Detecting and Recognizing Signage for Blind Persons to Access Unfamiliar Environments

Shuihua Wang, *Student Member*, *IEEE*, Chucai Yi, and YingLi Tian, *Senior Member*, *IEEE*

***Abstract*—Signage plays an important role in wayfinding and navigation to assist blind people accessing unfamiliar environments. In this paper, we present a novel camera-based approach to automatically detect and recognize restroom signage. Our method can handle symbolic signs, text, or both. For symbolic signage, we first employ shape detection to extract the attended areas which contain restroom signage. Then, local features are extracted from the attended areas to search for possible signage based on Scale-Invariant Feature Transform (SIFT) descriptor. Third, symbolic signage is detected and recognized as the regions with the SIFT matching scores larger than a threshold. For signage with text, a robust text extraction method is developed to detect the text regions, which are then binarized and recognized by off-the-shelf optical character recognition (OCR) software. The recognition results are then transformed into speech for blind users. Experiments show that the proposed method can handle multiple signage detection. Evaluation results on our collected restroom signage dataset demonstrate the effectiveness and efficiency of the proposed method.**

***Index Terms*— Blind people, Navigation and wayfinding, Signage detection and recognition, Text extraction and recognition.**

I. Introduction

There were about 161 million visually impaired people around the world in 2002, which occupied 2.6% of the entire population according to the study of World Health Organization (WHO). Among these statistics, 124 million were low vision and 37 million were blind [11]. Independent travel is well known to present significant challenges for individuals with severe vision impairment, thereby reducing quality of life and compromising safety. Based on our survey with blind users, detecting and recognizing signage has high priority for a wayfinding and navigation aid. In this paper, we focus on developing effective and efficient method for restroom signage detection and recognition from images captured by a wearable camera to assist blind people independently accessing unfamiliar environments.

Manuscript received March 13, 2012. This work was supported in part by NSF grant IIS-0957016, EFRI-1137172, and NIH 1R21EY020990.

Shuihua Wang is with the City College, City University of New York, New York, NY 10031 USA (e-mail: swang15@ccny.cuny.edu).

Chucai Yi is with the Graduate Center, City University of New York, New York, NY 10016 USA (e-mail: cyi@gc.cuny.edu).

YingLi Tian is with the City College, City University of New York, New York, NY 10031 USA (phone: 212-650-7046; fax: 212-650-8249; e-mail: ytian@ccny.cuny.edu). Prior to joining the City College in September 2008, she was with IBM T.J. Watson Research Center, Yorktown Heights, NY 10598 USA.

There are many camera-based technologies and systems developed to assist people who are blind or visually impaired. Ivanchenko and his colleagues developed a cell phone based prototype system to detect crosswalks at traffic intersections [8, 9]. The voice vision [7] technology for the totally blind offers sophisticated image-to-sound renderings by using a camera [12]. Manduchi *et al.* [14] also attempted to use a cell phone to help visually impaired persons for wayfinding. Many algorithms have been developed to detect and recognize text for assistive systems to help the visually impaired people understand the surrounding environments [18, 26-28, 31]. They mainly focus on text extraction and localization from camera-based images, and then combine with existing OCR techniques to detect and recognize text in the environment. Then recognized text will be converted to speech or audio for blind users. Everingham *et al.* [3] 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*. Shoval *et al.* [19] 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 will steer around it. The user can immediately perceive this steering action and follow the Guide-Cane's new path. Pradeep *et al*. [17] describes a stereo-vision based algorithm that estimates the underlying planar geometry of the 3D scene to generate hypotheses for the presence of steps. The authors group has developed a number of computer vision based technologies to help blind people including banknote recognition [4], clothing pattern matching and recognition [23, 29], text extraction [26-28], and navigation and wayfinding [20, 21, 24, 25]. In addition, some systems employed Quick Response (QR) codes or Radio-frequency identification (RFID) tags to guide blind persons to destinations. However, most of these systems require pre-installed devices and pre-marked QR codes and RFID tags [10, 22]. Although many efforts have been made, it is still an open research topic to properly apply computer vision technologies to help blind people understand their surroundings.

In this paper, we propose a computer vision-based method for restroom signage detection and recognition to assist blind persons independently accessing unfamiliar environments. The proposed method contains both detection and recognition procedures. Detection procedure obtains the location of a restroom signage in scene images. Recognition procedure is then performed to recognize the detected signage as three categories: "Men", "Women", and "Disabled". The signage detection is based on shape segmentation, which is widely employed and achieved great success in traffic signage and traffic light detection [16]. The signage recognition employs SIFT based matching, which is robust to variations of scale, translation and rotation, meanwhile partially invariant to illumination changes and 3D affine transformation. To read text information attached in restroom signage, a robust text extraction method is developed. Then the extracted text regions are binarized and recognized by off-the-shelf OCR software.

The prototype system of the proposed signage detection to assist blind persons for wayfinding and navigation consists of a camera, a computer, and an auditory output device. Visual information would be captured via a mini-camera mounted on a cap or sunglasses, while feature extraction and matching are processed by a wearable computer. The recognition results are presented to blind users by auditory signals (e.g., speech or sound), which are output via a Bluetooth earpiece.

The paper is organized as following: Section II describes the methodology of our proposed algorithm, including 1) method overview, 2) image preprocessing for signage detection, 3) signage detection based on shape and compactness, 4) signage recognition based on SIFT matching, and 5) text extraction, binarization, and recognition. Section III presents our experimental results which prove the effectiveness and efficiency of the proposed algorithm. Section IV will conclude the paper.

II. Methodology for Restroom Signage Detection and Recognition

A. Method Overview

The proposed restroom signage detection and recognition algorithm includes two main components: 1) symbolic signage detection, and 2) text extraction and recognition. The symbolic signage component contains three main steps: preprocessing of camera captured image, signage detection, and signage recognition as shown in Figure 1. Image preprocessing involves scale normalization, monochrome, binarization, and connected component labeling. Signage detection includes rule-based shape detection by detecting head and body parts of the signage respectively. The category of restroom signage (e.g., for "Men", "Women", or "Disabled") is recognized by SIFT based matching distance between the detected signage region and restroom signage templates. For the restroom signage with text, we first extract the text regions and then input the binarized text region into OCR software for recognition. The recognition results will be output to blind users as audio signal.

B. Image Preprocessing for Symbolic Signage Detection

As illustrated in Figure 2, the preprocessing of a camera captured image consists of three steps: 1) converting the input image to gray image; 2) transferring the gray image to a binary image; and 3) performing connected component analysis on the binary image to detect the possible signage and eliminate background noises.

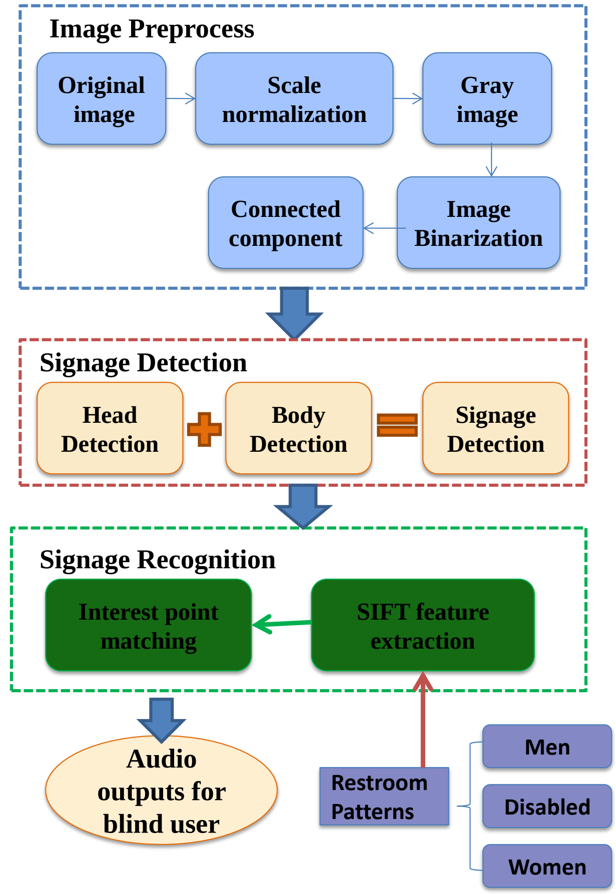


Figure 1. Flowchart of the proposed method for symbolic restroom signage detection and recognition.

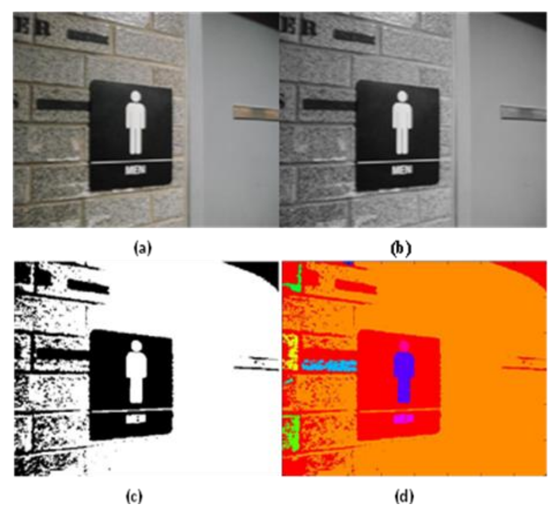


Figure 2. Image Preprocessing for symbolic restroom signage detection. (a) Original image, (b) Gray image; (c) Binary image; (d) Labeled connected components.

C. Symbolic Restroom Signage Detection based on Shape and Compactness

We observe that most signage in camera captured images stand upright. Furthermore, the shapes of the restroom signage for all the three types (i.e., Men, Women, and Disabled) in USA contain a circle-shaped "head" part and a more complicated "body" part (see Figure 3). In this section, we describe our new proposed rule-based method to locate restroom signage in images using shape information.

(a) (b) (c)

Figure 3. Templates of symbolic restroom signage for recognition. (a) Women, (b) Men, (c) Disabled

Detecting Head Part of a Restroom Signage: Figure 3 depicts that all restroom signage of “Men”, “Women”, and “Disabled” has circle-shaped head part. The most popular circle detection method is Hough transform which detects circles by voting procedures based on a-b-R space [1]. However, the computational complexity is high in a-b-R space. For example, if an image has 200x200 pixels, the size of a-b-R space is 200*200*100 = $4*10^6$, which is computing expensive [16]. Meanwhile, Hough transform accepts open circles, which do not represent the head part (closed circles) and may cause unpredicted recognition results. Thus we propose an efficient circle detection method based on the properties of connected components.

For each connected component which has a circle shape, the ratio of its perimeter and area is expected to be approximate to 4π. The connected component *CC* will be a “Head” part of a restroom signage if the following rule is satisfied:

$$If \quad \alpha_2 \le \frac{\text{CC.peri\^{}2}}{\text{CC.Area}} \le \alpha_1$$

where *CC.Area* is the area of the connected component and *CC.Peri* is the perimeter of the connected component.

(a) (b) (c)

Figure 4. Example results of symbolic restroom signage detection. The green components indicate the detected “head” part, and yellow components indicate detected the “body” part, and red boxes show the signage locations in images which will be used for recognition.

Detecting Body Part of a Restroom Signage: The body part of a restroom signage has more complicated shape which cannot be directly detected by simple shape detection method. Therefore, we detect a body part based on the positions of its corresponding head part of a restroom signage. A connected component is a body part if:

$$\beta_2 \le \frac{\text{CC.Area}}{\text{Head.Area}} \le \beta_1 \;\; \& \;\; \delta_2 \le \frac{\text{CC.Peri}}{\text{Head.Peri}} \le \delta_1$$

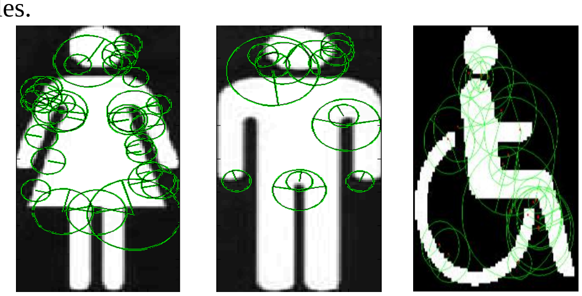
where *CC* is the nearest connected component to the head part, and *CC.Area* and *CC.Peri* represent its area and perimeter. All the parameters in the above equations are obtained from connected component statistics of the training samples in a database of restroom signs in good quality. Some example results of restroom signage detection are shown in Figure 4. The green components indicate the detected “head” part, the yellow components indicate the detected “body” part, and red boxes show the signage regions located in images which will be used for recognition. As shown in Figure 4(b), our method can also handle signage with rotations.

D. Symbolic Signage Recognition Based on SIFT Matching

SIFT Feature Extraction and Representation: Scale-invariant feature transform (SIFT) features have been widely employed for object detection and recognition due to the robustness to variations of scale, translation, rotation, illumination, and 3D affine transformation. In order to perform signage recognition, we employ SIFT feature descriptors as [5, 6]. SIFT feature extraction and representation contains two phases: (1) detect interest feature points and (2) feature point descriptor.

First, potential feature points are detected by searching overall scales and image locations through a difference-of-Gaussian (DoG) function pyramid. The DoG is a close approximation to the scale normalized Laplacian of Gaussian to find the most stable image features [12, 13]. Hence, the locations of the points correspond to these most stable features are identified as interest feature points.

Second, the feature descriptor is created for each interest point by sampling the magnitudes and orientations of image gradients in a 16x16 neighboring region. The region is centered at the location of the interest point, rotated on the basis of its dominant gradient orientation and scaled to an appropriate size, and evenly partitioned into 16 sub-regions of 4x4 pixels. For each sub-region, SIFT accumulates the gradients of all pixels to orientation histograms with eight bins [15]. A 4x4 array of histograms, each with eight orientation bins, captures the rough spatial structure of the neighboring region. This 128 dimensional vector, i.e. the feature descriptor for each interest point, is then normalized to a unit length. Figure 5 shows the SIFT features extracted from “Women”, “Men”, and “Disabled” templates. The detected interest points are the centers of the green circles, and the radius of the green circle indicates the scales.



(a) (b) (c)

Figure 5. Examples of SIFT feature extraction for different restroom signage (a) “Women”; (b) “Men”; and (c) “Disabled”. The center of each green circle indicates one detected interest point, and the radius of the green circle indicates the scales.

Signage Recognition by SIFT Matching: In order to recognize the detected signage, SIFT-based interest points are first extracted from the template images of restroom signage patterns which have no shape distortions (see Figure 3). Then, the features of the image region of the detected signage are matched with those from the template signage patterns based on nearest Euclidean distance of SIFT feature vectors. We employ the object location, scale, and orientation to identify good matches. Following Lowe's method, we employ the Best-bin-first (BBF) search method to identify the nearest neighbors with high probability by limited computation. The key point with minimum Euclidean distance from a descriptor vector is defined as its nearest neighbor. The BBF algorithm makes the bins in feature space searched in the order of their closest distance from the query location. In our method, two criteria are required to determine matching points: (1) similar descriptors for corresponding features; and (2) uniqueness for the correspondence.

For a detected signage, we calculate its number of matching points with each category of the template signage. The detected signage will be predicted as the category where maximum matching points are generated. Figure 6 demonstrates the extracted SIFT features (left column) and the matched features (right column) between the template signage patterns (the first row) and the detected signage regions (the second row) for signage of "Women" (Figure 6(a)), "Men" (Figure 6(b)), and "Disabled" (Figure 6(c)) respectively.

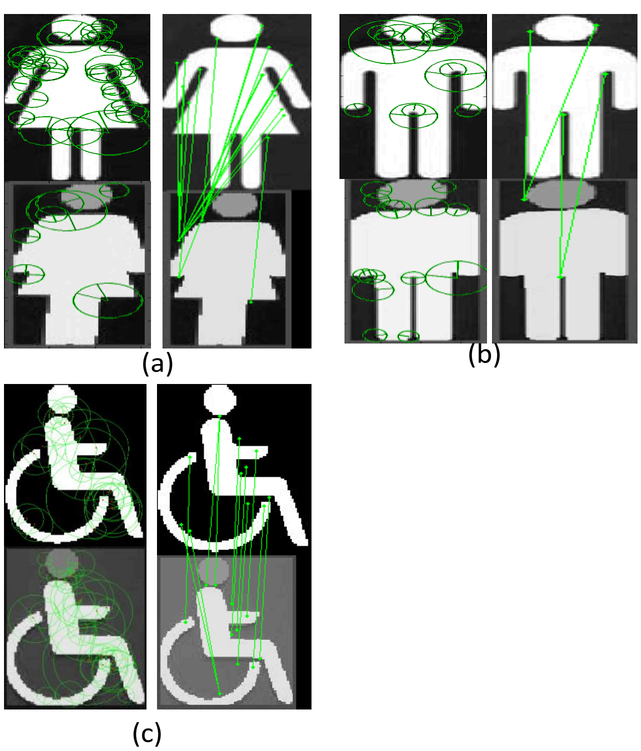


Figure 6. Examples of matched features between templates of restroom signage patterns (upper row) and the detected signage image regions (lower row).

E. Text Extraction, Binarization, and Recognition

In addition to symbolic signage appearances, text characters and strings in signage boards also play a significant role in wayfinding and navigation. We further design an algorithm to extract text information from camera-captured images.

In most scene images, text appears in the form of text strings that consist of three or more character members. These characters generally keep uniform colors and linear alignments. Thus we can employ color similarity and spatial layout analysis to extract text strings from scene images.

Firstly, color reduction is performed to group the pixels in the similar color into a color layer. Thus text strings can be separated from background objects such as signage boards and walls with different colors (see Figure 7(b)). Secondly, each color layer is actually a binary map, with color pixels as foreground. Then we perform connected component analysis to find out all possible text characters in the form of connected components. Their positions are marked by rectangle bounding boxes. Thirdly, spatial layout analysis is performed by using the adjacent character grouping methods in [28] to find out fragments of text strings from the connected components (see Figure 7(c)). Fourthly, all the text string fragments are merged into text regions.

In the detected text regions, we employ off-the-shelf OCR software to recognize the text characters and strings. Thus pixel-based text image is transformed into readable text code, and we notify blind user of text information by speech output.

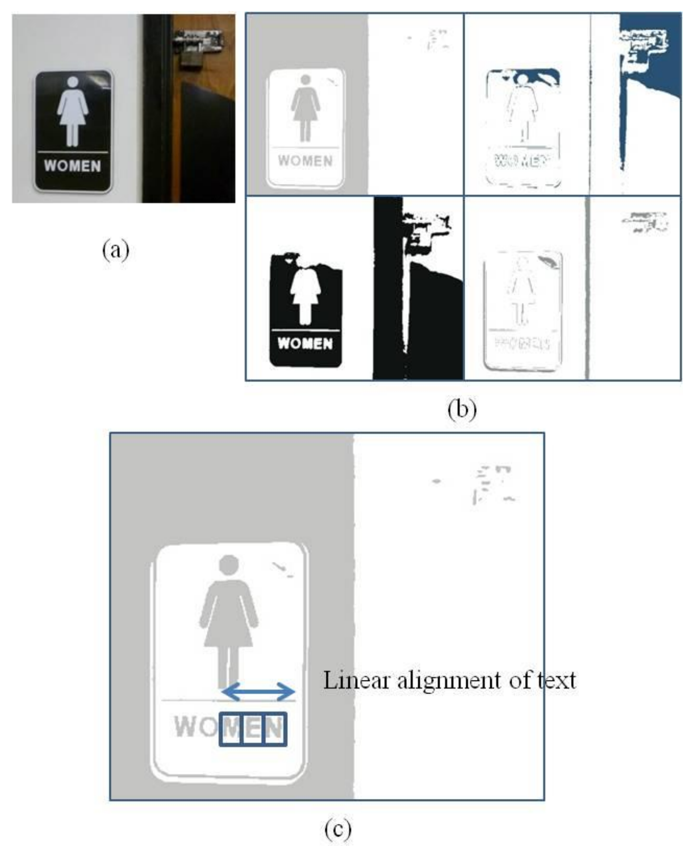


Figure 7. Text detection from scene images. (a) Original image. (b) Color layers obtained from color reduction. (c) Adjacent character grouping based linear text alignment in one of the color layers. Each group corresponds to a fragment of text string.

III. Experimental Results

A. Restroom Signage Database

To validate the effectiveness and efficiency of our method, we have collected a database containing 96 image samples of restroom signage including patterns of “Women”, “Men”, and “Disabled”. There are 50 “Men” signage, 42 “Women” signage, and 10 images of “Disabled” signage respectively. As shown in Figure 8, the database includes the changes of illuminations, scale, rotation, camera view, perspective projection, etc. Some of the images contain both signage of “Men” and “Women”, or combined signage of “Men” and “Disabled”, or combined signage of “Women” and “Disabled”. In addition, there are 30 images in the database which contain text information in the form of text strings, such as “MEN”, “WOMEN”, and “RESTROOM”. Our Text extraction method is evaluated on these images.

TABLE 1

Confusion matrix of restroom signage recognition

	Men(50)	Women(42)	Disable(10)
Men	41	3	0
Women	2	36	0
Disabled	0	0	9

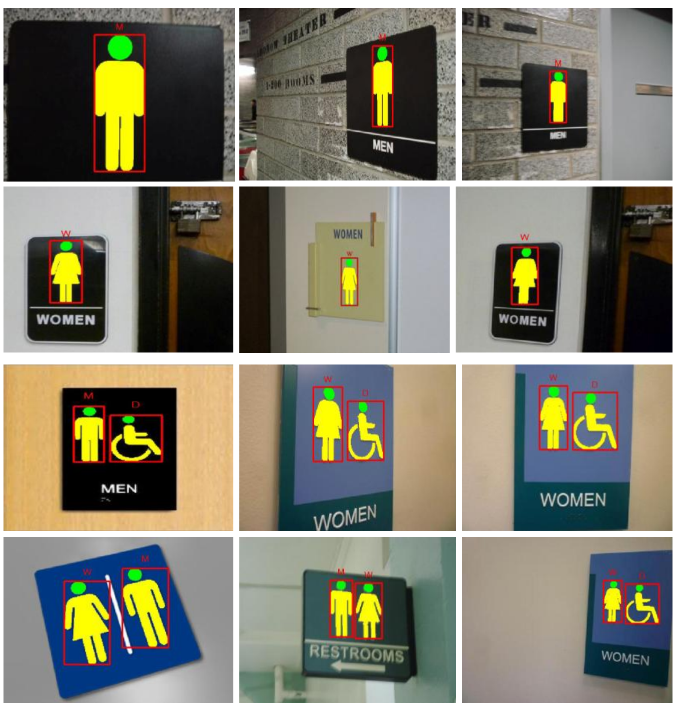


Figure 8. Sample images with signage detection and recognition in our database include changes of illuminations, scale, rotation, camera view, and perspective projection, etc. The red boxes show the detected signage region, while the letter above each red box indicates the recognition of the signage: “W” for “Women”, “M” for “Men”, and “D” for “Disabled”.

B. Results of Restroom Signage Recognition

Our method of signage detection and recognition can handle signage with variations of illuminations, scales, rotations, view angles, and perspective projections. The recognition accuracy of the proposed method is evaluated on the collected database, as shown in Table 1. Our proposed algorithm achieves accuracy of detection rate 89.2% and of recognition rate 84.3%. Of the 96 images in our dataset which contains 102 signage (some images have combined signage), 91 signs are correctly detected and 86 signs are correctly recognized. Some examples of the restroom signage detection and recognition from different environments are shown in Figures 8 and 9. The red boxes denote the detected signage region, while the letter above each red box indicates the recognized label of the signage: “W” for “Women”, “M” for “Men”, and “D” for “Disabled”.

C. Results of Text Detection and Recognition

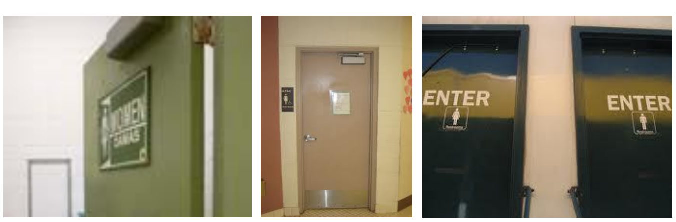
For the text information in signage, we apply the algorithm described in Section II-E to detect text strings. The proposed text localization method is evaluated by using the Robust Reading Dataset [32] from International Conference on Document Analysis and Recognition (ICDAR) 2003. We selected 420 images from the database which contain text strings with at least three characters and uniform colors. Our method achieves an average precision of 71%, where the precision is defined as the ratio of area of the successfully extracted text regions to the areas of the whole detected regions. Figure 10 demonstrates some detected regions of text strings marked in blue masks (the left column). OCR software recognizes the text information within the regions. The binarized text regions and their corresponding text codes recognized by OCR are displayed in the right column of Figure 10.

Figure 9. Sample restroom signage images of each step of the proposed method. Rows from top to bottom: original images, binarized images, images with connected components, and detected and recognized signage.

We further verify the computation time of the proposed method. The experiments are carried out on a computer with a 2GBHz processor and 1GB memory. The proposed algorithm is implemented in Matlab code without optimization. The average computation speed for detecting and recognizing restroom symbolic signage is about 150 frames per second at image resolution of 320x240. This ensures real-time processing for developing navigation and wayfinding systems to help blind and vision impaired users.

Figure 11 demonstrates several signage examples which our method fails to detect and recognize. We observe that the failures are caused by the following three reasons: 1) large camera view changes which can cause large shape distortion; 2) image blurry due to camera motion; and 3) low image resolution when the user is far from the signage.

Figure 10. Detected text regions are marked in blue masks in the images (left column). The detected text regions and the corresponding text codes recognized by off-the-shelf OCR are displayed in the right column.



(a) (b) (c)

Figure 11. Examples of images our method fails due to the extreme large perspective projections (a) and the small size of signage (b and c).

D. System Prototype Implementation

As shown in Fig. 12, we have developed a prototype system including a Logitech web camera with auto focus attached on a pair of sunglasses to capture images, a HP mini laptop connected to the camera by USB for algorithm-based processing, and a wireless Bluetooth earpiece for presenting detection results as speech outputs to the blind travelers. To avoid serious blocking or distracting the hearing of blind people, a sense that they rely upon heavily, the signage detection and recognition will be conducted only when blind users request. For speech output, we employ Microsoft speech software development kit (SDK). For text recognition, we integrate *Tesseract* [33], an open-source OCR engine, into our implementation. In pilot interface studies, we have conducted a survey and collected a testing dataset by 10 blind subjects using the prototype system with a wearable camera on sunglasses.

Figure 12. The developed prototype system includes a Logitech web camera with auto focus on sunglasses and tested by blind subjects for signage detection and recognition.

IV. Conclusion and Future Work

To assist blind persons independently accessing unfamiliar environments, we have proposed a novel method to detect and recognize restroom signage based on shape and appearance features and text strings associate with the signage. The proposed method can handle restroom signage with variations of scales, camera views, perspective projections, and rotations. The experiment results demonstrate the effectiveness and efficiency of our method.

Our future work will focus on detecting and recognizing more types of signage and context information to improve indoor navigation and wayfinding for blind people. We will also address the significant human interface issues including auditory displays and spatial updating of object location, orientation, and distance.

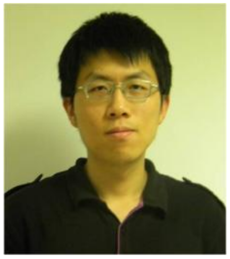
References

[1] M. Cao, C. Ye, O. Doessel and C. Liu. Spherical parameter detection based on hierarchical Hough transform, Pattern Recognition Letters, Vol.

27, pp. 980-986, 2006.
[2] N. Caporusso, L. Mkrtchyan, L. Badia, A Multimodal Interface Device for Online Board Games Designed for Sight-Impaired People, IEEE Trans on Information Technology in BioMedicine, Vol. 14, No.2, pp. 248-254, 2010.
[3] M. Everingham, B. Thomas and T. Troscianko, Wearable Mobility Aid for Low Vision Using Scene Classification in a Markov Random Field Model Framework, International Journal of Human Computer Interaction, vol. 15, pp.231-244, 2003.
[4] F. Hasanuzzaman, X. Yang and T. Tian, Robust and Effective Component-based Banknote Recognition for the Blind, IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews, 2011. 10.1109/TSMCC.2011.2178120
[5] C. Li and L. Ma, A new framework for feature descriptor based on SIFT, Pattern Recognition Letters, Vol. 30, pp. 544-557, 2009.
[6] D. Lowe, "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. pp. 1150–1157, 1999.
[7] Seeing with Sound – The voice, http://www.seeingwithsound.com/.
[8] V. Ivanchenko, J. Coughlan and H Shen. Crosswatch: a Camera Phone System for Orienting Visually Impaired Pedestrians at Traffic Intersections. 11th International Conference on Computers Helping People with Special Needs (ICCHP '08). 2008.
[9] V. Ivanchenko. J. Coughlan and H. Shen, Detecting and Locating Crosswalks using a Camera Phone." Fourth IEEE Workshop on Embedded Computer Vision, in conjunction with Computer Vision and Pattern Recognition, 2008.
[10] M. Kiers and T. Sovec, Ways4all: Indoor navigation for visually impaired and blind people, Proceedings of REAL CORP: CITIES FOR EVERYONE. Liveable, Healthy, Prosperous, 2010.
[11] I. Kocur, R. Parajasegaram and G. Pokharel, "Global Data on Visual Impairment in the Year 2002". Bulletin of the World Health Organization, Vol. 82, pp.844-851, 2004
[12] T. Lindeberg, Scale-space theory: "A basic tool for analyzing structures at different scales". J. Appl. Statist, vol. 21, pp. 224–270, 2004.
[13] K. Mikolajczyk and C. Schmid, An affine invariant interest point detector. In: Proc. European Conf. of Computer Vision (ECCV), pp. 128–142, 2002.
[14] R. Manduchi, J. Coughlan and V. Ivanchenko, Search Strategies of Visually Impaired Persons using a Camera Phone Wayfinding System. 11th International Conference on Computers Helping People with Special Needs (ICCHP '08), 2008
[15] Y. Matsui and Y. Miyoshi, Difference-of-Gaussian-Like Characteristics for Optoelectronic Visual Sensor, Signal Processing & Analysis, Vol. 7, pp. 1447 – 1452, 2007.
[16] M. Omachi and S. Omachi, Traffic light detection with color and edge information. Computer Science and Information Technology, 2nd IEEE International Conference in Beijing, pp.284-287, 2009.
[17] V. Pradeep, G. Medioni and J. Weiland, Piecewise Planar Modeling for Step Detection using Stereo Vision, Workshop on Computer Vision Applications for the Visually Impaired, 2008.
[18] H. Shen and J. Coughlan. Grouping Using Factor Graphs: an Approach for Finding Text with a Camera Phone. Workshop on Graph-based Representations in Pattern Recognition , 2007
[19] S. Shoval, I. Ulrich and J. Borenstein, Computerized Obstacle Avoidance Systems for the Blind and Visually Impaired, Invited chapter in "Intelligent Systems and Technologies in Rehabilitation Engineering." Editors: Teodorescu, H.N.L. and Jain, L.C., CRC Press, ISBN/ISSN: 0849301408, pp. 414-448, 2000.
[20] S. Wang and Y. Tian, Indoor signage detection based on saliency map and Bipartite Graph matching, International Workshop on Biomedical and Health Informatics, 2011.
[21] S. Wang and Y. Tian, Camera-based Signage Detection and Recognition for Blind Persons, 13th International Conference on Computers Helping People with Special Needs (ICCHP), 2012.
[22] S. Willis and A. Helal: "RFID Information Grid and Wearable Computing Solution to the Problem of Wayfinding for the Blin User in a Campus Environment"; Proceedings of the ninth annual IEEE International Symposium on Wearable Computers, 2005.
[23] X. Yang, S. Yuan and Y. Tian, Recognizing Clothes Patterns for Blind People by Confidence Margin based Feature Combination, International Conference on ACM Multimedia, 2011.
[24] X. Yang, Y. Tian, C. Yi and A. Arditi, Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environment", International Conference on ACM Multimedia, 2010.
[25] X. Yang and Y. Tian, "Robust Door Detection in Unfamiliar Environments by Combining Edge and Corner Features", 3rd Workshop on Computer Vision Applications for the Visually Impaired (CVAVI), 2010.
[26] 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.
[27] C. Yi and Y. Tian, Text Detection in Natural Scene Images by Stroke Gabor Words, The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011.
[28] 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. PMID: 21411405.
[29] S. Yuan, Y. Tian and A. Arditi, "Clothing Matching for Visually Impaired Persons", Technology and Disability 23, Page1-11, 2011.
[30] M. Zahid, A. Mittal, and R. Joshi, A pattern recognition-based approach for phylogenetic network construction with constrained recombination, Pattern Recognition, vol. 39, pp. 2312-232, 2006.
[31] A. Zandifar, R. Duraiswami, A. Chahine and L. Davis, A video based interface to textual information for the visually impaired. In: Proc. IEEE 4th international conference on multimodal interfaces, 2002.
[32] Robust Reading Dataset, http://algoval.essex.ac.uk/icdar/Datasets.html
[33] Tesseract OCR engine, http://code.google.com/p/tesseract-ocr/

Shuihua Wang (S'11) received her B.S. in Department of Information Science and Engineering from Southeast University, Nanjing, China in 2008. From 2011, she is a Master's student in Department of Electronic Engineering, the City University of New York, New York, NY, USA.

Her research focuses on object detection and recognition to assist blind persons travel independently. Her research interests include object detection, object recognition, and image processing.



Chucai Yi received his B.S. and the M.S. degrees in Department of Electronic and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007 and 2009, respectively. From 2009, he is a Ph.D. graduate student in Computer Science at the Graduate Center, the City University of New York, New York, NY, USA.

His research focuses on text detection and recognition in natural scene images. His research interests include object recognition, image processing, and machine learning.

YingLi Tian (M'99–SM'01) received her BS and MS from TianJin University, China in 1987 and 1990 and her PhD from the Chinese University of Hong Kong, Hong Kong, in 1996. After holding a faculty position at National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, she joined Carnegie Mellon University in 1998, where she was a postdoctoral fellow of the Robotics Institute. Then she worked as a research staff member in IBM T. J. Watson Research Center from 2001 to 2008. She is currently an associate professor in Department of Electrical Engineering at the City College of New York. Her current research focuses on a wide range of computer vision problems from motion detection and analysis, assistive technology, to human identification, facial expression analysis, and video surveillance. She is a senior member of IEEE.