Walk Environment Analysis Using MY VISION: Toward a Navigation System Providing Visual Assistance

doi: info:doi/10.24507/ijicic.15.03.861

International Journal of Innovative
Computing, Information and Control
Volume **15**, Number **3**, June **2019**

ICIC International ©2019 ISSN 1349-4198
pp. 861–871

WALK ENVIRONMENT ANALYSIS USING MY VISION: TOWARD A NAVIGATION SYSTEM PROVIDING VISUAL ASSISTANCE

Joo Kooi Tan[1], Tomoki Ishimine[2] and Shohei Arimasu[2]

[1]Department of Engineering
[2]Graduate School of Engineering
Kyushu Institute of Technology
1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan
etheltan@cntl.kyutech.ac.jp

Received July 2018; revised November 2018

Abstract. *This paper proposes a method of analyzing a human walk environment using MY VISION. MY VISION is an ego-camera and a computer system which analyzes a video obtained from the ego-camera to acquire certain visual information useful for human daily activities. The system is expected to be a virtual eye of a visually impaired person or the third eye of a pedestrian absorbed in a mobile phone. The proposed method keeps in a database the background images of key points along a sidewalk and judges if a MY VISION user is walking along the sidewalk or if he/she has come to a crosswalk by referring to the backgrounds. If the former, the method finds a safe road region on the sidewalk, whereas, if the latter, it searches for the crosswalk for finding an appropriate walk direction and a traffic light to know the proper timing to cross it. Experimental results show effectiveness of the proposed method.*
Keywords: Safe walk, Visually impaired person, Ego-camera, MY VISION, Background, Bag of features, Graph based segmentation, Computer vision

1. **Introduction.** Outdoor walk is not safe for visually impaired people. It is necessary, and probably expected, to develop a certain kind of navigation system that supports their safe walk. Although white canes, guide dogs and studded paving blocks are of some help, they will not be enough for impaired people to live a safe life. This paper proposes a walk environment analysis method, employing a computer vision system, which collects certain kinds of scenery information and detects a sidewalk, a crosswalk, a traffic light, etc., necessary for their safe outdoor walk.

MY VISION [1-4] is a developed virtual eye and a brain system composed of an ego-camera mounted on a human body and a computer. It is the abbreviation of 'a Magic eYe of a Visually Impaired for Safety and Independent actiON'. It analyzes a video provided from the ego-camera by computer in order to acquire some useful information for, mainly, outdoor walk. It assumes to be a virtual eye of a visually impaired person or the third eye of a pedestrian absorbed in a mobile phone.

Similar systems exist until now. Kanade and Hebert [5,6] proposed the first-person vision employing a pair of an out-looking camera and an eye-sight detection camera, by which the intention on the activities of a user, an ordinary person, is inferred. However, the system does not take visually impaired people into account. Kitani et al. [7] used a video captured by a head-mounted camera of a user to classify his/her action. Their concern is, however, ego-actions classification from an ego-camera video. Le et al. [8]

DOI: 10.24507/ijicic.15.03.861

used probability density function of a road color and a lane color on the road to extract pedestrian lanes for assistive navigation of the visually impaired. They only extract the lanes at traffic junctions, though. Treuillet et al. [9] proposed a body mounted vision system. It guides a visually impaired user walk along a learned path using a set of key points detected in successive image frames. However, their main concern is walking an intended road as precise as possible in comparison with a system using a GPS. Some Tongue Display Unit (TDU) prototypes [10-12] have been proposed for those visually impaired to feel four to eight directions/angles. This tool may be difficult in wide-spread use, since it must be kept in the mouth, though technically interesting.

All the above proposals have different purposes or insufficient performance for practical use. The proposed system focuses on the development of a method which guides a visually impaired person or a mobile phone user in outdoor walk by use of the MY VISION system. As is well known, there are many car vision systems nowadays to realize safe driving. The present system is its human version, i.e., a human vision system for safe walk.

Several MY VISION systems [1,2] have been studied which analyze a walk environment. However, these studies detect only a single class of objects separately such as a road [13], a traffic light [1], a pedestrian [2], and a public bus [3,4], which is still insufficient practically.

In this paper, a novel walk environment analysis method is proposed which infers approximate location of a MY VISION user (referred to as a 'user' hereafter) by analyzing the background in a frontal scene and switches the target of analysis according to the location of a user. The method is expected to let a user moves from one place to another by walking sidewalks and crossing crosswalks, if necessary, paying attention to traffic lights. The proposed method differs from the existent methods in that it is an integrated safe walk assistance system based on switching the target of analysis. It aims at realization of safe walk of not only a visually impaired person, but also those pedestrians absorbed in a mobile phone, which was not taken into account until it has gained recent large popularity. The method is described in Section 2, some experimental results are shown in Section 3, discussion is given in Section 4 and the paper is concluded in Section 5.

2. **Proposed Method.**

2.1. **Overview of the proposed method.** The proposed method extracts some features from the background and infers the situation (location and intension) of a user. Initially the method creates Background Models (BMs) of an area of interest. The background scene observed from the ego-camera changes depending on which road and which location on the road a user walks. Since the background changes gradually as the user walks on, Representative Background Models (RBMs) with every certain distance, or at certain key points, are stored in a database and referred to as a Background Model Database (BMD).

The model is then used to infer a user's walk situation. If the background in the image fed by the ego-camera is identified as a certain RBM in the BMD, the system knows the road and the location where the user is and the walk direction of the user. It then guides the user by informing which direction to walk. In order to escape from some obstacles such as rubbish bins or left bicycles, it detects safe part of the sidewalk the user is on.

On the other hand, if the background in the fed image is identified as that near a crosswalk in the BMD, the system tells the user the existence of the crosswalk. If the user intends to cross the road there, the system finds the zebra pattern and the traffic light to tell the user exact direction of the crosswalk and exact timing to cross it.

The above procedure is repeated until the user finally reaches his/her destination. The flow of the entire procedure is shown in Figure 1.

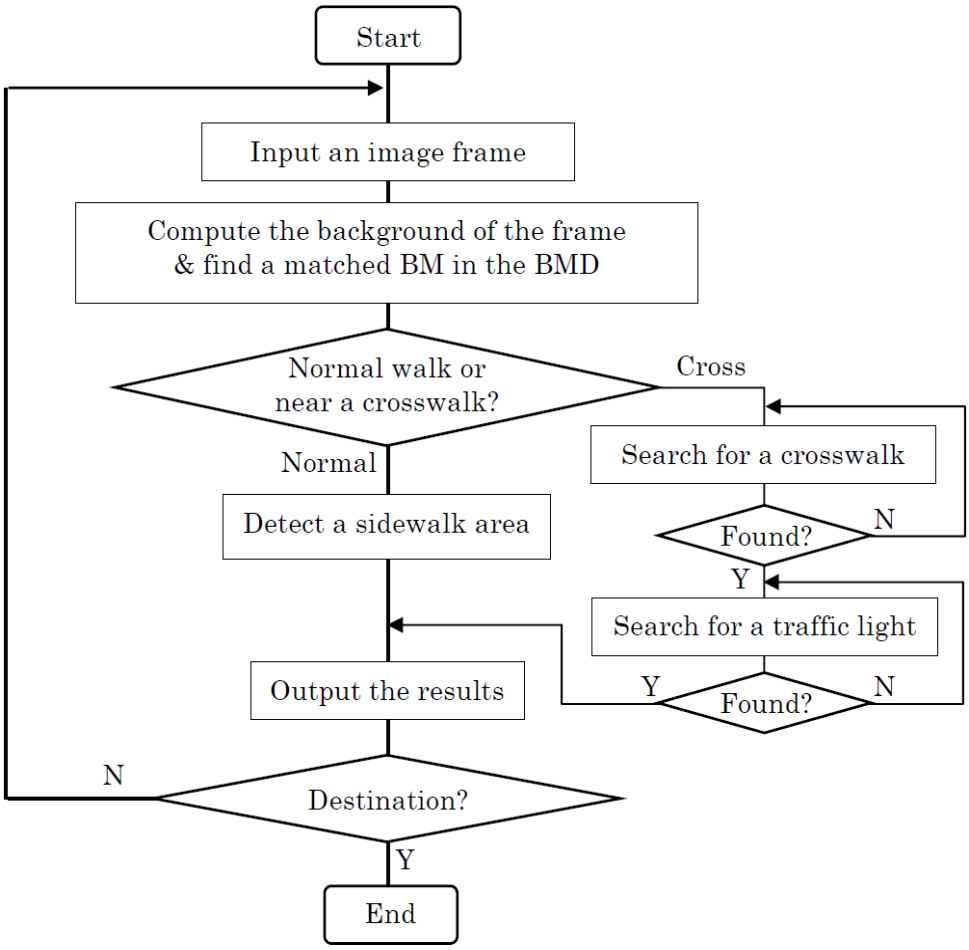


FIGURE 1. Flowchart of the procedure

2.2. **Database creation.** The BM in the proposed method is described by a Bag of Features (BoF) vector [14] based on AKAZE [15]. AKAZE is a powerful feature point extractor invariant to scale change, rotation, illumination change and image blur. These facts are advantageous particularly in applying it to the outdoor images provided from an ego-camera, since they receive disturbance in camera motion from a user or illumination change frequently. The background contains various objects and they may have many strong feature points, which can be well described by the BoF.

BoF vectors are calculated from the frames in a walk video and stored into a feature space. They are clustered in the space by the k-means method, resulting in k representative background classes. Each cluster center is saved as an RBM in a BMD and given labels such as 'normal walk' or 'near crosswalk' along with the road number and the location on the road.

2.3. **Inferring walk situation.** The images from an ego-camera are analyzed to find which RBM the present image frame corresponds to. This is found employing the k-nearest neighbor algorithm. The walk situation, or a user's location and his/her intension, is finally inferred by the label attached to the RBM. The system infers the situation; e.g., the user is walking on the sidewalk straight, or there is a crosswalk near the user and he/she might cross the road there.

2.4. **Extracting a sidewalk area.** When walking a sidewalk, obstacles such as bicycles or left objects must be found and safe walk areas on the sidewalk should be informed to a

user. For this purpose, the input image is segmented first by Graph Based Segmentation (GBS) [16] to the areas having an identical color. The above obstacles are normally segmented separately, as they have colors different from the road.

A small artificial area containing a user's foot is initially defined as shown in Figure 2(a). The area including this artificial area among those provided by the GBS (whose result is shown in Figure 2(b)) is then extracted as given in Figure 2(c). This area is referred to as 'a foot area'. Then it is combined with the area connected to the foot area and having the most similar gray values. The combined area is referred to as 'a candidate sidewalk area'. It is depicted in Figure 2(d) and its corresponding areas in the GBS image in Figure 2(e).

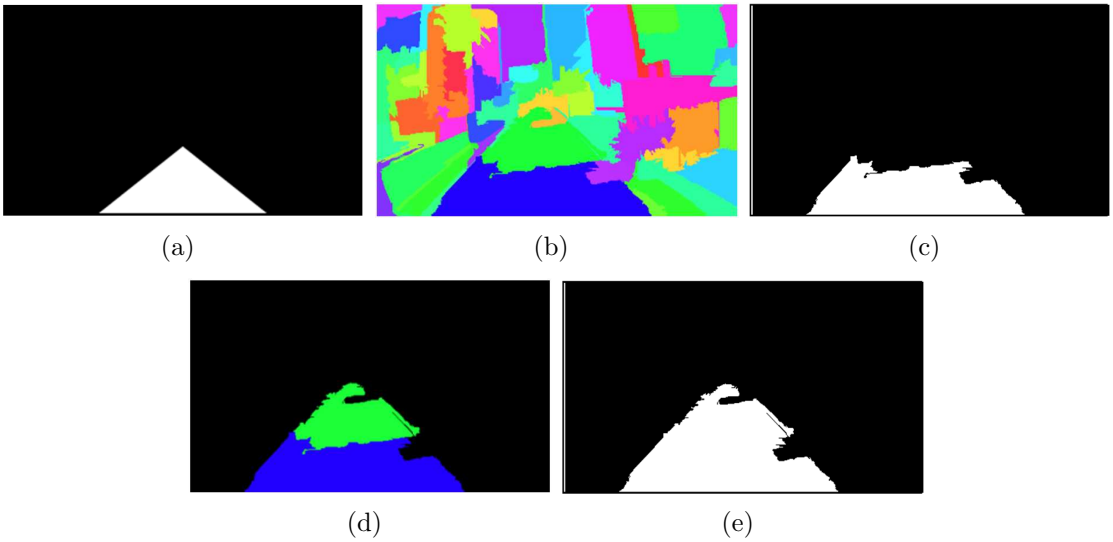


FIGURE 2. Extraction of a sidewalk area: (a) artificial foot area, (b) result of the GBS, (c) a foot area, (d) a candidate sidewalk area, (e) the areas in (b) corresponding to (d)

Let the gray value of a pixel (x, y) contained in an area A_p $(p = 1, 2, \ldots, P)$ adjacent to the foot area be denoted by $f_p(x, y)$ and that of the foot area A_f by $f_f(x, y)$. Then the area A_{p^*} satisfying the following equation is combined to the foot area to yield a sidewalk candidate area.

$$p^* \equiv \arg\min_p \left| \frac{1}{|A_p|} \sum_{(x,y)\in A_p} f_p(x, y) - \frac{1}{|A_f|} \sum_{(x,y)\in A_f} f_f(x, y) \right| \tag{1}$$

Here $|A_*|$ stands for the number of the elements in a set A_*. The set A_p must satisfy $|A_p| > T_1$ (> 0) to exclude trivial areas. T_1 is a threshold experimentally defined.

The above procedure assumes that a user is initially on the sidewalk. Equation (1) finds the area whose average gray value is the most similar to the foot area among all the areas connected to the foot area.

The sidewalk area is finally determined as the area containing those pixels (in the candidate sidewalk area) which belong to the sidewalk areas for 5 frames or more among the past 10 frames. If this is not satisfied, the sidewalk area of the last frame is chosen as the present sidewalk area.

2.5. **Detection of a crosswalk.** Suppose that the analysis of the video provided from the ego-camera of a user has found the background whose label is 'near crosswalk' and the

user wants to cross the road there. Then the system must give an appropriate instruction so that the user may stand in the right direction before crossing. Though the background image when a user is facing a crosswalk is memorized in BMD, the zebra pattern of a crosswalk is searched when a user has come close to a crosswalk in order to reach quickly to the position where the background image seen from the user's camera best matches the one in the BMD.

Two clues are taken into account as the features of a crosswalk: (i) high average luminance value, and (ii) high variance with the luminance value in the vertical direction and low in the horizontal direction.

A rectangular search window is scanned in the lower part of an image frame, and the window locations satisfying the above two clues are detected.

Let the average luminance value in a scanned window be denoted by l_{av}, and the vertical and the horizontal variance of the luminance be denoted by $\sigma_{\mathrm{v}}^2(\theta)$ and $\sigma_{\mathrm{h}}^2(\theta)$, respectively. Here the angle of a user's turn is denoted by θ $(0 \leq \theta < \pi)$. Then, obviously, the right direction for crossing the crosswalk is given by the direction when the ratio of $\sigma_{\mathrm{h}}^2(\theta)$ to $\sigma_{\mathrm{v}}^2(\theta)$ is the minimum under the change of θ on condition that $l_{\mathrm{av}} > T_2$ (> 0) (T_2 is an experimentally defined threshold). This is formulated as

$$\theta^* \equiv \arg\min_{\theta} \frac{\sigma_{\mathrm{h}}^2(\theta)}{\sigma_{\mathrm{v}}^2(\theta)} \tag{2}$$

Angle θ^* gives the right direction for the user to turn.

2.6. **Detection of a traffic light.** A traffic light is detected using Co-HOG (Co-occurrence Histograms of Oriented Gradients) feature [1] and random forest [17]. (Here the method focuses its attention on a pedestrian's traffic light.) First, a traffic light discriminator is designed by machine learning using traffic light images and the images other than a traffic light. Next, a search region is set on the upper part of an analyzed image and a traffic light is searched by scanning a window in the region and by applying the discriminator to the window.

2.6.1. *Co-HOG feature.* The Co-HOG feature is defined in the following way. The image in the scanning window is separated into cells each having 5×5 pixels. The size of the image is the one such that it contains 8×6 cells which do not overlap each other (See Figure 3). Considering that a traffic light for a pedestrian is vertically long rectangle and it is mirror symmetric w.r.t. the upper and the lower part, a cell $\boldsymbol{a}_{ij} = (a_1, a_2, \ldots, a_9)_{ij}$

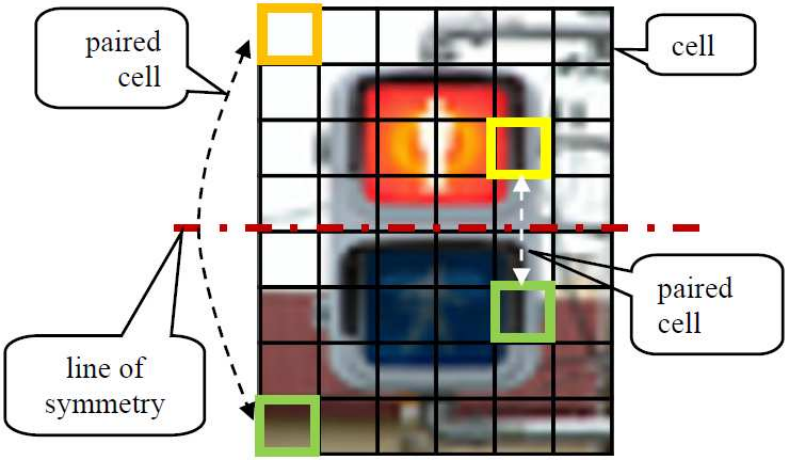


FIGURE 3. Image of a pedestrian's traffic light in a scanning window

(row: $i = 1, 2, 3, 4$; column: $j = 1, 2, \ldots, 6$) in the upper part and its corresponding (mirror symmetric) cell $\boldsymbol{b}_{9\text{-}i,j} = (b_1, b_2, \ldots, b_9)_{9\text{-}i,j}$ in the lower part are paired as shown in Figure 3 to make a 81-dimensional single feature vector $\boldsymbol{c}_{ij}$ defined by

$$\begin{aligned}\boldsymbol{c}_{ij} &= (a_1 + b_1, a_2 + b_1, \ldots, a_9 + b_1, a_1 + b_2, a_2 + b_2, \ldots, a_9 + b_2, \ldots, \\ &\quad a_1 + b_9, a_2 + b_9, \ldots, a_9 + b_9)_{ij} \\ &\equiv \boldsymbol{a}_{ij} \oplus \boldsymbol{b}_{ij}\end{aligned} \tag{3}$$

It is noted that a cell $\boldsymbol{a}_{ij}$ (or $\boldsymbol{b}_{9\text{-}i,j}$) above is described by a gradient histogram of 9 bins as in the original HOG feature.

Since the image in the scanning window contains 24 cells, the Co-HOG feature vector $\boldsymbol{c}$ is defined employing Equation (3) by

$$\boldsymbol{c} = (\boldsymbol{c}_{11}, \boldsymbol{c}_{12}, \ldots, \boldsymbol{c}_{16}, \ldots, \boldsymbol{c}_{41}, \boldsymbol{c}_{42}, \ldots, \boldsymbol{c}_{46}) \equiv (\boldsymbol{a}_{11} \oplus \boldsymbol{b}_{11}, \boldsymbol{a}_{12} \oplus \boldsymbol{b}_{12}, \ldots, \boldsymbol{a}_{46} \oplus \boldsymbol{b}_{46}) \tag{4}$$

which is a 1944 ($= 81 \times 24$)-dimensional vector. The Co-HOG feature vector $\boldsymbol{c}$ has an effect of emphasizing the symmetric nature of the traffic light.

2.6.2. *Random forest.* A traffic light discriminator is designed by the employment of random forest. The advantages of random forest include fast training and discrimination. It is also robust to the noise contained in training data. For the training, the Co-HOG feature vectors defined by Equation (4) are employed. The branch function is defined by

$$I_l = \{\boldsymbol{v} \in I | v_i > T_3, i = 1, 2, \ldots, N\} \qquad I_r = I \backslash I_l \tag{5}$$

Here I is a sample set fed to a branch node; I_l and I_r are the sample sets branched to the left and to the right, respectively; N is the dimension of a feature vector $\boldsymbol{v}$ and v_i is the ith component of $\boldsymbol{v}$. The number i and threshold T_3 are determined so that information gain is the minimum by branching at the node.

3. **Experimental Results.** The proposed method was applied to the videos obtained from the MY VISION system set to its user. A camera is mounted to a user as shown in Figure 4(a). Figure 4(b) is an image taken by the camera. The image contains a straight sidewalk in its center and a crosswalk on the far left of the sidewalk. Four videos are taken in this place at different times. A person with the ego-camera walks straight on the sidewalk to the position where the crosswalk is, turns to the left at the crosswalk, and turns to the left again to walk back to the initial position. In this particular experiment, the subject is a 24-year-old normal male graduate student.

(a) (b)

FIGURE 4. Experimental setup: (a) head mounted ego-camera, (b) walk environment used in the experiment

3.1. **Result of walk situation inference.** Three videos out of the 4 videos are employed for building a BMD, whereas the remaining single video is used for test. The RBM in the BMD is given three labels, i.e., 'normal walk', 'near crosswalk' and 'facing a crosswalk'. In the experiment, frames of the test video are examined their labels (i) manually (correct answers) and (ii) by the proposed method (employing the BMD). Let the set of the frames judged as having label k manually and by the proposed method be denoted by H_k and M_k, respectively. Then the success (or coincidence) rate S_k is defined by

$$S_k = \frac{|H_k \cap M_k|}{|H_k|} \tag{6}$$

Here $|\#|$ stands for the number of the elements in set $\#$.

The obtained success rates are given in Table 1 and an example of successful walk situation inference is depicted in Figure 5.

TABLE 1. Result of walk situation inference

k	S_k [%]
Normal walk	99.9
Near a crosswalk	94.9
Facing a crosswalk	94.7

(a) (b)

FIGURE 5. Result of RBM inference: (a) input image, (b) the image most similar to the inferred RBM

3.2. **Result of sidewalk area detection.** The four videos are employed in this experiment. From each video, 100 frames are chosen at regular intervals. Ground truth images of sidewalk areas are made from the frames manually and they are compared to the sidewalk areas detected by the proposed method. The result is evaluated by *Recall* and *Precision* defined by

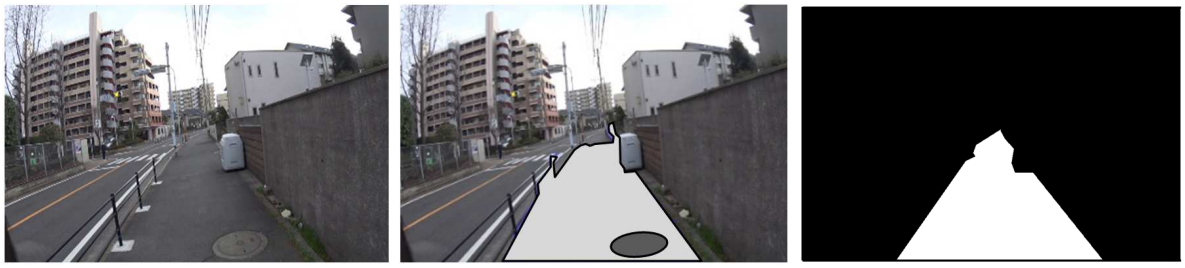
$$Recall = \frac{TP}{TP + FN}, \quad Precision = \frac{TP}{TP + FP} \tag{7}$$

Here TP, FN and FP represent true positive, false negative and false positive, respectively.

The average *Recall* and *Precision* rates over the 100 image frames are shown in Table 2 and an example of the results in Figure 6.

TABLE 2. Result of sidewalk area detection

Evaluation index	Rate [%]
Recall	93.5
Precision	91.2



(a) (b) (c)

FIGURE 6. Result of sidewalk area detection: (a) input image, (b) detected sidewalk area, (c) the ground truth image

3.3. **Result of traffic light detection.** In this experiment, a traffic light discriminator is produced employing 1,000 traffic light images and 2,000 images not containing a traffic light. The performance of the discriminator is compared between the method employing the original HOG features and the proposed Co-HOG features. The evaluation indexes are the recognition rate R_T and the miss recognition rate R_F. They are defined by

$$R_T = \frac{N_{TP}}{N_P}, \quad R_F = \frac{N_{FP}}{N_F} \tag{8}$$

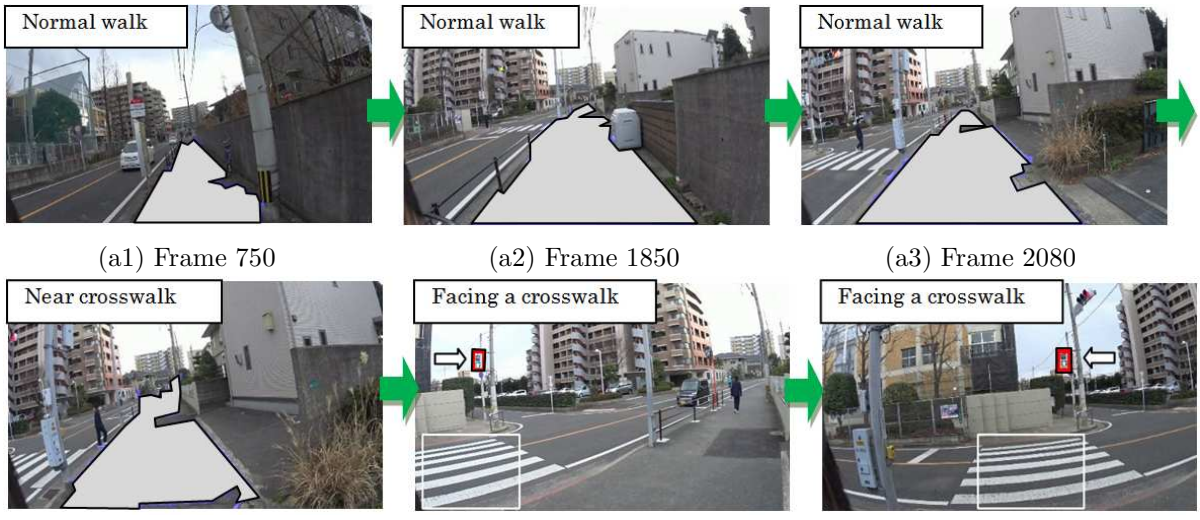
Here N_P is the number of traffic light images; N_{TP} is the number of correctly recognized traffic light images; N_F is the number of images not containing a traffic light; N_{FP} is the number of images incorrectly recognized as containing a traffic light.

The traffic light discriminator was applied to 3,000 images. The experiment was done by the employment of 5-fold cross validation. The results in average are given in Table 3.

TABLE 3. Result of traffic light detection in average

Feature	R_T [%]	R_F [%]
Original HOG	89.4	4.0
Proposed Co-HOG	91.4	3.1

3.4. **Result of navigation.** A demonstration video of the proposed navigation system for visually impaired people is depicted in Figure 7. The video is approximately 2 minutes and 50 seconds. In the video, a MY VISION user walks straight approximately 150 meters on the detected sidewalk which is shown by a white region and the walk situation is 'Normal walk' (a1-a3). Having approached a junction, the system informs the existence of a crosswalk by indicating 'Near crosswalk' (b). At the junction, the system searches and detects the crosswalk and a pedestrian's traffic signal, illustrated by a white rectangle and a black rectangle with an arrow, respectively (c1 and c2). Note that the guidance information shown on the top left of the video is only an experimental display. The user is going to be informed it vocally.



(a1) Frame 750 (a2) Frame 1850 (a3) Frame 2080

(b) Frame 2140 (c1) Frame 2300 (c2) Frame 2430

FIGURE 7. Result of navigation: (a) normal walk, (b) near a crosswalk, (c) facing a crosswalk

A junction without a crosswalk (and a traffic light) is not taken into account in the experiment. The present study assumes and recommends a user to pass a junction equipped with a crosswalk and a traffic light for safety.

4. **Discussion.** This paper proposed a method of inferring a person's, or the system user's, situation, location and intension, in outdoor walk by the employment of an ego-camera video and the background images as a clue. A database containing the background information at key points in an interested area helped to analyze a user's walk environment and infer his/her walk situation.

The BoF based on AKAZE was employed for the background description. Although the images provided from an ego-camera contain rotation, displacement and scale changes caused by the user's movement, these disturbances are well absorbed by BoF and AKAZE. The employment of Co-HOG feature is another successful procedure. It contributes to extracting a pedestrian's traffic light better than the original HOG feature, since the co-occurrence nature employed in the method well describes symmetry of a shape. Some traffic light images were, however, misrecognized on account of the disagreement w.r.t. the center of the traffic light and that of the scanning window. If the position of the traffic light is adjusted so as to be the center of the window, the recognition rate will be further improved.

The proposed method detects sidewalks without using the road lane markers (white lane) extraction employed in [8]. It employs neither Hough transform combined with contour detection [18,19], nor deep neural networks [20], which are commonly used in autonomous vehicles for driving support. The employment of a deep neural network scene parsing [21], object detection using YOLO and its update versions [22-24] may improve the recognition part in this study to a certain extent. It is, however, not adopted because of its black box nature. It is important to analyze and clarify which or what parameters are intrinsic to machine recognition problems.

The obtained information on the user's environment is going to be transferred to the MY VISION user. A vocal transfer is under consideration as a man-machine interface. Another subject to be tackled in the next stage is, and this is of great importance, to detect moving objects such as approaching pedestrians, and bicycles. A detection method

of approaching pedestrians is already proposed [2]. It will be expanded so as to detect bicycles and integrated into the present method. The interface will transfer all of these kinds of information to the user.

The proposed method is intended to be a virtual eye of a visually impaired person. However, it may also be applied to recent pedestrians who are absorbed in a mobile phone and therefore walking in danger, as his/her third eye.

Various conditions such as different sidewalks, times of a day, and climate, will be considered in the next experiments. The experiment on crosswalk detection was done once using a real environment video, because the emphasis was rather put on the detection of a sidewalk and a traffic light. Crosswalk detection is to be done employing substantial amount of crosswalk images to enhance the performance of the proposed method.

Demonstration experiments will be performed in the future by visually impaired people. It is necessary to obtain an informed consent from them w.r.t. the purpose of the study and to protect their privacy.

5. **Conclusion.** This paper proposed a walk environment analysis method using MY VISION system that recognizes and infers the surrounding environment of an ego-camera wearer and changes the target of analysis depending on his/her location and intension. The proposed system is intended to be a virtual eye of a visually impaired person. It may also be applicable to a pedestrian absorbed in a mobile phone as his/her third eye. The performed experiments confirmed the usefulness of the proposed method. Future work includes performing more experiments in various conditions and extending the method so that it may detect moving objects on the sidewalk.

Acknowledgment. This research was supported by JSPS Kakenhi, Grant Number 16K0 1554.

REFERENCES

[1] T. Kumano, J. K. Tan, H. Kim and S. Ishikawa, Traffic signs and signals detection employing the MY VISION system for a visually impaired person, *ICIC Express Letters, Part B: Applications*, vol.7, no.2, pp.385-391, 2016.

[2] R. Sakai, J. K. Tan, H. Kim and S. Ishikawa, Detecting pedestrian and extracting their attributes from self-mounted camera views, *ICIC Express Letters, Part B: Applications*, vol.7, no.2, pp.279-286, 2016.

[3] S. Hatano, J. K. Tan, H. Kim and S. Ishikawa, Detecting a specific moving object from self-mounted camera images considering occlusion, *Proc. of SICE Kyushu Conf.*, 2014 (in Japanese).

[4] K. Ishitobi, J. K. Tan and S. Ishikawa, Detection of a specific moving object from head-mounted camera images, *Proc. of IEEE International Symposium on System Integration*, Taipei, Paper WeD4.4, pp.1-6, 2017.

[5] T. Kanade, First-person inside-out vision, *IEEE Workshop on Egocentric Vision*, 2009.

[6] T. Kanade and M. Hebert, First-person vision, *Proc. of the IEEE*, vol.100, no.8, pp.2442-2453, 2012.

[7] K. Kitani, T. Obake, Y. Sato and A. Sugimoto, Fast unsupervised ego-action learning for first-person sports videos, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.3241-3248, 2010.

[8] M. C. Le, S. L. Phung and A. Bouzerdoum, Pedestrian lane detection for assistive navigation of blind people, *Proc. of IEEE the 21st International Conference on Pattern Recognition*, pp.2594-2597, 2012.

[9] S. Treuillet, E. Royer, T. Chateau, M. Dhome and J. Lavest, Body mounted vision system for visually impaired outdoor and indoor way finding assistance, *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments*, 2007.

[10] T. H. Nguyen, T. H. Nguyen, T. L. Le, T. T. H. Tran, N. Vuillerme and T. P. Vuong, A wearable assistive device for the blind using tongue-placed electrotactile display: Design and verification, *International Conference on Control, Automation and Information Sciences*, pp.42-47, 2013.

[11] M. Ptito, S. Moesgaard, A. Gjedde and R. Kupers, Cross-modal plasticity revealed by electrotactile stimulation of the tongue in congenitally blind, *Brain*, vol.128, pp.606-614, 2005.

[12] N. Vuillerme, N. Pinsault, O. Chenu, A. Fleury, Y. Payan and J. Demongeot, A wireless embedded tongue tactile biofeedback system for balance control, *Pervasive and Mobile Computing*, vol.5, no.3, pp.268-275, 2009.
[13] K. Murozono, J. K. Tan, H. Kim and S. Ishikawa, Analyzing a walk environment employing self-mounted camera images, *Proc. of SICE Kyushu Conf.*, pp.69-70, 2015 (in Japanese).
[14] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray, Visual categorization with bags of keypoints, *Workshop on Statistical Learning in Computer Vision*, pp.1-16, 2004.
[15] P. F. Alcantarilla and T. Solutions, Fast explicit diffusion for accelerated features in nonlinear scale spaces, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.34, no.7, pp.1281-1298, 2013.
[16] P. F. Felzenszwalb and D. P. Huttenlocher, Efficient graph-based image segmentation, *International Journal of Computer Vision*, vol.59, no.2, pp.167-181, 2004.
[17] L. Breiman, Random forests, *Machine Learning*, vol.45, no.1, pp.5-32, 2001.
[18] D. Schreiber, B. Alefs and M. Clabian, Single camera lane detection and tracking, *IEEE Conference on Intelligent Transportation Systems*, pp.302-307, 2005.
[19] T. F. Bente, S. Szeghalmy and A. Fazekas, Detection of lanes and traffic signs painted on road using on-board camera, *IEEE International Conference on Future IoT Technologies*, pp.1-7, 2018.
[20] A. Gurghian, T. Koduri. S. V. Bailur, K. J. Carey and V. N. Murali, DeepLanes: End-to-end lane position estimation using deep neural networks, *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.38-45, 2016.
[21] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, Pyramid scene parsing network, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.6230-6239, 2017.
[22] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You Only Look Once: Unified, real time object detection, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.779-788, 2016.
[23] J. Redmon and A. Farhadi, YOLO9000: Better, faster, stronger, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.6517-6525, 2017.
[24] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, *Technique Report, arXiv:1804.02767*, 2018.