Robust Sign Recognition System at Subway Stations Using Verification Knowledge

Dongjin Lee, Hosub Yoon, Myung-Ae Chung, and Jaehong Kim

In this paper, we present a walking guidance system for the visually impaired for use at subway stations. This system, which is based on environmental knowledge, automatically detects and recognizes both exit numbers and arrow signs from natural outdoor scenes. The visually impaired can, therefore, utilize the system to find their own way (for example, using exit numbers and the directions provided) through a subway station. The proposed walking guidance system consists mainly of three stages: (a) sign detection using the MCT-based AdaBoost technique, (b) sign recognition using support vector machines and hidden Markov models, and (c) three verification techniques to discriminate between signs and non-signs. The experimental results indicate that our sign recognition system has a high performance with a detection rate of 98%, a recognition rate of 99.5%, and a false-positive error rate of 0.152.

Keywords: Sign detection, sign recognition, character recognition, verification, support vector machine, SVM, hidden Markov model, HMM, natural-scene images.

I. Introduction

The automatic detection and recognition of both signs and text in real-world environments has been studied for several years [1]–[3]. One of the reasons for such studies is the importance in obtaining useful information about real-word environments so as to assist the visually impaired. According to the Korean Ministry of Health and Welfare, the number of visually impaired persons in South Korea grew from 180,526 in 2005 to 249,259 in 2011, which is an increase of 38%. However, there were only 65 guide dogs in South Korea in 2011. For this reason, we have been developing a walking guidance system for application at subway stations. This system allows us to extract signs from a real-world image and to classify them into specific categories. The final guiding message is then provided to the blind person through tiny speakers embedded in their hat. In this way, the system can help the visually impaired find their way around a subway station.

The main role of the walking guidance system is to detect candidates for subway signs from natural scenes and then to filter out non-sign regions correctly. One of the representative algorithms for this type of system is AdaBoost, which many researchers have adopted for the detection of signs and texts [4]–[6]. Although the algorithm extracts signs and texts from a cluttered background exceptionally well, it is sensitive to noisy data and outliers. Furthermore, it has a difficult time detecting most signs that are smaller than 20 pixels in diameter [7].

To solve these problems, as described in Section III of this paper, we suggest a detection method that is robust even for small or geometrically distorted signs in natural-scene images. In Section IV, we introduce the three verification techniques used to identify candidate sign regions. In addition, the results

Manuscript received Jan. 10, 2014; revised Sept. 2, 2014; accepted Sept. 5, 2014.

This research was supported by the Converging Research Center Program through the Ministry of Science, ICT and Future Planning, Korea (2013K000329) and the R&D program of MOTIE & KEIT [10041610, The development of the recognition technology for user identity, behavior, and location that has a performance approaching recognition rates of 99% on 30 people by using perception sensor network in the real environment].

Dongjin Lee (robin2002@etri.re.kr), Hosub Yoon (corresponding author, yoonhs@etri.re.kr), and Jaehong Kim (jhkim504@etri.re.kr) are with the IT Convergence Technology Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Myung-Ae Chung (machung@etri.re.kr) is with the Future Research Creative Laboratory, ETRI, Daejeon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.14.2214.0007

demonstrating the superiority of our proposed system are provided. We end this paper with our conclusions in Section V.

II. System Overview

The aim of our walking guidance system is to detect and recognize two different types of signs: exit numbers and arrows from the natural-scene images found at subway stations. As shown in Fig. 1, our system consists of the following three main stages:

1) Detection. Candidate sign (arrows and numbers) regions are detected from the natural subway scenes using the MCT-AdaBoost algorithm. A large set of images are then used to train the sign detector. A general sign detector can occasionally miss a small numeral, such as an exit number located behind a detected arrow. In this case, we can use an additional *a prior* knowledge-based detection algorithm that allows signs and exit numbers to be detected in a pairwise manner. Figure 2 illustrates the types of signs used in our work.

2) Recognition. The recognition process is based on the following three steps:

- Support vector machine (SVM)-based feature extraction and classification.
- Hidden Markov model (HMM)-based feature extraction and classification.
- SVM/HMM hybrid recognition.

3) Knowledge-based verification. Non-sign regions are excluded using the following three effective means of verification:

- HMM log-likelihood
- Hough transform (HT)
- Color information

To capture the scene images, we designed a hat with two embedded webcams, as shown in Fig. 3. The two webcams allow us to obtain a wider field of view (FOV) (100°) (1,280 pixels × 480 pixels) with a real-time processing of about 3 fps to 8 fps using a 1.8 GHz Samsung ATIV smart PC.

III. Sign Detection

1. AdaBoost-Based Detection

To detect candidate subway signs, we adopted the MCT-AdaBoost technique, which was originally used for face detection and aligning a detected face image [8]–[9]. This technique is divided into three steps. First, to train the classifiers of the proposed sign detector, it is necessary to obtain a large set of sign (positive) and non-sign (negative) images from subway stations. We therefore collected a large set of images in seven different subway stations in Daejeon, South Korea. Second, we used the modified census transform (MCT) feature, which compares the intensity of each pixel in a 3 × 3 kernel with the kernel mean. The resulting values of the MCT feature are integer indices representing a local structure of a pixel. Third, we trained the weak classifiers of AdaBoost by examining the distribution of the integer indices for the pixels.

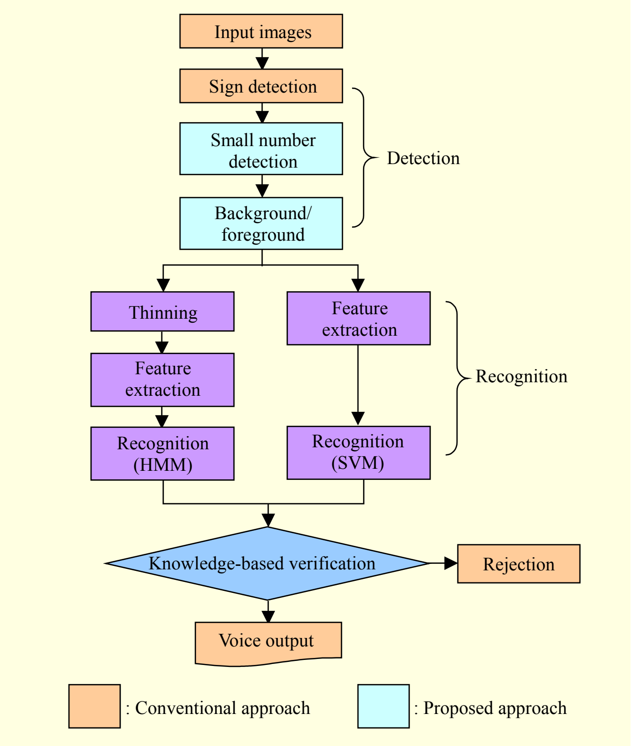


Fig. 1. Flowchart of proposed walking guidance system.

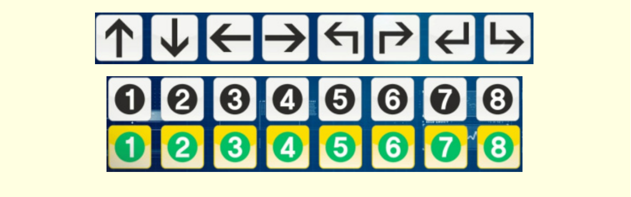


Fig. 2. Examples of eight directional arrows and exit number patterns.

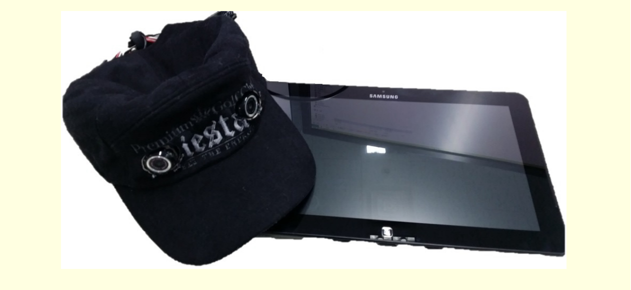
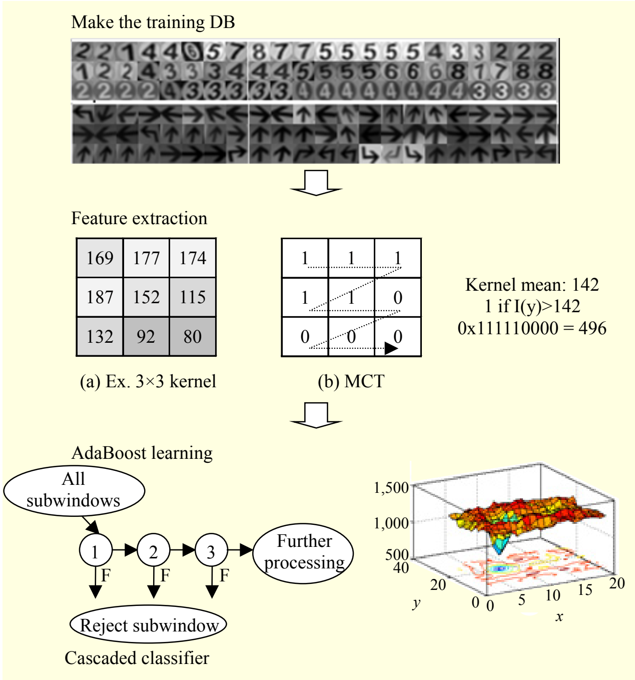


Fig. 3. Hat embedded with two webcams and a smart PC.



169	177	174
187	152	115
132	92	80

1	1	1
1	1	0
0	0	0

Fig. 4. Steps of MCT-AdaBoost detector.

A combination of these weak classifiers was used as an input for the AdaBoost algorithm so as to create a strong classifier. Figure 4 shows the steps of the proposed MCT-AdaBoost detector.

In our work, 30,000 training samples were determined to be positive images, and 80,000 were determined to be negative images. Thus, we were able to make a respectable sign detector that has a high success rate of 95%, with a false-positive rate of 5%. To reduce false-positive sign results, we introduce several verification methods in the next subsection.

2. Detection of Small Numbers

In this section, we introduce a detection technique that is robust with respect to small and geometrically distorted signs from subway scenes. The main idea is simple but very effective. We first needed to detect and recognize relatively large arrow signs, and then detect a small sign within a limited area of the scene next to this large sign. In our case, we used the prior knowledge that the small exit number signs were located next to the arrow signs. Moreover, a region including small signs has a black (or close to black) background such that the correct region is, therefore, mostly darker than other regions, as shown in Fig. 5(a). From this idea, both candidate regions (the same size as the arrow sign) were investigated to choose the correct one. We then worked out the coordinates of the selected region of interest (ROI) in the scene under the following conditions:

- ROI to the left of the arrow sign.
  - ROI.left = Arrow.left – (Arrow.width × 3)
  - ROI.top = Arrow.top – (Arrow.height / 2)
  - ROI.right = Arrow.left – 1
  - ROI.bottom = Arrow.bottom + (Arrow.height / 2)
- ROI to the right of the arrow sign.
  - ROI.left = Arrow.right + 1
  - ROI.top = Arrow.top – (Arrow.height / 2)
  - ROI.right = Arrow.right + (Arrow.width × 3)
  - ROI.bottom = Arrow.bottom + (Arrow.height / 2)

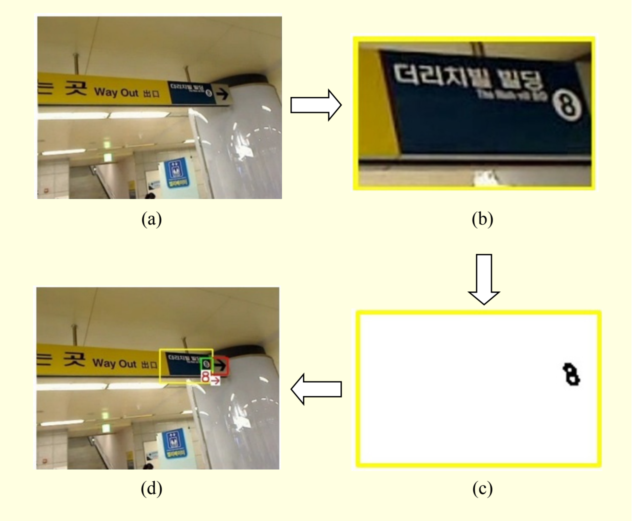


Fig. 5. Sequence of detecting and recognizing small signs: (a) detected arrow, (b) selected candidate ROI from the detected arrow, (c) result of a blob analysis on ROI, and (d) recognition results.

Figure 5(b) shows the selected candidate region. Next, Sauvola's locally adaptive binarization method was performed to binarize the selected region [10]. This method is an improvement of Niblack (a local masking method) and is especially for use on badly illuminated documents [11]. Then, a connected-component analysis (CCA) algorithm was performed to inspect all of the blobs and to remove non-sign blobs, as shown in Fig. 5(c) [12]. The CCA method is adopted in computer vision to find connected regions in binary images. Finally, we were able to recognize the sign blobs, as shown in Fig. 5(d), and verify the results using an adaptive Hough transform (AHT), the algorithm of which is described in detail in Section V-2. From this proposed approach, we can achieve a detection rate of 98%.

3. Foreground/Background Separation

The purpose of this step is to perform foreground/background separation. In Fig. 6, we show two types of exit numbers with different foreground/background structures. Our detection algorithm provides these two types of areas, and we then have to separate each type.

http://dx.doi.org/10.4218/etrij.14.2214.0007

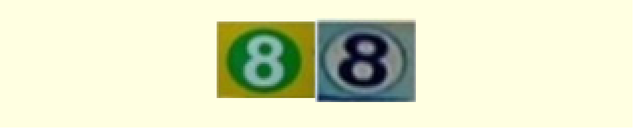


Fig. 6. Two types of exit numbers with different backgrounds.

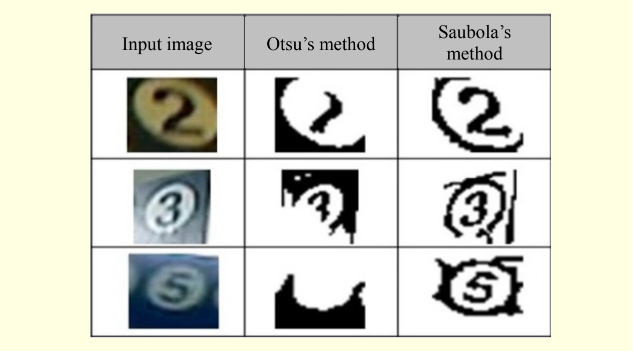


Fig. 7. Comparison of binarization with small signs.

In our previous research, we utilized the Otsu threshold, which automatically finds the optimal threshold value to minimize the intra-class variance of the white and black pixels. This method has a better performance than the Niblack for our system to binarize a candidate region [12]. However, Otsu's method occasionally did not work well with small signs under varying types of illumination. Moreover, the method assumes an image histogram to be bimodal; hence, the method breaks down if the distribution of two classes of data are largely unequal. To cope with this problem, we used Sauvola's algorithm (explained in Section III-2) when the width or height of the detected region was less than 37 pixels. The results of this algorithm are shown in Fig. 7.

Next, in the segmentation step, we performed the CCA (explained in Section III-2) to eliminate non-sign blobs from the labeled blobs. For this, we used the following conditions:

- Too small a blob: $W < 0.33 \times I_W, H < (0.33 \times I_H$ or 9 pixels).
- Too small or large an area compared to the size of the bounding box: $A < 0.18 \times W \times H, A > 0.62 \times W \times H$.

Here, the width (W), height (H), area (A), and center of the bounding box (X, Y) of the blobs are used, and I_W and I_H are the width and height of the detected sign region. When none of the blobs satisfied the above conditions, we rejected the detected sign region. Finally, the selected blob was normalized by 30 pixels × 30 pixels, and a median filter was then used to smooth the blob.

IV. Sign Recognition

1. Learning-Based Recognition

A. SVM-Based Recognition

In our previous work, we adopted the gradient-based feature extraction method proposed by Liu [13]. For this gradient feature extraction, we computed a gradient for each pixel to calculate a weighted vote for an orientation histogram based on the gradient magnitude, and the votes were then accumulated into eight orientation bins over the interval [0, 2π]. The gradient is computed by a Sobel Operator, which was employed to calculate the approximations of the horizontal and vertical derivatives. Next, to reduce the dimensionality of the feature vectors, the normalized 30 × 30 pixel blob in Section III-3 was divided into $N \times N$ blocks. Then, spatially connected pixels were merged into a block, so the combination of the pixel histograms created the block histogram. In our work, the block size was 5 × 5 pixels; thus, the dimensionality of the feature vectors was 200 (block size (5) × block size (5) × bin size (8)).

In the recognition step, the implementation of SVMs was based on the LIBSVM library, and we were therefore able to employ a multiclass classification using SVMs with a radial basis function (RBF) kernel, defined formally as [14]

$$\mathrm{K} = \exp\left(-\gamma \| X_i - X_j \|^2\right). \tag{1}$$

In this work, we trained 38,250 sample images, including 2,250 non-sign (noise) images, which were selected randomly from natural-scene images in subway stations.

B. HMM-Based Recognition

In the above section, we show an SVM-based classifier that can be used to classify sign and non-sign candidates well. However, we cannot train all samples in a real environment as there are a variety of signs and non-signs in subway stations. We therefore used an HMM-based classifier [15], which has several advantages. First, this classifier is a stochastic model and is robust to noise and shape variations. Second, we can use the log-likelihood of the HMM as a verification technique to discriminate signs from non-signs (noise).

In the first step, to obtain the skeleton of a normalized image, we compared two different types of thinning algorithms — Zhang's [16] and Ahmed's [17]. As a result, Ahmed's thinning algorithm is more suitable for our work because this algorithm preserves the shape of the normalized image such that it is invariant to rotation, as shown in Fig. 8. This method is a rule-based thinning algorithm to generate a single-pixel-wide skeleton in a binary image and has the advantage of low computational cost.

For this reason, we adopted Ahmed's thinning method. In the next step, we created a chain code by tracing the skeleton of a thinned image and storing the tracking information in a vector space. We then made the chain code smoother such that its effect is similar to a median filter [18]. Next, we found the most important points, such as end, curve, and branch points, in the

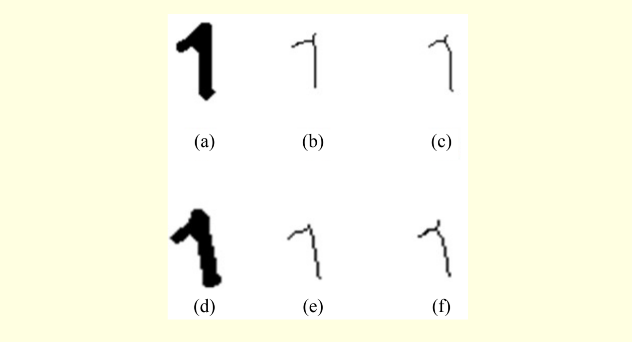


Fig. 8. Comparison of two thinning algorithms: (a) input image, (b) thinned image of (a) using Zhang's algorithm, (c) thinned image of (a) using Ahmed's algorithm, (d) rotated input image, (e) thinned image of (d) using Zhang's algorithm, and (f) thinned image of (d) using Ahmed's algorithm.

```
-  -  -  E  -  -  -     -  -  -  0  -  -  -
-  -  6  -  -  -  -     -  -  6  -  -  -  -    0 6 6 6 61 1 1 1 93 1 0 3 0 7 0
-  6  -  -  E  -  -     -  6  -  -  0  -  -
6  -  -  -  3  -  -     6  -  -  -  3  -  -                 (c)
C  1  1  1  B  1  E     61 1  1  1  93 1  E    0 0 0 6 6 6 61 61 61 61 61 1 1
-  -  -  -  7  -  -     -  -  -  -  7  -  -    1 93 93 93 93 93 93 93 93 1 0 0 3
-  -  -  -  E  -  -     -  -  -  -  0  -  -              0 0 7 0 0
          (a)                     (b)                        (d)
```

Fig. 9. Method for generating a feature vector: (a) chain code of eight directions with end, curve, and branch points; (b) change the important points of each feature value; (c) ordered feature set from (b); and (d) set of weighted feature values for the significant points.

chain code. Finally, we created the weighted chain code features described in our previous study [12]. Figure 9 shows how to generate feature vectors from a chain code. This set of weighted features was used to create the model for each class using HMMs.

2. Knowledge-Based Verification

In spite of high classification rates, we still experienced erroneous classifications of non-signs as a false-positive recognition. For this reason, we additionally introduced compatible techniques with the log-likelihood of the HMM; that is, an HT and color information.

A. HMM Log-Likelihood Verification

In our previous research, we introduced the log-likelihood of the HMM as a verification technique to discriminate signs from non-sign regions [12]. The decision to accept or reject was based on the log-likelihood score of the HMM [19]. To obtain the optimal threshold, we examined 40,000 images (2,500 per class). However, we still experienced erroneous classifications of non-signs.

B. Number Verification

The HT was used to verify the exit number signs, as all such signs at subway stations in Daejeon have a circular shape. The circle equation can be described as follows:

$$(x-a)^2+(y-b)^2=r^2, \tag{2}$$

where (a, b) is the coordinate of the circle's center and r is the radius of the circle. To find circles in the sign regions, we utilized the AHT, which can help reduce both storage and computational requirements compared to the HT [20]. This algorithm is divided into two steps. The first step involves a two-parameter HT to find the center of the circle, and the second step involves a one-dimensional HT, which is a simple histogram used to identify the radius of the circle. For evaluating the performance of this method, we tested 1,500 samples of exit numbers and compared the results with those obtained using the ellipse detection method in [21]. As a result, the detection result of the AHT was 99.8%, whereas, the ellipse detection method only yielded 96.2%. This means that the AHT method is more suitable than the ellipse detection method, though most of the data (images), collected from a hat embedded with two webcams, in Fig. 3 is a little skewed. Figure 10(a) shows the detected circle using the AHT.

C. Arrow Verification

In this stage, color information is used to verify the arrow signs. We found that all of the signs can be classified into four different types, as shown in Fig. 10(b).

To use the color information, we adopted a multiclor model based on a hue-saturation-value (HSV) color space, which is more robust against illumination changes than the RGB color space [22].

A histogram of the HSV color space was made of $N = N_hN_s + N_v$ bins, and we denoted $b_t(p) \in \{1, \dots, N\}$ as the bin index associated with the color vector $y_t(p)$ at pixel location p in sign images t. Here, the number of bins N is 110, where N_h, N_s, and N_v are set to 10. A histogram of four different signs is shown in Fig. 11.

To compare the color model between the digit and arrow, we defined the candidate region as $R(x_t)$ of the state vector x_t, while the kernel density estimate $q_t(x) = \{q_t(n;x)\}_{n=1,\dots,N}$ of the

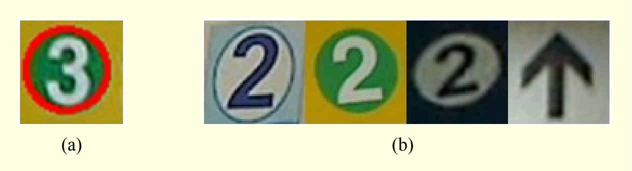


Fig. 10. Number verification techniques: (a) result of AHT and (b) a variety of signs from a subway station in Daejeon.

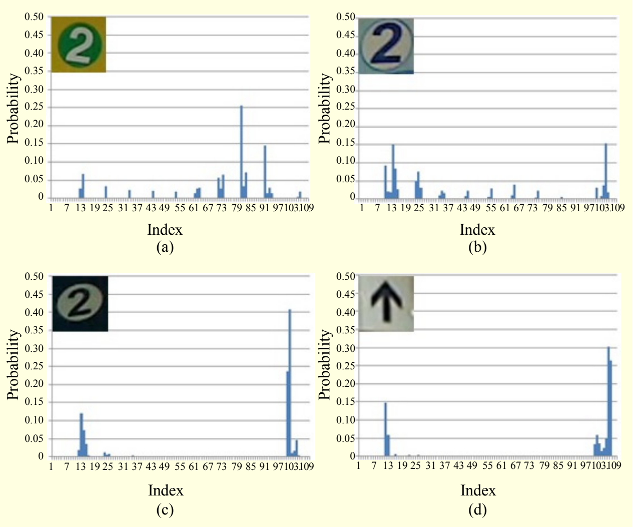


Fig. 11. Color (HSV) histograms of the four different signs.

color distribution at time *t* was then composed using [23]

$$q_t(n;x) = \mathrm{C} \sum_{p\in R(x)} \delta[b_t(p)-n], \qquad (3)$$

where δ is the Kronecker delta function, C is a normalizing constant that ensures $\sum_{n=1}^{N} q_t(n;x)=1$, and p is any pixel within $R(x_t)$. The number of bins N in the histogram is represented as a feature vector used to train a classifier using an SVM. In our work, we trained 16,560 samples of four different signs and tested the other 1,840 samples to evaluate the performance of the classifier. The classification result was 99.72%, and only five samples were misclassified.

Before the verification technique was applied, we combined the recognition results from two different types of classifiers, SVM and HMM, to get lower false positive rates. Because the low false positive rate is much more important than a high true positive rate to correctly inform the visually impaired person of sign information. For this work, we adopted Wu's pairwise coupling method using the SVM [24]. This method estimated the probabilities for multiclass classification by combining all pairwise comparisons of binary SVM classifiers. After that, when the recognition result was exit numbers and arrows and not noise, and the probability was less than a certain threshold, we rejected the result from the SVM. However, when the recognition result from the SVM and HMM was the same, we accepted this result without reference to the multiclass probabilities with SVMs.

In Table 1, we compared four different combinations. Table 1 shows the variation in average recognition rates and false positive error (FPE) from the four different cases. As we expected, adopting the proposed verification rules, we were able to obtain higher recognition rates and lower FPE rates,

Table 1. Average recognition rates and FPE of four different cases.

Verification case	FPE	Average recognition rate
SVM+HMM	1.087	97.9
SVM+HMM+log-likely	0.604	98.2
SVM+HMM+log-likely +number	0.257	98.9
SVM+HMM+log-likely + number+arrow	0.152	99.5

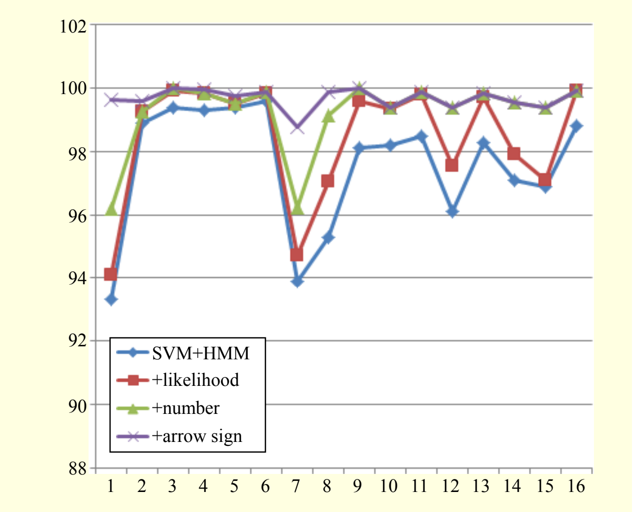
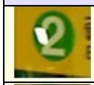
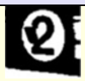
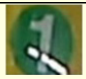
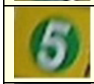
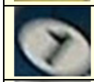
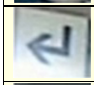
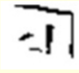
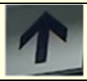


Fig. 12. Fluctuation in detailed recognition rate in accordance with each sign.

Table 2. Examples of misclassified sign candidates.

Original image	Binarized image	Recognition result	Correct recognition result	Cause of misclassification
		Noise	Number 2	Reflection
		Number 2	Number 1	Reflection
		Number 6	Number 5	Motion blur
		Number 7	Number 1	Pose changes/ motion blur
		Noise	Arrow down-left	Illumination
		Noise	Arrow up	Shadow

which can be diminished. In addition, Fig. 12 shows the detailed recognition rate fluctuation in accordance with each sign.

Finally, we show misclassified examples, which can occur for several different reasons, including changes in illumination, shadow, motion blur, and changes in pose (see Table 2).

V. Conclusion

We described a walking guidance system that can detect and recognize sixteen classes of exit numbers and arrow signs. Our system may be useful for assisting the visually impaired when finding their way through a subway station. The main contributions of this paper are divided into two parts. First, we exploited three verification methods: (a) HMM log-likelihood, (b) the Hough Transform, and (c) color information in Section IV-2. This significantly decreased the false-positive rate and hence, increased the reliability of the system. Second, we suggested a detection method for small signs. As described in Section III, our detector for small signs extracted an additional 1,057 candidate regions, including three noises that were not overlapped with other regions, detected by the MCT-AdaBoost detector. Additionally, the recognition result for small signs was 99.5%. These results indicate the robustness of our detector for small and geometrically distorted signs.

However, some improvements can be considered in future work. First, it is necessary to add more classes, such as toilets and elevators, as well as ticket-office pictograms in subway stations, allowing the proposed system to provide more subway-sign information. Second, the system must work in real time. The processing time of our system is currently about 7 fps to 15 fps, which is dependent on how many sign candidates are extracted from natural scenes. An improvement in the processing time will also, therefore, be made.

References

[1] D. Gerónimo et al., "Traffic Sign Recognition for Computer Vision Project-Based Learning," *IEEE Trans. Educ.*, vol. 25, no. 3, Aug. 2013, pp. 364–371.

[2] H.S. Park et al., "In-Vehicle AR-HUD System to Provide Driving-Safety Information," *ETRI J.*, vol. 35, no. 6, Dec. 2013, pp. 1038–1047.

[3] S. M¨uller-Schneiders, C. Nunny, and M. Meuter, "Performance Evaluation of a Real-Time Traffic Sign Recognition System," *IEEE Intell. Veh. Symp.*, Eindhoven, Netherlands, June 4–6, 2008, pp. 79–84.

[4] J. Lee et al., "AdaBoost for Text Detection in Natural Scene," *Int. Conf. Document Anal. Recogn.*, Beijing, China, Sept. 18–21, 2011, pp. 429–434.

[5] K. Choi et al., "State Machine and Downhill Simplex Approach for Vision-Based Nighttime Vehicle Detection," *ETRI J.*, vol. 36, no. 3, June 2014, pp. 439–449.

[6] S.M. Hanif and L. Prevost, "Text Detection and Localization in Complex Scene Images Using Constrained AdaBoost Algorithm," *Int. Conf. Document Anal. Recogn.*, Barcelona, Spain, July 26–29, 2009, pp. 1–5.

[7] C. Bahlmann et al., "A System for Traffic Sign Detection, Tracking, and Recognition Using Color, Shape, and Motion Information," *IEEE Intell. Veh. Symp.*, Las Vegas, NV, USA, June 6–8, 2005, pp. 255–260.

[8] B. Froba and A. Ernst, "Face Detection with the Modified Census Transform," *IEEE Conf. Autom. Face Gesture Recogn.*, Seoul, Rep. of Korea, May 17–19, 2004, pp. 91–96.

[9] K-D. Ban et al., "Tiny and Blurred Face Alignment for Long Distance Face Recognition," *ETRI J.*, vol. 33, no. 2, Apr. 2011, pp. 251–258.

[10] J. Sauvola and M. Pietikainen, "Adaptive Document Image Binarization," *Pattern Recogn.*, vol. 33, no. 1, 2000, pp. 149–160.

[11] Y. Yoon et al., "Best Combination of Binarization Methods for License Plate Character Segmentation," *ETRI J.*, vol. 35, no. 3, June 2013, pp. 491–500.

[12] D. Lee and H. Yoon, "Sign Recognition with HMM/SVM Hybrid for the Visually-Handicapped in Subway Stations," *Int. Joint Conf. Comput. Intell.*, Barcelona, Spain, Oct. 5–9, 2012, pp. 631–634.

[13] C-L. Liu, "*Handwritten Chinese Character Recognition: Effects of Shape Normalization and Feature Extraction*," *Lecture Notes Comput. Sci.*, vol. 4768, 2008, pp. 104–128.

[14] C.C. Chang and C.J Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 27, 2011, pp. 2–27.

[15] A. Schlapbach and H. Bunke, "Off-line Handwriting Identification Using HMM Based Recognizers," *Int. Conf. Pattern Recogn.*, Cambridge, UK, vol. 2, Aug. 23–26, 2004, pp. 654–658.

[16] T.Y. Zhang and C.Y. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns," *Commun. ACM*, vol. 27, no. 3, Mar. 1984, pp. 236–239.

[17] M. Ahmed and R. Ward, "A Rotation Invariant Rule-Based Thinning Algorithm for Character Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, Dec. 2002, pp. 1672–1678.

[18] J. Kim and H. Yoon, "Graph Matching Method for Character Recognition in Natural Scene Images: A Study of Character Recognition in Natural Scene image Considering Visual Impairments," *IEEE Int. Conf. Intell. Eng. Syst.*, Poprad, Slovakia, June 23–25, 2011, pp. 347–350.

[19] B.L. Van, S. Garcia-Salicetti, and B. Dorizzi, *Fusion of HMM's*

http://dx.doi.org/10.4218/etrij.14.2214.0007

Likelihood and Viterbi Path for On-line Signature Verification, Springer Berlin Heidelberg, vol. 3087, 2004, pp. 318–331.

[20] J. Illingworth and J. Kittler, "The Adaptive Hough Transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, Sept. 1987, pp. 690–698.

[21] Y. Xie and Q. Ji, "A New Efficient Ellipse Detection Method," *Int. Conf. Pattern Recogn.*, Quebec, Canada, Aug. 11–15, 2002, pp. 957–960.

[22] P. Perez et al., "Color-Based Probabilistic Tracking," *European Conf. Comput. Vis.*, Copenhagen, Denmark, May 28–31, 2002, pp. 661–675.

[23] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," *Comput. Vis. Pattern Recogn.*, Hilton Head, SC, USA, vol. 2, June 13–15, 2000, pp. 142–149.

[24] T. Wu, C. Lin, and R.C. Weng, "Probability Estimates for Multiclass Classification by Pairwise Coupling," *J. Mach. Learning Res.*, vol. 52, Dec. 2004, pp. 975–1005.

Dongjin Lee received his MS degree in computer software engineering (image processing) from the University of Science and Technology, Daejeon, Rep. of Korea, in 2013. From 2008 to 2010, he was an engineer at the Samsung S1 Coporation, Seoul, Rep. of Korea. He has been a research scientist at the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, since 2013. He is working toward his PhD degree in electronics at Chungnam National University, Daejeon, Rep. of Korea. His research interests are in computer vision, machine learning, and affective computing.

Hosub Yoon received his PhD degree in computer science (image processing) from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 2003. In 1991, he was with the System Engineering Research Institute, Korea Institute of Science and Technology, Seoul, Rep. of Korea. Since 1998, he has been a research scientist at the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea. His research interests include HRI, image processing, pattern matching, and speech processing.

Myung-Ae Chung received her BS and MS degrees in chemistry from Ewha Womans University, Seoul, Rep. of Korea, in 1986 and 1988, respectively. She moved to Germany during her PhD course at Ewha Womans University and received her PhD degree in physical chemistry from Clausthal Technical University, Lower Saxony, Germany, in 1997. From 1998 to 1999, she was a member of the research staff at the Max-Planck Institute for Polymer Research, Mainz, Germany. She joined the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, in 2000 and is engaging in research projects on IT convergence technologies, such as in nano-bio materials and in cognition systems.

Jaehong Kim received his PhD degree in computer engineering from Kyungpook National University, Daegu, Rep. of Korea, in 2006. He has been a research scientist at the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, since 2001. His research interests include socially assistive robotics for the elderly, human-robot interaction, and gesture/activity recognition.