# Feature Co-occurrence Representation Based on Boosting for Object Detection

Yuji Yamauchi[1], Masanari Takaki[2], Takayoshi Yamashita[1], Hironobu Fujiyoshi[1]
[1]Chubu University, [2]DENSO CORPORATION
yuu@vision.cs.chubu.ac.jp, MASANARI_TAKAGI@denso.co.jp
takayoshi@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

## Abstract

*This paper proposes a method of feature co-occurrence representation based on boosting for object detection. A previously proposed method that combines multiple binary-classified codes by AdaBoost to represent the co-occurrence of features has been shown to be effective in face detection. However, if an input feature is difficult to be assigned to a correct binary code due to occlusion or other factors, a problem arises here since the process of binary classification and co-occurrence representation may combine features that include an erroneous code. In response to this problem, this paper proposes a Co-occurrence Probability Feature (CPF) that combines multiple weak classifiers by addition and multiplication arithmetic operators using Real AdaBoost in which the outputs of weak classifiers are real values. Since CPF combines classifiers using two types of operators, diverse types of co-occurrence can be represented and improved detection performance can be expected. To represent even more diversified co-occurrence, this paper also proposes co-occurrence representation that applies a subtraction arithmetic operator. Although co-occurrence representation using addition and multiplication operators can represent co-occurrence between features, use of the subtraction operator enables the representation of co-occurrence between local features and features having other properties. This should have the effect of revising the probability of the detection-target class obtained from local features. Evaluation experiments have shown co-occurrence representation by the proposed methods to be effective.*

## 1. Introduction

Detecting object in image is being researched for application in a wide variety of fields including surveillance and ITS. In particular, many methods for detecting pedestrians have been proposed in recent years with the aim of improving accuracy.

As reflected by the face-detection framework proposed by Viola and Jones [17], object-detection methods generally combine local features representing the appearance of an object with statistical learning. The appearance of an object targeted for detection can change due to lighting, pose, viewpoint, etc., and to deal with these changes, features that focus on local region in the image have been proposed.

Local features that have been proposed include differences in intensities [17], edges [7][4][19][16], and texture [10][18]. For local features like these, the information contained by any one feature is very small, but the appearance of a target object can be represented by a very large number of features. In addition, a high-accuracy detector can be achieved through training by using a statistical-learning method to extract from a huge feature pool those that are common to training samples of target objects.

Many previous works are therefore making use of features obtained from local region to deal with aspects of appearance common to target objects. However, as the information included in a local region is small, only part of the appearance of a target object can be represented. As a result, background image similar to the partial appearance of a target object might be erroneously detected. To solve this problem, object-detection methods that consider the co-occurrence of multiple local features have been proposed in recent years [8][14][1]. These methods can capture not only partial appearance but also a global structure of object making for high-accuracy detection.

Also proposed in recent years are image categorization methods that make use of context [13][5][12]. These methods use category-classification results determined from appearance to evaluate the relationship between categories and revise category output. This makes for high-accuracy category classification even in the case of ambiguous appearance. However, some training samples may include co-occurrence representation even for objects in categories that have a low probability of existing simultaneously. If such samples should be included in test images, the possibility of erroneous classification as a result of co-occurrence representation is high. This is typical problems due to the representation of very strong relationships such as co-

978-1-4244-7028-0/10/$26.00 ©2010 IEEE

occurrence between categories. To deal with this problem, co-occurrence representation capable of even higher generalization is needed.

In this paper, we propose a Co-occurrence Probability Feature (CPF) that combines weak classifiers of Real AdaBoost [15][19] by arithmetic operator. A CPF is a discriminative feature achieved by combining multiple local features through a boosting. Here, the combining of weak classifiers by multiple types of different operators enables the representation of diverse types of co-occurrence. In addition, to represent even more diverse types of co-occurrence, we also propose co-occurrence representation that applies a subtraction operator. By using a co-occurrence representation method that applies a subtraction operator, we can expect co-occurrence such that the probability of the object class is revised from information having other properties.

This paper is organized as follows. We summarize related works in Section 2. We describe our proposed method, co-occurrence probability feature in Section 3, and report the experimental results in Section 4. Section 5 then describes the CPF extension that applies the subtraction operator and Section 6 concludes the paper.

## 2. Related work

Many methods have been recently proposed with the aim of improving classification accuracy by focusing on co-occurrence. This section surveys the research associated with those methods from two viewpoints: the co-occurrence of categories and the co-occurrence of features.

First, we look at methods that consider the co-occurrence of different categories [13][5][12]. Using categories determined and their degree of reliability from appearance, these methods revise classification results from the relationship between those categories. For example, if an "automobile" exists in the input image, the low possibility of a "cow" running alongside and the high possibility of a "motorcycle" running alongside can each be represented as a co-occurrence between categories. This approach enables a sample that would otherwise be difficult to classify based only on appearance to be correctly classified. However, as some training samples may include a co-occurrence representation of object categories that have a low probability of existing simultaneously, the possibly of a negative effect on classification accuracy exists.

Next, we look at methods that consider the co-occurrence of features [8][14][9]. Mita *et al.* have proposed a joint Haar-like feature [8] based on the co-occurrence of multiple Haar-like features. A joint Haar-like feature combines binary codes from multiple Haar-like features through boosting. As a result, the relationship between features based on structures such as the eye, nose, and mouth of a face can be determined. This makes erroneous detections difficult even if the background image should include a section similar to part of a face. However, in the case that an erroneous binary code is included in the features to be combined, the joint Haar-like feature will be adversely affected, and the classification accuracy deteriorates. This problem is thought to be especially prevalent for target objects like human whose shape can undergo dramatic changes and for objects affected by occlusion due to overlapping in the image. Sabzmeydani *et al.* have proposed a shapelet feature [14] that combines edge features in four directions by boosting. A shapelet feature can simultaneously capture edges that co-occur in a target object and edges that do not. Also proposed is a joint HOG feature [9] that aims to increase the accuracy of classification even further by combining the advantages of joint Haar-like features with those of shapelet features. These methods can generate discriminative feature effective for object detection that evaluate feature co-occurrence through boosting. Other methods that consider co-occurrence include local binary pattern (LBP)[11], which represents the intensity relationship between adjacent pixels, and methods that apply LBP [10][18]. However, LBP is not necessarily effective features for object detection because combinations to evaluate co-occurrence are prepared.

The basic idea of our approach presented in this paper is embodied by the second viewpoint above, that is, by the type of method that considers the co-occurrence of features. In conventional methods of this type, classification results obtained by boosting are binarized and combined to represent co-occurrence. As a result, the combination of erroneous classification results could have an adverse effect on final results. In this paper, we deal with this problem by combining weak classifiers from Real AdaBoost. Weak classifiers are combined by addition and multiplication operators having different properties, thus this approach enables the representation of co-occurrence with some fuzziness. And with the aim of improving detection performance even further, the first idea presented above of using the co-occurrence of categories is also deemed important. Therefore, in this study, we also use a subtraction operator to represent co-occurrence such that the probability of the target-object class is revised by information having other properties.

## 3. Co-occurrence Probability Feature(CPF)

In this section, we describe CPF for representing the co-occurrence of features and an efficient training for generating CPF.

### 3.1. Histograms of Oriented Gradients

The proposed CPF is a feature that combines multiple local features. In this study, we use Histograms of Oriented Gradients (HOG) [4] as local features. HOG have been reported to be effective in object detection, and many pedestrian detection methods using HOG [20][6] have been proposed because of the high performance that HOG offers. Since HOG features convert gradients with adjacent pixels into histograms at each local area, they are not easily affected by lighting and are robust to local geometric changes.

In the process for computing HOG features, 9-direction histograms of oriented gradients are created for each cell area ($8 \times 8$ pixels) in a detection window ($64 \times 128$ pixels). The features in each block area ($2 \times 2$ cells) are then normalized by the norm-L2 technique. This block-by-block normalization process shifts one cell at a time so that blocks overlay each other. The HOG after normalization $V_c^{HOG}$ is expressed as $V_c^{HOG} = \{v_c^{HOG}(1), v_c^{HOG}(2), \cdots, v_c^{HOG}(B \times N \text{ blocks})\}$. Here, $c$ denotes cell, $B$ the number of blocks, and $N$ the number of gradient directions. In the end, $3,780$ features can be obtained from one detection window in this way.

### 3.2. Probability of Detection-target Class

To generate CPF, the probability of the detection-target class is computed when inputting local features. In this study, we use the weak classifiers of Real AdaBoost [15][19] to compute the probability of the detection-target class. The output of weak classifiers can be used to obtain a statistical label confidence by using a large training set. First, the probability density functions $W_y$ is created using local features obtained from training samples. Probability density functions $W_\pm$ are represented by histograms and created by Eq. (1) and (2).

$$W_+^j = \sum_{i:j \in J \wedge y_i = +1} D_t(i) \tag{1}$$

$$W_-^j = \sum_{i:j \in J \wedge y_i = -1} D_t(i) \tag{2}$$

Here, $D_t(i)$ is a weight of a sample, $y \in \{+1, -1\}$ is the class label, $j$ is bin number of histogram.

Next, to obtain the probability of the detection-target class, weak classifier $f(v)$ is computed by

$$f(v) = \frac{1}{2} \ln \frac{W_+^j + \epsilon}{W_-^j + \epsilon}, \tag{3}$$

where $v$ is a local feature for input and $\epsilon$ is a coefficient to prevent a zero from occurring in the denominator. Weak classifier $f(v)$ expresses a statistical reliability.

### 3.3. Generation of CPF($+$ and $\times$)

To represent the co-occurrence of features, CPFs are generated using two types of operators having different

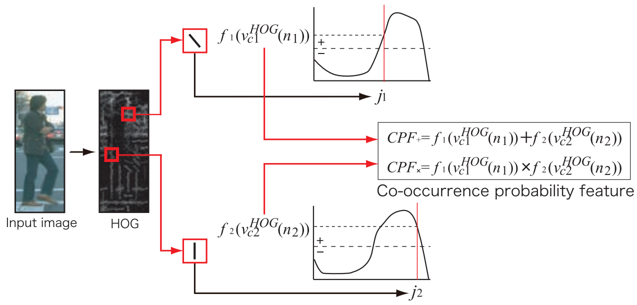


Figure 1. Representation of co-occurrence of HOG features. CPF is computed by combining weak classifiers $f(x)$ that input local features using an addition or multiplication operator.

properties. Specifically, they are generated by Eq. (4) and (5) using weak classifiers $f(x)$.

$$CPF_+ = f_1(v_{c_1}^{HOG}(n_1)) + f_2(v_{c_2}^{HOG}(n_2)) \tag{4}$$

$$CPF_\times = f_1(v_{c_1}^{HOG}(n_1)) \times f_2(v_{c_2}^{HOG}(n_2)) \tag{5}$$

The addition (+) operator expresses co-occurrence representing a weak relationship between features and the multiplication ($\times$) operator expresses co-occurrence representing a strong relationship between features. A CPF that combines weak classifiers by an addition operator is a feature that captures two features in a comprehensive way. Thus, if one feature should be occluded or affected by external disturbances such as noise but the other feature happens to represent the target class in a big way, the CPF will end up reflecting the target class. On the other hand, a CPF that combines weak classifiers by a multiplication operator is a feature that captures the simultaneity of two features. It is a feature that captures that target only when those two features $f1(v), f2(v)$ are both high. An example of feature co-occurrence is shown in Fig. 1. In this example, in which the detection target is pedestrian, the co-occurrence of a shoulder gradient and a torso gradient is represented. In methods that do not represent co-occurrence, erroneous detections would easily occur by reacting only to gradients that deal with the human torso. The proposed technique, however, by simultaneously observing the torso and shoulder gradients and representing their co-occurrence by an operator, can supress the erroneous detection of such samples.

### 3.4. Training by Nested Real AdaBoost

The CPFs described in section 3.3 are generated by combining local features. Weak classifiers can now be selected from a weak-classifier pool by evaluating generated CPFs by Real AdaBoost. It is difficult, however, to evaluate all CPFs by Real AdaBoost due to the huge number of combinations. Thus, to perform efficient combination of local features in this study, training is performed by the nested Real AdaBoost scheme shown in Fig. 2. The our training method by nested Real AdaBoost is given in **Algorithm 1**.

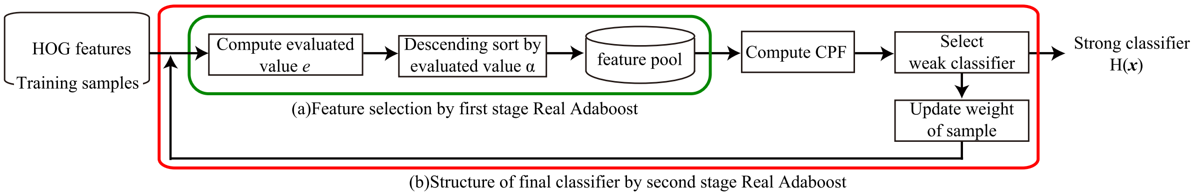


Figure 2. Training flow by nested Real AdaBoost.


Figure 3. Database examples.

**Algorithm1** Training algorithm by nested Real AdaBoost.

**Input:** Labeled training samples $\{x_i, y_i\}_{i=1\cdots N}, y_i \in \{0,1\}$.
**Initialization:** Initialize weights $D_1(i) = 1/N$.
**Training:**
For $t = 1, \cdots, T$.
·Select $CPF$ by Real AdaBoost in 3.4.1.
·Train a weak classifier $h_t(x)$ based on $CPF$.
·Update the sample weights:
$D_{t+1}(i) = \frac{D_t(i)[-y_i h_t(x)]}{Z_t}$
where $Z_t$ is a normalization factor.
**Output:**Final strong classifier $H(x) = \text{sign}\left[\sum_{t=1}^{T} h_t(x)\right]$

#### 3.4.1 Efficient Feature Selection by Real AdaBoost

Evaluating all local-feature-combining CPFs is difficult due to the large computational cost involved. To get around this problem, a feature pool can be prepared consisting of local features deemed effective in classification and CPFs can then be generated using only the local features in that pool. This approach can reduce computational cost while maintaining classification performance.

A weak classifier that inputs a local feature is given by Eq. (3). The classification performance of a weak classifier can be evaluated by the Bhattacharyya distance between the probability density functions (Eq. (1) and (2)) of the detection-target and non-detection-target classes. Specifically, error $e$ of a weak classifier is computed by Eq. (6).

$$e = \sum_j \sqrt{W_+^j W_-^j} \tag{6}$$

Table 1. Databases.

| Class | Training | | Testing | |
|---|---|---|---|---|
| | Pos. | Neg. | Pos. | Neg. |
| Pedestrian | 1,215 | 10,416 | 1,836 | 2,108 |
| Vehicle | 710 | 8,800 | 1,230 | 3,880 |

This error is calculated for all local features. Those with small error are selected for feature candidates to compute CPF. Then, CPF is generated using the feature candidates, and the sample weights are updated by the second Real AdaBoost. A strong classifier is trained through frequent repetition of above processes.

## 4. Evaluation Experiment

This section describes experiments to confirm the effectiveness of the addition and multiplication operators in CPF.

### 4.1. Database

A number of databases have become publically available in recent years for evaluating the performance of object-detection methods. In the field of pedestrian detection, for example, use of the INRIA database [4] has become common. However, as we will be describing object detection using a geometric context in section 5, we cannot perform an evaluation experiment using the INRIA database here. To perform a unified evaluation experiment, we decided to create a new database in which detection targets are taken to be pedestrians and vehicles.

This image database was created using HD video taken by an on-vehicle camera during the day and in the evening

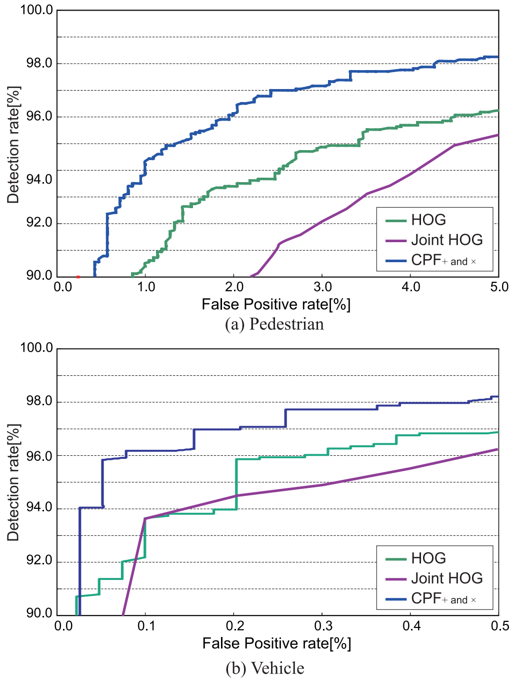


Figure 4. ROC curves of experimental results.

hours. Image samples were obtained by extracting people and vehicles from the video. The extracted images included various types of background and lighting, various poses of the detection target, partial occlusion due to overlapping of detection targets in the image, etc. Fig. 3 shows some of the samples used for training and Table 1 lists the number of pedestrian and vehicle samples used for training and testing.

### 4.2. Experiment

To compare the effectiveness of feature co-occurrence representation as proposed here, we compared results obtained by CPF with those obtained by HOG feature[4] and Joint HOG feature[9], the latter of which is a feature co-occurrence representation of state-of-the-art approaches.

Experiment results are compared using receiver operating characteristic (ROC) curves in which the false positive rate is shown on the horizontal axis and the detection rate on the vertical axis. A ROC curve indicates better detection performance as it approaches the upper-left area of the graph.

### 4.3. Experimental Results

The ROC curves of experimental results are shown in Fig. 4. These results show that the detection performance of the proposed method is the highest. First, we compare the proposed method with HOG feature. In the pedestrian test, we see that the proposed method shows an improvement of $3.7\%$ over HOG and Joint HOG features at a false positive rate of $1.0\%$, and in the vehicle test, an improvement of $2.4\%$ at a false positive rate of $0.1\%$. These results demonstrate the effectiveness of feature co-occurrence representation by the proposed method in object detection.

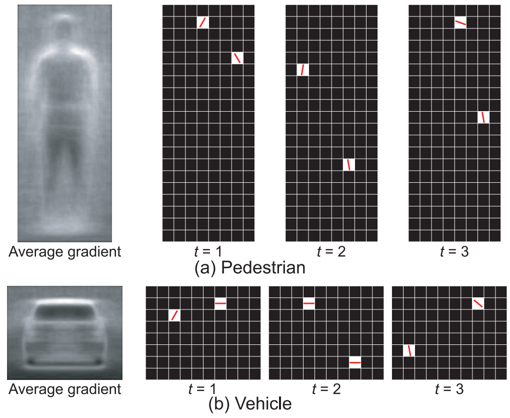


Figure 5. Visualization of CPF selected by training. On the left are average gradient images computed from training samples in the detection-target class. On the right are visualization results of CPF selected by the Real AdaBoost training.

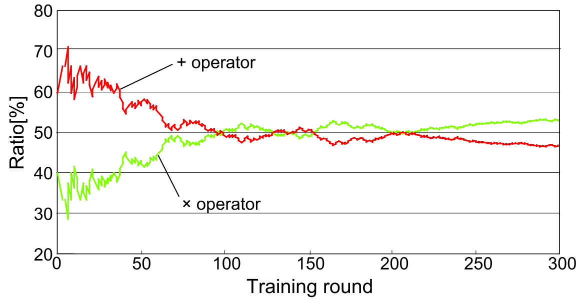


Figure 6. Ratios of selected operator by boosting.

Fig. 5 shows average gradient images computed from training samples of the detection target class and visualization images of CPF selected by training. The visualization images on the right show that the CPF selected by the training reflect the shape of the target object in both the pedestrian and vehicle cases. For example, at the first round weak classifier ($t = 1$) in the pedestrian case, the selected CPF simultaneously capture the right part of the head and the left shoulder of a pedestrian can be seen. In pedestrian detection, erroneous detection frequently occurs by reacting to side edges that represent a torso. We consider that this kind of erroneous detection can be controlled by representing the co-occurrence of features.

Fig. 6 shows how the ratio of operators selected during training changes. In the initial training period, more CPFs using the addition operator are selected than CPFs using the multiplication operator, but from about 100 training rounds on, the two ratios become about the same. This is thought to occur for the following reason. In the initial training period, many addition operators are selected so that many training samples can be correctly classified. In the middle

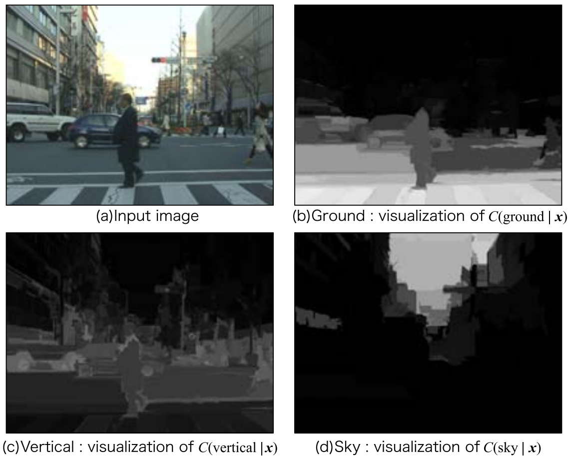


Figure 7. Example of geometric confidence map. Intensity in (b), (c), and (d) indicates confidence level.

of training, however, the multiplication operator is increasingly selected so that training samples with high weights can be correctly classified. From these results, we can say that these two types of co-occurrence representation have different properties and that operators optimal for classification are selected by boosting.

## 5. CPF using a Subtraction Operator

The CPFs described so far represent the co-occurrence of features by combining multiple local features by an addition or multiplication operator. In this section, with the aim of representing an even greater variety of co-occurrence, we describe a CPF that applies the subtraction operator. With this new type of CPF, co-occurrence in which the probability of the detection-target class is revised from information having other properties can be expected. In this paper, we use geometric context proposed by Hoiem *et al.* as local features giving us information having different properties.

### 5.1. Geometric Context

Geometric context (GC) [3] proposed by Hoiem *et al.* is a method for estimating and labeling three-dimensional scene structures from a single image. This method inputs color, texture, location, shape and geometry feature, and learns a model using a logistic regression form of AdaBoost [2]. At the time of classification, the model obtained by learning is used to classify the test image into "ground", "vertical", and "sky" classes. The results of estimating geometric labels by GC from the input image of Fig. 7(a) are shown in Figs. 7(b), 7(c), and (d) for "ground", "vertical", and "sky", respectively.

In this study, we use the confidences of "ground", "vertical", and "sky" obtained by GC as local features in each of the cells used for computing HOG features. That is, we determine average confidences for the pixels included in cell

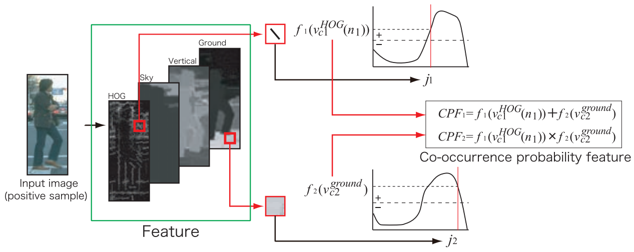


Figure 8. Representation of co-occurrence between HOG features and geometry context by addition and multiplication operators.

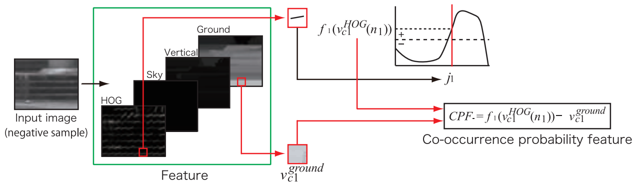


Figure 9. Representation of co-occurrence between HOG features and geometry context by subtraction operator.

$c$ and take them to be local features $v_c^{ground}$, $v_c^{vertical}$, and $v_c^{sky}$ obtained by GC.

### 5.2. Co-occurrence with Geometry Context

#### 5.2.1 Co-occurrence with GC by addition and multiplication operators

We represent the co-occurrence of HOG features and geometry context as shown in Fig. 8 based on the co-occurrence representation technique using addition and multiplication operators described in section 3.3. This achieves co-occurrence in which the probability of the target-detection class is revised by information having different properties obtained from different regions.

#### 5.2.2 Co-occurrence with GC by subtraction operator

We also represent co-occurrence with geometry context using the subtraction operator. CPF using the subtraction operator is generated combining obtained local features from the same local region. CPF using the subtraction operator are given by Eq. 7.

$$CPF_{-} = f_1(v_{c_1}^{HOG}(n_1)) - \begin{cases} v_{c_1}^{ground} \\ v_{c_1}^{vertical} \\ v_{c_1}^{sky} \end{cases} \tag{7}$$

An example of generating a CPF using the subtraction operator is shown in Fig. 9. In this negative sample, vehicle-like characteristics would be high on the basis of only HOG features. However, by representing co-occurrence with "ground" confidence using the subtraction operator, the excessive probability of a vehicle computed when focusing on HOG features can be controlled.

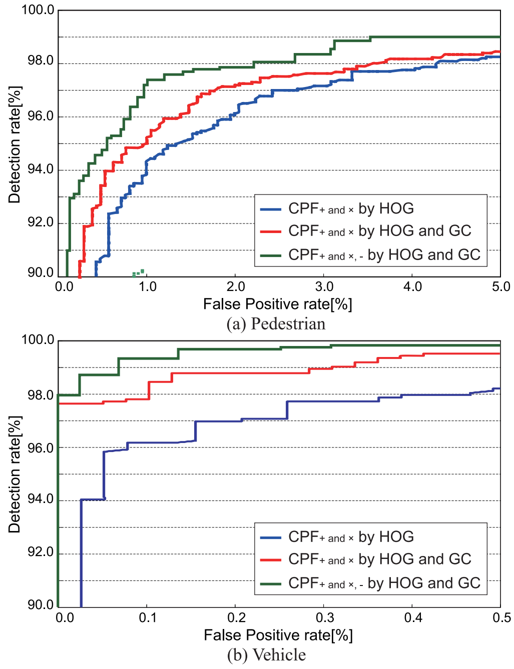


Figure 10. ROC curves of experimental results.

### 5.3. Evaluation Experiment

We performed an experiment to evaluate a CPF that applies the subtraction operator. The experiment compared (1) CPF$_{+and\times}$ using HOG features, (2) CPF$_{+and\times}$ using HOG and geometry context, and (3) CPF$_{+and\times,-}$ using HOG and geometry context.

The ROC curves of experimental results are shown in Fig. 10. To begin with, it can be seen from these results that the detection performance of CPF applying geometric context by addition and multiplication operators is higher than CPF using only HOG features. It can therefore be said that geometry information obtained from GC is effective in detecting people and vehicles. Fig. 11 shows evaluation samples that are judged to be a non-detection target by CPF using only HOG features but judged to be a detection target by CPF that also applies geometry context. Since these evaluation samples consist of images taken in the evening hours, the human contour, for example, cannot be clearly captured by only HOG features, i.e. appearance. When applying geometry information, however, a correct classification is obtained for these samples that are difficult to classify on the basis of appearance only.

Next, we evaluate CPF applying the subtraction operator. It can be seen from the ROC curves of Fig. 10 that applying the subtraction operator improves detection performance for both pedestrian and vehicle cases. This indicates that the use of geometry context was able to revise the output of the detection-target class obtained from HOG features. Fig. 12 shows examples of evaluation samples that were judged to be detection targets when not applying the subtraction operator but judged to be non-detection-targets when applying the subtraction operator. Here, introducing the subtraction operator in CPF reduces the values of output from the final strong classifier compared to the method not applying this operator. This is because the use of the subtraction operator during training can avoid the selection of HOG features ineffective for classification thereby helping to suppress erroneous detection.

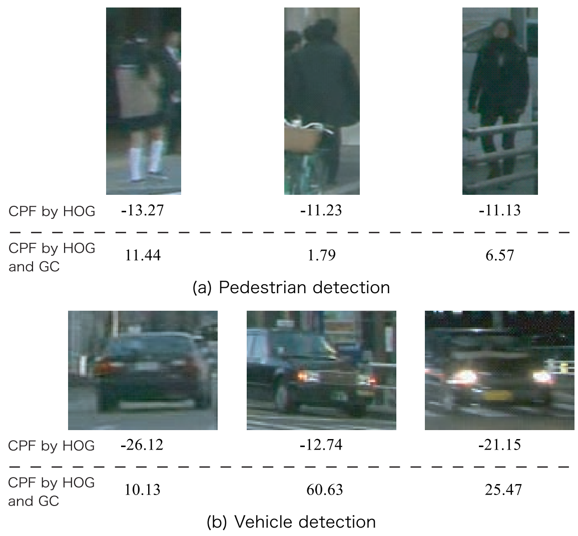


Figure 11. Examples of evaluation samples for which detection became possible by CPF using HOG and GC applying the addition and multiplication operators. The values in the figure indicate output of the final strong classifier.

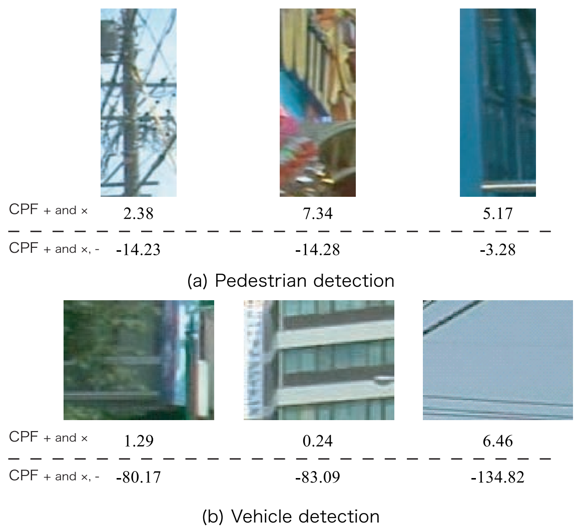


Figure 12. Examples of evaluation samples in which erroneous detection was controlled by CPF applying the subtraction operator.

### 5.4. Discussion

Here we discuss the CPFs selected by training in the experiments. Fig. 13 shows selected local features by boosting. The regions corresponding to those features are shown

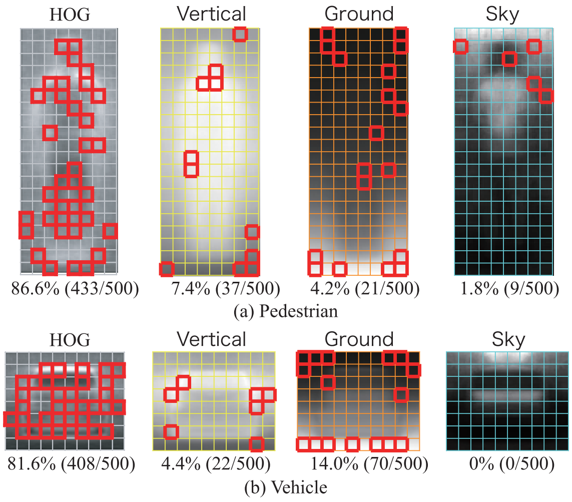


Figure 13. Local features selected by boosting. Ratios in parentheses indicate number of local features selected per total number of local features. The regions corresponding to those features are shown by red squares. A higher intensity level in these images indicates higher gradients in the case of HOG and higher confidence (in the "vertical", "ground", and "sky") in the case of GC.

by red squares. As for HOG features, there is a tendency to capture the entire silhouette of the detection target whether that be a pedestrian or vehicle. At about 80%, HOG features make up most of the selected features. As for GC, the feature that expresses ground-like characteristics was selected in greater number for the vehicle than for the pedestrian. In the case of pedestrian, feet are often occluded by guard rails or other objects, but in the case of vehicles, the occlusion of the lower portion of a vehicle occurs infrequently in comparison to pedestrians. Thus, when the detection target is a pedestrian, conditions conducive to selection of ground-like features occur less often compared to a vehicle. On the other hand, sky and vertical features, while few, are selected in greater number for the pedestrian than for the vehicle.

Next, we turn our attention to "ground" and see that low-confidence regions were also selected. In the case of pedestrian detection, which is taken from vehicle-mounted camera, this expresses that fact that not only is the area around a person's feet "ground" but also that the area around the person's torso is not "ground". Likewise for vehicle detection, the selection of low-confidence regions indicates that not only is the area below the vehicle "ground" but also that area around the upper part of the vehicle is not "ground" but "vehicle".

## 6. Conclusion

We proposed a co-occurrence probability feature that combines weak classifiers from Real AdaBoost by addition and multiplication operators. The co-occurrence probability feature becomes a discriminative feature by virtue of combining multiple local features through boosting achieved by operators. This combination of weak classifiers using multiple operators of different types enables the representation of various types of co-occurrence. In addition, to represent co-occurrence of even greater diversity, we proposed co-occurrence representation that applies the subtraction operator. The extended co-occurrence probability feature using the subtraction operator represents co-occurrence such that the probability of the target-object class is revised by information having different properties. We confirmed by evaluation experiments that detection accuracy could be improved by this extended approach. In future research, we plan to study a feature co-occurrence representation method applicable to multi-class category classification.

## References

[1] H. Chang, A. Haizhou, T. Yamashita, L. Shihong, and M. Kawade. Incremental learning of boosted face detector. In *ICCV*, pages 1–8, 2007.

[2] M. Collins, R. E. Schapire, and Y. Singer. Logistic regression, adaboost and bregman distances. *Machine Learning*, 48(1-3):253–285, 2002.

[3] A. D. Hoiem, A. Efros, and M. Hebert. Geometric context from a single image. In *ICCV*, pages 654–661, 2005.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

[5] C. Galleguillos, A. Rabinovich, and S. Belongie. Object categorization using co-occurrence, location and appearance. In *CVPR*, pages 1–8, 2008.

[6] C. Hou, H. Ai, and S. Lao. Multiview pedestrian detection based on vector boosting. In *Asian Conference on Computer Vision*, pages 18–22, 2007.

[7] K. Levi and Y. Weiss. Learning object detection from a small number of examples: the importance of good features. In *CVPR*, volume 2, pages 53–60, 2004.

[8] T. Mita, T. Kaneko, B. Stenger, and O. Hori. Discriminative feature co-occurrence selection for object detection. *PAMI*, 30(7):1257–1269, 2008.

[9] T. Mitsui and H. Fujiyoshi. Object detection by joint features based on two-stage boosting. In *Workshop on Visual Surveillance*, 2009.

[10] Y. Mu, S. Yan, Y. Liu, H. uang, and T. B. Zhou. Discriminative local binary patterns for human detection in personal album. In *CVPR*, pages 1–8, 2008.

[11] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[12] T. Okabe, Y. Kondo, K. M. Kitani, and Y. Sato. Recognizing multiple objects via regression incorporating the co-occurrence of categories. In *Pacific-Rim Symposium on Image and Video Technology*, pages 497–508, 2009.

[13] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *ICCV*, pages 1–8, 2007.

[14] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. In *CVPR*, pages 1–8, 2007.

[15] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, (37):297–336, 1999.

[16] O. Tuzel, F. Porikli, and P. Meer. Human detection via classification on riemannian manifolds. In *CVPR*, pages 1–8, 2007.

[17] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[18] X. Wang, T. X. Han, and S. Yan. An hog-lbp human detector with partial occlusion handling. In *ICCV*, pages 32–39, 2009.

[19] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *CVPR*, pages 90–97, 2005.

[20] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *CVPR*, pages 1491–1498, 2006.