Camera-based Signage Detection and Recognition for Blind Persons

Shuihua Wang and Yingli Tian

Department of Electrical Engineering
The City College, City University of New York,
160 Convent Ave., New York, NY 10031
{swang15, ytian}@ccny.cuny.edu

Abstract. Signage plays an important role for 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 from surrounding environments. Our method first extracts the attended areas which may content signage based on shape detection. Then, Scale-Invariant Feature Transform (SIFT) is applied to extract local features in the detected attended areas. Finally, signage is detected and recognized as the regions with the SIFT matching scores larger than a threshold. The proposed method can handle multiple signage detection. Experimental results on our collected restroom signage dataset demonstrate the effectiveness and efficiency of our proposed method.

Keywords: Blind people, Navigation and wayfinding, Signage detection and recognition.

1 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 [5]. 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.

Many disability and assistive technologies have been developed to assist people who are blind or visually impaired. The voice vision technology for the totally blind offers sophisticated image-to-sound renderings by using a live camera [10]. The Smith-Kettle well Eye Research Institute developed a series of camera phone-based technological tools and methods for the understanding assessment, and rehabilitation of blindness and visual impairment [14], such as text detection [12], crosswatch [4] and wayfinding [8]. To help the visually impaired, Everingham *et al.* [1] 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.* [13] 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. Pradeep *et al*. [11] 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 Media Lab at the City College of New York has been developed a number of computer vision based technologies to help blind people including banknote recognition [7], clothing pattern matching and recognition [16], text extract [18, 19], and navigation and wayfinding [15, 17]. Although many efforts have been made, how to apply this vision technology to help blind people understand their surroundings is still an open question.

In this paper, we propose a computer vision-based method for restroom signage detection and recognition. The proposed method contains both detection and recognition procedures. Detection procedure gets the location of a signage in the image. Recognition procedure is then performed to recognize the detected signage as ‘Men”, “Women”, or “Disabled”. The signage detection is based on effective shape segmentation, which is widely employed and achieved great success in traffic signage and traffic light detection [10]. The signage recognition employs SIFT feature-based matching, which is robust to variations of scale, translation and rotation, meanwhile partially invariant to illumination changes and 3D affine transformation.

Our proposed method in this paper is one component of a computer-vision based wayfinding and navigation aid for blind persons which consists of a camera, a computer, and an auditory output device. Visual information will be captured via a mini-camera mounted on a cap or sunglasses, while image processing and speech output would be provided by a wearable computer (with speech output via a Bluetooth earpiece). The recognition results can be presented to blind users by auditory signals (e.g., speech or sound).

2 Methodology for Restroom Signage Detection and Recognition

2.1 Method Overview

The proposed restroom signage recognition algorithm includes three main steps: image preprocessing, signage detection, and signage recognition as shown in Fig 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. Finally, the characteristic of restroom signage (e.g., for “Men”, “Women”, or “Disabled”) is recognized by SIFT feature based matching distance between the detected signage region and restroom signage templates.

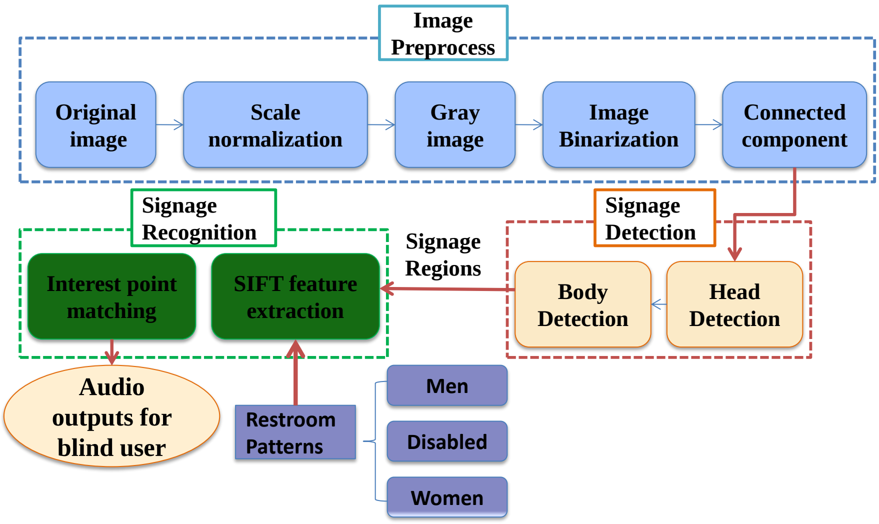

Fig 1. Flowchart of the proposed method of restroom signage detection.

2.2 Image Preprocessing for Signage Detection

To effectively detect signage from an image, an image preprocessing is first conducted which includes three main steps: 1) convert input image to gray image; 2) binarize gray image to a binary image; and 3) perform connected component processing on binary images to find the connected pixels and eliminate small noises.

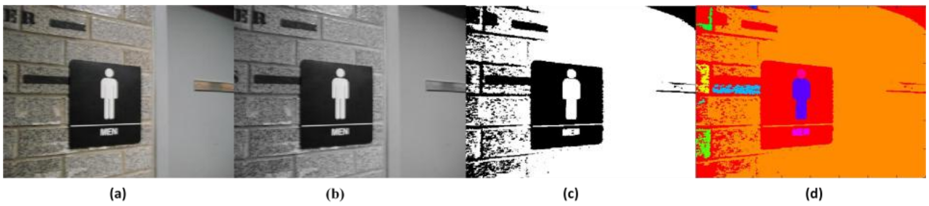

Fig 2. Image Preprocessing for restroom signage detection. (a) Original image, (b) Gray image; (c) Binary image; (d) Labeled connected components.

2.3 Signage Detection based on Shape and Compactness

We observe that most of images are upright and with relative stable illumination. Most important, the shape of the restroom signage in USA does not change much, which involves a circle-shaped "head" part and a more complicated "body" part as shown in Fig. 2. In this section, we describe an effective rule-based method to locate restroom signage in images using shape information.

2.3.1 Detecting Head Part of a Restroom Signage

As shown in Fig 2, the restroom signage of all "Men", "Women", and "Disabled" has a circle-shaped head part. The most popular circle detection method is Hough transform. However, Hough transform detects circles by voting procedures based on *a-b-R* space]. Suppose processing a 200-by-200 image, the size *a-b-R* space is $200*200*100=4*10^6$, which has high computation cost [10]. Meanwhile, Hough transform accept open circles, which do not represent the head part (closed circles), causing unpredicted recognition results. Thus we detect circles via 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. We set the rule as:

$$\text{If} \quad \alpha_2 \le \frac{\text{CC.peri\^{}2}}{\text{CC.Area}} \le \alpha_1 \text{, \textit{then the CC is} \textbf{\textit{Head}}.}$$

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

2.3.2 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 connected component if a body part based on the positions of the body part and head part of a restroom signage. A connected component is a body part if:

$$\beta_2 \le \frac{\text{CC.Area}}{\text{Head.Area}} \le \beta_1 \ \ \& \ \ \delta_2 \le \frac{\text{CC.Peri}}{\text{Head.Peri}} \le \delta_1 \ \ \& \ \textit{CC is the nearest component to the Head.}$$

where *CC.Area* and *CC.Peri* are the area and the perimeter of the connected component which is nearest the detected head part. All the parameters in the above equations are set by training of good quality sampled images from a restroom image database.

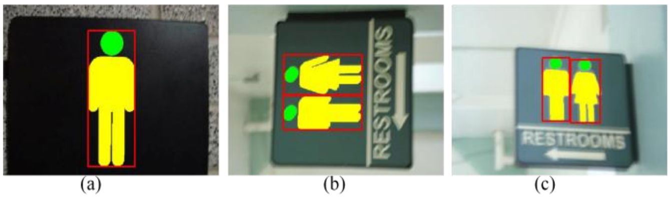

Fig 3. Example results of signage detection.

Fig 3 shows some example results of restroom signage detection. The green components indicate the detected "head" part, and yellow components indicate the "body" part, and red boxes show the signage locations in images which will be used for recognition.

2.4 Signage Recognition Based on SIFT Matching

2.4.1 SIFT Feature Extraction and Representation

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 features and descriptors. 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 [6, 7]. 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 neighbor 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 [9]. A 4x4 array of histograms, each with eight orientation bins, captures the rough spatial structure of the neighboring region. This 128-element vector, i.e. the feature descriptor for each interest point, is then normalized to unit length.

2.4.2 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 are stored in a database. Then, the features of the image region of the detected signage will be matched with those from the template signage patterns based on nearest Euclidean distance of their feature vectors. From the full set of matches, subsets of key points that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches. If two or more feature points in another image match a single point in the image, we assign the pair as the best match. In our method, two criteria are required for matching points (1) similar descriptors for corresponding feature; and (2) uniqueness for the correspondence.

Provided the number of matches between the signage template images and detected signage, the signage gets the maxima feature matches are selected as the most possible pattern.

3 Experimental Results

To validate the effectiveness and efficiency of our method, we have collected a database which contains 96 images of restroom signage including patterns of “Women”, “Men”, and “Disabled”. There are total 50 “Men” signage, 42 “Women” signage, and 10 images of “Disabled” signage. As shown in Fig. 4, 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 both signage of “Men” and “Disabled”, or “Women” and “Disabled”.

Our method can handle signage with variations of illuminations, scales, rotations, camera views, perspective projections. We evaluate the recognition accuracy of the proposed method. As shown in Table 1, the proposed algorithm achieves accuracy of detection rate 89.2% and of recognition rate 84.3% which correctly detected 91 and recognized 86 signage of total`102 signage in our dataset. Some examples of the detected restroom signage from different environments are shown in Figure 4. 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".

Table 1. Restroom signage recognition accuracy

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

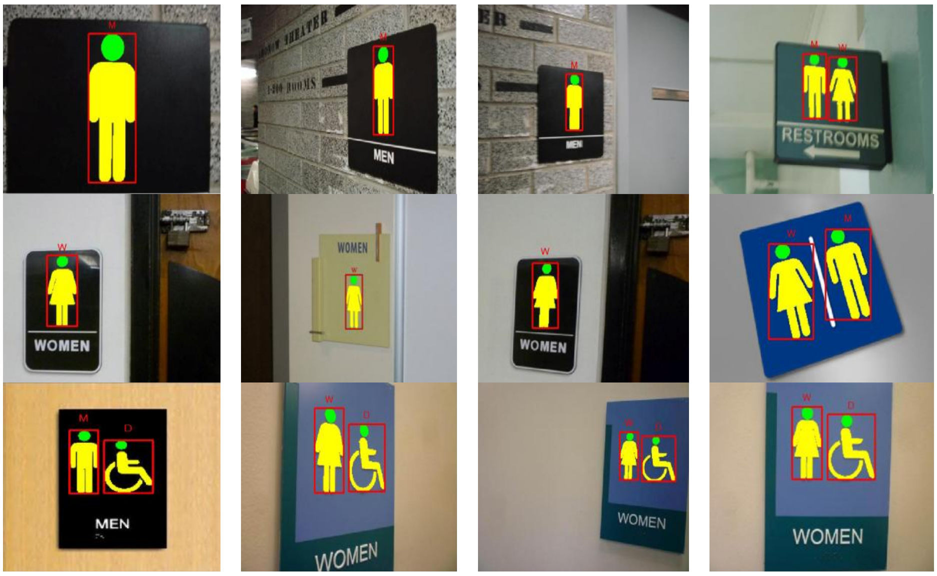

Fig 4. Sample images with signage detection and recognition in our database include changes of illuminations, scale, rotation, camera view, and perspective projection, etc. The red boxes show the detected signage region, while the letter above each red box indicates the recognition of the signage: "W" for "Women", "M" for "Men", and "D" for "Disabled".

Fig 5. Examples of Failures

Fig 5 demonstrates several signage examples which our method fails to detect and recognize. We observe that the failures are mainly 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.

We further verify the computation time of the proposed method. The experiments are carried on a computer with a 2GBHz processor and 1GB memory. The proposed algorithm is implemented in Matlab without optimization. The average time for detecting and recognizing signage from 30 testing images is 0.192s. This ensures real-time processing for developing navigation and wayfinding systems to help blind and vision impaired users.

5 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 both shape and appearance features. 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 incorporating 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.

Acknowledgement

This work was supported by NIH 1R21EY020990, NSF grants IIS-0957016 and EFRI-1137172.

References

1. 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, pp.231-244 (2003)
2. Hasanuzzaman, F., Yang, X., and Tian, T., Robust and Effective Component-based Banknote Recognition for the Blind, IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews, Vol. 41, Issue. 5, 2011. 10.1109/TSMCC.2011.2178120
3. Seeing with Sound – The voice, http://www.seeingwithsound.com/.
4. Ivanchenko, V., Coughlan J. and Shen, H., 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)
5. Kocur, I., Parajasegaram, R., Pokharel, G., Global Data on Visual Impairment in the Year 2002. Bulletin of the World Health Organization, vol. 82, pp.844-851

(2004)
6. Lindeberg, T., Scale-space theory: A basic tool for analyzing structures at different scales. J. Appl. Statist, vol. 21, 224–270 (2004)
7. Mikolajczyk, K., Schmid, C., An affine invariant interest point detector. In: Proc. European Conf. of Computer Vision (ECCV), pp. 128–142 (2002)
8. Manduchi, R., Coughlan, J., Ivanchenko, V., 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)
9. Matsui, Y., Miyoshi, Y., Difference-of-Gaussian-Like Characteristics for Optoelectronic Visual Sensor, Signal Processing & Analysis, vol. 7, pp. 1447 – 1452 (2007)
10. Omachi M, Omachi, S. Traffic light detection with color and edge information. Computer Science and Information Technology, 2nd IEEE International Conference in Beijing, pp.284-287 (2009)
11. Pradeep, V., Medioni, G., and J. Weiland, Piecewise Planar Modeling for Step Detection using Stereo Vision, Workshop on Computer Vision Applications for the Visually Impaired (2008)
12. Shen, H., Coughlan. J., Grouping Using Factor Graphs: an Approach for Finding Text with a Camera Phone. Workshop on Graph-based Representations in Pattern Recognition (2007)
13. 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.N.L. and Jain, L.C., CRC Press, ISBN/ISSN: 0849301408, Publication Date: 12/26/00, pp. 414-448.
14. The Smith-Kettlewell Rehabilitation Engineering Research Center (RERC) develops new technology and methods for understanding, assessment and rehabilitation of blindness and visual impairment. http://www.ski.org/Rehab/
15. Wang, S. H., Tian, Y. L. Indoor signage detection based on saliency map and Bipartite Graph matching, International Workshop on Biomedical and Health Informatics, (2011)
16. Yang, X., Yuan, S., and Tian, Y., Recognizing Clothes Patterns for Blind People by Confidence Margin based Feature Combination, International Conference on ACM Multimedia, (2011)
17. Yang, X., Tian, Y., Yi, C., and Arditi, A., Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environment", International Conference on ACM Multimedia, (2010)
18. Yi, C. and Tian, Y., Text Detection in Natural Scene Images by Stroke Gabor Words, The 11th International Conference on Document Analysis and Recognition (ICDAR), (2011)
19. Yi, C. and Tian, Y., Text String Detection from Natural Scenes by Structure-based Partition and Grouping, IEEE Transactions on Image Processing, Vol. 20, Issue 9, (2011) PMID: 21411405.