RGB-D Sensor-Based Computer Vision Assistive Technology for Visually Impaired Persons

Yingli Tian

Department of Electrical Engineering,
The City College of New York, New York, NY 10031 USA
Email: 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 chapter, we focus on RGB-D sensor-based computer vision technologies in application to assist blind and visually impaired persons. We first briefly review the existing computer vision based assistive technology for the visually impaired. Then we provide a detailed description of the recent RGB-D sensor based assistive technology to help blind or visually impaired people. Next, we present the prototype system to detect and recognize stairs and pedestrian crosswalks based on RGB-D images. Since both stairs and pedestrian crosswalks are featured by a group of parallel lines, Hough transform is applied 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. The detection and recognition results on our collected datasets demonstrate the effectiveness and efficiency of our developed prototype. We conclude the chapter by the discussion of the future directions.

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

1. Introduction

Of the 314 million visually impaired people worldwide, 45 million are blind [1]. In the United States, the 2008 National Health Interview Survey (NHIS) reported that an estimated 25.2 million adult Americans (over 8%) are blind or visually impaired [2]. This number is increasing rapidly as the baby boomer generation ages. Recent developments in computer vision, digital cameras, and portable computers make it feasible to assist these individuals by developing camera-based products that combine computer vision technology with other existing commercial products such OCR, GPS systems.

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.

We note that electronic technology developed over the last fifty years has been an enormous boon for visually-impaired people, allowing access to text through video magnification [3], reading machines [4], text-to-speech (TTS) and screen readers [5], increasing quality of life for millions of individuals with vision loss by allowing independent and private access to text. However, few of them use these for navigation and travel. For navigation and travel, nearly all blind people use a cane at least some of the time due to the effectiveness, convenience, and low-cost, even if they rely on another mobility aid. The user typically scans left and right along their forward directional path, gathering information about obstacles from tactile and sonic information. Additionally it gives information about drop-offs, stairs, ground textures and type of flooring. It also serves as an identifier so that sighted people may avoid collisions with the blind traveler. However, the long cane is not able to detect obstacles higher off the ground. We think that the cane is most likely to remain useful to blind users for the foreseeable future, together with other high-tech assistive devices.

Many efforts have been made in development of electronic assistive devices to help blind persons navigate which can be found in recent surveys [6-9, 59, 61]. In addition to develop effective, reliable, and robust technology, friendly human interface design is even more important for successful assistive devices. There are two central and persistent issues in the human interface design: 1) how easy a system can be operated by the user, and 2) how the system can best present nonvisual information to the user.

A computer vision-based assistive system can improve the mobility of blind and visually impaired people to reduce risks and avoid dangers, enhance independent living, and improve quality of life. Our research efforts are focused on developing a computer vision-based navigation aid, because we believe this approach holds the greatest long-term promise, given the continually rapid growth in capabilities of computer and robotic technology fields of computer vision and robotics [10, 11]. The need for robots to navigate in the environment, in particular, is fueling the development of computer vision techniques for object recognition and scene analysis, along with localization and mapping. As imaging techniques advance, such as RGB-D cameras of Microsoft Kinect [12] and ASUS Xtion Pro Live [13], it has become practical to capture RGB sequences as well as depth maps in real time. Depth maps are able to provide additional information of object shape and distance compared to traditional RGB cameras. It has therefore motivated recent research work to investigate computer vision based assistive technology using RGB-D cameras.

In this chapter, we focus on RGB-D sensor-based computer vision technologies in application to assist blind and visually impaired persons. We first briefly review the existing computer vision based assistive technology for the visually impaired. Then we provide a detailed description of the recent RGB-D sensor based assistive technology to help blind or visually impaired people. Next, we present the prototype system we developed to detect and recognize stairs and pedestrian crosswalks based on RGB-D images. We conclude the chapter by the discussion of the future directions.

2. Related Work of Computer Vision Based Assistive Technology for Visually Impaired

Many electronic mobility assistant systems are developed based on converting sonar information into an audible signal for the visually impaired persons to interpret [14-18]. 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.

The "vOICe" system [19] is a commercially available vision-based travel aid that displays imagery through sound using videos captured by a head-mounted camera to help them build a mental image about the environment. However, the vOICe system translates images into corresponding sounds through stereo headphones, which will seriously block and distract the blind users' hearing sense. In addition, a training and education process is must conducted to understand the meanings of different tones and pitches of sounds about the environment. For example, if a short beep indicates a bright speck of light, three specks will produces three beeps. A vertical line is a stack of specks, sounding all at the same time but all with different pitches since they are at different heights.

In real situation, an environment is generally contains different objects, the vOICe system will generate a complex and "noisy" sound map which will be too complex for blind users to build the mental image.

Very recently, the US Food and Drug Administration approved for sale a new device -- the Argus II retinal prosthesis [20], from Second Sight Medical Products – comprised of a small video camera mounted on the nose bridge of a pair of sunglasses, a transmitter mounted near one temple of the sunglasses, a worn or carried video processing unit and a 60-electrode array that is intended to replace the function of degenerated photoreceptor cells in the retinas of those with the disease retinitis pigmentosa. Although it does not fully restore vision, the Argus II can improve ability to perceive lights, images and movement, using the video-processing unit to transform images from the video camera into image data that is wirelessly transmitted to the retinal electrode array. However, the temporal dynamics of electrical retinal stimulation are likely very different from those of a normal retina, the image is extremely low resolution relative to normal vision, and most important, the user must aim the head rather than the eye, to move an object into the field of view.

VizWiz [21] is a free iPhone app to provide answers to questions asked by blind users about their surroundings through anonymous web workers and social network members. Based on the statistic data about 50,000 questions, most questions were answered in a minute or less. With VizWiz, a user takes a picture and records a question on their mobile phone, then sends their question to anonymous workers, object recognition software – IQ [22], Twitter, or an email contact. Once an answer is received from any of those services, it is sent back to the users' phone. The advantage of VizWiz is the fusion of automatic image processing software with human replies from other members in user's social network. However, there are several main limitations for blind navigation and wayfinding: 1) For blind users, it is very hard to aim their iPhone to the targeted objects; 2) For the answers the user received, there is not a way to validate whether the answer is accurate. 3) Some questions may not be answered, and 4) Questions may take some time to answer.

A product in development called BrainPort (from Wicab Inc.) and recently approved for sale in Europe [23], uses a camera mounted on a pair of sunglasses as its input device. After image processing, images are displayed on the tongue via a “lollipop”-like display as shown in Figure 1. The “image” has been described as “tasting” a bit like effervescent champagne bubble on the tongue. Studies have demonstrated that blind and blindfolded sighted subjects can localize and identify some objects [24, 25] and avoid obstacles while navigating [26] under favorable conditions of contrast and lighting. Drawbacks of this system are that it requires use of the mouth, which precludes concurrently engaging in other lingual activities such as speaking and eating, and its spatial resolution is still far worse (by orders of magnitude) than that of the visual system, posing limits to object recognition.

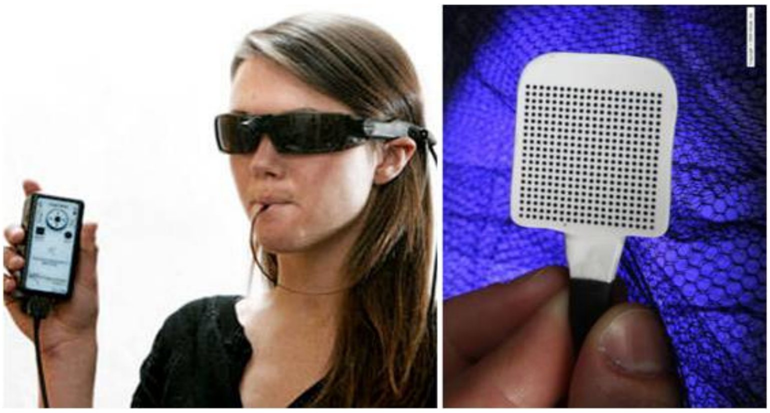

Figure 1: BrainPort vision substitution device [23].

Coughlan *et al.* [27] 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. As shown in Figure 2, Ivanchenko *et al.* [28] 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.

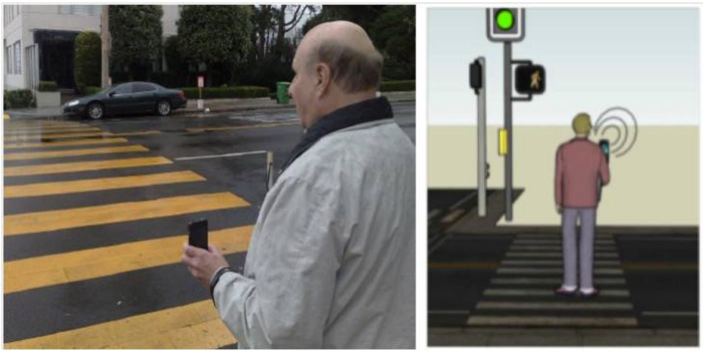

Figure 2: Crosswatch system for providing guidance to visually impaired pedestrians at traffic intersections by panning a cell phone camera left and right; the system provides feedback to help user align him/herself to crosswalk before entering it [28].

Advanyi *et al.* [29] 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.* [30] 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*. [31] 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. Uddin *et al.* [32] 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.

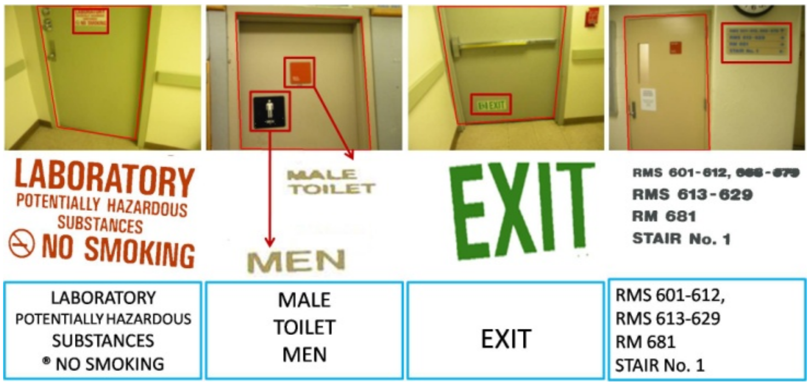

Figure 3: Top row: detected doors and regions containing text information. Middle row: extracted and binarized text regions. Bottom row: text recognition results of OCR in text regions [40].

Our own research group has developed a series of computer vision-based methods for blind people to recognize text and signage [33-36], recognize objects and clothes patterns [37-39], independently access and navigate unfamiliar environments [40-43], and under interface study [44]. Text and signage play important role in blind navigation and wayfinding. Tian *et al.* developed a proof-of-concept computer vision-based wayfinding aid for blind people to independently access unfamiliar indoor environments [40]. In order to find different rooms (e.g. an office, a lab, or a bathroom) and other building amenities (e.g. an exit or an elevator), the object detection is integrated with text recognition. A robust and efficient algorithm is developed to detect doors, elevators, and cabinets based on their general geometric shape which combines edges and corners. Then the text information associated with the detected objects is extracted and recognized. For text recognition, they first extracted text regions from signs with multiple colors and possibly complex backgrounds, and then applied character localization and topological analysis to filter out background interference. The extracted text is recognized using off-the-shelf optical character recognition (OCR) software products. The object type, orientation, location, and text information are presented to the blind traveler as speech. Some example results of door detection and text signage recognition are demonstrated in Figure 3. The first row of Figure 3 shows the detected door and signage regions. The second row displays the binarized signage. The last row displays that recognized text from OCR as readable codes on the extracted and binarized text regions.

Reading is obviously essential in today's society. Printed text is everywhere in the form of reports, receipts, bank statements, restaurant menus, classroom handouts, product packages, instructions on medicine bottles, etc. And while optical aids, video magnifiers and screen readers can help blind users and those with low vision to access documents, there are few devices that can provide good access to common hand-held objects such as product packages, and objects printed with text such as prescription medication bottles. The ability of people who are blind or have significant visual impairments to read printed labels and product packages will enhance independent living, and foster economic and social self-sufficiency.

Our group proposed a camera-based assistive framework to help blind persons to read text labels from hand-held objects in their daily life. As shown in Figure 4, a blind user wearing a camera captures the hand-held object from the cluttered background or other neutral objects in the camera view by slightly shaking the object for 1 or 2 seconds. This process solves the aiming problem for blind users. The hand-held object is detected from the background or other surrounding objects in the camera view by motion detection. Then a mosaic model is applied to unwarp the text label on the object surface and reconstruct the whole label for recognizing text information. This model can handle cylinder objects in any orientations and scales. The text information is then extracted from the unwarped and flatted labels. In the text localization method, the basic processing cells are rectangle image patches with fixed ratio, where features of text can be obtained from both stroke orientations and edge distributions [45-47]. The extracted text regions are then recognized by OCR software and communicate with the blind user in speech.

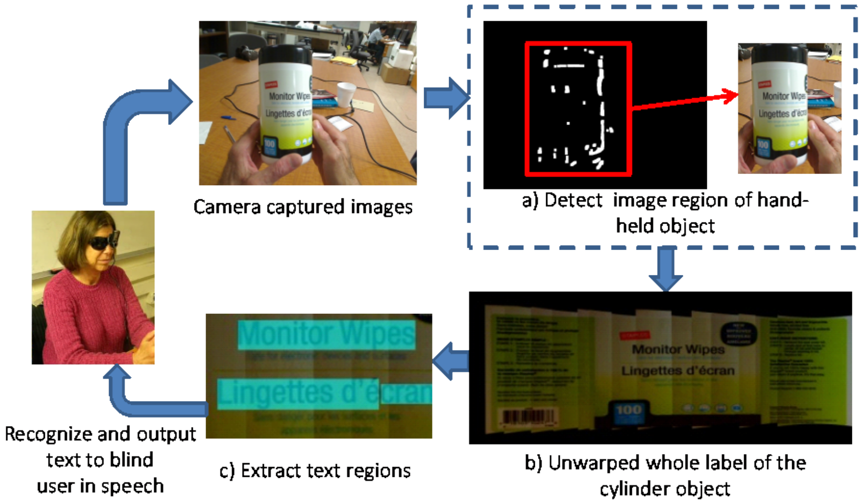
Camera captured images
a) Detect image region of hand-held object
Recognize and output text to blind user in speech
c) Extract text regions
b) Unwarped whole label of the cylinder object
Monitor Wipes
Lingettes d'écran

Figure 4: Flowchart of the framework to read labels from hand-held objects for blind users [36].

3. RGB-D Sensor-based Computer Vision Assistive Technology for Visually Impaired

As the release of RGB-D sensors and corresponding development toolkits, the applications of RGB-D sensor based computer vision technology has been extended far beyond gaming and entertainment. More reviews of the applications for multimedia and object recognition can be found in [62, 63]. In this section, we only focus on the research related to RGB-D camera-based assistive technology to help visually impaired people. Compared to the traditional RGB cameras or the stereo cameras, RGB-D sensors have the following advantages: a) RGB-D cameras contain both an RGB channel and a 3D depth channel which can provide more information of the scene; b) they work well in a low light environment; c) they are low-cost; and d) they are efficient for real-time processing. Currently, the RGB-D camera captures both RGB images and depth maps at a resolution of 640x480 pixels with 30 frames per second. The effective depth range of the Kinect RGB-D camera is from 0.8 to 4 meters. Although the Kinect for Windows Hardware can be switched to Near Mode which provides a range of 0.5 to 3 meters, currently the Near Mode is not supported for an Xbox Kinect for Windows SDK. The RGB-D cameras field of view is about 60 degrees. Since the RGB-D sensors use Infrared, they cannot be used reliably for obstacle avoidance of transparent objects such as glass doors. Also they will not work in outdoor environments with direct sunlight.

Theoretically, the traditional RGB camera based assistive technology for blind persons can be implemented using RGB-D sensors. However, since the limited resolution (640x480 pixels) of the current RGB-D sensors, some technologies may not work such as text detection and recognition especially for text with small size. In this section, we briefly summarize RGB-D sensor based technology for applications to assist visually impaired people.

Khan *et al.* developed a real time human and obstacle detection system for a blind or visually impaired user using an Xtion Pro Live RGB-D sensor [48]. As shown in Figure 5, the prototype system includes an Xtion Pro live (Kinect) sensor, waist assembly to mount the Kinect, a laptop for processing and transducing the data, a backpack to hold the laptop, and a set of headphone for providing feedback to the user. The system runs in two modes: 1) track and/or detect multiple humans and moving objects and transduce the information to the user; and

2) avoid obstacles for safe navigation for a blind or visually-impaired user in an indoor environment. They also presented a preliminary user study with some blind-folded users to measure the efficiency and robustness of their algorithms.

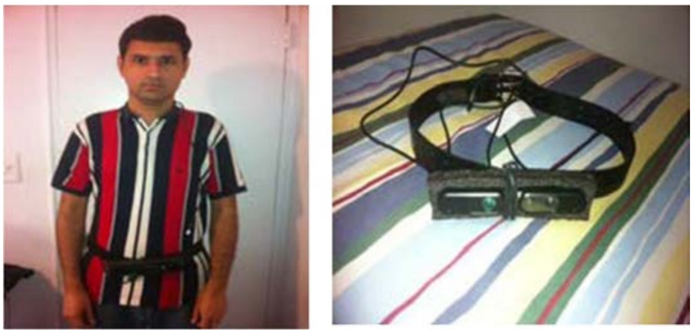

Figure 5. The Xtion Pro Live-Waist assemblies for detecting humans and obstacles [48].

Tang *et al.* presented a RGB-D sensor based computer vision device to improve the performance of visual prostheses [50]. First a patch-based method is employed to generate a dense depth map with region-based representations. The patch-based method generates both a surface based RGB and depth (RGB-D) segmentation instead of just 3D point clouds, therefore, it carries more meaningful information and it is easier to convey the information to the visually impaired. Then they applied a smart sampling method to transduce the important/highlighted information, and/or remove background information, before presenting to visually impaired people. They also reported some preliminary experiments with the BrainPort V100 [23] to investigate the effectiveness of both recognition and navigation that blind people can perform using such a low-resolution tactile device.

Lee and Medioni developed a wearable RGB-D camera based navigation aid for the visually impaired to navigate low textured environment as shown in Figure 6 [52]. To extract orientational information of the blind users, a visual odometer and feature based metric-topological SLAM (Simultaneous Localization and Mapping) are incorporated. A vest-type interface device with 4 tactile feedback effectors is used to communicate with the user for the presence of obstacles and provide the blind user with guidance along the generated safe path from the SLAM.

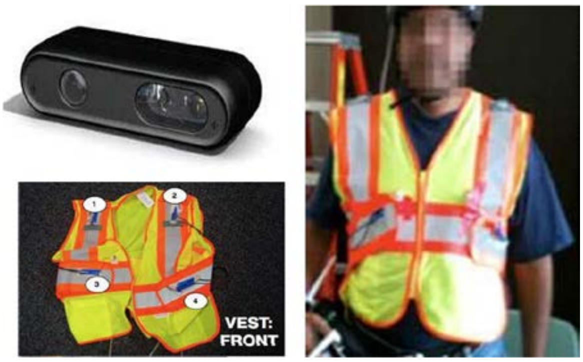

Figure 6. The RGB-D camera based navigation aid for the visually impaired which includes a RGB-D camera and a tactile vest interface device [52].

Park and Howard presented some preliminary results of development of a real-time haptic telepresence robotic system for the visually impaired to reach specific objects using a RGB-D sensor [58]. As shown in Figure 7, Tamjidi *et al.* developed a smart cane prototype by adding a SwissRanger SR4000 3D camera [60] for camera's pose estimation and object/obstacle detection in an indoor environment [59]. The SR4000 is an RGB-D sensor and provides intensity and range data of the scene. The SR4000 has a spatial resolution of 176×144 pixels and a field of view of 43.6º×34.6º.

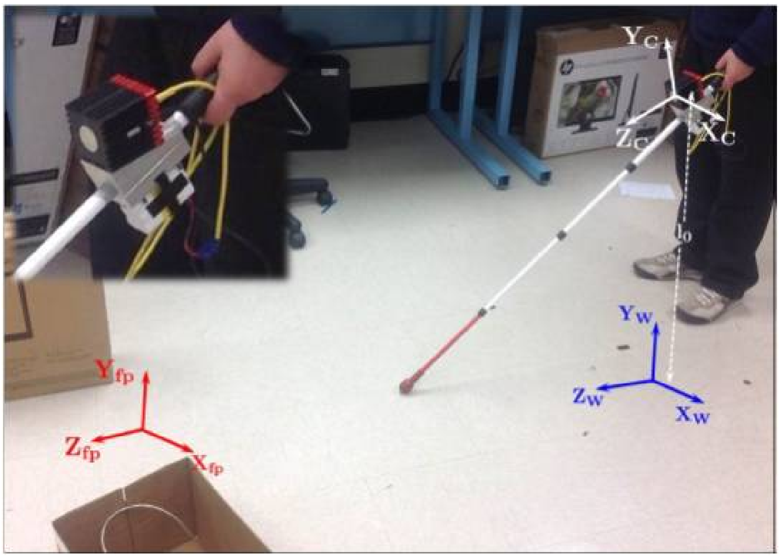

Figure 7. The Smart Cane prototype with a SwissRanger SR4000 3D camera [59].

In assistive system development, the user interface design plays a very important role. Without a friendly user interface, it is impossible for blind users to use the device even though the technology is perfect. How the system can best present spatial information nonvisually to the user is one of the center issues together these comprise the human interface for blind users. Ribeiro *et al.* developed a new approach for representing visual information with spatial audio to help a blind user building mental maps from the acoustic signals, and associating them with spatial data. [49]. As shown in Figure 8, the prototype device includes a Kinect RGB-D camera, an accelerometer, a gyroscope, and open-ear headphones. They applied computer-vision methods for plane decomposition, navigable floor mapping and object detection. Unlike previous work to create acoustic scenes by transducing low-level (e.g. pixel-based) visual information, their method only identifies high-level features of interest in an RGB-D stream. Then they rendered the location of an object by synthesizing a virtual sound source at its corresponding real-world coordinates. By sonifying high-level spatial features with 3D audio, users can use their inherent capacity for sound source localization to identify the position of virtual objects.

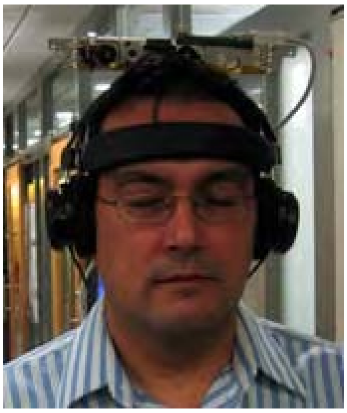

Figure 8. The prototype device of the auditory augmented reality proposed in [49].

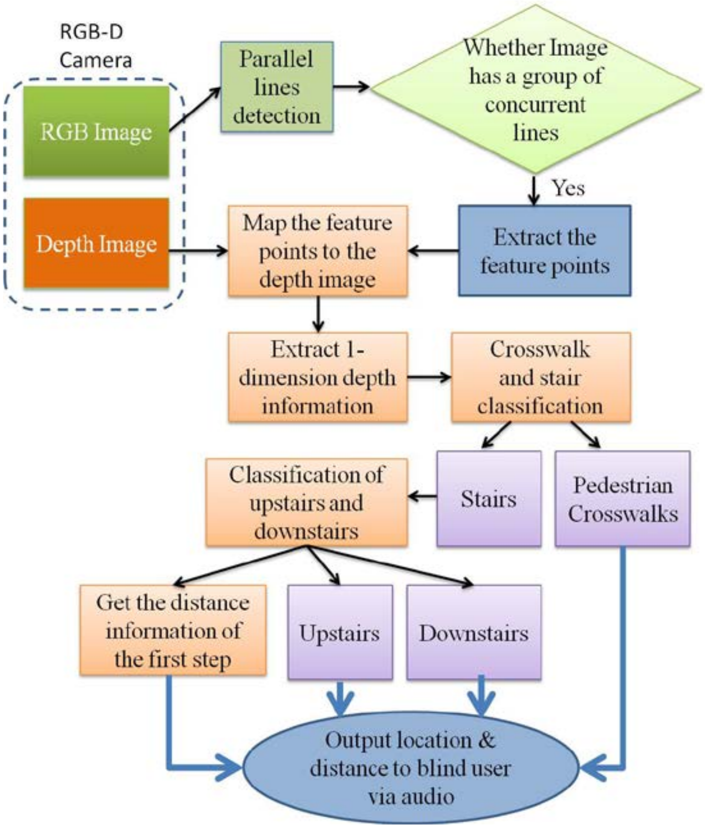
RGB-D Camera
RGB Image
Depth Image
Parallel lines detection
Whether Image has a group of concurrent lines
Yes
Extract the feature points
Map the feature points to the depth image
Extract 1-dimension depth information
Crosswalk and stair classification
Stairs
Pedestrian Crosswalks
Classification of upstairs and downstairs
Get the distance information of the first step
Upstairs
Downstairs
Output location & distance to blind user via audio

Figure 9: Flowchart of our proposed algorithm for stair and pedestrian crosswalk detection and recognition.

4. RGB-D Image-based Stair and Pedestrian Crosswalk Detection

4.1 System Overview

In this section, we describe a RGB-D based framework to detect stair-cases and pedestrian crosswalks for blind persons by integrating an RGB-D camera, a microphone, a portable computer, and a speaker connected by Bluetooth for audio description of objects identified. In our prototype system, a 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 mini laptop via a USB connection. The user can control the system by speech input via a microphone. Compared to existing work of staircase detection which only depends on RGB videos or stereo cameras [53-55], our proposed method is more robust and efficient to detect staircases and crosswalks.

As shown in Figure 9, our whole framework consists of 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 [56] 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.

4.2 Detecting Candidates of Pedestrian Crosswalks and Stairs from RGB images

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

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 first 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 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. We apply Hough transform to detect straight lines based on the edge points by the following steps:

Step1: Detect edge maps from the RGB image by edge detection.
Step2: Compute the Hough transform of the RGB image to obtain the direction of the line.
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.

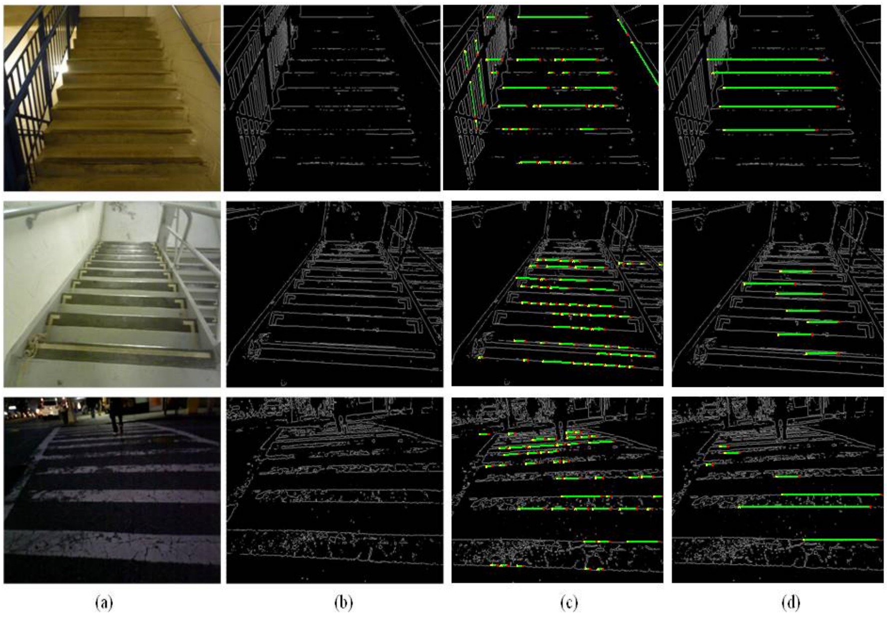

Figure 10: Example of upstairs (row 1), downstairs (row 2), and Pedestrian crosswalks (row 3). (a) Original image; (b) edge detection; (c) line detection; (d) concurrent parallel lines detection (yellow dots represent the starting points, red dots represent the ending points of the lines, and green lines represent the detected lines.)

As shown in Figure 10(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 is less than a threshold (set as 60 pixels in our system), then the line does not belong to the line group. And if the number of parallel lines less than 5 (more than two stair steps), the scene image does not contain stairs and pedestrian crosswalks.

4.3. Recognizing Pedestrian Crosswalks and Stairs from Depth Images

By detecting parallel lines under the constraints in a scene image captured by an RGB-D camera, we can detect the candidates of stairs and pedestrian crosswalks. 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 11. Considering the safeness for the visually impaired people, it is necessary to classify the different stairs and pedestrian crosswalks into the correct categories.

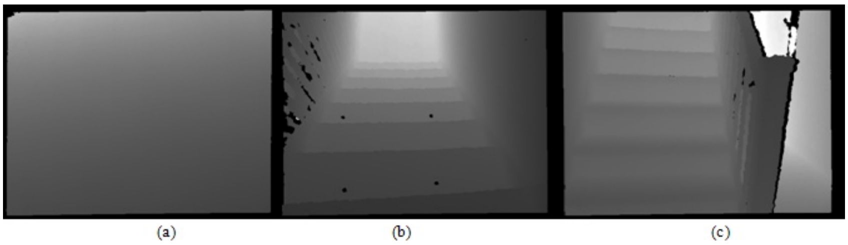

Figure 11. Depth images of (a) crosswalks, (b) downstairs, and (c) upstairs.

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 12(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 12(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 12(b).

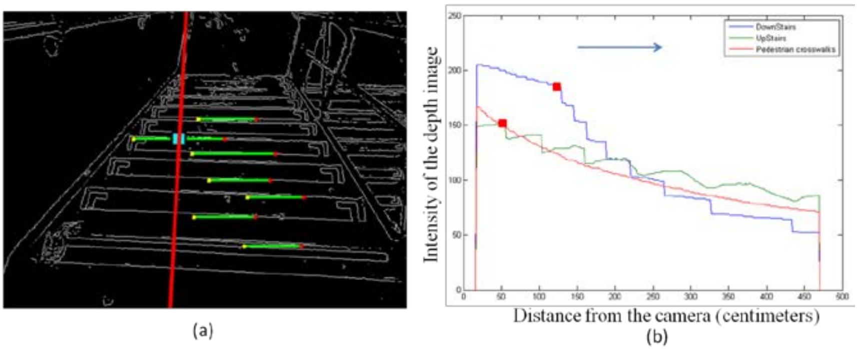

Figure 12. (a) The orientation and position of the feature line to extract one-dimensional depth features from the edge image. The blue square indicates the middle point of the longest line and the red line shows the orientation which is perpendicular to the detected parallel lines. (b) One-dimensional 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 captured by an RGB-D camera [12, 13] in Figure 11 is *640*480* pixels. The effective depth range of the RGB-D 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 12(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 RGB-D 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. In order to classify upstairs, downstairs, and pedestrian crosswalks, we employ a hierarchical SVM structure by using the extracted one-dimensional depth feature vector as the input. The classification processing includes two steps: 1) one classifier to identify pedestrian crosswalks from stairs. 2) For those detected stairs, one more classifier to further identify upstairs and downstairs.

4.4. 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 of the stairs to the blind or visually impaired individuals (i.e. the camera position) 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 12(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 12(b) along the one-dimensional depth feature, a point *x* satisfies the following two conditions is considered as a turning point:

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

where *f(x)* is the intensity value of the depth information, λ and ε are the thresholds which are determined by the RGB-D camera configuration. In our experiment, we observe that the best results can be obtained 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 RGB-D depth data and provided to the blind traveler by speech.

4.5. Experiment Results for Stair and Crosswalk Detection and Recognition

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 using an RGB-D camera [13]. The database is randomly divided into two subsets: 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) 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. Here, positive image samples indicate images containing either stairs or pedestrian crosswalks, and negative image samples indicate images containing neither stairs nor pedestrian crosswalks. Some of the negative images contain objects structured with a group of parallel lines such as bookshelves as shown in Figure 14. 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 13. The first row displays some RGB images of upstairs (Figure 13(a)), downstairs (Figure 13(b)), and crosswalks (Figure 13(c)) with different camera angles and the second row shows the corresponding depth images.

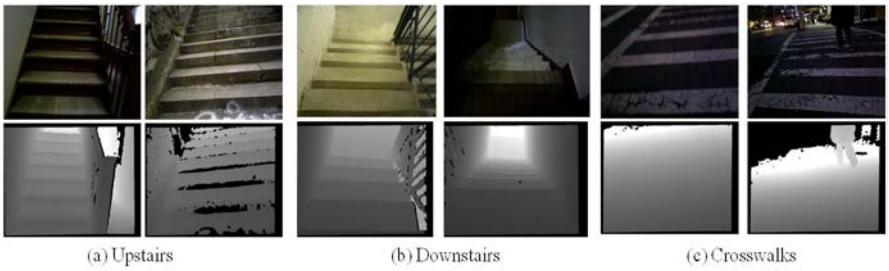

Figure 13**.** Examples of RGB (1st row) and depth images (2nd row) for (a) upstairs, (b) downstairs, and (c) pedestrian crosswalks in our database.

Table 1. Detection accuracy of candidates 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%**

Experiment Results: 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 of candidates of stairs and crosswalks, we correctly detect 103 stairs from 106 images, and 41 pedestrian crosswalks from 52 images of pedestrian crosswalks. Some of the negative samples are constructed similar edges as stairs and pedestrian crosswalks as shown in Figure 14. With the current RGB-D camera configuration, in general, only one to two shelves can be captured. The detected parallel lines do not meet the constraint conditions. Therefore, the bookshelves are not detected as candidates of stairs and pedestrian crosswalks.

In order to classify stairs and pedestrian crosswalks, the detected candidates of stairs and crosswalks 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.

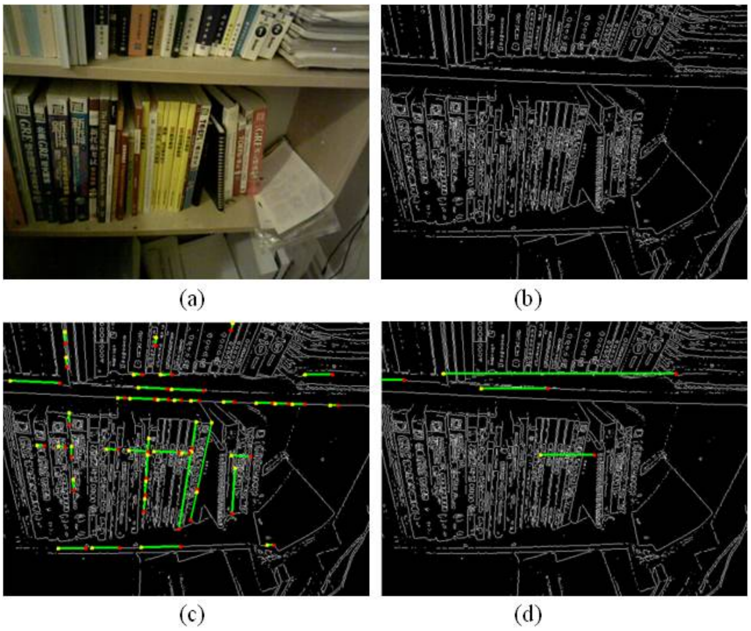

Figure 14. Negative examples of a bookshelf which has similar parallel edge lines to stairs and crosswalks.

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 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

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 RGB-D 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 15. For example, the depth information of some parts of the images is missing. Furthermore, as shown in Figure

15(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 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 15(d).

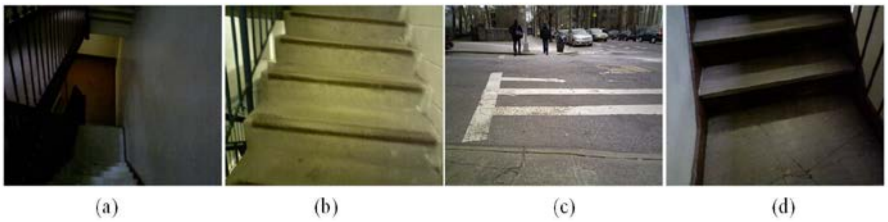

Figure 15. Examples of our proposed method of stair and crosswalk detection fails. (a) Downstairs with poor illumination; (b) Upstairs with less detected lines caused by noise; (c) Pedestrian crosswalks with missing zebra patterns; and (d) Stairs with less steps.

5. Conclusions and Future Work

In this chapter, we have reviewed several computer vision based assistive systems and approaches for blind or visually impaired people, especially using RGB-D sensors. There are mainly three main limitations for current RGB-D sensor based computer vision technology for the application of blind wayfinding and navigation: 1) It is difficult to achieve 100% accuracy to apply only computer vision based technology due to the complex environments and lighting changes. Based on the survey we conducted with blind users, we find 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. 2) For the current available RGB-D sensors, the size is still too large; the resolution is not high enough; the depth range is too short. In addition, they will not work in outdoor environments with direct sunlight. We believe that these limitations will be partially solved with design of next generation of RGB-D sensors. 3) It is hard to develop effective and efficient nonvisual display to blind users due to the huge amount of information images contain. Therefore, the user should be always included in the loop of the computer vision based blind assistive device.

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 are more desirable. Therefore, it is very important to design a user-friendly interface to provide meaningful feedback to blind users with the detected important information.

We think that 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. The future research should be focused on enhancing the robustness and accuracy of the computer vision technology as well as more user interface study for blind users.

Acknowledgments

This work was supported in part by NSF grant EFRI-1137172, IIP-1343402, FHWA DTFH61-12-H-00002, and Microsoft Research.

References:

1. "10 facts about blindness and visual impairment", World Health Organization: Blindness and visual impairment, 2009. www.who.int/features/factfiles/blindness/blindness_facts/en/index.html
2. Advance Data Reports from the National Health Interview Survey, 2008. http://www.cdc.gov/nchs/nhis/nhis_ad.htm.
3. S. M. Genensky, P. Baran, H. Moshin, and H. Steingold, "A closed circuit TV system for the visually handicapped," 1968.
4. R. Kurzweil, "The Kurzweil reading machine: A technical overview," Science, Technology and the Handicapped, pp. 3–11, 1976.
5. A. Arditi and A. Gillman, "Computing for the blind user.," Byte, vol. 11, no. 3, pp. 199–211, 1986.
6. D. Dakopoulos & N. G. Bourbakis, Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews, 40, 25–35. 2010.

7. N. Giudice & G. E. Legge, Blind navigation and the role of technology. In A. Helal, M. Mokhtari & B. Abdulrazak (Eds.) The Engineering Handbook of Smart Technology for Aging, Disability, and Independence. Hoboken, N.J.: Wiley. 2008.

8. R. Manduchi, & J. Coughlan, (Computer) vision without sight. Communications of the ACM, 55(1), 96–104, 2012.

9. U. R. Roentgen, G. J. Gelderblom, M. Soede & L. P. de Witte, Inventory of electronic mobility aids for persons with visual impairments: A literature review. Journal of Visual Impairment & Blindness, 102, 702-724, 2008.

10. F. Bonin-Font, A. Ortiz, & G. Oliver, Visual navigation for mobile robots: a survey. Journal of Intelligent & Robotic Systems, 53(3), 263–296, 2008.

11. G. N. DeSouza, & A. C. Kak, Vision for mobile robot navigation: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(2), 237–267, 2002.

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

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

14. 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.

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

16. 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.

17. 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.

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

19. Seeing with Sound – The vOICe: http//www.seeingwithdound.com/.

20. The Argus II retinal prosthesis system. Available: http://2-sight.eu/en/product-en.

21. "VizWiz - Take a Picture, Speak a Question, and Get an Answer." Available: http://vizwiz.org/.

22. "Image Recognition APIs for photo albums and mobile commerce | IQ Engines." Available: https://www.iqengines.com/.

23. "BrainPort lets you see with your tongue, might actually make it to market." Available: http://www.engadget.com/2009/08/14/brainport-lets-you-see-with-your-tongue-might-actually-make-it/.

24. D. R. Chebat, C. Rainville, R. Kupers, and M. Ptito, "Tactile–'visual' acuity of the tongue in early blind individuals," NeuroReport, vol. 18, no. 18, pp. 1901–1904, Dec. 2007.

25. M. D. Williams, C. T. Ray, J. Griffith, and W. De l 'Aune, "The Use of a Tac-

tile-Vision Sensory Substitution System as an Augmentative Tool for Individuals with Visual Impairments," Journal of Visual Impairment & Blindness, vol. 105, no. 1, pp. 45–50, Jan. 2011.

26. D. R. Chebat, F. C. Schneider, R. Kupers, and M. Ptito, "Navigation with a sensory substitution device in congenitally blind individuals," NeuroReport, vol. 22, no. 7, pp. 342–347, May 2011.

27. 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.

28. 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.

29. 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.

30. Se, S. Zebra-crossing Detection for the Partially Sighted, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2000; pp. 211-217.

31. Se S.; Brady, M. Vision-based Detection of Stair-cases, Proceedings of Fourth Asian Conference on Computer Vision (ACCV), 2000; pp. 535-540.

32. 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.

33. C. Yi and Y. Tian. Assistive Text Reading from Complex Background for Blind Persons, The 4th International Workshop on Camera-Based Document Analysis and Recognition (CBDAR), 2011.

34. 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.

35. C. Yi, Y. Tian, and A. Arditi, “Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons,” IEEE/ASME Transactions on Mechatronics, accepted, 2013. http://dx.doi.org/10.1109/TMECH.2013.2261083

36. Z. Ye, C. Yi and Y. Tian, Reading Labels of Cylinder Objects for Blind Persons, IEEE International Conference on Multimedia & Expo (ICME), 2013.

37. S. Yuan, Y. Tian, and A. Arditi, Clothing Matching for Visually Impaired Persons, Technology and Disability, Vol. 23, 2011.

38. F. Hasanuzzaman, X. Yang, and Y. Tian, Robust and Effective Component-based Banknote Recognition for the Blind, IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews, Jan. 2012.

39. X. Yang, S. Yuan, and Y. Tian, "Assistive Clothing Pattern Recognition for Visually Impaired People," IEEE Transactions on Human-Machine Systems, accepted, 2013.

40. Tian, Y.; Yang, X.; Yi, C.; Arditi, A. Toward a Computer Vision-based Wayfinding Aid for Blind Persons to Access Unfamiliar Indoor Environments, Ma-

chine Vision and Applications, 2012.

41. H. Pan, C. Yi and Y. Tian, A Primary Travelling Assistant System of Bus Detection and Recognition for Visually Impaired People, IEEE Workshop on Multimodal and Alternative Perception for Visually Impaired People (MAP4VIP), in conjunction with ICME 2013.

42. S. Joseph, X. Zhang, I. Dryanovski, J. Xiao, C. Yi, and Y. Tian, "Semantic Indoor Navigation with a Blind-user Oriented Augmented Reality," IEEE International Conference on Systems, Man, and Cybernetics, 2013.

43. S. Wang, H. Pan, C. Zhang, and Y. Tian, RGB-D Image-Based Detection of Stairs, Pedestrian Crosswalks and Traffic Signs, Journal of Visual Communication and Image Representation (JVCIR), Vol. 25, pp263-272, 2014. http://dx.doi.org/10.1016/j.jvcir.2013.11.005

44. A. Arditi and Y. Tian, “User Interface Preferences in the Design of a Camera-Based Navigation and Wayfinding Aid,” Journal of Visual Impairment & Blindness, Vol. 107, Number 2, pp118-129, March-April, 2013.

45. C. Yi and Y. Tian, Text String Detection from Natural Scenes by Structure-based Partition and Grouping, IEEE Transactions on Image Processing, Vol. 20, Issue 9, 2011.

46. C. Yi and Y. Tian, “Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification,”IEEE Transactions on Image Processing, Vol. 21, No. 9, pp4256-4268, 2012.

47. C. Yi and Y. Tian, Text Extraction from Scene Images by Character Appear-

ance and Structure Modeling, Computer Vision and Image Understanding, Vol. 117, No. 2, pp. 182-194, 2013.

48. A. Khan, F. Moideen, W. Khoo, Z. Zhu, and J. Lopez, KinDetect: Kinect Detecting Objects , 13th International Conference on Computers Helping People with Special Needs, 7383, Miesenberger, Klaus and Karshmer, Arthur and Penaz, Petr and Zagler, Wolfgang, Springer Berlin Heidelberg, Linz, Austria, July 11-13, 2012 , 588-595

49. F. Ribeiro, D. Florêncio, P. A. Chou, and Z. Zhang, Auditory Augmented Reality: Object Sonification for the Visually Impaired, IEEE 14th International Workshop on Multimedia Signal Processing (MMSP), 2012.

50. H. Tang, M. Vincent, T. Ro, and Z. Zhu, From RGB-D to Low resolution Tactile: Smark Sampling and Early Testing, IEEE Workshop on Multimodal and Alternative Perception for Visually Impaired People (MAP4VIP), in conjunction with ICME 2013.

51. Z. Wang, H. Liu, X. Wang, Y. Qian, Segment and Label Indoor Scene Based on RGB-D for the Visually Impaired, MultiMedia Modeling, Lecture Notes in Computer Science Volume 8325, pp 449-460, 2014.

52. Y. H. Lee and G. Medioni, A RGB-D camera Based Navigation for the Visually Impaired, RGB-D: Advanced Reasoning with Depth Camera Workshop, June 2011.

53. Y. H. Lee, T. Leung and G. Médioni, Real-time staircase detection from a wearable stereo system, ICPR, Japan, Nov 11-15, 2012

54. Lu X. and Manduchi, R., “Detection and Localization of Curbs and Stairways Using Stereo Vision” ICRA, 2005.

55. S. Wang and H. Wang, “2D staircase detection using real Adaboost”, Information, Communications and Signal Processing (ICICS), 2009.

56. Chang C.; Lin, C. LIBSVM: a Library for Support Vector Machine, 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm

57. R. Velazquez, Wearable Assistive Devices for the Blind. Book chapter in A. Lay-Ekuakille & S.C. Mukhopadhyay (Eds.), Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization, LNEE 75, Springer, pp 331-349, 2010.

58. C. H. Park and A. M. Howard, Real-time Haptic Rendering and Haptic Telepresence Robotic System for the Visually Impaired, World Haptics Conference (WHC), 2013.

59. A. Tamjidi, C. Ye, and S. Hong, 6-DOF Pose Estimation of a Portable Navigation Aid for the Visually Impaired, IEEE International Symposium on Robotic and Sensors Environments, 2013.

60. SR4000 User Manual (http://www. mesa-imaging).

61. Assistive Technology for Visually Impared and Blind People, Edited by Marion A. Hersh and Michael A. Johnson, Springer-Verlag London Limited, 2008. ISBN 978-1-84628-866-1.

62. Z. Zhang, Microsoft Kinect Sensor and Its Effect, IEEE MultiMedia, Volume: 19, Issue: 2, Page(s):4 - 10, 2012.

63. Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton, Enhanced Computer Vision with Microsoft Kinect Sensor: A Review , IEEE Transactions on Cybernetics, Oct. 2013.