

Local descriptors based random forests for human detection

- **Van-Dung Hoang**

Quang Binh University, Vietnam

- **My-Ha Le**

University of Technical Education Ho Chi Minh City, Vietnam

- **Hyun-Deok Kang**

Ulsan National Institute of Science and Technology, Korea

- **Kang-Hyun Jo**

University of Ulsan, Korea

(Manuscript Received on July 15, 2015, Manuscript Revised August 30, 2015)

ABSTRACT

This paper presents a framework based on Random forest using local feature descriptors to detect human in dynamic camera. The contribution presents two issues for dealing with the problem of human detection in variety of background. First, it presents the local feature descriptors based on multi scales based Histograms of Oriented Gradients (HOG) for improving the accuracy of the system. By using local feature descriptors based multiple scales HOG, an extensive feature space allows obtaining

high-discriminated features. Second, machine detection system using cascade of Random Forest (RF) based approach is used for training and prediction. In this case, the decision forest based on the optimization of the set of parameters for binary decision based on the linear support vector machine (SVM) technique. Finally, the detection system based on cascade classification is presented to speed up the computational cost.

Keywords: Multi scales based HOG, Support vector machine, Random decision forest, Local descriptor.

1. INTRODUCTION

In recent years, human detection systems using vision sensors have been become key task for a variety of applications, which have potential influence in modern intelligence systems knowledge integration and management in autonomous systems[1, 2]. However, there are many challenges in the detection procedures such as various articulate poses, appearances, illumination conditions and complex backgrounds of outdoor scenes, and occlusion in crowded

scenes. Up to day, several successful methods for object detection have been proposed. The state of the art of human detection was presented by Dollar *et al.* in [3]. The standard approach investigated Haar-like features using the classification SVM for object detection [4]. However, the performance of Haar-like features is limited in human detection applications [5,6] due to it is sensitive to a high variety of human appearances, complex backgrounds, and

illuminative dynamic in outdoor environments. Other authors proposed the Histograms of Oriented Gradients descriptor (HOG) [7-9] to deal with that problem. In another approach, Schwartz *et al.* [10] proposed the method for integrating whole body detection with face detection to reduce the false positive rate. However, the camera pose is not always opposite with the human, therefore the face is not always appearance. In terms of learning algorithms used in object detection, SVM and boosting methods are the most popular algorithms which have been successfully applied to classification problems.

Recently, some groups focused on combining classification algorithms. They proposed a new hybrid algorithm combining SVM with boosting techniques in order to create a better classification benefitting from the desirable properties of both methods [11]. In order to improve the capability of mechanism system, the heuristic process is added for enforcing the selection of proper subset of training set to avoid the duplication examples and emphasizes the probabilities of examples that hard to learn. However, that paper did not explore the relation of data structure that allows sufficient combining features of data fed to each SVM learner. In other investigation, the system based on AdaBoost and SVM is presented for pedestrian detection [12]. The authors used the SVM technique instead of a one-cascade AdaBoost classifier layer when the number of weak classifiers of the current layer exceeded a preset threshold. That mean the SVM is only used when the number of weak classifiers larger than the threshold value. The strengths of SVM will be omitted when the number of weak classifiers less than preset value. By contrast, the system using AdaBoost and SVM as two stages was proposed for pedestrian detection [13]. The classification system consists of two stages. The AdaBoost is first used to raw classify, and then the output classification is fed to the SVM machine. That

mean SVM is used to confirm all positive examples, which pass the first stage. This method can help to reduce the false alarm rate, but it also reduces the detection rate. The miss-detection examples at first stage will not be rescued at later stage. On the other hand, the system also consumes high computational time because it has to solve the problem in two stages.

On the contrary, this paper focuses on enhancing the accuracy and improving the speed of a pedestrian detection system by using variant scale block-based HOG features along with a hybrid of Random Forests and SVM techniques. The Random Forests technique is used as global system, while the SVM is used as classifier inside of the Random Forests. Vector data input for SVM is blocks of HOG feature vector, this represent data structure for SVM can avoid the duplication common data and guarantee the independence of SVM machines in global system.

2. PRELIMINARY RANDOM FOREST

Random forest (RF) is an ensemble model in machine learning, which is used for classification and regression. The basic idea based on construction of multiple decision trees at the training step. The prediction output is combination of all individual trees in forest. In the training step, the selection subset of sample features for each tree is randomly processed.

The trees are grown very deep tend to learn highly irregular patterns, which can made over-fitting the model with training data. The RF is averaging multiple deep decision trees, trained on different parts of the same training data, with the objective of reducing the variance.

The training algorithm for random forest applies the general technique of bootstrap aggregating to tree learners, which is summarized as follows.

Given a training data set $T=(X,Y)$ with $X=\{x_1, \dots, x_n\}$ and $Y=\{y_1, \dots, y_n\}$ are the samples and

labels, respectively. The label Y is a set of classes ($Y=\{0,1\}$ for binary classification). The bagging repeatedly selects a random sample feature with replacement of the training set and fits trees to these samples.

For $t = 1, \dots, T$:

(a) Randomly sample a small subset of features, called $\Gamma_s \subset \Gamma$

(b) For each $\rho_j \in \Gamma_s$

(b-1) Split the set of Y_j into two subsets by split function $h(x, \xi_j)$, which ξ is the set of defined parameters of split function, with the feature selector ϕ .

$$\begin{cases} Y_j^R = \{x \in Y_j \mid h(\phi_j(x), \xi) = 1\} \\ Y_j^L = \{x \in Y_j \mid h(\phi_j(x), \xi) = -1\} \end{cases} \quad (28)$$

(b-2) Evaluation for goodness of partition by using purity measurement, which called as information gain.

$$\tau(\xi) = H(Y_t) - \sum_{c \in \{L, R\}} \frac{|Y_t^c|}{|Y_t|} H(Y_t^c) \quad (29)$$

where the entropy $H(Y)$ is

$$H(Y) = - \sum_{c \in \text{classes}} p(c \mid \phi_j(x)) \log(p(c \mid \phi_j(x)))$$

(c) The objective is finding the parameters ξ for each node j to maximal information gain

$$\xi_j^* = \underset{\xi_j \in \xi}{\operatorname{argmax}} (\tau(\xi_j)) \quad (30)$$

The ensemble prediction of RF is presented as follows:

$$p(c \mid x) = \sum_{t=1}^T p_t(c \mid \phi_t(x)) \quad (31)$$

where p_t is the decision prediction of each tree in the forest.

Training decision tree includes all training data $\{x\}$, the feature selector $\phi: R^d \rightarrow R^{d'}$ with $d' \ll d$. The forest of tree can be process parallel. Due to $d' \ll d$, the RF can deal with the expensive

consuming time in the case of huge dimensional data.

3. LOCAL DESCRIPTORS

In this contribution, a feature descriptor based on HOG features is applied [7]. The general flowchart of feature extraction is presented in Fig. 1. Difference to other approaches, the split function of weak classifier based on optimization of maximum margin hyperplane of the *feature descriptor in local patch* is used. The ensemble of local descriptor is solved by appropriate feature selector $\phi(x)$. Fig. 2 demonstrates the idea of the use local descriptors based ensemble approach. In this work, the set of local feature block is used at a node for split function. The optimization ξ parameter is solved by the linear SVM learning method.

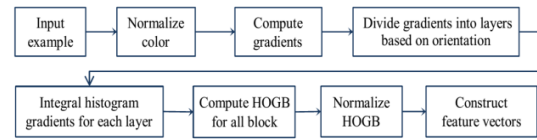


Figure 1. Feature extraction flowchart

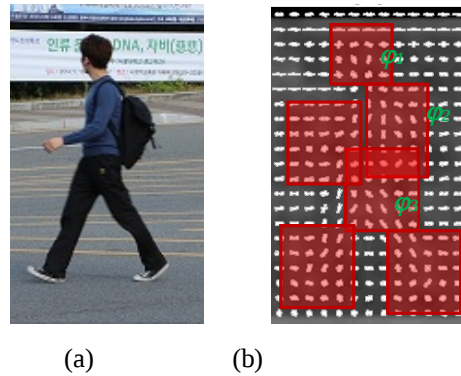


Figure 226. Random forest based local feature descriptors: (a) image sample, (b) feature selector for partial block descriptor.

The extended descriptor is improved based on the original HOG [7] by using multiple scale block based HOG feature. There was no limitation in the scale degrees of block size for constructing HOG features, providing an extensive feature

descriptor space, which helps in obtaining highly discriminative features for high accuracy detection. Because of the use of multiple scale levels, histograms of gradients are repeatedly computed many times around the sample region. Therefore, to speed up the system, a cumulative sum of histogram gradients method is used for rapidly computing the feature descriptor. Similarly, the histogram of each oriented gradient within an arbitrary region is computed with four accesses using the cumulative sum gradient table (CS). In accordance with the characteristics of the cumulative sum table, gradients are separated into groups based on orientation, with each group organized into one table for computing cumulative sums. Each CS table is used to compute the histogram of gradients with respect to each orientated interval, e.g., each 20 degrees for one group, which is known as one layer, illustrated in Fig. 5. Finally, the histogram of gradients within any block only requires four operations multiplying with the number of oriented gradient layers, e.g. 4 operations/layer $\times$ 9 layers, with respect to 9 groups of orientation gradients.

In coherence with our argument, the HOG feature descriptor as well as the fast computation based on the cumulative sum of histogram gradients method is briefly presented [9]. The gradient values at each pixel in the sample image are computed by discrete derivations. The filter kernels $[-1 \ 0 \ 1]$ and $[-1 \ 0 \ 1]^T$ are used to compute discrete derivations on horizontal and vertical axes, respectively. G_x and G_y are directional gradients on the x and y axis, respectively. The gradient magnitude and gradient orientation are computed as follows:

$$\theta = \arctan(G_y / G_x) \quad (1)$$

The gradient magnitudes are separated into 9 tables based on their oriented angles. The unsigned orientation of the gradients (spaced from 1 degree through 180 degrees, in conjunction with 9 bins, 20 degrees/bin) is used to construct the histogram of oriented gradients, as depicted in Fig. 3. Each table of gradients is used to compute the cumulative sum gradients. Finally, 9 CS tables are used for computing the HOGBs and constructing the feature vector, which feed into training and classification.

Fig. 4 presents the visualization of HOG using different size of basic cells. As the use of multiple scales of cell size is inevitable, several HOGBs are highly discriminative between positive (person) and negative (non-person) regions, besides that also there are many low distinctive HOGBs. To select for the highly discriminative blocks, which are used for classification stage, the SVM technique is applied on each individual HOGB for training and evaluation. Only blocks, so that SVM results high accuracy, would be selected for detection system. This preprocessing step is applied for both full-body and component detections.

4. EVALUATION

In this session, the affection of some criteria to the time consuming and accuracy of the RF for object detection is analyzed and tested. The training data consists of 1,500 positive samples and 1,500 negative samples. In classification stage, the evaluation data includes 15,000 positive samples and 15,000 negative samples. Fig. 5 shows testing results of 15 times and the mean values on the same data. The results show that, there is a tradeoff of the RF, the large number of trees results in high accuracy, also expensive computational time and vice versa. Therefore, the number of tree in forest is defined based on the

objective of the system, which is balance accuracy and time processing target.

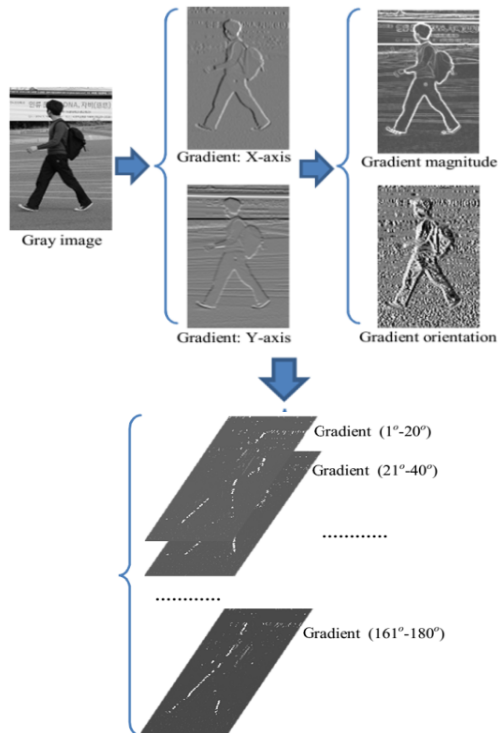


Figure 3. Gradient process based on orientation for the cumulative sum method.

Fig. 6 presents the comparison results of the SVM and the RF classification method. The results figure out that the SVM results higher detection rate than the RF at low false detection rate. However, the RF results higher that of at high false detection. In other comparison criteria, SVM is usually faster in training stage, and slower in classification stage than the RF. Fig. 7 presents the comparison results of our feature descriptor with

original HOG with LBP feature descriptors using SVM classification method. Fig. 8 presents some results of people detection.

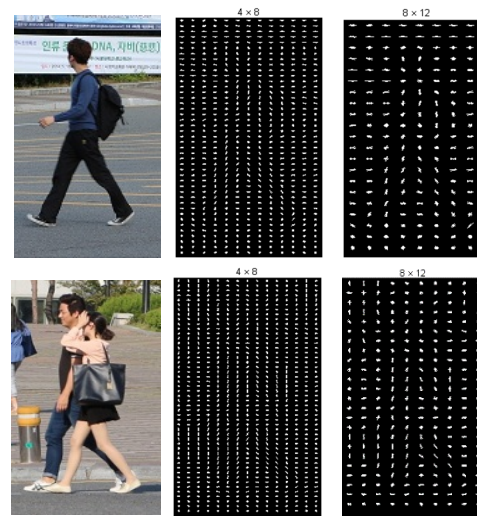


Figure 4. Intuitive histogram of oriented gradients using HOG based on different sizes.

5. CONCLUSION

The classification approach based on local feature descriptors and the RF frame-work is presented for human detection. The approach utility of advantage of fast processing based forest of decision trees and robustness of the SVM for estimating the optimal parameters for split function. The classification method is based on the RF ensemble using multiple local feature descriptors. The proposed method utilizes the rich block-based descriptor. The computing time of the variety block sizes based feature descriptor is speeded up using heuristic stored data structure

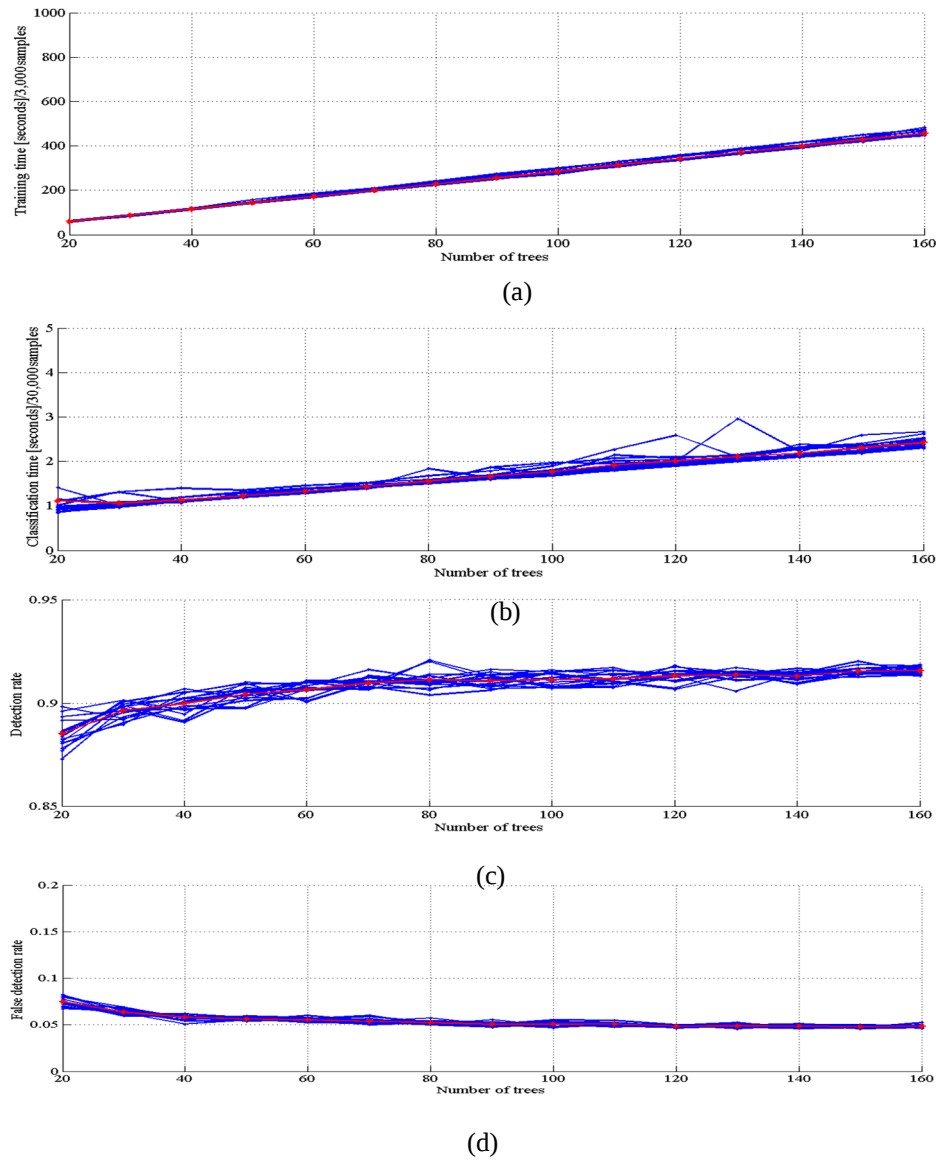


Figure 5. Affection of the number of trees to (a) training time, (b) classification, (c) detection rate, and (d) miss detection rate

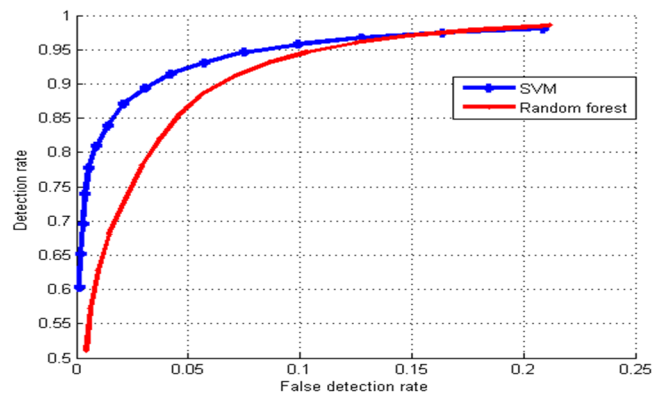


Figure 6. Comparison of accuracy result by using SVM and RF methods.

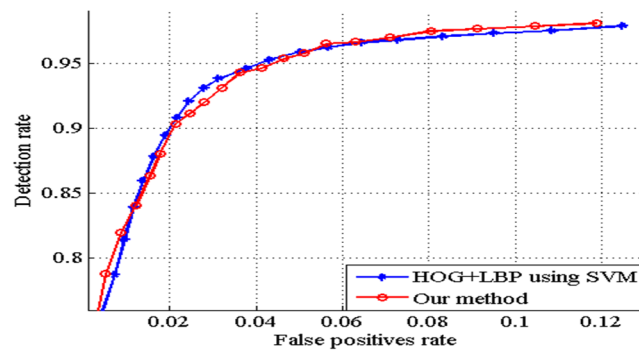


Figure 7. The comparison of our method with the standard approach HOG+ SVM method.

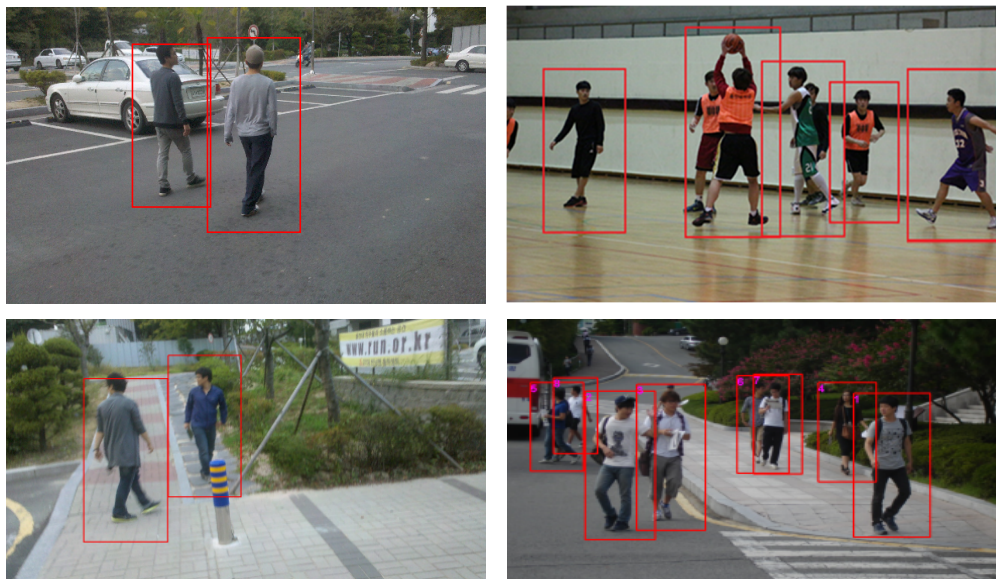


Figure 8. Some detection results.

Kết hợp phương pháp biểu diễn đặc trưng cục bộ và kỹ thuật random forests trong nhận dạng người

- **Hoàng Văn Dũng**

Trường Đại học Quảng Bình, Việt Nam

- **Lê Mỹ Hà**

Trường Đại học Sư phạm Kỹ thuật thành phố Hồ Chí Minh, Việt Nam

- **Kang Hyun Deok**

Viện Khoa học và Công nghệ quốc gia Ulsan, Hàn Quốc

- **Jo Kang Hyun**

Trường Đại học Ulsan, Hàn Quốc

TÓM TẮT

Bài báo trình bày hệ thống phân loại dựa trên kỹ thuật Random forest sử dụng phương pháp biểu diễn đặc trưng cục bộ áp dụng trong nhận dạng người. Có hai nội dung chính được trình bày trong bài này để giải quyết vấn đề nhận dạng trong trường hợp cảnh nền thay đổi đa dạng. Thứ nhất, chúng tôi trình bày kỹ thuật biểu diễn đặc trưng HOG đa mức độ kích thước vùng cục bộ nhằm tăng độ chính xác của hệ thống phân loại. Phương pháp này cho phép trích rút ra một tập lớn các đặc trưng, sau đó lọc ra chỉ những phần tử có độ khác biệt cao giữa tập positive và negative

dựa vào bộ dữ liệu huấn luyện. Thứ hai, máy phân loại sử dụng cấu trúc thác nước dựa trên kỹ thuật RF được đề xuất sử dụng để huấn luyện và nhận dạng. Trong trường hợp này, kỹ thuật decision forest dựa trên việc kết hợp các quyết định yếu sử dụng nhân phân loại là các SVMs. Mỗi phân loại yếu sử dụng tập đặc trưng trong một vùng cục bộ của mẫu. Hệ thống sử dụng cấu trúc thác nước cho phép tăng tốc độ phân loại nhờ vào việc loại bỏ được các mẫu negatives mà chỉ cần một tập nhỏ đặc trưng cục bộ.

Từ khóa: Multi scales based HOG, Support vector machine, Random decision forest, Local descriptors

REFERENCES

- [1] V.-D. Hoang, D. C. Hernández, M.-H. Le, and K.-H. Jo, "3D Motion Estimation Based on Pitch and Azimuth from Respective Camera and Laser Rangefinder Sensing", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, pp. 735-740, 2013.
- [2] V.-D. Hoang, D. Hernández, and K.-H. Jo, "Combining Edge and One-Point RANSAC Algorithm to Estimate Visual Odometry",

- Intelligent Computing Theories*. vol. 7995, D.-S. Huang, *et al.*, Eds., ed, pp. 556-565, 2013.
- [3] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 743-761, 2012.
- [4] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance", *International Conference on Computer Vision*, pp. 734-741, 2003.
- [5] S. Munder and D. M. Gavrila, "An Experimental Study on Pedestrian Classification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1863-1868, 2006.
- [6] V.-D. Hoang, A. Vavilin, and K.-H. Jo, "Fast Human Detection Based on Parallelogram Haar-Like Feature", *The 38th Annual Conference of The IEEE Industrial Electronics Society*, Montréal, Canada, pp. 4220-4225, 2012.
- [7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", *Conference on Computer Vision and Pattern Recognition*, pp. 886-893, 2005.
- [8] V.-D. Hoang, M.-H. Le, and K.-H. Jo, "Robust Human Detection Using Multiple Scale of Cell Based Histogram of Oriented Gradients and AdaBoost Learning", *Computational Collective Intelligence. Technologies and Applications*. vol. 7653, N.-T. Nguyen, *et al.*, Eds., ed, pp. 61-71, 2012.
- [9] 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.
- [10] W. Schwartz, R. Gopalan, R. Chellappa, and L. Davis, "Robust Human Detection under Occlusion by Integrating Face and Person Detectors", *Advances in Biometrics*. vol. 5558, M. Tistarelli and M. Nixon, Eds., ed: Springer Berlin Heidelberg, pp. 970-979, 2009.
- [11] T. T. Maia, A. P. Braga, and A. F. de Carvalho, "Hybrid classification algorithms based on boosting and support vector machines", *Kybernetes*, vol. 37, pp. 1469-1491, 2008.
- [12] W.-C. Cheng and D.-M. Jhan, "A self-constructing cascade classifier with AdaBoost and SVM for pedestrian detection", *Engineering Applications of Artificial Intelligence*, vol. 26, pp. 1016 - 1028, 2013.
- [13] L. Guo, P.-S. Ge, M.-H. Zhang, L.-H. Li, and Y.-B. Zhao, "Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine", *Expert Systems with Applications*, vol. 39, pp. 4274-4286, 2012.