Traffic signs and signals detection employing the MY VISION system for a visually impaired person

doi: http://dx.doi.org/10.24507/icicelb.07.02.385

ICIC Express Letters
Part B: Applications
Volume **7**, Number **2**, February **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 385–391

TRAFFIC SIGNS AND SIGNALS DETECTION EMPLOYING THE MY VISION SYSTEM FOR A VISUALLY IMPAIRED PERSON

Takaki Kumano[1], Joo Kooi Tan[2], Hyougnseop Kim[2] and Seiji Ishikawa[2]

[1]Graduate School of Engineering
[2]Faculty of Engineering
Kyushu Institute of Technology
Sensuicho 1-1, Tobata, Kitakyushu, Fukuoka 804-8550, Japan
kumano@ss10.cntl.kyutech.ac.jp; etheltan@cntl.kyutech.ac.jp

Received July 2015; accepted September 2015

Abstract. *In recent years, the equipment on walk support for visually impaired persons has spread to a certain extent. A studded paving block and a sound-type signal are set on the sidewalk or on the road in many places. However, there are still problems such that the studded paving block is still installed at limited places, or the sound of the signal is obscure because of the environmental noise such as roaring traffic or heavy rain. Therefore, a system that can support visually impaired persons in more effective way is necessary, such as a system which finds a signal or a traffic sign automatically and provides its information to the visually impaired. In order to realize such a system, this paper proposes a method of detecting pedestrian signals and crosswalk signs by a camera and a computer which we call MY VISION system. In the proposed method, color information is used in the first step to restrict the search for traffic signal and sign candidates, and then the HOG feature is introduced to describe the candidates by feature vectors. Recognition of the signals and the signs is performed by applying randomized trees to the candidates. The color sign of the pedestrian signal is also discriminated by using color information. Experimental results are shown and the method is evaluated.*
Keywords: MY VISION, HOG, Randomized trees, HSV model, Computer vision

1. **Introduction.** The number of Japanese visually impaired persons is about 310,000 now [1]. They make the judgment on the color sign of a traffic signal at a crosswalk by auditory information, such as vehicle sound running at the crosswalk. However, the judgment by auditory information is sometimes not safe, if environmental noise is large or vehicles do not pass through. In recent years, sound-type signals have become widespread, but there is a problem with the number of installation. If there is no traffic light in a crosswalk, it is necessary to determine the presence or absence of a pedestrian crossing by studded paving block. However, there is a problem of the installation number and the difficulty in recognizing. Therefore, there is a need for a system by which visually impaired persons can obtain the signal or traffic signs information.

To detect and recognize traffic signals and road signs, studies using in-vehicle vision have been carried out. There are methods of traffic signals detection using the shape and color information of the traffic signal light [2,3]. In addition, there are methods of road signs detection by feature points matching and a classifier [4,5]. However, these studies are intended for safe support of a vehicle driver. Obviously, vehicle drivers pay attention to the traffic signals for vehicles, whereas pedestrians rather watch pedestrians' traffic signals. To the best of our knowledge, studies on the detection of pedestrians' signals and crosswalk signs for the purpose of safe support of pedestrians have not yet been performed. It is vital to visually impaired pedestrians, or even ordinary pedestrians inattentive to surroundings, to provide with the information on the location and the indicating color sign (red or green) of a pedestrian's signal for their safety.

This paper proposes a method of detecting pedestrian signals and traffic signs at a crosswalk or on a sidewalk by MY VISION system. MY VISION is a self virtual viewpoint produced by a self-wearable camera-computer system. A self virtual viewpoint means the second viewpoint of a person who employs this system. The user may include a visually impaired person and an ordinary person inattentive to his/her front. In the MY VISION system, a camera is mounted on the head of a user, and pedestrians signals and crosswalk signs are detected by analyzing the video provided from the camera.

The proposed method consists of two stages: learning and identification. We use the HOG features in learning to build a decision tree employing a randomized tree which is a multi-class classifier. In identification, after extracting the HOG feature vector from an object candidate region detected by HSV values from the input image, the feature vector is identified by the randomized trees, by which the signals and the signs are detected. The indicating color sign of the signal is also identified.

2. **Extraction of Feature.** This section describes the HOG feature [6] to be used in the proposed method. First, an image is separated into cell areas, and gradient magnitudes and gradient directions are computed with each pixel. Second, the gradient direction is separated into several orientations to create a histogram with each cell by adding the gradient magnitude with respective bins. In the proposed method, precise angle is not necessary, because the rigid body such as traffic signals and signs is the target. Therefore, the gradient direction is separated as shown in Figure 1 and the number of the bin is 4. Horizontal and vertical edges are easily captured by performing the division of the gradient direction such as shown in Figure 1. Then 4-dimensional feature vector $\boldsymbol{a}_{ij} = (a_1, a_2, a_3, a_4)_{ij}$ is obtained from the cell located at (i, j).

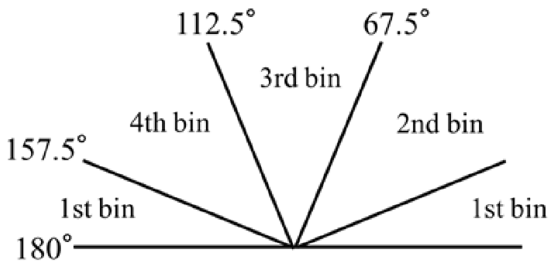

Figure 1. Division of the gradient direction

3. **Detecting Object Candidates.** This section describes the method of detecting object candidates region from an input image. An input image is expressed by HSV color model, and red pixels, green pixels of the signal, and green pixels of the traffic sign are extracted. The range of the investigation in the search window is assumed to be only the upper half of the input image. The candidate region is detected by considering the ratio of particular color pixels to other pixels and detecting the center of gravity of the color pixels. Each technique is explained in the following.

3.1. **Detection of a candidate region by the ratio of color pixels.** A search window is scanned as shown in Figure 2, and the pixels having particular color in the window are counted. Red pixels are counted in the upper half region of the search window and green pixels are counted in the lower half region of the search window. On the other hand, blue pixels are counted in the whole region of the window. A pedestrian signal and a crosswalk sign are discriminated by the above way of counting color pixels. Let us denote the entire number of the pixels in a search window by N_W, the number of red pixels by N_R, the number of green pixels by N_G, and the number of blue pixels by N_B. Then, if either of the

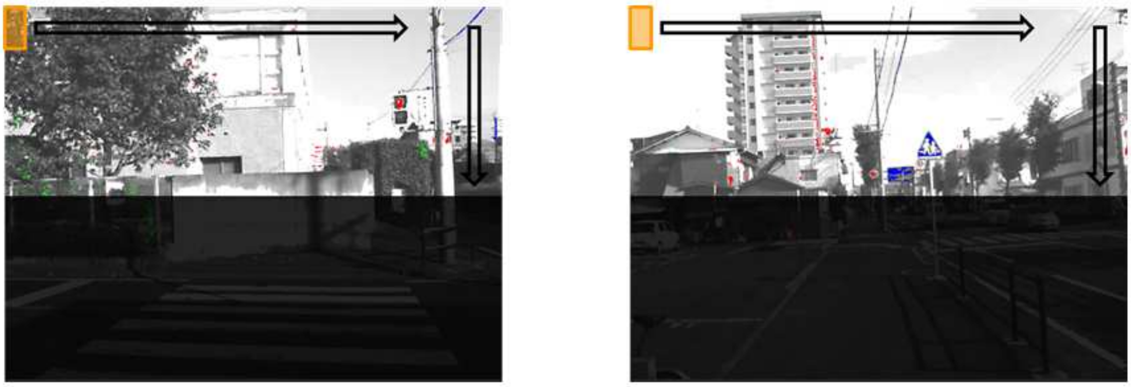

Figure 2. Scanning method of the search window

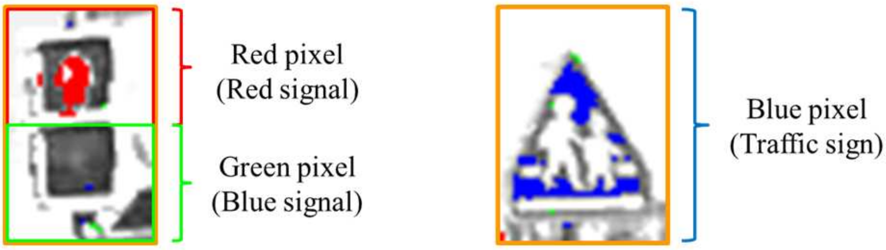

Figure 3. Range of counting pixels

following inequalities holds, the search window area is identified as an object candidate region.

$$\begin{cases} N_R > \alpha N_W \\ N_G > \beta N_W \\ N_B > \gamma N_W \end{cases} \tag{1}$$

Here $0 < \alpha, \beta, \gamma < 1$. In Equation (1), the first two inequalities are for detecting a pedestrian signal, whereas the last inequality is for a crosswalk sign. In the performed experiment, $\alpha = \beta = 0.02$ and $\gamma = 0.06$.

Since the size of detected objects changes in images, the width of the search window is varied from 20 to 60 pixels with every 5 pixels, whereas the height of the window is 1.5 times of the width.

3.2. **Detection of a candidate region by the center of gravity of color pixels.** Once an object candidate region is detected by the procedure stated in Subsection 3.1, the center of gravity of each set of color pixels in the window is calculated. A rectangular area is defined in a window as shown in Figure 4 and, if the center of gravity exists within the rectangular area, the window is assumed to include a candidate region of the object. In this way, an object candidate region is detected through two stages, 3.1 and 3.2.

3.3. **Identification.** The search window obtained from the procedure described in Subsections 3.1 and 3.2 is an object candidate region, and it is identified by the randomized trees [7]. Then, the number of red and blue pixels of a signal in the region is compared and the color sign of the signal is recognized.

4. **Experiment.** In this section, we describe the experiment on the detection of pedestrian signals and traffic signs, in particular, crosswalk signs.

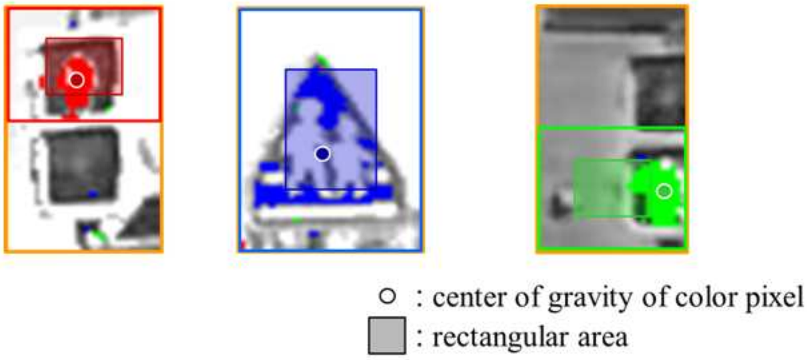

FIGURE 4. The center of gravity of color pixels and a rectangular area

4.1. **Experimental setup.** In this experiment, images shown in Figure 5 are used for learning. The pedestrian signals and the crosswalk signs such as in Figures 5(a) and 5(b) are positive samples and the background images shown in Figure 5(c) are negative samples. The number of each positive samples is 1,000, and the number of negative samples is 2,000. These sample images are normalized into 30×45 pixels' images on which the HOG feature is calculated. A sample image has 28 blocks (4 blocks to the horizontal direction and 7 blocks to the vertical direction), a block has 3×3 cells and a cell has 4 bins. So the dimension of the used HOG feature is 1008 ($= 28 \times 9 \times 4$). Some learning parameters of the employed randomized trees are given in Table 1.

The test videos used in the experiment are obtained outdoors using the MY VISION system. Namely, a camera is mounted on the right ear of a user (a pedestrian) and takes a video of a frontal scene of the user (See Figure 6).

(a) Pedestrian signals (b) Crosswalk signs (c) Background

FIGURE 5. Examples of learning images

TABLE 1. Learning parameters of randomized trees

Parameters	Value
The number of decision trees	30
The maximum depth of the decision tree	5
The number of selected features	32
The number of selected thresholds	5
The ratio of the number of samples to all the image samples	30%

Figure 6. Scenery of taking a video by an ego camera

4.2. **Experiment on the detection of pedestrian signals and crosswalk signs.** A user equipped with the MY VISION system walks a little slowly on a sidewalk, during which frontal scenes of the user are taken a video (See Figure 6). The video is processed to detect pedestrian signals and their indicating color signs. Crosswalk signs are also detected on the side of a sidewalk. Some results are shown in Figure 7, where detected red signals are enclosed by a red rectangle frame, blue signals are enclosed by a green frame and crosswalk signs are enclosed by a yellow frame. In the upper row of Figure 7, a user is standing on a sidewalk, whereas the user is walking in the lower row of Figure 7 and in Figure 8.

Figure 7. Experimental results on a pedestrian signal detection

Figure 8. Experimental results on a crosswalk sign detection

4.3. **Evaluation.** In order to evaluate the proposed method, following indices are defined:

N_A: The number of all processed frames;

N_D: The number of the frames in which pedestrian signals and crosswalk signs (referred to as interested objects hereafter) were correctly detected;

N_F: The number of the frames in which wrong objects were detected;

N_U: The number of the frames in which nothing was detected on account that the identification process using the randomized trees was not active as the pedestrian signal just turned off the light;

N_N: The number of the frames in which interested objects were not detected.

Note that N_D, N_F, N_N correspond to true positive, false positive and false negative, respectively. Since the purpose of the present experiment is to evaluate the performance of the proposed detection algorithm, test videos are taken so that they include a single pedestrian signal or a single crosswalk sign in every frame. Hence, there are no frames in the test videos corresponding to true negative.

The detection rate R is defined by

$$R = \frac{N_D}{N_A - N_U} \times 100 \tag{2}$$

Note that $N_A - N_U = N_D + N_F + N_N$ holds. As stated above, N_U is the number of the frames from which an interested object was not detected, though it exists, because no color information was available. These frames are out of those frames to which the proposed detection method can be applied. Hence, they are excluded from N_A in Equation (2).

The results on the evaluation with respect to a pedestrian signal detection and a crosswalk sign detection are shown in Table 2 and Table 3, respectively. They provide the values of the indices explained above and the processing time. It is noted that index N_U is not given in Table 3, since a crosswalk sign does not use a light. With the detection of a pedestrian signal, the average detection rate was 84.2% and the average processing time was 1.07 sec/frame. This average detection rate is, instead of taking the average of R, calculated by Equation (2), in which N_A, N_U and N_D are the sum of the frames with respect to the three scenes. This is because the number of frames N_A is different with each scene. On the other hand, the average detection rate of a crosswalk sign was 93.2% and the processing time was 1.26 sec/frame in average.

Table 2. Evaluation on a pedestrian signal detection

Scene	N_A [frame]	N_D [frame]	N_F [frame]	N_U [frame]	N_N [frame]	R [%]	Processing time [sec/frame]
1	350	172	2	162	14	91.5	1.31
2	480	214	0	193	73	74.6	0.83
3	200	94	0	105	1	98.9	1.08
					Average	84.2*	1.04

* The calculation of this average is explained in 4.3.

Table 3. Evaluation on a crosswalk sign detection

Scene	N_A [frame]	N_D [frame]	N_F [frame]	N_N [frame]	R [%]	Processing time [sec/frame]
4	270	270	0	0	100	1.33
5	270	233	0	37	86.3	1.19
				Average	93.2	1.26

5. **Conclusions.** The purpose of this study is to develop an ego vision system, named MY VISION system, which gives hand to a visually impaired person as the second virtual eye by providing various visual information to him/her. In this particular paper, a method was proposed of detecting pedestrian signals and crosswalk signs by the system. In the system, a camera was mounted on the ear of a user (a pedestrian) and it captured a video of the frontal scene of the user. By analyzing the video by the proposed method, those signals and signs were detected. For the detection, an HSV image expression, HOG features and a randomized trees classifier were employed. As a result, the average detection rate of the signal was 84.2 [%] and the average detection rate of the crosswalk sign was 93.2 [%]. In this way, as the first stage of the present research, satisfactory results were obtained.

In the use of the HOG feature, processing time and detection accuracy have been improved by reducing the separation of the gradient directions into four.

Future work includes raising the detection rate of interested objects further by the improvement of their search method and reduction of processing time. Since the proposed method uses an ego camera, the captured video may be annoyed with vibration caused by the walk of a user. Compensation of the vibration on the video also remains for future study.

Acknowledgment. This work was supported by JSPS KAKENHI Grant No. 25350477, which is greatly acknowledged.

REFERENCES

[1] Ministry of Health, *Labour and Welfare: The Investigation on Physically Impaired Children and Adults in 2006*, pp.3-16, 2008.
[2] H. Nakada, Recognition of traffic signal from on-board captured image stream, *IEICE ITS*, vol.107, no.486, pp.121-125, 2008.
[3] T. Umezawa and K. Eguchi, A study on recognition of traffic lights from video image in night-time, *Research Paper of Aichi Institute of Technology*, no.43, pp.41-47, 2008.
[4] K. Doman, Construction of a traffic sign detector using generative learning considering color variation, *IEICE*, vol.J93-D, no.8, pp.1375-1385, 2010.
[5] A. Ihara, Road sign detection by PCA-SIFT in different feature subspaces, *The 14th Image Sensing Symposium SSII08*, 2008.
[6] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Proc. of the IEEE Conference on CVPR*, pp.886-893, 2005.
[7] L. Breiman, Random forests, *Machine Learning*, vol.45, no.1, pp.5-32, 2001.