An Analysis of Edge Orientation and Magnitude in Co-occurrence Feature Descriptor

Hirokatsu Kataoka[1 2†] and Yoshimitsu Aoki[2]

[1]Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan
(E-mail: kataoka@aoki-medialab.org)
[2]Graduate School of Science and Technology, Keio University, Kanagawa, Japan

Abstract: Fundamental study on image feature descriptor for robust object detection is one of the essential issues in computer vision field. Local image features are effectively utilized for classifying a target object and background so that they are applied to face, human, vehicle detection systems. Many types of feature descriptor have been proposed in these days. Co-occurrence feature is known as the highest recognition performance among the feature descriptors. This paper analyzes edge orientation and magnitude in co-occurrence feature by adopting co-occurrence histograms of oriented gradients (CoHOG) [6] and its extended feature [7]. Extended CoHOG (ECoHOG) additionally extracts edge magnitude against CoHOG that acquires only edge direction pairs. The effectiveness of the magnitude in co-occurrence feature is proved on pedestrian detection performance test using INRIA person dataset and Daimler pedestrian benchmark dataset.

Keywords: Local Feature Descriptor, Human Detection, CoHOG, ECoHOG.

1. INTRODUCTION

The research about local feature descriptor is one of important topic for detection problem. Recently, local feature descriptor is applied to localize the position of human, vehicle and face and so on. Through the years, various local feature descriptors are proposed and used for object detection. To detect pedestrian, machine learning can be applied in computer vision technique. In machine learning, positive and negative images are prepared for creating classifiers. We must set a feature descriptor to capture the feature from a human. Many feature descriptors have been studied in computer vision and machine learning community [1], and we can apply various of feature descriptors. In human detection study, for example, Gandhi *et al.* [2], Dollar *et al.* [3] and Geronimo *et al.* [4][5] have published survey papers respectively. They have argued local feature (i.e. shape, texture and color descriptor) based on machine learning is effective approach for human detection.

HOG (Histograms of Oriented Gradients) is the famous and effective method as the previous work for object detection. HOG can represent rough human shape from local patch in image. Though there are many types of feature descriptor about HOG, CoHOG (Co-occurrence Histograms of Oriented Gradients) is the state-of-the-art method in object detection [6]. CoHOG describes edge-pair from two different pixels, i.e. head and shoulder in human detection. CoHOG reduces over-detection from HOG considering orientation-pair counting for human detection. On the other hand, ECoHOG additionally extracts edge magnitude on behalf of orientation-pair counting [7]. ECoHOG indicates better performance than CoHOG on the previous paper.

In this paper, we analyze edge orientation and magnitude in co-occurrence feature descriptor through the difference between CoHOG and ECoHOG. At the beginning, we present ECoHOG improvements of (i) edge magnitude accumulation to feature histogram (ii) normalization of feature histogram from CoHOG. In detection and analyzing experiment, we show the effectiveness of edge magnitude in co-occurrence feature descriptor.

† Hirokatsu Kataoka is the presenter of this paper.

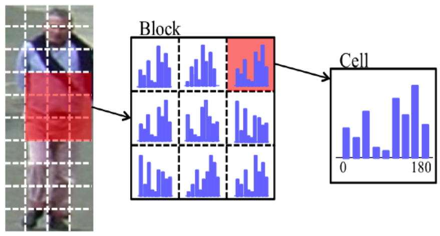

Fig. 1 Feature extraction of HOG

The rest of the paper is organized as follow. In Section 2, we explain the related works about HOG and CoHOG. The description of ECoHOG is described in Section 3, and the detection and analyzing experiments are given in Section 4. Finally, Section 5 concludes the paper.

2. RELATED WORKS: HOG & COHOG

In this chapter, we explain about HOG and CoHOG feature, based on the approrach to the proposed framework.

2.1. Histograms of Oriented Gradients (HOG)

The HOG feature expresses that edge magnitude and orientation accumulating edge information to feature histogram. Feature extraction window is scanned in a image and divided to block and cell in order to acquire edge information (Figure.2). Therefore, the HOG feature can represent rough shape of an object.

The process flow of the HOG feature is described here. The feature is extracted scanning window in a image. In the feature extraction window, a edge magnitude is captured and accumulated to feature histogram. The equa-

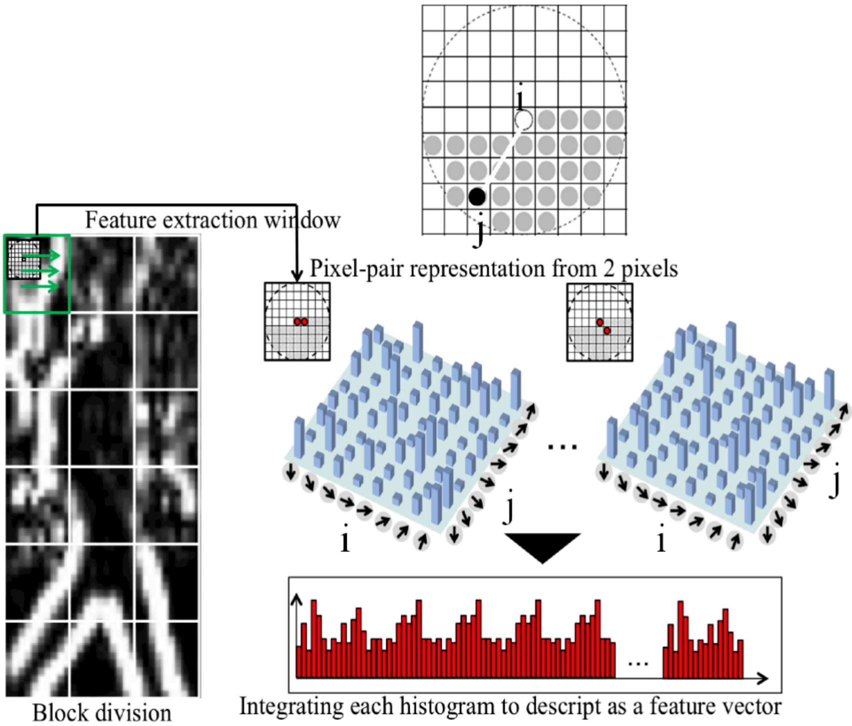

Fig. 3 Flow of CoHOG

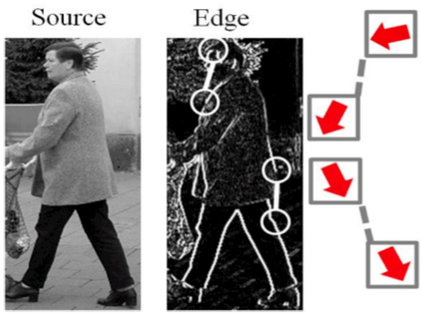

Fig. 2 Pair of edge orientation

tion of calculating edge orientation and magnitude are given below:

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \quad (1)$$

$$g(x,y) = tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \quad (2)$$

$$f_x(x,y) = I(x+1,y) - I(x-1,y) \quad (3)$$

$$f_y(x,y) = I(x,y+1) - I(x,y-1) \quad (4)$$

where $I(x,y)$ is the brightness a pixel, $m(x,y)$ and $g(x,y)$ are edge magnitude and orientation. Generally, a block consist of three × three cells. Edge orientation is divided 180 degree into nine bins, therefore, 1 cell is given as nine dimension. After the feature extraction step, histograms are set off with normalization. The feature histogram takes care about brightness changing with normalization. The normalization is shown below:

$$h' = \frac{h}{\sqrt{\sum_{i=0}^{k} h_i^2 + \epsilon}} \quad (5)$$

where h' is the histogram after normalization, h gives the histogram before normalization. k shows the number of dimension, and we set ϵ as 1.0.

2.2. Co-occurrence Histograms of Oriented Gradients (CoHOG)

The CoHOG feature is the feature descriptor that considers a co-occurrence between two pixels and counts the number of pixel pairs. In HOG feature description, an edge orientation and magnitude is accumulated to feature histogram. On the other hand, the CoHOG feature reduces a lot of over detection describing pixel pair feature in the CoHOG feature descriptor, for example, a pixel pair of head and shoulder is described at the same time (Figure2). The CoHOG feature extracts an edge pair scanning window in a image. Edge feature is sampled to histogram from pixel pair. Edge orientation is divided into eight direction, and pixel pair represents the number of dimension is 64. Edge orientation is calculated shown in below:

$$g(x,y) = tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \quad (6)$$

$$f_x(x,y) = I(x+1,y) - I(x-1,y) \quad (7)$$

$$f_y(x,y) = I(x,y+1) - I(x,y-1) \quad (8)$$

where $I(x,y)$ is the brightness, $g(x,y)$ is edge orientaiton. The process of the CoHOG is shown in Figure.2.2

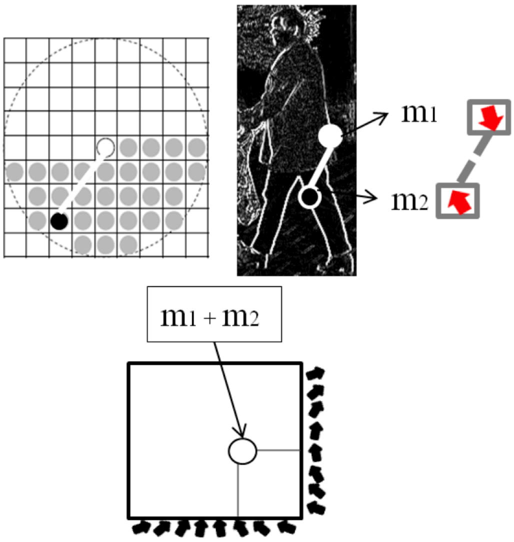

Fig. 4 Orientation magnitude accumulation

CoHOG can express co-occurrence edge orientation acquiring from two pixels. CoHOG indicates high accuracy better than HOG because of the co-occurrence edge representation. However, CoHOG evenly counts all co-occurrence edges regardless of the edge magnitude. Human detection with CoHOG includes over-detection depending on presence of edge on a local image. Similar objects (e.g. tree, traffic sign) from human have a lot of the same elements to be consist of a histogram. We believe edge magnitude is effective for creating feature vector in human detection.

3. EXTENDED COHOG

We added the process of edge magnitude accumulation and histogram normalization. In this section, we explain the method for edge magnitude accumulation and histogram normalization. The ECoHOG, that is the improvement of feature descriptor is described below:

3.1. Accumulating Edge Magnitude

Edge magnitude pair represents a detailed human shape. In the proposed framework, we accumulate a sum of two pixel edge magnitude. The sum of edge magnitude represents the difference of two pixel edge magnitude. The sum of edge magnitude represents the difference between pedestrians and backgrounds representing total feature from an image. The description of the ECoHOG is below:

$$C_{x,y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m} \begin{cases} m_1(x_1,y_1)+m_2(x_2,y_2) \\ (if\ d(p,q)=i\ and \\ \ d(p+x,q+y)=j) \\ \\ 0 \\ (otherwise) \end{cases} \quad (9)$$

where $m(x,y)$ is edge magnitude, $C(i,j)$ shows the co-occurrence value of each element, coordinate (p,q) is the center of the window, coordinate $(p+x,p+y)$ is the position of pixel pair. $d(p,q)$ is the divided orientation representing 0-7.

ECoHOG describes edge-magnitude at each direction in order to create co-occurrence histogram. ECoHOG strongly express boundary between human/background and upper/lower body. Edge magnitude representation can evaluate depending on the strength of edge. The feature descriptor represents not only combination of curvature, straight line but the degree to create better feature vector than CoHOG.

3.2. Histogram Normalization

Brightness of a image is always changing depending on light sources. Feature histogram should be normalized for a robust detection under the various light situations. The range of normalization is 64 dimensions, that is the co-occurrence histogram. The equation of normalization is given below:

$$C'_{x,y}(i,j) = \frac{C_{x,y}(i,j)}{\sum_{i'=1}^{8}\sum_{j'=1}^{8} C_{x,y}(i',j')} \quad (10)$$

4. EXPERIMENT

In this section, we carried out the experiment one with INRIA person dataset [9] and Daimler pedestrian benchmark dataaset [10], and the experiment two to compare the proposed and previous framework in real world dataset.

4.1. Datasets & Implementation

We applied INRIA person dataset [9] and Daimler pedestrian benchmark dataset which includes pedestrian and background images. The positive and negative images are shown in Figure 5, Figure 6 and Figure 7. Table 1 and Table 2 gives the detail of the INRIA person dataset and the Daimler pedestrian benchmark dataset respectively. The number of pixel for feature extraction is 18 pixels, the offset window length is 7 pixels, and the dimensions of ECoHOG is 4608. We applied detection error tradeoff (DET) curve for the verification. The vertical axis of DET curve is miss rate, and the horizontal axis is false positive rate, therefore, bottom-left of the DET curve shows higher performance. And receiver operating characteristic (ROC) curve is adopted in the second experiment. The vertical axis of ROC curve is detection rate, and the horizontal axis is false positive rate, top-left of the ROC curve shows higher performance.

4.2. The Experiment for Feature Improvement

At the beginning, we compare the proposed and previous frameworks Figure.8. Figure.9 shows the comparison with CoHOG and ECoHOG on the Daimler pedestrian benchmark dataset. Table 3 gives the CoHOG and ECoHOG process time at each frame.

The Figure 8 shows the proposed framework is the highest value, ECoHOG accumulates edge magnitude

Fig. 5 Positive images of INRIA person dataset

Fig. 6 Negative images of INRIA person dataset

Fig. 7 Positive and negative images of Daimler pedestrian benchmark dataset

Table 1 INRIA person dataset

Training data	2,415 positive images 12,180 negative images
Test data	1,132 positive images 453negative images
Image size	Positive image : 64 × 128 pixels Negative image : 214 × 320 - 648 × 486 pixels

and normalize an image. The effectiveness is much higher than CoHOG. From HOG to CoHOG and CPF, we consider the feature as co-occurrence from two different pixels with offset window or returning value of classifier. Co-occurrence histogram allows us to reduce miss detection from HOG feature vector. And the difference between CoHOG and ECoHOG, we put edge-magnitude into co-occurrence histogram. We evaluate the strength of edge-magnitude in the feature vector.

Table 2 Daimler pedestrian benchmark dataset

Training data	4,800 positive images 5,000 negative images
Test data	4,800 positive images 5,000 negative images
Image size	Positive image : 18 × 36 pixels Negative image : 18 × 36 pixels
Dataset	5 training sets for cross-validation

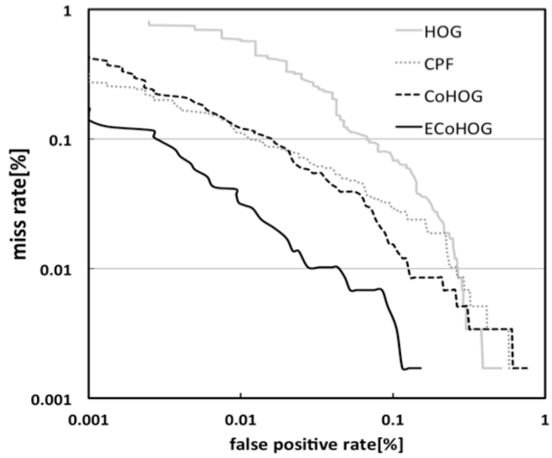

Fig. 8 The comparison of four feature descriptors applying Detection Error Tradeoff (DET) curve : ECoHOG, CoHOG, CPF, HOG

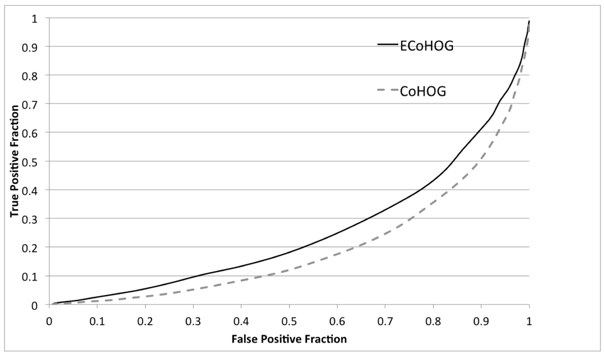

Fig. 9 The ROC curve of CoHOG and ECoHOG on Daimler pedestrian benchmark dataset

Table 3 Processing time of feature descriptors (CoHOG, ECoHOG)

Edge-based feature descriptor	Prodessing time
CoHOG	49.59 msec
ECoHOG	51.67 msec

The Figure 9 shows the comparison of CoHOG and ECoHOG on the Daimler pedestrian benchmark dataset. We applied Daimler pedestrian benchmark dataset in order to verificate versatility in small human images and complicated situations. The DET curves indicate our proposed approach is better than CoHOG in traffic scenes.

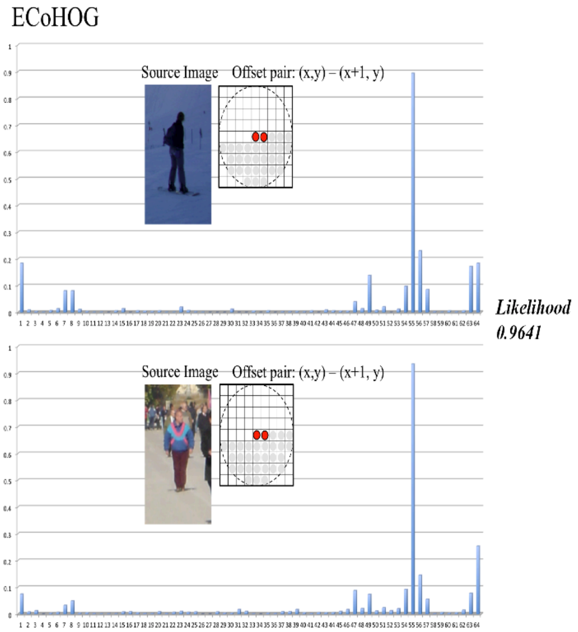

Fig. 10 ECoHOG feature vectors between simple texture and complicated texture image and the degree of histogram similarity (0.9641)

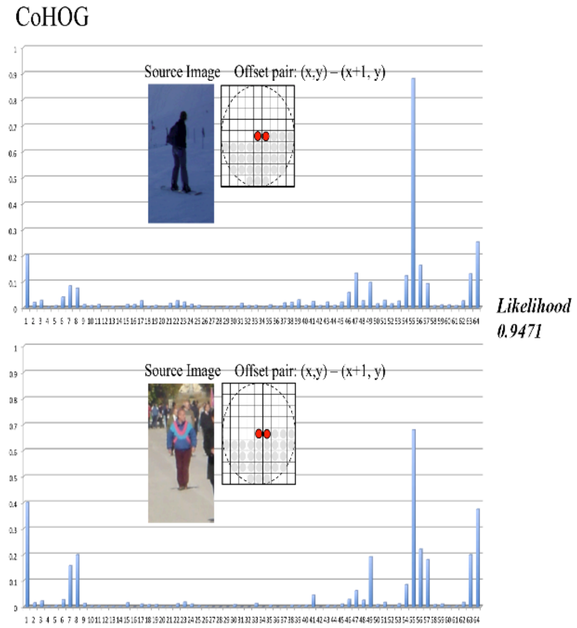

Fig. 11 CoHOG feature vectors between simple texture and complicated texture image and the degree of histogram similarity (0.9471)

ECoHOG can capture extremely minute feature even if local image is small. In the case of Daimler pedestrian benchmark dataset, the pedestrian image is only 18×36 pixels.

4.3. Analysis of CoHOG and ECoHOG

We carried out the analysis of CoHOG and ECoHOG on general used dataset. In this paper, we show histogram comparison from positive-positive and positive-negative samples. Positive-positive is preferable if two histograms are similar cause of two histograms captured from humans are the same class. And positive-negative is desirable to separate from each other. We adopt Bhattacharyya coefficient [12] in order to calculate two histogram similarity as below:

$$S = \sum_{u=1}^{m} \sqrt{h_u^1 h_u^2} \tag{11}$$

where S is similarity value ($0 \leq S \leq 1$), h^1 and h^2 are feature vectors normalized as $\sum_{u=1}^{m} h_u^1 = \sum_{u=1}^{m} h_u^2 = 1.0$, m shows the number of histogram bin.

Figure 10 and Figure 11 show feature vectors between simple texture and complicated texture image and the degree of histogram similarity. The images in Figure 10 and Figure 11 come from INRIA person dataset. These figures indicate the Bhattacharyya coefficient are "0.9641 (ECoHOG)" and "0.9471 (CoHOG)" respectively. ECoHOG histograms are more similar in spite of human texture changing. For example, the peaks of ECoHOG histogram (55th bin) have almost the same value. From this result, ECoHOG can evaluate only high strength edge in a local image. CoHOG tends to confuse histograms that extract existence of edge.

Figure 12 and Figure 13 show feature vectors between human and similar background to human. The degrees of histogram similarity indicate "0.9485 (ECoHOG)" and "0.9576 (CoHOG)" in Figure 12 and Figure 13. In the case of positive-negative similarity, low value is more preferable to separate the two class in machine learning step. ECoHOG is better approach to classify human and background samples.

Histogram analysis allows us to understand that proposed approach ECoHOG is better way to classify human and background classes in machine learning. ECoHOG gets a positive-positive pair together, and separate a positive-negative pair from each other by applying edge magnitude accumulation. Although the histograms in Figure 10 - Figure 13 are almost the same value, a classifier weights learning value to classify positive and negative. The results show in Figure 8 and Figure 9 as ECoHOG is better than CoHOG.

5. CONCLUSION AND FUTURE WORKS

In this paper, we analyzed state-of-the-art detection method, ECoHOG and CoHOG. We showed the effectiveness of detection approach on INRIA person dataset and Daimler pedestrian benchmark dataset. The experiment verified ECoHOG is better detection approach than CoHOG on the datasets. Moreover, analyzed characteristics of ECoHOG for machine learning based detection. ECoHOG gets a positive-positive pair together, and separate a positive-negative pair from each other by applying edge magnitude accumulation.

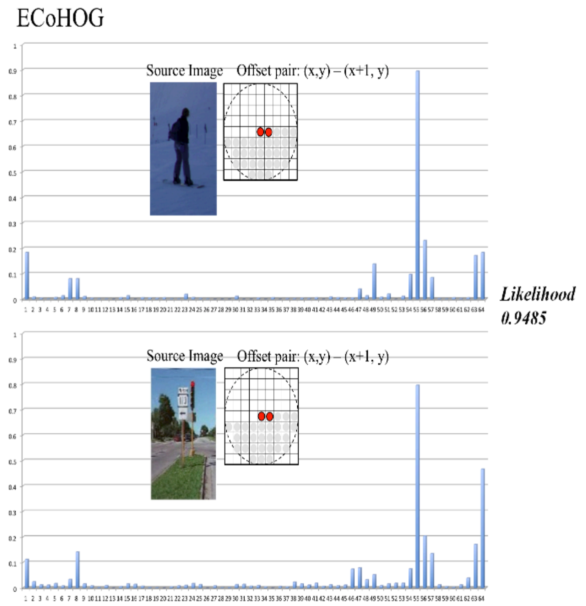

Fig. 12 ECoHOG feature vectors between human and similar background to human and the degree of histogram similarity (0.9485)

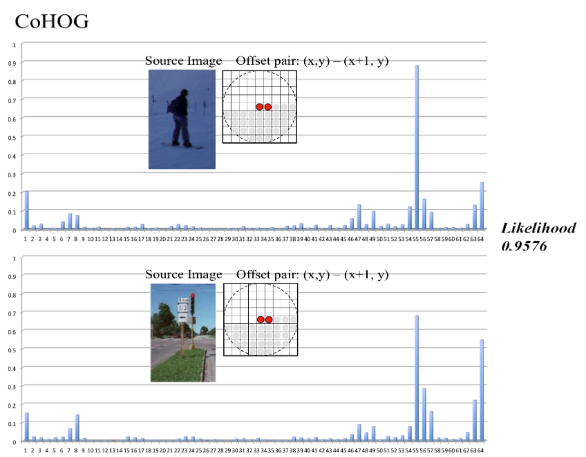

Fig. 13 CoHOG feature vectors between human and similar background to human and the degree of histogram similarity (0.9576)

In the future, we solve the multi-class classification problem, not only two class classification. Multi-class classification is important to apply general object categorization and activity recognition. And we also try to implement three or more pixels co-occurrence in edge based feature descriptor. Multi-cooccurrence feature generally make a large amount of dimensions which is worse in a machine learning situation. We simultaneously tackle the ”feature mining” problem to reduce processing time and feature dimension by using only effective feature in a feature vector.

REFERENCES

[1] T. B. Moeslund, A. Hilton, V. Kruger, L. Sigal, ”Visual Analysis of Humans: Looking at People”,Springer, 2011.

[2] T. Gandhi, M. M. Trivedi, ”Pedestrian Collision Avoidance Systems: A Survey of Computer Vision Based Recent Studies”, IEEE Intelligent Transportation Systems Conference (ITSC), pp.976-981, 2006.

[3] P. Dollar, C. Wojek, B. Schiele, P. Perona, ”Pedestrian Detection: An Evaluation of the State of the Art”, IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), Vol.34, No.4, pp.743-761, 2012.

[4] D. Geronimo, A. M. Lopez, A. D. Sappa, T. Graf, ”Survey of Pedestrian Detection for Advanced Driver Assistance Systems”, IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), Vol.32, No.7, pp.1239-1258, 2010.

[5] D. Geronimo, A. M. Lopez, ”Vision-based Pedestrian Protection Systems for Intelligent Vehicles”, ISBN 978-1-4614-7987-1, SpringerBriefs in Computer Science, 2014.

[6] T.Watanabe *et al*,”Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection”,PSIVT2009, pp.37-47, 2009

[7] H. Kataoka, K. Tamura, K. Iwata, Y. Satoh, Y. Matsui, Y. Aoki, ”Extended Feature Descriptor and Vehicle Motion Model with Tracking-by-detection for Pedestrian Active Safety”, IEICE Trans. Information and Systems, Vol.E97-D, No.2, pp.296-304, 2014.

[8] N. Dalal, B. Triggs: Histograms of oriented gradients for human detection, IEEE Conference on Computer Vision and Pattern Recognition, pp886-893(2005)

[9] INRIA Person Dataset, http://pascal.inrialpes.fr/data/human

[10] S. Munder, D. M. Gavrila, ”An Experimental Study on Pedestrian Classification”, IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), Vol.28, No.11, pp.1863-1868, 2006.

[11] Y. Yamauchi, M. Takagi, T. Yamashita, H. Fujiyoshi, ”Feature Co-occurrence Representation based on Boosting for Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)

[12] A. Bhattacharyya, ”On a measure of divergence between two statistical populations defined by their probability distributions”, Bulletin of the Calcutta Mathematical Society, Vol.35, pp.99-109, 1943.