

Detecting Signage and Doors for Blind Navigation and Wayfinding

Shuihua Wang, Xiaodong Yang, and Yingli Tian

Department of Electrical Engineering,

The City College of New York, CUNY

New York, NY, 10031, USA

{swang15, xyang02, ytian}@ccny.cuny.edu

Abstract

Signage plays a very important role to find destinations in applications of navigation and wayfinding. In this paper, we propose a novel framework to detect doors and signage to help blind people accessing unfamiliar indoor environments. In order to eliminate the interference information and improve the accuracy of signage detection, we first extract the attended areas by using a saliency map. Then the signage is detected in the attended areas by using a bipartite graph matching. The proposed method can handle multiple signage detection. Furthermore, in order to provide more information for blind users to access the area associated with the detected signage, we develop a robust method to detect doors based on a geometric door frame model which is independent to door appearances. Experimental results on our collected datasets of indoor signage and doors demonstrate the effectiveness and efficiency of our proposed method.

1 Introduction

Based on the study of the World Health Organization (WHO), there were about 161 million visually impaired people around the world in 2002, about 2.6% of the total population. Among these statistics, 124 million had low vision and 37 million were blind [13]. Robust and efficient indoor object detection can help people with severe vision impairment to independently access unfamiliar indoor environments and avoid dangers [1]. Signage and other visual information provide important guidance in finding a route through a building by showing the direction and location of the desired end-point. Visual information is particularly important to distinguish similar objects such as elevators, bathrooms, exits, and normal doors. There are two categories of visual information: 1) directional and locational signage, and 2) objects (e.g. elevators, bathrooms, etc). A signage may include the text with or without an arrow. Figure 1 shows some examples of directional and locational signage of indoor environments.

Object detection is a computer vision technique that detects instances of a priori objects of a certain class (such as faces, signs, buildings, etc) in digital images and videos captured by cameras [20]. Camera-based indoor signage detection is a challenging problem due to the following factors: 1) large variations of appearance and design (shape, color, texture, etc.) of signage in different buildings; and 2) large variations in the camera view and image resolution of signage due to changes in position and distance between the blind user with wearable cameras and the targeted signage.



Figure 1: Typical indoor signage: (a) directional signage; (b) locational signage.

Object detection and recognition is a fundamental component for scene understanding. The human visual system is powerful, selective, robust, and fast [14]. It is not only very selective, which allows us to distinguish among very similar objects, such as the faces of identical twins, but also robust enough to classify same category objects with large variances (e.g. changes of position, scale, rotation, illumination, color, occlusion and many other properties). Research shows that the human visual system can discriminate among at least tens of thousands of different object categories [2]. Object recognition processes in the human visual system are also very fast: it can take as little as 100 to 200ms [8, 23, 34]. However, it is extremely difficult to build robust and selective computer vision algorithms for object recognition which can handle very similar objects or objects with large variations. For example, state-of-the-art object detection methods require hundreds or thousands of training examples and very long durations to learn visual models of one object category [28, 37, 38].

Network Modeling Analysis in Health Informatics and Bioinformatics

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 [32]. The Smith-Kettlewell 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 [9, 10, 17, 29, 33], such as text detection [29], crosswatch [9, 10], and wayfinding [17]. To help the visually impaired, Zandifar *et al.* [47] used one head-mounted camera together with existing OCR techniques to detect and recognize text in the environment and then convert the text to speech. Everingham *et al.* [6] 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.* [31] discussed the use of mobile robotics technology in the GuideCane device, a wheeled device pushed ahead of the user via an attached cane for the blind to avoid obstacles. When the GuideCane detects an obstacle it steers around it. The user immediately feels this steering action and can follow the GuideCane's new path. Pradeep *et al.* [25] describes a stereo-vision based algorithm that estimates the underlying planar geometry of the 3D scene to generate hypotheses for the presence of steps. Recently, we have developed a number of technologies for assist people who are visually impaired including navigation and wayfinding [35, 42], clothing pattern and color matching and recognition [43, 46], banknote recognition [7], text reading [44, 45], signage detection [40, 41], and etc. Although many efforts have been made, how to apply this vision technology to help blind people understand their surroundings is still an open question.

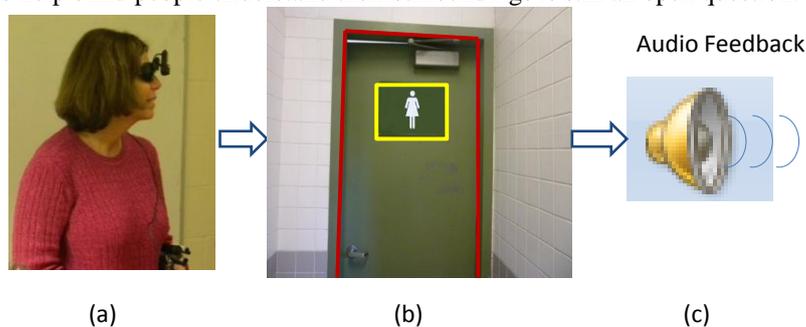


Figure 2: Query-based signage and door detection by a wearable camera mounted on a pair of sunglasses. (a) A blind user is looking for a bathroom; (b) “WOMEN” bathroom signage and door detection; (c) audio feedback is provided to the user.

Detection of signage and doors can help blind people to find their destinations in unfamiliar environments. As illustrated in Figure 2, in this paper, we propose a new framework based on camera captured visual information to provide guidance for blind people in unfamiliar indoor environments. The proposed framework including two main components: 1) signage detection by combining saliency map based attended area extraction and bipartite graph matching based signage recognition; 2) door detection based on a geometric door frame model by combining edges and corners. The hardware includes a camera, a microphone, a portable computer, and a speaker connected by Bluetooth for audio description of detection results. The user can control the system by speech input via microphone.

An earlier version of this paper can be found in [40]. Compared to our previous work, there are two major extensions that merit being highlighted:

- We recognize more types of signage. In this extended version, we recognize total of 9 categories of signage including three types of bathroom signs (Men, Women, and Disable), two types of elevator buttons (Open and Close), and 4 types of direction signs (Up, Down, Left, and Right).
- We add a new section of door detection to provide the entrance information for the blind users accessing the destinations associated with the detected signage.

The paper is organized as following: Section 2 describes the methodology of signage detection. Section 3 introduces the door detection method based on a general door geometric model. Section 4 analyzes experimental results and demonstrates the effectiveness and efficiency of the proposed algorithm. Section 5 concludes the paper and lists the future research directions.

2 Method of Indoor Signage Detection

Scholars tend to combine the window-sliding technique with the classifier to detect regions of an image at all locations and scales that contain the given objects. However, the window sliding

Network Modeling Analysis in Health Informatics and Bioinformatics

method suffers from two shortcomings [36]: 1) high processing time; and 2) inaccuracy of detection results due to different background. Therefore, we propose a new method to detect indoor signage which first employs the saliency map [15, 30] to extract attended areas and then applies bipartite graphic matching [16, 26, 48] to recognize indoor signage only at the attended areas instead of the whole image, which can increase the accuracy and reduce the computation cost. Our signage detection method consists of two phases as shown in Figure 3. In the first phase, attended areas are detected via saliency map based on color, intensity, and orientation. Then, the scaled patterns are detected within attended areas using bipartite graph matching. To localize the patterns, a window sliding method is employed to search the attended areas.

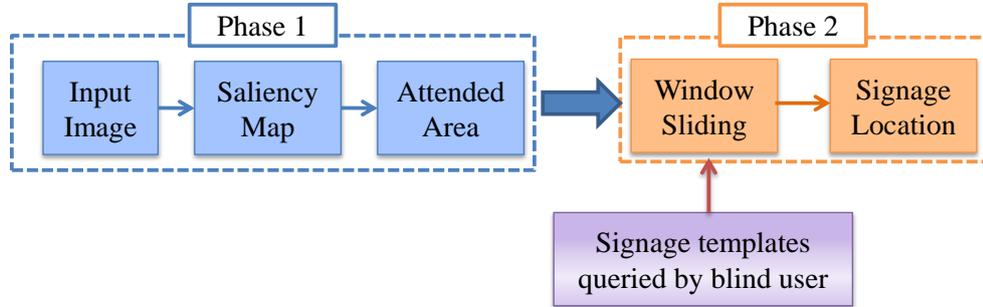


Figure 3. Flow chart of the proposed algorithm for indoor signage detection.

2.1 Building Image Saliency Maps

Saliency maps are used to represent the conspicuity at every location in the visual field by a scalar quantity and to guide the selection of attended locations based on the spatial distribution of saliency [4]. In analogy to the center-surround representations of elementary visual features, bottom-up saliency is thus determined by how different a stimulus is from its surround, in many sub-modalities and at many scales [24]. Saliency at a given location is determined primarily by how different this location is from its surroundings in color, orientation, motion, depth, etc.

As shown in Figure 4, the different visual features that contribute to attentive selection of a stimulus (color, intensity, and orientation) are combined into one saliency map. The saliency map which integrates the normalized information from the individual feature maps into one global measure of conspicuity. The detailed procedures are discussed in following sections.

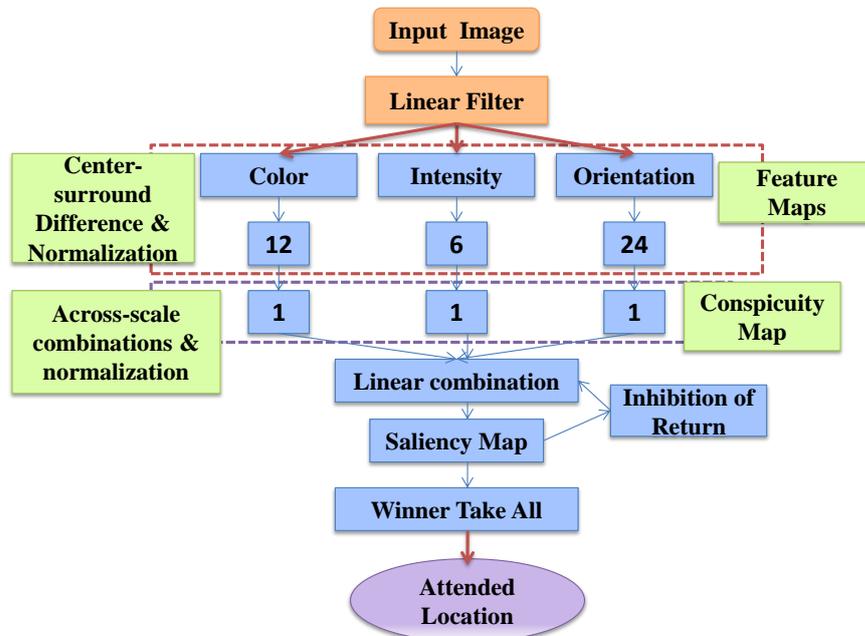


Figure 4. Architecture of building saliency maps based on three types of image features: intensity, color, and orientation.

Initialization:

We first resize the input color images at 640×480 resolution. Then, for each image, nine

Network Modeling Analysis in Health Informatics and Bioinformatics

spatial scales are created using dyadic Gaussian pyramids, which progressively low-pass filter and sub-sample the input image, yielding horizontal and vertical image-reduction factors ranging from 1:1 to 1:256 in eight octaves.

We compute saliency maps based on intensity, color, and orientation for each pixel by a linear “center-surrounded” operation similar to visual receptive fields. Center-surround feature extraction is implemented in the function of the difference between fine and coarser scales; the center is a pixel at the scales $c \in \{2,3,4\}$, and the surround is a pixel at scale $s=c+\delta$, with $\delta \in \{3,4\}$. The across-scale difference between two maps is gained by interpolation to the finer scale and point-by-point subtraction. Using several scales for both c and $\delta = s-c$ yield truly multi-scale feature extraction, by including different size ratios between the centers and surround regions.

Intensity-based Saliency Map:

With r , g , and b being the red, green, and blue channels of the input image, respectively, an intensity image I is achieved as $I = (r+g+b)/3$. Here I is used to create a Gaussian pyramid $I(\sigma)$, where $\sigma \in \{0,1,2,\dots,8\}$ denotes the scale. The r , g , and b channels are normalized by I in order to decouple hue from intensity. However, the hue variation are not perceivable at very low luminance (and hence are not salient), so normalization is only applied at the locations where I is larger than 10% of its maximum over the entire image and other locations yield zero. Four broadly-tuned color channels are created: $R = r-(g+b)/2$ for red, $G = g-(r+b)/2$ for green, $B = b-(r+g)/2$ for blue, and $Y = (r+g)/2-(r-g)/2-b$ for yellow, and negative values are set to zero for the R , G , B and Y values gained from above equations, 4 Gaussian pyramids $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, and $Y(\sigma)$ are created from these color channels.

Center-surround differences between a ‘center’ fine scale c and a ‘surround’ coarser scale s yield the feature maps. The first group of feature maps is concerned with intensity contrast, which is detected by neurons sensitive either to dark centers on bright surrounds or to bright centers on dark surrounds. In this paper, both types of sensitivities are simultaneously computed by using a rectification in six maps $I(c,s)$.

$$I(c,s) = |I(c) - I(s)| \quad (1)$$

Color-based Saliency Map:

The second group of saliency maps is similarly built for the color channels, which in cortex are represented by a so-called “color double-opponent” system: in the center of their receptive fields, neurons are excited by one color such as red and inhibited by another such as green, while the converse is true in the surround [27]. Such spatial and chromatic opponency exists for the red/green, green/red, blue/yellow, and yellow/blue color pairs in human primary visual cortex. Accordingly, maps $RG(c,s)$ are created in the function to simultaneously account for red/green and green/red double opponency Eq. (2) and $BY(c,s)$ for blue/yellow and yellow/blue double opponency Eq. (3)

$$RG(c,s) = |(R(c) - G(c)) - (G(s) - R(s))| \quad (2)$$

$$BY(c,s) = |(B(c) - Y(c)) - (Y(s) - B(s))| \quad (3)$$

Orientation-based Saliency Map:

Local orientation information is gained from I by oriented Gabor pyramids $O(\sigma, \theta)$ where $\sigma \in [0, 1, \dots, 8]$ represents the scale and $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ stands for the preferred orientation. Gabor filters are the product of a cosine grating and a 2D Gaussian envelope, approximating the receptive field sensitivity profile of orientation-selective neurons in primary visual cortex. Orientation feature maps $O(s, c, \theta)$ encode local orientation contrast between the centers and surround scales. In total, 42 feature maps are created: 6 from intensity, 12 from color, and 24 from orientation.

$$O(c,s,\theta) = |O(c,\theta) - O(s,\theta)| \quad (4)$$

Combination of Saliency Maps from Different Features:

The difficulty in combining different maps is that they represent a priori not comparable modalities, with different dynamic ranges and extraction mechanisms [12]. Furthermore, because all 42 maps are combined, salient objects which are strong in only a few maps may be masked by

noise or by less-salient objects present in a larger number of maps.

Because of the absence of top-down supervision, a map normalization operator $N(\cdot)$ is proposed, which globally promotes maps in which a small number of strong peaks of activity is present, meanwhile globally suppressing maps which contain numerous comparable peak responses. $N(\cdot)$ consists of:

- (1) Normalizing the values range in the map to a fixed range $[0, \dots, M]$, in order to eliminate modality dependent amplitude differences;
- (2) Finding the location of the map's global maximum M and computing the average \bar{m} of all its other local maxima;
- (3) Globally multiplying the map by $(M - \bar{m})^2$.

Only local maxima activities are considered, such that $N(\cdot)$ compares responses associated with meaningful 'active spots' in the map and ignores homogeneous areas. Comparing the maximum activity in the whole map to the average overall activity measures the difference between the most active location and the average. If the difference is large, the most active location stands out, and the map is strongly promoted. Otherwise, the map contains nothing unique and is suppressed.

Feature maps are combined into three "conspicuity maps", \bar{I} for intensity (5), \bar{C} for color (6), and \bar{O} for orientation (7), at the scale $\sigma = 4$ of the saliency map. They are gained by across-scale addition \oplus which consists of reduction of each map to scale four and point-by-point addition.

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c=4} N(I(c, s)) \quad (5)$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c=4} N(\text{RG}(c, s) + N(\text{BY}(c, s))) \quad (6)$$

For orientation, 4 intermediary maps are created by combining the six maps for a given θ and then are combined into a single orientation conspicuity map:

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c=4} N(I(c, s)) \quad (7)$$

The motivation of the creation of three separate channels (\bar{I} , \bar{C} , and \bar{O}) is based on the hypothesis that similar features compete strongly for saliency, meanwhile different modalities contribute separately to the saliency map. The three conspicuity maps are normalized and grouped into the final input S to the saliency map.

$$S = \frac{1}{3} (N(\bar{I}) + N(\bar{C}) + N(\bar{O})) \quad (8)$$

From above equations, we get the most salient image locations based on the maximum of the saliency map. We can simply detect the attended areas as the connected points by comparing the value of the saliency map and a threshold.

2.2 Bipartite Graph Matching Based Indoor Signage Detection

Detecting Signage in Attended Areas:

The sliding method window is a popular technique for identifying and localizing objects in an image [5]. The traditional approach involves scanning the whole image with a fixed-size rectangular window and applying a classifier to the sub-image defined by the window. The classifier extracts image features within the window and returns the probability that the window bounds a particular object. The process is repeated on different scales so that objects can be detected at any size [21]. Usually non-maximal neighborhood suppression is applied to the output to remove multiple detections of the same object. In order to improve the efficiency, our method only scans the attended areas which can significantly reduce the processing time of the sliding window algorithm [39].

Bipartite Graph Matching for Queried Signage Recognition:

To detect indoor signage queried by a blind user, we employ the "Bipartite Graph Matching (BGM)" algorithm to compare the query patterns and the slide-windows at the attended areas. The

Network Modeling Analysis in Health Informatics and Bioinformatics

detailed procedures of the BGM algorithm can be found at papers [11, 22]. Suppose A denotes the window and B denotes the sub-image covered by window A . The BGM algorithm first calculate the edges of the window and the sub-image, followed by evaluating the degree of the match as

$$BGM(A, B) = \frac{\omega}{N} \quad (9)$$

where ω denotes the sum of the pixels that exist in the edge images of both A and B , and N denotes the size of the window. Larger values of function (9) correspond to a better match; therefore, the BGM algorithm scans the image and finds the location with the largest matching score:

$$B^* = \arg \max (BGM(A, B)) \quad (10)$$

The maximization can be solved by a gradient-based optimization technique. However, the function (10) is non-convex and multi-modal, so the gradient-based optimization technique is easy to mislead in local extreme. In this paper, an exhaustive searching method is employed to find the global maxima.

2.3 Generalization to Multiple Signage Detection

Our method can handle both single signage and multiple signage detection. The pseudo-codes of multiple signage detection are listed below, which are based on multiple runs of aforementioned single pattern detection described above. Here, I denotes the original image; P_i denotes the i th signage pattern; S denotes the saliency map; BGM denotes the value of bipartite graph matching; L_i denotes the location found for the i th pattern.

```

Step 1 Initialization. Input  $I, P_1, P_2, \dots, P_N$ ;
Step 2  $S = \text{GetSaliencyMap}(I)$ ;
Step 3 for  $i = 1 : N$ 
         $[BGM_i, L_i] = \text{SinglePatternDetection}(I, S, P_i)$ ;
    end
Step 4  $i^* = \text{argmax} \{BGM_i\}$ ;
Step 5 Output  $i^*$  and  $L_{i^*}$ .

```

3 Method for Door Detection

In order to provide more information for blind users, we further develop a robust method to detect doors associated to the detected signage because doors are important landmarks for wayfinding and navigation. Doors provide transition points between separated spaces as well as entrance and exit information. Therefore, reliable and efficient door detection is a key component of an effective indoor wayfinding aid. There are several existing image-based door detection algorithms [3, 18, 19]. Chen and Birchfield developed a door detector based on the features of pairs of vertical lines, concavity, gap between the door and floor, color, texture, kick plate, and vanishing point [3]. Murillo *et al.* [19] proposed a door detection algorithm to handle all the doors with a similar color. Both above algorithms would fail while the colors of the doors varied. Munoz-Salinas *et al.* [18] developed a doorframe model-based algorithm which can handle doors with different colors. However, their algorithm cannot discriminate doors from other large rectangular objects, such as bookshelves, cabinets, etc.

As shown in Figure 5, our algorithm is based on the general geometric shape of door frame by combining edges and corners which can handle doors with a wide variety of color, texture, occlusions, illumination, scales, and viewpoints among different indoor environments. Furthermore, our algorithm can differentiate doors from other door-like objects such as bookshelves and cabinets.

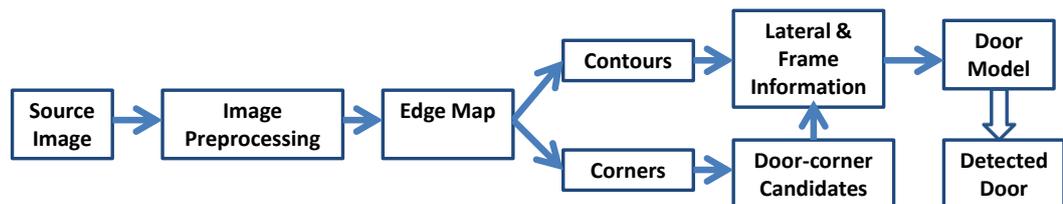


Figure 5. Flow chart of our proposed door detection method based on geometric door frame model.

3.1 Edge and Corner based Geometric Door Model

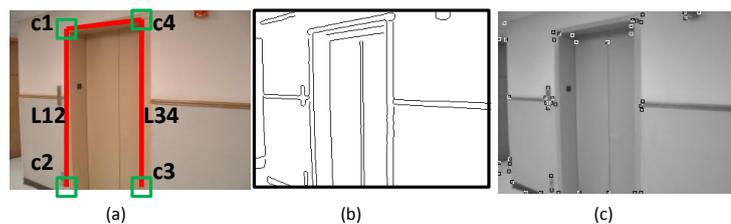


Figure 6. (a) Geometric door model; (b) detected edge map; (c) Extracted corners.

As shown in Figure 6(a), the geometric model of a door consists of four corners and four lines. However, in our application, it is often that only a part of a door is captured by a wearable camera, especially for visually impaired users who cannot aim the camera to “frame” the door. To detect a door from images, we made the following assumptions which are easy to achieve in practice: 1) At least two door corners are visible. 2) Both vertical lines of a door frame are visible. 3) Vertical lines of a door frame are nearly perpendicular to the horizontal axis of an image. 4) A door has at least a certain width and length.

For the algorithm implementation, we first extract edges and corners as shown in Figure 6(b) and (c). Edges and corners are insensitive to variations in color, viewpoints, scales, and illumination. Then the door-corner candidates are detected based on the relationship of four corners of the doorframe in the geometric door model. To determine whether a door-corner candidate represents a real doorframe, we further check if there are matching edges of the door frame between corners of each door-corner candidate. If the pixels of matching edges are larger than a threshold, then this door-corner candidate corresponds to a real door in the image. If there is more than one door-corner candidate with matching edges of same doorframe, they will be merged as one detected door.

3.2 Detecting Doors from Door-like Objects

Doors are generally recessed or inset into a wall while other door-like objects such as bookshelves and cabinets normally protrude in relief from a wall. To distinguish inset doors from protruding bookshelves and cabinets, we utilize information from faces formed by an inset or protrusion to obtain the depth information with respect to the wall.

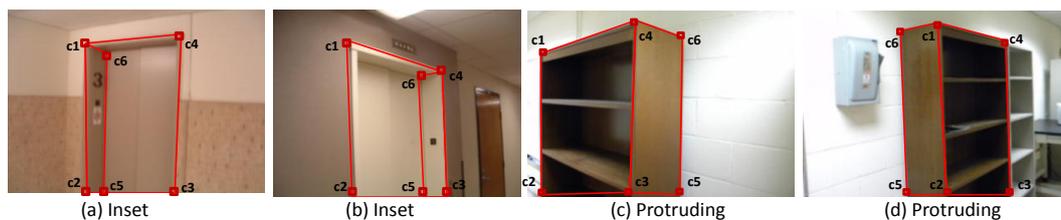


Figure 7. Inset and protruding models. (a) Inset object: door frame located on user’s right side and lateral face located on doorframe’s left side. (b) Inset object: frame located on user’s left side and lateral face located on frame’s right side. (c) Protruding object: frame located on user’s right side and lateral face located on frame’s right side. (d) Protruding object: frame located on the left side; lateral face located on the left side.

As shown in Figure 7, the different geometric characteristics and different relative positions of lateral faces and the detected door frame. In Figure 7(a) and (c), the door or door-like object is located on the right side with respect to the camera if $L_{12} < L_{34}$ of the detected frame. The lateral ($C_1C_2C_5C_6$) position of the doorframe ($C_1C_2C_3C_4$) indicates that the object is inset (left side) or protruding (right side). Similarly as shown in Figure 7(b) and (d), the bookshelf and the elevator are located on the left side with respect to the camera, and the lateral positions are different for the inset elevator (right side) and protruding bookshelf (right side). Therefore, combining the position of a frame and the position of a lateral, we can determine the inset or relief of a frame-like object. More details can be found in paper [35].

4 Experimental Results and Discussions

4.1 Single Signage Detection

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, detection of indoor signage such as elevator buttons and restrooms has high priority. Therefore, our experiments focus on the detection of signage.

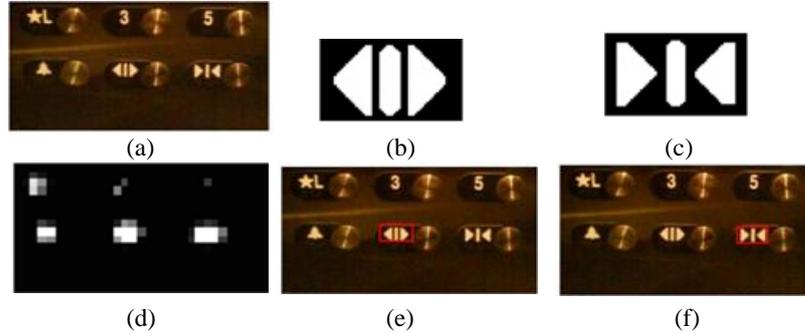


Figure 8. Elevator button detection: (a) Original image; (b) Query elevator button of “Open” symbol; (c) Query elevator button of “Close” symbol; (d) Saliency map of the original image; (e) Detected “Open” button; and (f) Detected “Close” button.

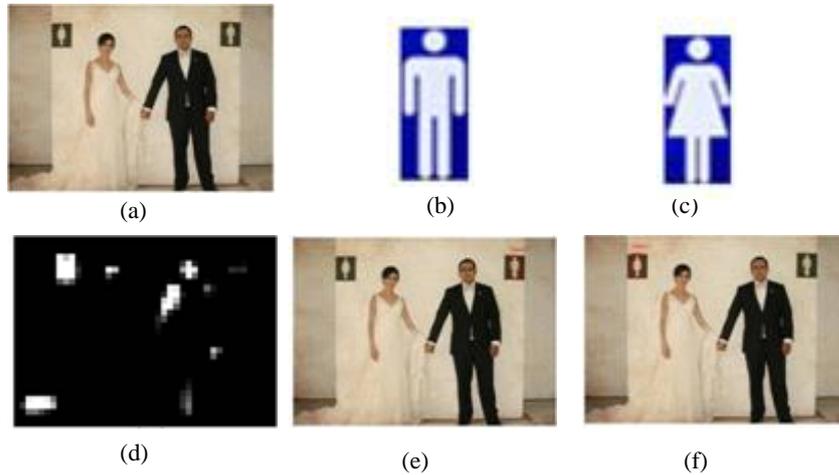


Figure 9. Restroom signage detection: (a) Original image; (b) Query pattern of “Men’s” signage; (c) Query pattern of “Women’s” signage; (d) Saliency map; (e) Detected “Men’s” signage; and (f) Detected “Women’s” signage.

Figure 8 displays detection of elevator buttons. The camera-captured image of six different elevator buttons is shown in Figure 8(a). Figure 8(d) is the corresponding saliency map extracted from the original image by using intensity, color, and orientation. The bright pixels indicate the attended areas. Figure 8(b) and Figure 8(c) show the query symbols of “open” and “close” buttons which are recorded in the query database of indoor signage. Figure 8(e) and Figure 8(f) demonstrate the final detection results (the red rectangular boxes).

Similarly, the detection of restroom signage is displayed in Figure 9 as (a) shows the original image including the signage of both “Women’s” and “Men’s” restrooms; (b) shows the saliency map of the original image where the bright regions indicate the attended areas; (c) and (e) are the query patterns; (d) and (f) are the final detection signage results of the “Men’s” and “Women’s” restroom.



Figure 10. Multiple signage detection results. First row: Multiple queried signage patterns; Second row: Original image, Saliency map, and the detected result of pattern 2 (signage of Men’s restroom); Third row: Original image, Saliency map, and the detected result of pattern 1 (signage of Women’s restroom). Fourth row: original image, Saliency map, and the detected result of pattern 3 (signage of Disabled restroom).

4.2 Multiple signage Detection

We further extend our algorithm to detect multiple patterns. In this case, the blind user will give multiple query patterns. As shown in Figure 10 (first row), the query patterns include both “Men’s” and “Women’s” restroom. We need to detect the patterns as well as recognize which pattern is found. The original image, corresponding saliency map, and the detected signage are shown in (the second row) for a “Men’s” and a “Women’s” restroom (the third row) respectively.

We further evaluate the multi-pattern method to detect the four emergency exit signage of up, down, left, and right directions. The detection results of directional signage are demonstrated in Figure 11.

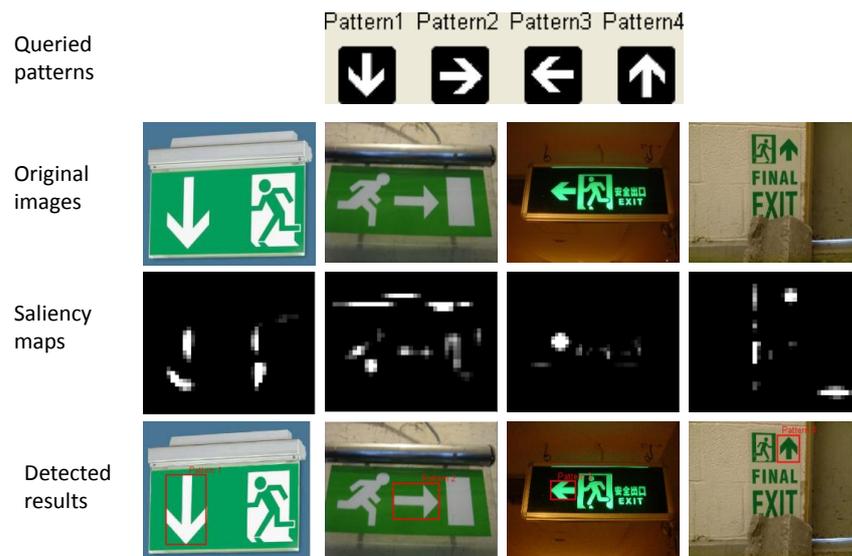


Figure 11. Examples of detected directional signage. 1st row: queried directional signage patterns (“Down”, “Right”, “Left”, and “Up”); 2nd row: original images; 3rd row: saliency maps; 4th row: detected results of the directional signage.

4.3 Experimental Results

Network Modeling Analysis in Health Informatics and Bioinformatics

Our proposed method is evaluated by a database of 182 indoor signage images with 9 different types of signage including restrooms (Men’s, Women’s and Disables’), elevator buttons (Open and Close), and exit directions (left, right, up, and down). Some examples are shown in Figure 12 which contains variations in lighting, resolution, and camera angle.

In single-pattern detection, we use 40 restroom sign images (20 Women’s, 20 Men’s and 20 Disables’) and 40 elevator button images (20 close and 20 open). We correctly detected 18 “Women’s” signs, 16 ‘Men’s’ signs, 18 ‘Disables’ signs, 17 “Close” signs, and 17 “Open” signs as shown in Table 1.

In multi-pattern detection, we use 82 direction signage (20 up, 20 down, 20 left, and 22 right) and 40 restroom signage (20 women, 20 men and 20 disable). In this experiment, 14 “Women’s” signs, 16 ‘Men’s’ signs, 18 “Disables” signs, 18 “Up” directions, 16 “Down” directions, 15 “left” directions and 19” right” directions are correctly detected from the original image. As shown in Table 2, the multi-patterns detection has a lower successful rate than those of single pattern detection.

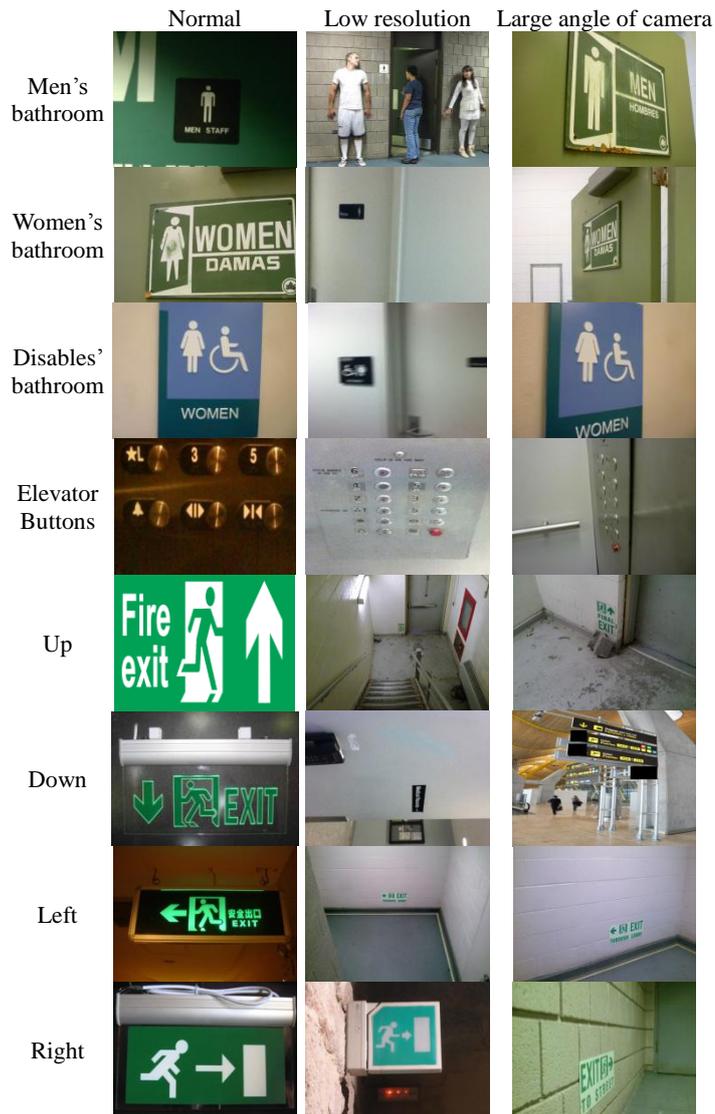


Figure 12. Example images of our indoor signage dataset.

Table 1. Accuracy of single pattern detection for bathroom signs

Signage Classes	No. of Samples	Correctly Detected	Detection Accuracy
Men’s	20	16	80%
Women’s	20	18	90%
Disables’	20	18	90%
Close	20	17	85%
Open	20	17	85%

Network Modeling Analysis in Health Informatics and Bioinformatics

Total	100	86	86%
--------------	------------	-----------	------------

Table 2. Accuracy of detect and recognize multiple patterns

Signage classes	No. of samples	Correct recognized	Wrong Recognized	Missed	Recognition accuracy
Women's	20	14	2	4	70%
Men's	20	16	1	3	80%
Disables'	20	18	2	0	90%
Up	20	18	0	2	90%
Down	20	16	2	2	80%
Right	22	19	1	2	86.3%
Left	20	15	2	3	75%
Total	142	116	10	16	81.6%

The wrong signage detections fall into the following two categories. 1) When we build the saliency map, the query pattern will be ignored because of low resolution of the original images. 2) The bipartite graph matching method is sensitive to the resolution, perspective projection, and angle of the camera views. It is difficult to distinguish different types of signage and the success rate of the detection decreases if the image resolution is too low or captured with an extreme camera view. As shown in Figure 13, the "close" signage is not detected as attended area in the saliency map due to the low image resolution. Figure 14 demonstrates that "Men" signage is correctly detected by the saliency map, but is missed by the bipartite graph matching method.

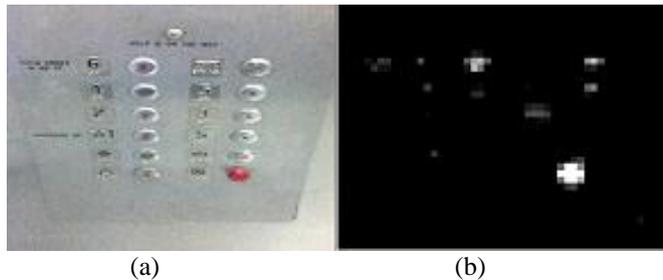


Figure 13. "Open" button of elevator: (a) Original image; (b) Saliency map

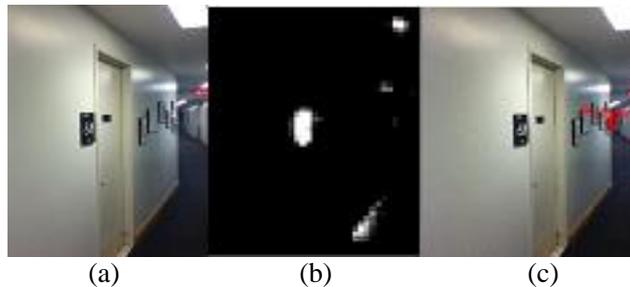


Figure 14. "Men" restroom: (a) Original image; (b) Saliency map; (c) Detection result

Table 3. Effectiveness of Saliency Map for Indoor Signage Detection

Classes	No. of Samples	Correct detected by BGM	Correct detected by SM and BGM
Men	20	11	14
Women	20	12	16
Total	40	23	30
Accuracy		57.5%	75%

4.4 Effectiveness of Employing Saliency Map

To evaluate the effectiveness of a saliency map for the indoor signage detection, we compare the detection results with and without using saliency map detection by using the "Women's" and "Men's" restroom signage. As shown in Table 3, only 11 "Men's" and 12 "Women's" restroom signs are correctly detected by applying bipartite graph matching on the image without performing saliency map based attended area detection. However, with saliency maps, we correctly detected 14 "Men's" and 16 "Women" restroom signs. The accuracy is increased from 57.5% (without

Network Modeling Analysis in Health Informatics and Bioinformatics

using saliency maps) to 75% (with saliency maps). Saliency map can effectively eliminate disturbing information which would decrease the accuracy of the bipartite graph matching method.

4.5 Results of Door Detection

We evaluated the proposed algorithm on a database we collected which contains 221 images collected from a wide variety of environments. The database includes 209 door images and 12 non-door images. Door images include doors and elevators with different colors and texture, and doors captured from different viewpoints, illumination conditions, and occlusions, as well as open and glass doors. Non-door images include door-like objects, such as bookshelves and cabinets. We categorized the database into three groups: *Simple* (57 images, see examples of the first row in Figure 15), *Medium* (113 images, see examples of the second row in Figure 15), and *Complex* (51 images, see examples of the third row in Figure 15), based on the complexity of backgrounds, intensity of deformation, and occlusion, as well as changes of illumination and scale. As shown in Table 4, the proposed algorithm achieves an average detection rate of 89.5% with a false positive rate of 2.3%. Some examples of door detection are illustrated in Figure 15.

Table 4. Door Detection Results

Data Category	True Positive Rate	False Positive Rate
<i>Simple</i>	98.2%	0%
<i>Medium</i>	90.5%	3.5%
<i>Complex</i>	80.0%	2.0%
Total	89.5%	2.3%

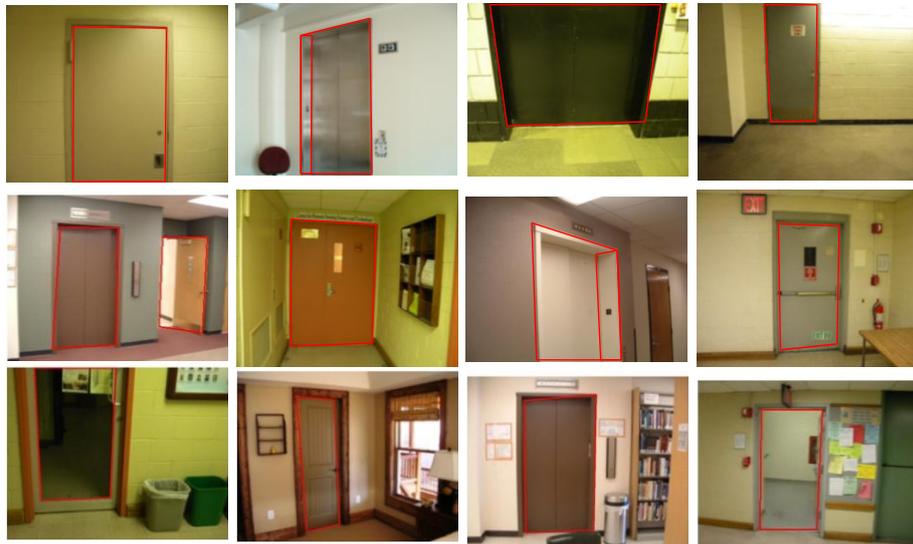


Figure 15. Examples of successfully detected doors in different environments. First row: Simple; Second row: Medium; Third row: Complex.

5 Conclusions

In this paper, we have proposed a framework for signage recognition and door detection to assist blind people accessing unfamiliar indoor environments. The signage recognition approach employs both saliency maps and bipartite graph matching to recognize the signage patterns queried by blind users. The door detection method is based on a general geometric door frame model which is robust to handle doors with different appearances. The effectiveness and efficiency of the proposed methods have been evaluated by the databases of indoor signage and doors.

Our future work will focus on extending our method to handle signage and objects with larger variations of perspective projection, scale, point view, occlusions, etc. We will also address the significant human interface issues associated with wayfinding for blind users.

Acknowledgement

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

References

- [1] A. Baker, Blind Man is Found Dead in Elevator Shaft, The New York Times, City Room, May 2010.
- [2] Biederman, I. 1987, "Recognition-by-Components: A Theory of Human Image Understanding," *Psychological Rev.*, vol. 94, pp. 115-147, 1987.
- [3] Z. Chen and S. Birchfield, Visual Detection of Lintel-Occluded Doors from a Single Image. *IEEE Computer Society Workshop on Visual Localization for Mobile Platforms*, 2008.
- [4] C. Davies, W. Tompkinson, N. Donnelly, L. Gordon, and K. Cave, "Visual saliency as an aid to updating digital maps," *Computers in Human Behavior*, July 2006. Vol. 22, No. 4, pp. 672-684.
- [5] F. Davignon, J. F. Deprez, and O. Basset, "A parametric imaging approach for the segmentation of ultrasound data," *Ultrasonics*, December 2005, Vol. 43, No. 10, pp. 789-801
- [6] 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*, Volume 15, Issue 2, 2003
- [7] F. Hasanuzzaman, X. Yang, and Y. Tian, Robust and Effective Component-based Banknote Recognition for the Blind, Accepted, *IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews*, 2011.
- [8] C. Hung, G. Kreiman, T. Poggio, and J. DiCarlo, Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. *Science*, 2005. 310: 863-866.
- [9] 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.
- [10] [Ivanchenko08] 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 (CVPR '08). 2008
- [11] N. Kakimura, "Matching structure of symmetric bipartite graphs and a generalization of Pólya's problem," *Journal of Combinatorial Theory, Series B*, November 2010, Vol. 100, No. 6, pp. 650-670.
- [12] C. Kayser, C. I. Petkov, M. Lippert, and N. K. Logothetis, "Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map," *Current Biology*, 8. November 2005, Vol. 15, No. 21, pp. 1943-1947.
- [13] I. Kocur, R. Parajasegaram, and G. Pokharel, Global Data on Visual Impairment in the Year 2002. *Bulletin of the World Health Organization*, 82, pp.844-851, 2004.
- [14] G. Kreiman, 2008, "Biological object recognition," *Scholarpedia*, 3(6):2667. http://www.scholarpedia.org/article/Biological_object_recognition.
- [15] S. Leyk, R. Boesch, and R. Weibel, "Saliency and semantic processing: Extracting forest cover from historical topographic maps," *Pattern Recognition*, May 2006, Vol. 39, No. 5, pp. 953-968.
- [16] K. Makino, T. Takabatake, and S. Fujishige, "A simple matching algorithm for regular bipartite graphs," *Information Processing Letters*, November 2002, Vol. 84, No. 4, pp. 189-193.
- [17] 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
- [18] R. Munoz-Salinas, E. Aguirre, M. Garcia-Silvente, and A. Gonzalez, Door-detection using computer vision and fuzzy logic. In *Proceedings of the 6th WSEAS International Conference on Mathematical Methods & Computational Techniques in Electrical Engineering*, 2004.
- [19] A. Murillo, J. Kosecka, J. Guerrero, and C. Sagues, Visual door detection integrating appearance and shape cues. *Robotics and Autonomous Systems*, 2008.
- [20] A. Neves, A. Pinho, D. Martins, and B. Cunha, "An efficient omnidirectional vision system for soccer robots: From calibration to object detection," *Mechatronics*, March 2011. Vol. 21, No. 2, pp. 399-410.
- [21] A. Noureldin, A. El-Shafie, and M. R. Taha, "Optimizing neuro-fuzzy modules for data fusion of vehicular navigation systems using temporal cross-validation," *Engineering Applications of Artificial Intelligence*, February 2007. Vol. 20, No. 1, pp. 49-61.
- [22] C. R. Pranesachar, "A class of matching-equivalent bipartite graphs," *Discrete Mathematics*, May 1999. Vol. 203, No. 1-3, pp. 207-213.
- [23] M. Potter and E. Levy, Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 1969. 81: 10-15.
- [24] S. J. Park, K. H. An, and M. Lee, "Saliency map model with adaptive masking based on independent component analysis," *Neurocomputing*, December 2002, Vol. 49, No. 1-4, pp. 417-422.
- [25] 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
- [26] K. Riesen, and H. Bunke, "Approximate graph edit distance computation by means of bipartite

Network Modeling Analysis in Health Informatics and Bioinformatics

- graph matching," *Image and Vision Computing*, June 2009. Vol. 27, No. 7, pp. 950-959.
- [27] A. Shokoufandeh, I. Marsic, and S. J. Dickinson, "View-based object recognition using saliency maps," *Image and Vision Computing*, April 1999. Vol. 17, No. 5-6, pp. 445-460.
- [28] H. Schneiderman and T. Kanade. A statistical approach to 3d object detection applied to faces and cars. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, 2000.
- [29] 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.
- [30] H. Shi, and Y. Yang, "A computational model of visual attention based on saliency maps," *Applied Mathematics and Computation*, May 2007, Vol. 188, No. 2, pp. 1671-1677.
- [31] 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, Publication Date: 12/26/00, pp. 414-448.
- [32] Seeing with Sound – The vOICe, <http://www.seeingwithsound.com/>
- [33] 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/>
- [34] S. Thorpe, D. Fize, and C. Marlot, Speed of processing in the human visual system. *Nature*, 1996. 381: 520-522.
- [35] Y. Tian, X. Yang, C. Yi, and A. Arditì, "Toward a Computer Vision-based Wayfinding Aid for Blind Persons to Access Unfamiliar Indoor Environments," *Machine Vision and Applications*, 2012.
- [36] W. Tsao, A. Lee, Y. Liu, T.-W. Chang, and H. -H. Lin, "A data mining approach to face detection," *Pattern Recognition*, March 2010, Vol. 43, No. 3, pp. 1039-1049.
- [37] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '01)*, 2001.
- [38] M. Weber, M. Welling, and P. Perona, "Unsupervised Learning of Models for Recognition," *Proc. European Conf. Computer Vision*, vol. 2, pp. 101-108, 2000.
- [39] Y. Wang, W. Lin, and S. Horng, "A sliding window technique for efficient license plate localization based on discrete wavelet transform," *Expert Systems with Applications*, April 2011, Vol. 38, No. 4, pp. 3142-3146.
- [40] S. Wang and Y. Tian, Indoor Signage Detection Based on Saliency Map and Bipartite Graph Matching, International Workshop on Biomedical and Health Informatics (BHI), 2011.
- [41] S. Wang, C. Yi, and Y. Tian, "Signage Detection and Recognition for Blind Persons to Access Unfamiliar Environments," *Journal of Computer Vision and Image Processing*, Vol. 2, No. 2, 2012.
- [42] X. Yang, Y. Tian, C. Yi, and A. Arditì, Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environments, International Conference on ACM Multimedia, 2010.
- [43] 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.
- [44] 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.
- [45] C. Yi and Y. Tian, Assistive Text Reading from Complex Background for Blind Persons, In book *Camera-Based Document Analysis and Recognition*, the 4th International Workshop on (CBDAR'11), Revised Selected Papers, Lecture Notes in Computer Science series, Vol. 7139, by Springer, Masakazu Iwamura and Faisal Shafait (Eds), ISSN 0302-9743, 2012.
- [46] S. Yuan, Y. Tian, and A. Arditì, Clothing Matching for Visually Impaired Persons, *Technology and Disability*, Vol. 23, 2011.
- [47] A. Zandifar, R. Duraiswami, A. Chahine, L. Davis, "A video based interface to textual information for the visually impaired". In: *Proc. IEEE 4th international conference on multimodal interfaces*, 2002.
- [48] Y. D. Zhang, and L. N. Wu, "Pattern recognition via PCNN and Tsallis entropy," *Sensors*, 2008. Vol. 8, No. 11, pp. 7518-7529.