Object Detection by Joint Features based on Two-Stage Boosting

Tomokazu Mitsui, Hironobu Fujiyoshi
Dept. of Computer Science, Chubu Univ.
Matsumoto 1200, Kasugai, Aichi, 487-8501 Japan
mitsui@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp
http://www.vision.cs.chubu.ac.jp

Abstract

In this paper, we propose an object detection method that uses Joint features combined from multiple Histograms of Oriented Gradients (HOG) feature using two-stage boosting. There has been much research in recent years on statistical training methods and object detection methods that combine low-level features obtained from local areas. In our approach, multiple low-level HOG features are combined by using Real AdaBoost to automatically generate Joint features. Joint features represent the co-occurrence of the HOG features of multiple cells combined by the first-stage Real AdaBoost. Next, the generated Joint features pool is input to the second-stage Real AdaBoost, which constructs the final classifier. In this way, it is possible to capture shape symmetry and edge continuity, which single HOG features cannot do, so highly accurate detection is possible. In this paper we report experiments involving the detection of humans and vehicles performed to test the effectiveness of the proposed method. In addition, two-stage boosting classifier is used to represent the co-occurrence of the appearance(HOG) and spatiotemporal(PSA) features for detecting pedestrian. The use of feature co-occurrence, which captures the similarity of appearance, motion, and spatial information within the human class, makes it an effective detector.

1. Introduction

Object detection, which is searching for particular objects in an image, is one of the biggest problems in the field of computer vision, and many methods for achieving it have already been proposed. Most object detection methods of recent years have used object classification methods that apply statistical training methods for local features selected from several thousand training samples. For the local features used by such methods, there are low-level features such as Haar-like features [1], Edge Orientation Histograms (EOH) [2] and Edgelet features [3]. Of those features, the Histogram of Oriented Gradients (HOG) [4] has been demonstrated to be robust to changes in illumination and local changes in geometry to detect a human in the images. With only these low-level features, however, the recognition capability is limited in the case of complex scene. Therefore, methods for generating features by using statistical learning algorithms such as AdaBoost[5] to combine low-level features based on relations between features (relatedness) to achieve highly accurate object detection have been proposed in recent years[6], [7].

The features of the human form are broadly classified into the following two types：

(1) The head and shoulders have a shape similar to the Greek letter Ω, and the shape from the upper half to the lower half is continuous

(2) Left-right symmetry of the head and shoulders, torso, and legs, etc.

For the first of the above features, Shapelet [6], which are AdaBoost combinations of four-direction edge features within local areas that represent surface information on local areas, have been proposed. However, there are very many low-level features, so they are restricted to only combinations of features within a local area. Therefore, it is not possible to observe features outside of the local area at the same time, and the important features of symmetry and continuous edges cannot be captured. For the second type of feature listed above, Joint Haar-like features that represent co-occurrence observed simultaneously by the AdaBoost weak classifier have been proposed [7]. This method can be faster and more accurate than conventional face detection, but the features depend on the intensity, so the method is not suitable for humans and other objects that have diverse shapes and texture.

Those two methods can capture the relations between features by using boosting to combine multiple low-level features, making highly accurate detection possible. Although both of the above two types of shape features are important, no method can capture both has been proposed.

We therefore propose here an object detection method in a image that uses Joint features that can automatically

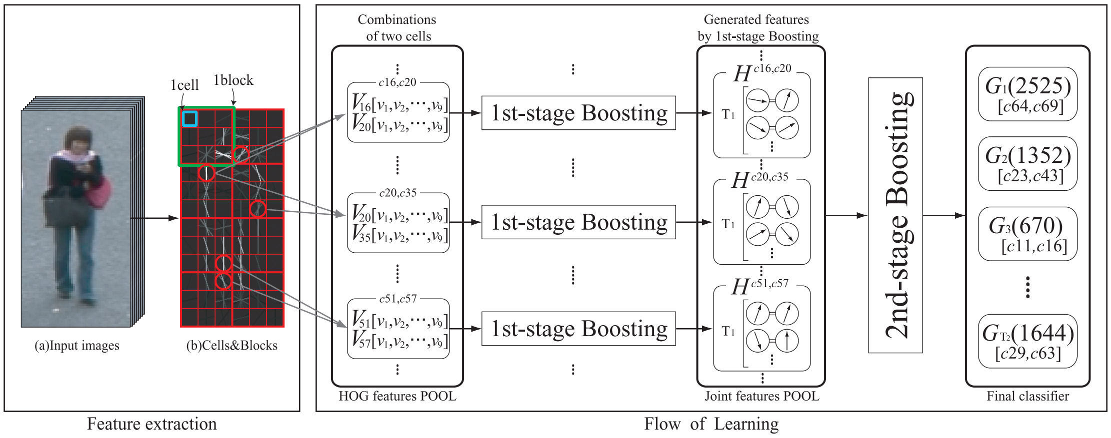


Figure 1. The flow of proposed method.

capture object shape symmetry and continuity. The Joint features are first obtained by combining the HOG features for several different local areas by means of the first-stage Boosting. Then, the generated Joint features pool is input to the second-stage Boosting to construct the final classifier and detect the object. Thus, the proposed method can automatically generate a Joint feature that represents symmetry and continuity, which are difficult to represent with a single feature. From the next chapter, we demonstrate the effectiveness of the proposed method experimentally, taking as the objects of detection the non-rigid human body and vehicles, which have greatly different appearances from different points of view.

2. Joint features based on two-stage Boosting

The flow of the proposed method is illustrated in Figure 1. This method uses two-stage Real AdaBoost to construct the final classifier. For the boosting, we used Real AdaBoost[8], which can obtain continuous output of the weak classifier. Real AdaBoost allows highly accurate detection with fewer weak classifiers than AdaBoost. First, we prepare a pool of Joint features that combine several low-level HOG features obtained for different locations by the first-stage Real AdaBoost. Doing so allows multiple HOG features to be observed at the same time. This makes it possible to automatically generate Joint features that can represent shape symmetry and edge continuity, which cannot be grasped by conventional single HOG features alone. Next, the second-stage Real AdaBoost selects from the Joint feature pool the features that are best for automatic human detection, and then the final classifier is used to perform the detection.

2.1. Low-level feature : HOG

In this paper, we use the Histogram of Oriented Gradients (HOG) proposed by Dalal *et al.* as the low-level feature [4]. HOG features are calculated from gradient orientations in local areas called cells (Figure 1(b)) converted into histograms. They can capture the shape of an object and are robust to changes in illumination and local changes in geometry. The procedure for calculating the HOG features is described below.

From the brightness L of each pixel, compute the gradient magnitude m and orientation θ with the following formula.

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

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

$$\begin{cases} f_x(x,y) = L(x+1,y) - L(x-1,y) \\ f_y(x,y) = L(x,y+1) - L(x,y-1) \end{cases} \tag{3}$$

The brightness gradient orientation histogram of each cell is generated from the calculated gradient magnitude m and orientation θ. The obtained gradient orientations are divided into 20-degree groups to create the gradient orientation histogram.

Finally, the features are normalized to each block area (Figure 1(b)) with the following equation.

$$v = \frac{v}{\sqrt{\left(\sum_{i=1}^{k} v_i^2\right) + \epsilon}} \qquad (\epsilon = 1) \tag{4}$$

Here, v is the HOG feature, k is the number of HOG features in the block, and ϵ is a coefficient for preventing division by zero problems.

2.2. features co-occurrence

To generate the Joint features, we represent the co-occurrence of multiple HOG features [7]. Representing co-occurrence makes it possible to observe several features(more than two) at the same time. We explain here how to combine two HOG features.

First, we calculate binary symbols s that represent detection objects and non-detection objects with the following equation.

$$s(\mathbf{V}) = \begin{cases} 1 & p \cdot v_o > p \cdot \theta \\ 0 & \texttt{otherwise} \end{cases} \tag{5}$$

Here, θ is the threshold value, p is a parity indicating the direction of the inequality sign, o is the orientation of gradient, and takes the values $p \in \{+1, -1\}$. The value of θ and p are determined so that there error rate is minimized. $\mathbf{V} = [v_1, v_2, \cdots v_q]$ is the HOG feature calculated from one cell, and q is the number of orientation of the gradient in a cell. By combining the two binary symbols obtained in this way, we get features j, which represent co-occurrence [7]. For example, when HOG feature binary symbols $s_1 = 1$ and $s_2 = 1$ are observed in an input image such as shown in Figure 2, the co-occurrence feature j is $j = (11)_2 = 3$. This j is an index number for a binary representation of combined features. In this case, there are four values because we are dealing with combinations of two features.

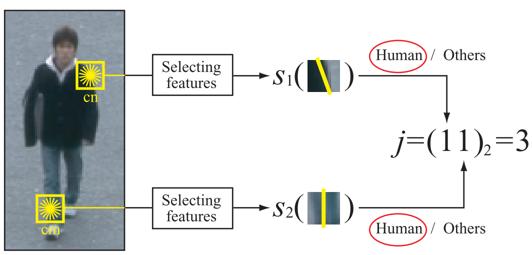


Figure 2. Co-occurrence of features.

2.3. Joint features

The HOG feature co-occurrence values calculated in section 2.2 are used to generate Joint features in the first-stage Real AdaBoost. This captures the relations of cells as well as the symmetry of object shape and edge continuity.

First, from the features that represent co-occurrence for cells at two different locations, c_m, c_n, Real AdaBoost selects those that are effective in discrimination. Here, a set of N labeled training samples is given as (x_1, y_1), . . ., (x_N, y_N), where $y_i \in \{+1, -1\}$ is the class label associated with a training sample x_i. The function for observing HOG feature co-occurrence in training sample x_i is expressed as $J_t(x_i)$. When feature $J_t(x_i) = j$ is observed , the weak classifier $h_t(x)$ of the first-stage Real AdaBoost is expressed as follows:

$$h_t(x) = \frac{1}{2} \ln \frac{P_t(y = +1|j) + \epsilon}{P_t(y = -1|j) + \epsilon}. \tag{6}$$

Here, t is the number of training rounds and ϵ is a coefficient for preventing division by zero problems. We determined by experiment that $\epsilon = 0.0000001$. $P_t(y = +1 \mid j)$ and $P_t(y = -1 \mid j)$ are the respective conditional probability distributions for when the features j that represent HOG feature co-occurrence are observed. The conditional probability distributions are calculated with the following equation from the weights $D_t(i)$.

$$P_t(y = +1|j) = \sum_{i:J_t(x_i)=j \wedge y_i=+1} D_t(i) \tag{7}$$

$$P_t(y = -1|j) = \sum_{i:J_t(x_i)=j \wedge y_i=-1} D_t(i) \tag{8}$$

$$D_{t+1}(i) = D_t(i) \exp[-y_i h_t(x_i)] \tag{9}$$

$D_t(i)$ is a weight of a training sample x_i. The weights are initialized by $D_1(i) = 1/N$. The conditional probability distributions $P_t(y = +1 \mid j)$ and $P_t(y = -1 \mid j)$ are represented by one-dimensional histograms. The distributions are created by calculating the features that represent co-occurrence from the training samples x and adding the training sample weights D_t to the corresponding one-dimensional histogram BIN [1] numbers j. Because the BIN numbers j correspond to index numbers for the features that represent co-occurrence, there are 4 BIN.

Next, we use the conditional probability distribution to obtain an evaluation value z_1 that represents the separation of the distributions with the following equation:

$$z_1 = 2 \sum_j \sqrt{P_t(y = +1|j) P_t(y = -1|j)}. \tag{10}$$

Smaller values of z_1 indicate greater separation of the positive class and negative class distributions. The smallest of z_1 is used in the selection of a weak classifier from among the many candidates in each round.

Finally, the Joint feature $H^{c_m,c_n}(x)$, which is the strong classifier of the first-stage Real AdaBoost, is constructed with the following equation:

$$H^{c_m,c_n}(x) = \sum_{t=1}^{T} h_t^{c_m,c_n}(x). \tag{11}$$

The processing described above is applied to all combinations of cells to generate as many Joint features as there are cell combinations. For example, taking a 30×60 pixel input image and a cell size of 5×5 pixels, the number of cell combinations is ${}_{72}C_2 = 2,556$ and $2,556$ Joint features $H^{c_m,c_n}(i)$ can be generated. All of the generated Joint features are put into a single pool for input to the second-stage Real AdaBoost to construct the final classifier as described below.

2.4. Constructing the final classifier from the second-stage Real AdaBoost

The pool of Joint features generated in the first-stage Real AdaBoost is input to the second-stage Real AdaBoost to construct the final classifier (Figure 3). In this way, the Joint features that are effective in discrimination can be selected automatically.

[1] Number of partitions of the histogram.

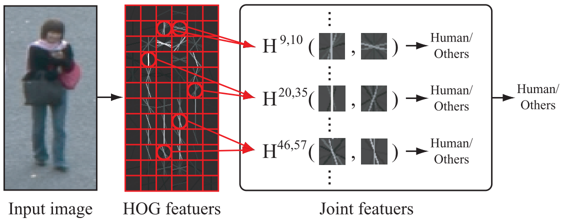


Figure 3. Joint features.

First, we input the Joint features $H^{c_m,c_n}(x)$ and create positive class and negative class probability density distributions W_+ and W_-. The probability density distribution W_+ and W_- is represented by a one-dimensional histogram that is generated from the training sample weights D_t with the following equation:

$$W_+^k = \sum_{i:k \in K \wedge y_i = +1} D_t(i), \tag{12}$$

$$W_-^k = \sum_{i:k \in K \wedge y_i = -1} D_t(i), \tag{13}$$

where, t is the number of training rounds, i is the number of training samples, k is the BIN number of the one-dimensional histogram, and y_i is the class label $y_i \in \{+1, -1\}$. The calculation method of D_t is in the same way as the first-stage Real AdaBoost. The probability density distribution W_+^k and W_-^k can be created by calculating the features from the training samples x_i and applying training sample weights $D_t(i)$ to the BIN numbers k of the one-dimensional histograms that correspond to the feature values. The BIN count of the one-dimensional histogram must be set to a value that is suitable for the number of training samples. In the work reported here, the one-dimensional histogram BIN count was set to 64 by experiment at the second-stage Real AdaBoost. The created probability density distribution W_+^k and W_-^k is normalized so that the sum of all of the probability density distributions of each class is 1.

That probability density distribution is used to obtain the evaluation value z_2, which represents the degree of separation of the distributions.

$$z_2 = 2 \sum_j \sqrt{W_+^k W_-^k} \tag{14}$$

Smaller values of z_2 indicate greater separation between the positive and negative class distributions. The minimum value of z_2 is used to select a single weak classifier from among the many candidates in each round.

Next, we use the created probability density distribution W_+^k and W_-^k to calculate the output $g_t(c)$ of the second-stage Real AdaBoost weak classifier. From the values of the HOG feature co-occurrence j obtained from the training samples, we calculate one-dimensional histogram BIN numbers j, and from their corresponding probability density distributions W_+^k and W_-^k we calculate the weak classifier output $g_t(c)$ by using the following equation:

$$g_t(c) = \frac{1}{2} \ln \frac{W_+^k + \epsilon}{W_-^k + \epsilon}. \tag{15}$$

The c is a serial number that represents combinations of cells. The ϵ is a coefficient for preventing division by zero problems and it is given the same value as used in the first-stage Real AdaBoost, $\epsilon = 0.0000001$.

Finally, the following equation is used to construct the final strong classifier $G(c)$ in the second-stage Real AdaBoost.

$$G(c) = \begin{cases} 1 & \sum_{t=1}^{T} g_t(c) > \lambda \\ 0 & \texttt{otherwise} \end{cases} \tag{16}$$

The λ is a threshold value of the classifier. The second-stage Real AdaBoost constructs the classifier by selecting from the Joint feature pool only the features that are effective in discrimination.

3. Discrimination experiments using Joint features

To demonstrate the effectiveness of the proposed method, we conducted experiments to evaluate the detection of humans and vehicles.

3.1. Databases

We construct a database for human and vehicle, respectively. A part of the database is shown below in Figure 4.

Human database Vehicle database

Figure 4. Databases.

Human database The positive samples for human class are taken in multiple different locations. The negative samples are random areas from the background. The database used for training contains 2,054 positive samples and 6,258 negative samples, and we used 1,000 positive samples and 1,235 negative samples for the evaluation database.

Vehicle database The positive samples for vehicle class are vehicle areas cut out from video images taken by a vehicle-mounted camera pointed rearward. The negative samples are random areas from the background. The training database comprised 2,464 positive samples and 16,158 negative samples, the evaluation database comprised 1,900 positive samples and 5,153 negative samples.

3.2. Experiments

We used the evaluation databases to conduct the human and vehicle discrimination experiments. The parameters for the experiments are listed in Table 1.

Table 1. Experiment parameter.

	Human	Vehicle
# of Orientations	9(0° ∼180°)	18(0° ∼360°)
Patch size (pixels)	30×60	90×72
Cell size (pixels)	5×5	9×9
Block size (cells)	3×3	2×2
# of cells	72	80
# of 2 combinations	2,556	3,160
# of 3 combinations	59,640	82,160

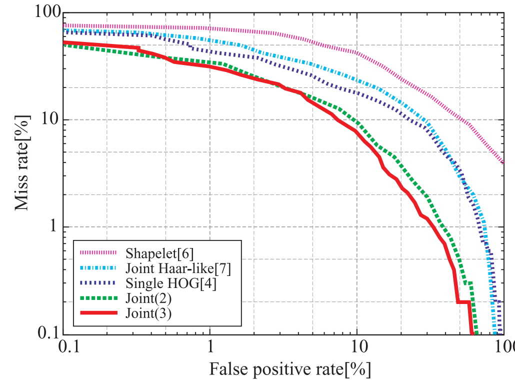
The gradient orientations shown in Table 1 are from 0 degrees to 360 degrees for vehicles and from 0 degrees to 180 degrees for humans. The reason for this conversion is to obtain orientations for which a human's clothing does not affect the results, because there is sometimes an inverse relation of clothing and the background color. In this experiments, we set the number of combined features to two(Joint(2)) and three(Joint(3)).

In the evaluation, we used the Detection Error Tradeoff (DET) graph, which is a dual-log plot with false positive rate on the horizontal axis and miss rate on the vertical axis. In a DET plot, values closer to the origin indicate better performance.

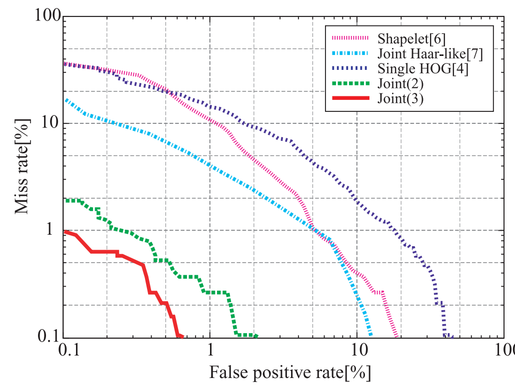
3.3. Experimental results

The discrimination results are presented in Figure 5(a) for humans and in Figure 5(b) for vehicles. We see that the proposed method discriminates more accurately than the conventional method using single HOG [4], Shapelet[6] and Joint Haar-like[7]. At a false positive rate of 5.0% for humans, the improvement in discrimination performance was about 12.1%(Joint(3)). At a false positive rate of 0.5% for vehicles, the improvement in discrimination was about 18.9%(Joint(3)). These improvements came with the automatic selection of new effective features by combining the HOG features of more than two different locations.

Figure 6 shows the detection results by window raster scanning over the image from the upper left multiple times at different scales. Windows in which the target was detected are finally integrated by Mean Shift clustering [9]. We see from Figure 6 that it is possible to detect individual objects even if they are overlapped.

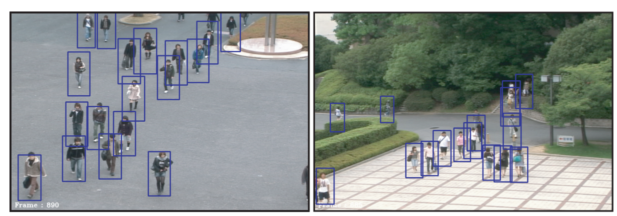


(a)Human

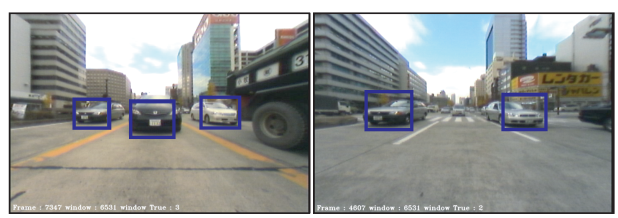


(b)Vehicle

Figure 5. DET curve of discrimination result.



(a)Human

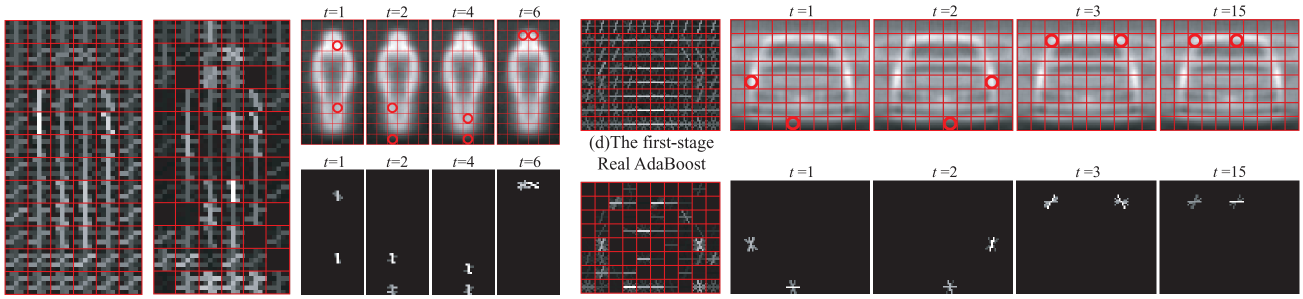


(b)Vehicle

Figure 6. Examples of object discrimination.

3.4. Discussion

Here we discuss the Joint features selected by Real AdaBoost in the experiments on discriminating humans and vehicles. The visualized results of the selected HOG features are shown in Figure 7(a) and Figure 7(d) for the first-stage Real AdaBoost and in Figure 7(b) and Figure 7(e) for the second-stage Real AdaBoost. In (c) and (f) of the same figure, the two selected cells and the Joint features from the second-stage Real AdaBoost are shown for each round. The HOG feature gradient orientations are represented by 9 directions for humans and 18 directions for vehicles. Higher intensity corresponds to lower z values of the weak classifier in Real AdaBoost, indicating that the features are effective in discrimination.



(a)The first-stage Real AdaBoost (b)The second-stage Real AdaBoost (c)Two cells and Joint features by the second-stage Real AdaBoost (e)The second-stage Real AdaBoost (f)Two cells and Joint features by the second-stage Real AdaBoost

Figure 7. Visualization of selected Joint features.

When humans are the target for detection, from the Figure 7(b), there is a tendency to not select features that are not part of the human outline, even though they were selected at the first-stage Real AdaBoost as shown in Figure 7(a). This probably results from the judgment that those features are ineffective for discrimination in the feature selection of the second-stage Real AdaBoost. Finally, consider Figure 7(c). We can see that the selected Joint feature at each round that follows the outline of a human tend to be selected in the second-stage Real AdaBoost. This demonstrates that our method is effective in detecting the human form, which is a non-rigid body.

When vehicles are the target for detection, we see in Figure 7(d) that many horizontal edges inside the vehicle and edges that follow the outline of the vehicle are selected by the Joint features of the first-stage Real AdaBoost. In Figure 7(e), we see that from the HOG features selected in Figure 7(d), the Joint features that follow the outline of the vehicle are selected by the final classifier obtained from the second-stage Real AdaBoost. We thus see that the HOG features that follow the vehicle outline are effective for distinguishing the vehicle from the background. Finally, consider Figure 7(f). In the first and second rounds of training, the positional relations of the vertical and horizontal edges are selected. In the third round, the cells that have left-right symmetry are selected. In round 15, the cells whose positional relations capture continuity are selected and horizontally oriented features are selected. The proposed Joint features make it possible to automatically select cells whose positional relationships represent symmetry and continuity through training, without advance preparation of features that capture vehicle shape symmetry and continuity, and thus obtain an effective feature set for object discrimination.

4. Experiments on pedestrian detection with appearance and spatiotemporal features

It is also possible to use other features together with the HOG features that we use here for the low-level features. For example, in the case of using a fixed camera, we can use motion information detected by background subtraction. In this section, we describe a method for detecting pedestrians by combining HOG features and spatiotemporal features obtained by pixel state analysis (PSA).

4.1. Pixel State Analysis(PSA)

Objects similar to human are done false detection when only appearance feature is used. Therefore, we use features obtained from the result of pixel state analysis(Figure 8)[10] that represent object motion and spatial information.

To capture the nature of changes in pixel intensity profiles, two factors are important: the existence of a significant step change in intensity, and the intensity value to which the profile stabilizes after passing through a period of instability.

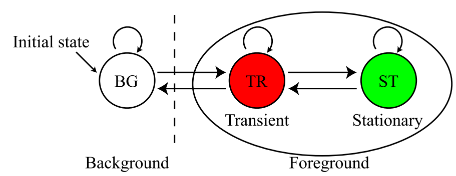


Figure 8. Diagram of state transition for a pixel.

Let I_t be some pixel's intensity at a time t occurring k frames in the past. Two functions are computed: a motion trigger T just prior to the frame of interest t, and a stability measure S computed over k frames from time t to the present. The motion trigger is simply the maximum absolute difference between the pixel's intensity I_t and its value in the previous five frames:

$$T = max\{|I_t - I_{(t-j)}|, \forall_j \in [1,5]\}. \quad (17)$$

The stability measure is the variance of the intensity profile from time t to the present:

$$S = \frac{k\sum_{j=0}^{k} I_{(t+j)}^2 \left(\sum_{j=0}^{k} I_{(t+j)}\right)^2}{k(k-1)}. \quad (18)$$

Transient map M is defined by the algorithm below (Figure 9) for each pixel, using three possible values : background = (bg); transient = (tr) and stationary = (st).

The background intensity is prepared in advance as a background image.

```
if((M = st OR M = bg) AND (T > th_t)){
    M = tr
}
if((M = tr) AND (S < th_s)){
    if(I = background intensity){
        M = bg
    }else{
        M = st
    }
}
```

Figure 9. Algorithm for pixel state analysis.

4.2. Use together with spatiotemporal features

We use pixel state analysis (PSA) to obtain spatiotemporal features. Pixel state analysis involves distinguishing pixels according to three states, background, stationary, and transient, by modeling the temporal changes in pixel states. Figure 10 shows an example of pixel state analysis.

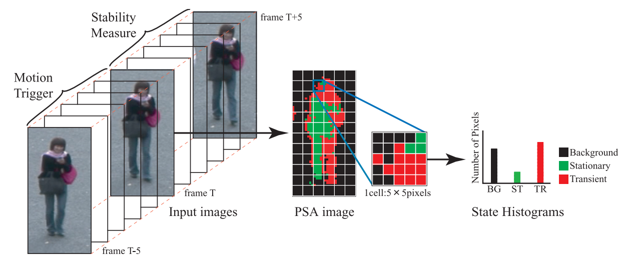


Figure 10. An example of Pixel State Analysis and a state histogram.

A state histogram is created from the results of the pixel state analysis for the state of each cell as shown in Figure 10. Each pixel is classified into one of three states, so three features can be calculated from a single histogram. By adding three PSA features to the nine HOG features, 12 features can be obtained from each cell.

Using PSA features makes it possible to handle features that capture movement, which differ in nature from the HOG features. For example, the human in Figure 10 is walking, and in the five frames before and after this image, the person pivots on the right leg and steps forward with the left leg. That series of movements cannot be captured by edge-based features such as HOG features. The results of pixel state analysis, however, reveal the static state of the right leg and the moving state of the left foot, and thus can capture how the human is moving from a single image. Combining HOG features as appearance information and PSA features as spatiotemporal information can construct a object detector with high accuracy.

A final classifier that uses Joint features that involve both HOG features and PSA features is shown in Figure 11.

4.3. Discrimination experiments

We performed the evaluation experiments using the human database described in section 3.1. The results of the pixel state analysis were extracted as features. We used

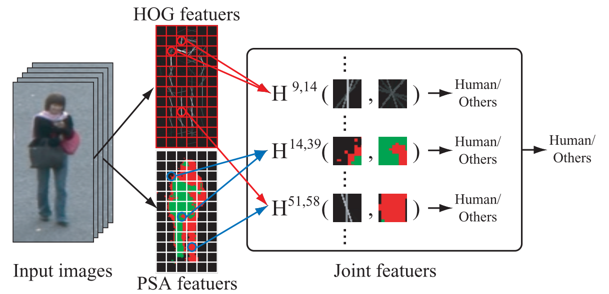


Figure 11. Human detection by Joint features using HOG features and PSA features.

the same experimental parameters as presented in Table 1. In the experiments, we compared Joint(HOG+PSA), Joint(HOG+HOG), and single HOG. The detection results are shown by DET in Figure 12. For a false positive rate of 1.0%, the improvement in detection rate was 39.2%. The proposed method exhibits higher detection performance than the conventional method of using HOG feature and PSA feature co-occurrence. This indicates that Joint features of HOG features and PSA features is effective to detect the pedestrians, compared to using single HOG features alone.

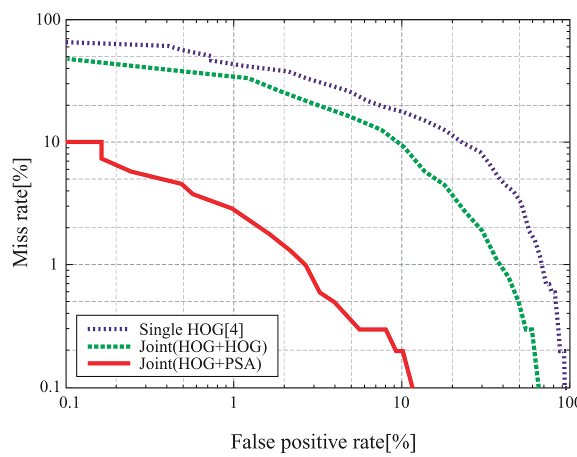


Figure 12. DET curve of discrimination result for Joint features.

We conducted the pedestrian detection experiment using the constructed classifier. The detection method is the same as used in section 3.3. For comparison, the detection results from using the single HOG features are shown in Figure 13(a) and the results for Joint feature using HOG feature and PSA features are shown in Figure 13(b).

In Figure 13(a) we see that there is a false positive detection with the sign in the center of the image. The result of adding the PSA features to the HOG features is to eliminate that false positive, as we see in Figure 13(b).

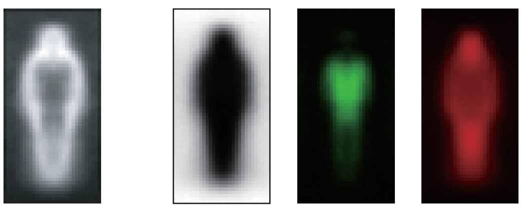
In the detection of humans, PSA features of movement are used, so highly accurate detection is possible even for when there are objects similar in shape to humans or when the background is complex.

(a)HOG features (b)Joint features(HOG+PSA)

Figure 13. Examples of pedestrian detection for Joint features with HOG features and PSA features.

4.4. Discussion

Here we discuss the proposed method capturing co-occurrence HOG features and PSA features in terms of the features selected in training. The average gradient image for when the positive samples used for training are used is shown in Figure 14(a).



(a)Mean gradient (b)Background (c)Stationary (d)Transient

Figure 14. Visualization of occurrence rate of each state.

In (b), (c) and (d) of the same figure, the visualized images of the frequency of occurrence of the three states are shown using the images of the results of the pixel state analysis of the positive samples used for training. We see from Figure 14 that higher intensity indicates a stronger gradient or higher frequency. In the average image for the background state in Figure 14(b), we can see that a human silhouette is represented. Furthermore, from Figure 14(c) and (d) we can see that stationary state pixels have high frequencies in the upper half of the human shape, while transient state pixels have high frequencies in the lower half.

Features that are selected earlier in the Real AdaBoost training round can be the more effective ones. To reveal this tendency to select the features that are effective for discrimination, the proportions of HOG features and PSA features that are selected in each actual training round of Real AdaBoost and an example of visualization of the selected features are presented in Figure 15.

At the beginning of the training, many PSA features are selected. This represents the fact that whether the pixels in background state are many or few is used for detecting human objects. Furthermore, we see that many HOG features begin to be selected from the fifth round. We believe that the PSA features, which can represent object movement, are selected early to first roughly distinguish between human and non-human, and then the HOG features, which capture appearance information, are selected to form detailed discrimination boundaries.

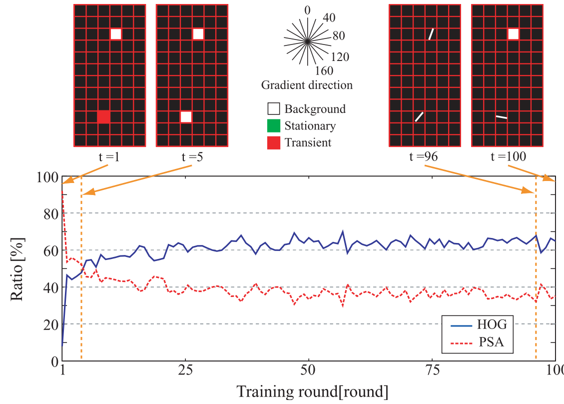


Figure 15. Selected Features and the ratio by training.

5. Conclusion

In this paper, we presented an object detection method that uses Joint features, which combine multiple HOG features, and two-stage Real AdaBoost training. The Joint features are effective for discrimination because they can capture relations among cells as features as well as object symmetry in shape and edge continuity. In future work, we plan to accomplish human detection in images from vehicle-mounted cameras using scene context information.

References

[1] Papageorgiou, C. P., Oren, M. and Poggio, T. "A general framework for object detection", *IEEE ICCV*, pp. 555–562 (1998). 1

[2] K. Levi and Y. Weiss, "Learning Object Detection from a Small Number of Examples: the Importance of Good Features.", *IEEE CVPR*, vol. 2, pp. 53-60(2004). 1

[3] B. Wu and R. Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors", *IEEE ICCV*, vol. 1, pp. 90-97 (2005). 1

[4] Dalal, N. and Triggs, B. "Histograms of Oriented Gradients for Human Detection", *IEEE CVPR*, pp. 886–893 (2005). 1, 2, 5

[5] Freund, Y. and Schapier, R. E., "Experiments with a new boosting algorithm", *Machine Learning*,(1996). 1

[6] P. Sabzmeydani and G. Mori, "Detecting Pedestrians by Learning Shapelet Feature", *IEEE CVPR*, pp. 511-518 (2007). 1, 5

[7] Mita, T, Kaneko, T, Stenger, B, and Hori, O, "Discriminative Feature Co-Occurrence Selection for Object Detection", *IEEE PAMI*, vol. 30, no. 7, pp. 1257-1269, (2008). 1, 2, 3, 5

[8] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions", *Machine Learning*, No. 37, pp. 297-336, (1999). 2

[9] D. Comaniciu, P. Meer : "Mean Shift: A Robust Approach toward Feature Space Analysis", *IEEE PAMI*, vol. 24, No. 5, pp. 603-619, (2002). 5

[10] H. Fujiyoshi and T. Kanade, "Layered Detection for Multiple Overlapping Objects", *IEICE Transactions on Information and systems*, vol. E87-D, pp. 2821-2827 (2004). 6