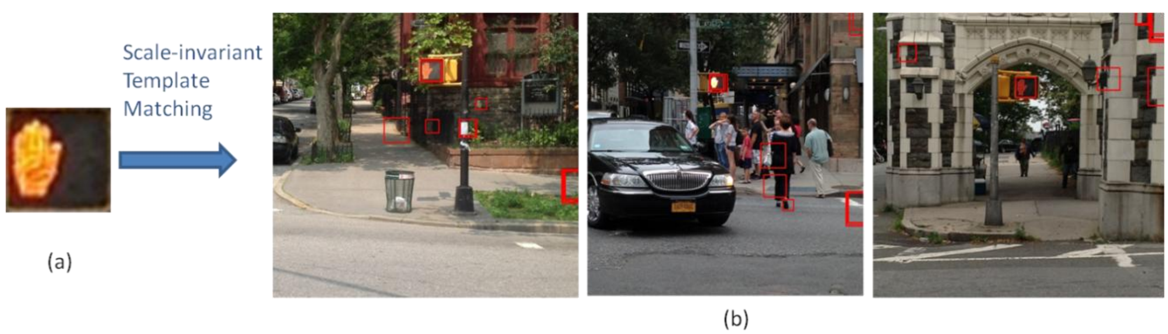


Figure 5. Examples of the detected candidates of crosswalk "red" traffic sign by the scale-invariant template matching. (a) Template of crosswalk "red" traffic sign; (b) detected candidates (in red boxes).

3.3. Color-based Filter for Candidate Regions

As shown in Figure 5(b), there are many false red sign regions in the detected candidates after performing the scale-invariant template matching method. To reduce the number of the false candidates, a color-based filter is further performed in the candidate regions. In our framework, the normalized color *rgb* is employed to eliminate the brightness variation which can be calculated by:

$$\mathrm{r} = \frac{\mathrm{R}}{\mathrm{R+G+B}} \times 255, \mathrm{g} = \frac{\mathrm{G}}{\mathrm{R+G+B}} \times 255, \mathrm{b} = 255 - \mathrm{r} - \mathrm{b} \qquad (6)$$

where R, G, and B is the intensity of the red, green and blue channel respectively. We keep the satisfied candidate regions if the ratio of red pixel exceeds a certain threshold. In our experiments, a pixel is detected as a red pixel if $r > 130$, $g < 110$, and $b < 50$. As shown in Figure 6, many false "*red*" traffic sign candidates are removed after applying the color-based filter. However, there are still some false candidates. To improve the detection accuracy, we further classify the detected "*red*" traffic sign candidates by a SVM classifier.

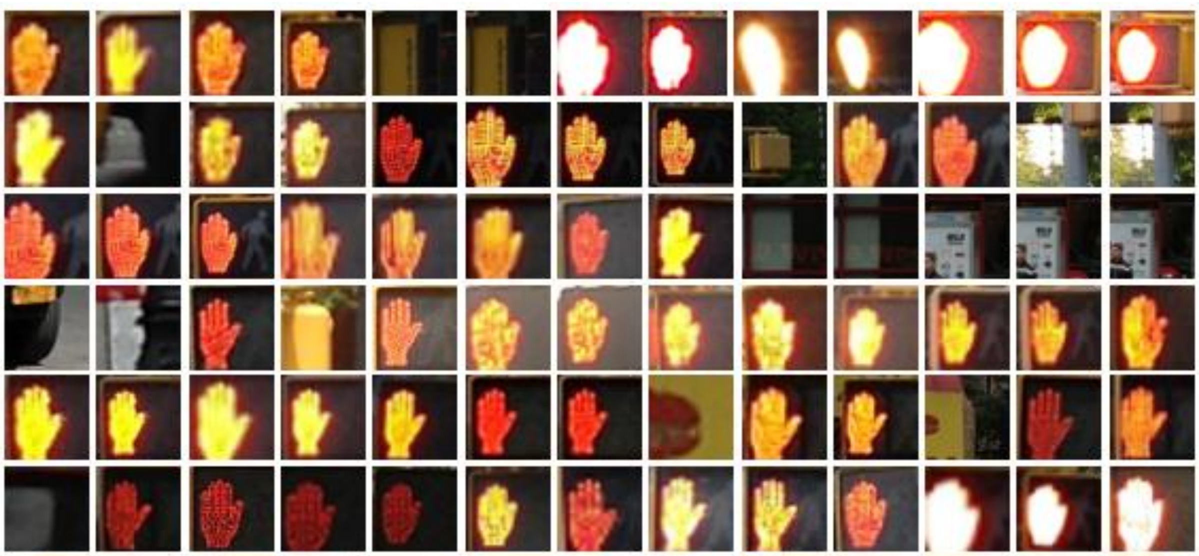


Figure 6. Examples of the detected candidates of crosswalk "red" traffic sign by the scale-invariant template matching and color segmentation.

3.4. Recognizing Red Traffic Sign based on HOG Features of the Detected Candidate Regions

As shown in Figure 7, to further recognize if a detected "*red*" traffic sign candidate is correct, we first extract HOG features for each candidate region. The appearance and shape information of the detected candidate regions is described by the distribution of gradients. In order to obtain more accurate results, each candidate region is further divided into small sub-regions, and then the histogram of oriented gradients for the pixels within each sub-region is combined as the HOG features. In our experiment, each detected candidate region is divided into 6 by 6 = 36 small sub-regions and oriented gradients are divided into 32 bins. So the dimension of the HOG feature vector for each detected candidate region is $6 \times 6 \times 32 = 1152$.

Since the shape information of the crosswalk red sign or non-red sign is mainly contained in the *R* (red) channel of the detected image region, the HOG features are extracted only from the R channel to reduce the computation cost. After obtaining the HOG features of each candidate region, the feature vectors are used as the input to train and test a SVM classifier with linear kernel for classification.

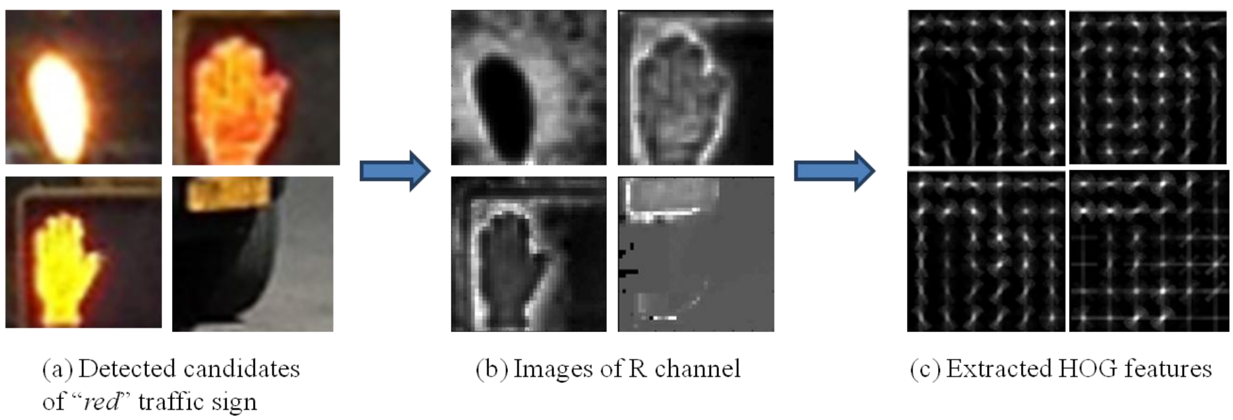


Figure 7. Extracting HOG features for the detected candidate regions of crosswalk "red" traffic sign from the Red channel.

4. Experiment Results

4.1. Experiment Results for Stair and Crosswalk Detection and Recognition

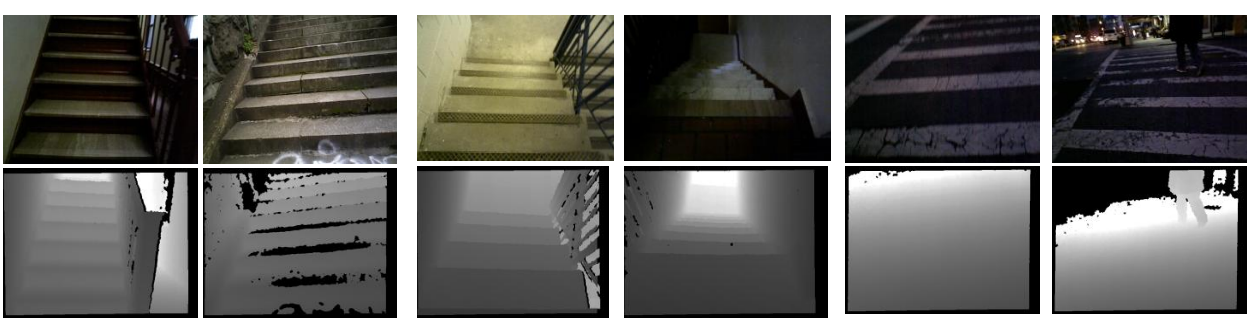
4.1.1. Stair and Crosswalk Database

To evaluate the effectiveness and efficiency of the proposed method, we have collected a database for stair and crosswalk detection and recognition. The database is divided into two sub-datasets: a testing dataset and a training dataset. The training dataset contains 30 images for each category (i.e. upstairs, downstairs, crosswalks, and nagative images which contain neither stairs nor pedestrian crosswalks) which are randomly selected to train the SVM classifiers. Then the remaining images are used for testing which contains 106 stairs including 56 upstairs and 50 downstairs, 52 pedestrian crosswalks, and 70 negative images. Some of the negative images contain objects structured with a group of parallel lines such as bookshelves. The images in the dataset include small changes of camera view angels $[-30^o, 30^o]$. Some of the experiment examples used in our algorithm are shown in Figure 8. The first row displays some RGB images of upstairs (Figure 8(a)), downstairs (Figure 8(b)), and crosswalks ((Figure 8(c)) with different camera angles and the second row shows the corresponding depth images.

4.1.2. Experiment Results of Detection and Recognition of Stairs and Crosswalks

We have evaluated 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 indicate 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 9. With the current camera configuration, in general, only one to two shelves can be captured. The detected parallel lines do not meet the constraint conditions as described in Section 2.1. Therefore, the bookshelves are not detected as candidates of stairs and pedestrian crosswalks.



(a) Upstairs (b) Downstairs (c) Crosswalks

Figure 8. Examples of RGB (1st row) and depth images (2nd row) for (a) upstairs, (b) downstairs, and (c) pedestrian crosswalks in our database.

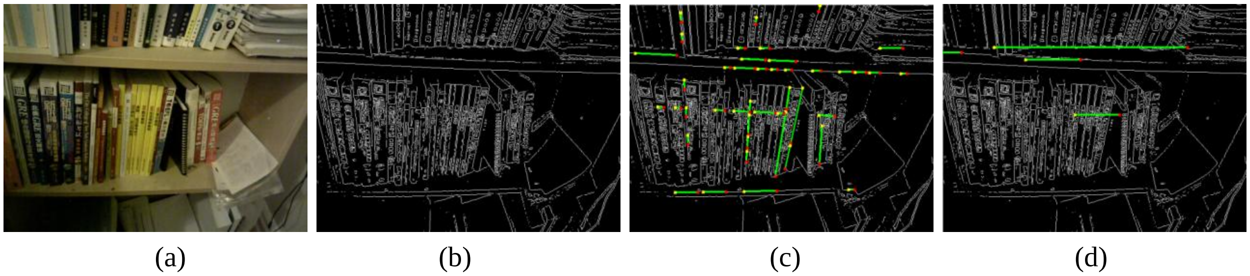


Figure 9. Negative examples of a bookshelf which has similar parallel edge lines to stairs and crosswalks.

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. A total of 6 images of stairs are wrongly classified as pedestrian crosswalks. All the detected pedestrian crosswalks are correctly classified.

For stairs, we further classify them as upstairs or downstairs by inputting the one-dimensional depth feature 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.

Our system is implemented by using MATLAB without optimization. The average processing time for stair and crosswalk detection and recognition of each image is about 0.2 seconds on a computer with 2.4GHz processor. This can be easily sped up 10-100 times in C++ with optimization.

Table 1. Detection accuracy of stairs and pedestrian crosswalks

Classes	No. of Samples	Correctly Detected	Missed	Detection Accuracy
Stairs	106	103	3	97.2%
Crosswalks	52	41	11	78.9%
Negative samples	70	70	0	100%
Total	**228**	**214**	**14**	**93.9%**

Table 2. Accuracy of classification between stairs and pedestrian crosswalks. In a total of 144 detected candidates of stairs (103) and crosswalks (41), all 41 crosswalks and 97 stairs are correctly classified. 6 stairs are wrongly classified as crosswalks.

Category	Total	Classified as Stairs	Classified as Crosswalks
Stairs	103	97	6
Crosswalks	41	0	41

Table 3. Accuracy of classification between upstairs and downstairs. In a total of 103 detected candidates of stairs (53 for upstairs and 50 for downstairs), 48 upstairs and 45 downstairs are correctly classified. 5 upstairs and 5 downstairs are wrongly classified.

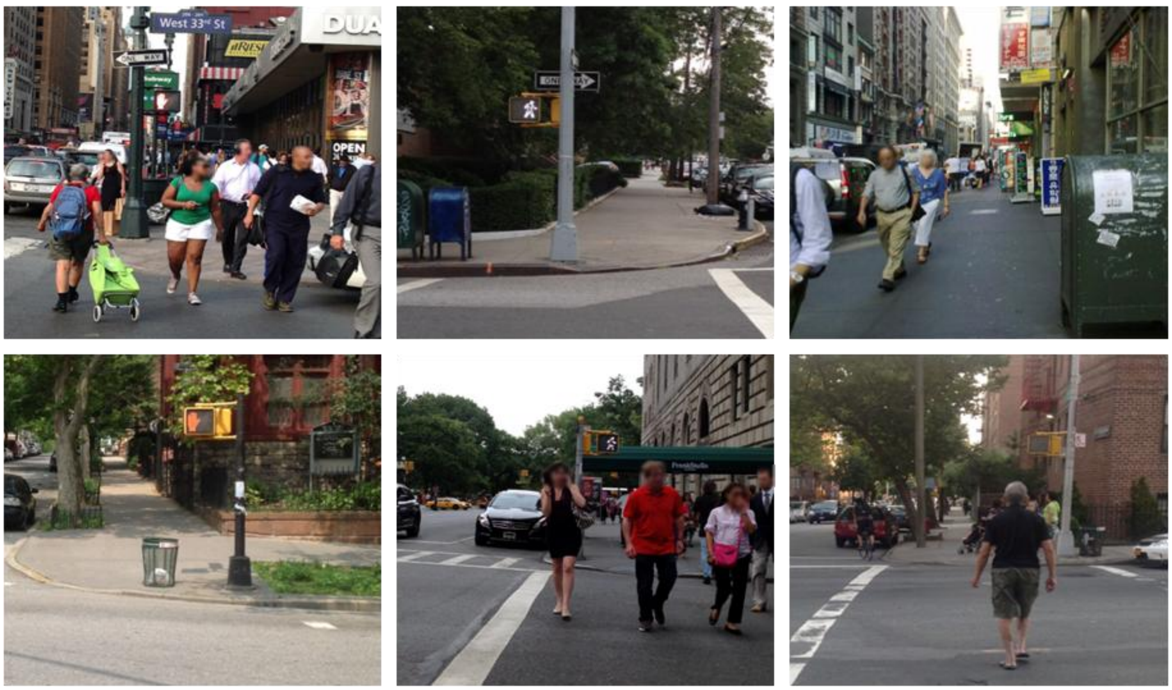
Category	Total	Classified as Upstairs	Classified as Downstairs
Upstairs	53	48	5
Downstairs	50	5	45

4.1.3. Limitations of the Proposed method of Stair and Crosswalk Recognition

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 10. For example, the depth information of some parts of the images is missing. Furthermore, as shown in Figure 10(c), the zebra patterns of pedestrian crosswalks are not always visible caused by the long time use. In this case, it is hard to extract enough number of parallel lines to satisfy the candidate detection constraints we described in Section 2.1 for stair and crosswalk detection. In our method, stairs with less than 3 steps (only have 3 or 4 parallel lines) cannot be detected, as shown in Figure 10(d).

(a) (b) (c) (d)

Figure 10. Examples of our proposed method of stair and crosswalk detection fails. (**a**) Downstairs with poor illumination; (**b**) Upstairs with less detected lines caused by noise; (**c**) Pedestrian crosswalks with missing zebra patterns; and (**d**) Stairs with less steps.



(a) (b) (c)

Figure 11. Example images from our crosswalk traffic sign database. (a) Images with "*red*" traffic sign; (b) images with "*walking*" traffic sign; (c) negative images without traffic signs. The faces are blurred for privacy protection.

4.2. Experiment Results for Crosswalk Traffic Sign Detection and Recognition

4.2.1. Crosswalk Traffic Sign Database

Our crosswalk traffic sign database consists of 412 images captured at intersection areas in New York City. Among them, 230 pictures are used for training and 182 pictures are used for testing. The images in the training and testing datasets are randomly selected. As shown in Figure 11, the database includes images with "*red*" crosswalk traffic sign (Figure 11(a)), images with "*walking*" crosswalk traffic sign (Figure 11(b)), and negative images without any crosswalk traffic sign (Figure 11(c)). Since our work only focuses on the detection and recognition of the "*red*" crosswalk traffic sign which may cause dangers for blind people, the images with "*walking*" crosswalk traffic sign and without crosswalk traffic sign are categorized as images as "*non-red*" crosswalk traffic sign during the experiment. The number of images with "*red*" sign and "*non-red*" sign is 101 and 129 in the training dataset, and 80 and 102 in the testing dataset. The images in our database also contain changes of intensity, scales, view angles, and slight rotation.

4.2.2. Experiment Results of "Red" Crosswalk Traffic Sign Detection and Recognition

As described in the methodology part (see Section 3.1), the crosswalk "*red*" sign detection and recognition system includes four main processing steps to achieve the final results. The first two steps fulfill the task of traffic sign candidate detection and the next two steps fulfill the task of recognition of red sign images from the non-red sign images. The testing dataset contains 80 images of the "red" crosswalk traffic sign and 102 images of non-red crosswalk traffic sign. As shown in Table 4, the proposed "*red*" crosswalk traffic sign detection and recognition method achieves an average accuracy rate of 95.05% , with the accuracy rate of 92.5% for "*red*" traffic sign and 97% accuracy rate for "*non-red*" traffic sign. Figure 12 displays some images with correctly detected "*red*" traffic signs (marked in red boxes).

Figure 13 displays some failure examples of our proposed method of "*red*" crosswalk traffic sign recognition under the following situations: a) images with low contrast between the lighting changes of "*red*" and "*walking*" signs; b) images with large camera view angle changes; c) images with occluded crosswalk traffic signs by passing vehicles or pedestrians; and d) images with large camera rotations.

Our system is implemented by using MATLAB without optimization. The average processing time for "*red*" traffic sign detection and recognition of each image is about 2 seconds on a computer with 2.4GHz processor. This can be easily sped up 10-100 times in C++ with optimization.

Table 4. Detection and recognition accuracy of "*red*" and "*non-red*" crosswalk traffic signs

Classes	No. of Samples	Correctly Detected & Recognized	Accuracy
Red	80	74	92.50%
Non-red	102	99	97.06%
Total	**182**	**173**	**95.05%**

Figure 12. Example results of our proposed method of "*red*" crosswalk traffic sign recognition (marked in red boxes).

(a) (b) (c) (d)

Figure 13. Examples of our proposed method of "*red*" crosswalk traffic sign recognition fails. (a) Image with low contrast between the changes of "*red*" and "*walking*" signs; (b) image with large camera view angle change; (c) image with occluded crosswalk traffic sign by passing vehicles or pedestrians; (d) image with large camera rotation.

4.3. Discussions for the Application of Blind Wayfinding and Navigation

For the application of blind wayfinding and navigation, an assistive system is not intended to replace the white cane while most blind users using. Instead, a navigation aid can help blind users to gain improved perception and better understanding of the environment so that they can aware the dynamic situation changes. Blind users are the final decision makers who make travel decision and react to local events within the range of several meters.

As described in the above sections, the proposed method achieved 97.2% detection accuracy for stairs, 78.9% detection accuracy of pedestrian crosswalks, and 92.5% detection accuracy for red crosswalk signs. Theoretically, higher detection accuracy is always better. However, in reality, it is very hard to achieve 100% detection accuracy in particular for computer vision-based methods due to the complex situations and the lighting changes. For the application to assist blind users, a high detection accuracy and a lower false negative rate for detections of red crosswalk signs and downstairs are more desirable. For example, a wrongly detected crosswalk sign (e.g. a real "red" crosswalk sign is detected as a "green" crosswalk sign) may cause serious dangers for blind users. Therefore, it is very important to design a user-friendly interface to provide meaningful feedback to blind users with the detected important information. To meet the real needs of blind users, we have conducted a survey with 10 blind subjects. We observe that most of blind users prefer higher detection accuracy but are willing to accept more meaningful information especially for users who recently lost their vision.

5. Conclusions and Future Work

We have developed a framework for automatic detection of stairs, pedestrian crosswalks and traffic signs from images to improve the travel safeness of the blind and visually impaired people. Our proposed framework has been evaluated on the databases of stairs, pedestrian crosswalks and traffic signs, and has achieved average accuracy rates of 91.1% for detecting stairs and pedestrian crosswalks from scene images, 95.8% for classification of stairs and pedestrian crosswalks, 90.3% for classification of upstairs and downstairs, and 95.05% for "red" crosswalk traffic sign recognition. Our future research will focus on enhancing our algorithm to handle stairs, pedestrian crosswalks and traffic signs with large perspective projections, to recognize more types of objects, and to conduct more user interface study with evaluation by blind subjects.

Acknowledgments

This work was supported in part by NSF grant EFRI-1137172, IIP-1343402, FHWA DTFH61-12-H-00002, ARO grant W911NF-09-1-0565, and Microsoft Research. The authors thank the anonymous reviewers for their constructive comments and insightful suggestions that improved the quality of this paper.

References:

1. Advanyi, R.; Varga, B.; Karacs, K. Advanced crosswalk detection for the Bionic Eyeglass, 12th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA), 2010; pp.1-5.
2. Alefs, B.; Eschemann, G.; Ramoser and H.; Beleznai, C. Road Sign Detection from Edge Orientation Histogram, IEEE Intelligent Vehicles Symposium, June 2007
3. Bousbia-Salah, M.; Redjati, A.; Fezari, M.; Bettayeb, M. An Ultrasonic Navigation System for Blind People, IEEE International Conference on Signal Processing and Communications (ICSPC), 2007; pp. 1003-1006.
4. Chang C.; Lin, C. LIBSVM: a Library for Support Vector Machine, 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm
5. Charette R.; Nashashibi, F. Traffic light recognition using image processing compared to learning processes, IEEE International Conference on Intelligent Robots and Systems, Oct. 2009; pp. 333-338.
6. Coughlan J.; Shen, H. A fast algorithm for finding crosswalks using figure-ground segmentation. The 2nd Workshop on Applications of Computer Vision, in conjunction with ECCV, 2006.
7. Dalal N.; Triggs, B. Histograms of Oriented Gradients for Human Detection, *Computer Vision and Pattern Recognition*, vol. 1, June 2005; pp. 886-893.
8. Everingham, M.; Thomas, B.; Troscianko, T. Wearable Mobility Aid for Low Vision Using Scene Classification in a Markov Random Field Model Framework, *International Journal of Human Computer Interaction*, vol. 15, 2003; pp.231-244.

9. Ivanchenko, V.; Coughlan, J.; Shen, H. Detecting and Locating Crosswalks using a Camera Phone, Computers Helping People with Special Needs Lecture Notes in Computer Science, Vol. 5105, 2008; pp. 1122-1128.

10. Kao, G. FM sonar modeling for navigation, Technical Report, Department of Engineering Science, University of Oxford. 1996.

11. Kim, H.; Araujo, S. Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast, *Advances in Image and Video Technology*, vol 4872, 2007; pp. 100-113.

12. Kuc, R. A sonar aid to enhance spatial perception of the blind: engineering design and evaluation, *IEEE Transactions on Biomedical Engineering,* Vol. 49 (10), 2002; pp. 1173–1180.

13. Laurent B.; Christian, T. A sonar system modeled after spatial hearing and echolocating bats for blind mobility aid, *International Journal of Physical Sciences*, Vol. 2 (4), April, 2007; pp. 104-111.

14. Lausser, L.; Schwenker, F.; Palm, G. Detecting zebra crossings utilizing AdaBoost, European Symposium on Artificial Neural networks, Advances in computational intelligence and learning, 2008.

15. Lee, Y., Leung, T., Medioni, G., "Real-time staircase detection from a wearable stereo system", ICPR 2012.

16. Lu X. and Manduchi, R., "Detection and Localization of Curbs and Stairways Using Stereo Vision" ICRA, 2005.

17. Microsoft. http://www.xbox.com/en-US/kinect, 2010.

18. Morland, C.; Mountain, D. Design of a sonar system for visually impaired humans, The 14th International Conference on Auditory Display, June, 2008.

19. Omachi M.; Omachi, S. Traffic Light Detection with Color and Edge Information, IEEE International Conference on Computer Science and Information Technology, vol.2, pp. 284-287, Aug 2009.

20. Omachi M.; Omachi, S. Detection of Traffic Using Structure Information, IEEE International Conference on Computer Science and Information Technology, Oct. 2010; pp 809-812.

21. Pallejà, T., Tresanchez, M., Teixidó, M., and Palacin, J., Bioinspired Electronic White Cane Implementation Based on a LIDAR, a Tri-Axial Accelerometer and a Tactile Belt, Sensors 2010, 10(12), 11322-11339; doi:10.3390/s101211322

22. PrimeSense. http://www.primesense.com/.

23. Shoval, S.; Ulrich, I.; Borenstein, J. Computerized Obstacle Avoidance Systems for the Blind and Visually

Impaired, Invited chapter in *Intelligent Systems and Technologies in Rehabilitation Engineering*. Editors: Teodorescu, H.; Jain, L.; CRC Press, ISBN/ISSN: 0849301408, pp. 414-448.

24. Se, S. Zebra-crossing Detection for the Partially Sighted, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2000; pp. 211-217.
25. Se S.; Brady, M. Vision-based Detection of Stair-cases, Proceedings of Fourth Asian Conference on Computer Vision (ACCV), 2000; pp. 535-540.
26. Tian, Y.; Yang, X.; Yi, C.; Arditi, A. 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.
27. Tian, Y.; Yi, C.; Arditi, A. Improving Computer Vision-Based Indoor Wayfinding for Blind Persons with Context Information, 12th International Conference on Computers Helping People with Special Needs (ICCHP), 2010.
28. Tian, Y.; Yang, X.; Arditi, A. Computer Vision-Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons, 12th International Conference on Computers Helping People with Special Needs (ICCHP), 2010.
29. Seeing with Sound – The vOICe: http//www.seeingwithdound.com/
30. Uddin, M.; Shioyama, T. Bipolarity and Projective Invariant-Based Zebra-Crossing Detection for the Visually Impaired, 1st IEEE Workshop on Computer Vision Applications for the Visually Impaired, 2005.
31. Wang S.; Tian, Y. Indoor Signage Detection Based on Saliency Map and Bipartite Graph Matching, International Workshop on Biomedical and Health Informatics (BHI), 2011.
32. Wang, S.; Yi, C.; Y. Tian, Y. Signage Detection and Recognition for Blind Persons to Access Unfamiliar Environments, *Journal of Computer Vision and Image Processing*, Vol. 2, No. 2, 2012.
33. Wang S.; Tian, Y. Detecting Stairs and Pedestrian Crosswalks for the Blind by RGB-D Camera, International Workshop on Biomedical and Health Informatics (BHI), 2012.
34. Wang S. and Wang, H., “2D staircase detection using real Adaboost”, Information, Communications and Signal Processing (ICICS), 2009.
35. Yang, X.; Tian, Y.; Yi, C.; Arditi, A. Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environments, International Conference on ACM Multimedia, 2010.

36. Yi, C.; Tian, Y. Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification, *IEEE Transactions on Image Processing*, 2012.