

Co-occurrence of Intensity and Gradient Features for Object Detection

Akinori Hidaka¹ and Takio Kurita^{2,3}

¹ Tokyo Denki University

² National Institute of Advanced Industrial Science and Technology (AIST)

³ University of Tsukuba

Abstract. For the purpose of object detection, Haar-Like Features (HLF) proposed by Viola [13][14] are very famous. To classify images, usually HLF and its extensions used only image intensity. However, it is well known that the gradient information of image intensity is very important for the object recognition [2][9]. So in this paper, we propose a feature which uses both intensity and gradient informations. Our feature, called “Co-Occurrence Feature (COF)”, can treat the co-occurrence of salient regions in both of intensity domain and gradient domain. We use an extended image set that consists of original (intensity) image and oriented gradient images which are extracted from original images. COF is composed from a pair of arbitrary rectangles on arbitrary image channel in the extended image set. As a result of face/nonface classification experiments, it is confirmed that our feature has good classification performance, especially in the high true positive rate zone of ROC curves, the false detection rate is significantly better than Viola’s HLF.

1 Introduction

For the purpose of object detection, Haar-like features (HLF) proposed by Viola [13][14] are very famous and many extensions have been studied [4][5][6][7][8][10][12]. HLF is one of local image features and it is composed from several elemental rectangles in an image plane. The feature value of HLF is a difference of average intensity between elemental rectangles on an inputted image. HLF classifies the image by using the feature value as an input of simple classification methods, e.g. decision stumps or naive bayes. Since there are too many HLF in an image plane according to the degree of freedom of elemental rectangles’ coordinates, Viola proposed feature selection by Adaboost to choose a small subset of HLF which has powerful classification performance.

The original HLF [13][14] consists of up to four elemental rectangles of the same size which are neighboring each other, and the configuration patterns of rectangles were limited to four types in Fig. 1. Afterwards, the configuration patterns were extended to 14 types by Lienhart [8], and 19 types by Pham [12], to improve the classification power of HLF (Fig. 2). The extended HLFs are composed from rectangles of not always the same size, and it can represent more complex local structures of targets. Hidaka extended HLF to the set of



Fig. 1. Configurations of elemental rectangles of original HLF

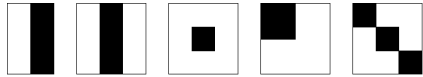


Fig. 2. Examples of configurations of extended HLF in [8] or [12]

several arbitrary rectangles in the image plane [5]. The feature can catch a relation between not only the close regions but also distant regions, and it can represent complex structures of targets more flexibly. Mita proposed an extended classifier constructed from multiple HLF [10][11]. The classifier is based on the co-occurrence of HLF and made it possible to construct an effective classifier.

It is well known that the gradient information of image intensity is very important for many image recognition tasks. SIFT descriptor proposed by Lowe [9] and Histograms of Oriented Gradients (HOG) proposed by Dalal [2] achieved strong performance in many computer vision applications, such as matching, retrieval, classification, detection, tracking and motion recognition. These descriptor or feature calculate oriented gradient informations of image intensity. The gradient informations are accumulated into histograms of quantized edge orientations. The histogram is made from a certain local measurement region (called “block”) in an image. In the case of HOG, usually a thousands of histograms are calculated from a set of many blocks which are covering the image entirely. Therefore, usually HOG is not computationally efficient. Zhu proposed feature selection for HOG blocks based on Adaboost [3], and they showed it is possible to achieve comparable classification power by using very small number of blocks compared with the case of usual HOG [15].

Perhaps to suppress the computational cost and system complexity, most of described methods employed either intensity or gradient informations. However, the additional information channels of images will lead to improvement of classification performance. So in this paper, we propose a feature which uses both intensity and gradient informations. We consider the extended image set $\{I_0, \dots, I_B\}$ where B is the number of quantized edge orientation and $\{I_1, \dots, I_B\}$ is the set of oriented edge images which are made from original image I_0 . Our feature handles the co-occurrence of rectangles of images, just like [5][10][11]. As a novel point, our feature is composed from a pair of arbitrary rectangles on arbitrary channels in the extended image set. In this paper, we call proposal feature “Co-Occurrence Feature (COF)”.

In our case, the total number of elemental rectangles increases according to the increasing of image channel. Hidaka showed that it is possible to achieve comparable classification power against Viola’s method by using a randomized small subset of HLFs for the feature selection of Adaboost [4]. As well as such method, we use random candidates approach; For the feature selection training by Adaboost, the set of candidate COF is made by selecting a pair of arbitrary rectangles from arbitrary image channel, at random. To maintain a variety of the features, the subset of COF will be regenerated in each iteration of Adaboost.

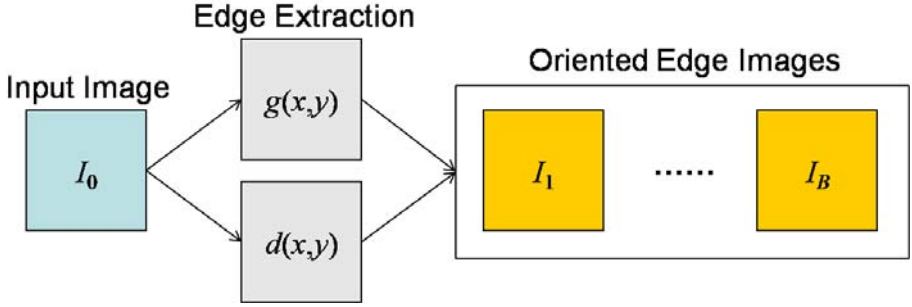


Fig. 3. The making process of the extended image set. The process of edge extraction is performed by using Eqs. (1) to (4). Given an image length W and height H , $g(x, y)$ and $d(x, y)$ ($1 \leq x \leq W, 1 \leq y \leq H$) represent the gradient magnitude and gradient orientation of a pixel (x, y) , respectively. The oriented edge images are made from $g(x, y)$ and $d(x, y)$ by using Eq. 5. As shown in Eq. 5, $i_b(x, y)$, the pixel value of b -th oriented edge image I_b , is the weighted gradient magnitude where the weight $\cos |d(x, y) - c_b|$ indicates the degree of similarity between central angle of b -th orientation bin and actual gradient orientation of a pixel (x, y) .

We valid the performance of COF by face/nonface classification experiments. Our feature has good classification performance compared with Viola’s HLF, especially in the zone of high true positive rate of ROC curve, the false detection rate is significantly better.

2 Co-Occurrence Feature for Extended Image Set

We propose the novel image feature, called Co-Occurrence Feature (COF), for the extended image set.

2.1 Extended Image Set

At first, oriented edge image set $\{I_1, \dots, I_B\}$ is made from original image I_0 (B is the number of bins of the edge orientation to be quantized). The central angle of b -th orientation bin is obtained as $c_b = (b - 1) * \frac{180}{B}$. To make each edge image I_b , every pixels $i(x, y)$ of I_0 is filtered by 3×3 sobel filter, and then the edge magnitude $g(x, y)$ and edge direction $d(x, y)$ are calculated as follows:

$$g_v(x, y) = \sum_{k=-1}^{+1} 2^{1-|k|} (i(x+1, y+k) - i(x-1, y+k)), \quad (1)$$

$$g_h(x, y) = \sum_{k=-1}^{+1} 2^{1-|k|} (i(x+k, y+1) - i(x+k, y-1)), \quad (2)$$

$$g(x, y) = \sqrt{g_v^2(x, y) + g_h^2(x, y)}, \quad (3)$$

$$d(x, y) = \arctan \frac{g_h}{g_v}. \quad (4)$$

The pixel value $i_b(x, y)$ of b -th oriented edge image I_b is calculated as follows:

$$i_b(x, y) = g(x, y) \cos |d(x, y) - c_b|. \quad (5)$$

As a result, the extended image set $\{I_0, I_1, \dots, I_B\}$ is obtained.

2.2 COF on Extended Image Set

In our method, feature extraction is performed on the extended image set $\{I_0, I_1, \dots, I_B\}$. A elemental rectangle r_b on I_b is defined as follows:

$$r_b = (x_s, y_s, x_e, y_e) = \{(x, y) \in I_b | x_s \leq x \leq x_e, y_s \leq y \leq y_e\} \quad (6)$$

where (x_s, y_s) and (x_e, y_e) are the diagonal apexes of the rectangle. A COF is composed from a set of R elemental rectangles $\{r_{b_p}^p\}_{p=1}^R$. Fig. 4 illustrates the difference between HLF and COF.

The feature value of COF is based on the average pixel values $\{m(r_{b_p}^p)\}_{p=1}^R$ on the elemental rectangles $\{r_{b_p}^p\}_{p=1}^R$. The average pixel value on a rectangle r can be calculated as follows:

$$m(r_b) = \iint_{r_b} v(x, y) dx dy \quad (7)$$

where $v(x, y)$ is $i(x, y)$ if $b = 0$ and $i_b(x, y)$ in Eq. (5) if $b > 0$.

HLF employed the difference of average intensity of positive (white) rectangles and negative (black) rectangles, as a feature value. Therefore HLF has

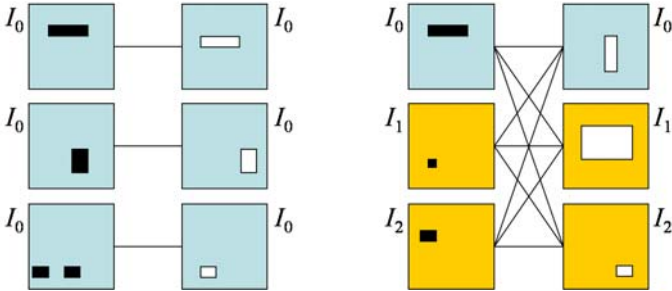


Fig. 4. The difference of HLF and COF. (Left) HLF is composed from only the combinations of the same size and neighboring rectangles on the intensity channel. (Right) On the other hand, COF is composed from arbitrary rectangles in arbitrary (i.e., intensity and oriented edge) channels.

robustness about additive illumination changes. On the contrary, in our case, the feature value will be made from different information channels (i.e. intensity and oriented edge). It is hard to consider that a robustness for additive change of intensity image and oriented edge image is useful in a real applications. So we should not adhere about the difference of the averages.

In this paper, we make the feature value of COF by two simple ways: to take the “sum” and the “difference” of the average pixel values of elemental rectangles. In the feature selection training, the classifiers based on each feature type are independently trained, and the most superior one is adopted.

In the case of “difference” type, a combinatorial problem of elemental rectangles is caused. However, because of the computational complexity problem indicated in Sec. 3.1, we set the number of elemental rectangles $R = 2$. Therefore the combinatorial problem is not occurred in this paper.

The feature values of COF $\{r_{b_1}^1, r_{b_2}^2\}$ for an image I is represented as follows:

$$f^+(I) = f^+(\{I_b\}_{b=0}^B) = m(r_{b_1}^1) + m(r_{b_2}^2) \quad (8)$$

$$f^-(I) = f^-(\{I_b\}_{b=0}^B) = m(r_{b_1}^1) - m(r_{b_2}^2) \quad (9)$$

The classification function of COF for an image I is written as follows:

$$h(p, \theta, f(I)) = \begin{cases} 1 & \text{if } p(f(I) - \theta) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where $p \in \{1, -1\}$ and $\theta \in R$ are parameters determined by training (Fig. 6).

3 Adaboost for COF Selection

Adaboost [3] is the ensemble learning method that trains multiple base-classifiers and assembles them to create a more powerful classifier. Viola used HLF as base-classifiers for the face detection [13][14]. The algorithm of Viola’s Adaboost is composed from two phases as follows:

1. Weak training: optimizes all HLF independently,
2. Feature selection: selects the best one from those optimized HLF.

Therefore the training time of this algorithm depends on the number of HLF.

The training algorithm for COF is described in the following sections.

3.1 Random Candidates Approach for COF Selection

In $W_b \times H_b$ pixels image, there are $P_b = W_b(W_b + 1) \times H_b(H_b + 1)/4$ elemental rectangles. Therefore the total number of elemental rectangles on the extended image set is $P = P_0 + P_1 + \dots + P_B$. As described in Sec. 2.2, COF is composed from R arbitrary elemental rectangles. So the total number of COF is $K = P^R$. For example $B = 2$, $R = 2$ and $W_0 = H_0 = 19$, the case that parameters about training condition are pretty small, K becomes about 7 billion. This is 10,000 times larger than the number of Viola’s HLF, and their feature

-
- Input N labeled samples $\{I_i, y_i\}_{i=1}^N$ where I_i is a sample and y_i is a label.
 - $y_i = 1 \Leftrightarrow I_i$ is a positive sample, $y_i = 0 \Leftrightarrow I_i$ is a negative sample.
 - $w_i > 0$, $\sum_{i=1}^N w_i = 1$.
 - Initialize samples weights: if $y_i = 1$ then $w_i = \frac{1}{2p}$, otherwise $w_i = \frac{1}{2q}$.
(p, q : # of positive samples and negative samples, respectively)
 - for $t = 1, \dots, T$
 - Normalize weights: $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{i=1}^N w_{t,i}}$.
 - Make the subset of COF at random: $\{h_1, h_2, \dots, h_{10000}\}$.
 - for $c = 1, \dots, 10000$
 - * Perform weak-training to COF $h_c(p, \theta, f^+)$.
 - * Evaluate weighted error rate $we(h_c(p, \theta, f^+))$ of optimized $h_c(p, \theta, f^+)$.
 - * Optimize COF $h_c(p, \theta, f^-)$.
 - * Evaluate weighted error rate $we(h_c(p, \theta, f^-))$ of optimized $h_c(p, \theta, f^-)$.
 - * Set $(h_c, we_c) = (h_c(p, \theta, f^+), we_c^+)$;
If $we^- < we^+$ then set $(h_c, we_c) = (h_c(p, \theta, f^-), we_c^-)$.
 - Select the t -th base-classifier: $b_t = \arg \min_{h_c} we(h_c)$, $e_t = \min_{h_c} we(h_c)$.
 - Compute $\alpha_t = \log((1 - e_t)/e_t)$.
 - Update samples weights: $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot \delta(y_i - b_t(I_i))]$ where
 $\delta(x) = 1$ if $x = 0$, or $\delta(x) = 0$ otherwise.
 - Final classification function is

$$H(I) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t b_t(I) \geq \Theta \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise.} \end{cases} \quad (\Theta: \text{Threshold})$$
-

Fig. 5. Adaboost for COF selection. The weak-training algorithm is shown in Fig. 6.

-
- Input COF $h(p, \theta, f)$.
 - Input N labeled and weighted samples $\{I_i, w_i, y_i\}_{i=1}^N$ where I_i , w_i and y_i is a sample, a weight and a label.
 - $y_i = 1 \Leftrightarrow I_i$ is a positive sample, $y_i = 0 \Leftrightarrow I_i$ is a negative sample.
 - $w_i > 0$, $\sum_{i=1}^N w_i = 1$.
 - Let w^+ and w^- be sum of weights of positive and negative samples, respectively.
 - Extract feature value $f_i = f(I_i)$ from every samples.
 - Make up-sorted array $\{g_j\}_{j=1}^N$ from $\{f_i\}_{i=1}^N$.
 - Set $\theta = 0$, $e^+ = w^-$, $e^- = w^+$, $\theta_{best}^+ = \theta_{best}^- = 0$, $e_{best}^+ = e_{best}^- = 1$.
 - For $j = 1, \dots, N - 1$
 - Set $\theta = \frac{1}{2}(g_j + g_{j+1})$.
 - If $y_j = 1$, then
 - * Set $e^+ \leftarrow e^+ + w_j$ and $e^- \leftarrow e^- - w_j$.
 - * If $e^- < e_{best}^-$ then set $e_{best}^- \leftarrow e^-$ and $\theta_{best}^- \leftarrow \theta$.
 - Else if $y_j = 0$, then
 - * Set $e^+ \leftarrow e^+ - w_j$ and $e^- \leftarrow e^- + w_j$.
 - * If $e^+ < e_{best}^+$ then set $e_{best}^+ \leftarrow e^+$ and $\theta_{best}^+ \leftarrow \theta$.
 - If $e_{best}^+ < e_{best}^-$ then set $p = +1$ and $\theta = \theta_{best}^+$, else set $p = -1$ and $\theta = \theta_{best}^-$.
-

Fig. 6. The algorithm of weak-training for COF

selection algorithm can not handle such the huge classifier candidates. Therefore, we use random candidate approach like in [4]. In our experiment, 10,000 features (HLF for Viola’s method and COF for our method) are generated at random as the candidate of a base-classifier. After the feature optimization, the best feature in those candidates is selected as the base-classifier. The random candidates are regenerated in each iteration of Adaboost to keep the variety of features.

Figs. 5, 6 show the algorithm of Adaboost by random candidates approach.

4 Experiments

In this paper, we used MIT CBCL face database [1] for our experiments. The database consists of 2,901 facial and 28,121 background 19×19 pixel images. We divided all images into training and test set; 2,000 face and 5,000 background images are used for training, and the remaining images are used for testing. Five pairs of training and test sets were generated at random.

We used four types HLF illustrated in Fig. 1 and then the total number of HLF becomes about 53,000. The parameters of COF are fixed to $B = 2$ and $R = 2$. There are about 83,000 elemental rectangles in the extended image set, so we can use about 7 billion ($83,000^2$) COF potentially.

In each iteration of Adaboost for HLF/COF training, the set of 10,000 HLF or COF as the base-classifier candidates is generated at random. The iteration steps of Adaboost (equivalent to the number of the base-classifiers) is up to 200.

Fig. 7 shows the false positive rate (FP) and false negative (FN) rate when true positive rate (TP) equals 99% and true negative rate (TN) equals 99%. When $TN = 99\%$, HLF and COF have comparable results. However, when $TP = 99\%$, the results of COF are significantly better than the results of HLF. In the case of 200 classifiers, FP rate of COF is improved 5% against HLF.

Fig. 8 shows the ROC curves for the five test sets created by using 200 HLF or COF. It can be seen that there are no significant difference between the results of HLF and COF when TP is smaller than 0.95. On the contrary, especially when TP is larger than 0.99, the results of COF are significantly better.

| RCS-10000 | 50 classifiers | | 100 classifiers | | 200 classifiers | |
|------------------|----------------|-------|-----------------|-------|-----------------|------|
| | HLF | COF | HLF | COF | HLF | COF |
| FP (%) at TP 99% | 23.39 | 19.47 | 17.32 | 12.93 | 12.32 | 7.35 |
| FN (%) at TN 99% | 9.90 | 10.61 | 6.84 | 6.36 | 4.83 | 4.00 |

Fig. 7. Average error rates (%) for negative (positive) samples at TP (TN) = 99%

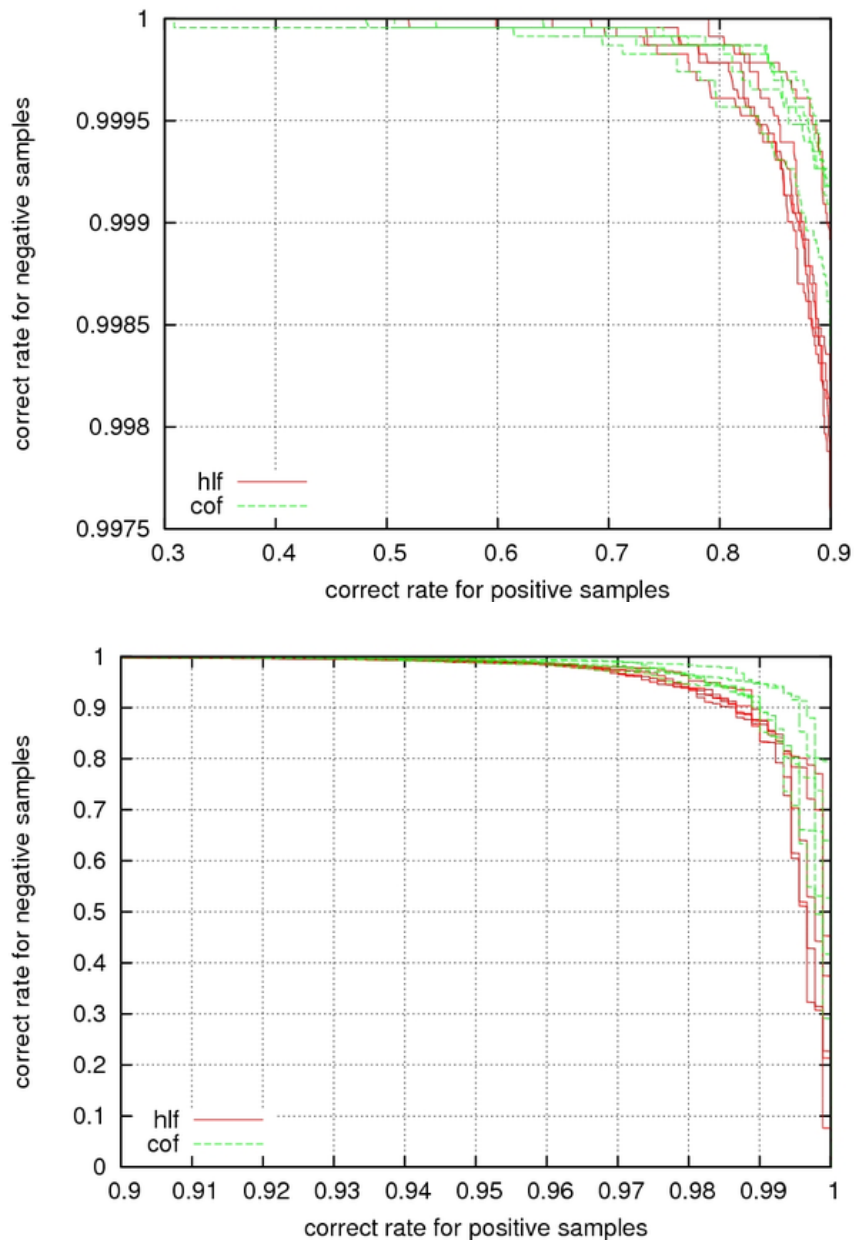


Fig. 8. ROC curves for five test sets. Red and green curves show the results of HLF and COF, respectively. Both methods used 200 weak classifiers. Horizontal axis shows the true positive (TP) rate and vertical axis shows the true negative (TN) rate. Top and bottom figures show different TP zones (from 0.3 to 0.9 and from 0.9 to 1.0).

5 Conclusions

In this paper, we proposed a new image feature which uses both intensity and gradient informations and can treat co-occurrence of salient regions in both of the information domains. In our experiments, we showed that the more powerful features than the best HLF are certainly included in our 7 billion of COF. We also showed that the feature selection based on random candidates manner can efficiently find such powerful features from those extremely abundant candidates. As a result, the classification accuracy is improved compared with Viola's HLF.

References

1. MIT Center For Biological and Computation Learning. CBCL Face Database #1, <http://www.ai.mit.edu/projects/cbcl>
2. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: Proc. of CVPR 2005, pp. 886–893 (2005)
3. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. *J. of Computer and System Sciences* 55(1), 119–139 (1997)
4. Hidaka, A., Kurita, T.: Fast Training Algorithm by Particle Swarm Optimization and Random Candidate Selection for Rectangular Feature Based Boosted Detector. In: Proc. of IJCNN 2008 on WCCI 2008, pp. 1164–1170 (2008)
5. Hidaka, A., Kurita, T.: Non-Neighboring Rectangular Feature Selection using Particle Swarm Optimization. In: Proc. of ICPR 2008 (December 2008)
6. Huang, C., Ai, H., Li, Y., Lao, S.: High-performance rotation invariant multiview face detection. *IEEE Trans. on PAMI* 29, 671–686 (2007)
7. Li, S.Z., Zhu, L., Zhang, Z., Blake, A., Zhang, H., Shum, H.: Statistical learning of multi-view face detection. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 67–81. Springer, Heidelberg (2002)
8. Lienhart, R., Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection. In: Proc. of IEEE ICIP, vol. 1, pp. 900–903 (2002)
9. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. of Computer Vision* 60(2), 91–110 (2004)
10. Mita, T., Kaneko, T., Hori, O.: Proc. of ICCV 2005, pp. 1619–1626 (2005)
11. Mita, T., Kaneko, T., Stenger, B., Hori, O.: Discriminative Feature Co-occurrence Selection for Object Detection. *IEEE Trans. PAMI* 30(7), 1257–1269 (2008)
12. Pham, M., Cham, T.: Fast training and selection of Haar features using statistics in boosting-based face detection. In: Proc. of ICCV 2007, pp. 1–7 (2007)
13. Viola, P., Jones, M.: Rapid Object Detection Using a Boosted Cascade of Simple Features. In: Proc. of CVPR 2001, December 2001, vol. 1, pp. 511–518 (2001)
14. Viola, P., Jones, M.: Robust real-time face detection. *Int. J. of Computer Vision* 57(2), 137–154 (2004)
15. Zhu, Q., Avidan, S., Yeh, M., Cheng, K.: Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. In: CVPR 2006, vol. 2, pp. 1491–1498 (2006)