Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 196415, 8 pages
http://dx.doi.org/10.1155/2014/196415

Research Article

Moving Object Localization Using Optical Flow for Pedestrian Detection from a Moving Vehicle

Joko Hariyono, Van-Dung Hoang, and Kang-Hyun Jo

Graduate School of Electrical Engineering, University of Ulsan, Ulsan 680-749, Republic of Korea

Correspondence should be addressed to Kang-Hyun Jo; acejo@ulsan.ac.kr

Received 9 April 2014; Revised 7 June 2014; Accepted 8 June 2014; Published 10 July 2014

Academic Editor: Yu-Bo Yuan

This paper presents a pedestrian detection method from a moving vehicle using optical flows and histogram of oriented gradients (HOG). A moving object is extracted from the relative motion by segmenting the region representing the same optical flows after compensating the egomotion of the camera. To obtain the optical flow, two consecutive images are divided into grid cells 14×14 pixels; then each cell is tracked in the current frame to find corresponding cell in the next frame. Using at least three corresponding cells, affine transformation is performed according to each corresponding cell in the consecutive images, so that conformed optical flows are extracted. The regions of moving object are detected as transformed objects, which are different from the previously registered background. Morphological process is applied to get the candidate human regions. In order to recognize the object, the HOG features are extracted on the candidate region and classified using linear support vector machine (SVM). The HOG feature vectors are used as input of linear SVM to classify the given input into pedestrian/nonpedestrian. The proposed method was tested in a moving vehicle and also confirmed through experiments using pedestrian dataset. It shows a significant improvement compared with original HOG using ETHZ pedestrian dataset.

1. Introduction

Vision-based environment detection methods have been actively developed in robot vision. Detecting pedestrian is one of the essential tasks for understanding environment. Pedestrian detection in images could be used in video surveillance systems and driver assistance systems. It is more challenging to detect moving objects or pedestrian in order to avoid an obstacle and control locomotion of the vehicle in the real-world environment.

In the past few years, moving object and pedestrian detection methods for a mobile robot or moving vehicle have been actively developed. For a practical real-time pedestrian detection system, Gavrila and Munder [1] employed hierarchical shape matching to find pedestrian candidates from moving vehicle. Their method uses a multicue vision system for the real-time detection and tracking of pedestrians. Nishida and Kurita [2] applied SVM with the automated selection process of the components by using AdaBoost. These researches show that the selection of the components and their combination are important to get a good pedestrian detector.

Many local descriptors are proposed for object recognition and image retrieval. Mikolajczyk and Schmid [3] compared the performance of several local descriptors and showed that the best matching results were obtained by the scale invariant feature transform (SIFT) descriptor [4]. Dalal et al. [5, 6] proposed a human detection algorithm using histograms of oriented gradients (HOG) which are similar to the features used in the SIFT descriptor. HOG features are calculated by taking orientation histograms of edge intensity in a local region. They are designed by imitating the visual information processing in the brain and have robustness for local changes of appearances and position. Dalal et al. extracted the HOG features from all locations of a dense grid on an image region and the combined features are classified by using linear SVM. They showed that the grids of HOG descriptors significantly out-performed existing feature sets for human detection. Kobayasi et al. [7] proposed selected

feature of HOG using PCA to decrease the number of features. It could reduce the number of features by less than half without lowering the performance.

Moving object detection and motion estimation methods using the optical flow for a mobile robot also have been actively developed. Talukder et al. [8] proposed a qualitative obstacle detection method that was proposed using the directional divergence of the motion field. The optical flow pattern was investigated in perspective camera and this pattern was used for moving object detection. Also, real-time moving object detection method was presented during translational robot motion.

Several researchers also developed methods for egomotion estimation and navigation from a mobile robot using an omnidirectional camera [9, 10]. They tried to measure camera egomotion itself using omnidirectional vision. They used Lucas Kanade optical flow tracker and obtained corresponding features of background in the consecutive two omnidirectional images. The motion of feature points analysis is used to calculate camera egomotion, however they didn't use for moving object detection. They set up an omnidirectional camera on a mobile robot and obtained panoramic image transformed from omnidirectional image. They obtained camera egomotion compensated frame difference based on an affine transformation of two consecutive frames where corner features were tracked by Kanade-Lucas-Tomasi (KLT) optical flow tracker [11]. However, detecting moving objects resulted in a problem that only one affine transformation model could not represent the whole background changes since the panoramic image has many local changes of scaling, translation, and rotation of pixel groups. For this problem, our previous work [12] proposed that each affine transformation of local pixel groups should be tracked by KLT tracker. The local pixel groups are not a type of image features such as corner or edge. We use grid windows-based KLT tracker by tracking each local sector of panoramic image (Figure 2) while other methods use sparse features-based KLT tracker. Therefore, we can segment moving objects in panoramic image by overcoming the nonlinear background transformation of panoramic image [13].

2. Related Works

Proposed method is inspired by the works on pedestrian detection from moving vehicle [1, 8], using optical flow [11] and egomotion estimation [9], we called it is egomotion compensate [12]. Pedestrian as a moving object is extracted from the relative motion by segmenting the region representing the same optical flows after compensating the egomotion of the camera. To obtain the optical flow, image is divided into grid windows and affine transformation is performed according to each window, so that conformed optical flows are extracted. The regions of moving object are detected as transformed objects are different from the previously registered background. Morphological process is applied to get the candidate region of human shape. In order to recognize the object, HOG features were extracted on a candidate region and classified using linear SVM [5, 14]. The HOG feature vectors are used as an input of linear SVM to classify the given input into pedestrian/nonpedestrian. For the performance evaluation, comparative study was presented in this paper.

3. Moving Object Segmentation

This section presents how to detect moving object from the camera mounted on the vehicle. In order to obtain moving object area from video or sequent of images, it is not easy to segment out only moving object area, because the camera moving is also caused by camera egomotion. So, we proposed a method to deal with this situation [12]. We used optical flow analysis to segment independent motion of moving object from egomotion caused by camera. It is called egomotion compensated. The optical flow caused by independent motion of moving object will have different pattern compared with flow caused by egomotion from camera; then, we localize those different pattern as a region of moving object. This region is candidate of detected human/pedestrian after we apply HOG. The overview of the pedestrian detection algorithm is shown in Figure 1.

3.1. Egomotion Compensated. In our previous work [12], we apply KLT optical flow tracker [11] in order to deal with several conditions. Brightness constancy, which is projection of the same point, looks the same in every frame; small motion that points do not move very far and spatial coherence that points move like their neighbors.

The frame difference represents all motions caused by camera egomotion and moving object in the scene. It needs to compensate this effect from frame difference to segment out only independent motion of moving object, so how much the background image has been transformed in two sequences of images. Affine transformation represents the pixel movement between two sequence images as follows:

$$P' = AP + t, \tag{1}$$

where P and P' are pixel location in the first and the second frame. A is transformation matrix and t is translation vector. Affine parameters are calculated by least square method using at least three corresponding features in two images.

In this work, the original input images are converted to grayscale images, and one channel intensity pixel value from the input images is obtained. Then, use two consecutive images which are divided into grid cells of size 14×14 pixels; then compare and track each cell in current frame to find corresponding cell in the next frame. The cell that has the most similar intensity value in a group will be selected as corresponding value. Using method from [11], find the motion distance of each pixel in a group of cells, the motion d in x-axis and y-axis of each cell $g_{t-1}(i, j)$, by finding most similar cell $g_t(i, j)$ in the next frame,

$$g_{t-1}(i, j) = g_t\left(i + d_x, j + d_y\right), \tag{2}$$

where d_x and d_y are motion distances in x-axis and y-axis, respectively. At least three corresponding features are used to estimate the affine parameters using the least square method.

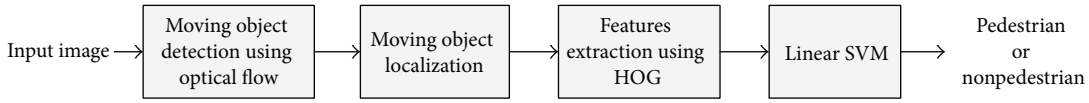


Figure 1: The overview of the pedestrian detection algorithm.

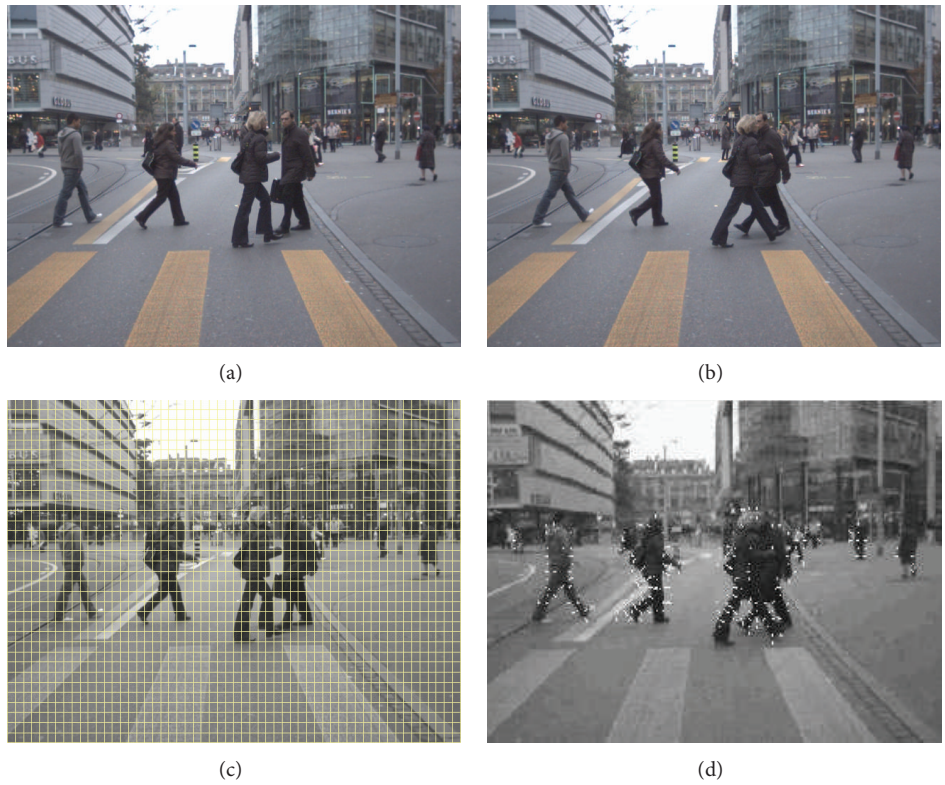


Figure 2: From two consecutive image sequences (a) and (b), we decide grid windows (c) and track each window in the next consecutive image (d).

Equation (2) is rewritten by affine transformation of each pixel in the same cell as follows:

$$I_t(x, y) = AI_{t-1}(x, y) + d, \tag{3}$$

where $I_t(x, y)$ and $I_{t-1}(x, y)$ are vector 2×1 which represent pixel location in the current and previous frame, respectively; A is 2×2 projection matrix and d is 2×1 translation vector. The results are shown in Figure 3.

To obtain the camera egomotion compensated, frame difference is applied in two consecutive input images by calculating based on the tracked corresponding pixel cells using

$$I_d(x, y) = \left| I_{t-1}(x, y) - I_t(x, y) \right|, \tag{4}$$

where $I_d(x, y)$ is a pixel cell located at (x, y) in the grid cell.

Suppose that two consecutive images shown in Figures 3(a) and 3(b) cannot segment out moving object using frame difference Figure 3(c), however when we apply frame difference with egomotion compensate could obtain moving objects area shown in Figure 3(d).

3.2. Moving Object Localization. Each pixel output from frame difference using egomotion compensated cannot show clearly as silhouette. It just gives information of motion areas from moving objects. Those moving areas are applied to morphological process to obtain region of moving object and noise removal.

Ideally, we would seek to devise a region segmentation algorithm that accurately locates the bounding boxes of the motion regions in the difference image. Given the sparseness of the data, however, accurate segmentation would involve the enforcement of multiple constraints, making fast implementation difficult. To achieve faster segmentation, we assumed the fact that humans usually appear in upright positions and conclude that segmenting the scene into vertical strips is sufficient most of the time. In this work, we define detected moving objects that are represented by the position

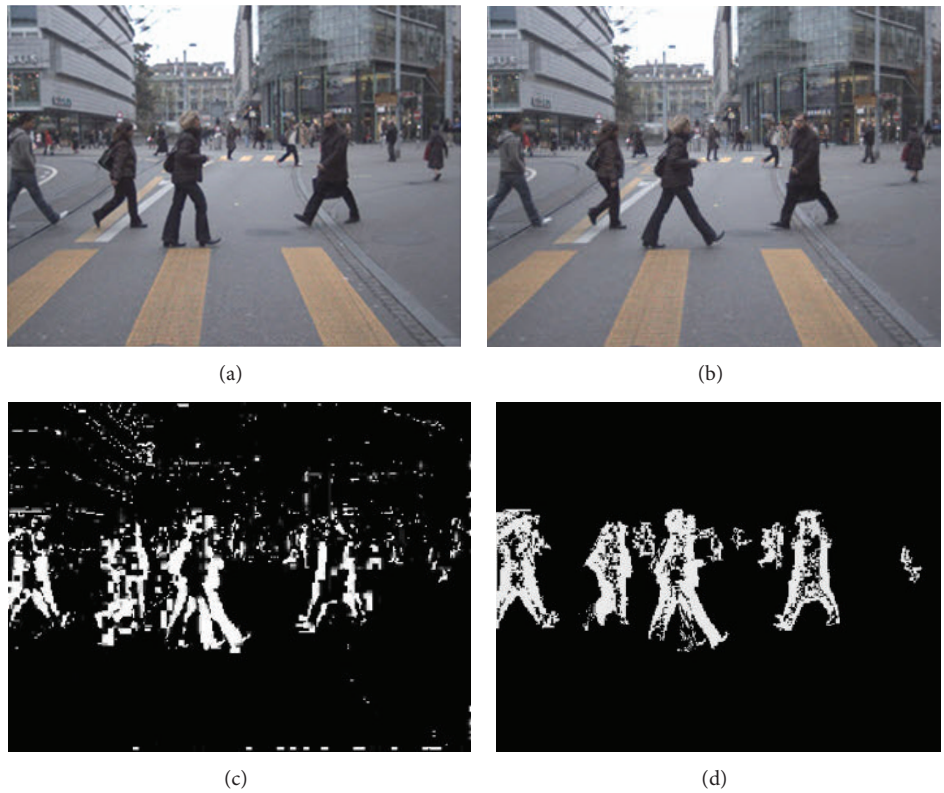


FIGURE 3: From two consecutive images (a) and (b), we applied frame difference (c) and comparing when we applied frame difference with egomotion compensated (d).

in width in x-axis. Using projection histogram h_x by pixel voting vertically projects image intensities into x-coordinate.

Adopting the region segmentation technique proposed in [15], we define the region using boundary saliency. It measures the horizontal difference of data density in the local neighborhood. The local maxima, which correspond to where maximal change in data density occurs, are candidates for region boundaries of pedestrian in moving object detection.

4. Feature Extraction

In this section, we present how we extract feature from candidate region obtained from previous section. In this work, we use histogram of oriented gradients (HOG) to extract features from moving object area localization. Local object appearance and shape usually can be characterized well by the distribution of local intensity gradients or edge direction. HOG features are calculated by taking orientation histograms of edge intensity in local region.

4.1. HOG Features. In this work, we extract HOG features from 16×16 local regions as shown in Figure 4. At first, we use Sobel filter to obtain the edge gradients, and orientations were calculated from each pixel in this local region. The gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ are calculated using directional gradients $dx(x, y)$ and $dy(x, y)$ computed by Sobel filter as

$$m(x, y) = \sqrt{dx(x, y)^2 - dy(x, y)^2}, \tag{5}$$

$$\theta(x, y) = \begin{cases} \tan^{-1}\left(\dfrac{dy(x, y)}{dx(x, y)}\right) - \pi, \\ \quad \text{if } dx(x, y) < 0, dy(x, y) < 0, \\ \tan^{-1}\left(\dfrac{dy(x, y)}{dx(x, y)}\right) + \pi, \\ \quad \text{if } dx(x, y) < 0, dy(x, y) > 0, \\ \tan^{-1}\left(\dfrac{dy(x, y)}{dx(x, y)}\right), \\ \quad \text{otherwise.} \end{cases} \tag{6}$$

The local region is divided into small spatial or cell, each size is 4×4 pixels. Histograms of edge gradients with 8 orientations are calculated from each of the local cells. Then the total number of HOG features becomes $128 = 8 \times (4 \times 4)$ and they constitute a HOG feature vector. To avoid sudden changes in the descriptor with small changes in the position of the window and to give less emphasis to gradients that are far from the center of the descriptor, a Gaussian

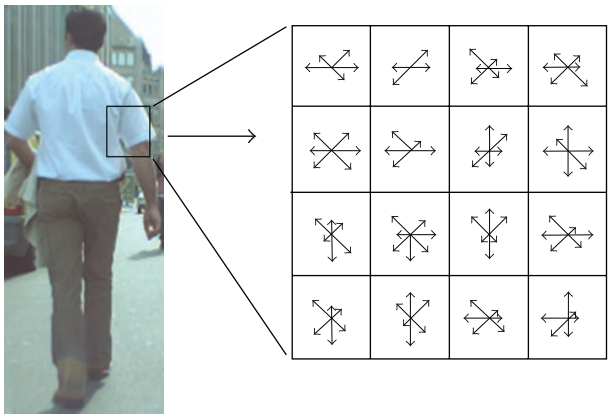
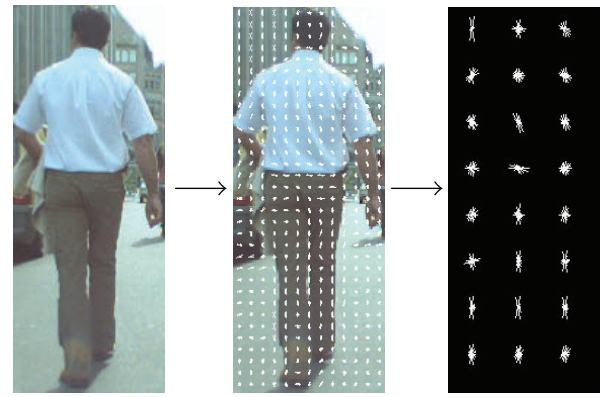


Figure 4: Extraction process of HOG features. The HOG features are extracted from local regions with 16 × 16 pixels. Histograms of edge gradients with 8 orientations are calculated from each of the 4 × 4 local cells.



(a) (b) (c)

Figure 5: From a candidate input image of size 150 × 382 (a), HOG features are extracted from all locations on the candidate region of an input image with 16 × 16 pixels region (b), and the result is shown in (c).

weighting function with σ equal to one-half of the width of the descriptor window is used to assign a weight to the magnitude of each pixel.

A vector of HOG feature represents local shape of an object, it has edge information at plural cells. In flatter regions like a ground or a wall of a building, the histogram of the oriented gradients has flatter distribution. On the other hand, in the border between an object and background, one of the elements in the histogram has a large value and it indicates the direction of the edge. Even though the images are normalized to position and scale, the positions of important features will not be registered with the same grid positions. It is known that HOG features are robust to the local geometric and photometric transformations. If the translations or rotations of the object are much smaller than the local spatial bin size, their effect is small.

Dalal and Triggs [5] extracted a set of HOG feature vectors from all locations in an image grid and that are used for classification. In this work, we just extract the HOG features from all locations on the candidate region localization from an input image as shown in Figure 5.

4.2. Linear SVM Classifier. In the human detection algorithm proposed by Dalal and Triggs [5], the HOG features are extracted from all locations of a dense grid and the combined features are classified using linear support vector machine (SVM). HOG shows significantly outperformed existing feature sets for human detection. This work also used the linear SVM to perform work in various data classification tasks. Let $\{f_i, t_i\}_{i=1}^{N} (f_i \in R^D, t_i \in \{-1, 1\})$ be the given training sample in D-dimensional feature space. The classification function is given as

$$z = \text{sign}\left(\omega^T f_i - h\right), \tag{7}$$

where ω and h are the parameters of the model. For the case of soft-margin SVM, the optimal parameters are obtained by minimizing

$$L(\omega, \xi) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \tag{8}$$

under the constraints

$$\xi_i \geq 0, \quad t_i\left(\omega^T f_i - h\right) \geq 1 - \xi_i \quad (i = 1, \ldots, N), \tag{9}$$

where ξ_i (≥ 0) is the error of the ith sample measured from the separating hyperplane and C is the hyperparameter which controls the weight between the errors and the margin. The dual problem of (8) is obtained by introducing Lagrange multipliers $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N), \alpha_k \geq 0$ as

$$L_D(\boldsymbol{\alpha}) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j t_i t_j f_i^T f_i \tag{10}$$

under the constraints

$$\sum_{i=1}^{N}\alpha_i t_i = 0, \quad 0 \leq \alpha_i \quad (i = 1, \ldots, N). \tag{11}$$

By solving (10), the optimum function is obtained as

$$z = \text{sign}\left(\sum_{i\in S}\alpha_i^* t_i f_i^T f_i - h^*\right), \tag{12}$$

where S is the set of support vectors.

To get a good classifier, we have to search the best hyperparameter C. The cross-validation is used to measure the goodness of the linear SVM classifier.

5. Experimental Results

In this work, our vehicle system is run in outdoor environment with speed that varies from around 0 to 50 kilometers

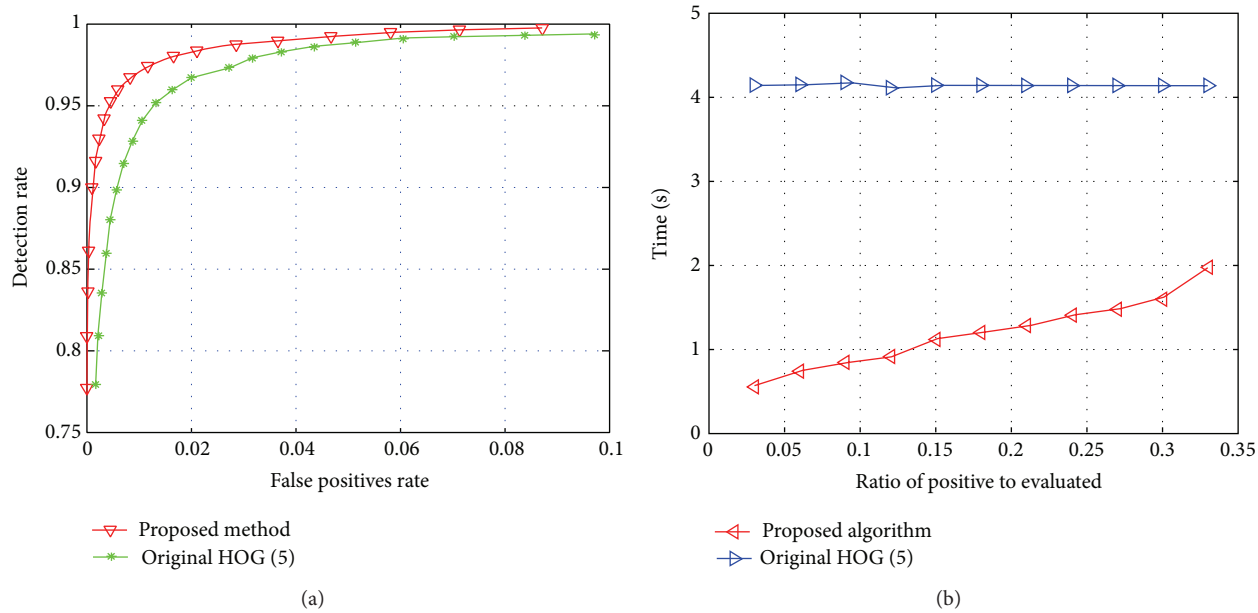


FIGURE 6: Comparison result when we tested our proposed method and original HOG by Dalal et al. (a) Comparison of detection rate and (b) comparison of time consumption.

per hour and detected object moving surround its path. Proposed algorithm was programmed in MATLAB and executed on a Pentium 3.40 GHz, 32-bit operating system with 8 GB random access memory. The proposed algorithm was evaluated by using five sequences of images from ETHZ pedestrian datasets which contain around 5,000 images of pedestrians in city scenes [15]. It contains only front or back views with relatively limited range of poses and the position and the height of human in the image are almost adjusted. The size of the image is 640 × 480 pixels. For the training process, we used person INRIA datasets [5]. These images were used for positive samples in the following experiments. The negative samples were originally collected from images of sky, mountain, airplane, building, and so forth. The number of images is 3,000. From these images, 1,000 person images and 2,000 negative samples were used as training samples to determine the parameters of the linear SVM. The remaining 100 pedestrian images and 200 negative samples were used as test samples to evaluate the recognition performance of the constructed classifier.

We studied methods for detecting human, and one of the objectives of this work is that we want a method that can detect people reliably whether they are moving or not. We were concerned that it might be sensitive to the relative proportion of moving and of static people in the videos. We check reliability of the proposed method that the combination of optical flow and HOG not only on the pure video contain of moving object, but also on objects without moving on the sequent images again with static object flows being zero. The results are diluting the fraction of motion regions naturally reduces the advantage of the combination of methods relative to the static ones; however, using the combination of methods, the relative ranking of the methods remains unchanged. Table 1 shows that when we used on relatively the objects without moving on the images for which there are a less flow field, the best combination of methods detectors do marginally better than the best of original HOG detectors done.

The reliability of our moving object detection system was evaluated whether it still works well in the case if the vehicle ran in varying speed. Outdoor application with speed of vehicle that varies from around 0 to 50 kilometers per hour was performed; then we evaluated the proposed window cells based flow estimation which are still visible at several levels. We tested reliability of the window cells for optical flow tracking in several sizes; it will determines from the relative distance of the object from the camera, so that we consider to choose the flow field window tracking which is more accurate for larger people and also well tracking for smaller people in the image. As a counterweight parameter, computational cost was considered for performance balancing. Table 2 shows the miss detection rate and computational cost of several windows size. However, 10 × 10 cells are the lowest on the miss detection rate but the slowest in computational cost; size 14 × 14 is selected based on low in miss detection rate and faster computational speed.

After all, we implemented original HOG by Dalal et al. using those datasets; the recognition rate for test dataset is 98.3%. Then, we test the combination of methods based on optical flow and HOG feature. HOG feature vectors were extracted from all locations of the grid for each training sample. Then, the selected feature vectors were used as input of the linear SVM. The selected subsets were evaluated by cross validation. Also, we evaluated the recognition rates of the constructed classifier using test samples.

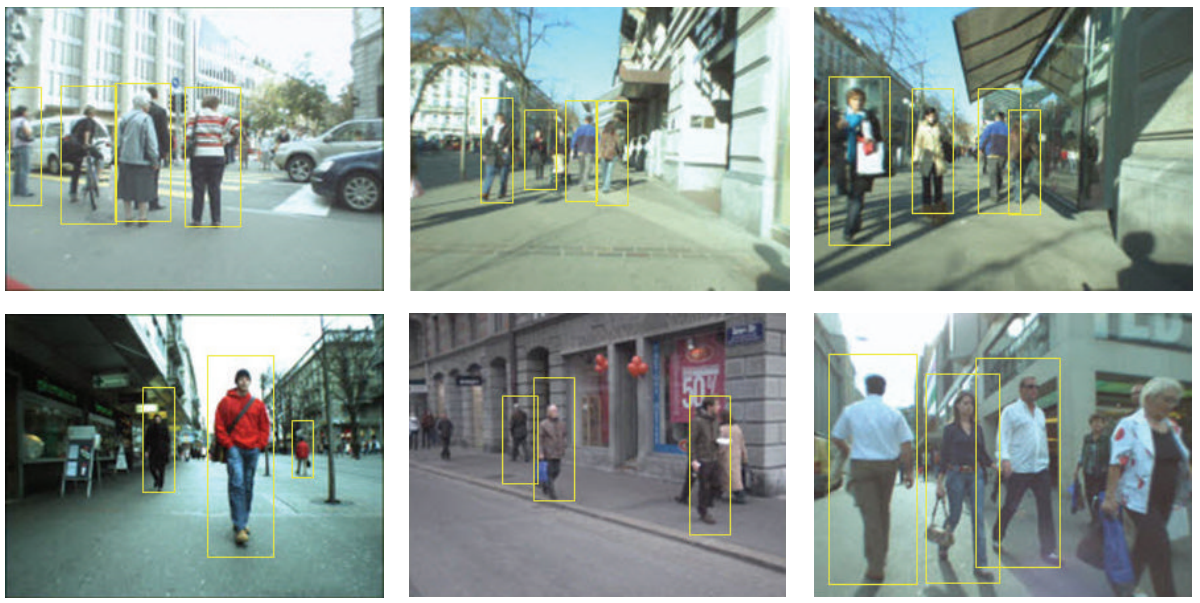
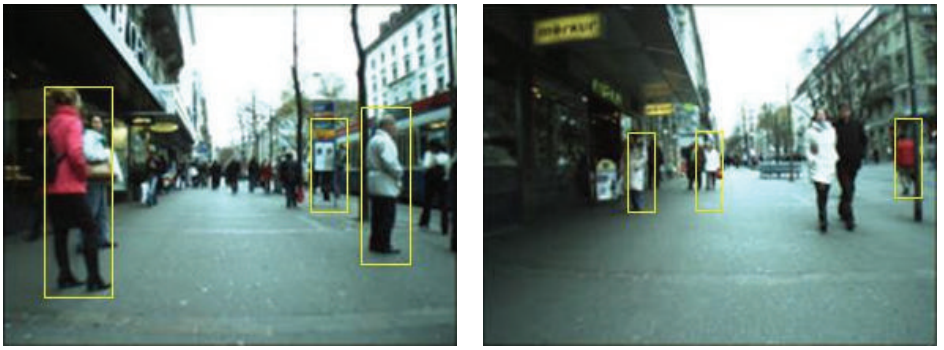


Figure 7: Successful moving objects detection results.



(a) (b)

Figure 8: (a) False positives detection and (b) false negative detection.

Table 1: The detection rates of various detectors.

False positive rate	0.02	0.04	0.06	0.08	0.1
HOG	0.965	0.980	0.985	0.990	0.992
Proposed method	0.980	0.986	0.989	0.992	0.994

Table 2: The miss rates of various cell windows size.

Cell window size	False positive rate (0.09)	Computational cost (fps)
10×10	0.023	11.89
12×12	0.024	11.97
14×14	0.024	12.35
16×16	0.026	12.55
18×18	0.028	12.83

The relation between the detection rates and the number of false positive rate is shown in Figure 6. The best recognition rate, 99.3%, was obtained at 0.09 false positive rates. It means that we obtain higher detection rate with smaller false positives rate. The computational cost also reduces eight times better when we use small ratio of positive to evaluated data. However, if we increase the number of ratios it also reduces time consuming significantly. The detection results are shown in Figure 7 and false detection is shown in Figure 8.

6. Conclusion

This paper addressed the problem for detecting pedestrian from moving vehicle using optical flow and HOG. The moving object is segmented out through the relative evaluation of optical flows to compensate egomotion of camera. Morphological process is applied to get the candidate region of pedestrian. In order to recognize the object, HOG features were extracted on a candidate region and classified using linear SVM. The HOG feature vectors are used as an input of linear SVM to classify the given input into

pedestrian/nonpedestrian. The proposed algorithm achieved comparable results compared with original HOG and also reduces computational cost significantly using moving object localization. In the future work, we consider the combination methods [16] compared with modification of HOG, such as LBP HOG and feature selection HOG.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by the 2014 Research Fund of University of Ulsan.

References

[1] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *International Journal of Computer Vision*, vol. 73, no. 1, pp. 41–59, 2007.

[2] K. Nishida and T. Kurita, "Boosting soft-margin SVM with feature selection for pedestrian detection," in *Proceeding of International Workshop on Multiple Classifier Systems*, vol. 13, pp. 22–31, 2005.

[3] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 886–893, San Diego, Calif, USA, June 2005.

[6] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proceedings of the 9th IEEE European Conference on Computer Vision (ECCV '06)*, Graz, Austria, 2006.

[7] T. Kobayasi, A. Hidaka, and T. Kurita, "Selection of histograms of oriented gradients features for pedestrian detection," in *Proceedings of the 14th International Conference on Neural Information Processing (ICONIP '08)*, vol. 4985 of *Lecture Notes in Computer Science*, pp. 598–607, Kitakyushu, Japan, 2008.

[8] A. Talukder, S. Goldberg, L. Matthies, and A. Ansar, "Real-time detection of moving objects in a dynamic scene from moving robotic vehicles," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1308–1313, Las Vegas, Nev, USA, October 2003.

[9] R. F. Vassallo, S. Victor, and H. Schneebeli, "A general approach for egomotion estimation with omnidirectional images," in *Proceedings of the 3rd Workshop on Omnidirectional Vision*, pp. 97–103, Copenhagen, Denmark, 2002.

[10] H. Liu, N. Dong, and H. Zha, "Omni-directional vision based human motion detection for autonomous mobile robots," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2236–2241, Hawaii, Hawaii, USA, October 2005.

[11] C. Tomasi and T. Kanade, "Detection and tracking of point features," *International Journal of Computer Vision*, vol. 9, pp. 137–154, 1991.

[12] J. Hariyono, V.-D. Hoang, and K.-H. Jo, "Human detection from mobile omnidirectional camera using ego-motion compensated," in *Intelligent Information and Database Systems*, vol. 8397 of *Lecture Notes in Computer Science*, pp. 553–560, 2014.

[13] J. Hariyono, L. Kurnianggoro, D. C. Wahyono, and K. H. Jo, "Ego-motion compensated for moving object detection in a mobile robot," in *Proceedings of the 27th IEA-AEI International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems*, vol. 8482 of *Lecture Notes in Computer Science*, pp. 289–287, Kaohsiung, Taiwan, 2014.

[14] V.-D. Hoang, M.-H. Le, and K.-H. Jo, "Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection," *Neurocomputing*, vol. 135, pp. 357–366, 2014.

[15] A. Ess, B. Leibe, and L. Van Gool, "Depth and appearance for mobile scene analysis," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, Rio de Janeiro, Brazil, October 2007.

[16] Y.-Y. Lu and H.-C. Huang, "Adaptive reversible data hiding with pyramidal structure," *Vietnam Journal of Computer Science*, pp. 1–13, 2014.