### ORIGINAL ARTICLE

# Detecting signage and doors for blind navigation and wayfinding

Shuihua Wang · Xiaodong Yang · Yingli Tian

Received: 18 September 2012/Revised: 26 November 2012/Accepted: 19 January 2013/Published online: 6 February 2013 © Springer-Verlag Wien 2013

**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 using a saliency map. Then the signage is detected in the attended areas 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

S. Wang · X. Yang · Y. Tian (☒)
Department of Electrical Engineering,
The City College of New York, CUNY,
New York, NY 10031, USA
e-mail: ytian@ccny.cuny.edu

S. Wang

e-mail: swang15@ccny.cuny.edu

X. Yang

e-mail: xyang02@ccny.cuny.edu

detects instances of a priori objects of a certain class (such as faces, signs, buildings, etc.) in digital images and videos captured by cameras (Neves et al. 2011). 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

vision and 37 million were blind (Kocur et al. 2004).

Robust and efficient indoor object detection can help people with severe vision impairment to independently

access unfamiliar indoor environments and avoid dangers

(Baker 2010). Signage and other visual information pro-

vide 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 cate-

gories 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 loca-

Object detection is a computer vision technique that

tional signage of indoor environments.

and the targeted signage.

Object detection and recognition is a fundamental component for scene understanding. The human visual system is powerful, selective, robust, and fast (Kreiman 2008). 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

Springer

Fig. 1 Typical indoor signage: a directional signage and b locational signage



visual system can discriminate among at least tens of thousands of different object categories (Biederman 1987). Object recognition processes in the human visual system are also very fast: it can take as little as 100–200 ms (Hung et al. 2005; Potter and Levy 1969; Thorpe et al. 1996). 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 (Schneiderman and Kanade 2000; Viola and Jones 2001; Weber et al. 2000).

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 using a live camera 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 (Ivanchenko et al. 2008a, b; Manduchi et al. 2008; Shen and Coughlan 2007), such as text detection (Shen and Coughlan 2007), crosswatch (Ivanchenko 2008a, b), and wayfinding (Manduchi 2008). To help the visually impaired, Zandifar et al. (2002) 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. (2003) 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. 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. (2008) 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 to assist people who are visually impaired including navigation and wayfinding (Tian et al. 2012; Yang et al. 2010), clothing pattern and color matching and recognition (Yang and Yuan 2011; Yuan et al. 2011), banknote recognition (Hasanuzzaman et al. 2011), text reading (Yi and Tian 2011, 2012), signage detection (Wang and Tian 2011; Wang 2012), 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.

Detection of signage and doors can help blind people to find their destinations in unfamiliar environments. As illustrated in Fig. 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 and (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.

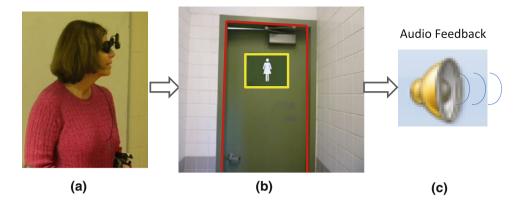
An earlier version of this paper can be found in (Wang and Tian 2011). Compared to our previous work, there are two major extensions whose merit being highlighted.

- 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 follows: Sect. 2 describes the methodology of signage detection. Sect. 3 introduces the



Fig. 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, and c audio feedback is provided to the user

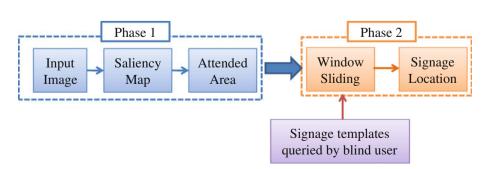


door detection method based on a general door geometric model. Sect. 4 analyzes experimental results and demonstrates the effectiveness and efficiency of the proposed algorithm. Sect. 5 concludes the paper and lists the future research directions.

#### 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 method suffers from two shortcomings (Tsao et al. 2010): (1) high processing time and (2) inaccuracy of detection results due to different backgrounds. Therefore, we propose a new method to detect indoor signage which first employs the saliency map (Levk et al. 2006; Shi and Yang 2007) to extract attended areas, and then applies bipartite graphic matching (Makino et al. 2002; Riesen and Bunke 2009; Zhang and Wu 2008) 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 Fig. 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.

**Fig. 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 (Davies et al. 2006). 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 surroundings, in many sub-modalities and at many scales (Park et al. 2002). Saliency at a given location is determined primarily by how different this location is from its surroundings in color, orientation, motion, depth, etc.

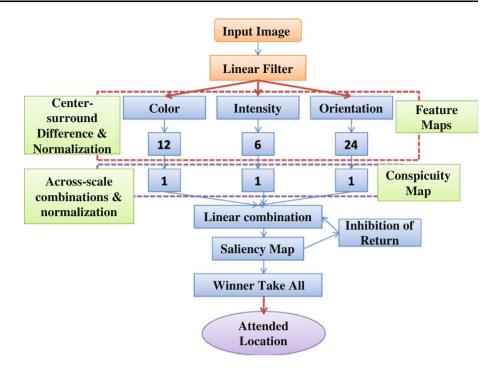
As shown in Fig. 4, the different visual features that contribute to attentive selection of a stimulus (color, intensity, and orientation) are combined into one saliency map. The saliency map which integrates the normalized information from the individual feature maps into one global measure of conspicuity.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

#### 2.1.1 Initialization

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



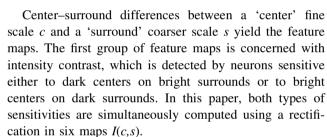
Fig. 4 Architecture of building saliency maps based on three types of image features: intensity, color, and orientation



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 yields truly multi-scale feature extraction, by including different size ratios between the centers and surround regions.

#### 2.1.2 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 variations 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|/2g/2-b for yellow, and negative values are set to zero for the R, G, B, and Y values gained from above equations. Four Gaussian pyramids  $R(\sigma)$ ,  $G(\sigma)$ ,  $B(\sigma)$ , and  $Y(\sigma)$  are created from these color channels.



$$I(c,s) = |I(c) - I(s)| \tag{1}$$

### 2.1.3 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 (Shokoufandeh 1999). 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).

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

$$BY(c,s) = |(B(c) - Y(c)) - (Y(s) - B(s))| \tag{3}$$

### 2.1.4 Orientation-based saliency map

Local orientation information is gained from *I* by oriented Gabor pyramids  $O(\sigma, \theta)$  where  $\sigma \in [0, 1, ..., 8]$  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.

$$O(c, s, \theta) = |O(c, \theta) - O(s, \theta)| \tag{4}$$

# 2.1.5 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 (Kayser et al. 2005). 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(.) 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(.) consists of

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

Only local maxima activities are considered, such that N(.) 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",  $\overline{I}$  for intensity (5),  $\overline{C}$  for color (6), and  $\overline{O}$  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.

$$\overline{I} = \bigoplus_{c=2}^{4} \sum_{s=c+3}^{c=4} N(I(c,s))$$

$$\tag{5}$$

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

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

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

The motivation of the creation of three separate channels  $(\overline{I}, \overline{C})$ , and  $(\overline{O})$  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.

$$S = \frac{1}{3} \left( N(\overline{I}) + N(\overline{C}) + N(\overline{O}) \right) \tag{8}$$

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

#### 2.2.1 Detecting signage in attended areas

The sliding method window is a popular technique for identifying and localizing objects in an image (Davignon et al. 2005). 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 (Noureldin et al. 2007). 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 (Wang et al. 2011).

# 2.2.2 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 detailed procedures of the BGM



algorithm can be found at papers (Kakimura 2010; Pranesachar 1999). Suppose *A* denotes the window and *B* denotes the sub-image covered by window A. The BGM algorithm first calculates the edges of the window and the sub-image, followed by evaluating the degree of the match as

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

where  $\omega$  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:

$$B^* = \arg\max(BGM(A, B)) \tag{10}$$

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 ith signage pattern; S denotes the saliency map; BGM denotes the value of bipartite graph matching; and  $L_i$  denotes the location found for the ith pattern.

```
Step1 Initialization. Input I, P_1, P_2, ... P_N;

Step2 S = GetSaliencyMap(I);

Step3 for i = 1: N [BGM_i, L_i] = SinglePatternDetection

(I, S, P_i); end

Step4 i^* = argmax \{BGM_i\};

Step5 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 (Chen and Birchfield 2008; Munoz-Salinas et al. 2004; Murillo et al. 2008). 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 (Chen and Birchfield 2008). Murillo et al. (2008) 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. (2004) 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 Fig. 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.

#### 3.1 Edge and corner-based geometric door model

As shown in Fig. 6a, 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 the 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 Fig. 6b, 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



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

Fig. 6 a Geometric door

model, **b** detected edge map, and **c** extracted corners

(a) Inset

Lateral & Door Frame Contours Model Information Edge Map Source **Image** Image Preprocessing Door-corner Detected Corners **Candidates** Door **■**c4 **L34** L12 (a) (b) (c)

(c) Protruding

**Fig. 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

(b) Inset

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

(d) Protruding

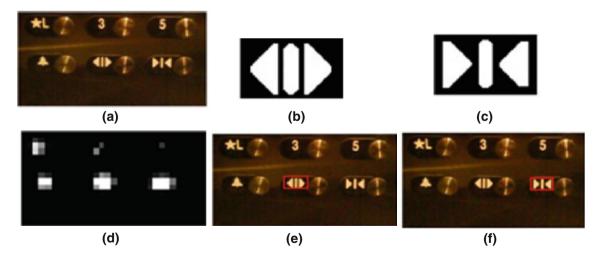


Fig. 8 Elevator button detection: a original image; b query elevator button of "Open" symbol; c query elevator button of "Close" symbol; d saliency map of the original image; e detected "Open" button; and f detected "Close" button

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 illustrates the different geometric characteristics and different relative positions of lateral faces and the detected door frame. In Fig. 7a, c, the door or door-like object is located on the right side with respect to the



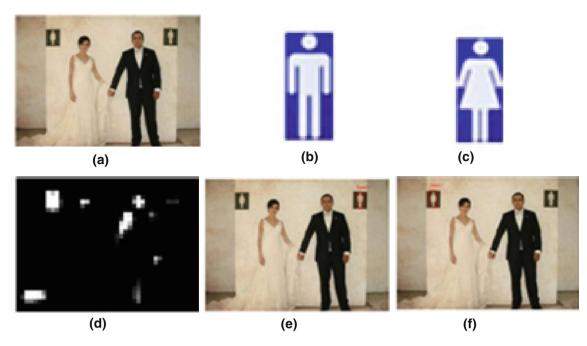


Fig. 9 Restroom signage detection: a original image, b query pattern of "Men's" signage, c query pattern of "Women's" signage, d saliency map, e detected "Men's" signage, and f detected 'Women's" signage

Fig. 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)





camera if  $L_{12} < L_{34}$  of the detected frame. The lateral  $(C_1 \ C_2 \ C_5 \ C_6)$  position of the doorframe  $(C_1 \ C_2 \ C_3 \ C_4)$  indicates that the object is inset (left side) or protruding (right side). Similarly as shown in Fig. 7b, 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 a paper by Tian et al. (2012).

### 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 displays detection of elevator buttons. The camera-captured image of six different elevator buttons is

shown in Fig. 8a. Figure 8d is the corresponding saliency map extracted from the original image using intensity, color, and orientation. The bright pixels indicate the attended areas. Figure 8b, c show the query symbols of "open" and "close" buttons which are recorded in the query database of indoor signage. Figure 8e, f demonstrates the final detection results (the red rectangular boxes).

Similarly, the detection of restroom signage is displayed in Fig. 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.

# 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 Fig. 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 in the third row for a "Women's" restroom.

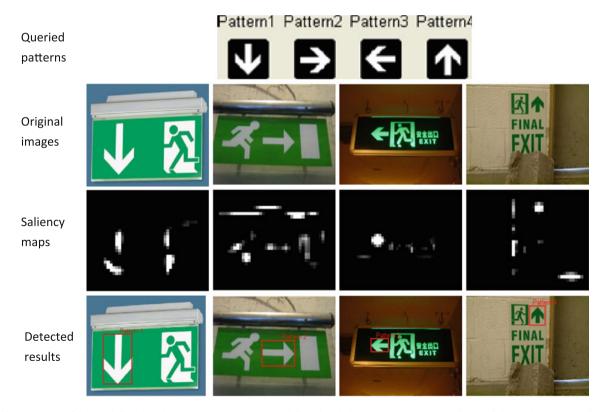


Fig. 11 Examples of detected directional signage. First row: queried directional signage patterns ("Down", "Right", "Left", and "Up"); second row: original images; third row: saliency maps; fourth row: detected results of the directional signage



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 Fig. 11.

### 4.3 Experimental results

Our proposed method is evaluated by a database of 182 indoor signage images with nine 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 Fig. 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-pattern detection has a lower successful rate than that of single-pattern detection.

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 and (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 Fig. 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.

#### 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-

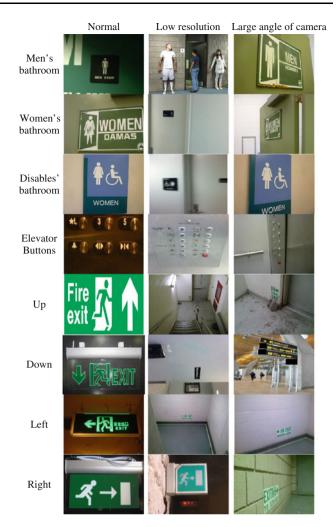


Fig. 12 Example images of our indoor signage dataset

Table 1 Accuracy of single pattern detection for bathroom signs

| Signage classes | No. of samples | Correctly detected | Detection<br>accuracy (%) |
|-----------------|----------------|--------------------|---------------------------|
| Men's           | 20             | 16                 | 80                        |
| Women's         | 20             | 18                 | 90                        |
| Disables'       | 20             | 18                 | 90                        |
| Close           | 20             | 17                 | 85                        |
| Open            | 20             | 17                 | 85                        |
| Total           | 100            | 86                 | 86                        |

The numbers in bold indicate the total numbers and the average accuracies of all categories

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 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.



| Table 2   | Accuracy of detect      |   |
|-----------|-------------------------|---|
| and recog | gnize multiple patterns | s |

| 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 numbers in bold indicate the total numbers and the average accuracies of all categories

Fig. 13 "Open" button of elevator: a original image, b saliency map

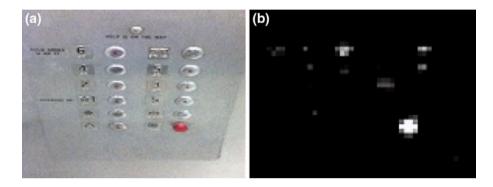


Fig. 14 "Men" restroom: a original image, b saliency map, c detection result

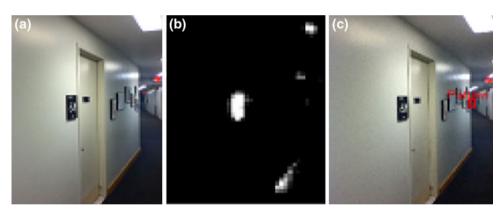


Table 3 Effectiveness of saliency map for indoor signage detection

| Classes  | No. of<br>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 %                           |

The numbers in bold indicate the total numbers and the average accuracies of all categories

# 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 textures, and doors captured from different viewpoints, illumination conditions, and occlusions, as well as open and glass doors. Non-door images include doorlike objects, such as bookshelves and cabinets. We categorized the database into three groups: *Simple* (57 images, see examples of the first row in Fig. 15), *Medium* (113 images, see examples of the second row in Fig. 15), and *Complex* (51 images, see examples of the third row in Fig. 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 Fig. 15.



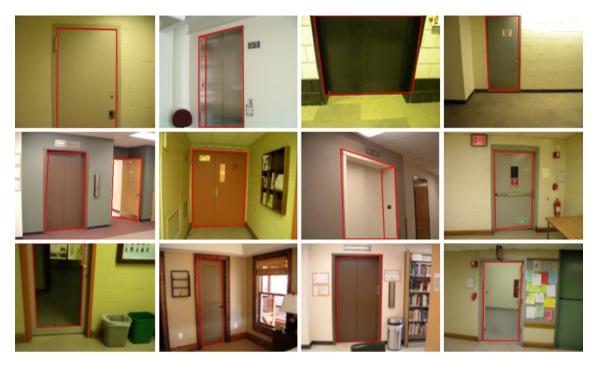


Fig. 15 Examples of successfully detected doors in different environments. First row: simple; second row: medium; third row: complex

Table 4 Door detection results

| Data category | True positive rate (%) | False positive rate (%) |
|---------------|------------------------|-------------------------|
| Simple        | 98.2                   | 0                       |
| Medium        | 90.5                   | 3.5                     |
| Complex       | 80.0                   | 2.0                     |
| Total         | 89.5                   | 2.3                     |

The numbers in bold indicate the total numbers and the average accuracies of all categories

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

# References

Baker A (2010) Blind man is found dead in elevator shaft. The New York Times, City Room, May

Biederman I (1987) 1987, Recognition-by-components: a theory of human image understanding. Psychological Rev 94:115–147

Chen Z, Birchfield S (2008) Visual detection of lintel-occluded doors from a single image. IEEE computer society workshop on visual localization for mobile platforms

Davies C, Tompkinson W, Donnelly N, Gordon L, Cave K (2006) Visual saliency as an aid to updating digital maps. Comput Hum Behav 22(4):672–684

Davignon F, Deprez JF, Basset O (2005) A parametric imaging approach for the segmentation of ultrasound data. Ultrasonics 43(10):789–801

Everingham M, Thomas B, Troscianko T (2003) Wearable mobility aid for low vision using scene classification in a markov random field model framework. Int J Hum Comput Interact 15(2)

Hasanuzzaman F, Yang X, Tian Y (2011) Robust and effective component-based banknote recognition for the blind, Accepted, IEEE transactions on systems, man, and cybernetics—Part C: applications and reviews

Hung C, Kreiman G, Poggio T, DiCarlo J (2005) Fast read-out of object identity from macaque inferior temporal cortex. Science 310:863–866

Ivanchenko V, Coughlan J, Shen H (2008) 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)



- Ivanchenko V, Coughlan J, and Shen H (2008) 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)
- Kakimura N (2010) Matching structure of symmetric bipartite graphs and a generalization of Pólya's problem. J Combinatorial Theory Ser B 100(6):650–670
- Kayser C, Petkov CI, Lippert M, Logothetis NK (2005) Mechanisms for allocating auditory attention: an auditory saliency map. Curr biol 8 15(21):1943–1947
- Kocur I, Parajasegaram R, Pokharel G (2004) Global data on visual impairment in the year 2002. Bull World Health Organ 82:844–851
- Kreiman G (2008) Biological object recognition, Scholarpedia. 3(6):2667. http://www.scholarpedia.org/article/Biological\_object\_recognition
- Leyk S, Boesch R, Weibel R (2006) Saliency and semantic processing: extracting forest cover from historical topographic maps. Pattern Recogn 39(5):953–968
- Makino K, Takabatake T, Fujishige S (2002) A simple matching algorithm for regular bipartite graphs. Inf Process Lett 84(4): 189–193
- Manduchi R, Coughlan J, Ivanchenko V (2008) Search strategies of visually impaired persons using a camera phone wayfinding system. 11th International conference on computers helping people with special needs (ICCHP '08)
- Munoz-Salinas R, Aguirre E, Garcia-Silvente M, Gonzalez A (2004) Door-detection using computer vision and fuzzy logic. In: Proceedings of the 6th WSEAS international conference on mathematical methods and computational techniques in electrical engineering
- Murillo A, Kosecka J, Guerrero J (2008) c. Visual door detection integrating appearance and shape cues. Robotics and autonomous systems, Sagues
- Neves A, Pinho A, Martins D, Cunha B (2011) An efficient omnidirectional vision system for soccer robots: from calibration to object detection. Mechatronics 21(2):399–410
- Noureldin A, El-Shafie A, Taha MR (2007) Optimizing neuro-fuzzy modules for data fusion of vehicular navigation systems using temporal cross-validation. Eng Appl Artificial Intelligence 20(1):49–61
- Park SJ, An KH, Lee M (2002) Saliency map model with adaptive masking based on independent component analysis. Neurocomputing 49(1–4):417–422
- Potter M, Levy E (1969) Recognition memory for a rapid sequence of pictures. J Exp Psychol 81:10–15
- Pradeep V, Medioni G, Weiland J (2008) Piecewise planar modeling for step detection using stereo vision, workshop on computer vision applications for the visually impaired
- Pranesachar CR (1999) A class of matching-equivalent' bipartite graphs. Discrete Math 203(1-3):207-213
- Riesen K, Bunke H (2009) Approximate graph edit distance computation by means of bipartite graph matching. Image Vis Comput 27(7):950–959
- Schneiderman H, Kanade T (2000) A statistical approach to 3d object detection applied to faces and cars. In IEEE conference on computer vision and pattern recognition (CVPR'00)
- Shen H, Coughlan J (2007) Grouping using factor graphs: an approach for finding text with a camera phone. Workshop on graph-based representations in pattern recognition
- Shi H, Yang Y (2007) A computational model of visual attention based on saliency maps. Appl Math Comput 188(2):1671–1677

- Shokoufandeh A, Marsic I, Dickinson SJ (1999) View-based object recognition using saliency maps. Image Vision Comput 17(5–6): 445–460
- 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., Jain L.C., CRC Press, ISBN/ ISSN: 0849301408, Publication Date: 12/26/00, 414–448
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520–522
- Tian Y, Yang X, Yi C, Arditi A (2012) Toward a computer visionbased Wayfinding aid for blind persons to access unfamiliar indoor environments machine vision and applications
- Tsao W, Lee A, Liu Y, Chang T-W, Lin H-H (2010) A data mining approach to face detection. Pattern Recogn 43(3):1039-1049
- Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In IEEE conference on computer vision and pattern recognition (CVPR'01)
- Wang S, Tian Y (2011) Indoor signage detection based on saliency map and bipartite graph matching, International workshop on biomedical and health informatics (BHI)
- Wang Y, Lin W, Horng S (2011) A sliding window technique for efficient license plate localization based on discrete wavelet transform. Expert Syst Appl 38(4):3142–3146
- Wang S, Yi C, Tian Y (2012) Signage detection and recognition for blind persons to access unfamiliar environments. J Comput Vis Image Process 2(2)
- Weber M, Welling M, Perona P (2000) Unsupervised learning of models for recognition. Proc Eur Conf Comput Vision 2: 101–108
- Seeing with sound—the voice, http://www.seeingwithsound.com/
- 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/
- Yang X, Tian Y, Yi C, Arditi A (2010) Context-based indoor object detection as an aid to blind persons accessing unfamiliar environments, International conference on ACM multimedia
- Yang X, Yuan S, Tian Y (2011) Recognizing clothes patterns for blind people by confidence margin based feature combination, International conference on ACM multimedia
- Yi C and Tian Y (2011) Assistive Text Reading from Complex Background for Blind Persons, The 4th International Workshop on Camera-Based Document Analysis and Recognition (CBDAR)
- Yi C, Tian Y (2012) 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, Iwamura M, Shafait F (eds), ISSN 0302-9743
- Yuan S, Tian Y, Arditi A (2011) Clothing matching for visually impaired persons, technology and disability, 23
- Zandifar A, Duraiswami R, Chahine A, Davis L (2002) A video based interface to textual information for the visually impaired. In: Proceeding of IEEE 4th international conference on multimodal interfaces
- Zhang YD, Wu LN (2008) Pattern recognition via PCNN and Tsallis entropy. Sensors 8(11):7518–7529

