Human Detection by Boosting-Based Co-occurrence of HOG and Color Feature

Ryo Matsumura[1] and **Akitoshi Hanazawa**[2], Non-members

ABSTRACT

In this paper, we propose a method for human detection using co-occurrence of Histograms of Oriented Gradients (HOG) features and color features. This method expresses the co-occurrence between HOG and color features by Adaboost and generates the combination of the features effective for the identification automatically. Color features were calculated by making histograms that quantized hue and saturation in local areas. We show the effectiveness of the proposed method by identification experiments for human and non-human images.

Keywords: Human detection, Histograms of Oriented Gradients, Color feature, Adaboost

1. INTRODUCTION

Technologies to detect humans in images are very important image technologies because they can be applied on various fields such as robot vision, surveillance system and Intelligent Transport System, etc. In recent methods for human detection, many methods combine local features that represent human appearances like Histograms of Oriented Gradients (HOG) [1] with Adaboost [2], a statistical learning method. Methods combining appearance and other features to improve accuracy has also been proposed, and effectiveness has been reported [3,4,5]. However, these methods [3,4] require consecutive images or a background image.

Color features are effective for object detection and can be extracted from a single image. Takayanagi et al. [5] showed that color features, e.g. skin color of face, are automatically selected and effective for human identification by Adaboost. However, co-occurrence between the features was not considered in this method. In this paper, we propose a method for human detection using co-occurrence of HOG features and color features by Adaboost. Because this study is the first step to examine the efficacy of the co-occurrence of HOG and color features, we employed Adaboost due to its simpleness. This method can generate the combination of the features effective for identification automatically, and highly accurate detection can be expected. We show the effectiveness of our method by identification experiments for human and non-human images.

2. FEATURES

2.1 Histograms of Oriented Gradients

Histograms of Oriented Gradients (HOG) [1] is a feature vector made by histograms of gradient direction of brightness in local areas called cell (Fig.1). HOG has a characteristic robust for local geometry change and illumination change. The calculation procedure of HOG is shown as follows.

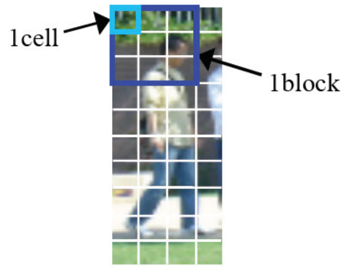

Fig.1: *Cell and block.*

First, calculates gradient intensity m and gradient direction θ by Eq. (1) and (2) from the brightness value L of each pixel.

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

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

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

Next, the gradient direction histograms of brightness are made by using calculated gradient intensity m and gradient direction θ in the cell. The calculated gradient direction θ is converted from 0 to 180 degrees. The obtained gradient direction θ divides into 20 degrees and makes histograms in 9 directions. Finally, the feature value is normalized by Eq. (4) at each block (Fig.1)

Manuscript received on February 28, 2011
[1,2] The authors are with the Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology 2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0916, Japan, E-mail: matsumura-ryo@edu.brain.kyutech.ac.jp and hanazawa@brain.kyutech.ac.jp

$$v' = \frac{v}{\sqrt{\left(\sum_{i=0}^{k} v(i)^2\right) + \varepsilon}} \quad (4)$$

Here, v and v are the values of HOG and normalized HOG feature, k is the number of HOG features in a block, ε is the coefficient that a denominator prevents from being impossible of a calculation in the case of 0.

2.2 Color Feature

The color feature is calculated in each cell as well as HOG. The calculation procedure of the color feature is shown as follows. First, hue and saturation are calculated in a cell. Next, histograms of 9 bins are made by using calculated hue and saturation (Fig.2). Finally, the feature value is normalized at each block as well as HOG. HOG and color feature are independently calculated. Fig.2 shows an example of the feature extraction of HOG and color feature.

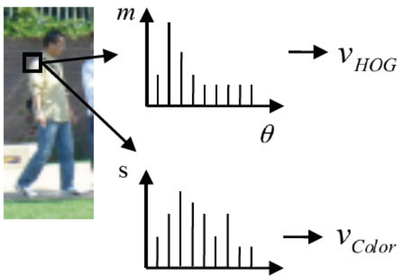

Fig.2: *Example of the feature extraction.*

2.3 Joint Feature

In this method, the co-occurrence of the features were expressed in each weak classifier of Adaboost by observing multiple features at a time [6]. To express the co-occurrence of the features, the sign of binary s that represent a human or non-human is calculated by Eq. (5).

$$s(x) = \begin{cases} 1 & \text{if } p \cdot z(x) > p \cdot \theta \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Here, x is a sample, and z is value of feature to be obtained from the sample. θ is the threshold. p is a sign deciding the direction of inequality sign and takes +1 or -1. As a result, joint features [6] that expresses the co-occurrence of features are obtained by combining observed multiple sign of binary s.

Fig.3 shows an example of two features observed from the sample. When the sign such as 1 is observed for two features, the value of the joint features is calculated by Eq. (6).

$$j = (11)_2 = 3 \quad (6)$$

j is the index number of the feature combinations of the binary representation. In this example, the combination of two features takes four kinds of value (0, 1, 2, 3).

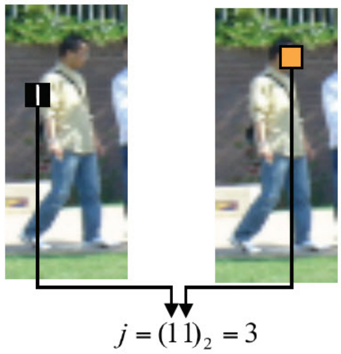

Fig.3: *Example of the joint features.*

3. LEARNING METHOD

3.1 Selecting Joint Features Based on Adaboost

$(x_i, y_i), \ldots, (x_N, y_N)$: N training data.
$y_i \in \{+1, -1\}$: The class label.

1. Initialize weights $D_t(i) = \frac{1}{N}$
2. For $t = 1, \ldots, T$
   A) Binarize feature value by eq.(5).
   B) Train a weak classifier based on a combination features.

The error is calculated by

$$\varepsilon_t = \sum_{i: y_i \neq h_t(x_i)} D_t(i)$$

   C) Choose $h_t(x)$ with lowest error ε_t .
   D) The confidence is calculated by

$$\alpha_t = \frac{1}{2} \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

   E) Update the weights:

$$D_{t+1}(i) = D_t(i) \exp\left(-\alpha_t y_i h_t(x_i)\right)$$

   F) Normalize the weights:

$$D_{t+1}(i) = \frac{D_{t+1}(i)}{\sum_{i=1}^{N} D_{t+1}(i)}$$

3. The strong classifier is:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

Fig.4: *Learning algorithm based on Adaboost*

Fig.4 shows the learning algorithm that selects effective joint features for identification by Adaboost. is a number of training data, is training sample, is the class label, and is a weight of training sample . In each boosting process, best suited joint features are selected by repeating steps from A) to F) in Fig.4.

The function that observes joint features from sample is shown with . When feature value is observed from , weak classifier of Adaboost is shown by Eq. (7).

$$h_t(x) = \begin{cases} +1 & \text{if } P_t(y=+1|j) > P_t(y=-1|j) \\ -1 & \text{otherwise} \end{cases} \quad (7)$$

$P_t(y = +1|j)$ and $P_t(y = -1|j)$ are conditional probability which is human or non-human. Because j is a value to express the co-occurrence of multiple features, $P_t(y = +1|j)$ and $P_t(y = -1|j)$ are joint probabilities of the combined features.

These are calculated based on the weight $D_t(i)$ of the training sample by Eq. (8) and (9).

$$P_t(y=+1|j) = \sum_{p:J_t(x_p)=j \wedge y_i=+1} D_t(i) \quad (8)$$

$$P_t(y=-1|j) = \sum_{p:J_t(x_p)=j \wedge y_i=-1} D_t(i) \quad (9)$$

Fig.5 shows an example of the joint probabilities $P_t(y = +1|j)$ and $P_t(y = -1|j)$ observed from two features.

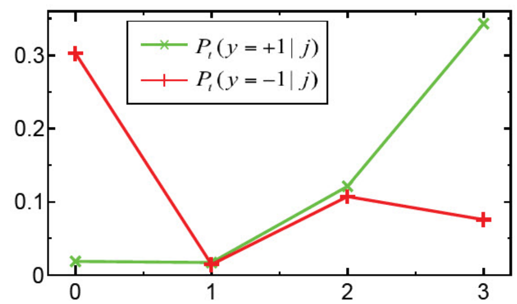

Fig.5: *Example of the joint probabilities.*

3.2 Searching for Combination of Feature

In this paper, we use Sequential Forward Selection (SFS) for the combination search of the feature. SFS can perform an effective search. The combination search of the feature by SFS is shown as follows.

First, feature with the lowest identification error to the training sample is selected. Next, adding another feature to become the lowest identification error. The above is repeated in times of F. As a result, F features are efficiently combinable.

4. EXPERIMENTS

4.1 Experimental Summary

This chapter verifies the property and the effectiveness of proposed method by the following three experiments.

(1) Evaluate the differences in identification accuracy by the number of combined features F.

(2) Evaluate the differences in identification accuracy by the cell size of color feature.

(3) Comparison in performance of our method and HOG.

In all experiments, the total of the feature selected by Adaboost is fixed at 300. Our method uses multiple features in each weak classifier. Therefore the total of the features becomes $T \times F$.

We use the Detection Error Tradeoff (DET) curve [7] for evaluation of identification performance. The DET curve is plotted of false positive rate and miss rate in double logarithmic graphs. That graph shows false positive rate in a horizontal axis and shows miss rate in a vertical axis. The DET curve shows that closer to origin is good performance.

4.2 Training and Test Data

In this paper, we use NICTA Pedestrian Dataset [8] for both the training and test data. The training data consist of 2,500 images of the positive class and 6,500 images of the negative class. The test data consists of 1,000 images of the positive class and 1,000 images of the negative class. Fig.6 shows the example of NICTA Pedestrian Dataset.

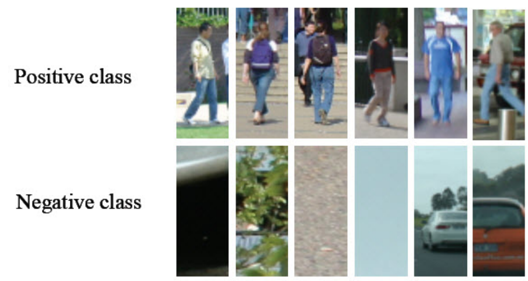

Fig.6: *Example of NICTA Pedestrian Dataset*

4.3 Experiment (1) : Effect of Number of Combined Features

First, the difference in the identification performance by the number of features used for each weak classifier is verified. This experiment compared the performance of the two types of classifier shown in Tab.1.

Fig.7 shows the identification result. Tab.2 shows the performance at false positive 1%. Because DET curve of Classifier B becomes closer to the origin, the identification accuracy of Classifier B is higher. That is, the identification accuracy improves by increasing number of combined features F. However, there is a possibility of over fitting the training data when F is increased.

Table 1: *Experiment (1) parameter.*

	Classifier A	Classifier B
HOG cell size (pixels)	6×6	6×6
Color feature cell size (pixels)	6×6	6×6
Block size (cells)	3×3	3×3
# of combined features F	2	3
# of weak classifier T	150	100
# of features	300	300

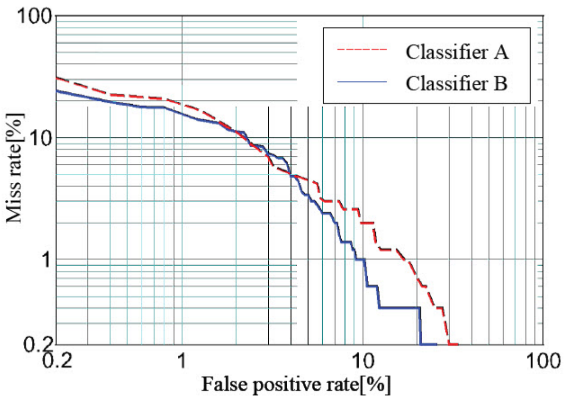

Fig.7: *DET curve of experiment (1).*

Table 2: *Performance at false positive 1% for experiment (1).*

	Classifier A	Classifier B
Miss rate [%]	17.9	15.9
Identification rate [%]	90.5	91.5

4.4 Experiment (2) : Effect of Cell Size of Color Feature

Next, the difference in the identification performance by the cell size of color feature is verified. This experiment compared the performance of Classifier B shown in Tab.1 and three types of classifier shown in Tab.3.

Fig.8 shows the identification result. Tab.4 shows the performance at false positive 1%. We expected color feature effective for human identification can be extracted by changing the cell size of the color feature and subdividing the cell area.

However, from the result, we can see that doesn't contribute to improvement of identification accuracy even if the cell size of color feature is changed. The reason for this result is that Classifier C, D and E are over fitting the training data. Classifier B is the best performance in this experiment.

Table 3: *Experiment (2) parameter.*

	Classifier C	Classifier D	Classifier E
HOG cell size (pixels)	6×6	6×6	6×6
Color feature cell size (pixels)	3×3	4×4	5×5
Block size (cells)	3×3	3×3	3×3
# of combined features F	3	3	3
# of weak classifier T	100	100	100
# of features	300	300	300

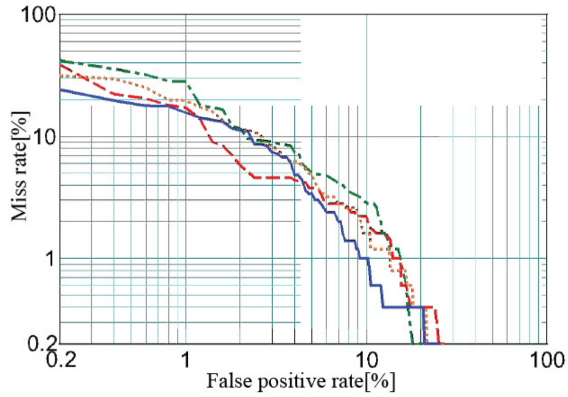

Fig.8: *DET curve of experiment (2).*

Table 4: *Performance at false positive 1% for experiment (2).*

	Classifier C	Classifier D	Classifier E
Miss rate [%]	19.6	17.4	26.8
Identification rate [%]	89.7	90.8	86.1

4.5 Experiment (3) : Comparison with HOG

Finally, the identification performances of our method and HOG are compared. Classifier B shown in Tab.1 is used for the comparison in this experiment. The parameter of the classifier using HOG is shown in Tab.5.

Table 5: *Experiment (3) parameter.*

cell size (pixels)	6×6
Block size (cells)	3×3
# of weak classifier T	300
# of features	300

Table 6: *Performance at false positive 1% for experiment (3).*

	HOG	Our method
Miss rate [%]	23.4	15.9
Identification rate [%]	87.7	91.5

Fig.9 shows the identification result. Tab.6 shows

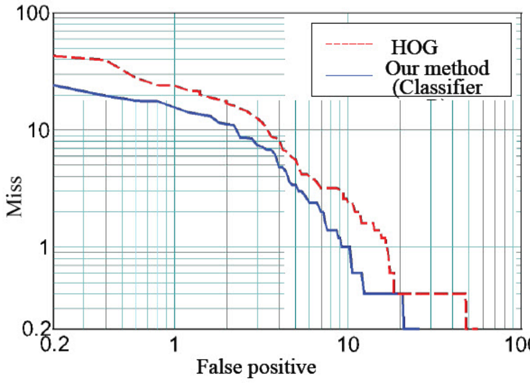

Fig.9: *DET curve of experiment (3).*

the performance at false positive 1%. Because the DET curve of Classifier B is closer to the origin, the identification accuracy of our method is higher than HOG. When false positive is 1%, miss rate of our method is 15.9% where as miss rate of HOG is 23.4%. Our method is improved the identification accuracy by 3.8%. When false positive is 4.8%, our method recorded the best identification rate, that is, 95.9%. Because effective feature for identification can be extracted by combine HOG with the color features, our method can identify difficult patterns for the identification by HOG only.

5. CONCLUSION

In this paper, we proposed a method for human detection using co-occurrence of Histograms of Oriented Gradients (HOG) features and color features. The accuracy of human detection was improved by expressing the co-occurrence of these features. We have not analyzed what features are selected and combined by Adaboost. We are going to analyze it to improve the performance of our algorithm. Further, by extending the learning algorithm to Real Adaboost may improve the identification accuracy.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection, *IEEE Computer Vision and Pattern Recognition*, vol.1, pp.886-893, 2005.

[2] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, vol.55, pp.119-139, 1997.

[3] P. Viola, M. Jones and D. Snow, "Detecting pedestrians using patterns of motion and appearance, *IEEE International Conference on Computer Vision*, pp.734-741, 2003.

[4] Y. Yamauchi, H. Fujiyoshi, H. Hwang, B.-W. and T. Kanade, "People detection based on co-occurrence of appearance and spatiotemporal features, *IEEE International Conference on Pattern Recognition*, 2008.

[5] Y. Takayanagi and J. Katto, "Human body detection using HOG with additional color features, *International Workshop on Advanced Image Technology*, 2010.

[6] T. Mita, T. Kaneko, and O. Hori, "Joint haar-like features for face detection, *IEEE International Conference on Computer Vision*, 2005.

[7] A. Martin, G. Doddington, T. Kamm, M. Ordowski and M. Przydocki, "The DET curve in assessment of detection task performance, *European Conference on Speech Communication and Technology*, pp.1895-1898, 1997.

[8] NICTA Pedestrian Dataset, URL: http://nicta.com.au/research/project_list/completed_projects/smart_cars/computer_vision_datasets/

Ryo Matsumura received M.S. degree from Kyushu Institute of Technology, Japan, in 2011.

Akitoshi Hanazwa received B.S., M.S. and Ph.D digree from Kyoto University, Japan, in 1990, 1992 and 1995, respectively. From 1995 to 2002, he was an Assistant Professor of National Institute of Physiological Sciences. Since 2002, he has been an Associate Professor of Kyushu Institute Technology. His research interests includes Visual Neuroscience, Psychophysics and Image Processing.