International Journal of Innovative Computing, Information and Control
Volume **15**, Number **4**, August **2019**

ICIC International ©2019 ISSN 1349-4198
pp. 1211–1222

HUMAN DETECTION USING COLOR CONTRAST-BASED HISTOGRAMS OF ORIENTED GRADIENTS

Ryo Matsumura[1] and Akitoshi Hanazawa[2]

[1]Information Science and Technology Department
National Institute of Technology, Oshima College
1091-1 Oaza-komatsu, Suooshima-cho, Oshima-gun, Yamaguchi 742-2193, Japan
matumura@oshima-k.ac.jp

[2]Faculty of Engineering
Kyushu Institute of Technology
1-1 Sensuicho, Tobata-ku, Kitakyushu 804-8550, Japan
hanazawa@mns.kyutech.ac.jp

Received October 2018; revised February 2019

Abstract. *In this paper, we propose a method for human detection using color contrast-based Histograms of Oriented Gradients (HOG). The proposed method calculates the color similarities between a pixel of interest and the pixels in its neighborhood. By applying the same gradient calculations as HOG on these color similarities, gradient orientation histogram can be made for color contrast. It can capture edge information derived from color contrast even when the luminance contrast is small. We evaluate the proposed method using the following three types of classifiers in the INRIA and NICTA datasets: Real Adaboost, Support Vector Machine (SVM), and random forest. As a result, the proposed method exhibits higher performance than HOG.*

Keywords: Human detection, Histograms of oriented gradients, Color contrast, Color similarity, Real Adaboost, Support vector machine, Random forest

1. Introduction.

Human detection technologies in the image recognition field can be applied to surveillance systems and Intelligent Transport System (ITS) [1, 2, 3]. In addition, these technologies are applied to the preprocessing of human tracking in human behavior analysis systems that are used in fields such as marketing. Practical use of the human detection technology can be found in various fields.

Most of the current methods that are used for object and human detection combine local features with the statistical machine learning method as typified by the Viola-Jones face detection method [4]. For human detection, effective local feature based on luminance information has been proposed. One of the methods employed is the Edge of Orientation Histogram (EOH) [5]. It generates the orientation histogram by extracting edges using the Sobel filter and calculating its gradient. Edgelet [6] performs calculations based on the difference in edge orientation (calculated based on Sobel filter) between the defined shape pattern and the local areas of the image.

In addition to these, Histograms of Oriented Gradients (HOG) [7] has been proposed. HOG is also based on luminance information. This feature calculates the gradient orientations and intensities of luminance and can capture the rough shape of the objects to be detected.

DOI: 10.24507/ijicic.15.04.1211

HOG is the most successful feature for object detection and various extensions have been proposed. Extended Histograms of Oriented Gradients (EHOG) [8] reduce the calculation cost to a greater extent than HOG because it involves dimension reduction. Co-occurrence Histograms of Oriented Gradients (CoHOG) [9] represent co-occurrence by making histograms using a combination of gradient orientations. S-HOG [10] captures the symmetry of the object by dividing the image into three blocks and making the gradient direction histogram in these blocks. In addition, HOG or HOG-based features are used in various applications such as gait recognition [11], ship detection [10], and landmine detection [12].

To utilize color information for object recognition, Color Self-Similarity (CSS) [13] has been proposed. The CSS feature calculation is based on color similarity between a local area of interest and other local areas, thereby capturing the probability that the two areas belong to a single object.

In the case of human detection, it is difficult to extract an effective color feature because the colors of clothing vary depending on the individual. However, the CSS feature makes color information effective for human detection.

Present study is an attempt to utilize color-defined edge information for object recognition because both HOG and CSS cannot capture color-defined edge information. Although HOG can capture edge information by luminance contrast, it cannot capture color information. In contrast, CSS can capture color correlation, but it cannot capture the edge information derived from color contrast.

Multiple color-defined edge detection methods have been proposed [14, 15]. However, these are only proposals and there are currently no study cases that have been applied to object recognition.

Wang et al. [16] used color information for object recognition. They used a color HOG and an HOG calculated using an integral histogram to detect traffic signs. Note that the color HOG only calculates the HOG for each RGB channel; thus, it cannot capture the color-defined edge information derived from color contrast.

In a study by Goto et al. [19] color-defined edge information for object recognition is utilized. A Color Similarity-based Histogram of Oriented Gradients (CS-HOG) [19] uses the CSS approach to calculate the color similarity between the area of interest and all the pixels in the image. It generates multiple color similarity images based on each region; HOG is then calculated from these color similarity images. However, this feature has the largest number of feature dimensions because it calculates HOG from several color similarity images that are generated from an input image. The reduction of memory resources is also required for embedded systems such as ITS as it is important to reduce feature dimensions.

In this paper, we propose a method for human detection using color contrast-based HOG. The proposed method extracts edge information for calculating gradient orientations and intensities based on the color similarity between a pixel of interest and the pixels in its neighborhood. In addition, by adopting the pixel-based similarity calculation to capture the edge information derived from color contrast, it is also possible to reduce the feature dimension.

This paper is organized as follows. In Section 2, the proposed method is described. Section 3 shows the experimental results. Discussion is presented in Section 4. Conclusions are given in Section 5.

2. Color Contrast-Based HOG.

In the proposed method, gradient orientation histograms are made by calculating gradient based on color similarities in local areas (called cells) of an image. Neighboring cells overlap with a position shift of one pixel (as shown in

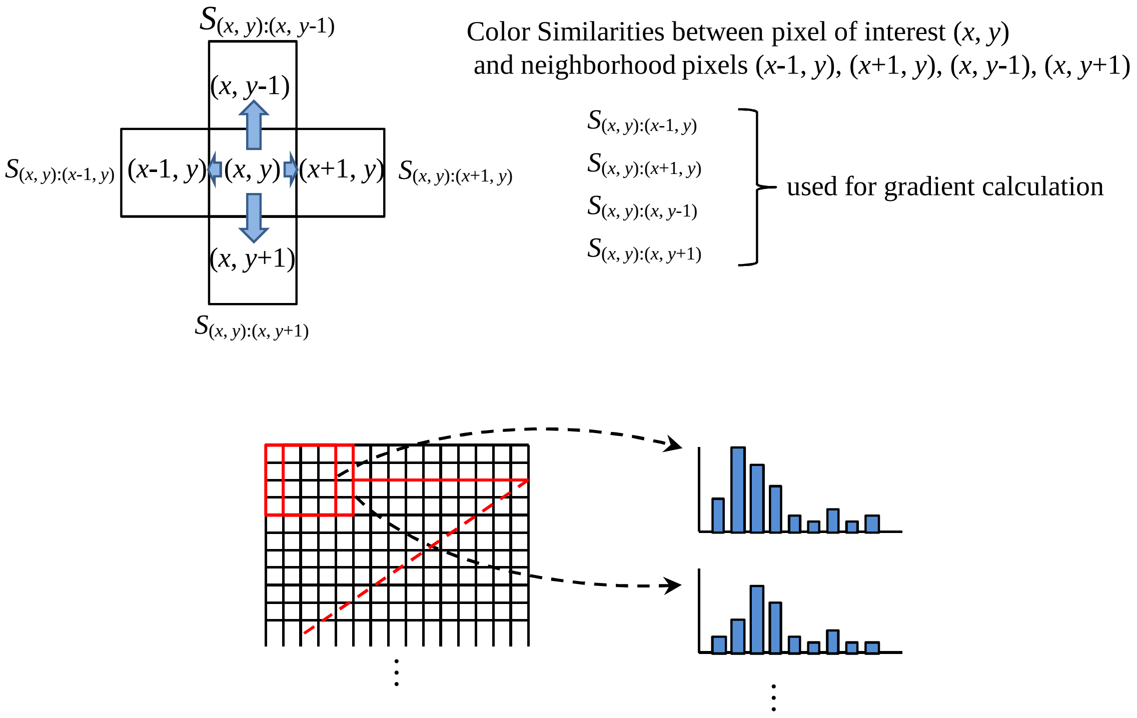


Figure 1. Proposed method overview

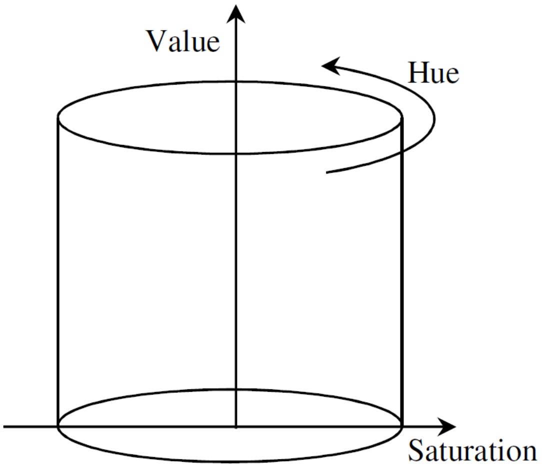


Figure 2. HSV color space

Figure 1). Figure 1 shows the proposed method overview. The upper row shows the color similarities that are used for gradient calculation. The lower row exhibits cell overlap and the histograms made by them. Cell overlap is used to generate various features.

The HSV color space (Figure 2) is used for the calculation of color similarities. That color space is represented using a cylindrical model where H is hue and it is represented in angular degree. S and V are saturation and value (or brightness), respectively. S and V values range from 0 to 1.0. Color similarity is calculated based on Euclidean distance. The value of H, S and V cannot be used for distance calculation directly because the HSV color space is in a polar coordinate system. Therefore, we convert the H, S and V

elements into Cartesian coordinate system represented by $\{u, r, v\}$ in Equation (1).

$$\begin{cases} u = \cos H \times S \\ r = \sin H \times S \\ v = V \end{cases} \tag{1}$$

After this coordinate conversion, Euclidean distance D is calculated using Equation (2), and similarity S is calculated using Equation (3). Subscripts i and n in Equations (2) and (3) are the position (x, y) of the pixel of interest and those of neighborhood pixels. Note that n includes positions $(x-1, y)$, $(x+1, y)$, $(x, y-1)$ or $(x, y+1)$.

$$D_{i:n\in\{(x-1,y),(x+1,y),(x,y-1),(x,y+1)\}} = \sqrt{(u_i - u_n)^2 + (r_i - r_n)^2 + (v_i - v_n)^2} \tag{2}$$

$$S_{i:n\in\{(x-1,y),(x+1,y),(x,y-1),(x,y+1)\}} = \frac{1}{D_{i:n} + 1} \tag{3}$$

Gradient orientation histograms are made by calculating gradient orientations and intensities based on the color similarities calculated using the above procedure. Gradient intensity m and orientation θ are calculated using Equations (4) and (5), respectively. $f_x(x, y)$ and $f_y(x, y)$ are calculated using Equation (6).

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \tag{4}$$

$$\theta = \tan^{-1} \frac{f_y(x, y)^2}{f_x(x, y)^2} \tag{5}$$

$$\begin{cases} f_x(x, y) = S_{(x,y):(x+1,y)} - S_{(x,y):(x-1,y)} \\ f_y(x, y) = S_{(x,y):(x,y+1)} - S_{(x,y):(x,y-1)} \end{cases} \tag{6}$$

As for color similarity, there is no variation for each image; this is unlike luminance value, in which variation exists. Therefore, the proposed method does not perform normalization processing. To prevent gradient reversal, the orientation is converted between 0 to 180 degrees. Gradient orientation histogram consists of nine bins which divide the range 0 to 180 degrees into segments of 20 degrees each.

We adopted the Euclidean distance for the color similarity calculation for the following reasons. The color similarity can be calculated as a color difference. The difference between two colors in a three-dimensional color space, such as the CIE-L*a*b* and L*u*v* color spaces, can be calculated as the Euclidean distance [17]. The similarity between two colors in the HSV color space can also be calculated using the Euclidean distance [18]. The aim of the proposed method is to enable edge detection under the condition where the luminance difference is small but a color contrast exists. In the case of the detection of a human, there is no regularity concerning the luminance and color differences between a human in the foreground (primarily clothing) and the background. Therefore, we believe that it is straightforward to simplify vector information, such as luminance and color differences of an edge part, to the Euclidean distance, which is scalar information. In addition, the Euclidean distance has the advantages of low calculation costs and ease of manipulation in feature calculations.

The methods introduced in [14] require calculations to integrate the results obtained from individual channels or because they detect edges from vector information; accordingly, the calculation costs for detecting color-defined edges are high. In addition, because many methods use the RGB color space [14, 15], there is the possibility that such methods are not robust to illumination variations. The proposed method simplifies vector information, such as luminance and color differences, to scalar information; therefore, the calculation costs for detecting color-defined edges are low. Further, because the HSV color space is adopted, the proposed method is robust to illumination variations.

3. Experimental Results.

3.1. Experimental overview. This section verifies the effectiveness of the proposed method using evaluation experiments. We use Real Adaboost [20], Support Vector Machine (SVM) [21], and random forest [22] for training. We perform the following two experiments: (i) comparison of classification performance of the proposed method and HOG, (ii) comparison of the detection performance of the proposed method and HOG. The evaluation methods in these experiments are False Positive Per Window (FPPW) and False Positive Per Image (FPPI), and the Detection Error Tradeoff (DET) curve is used.

In this paper, we use the INRIA Person dataset [7] and NICTA Pedestrian dataset [23]. The training data in the INRIA dataset consists of 2,416 and 4,832 images of positive and negative samples, respectively. The training data in NICTA dataset consists of 3,000 and 6,500 images of positive and negative samples, respectively. The test data in the INRIA dataset consists of 1,132 images of positive and negative samples, respectively. The test data in NICTA dataset consists of 1,000 images of positive and negative samples, respectively. Figure 3 shows an example of INRIA and NICTA datasets. The upper row shows an example of positive and negative samples of the INRIA Person dataset. The lower row shows an example of the positive and negative samples of the NICTA Pedestrian dataset. For details regarding each dataset, see [7, 23].

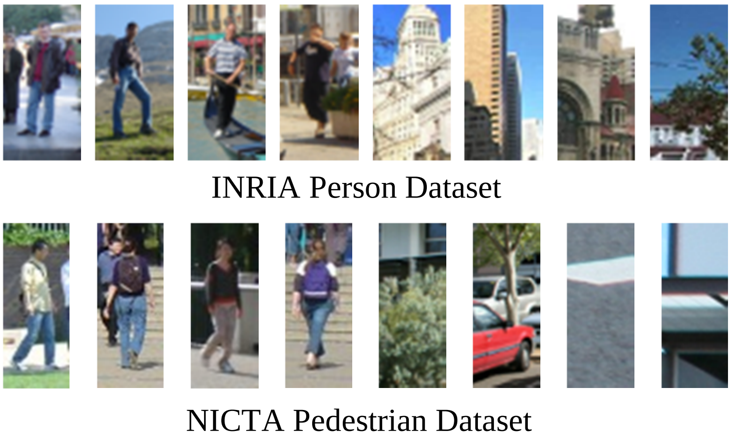
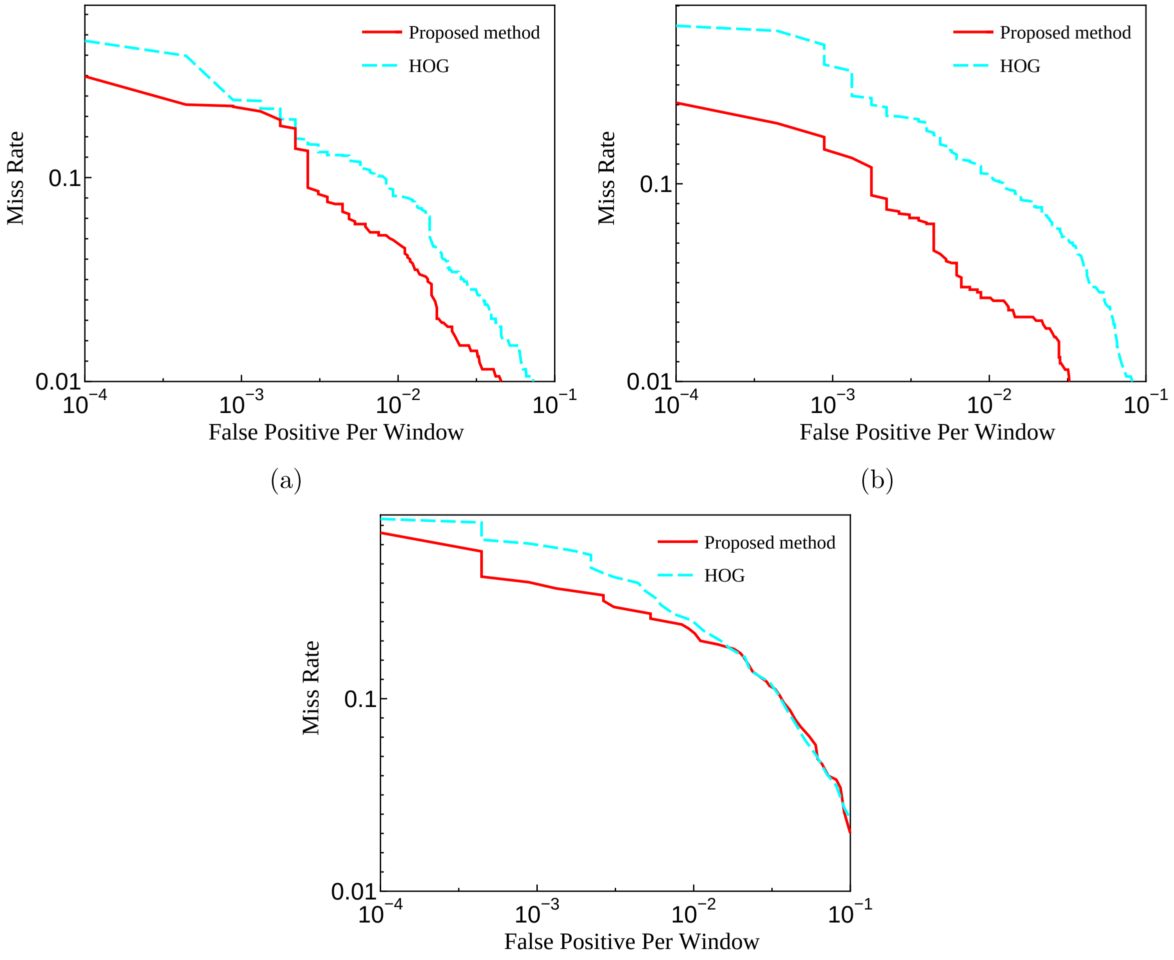


Figure 3. An example of the INRIA Person and NICTA Pedestrian datasets

3.2. Comparison with HOG in FPPW. Figure 4 and Figure 5 show the comparison between the proposed method and HOG. Figure 4 shows the results of the INRIA Person dataset, and Figure 5 shows the results of the NICTA Pedestrian dataset. (a), (b), and (c) show the results of Real Adaboost, SVM, and random forests. The solid line is the proposed method and the dashed line is HOG.

We use under 10^{-3} FPPW for evaluation. In surveillance systems, and ITS etc., it is important to reduce both false positives and miss detection. SVM and random forest is trained with LIBSVM [24] and Scikit-learn [25], respectively. Real Adaboost is trained in our implementation.

It can be seen from Figure 4 and Figure 5 that the proposed method improves the performance. In all results, the proposed method reduced the miss rate to under 10^{-3} FPPW. Table 1 and Table 2 show the comparison of miss rate with HOG in INRIA and NICTA datasets, respectively. (a), (b) and (c) show the results of Real Adaboost,



(a)

(b)

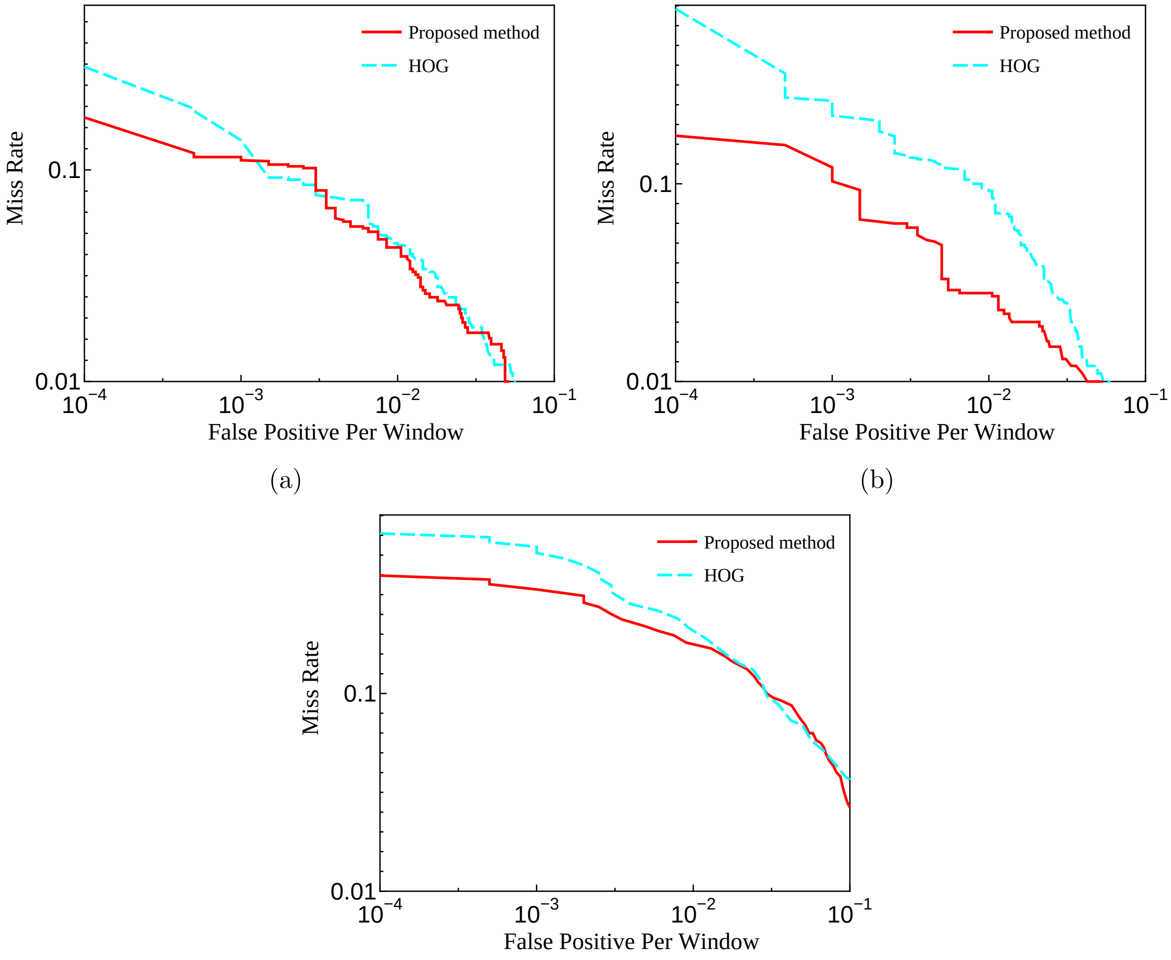
(c)

Figure 4. DET curves of the INRIA Person dataset

SVM, and random forest, respectively. In INRIA Person dataset, the Real Adaboost-based classifier reduces the miss rate by 1.8% points at 9×10^{-4} FPPW. The SVM-based classifier reduces the miss rate by 25.2% points at 9×10^{-4} FPPW. The random forest-based classifier reduces the miss rate by 23.7% points at 9×10^{-4} FPPW.

In the NICTA Pedestrian dataset, the Real Adaboost-based classifier reduces the miss rate by 2.7% points at 10^{-3} FPPW. The SVM-based classifier reduces the miss rate by 11.8% points at 10^{-3} FPPW. The random forest-based classifier reduces the miss rate by 17.8% points at 10^{-3} FPPW. Figure 6 shows the graphs of Table 1 and Table 2 and the miss rate comparison of each classifier. (a) and (b) show the results in the INRIA and NICTA datasets, respectively. The vertical axis of the graph indicates the miss rate. In (a) and (b), the miss rates of Real Adaboost, SVM, and random forest are arranged in order from left to right. The miss rates of HOG and the proposed method are represented by the left and right bars, respectively, for the Real Adaboost, SVM, and random forest. The result shows that SVM has the best performance in the proposed method.

3.3. **Comparison with HOG in FPPI.** In FPPI evaluation, we use frame images and annotation data in INRIA Person dataset. The NICTA Pedestrian dataset has no frame images and annotation data. In this experiment, SVM is used because it showed the best results in Section 3.2. Non maximum suppression [26] is used for integration of detected windows.



(a)

(b)

(c)

Figure 5. DET curves of NICTA Pedestrian dataset

Table 1. Miss rate comparison with HOG in INRIA Person dataset at 9×10^{-4} FPPW

(a)

	miss rate
HOG	0.240
Proposed method	0.222

(b)

	miss rate
HOG	0.401
Proposed method	0.149

(c)

	miss rate
HOG	0.639
Proposed method	0.402

Table 2. Miss rate comparison with HOG in NICTA Pedestrian dataset at 10^{-3} FPPW

(a)

	miss rate
HOG	0.138
Proposed method	0.111

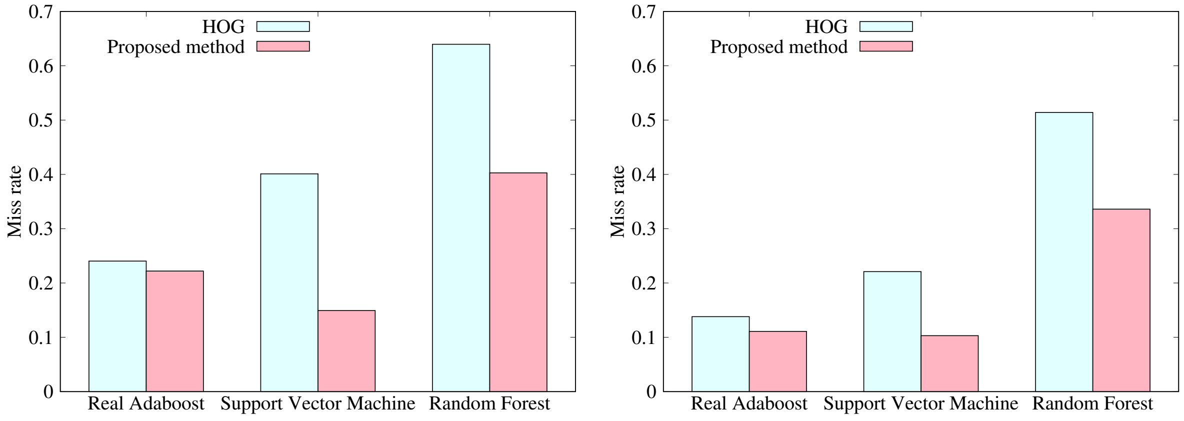
(b)

	miss rate
HOG	0.221
Proposed method	0.103

(c)

	miss rate
HOG	0.514
Proposed method	0.336

Figure 7 shows the DET curve in this experiment. The solid line is the proposed method and the dashed line is HOG. Even in FPPI, the proposed method can improve



(a) (b)

Figure 6. Miss rate comparison of Real Adaboost, SVM, and random forest

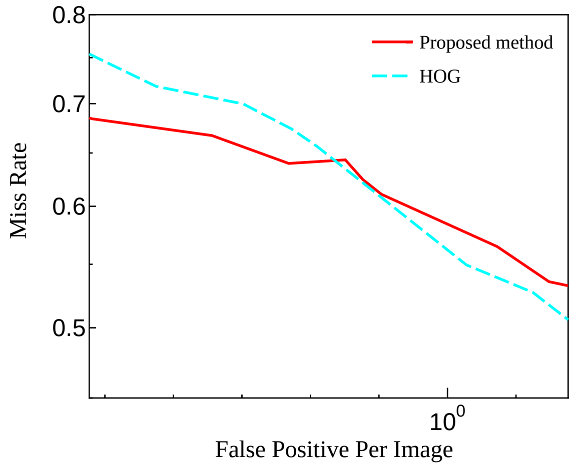


Figure 7. DET curve of the INRIA Person dataset in FPPI

performance and reduces the miss rate by approximately 8.2% points at approximately 25×10^{-2} FPPI. After approximately 9×10^{-1} FPPI, the performance of HOG exceeds that of the proposed method. However, as mentioned above, because it is important to reduce both false positives and miss detection, it is assumed that miss rate reduction at low FPPI is of importance for human detection.

4. **Discussion.** Experimental results show the effectiveness of the proposed method. Figure 8 shows the gradient images based on the gradient calculations in HOG and the proposed method. (a) is an input image. (b) and (c) are gradient images based on the gradient calculations in HOG and the proposed method, respectively. The red rectangle in the figure shows the area of interest. In Figure 8(c), the brightness is set higher than the original in order to be easy to see (Figure 9(c) and Figure 10(c) do the same). In Figure 8(b), gradient calculation in HOG cannot detect the edge of the shoulder part of the human because of the low luminance contrast condition, whereas in Figure 8(c), gradient calculation in the proposed method can detect the edge even under such condition.

Figure 9 shows examples of samples which were classified correctly by the proposed method but not by HOG. (a) is an input image. (b) and (c) are gradient images based on the gradient calculation in HOG and the proposed method, respectively. Shown in Figure 9(b) are samples containing low luminance contrast edges in the human body.

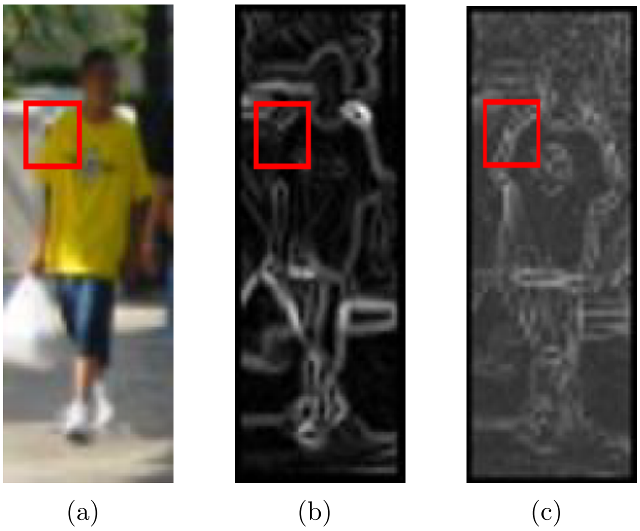


FIGURE 8. Gradient images based on the gradient calculations in HOG and the proposed method

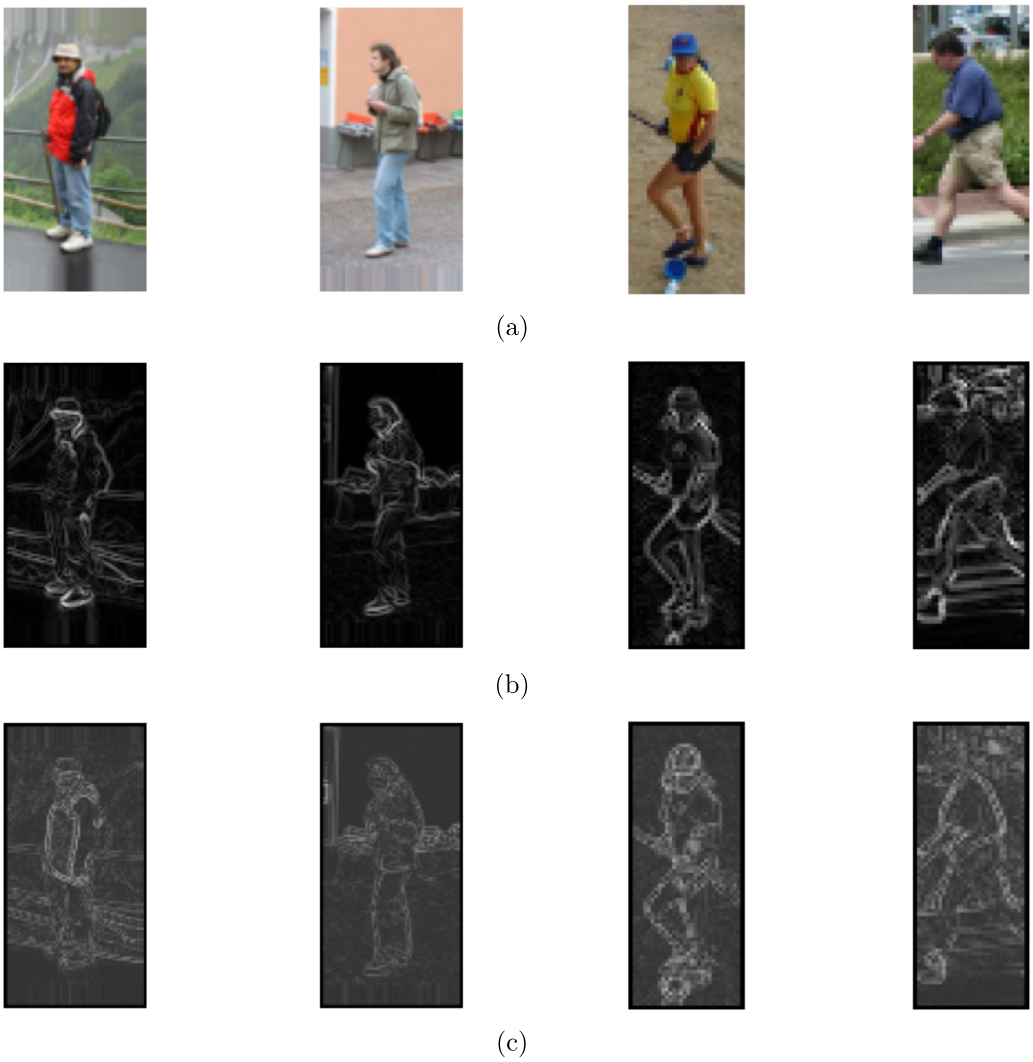
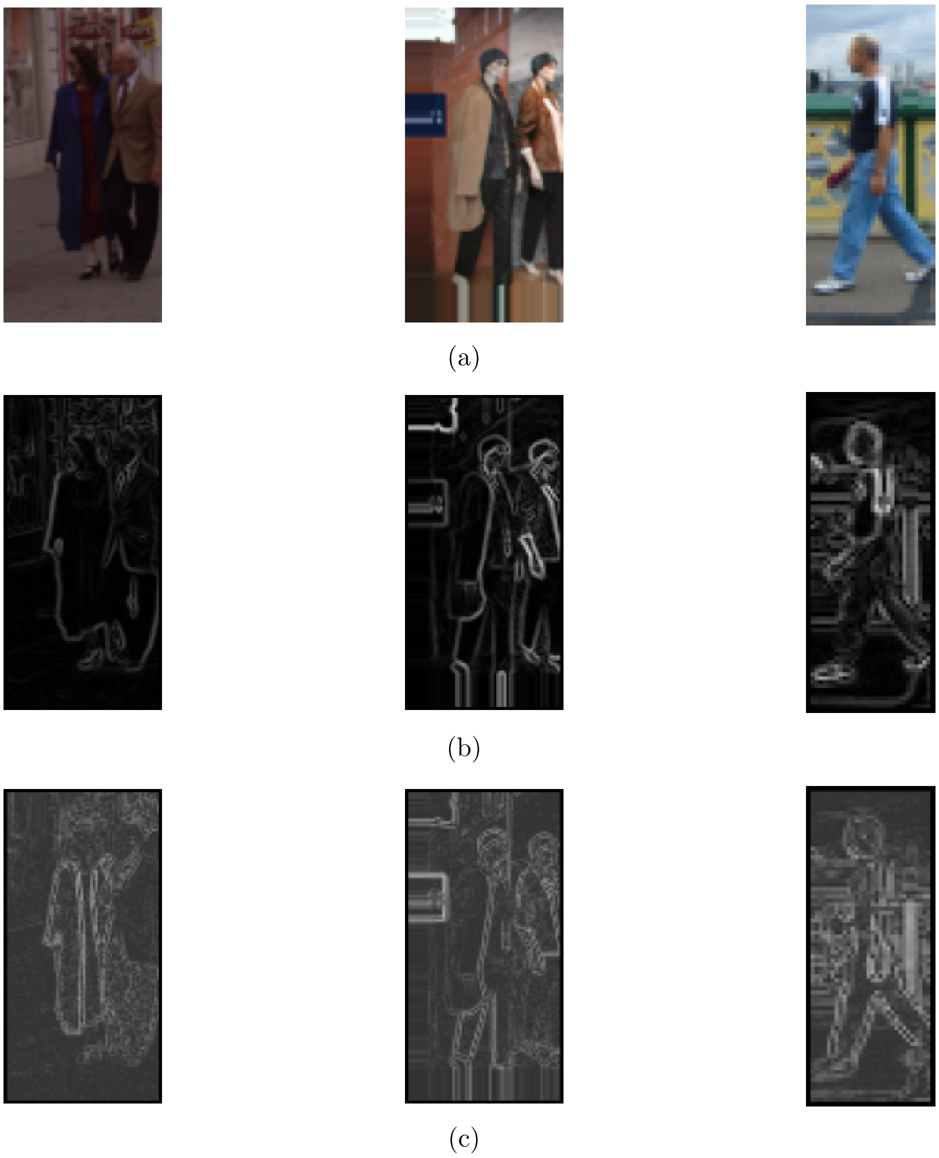


FIGURE 9. Examples of samples which were classified correctly by the proposed method but not by HOG



(a)

(b)

(c)

Figure 10. Examples of samples which were classified correctly by HOG but not by the proposed method

Gradient calculation in the proposed method can detect the edge of the body part (see the first, second, and fourth column in Figure 9(c)). It can also detect the edge of the head and right shoulder (see the third columns in Figure 9(c)). Thus, the proposed method classified these samples correctly.

Figure 10 shows examples of samples which were classified correctly by HOG but not by the proposed method. (a) is an input image. (b) and (c) are gradient images based on the gradient calculation in HOG and the proposed method, respectively. The edges of the inner and outer boundaries of clothing (see the first column in Figure 10(c)) and edges of the background structure (see the second and third column in Figure 10(c)) were detected. It is assumed that they become noise and cannot classify correctly. However, since there are many samples that can be classified correctly using the proposed method, it is effective for improving the performance of human detection.

The number of feature dimensions of CS-HOG is 967,680 dimensions for an image of 64 × 128 pixels of the INRIA Person dataset [19]. On the other hand, the number of feature dimensions of the proposed method is 62,073 dimensions when using the same image and cell size: a significant reduction is thus achieved.

The HSV color space is not designed to be a uniform color space. Therefore, there is the possibility that inhomogeneities may occur in the color difference. We compared the performance of the proposed method using the Euclidean distance and the normalized Euclidean distance for the color similarity calculation. We found no significant performance difference between the methods. (Even though the results for the INRIA dataset show some performance improvement, the results for the NICTA dataset show a performance degradation.) Considering the above result and the calculation cost, a color similarity calculation based on the Euclidean distance is preferable.

5. **Conclusions.** We proposed a method for human detection using color contrast-based HOG. The proposed method can capture edge information derived from color contrast. It can detect human even under low luminance contrast conditions, which cannot be detected by HOG.

We confirmed that the proposed method improves human detection performance through experiments with FPPW and FPPI. Additionally, we showed that the performance improvement through the proposed method does not depend on learning methods and datasets by performing experiments using Real Adaboost, SVM, and random forest in the INRIA and NICTA datasets. It is assumed that this method is particularly effective under low luminance conditions such as a cloudy day.

Future work aims to speed up the calculation of the proposed method and to perform human detection in real time.

REFERENCES

[1] C. Papageorgiou and T. Poggio, Trainable pedestrian detection, *Proc. of International Conference on Image Processing*, vol.4, pp.35-39, 1999.

[2] D. M. Gavrila and S. Munder, Multi-cue pedestrian detection and tracking from a moving vehicle, *International Journal of Computer Vision*, vol.73, no.1, pp.41-59, 2007.

[3] H. Cho, Y. W. Seo, B. V. Kumar and R. R. Rajkumar, A multi-sensor fusion system for moving object detection and tracking in urban driving environments, *IEEE International Conference on Robotics and Automation (ICRA)*, pp.1836-1843, 2014.

[4] P. Viola and M. J. Jones, Robust real-time face detection, *International Journal of Computer Vision*, vol.57, no.2, pp.137-154, 2004.

[5] K. Levi and Y. Weiss, Learning object detection from a small number of examples: The importance of good features, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, 2004.

[6] B. Wu and R. Nevatia, Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors, *IEEE International Conference of Computer Vision*, vol.1, pp.90-97, 2005.

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

[8] C. Hou, H. Ai and S. Lao, Multiview pedestrian detection based on vector boosting, *Asian Conference on Computer Vision*, 2007.

[9] T. Watanabe, S. Ito and K. Yokoi, Co-occurrence histograms of oriented gradients for pedestrian detection, *Pacific-Rim Symposium on Image and Video Technology*, 2009.

[10] S. Qi, J. Ma, J. Lin, Y. Li and J. Tian, Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images, *IEEE Geoscience and Remote Sensing Letters*, vol.12, no.7, pp.1451-1455, 2015.

[11] Y. Liu, J. Zhang, C. Wang and L. Wang, Multiple HOG templates for gait recognition, *The 21st International Conference on Pattern Recognition (ICPR)*, 2012.

[12] P. A. Torrione, K. D. Morton, R. Sakaguchi and L. M. Collins, Histograms of oriented gradients for landmine detection in ground-penetrating radar data, *IEEE Trans. Geoscience and Remote Sensing*, vol.52, no.3, pp.1539-1550, 2014.

[13] S. Walk, N. Majer, K. Schindler and B. Schiele, New features and insights for pedestrian detection, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[14] A. Koschan and M. Abidi, Detection and classification of edges in color images, *IEEE Signal Processing Magazine*, vol.22, no.1, pp.64-73, 2005.

[15] S. Dutta and B. B. Chaudhuri, A color edge detection algorithm in RGB color space, *International Conference on Advances in Recent Technologies in Communication and Computing*, 2009.

[16] D. Wang, X. Hou, J. Xu, S. Yue and C. L. Liu, Traffic sign detection using a cascade method with fast feature extraction and saliency test, *IEEE Trans. Intelligent Transportation Systems*, vol.18, no.12, pp.3290-3302, 2017.

[17] M. Tkalcic and J. F. Tasic, Colour spaces: Perceptual, historical and applicational background, *The IEEE Region 8 EUROCON 2003. Computer as a Tool*, Ljubljana, Slovenia, vol.1, pp.304-308, 2003.

[18] J. R. Smith, *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis*, Ph.D. Thesis, Columbia Univ., New York, 1997.

[19] Y. Goto, Y. Yamamoto and H. Fujiyoshi, Color similarity-based HOG, *The IEICE Trans. Information and Systems (Japanese Edition)*, vol.96, no.7, pp.1618-1626, 2013 (in Japanese).

[20] R. E. Schapire and Y. Singer, Improved boosting algorithms using confidence-rated predictions, *Machine Learning*, vol.37, no.3, pp.297-336, 1999.

[21] B. E. Boser, I. M. Guyon and V. N. Vapnik, A training algorithm for optimal margin classifiers, *Proc. of the 5th Annual Workshop on Computational Learning Theory*, 1992.

[22] L. Breiman, Random forests, *Machine Learning*, vol.45, no.1, pp.5-32, 2001.

[23] G. Overett, L. Petersson, N. Brewer, L. Andersson and N. Pettersson, A new pedestrian dataset for supervised learning, *Intelligent Vehicles Symposium*, 2008.

[24] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intelligent Systems and Technology (TIST)*, vol.2, no.3, 2011.

[25] F. Pedregosa et al., Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, pp.2825-2830, 2011.

[26] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, Object detection with discriminatively trained part-based models, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.32, no.9, pp.1627-1645, 2010.