

SCHOG Feature for Pedestrian Detection

Ryuichi Ozaki^(✉) and Kazunori Onoguchi

Graduate School of Science and Technology, Hirosaki University, 3 Bunkyo-cho,
Hirosaki, Aomori 036-8561, Japan
h13ms504@stu.hirosaki-u.ac.jp, onoguchi@eit.hirosaki-u.ac.jp

Abstract. Co-occurrence Histograms of Oriented Gradients (CoHOG) has succeeded in describing the detailed shape of the object by using a co-occurrence of features. However, unlike HOG, it does not consider the difference of gradient magnitude. In addition, the dimension of the CoHOG feature is also very large. In this paper, we propose Similarity Co-occurrence Histogram of Oriented Gradients (SCHOG) considering the similarity and co-occurrence of features. Unlike CoHOG, SCHOG quantize edge gradient direction to four directions. Therefore, the feature dimension for the co-occurrence between edge gradient direction decreases greatly. In addition, the binary code representing the similarity between features is introduced. In spite of reducing the resolution of the edge gradient direction, SCHOG realizes higher performance and lower dimension than CoHOG by adding this similarity. In experiments using the INRIA Person Dataset, SCHOG is evaluated in comparison with the conventional CoHOG.

Keywords: Pedestrian detection · Co-occurrence histograms of oriented gradients · Similarity · Support vector machine

1 Introduction

Recently, a pedestrian detection system have been put to practical use as a vehicle safety device [1]. Since features expressing characteristics of a person well is important in these system, various features for pedestrian detection have been proposed. T. Ojala et al. proposed the Local Binary Pattern (LBP) [2] representing the relation between the intensity of an interest pixel and the intensity of eight adjacent pixels. This feature has been studied in various ways because it's robust to illumination change and it's implemented easily. Y. Cao et al. proposed the Advanced LBP [3] which is robust to noise and low intensity. N. Dalal proposed HOG [4] feature which is robust to the change of the pedestrian's posture and the change of the illumination by generating the histogram of the edge gradient orientation in each block and normalizing each block for every cell. They also proposed the feature focusing on the gradient orientation of the time series [5]. T. Watanabe et al. proposed CoHOG [6] feature that represented the co-occurrence of gradient orientation and showed high performance for pedestrian detection. As other features using the co-occurrence, T. Kobayashi et al. proposed a Gradient Local Auto-Correlation (GLAC) [7] which calculated

the autocorrelation of the position and edge gradient orientation. K. Yamaguchi et al. proposed a two-dimensional gradient orientation histogram using polar coordinates which can express small difference [8]. S. Walk et al. proposed the Color Self-Similarity (CSS) [9] feature using the similarity of HSV histogram in the local area. As mentioned above, the co-occurrence of feature is effective for improving the performance of pedestrian detection. However, there is a problem that the dimension of the feature increases significantly.

In this paper, we propose SCHOG which consists of the co-occurrence of edge gradient direction and the similarity. Although SCHOG quantizes edge gradient direction to the half of CoHOG, it can represent the shape of the object more finely than CoHOG by adding the similarity to the co-occurrence of edge gradient direction. Because the similarity is represented by the binary code, the dimension of SCHOG is a half of the conventional CoHOG in spite of adding the similarity. We evaluate three kind of similarity, such as the pixel intensity, the edge gradient magnitude and the edge gradient direction in the experiment. These values are not used directly in CoHOG. Therefore, the proposed feature can compensate information lost in CoHOG. In the experiment, the edge gradient magnitude showed the best performance.

Experimental result using the INRIA Person Dataset and the Support Vector Machine shows that the performance of SCHOG is better than the conventional CoHOG.

The rest of this paper is organized as follow. In Sect. 2, the proposed method is explained in detail and its extensibility is discussed. In Sect. 3, the performance of the SCHOG is evaluated by comparing with the conventional CoHOG. In Sect. 4, this paper is summarized.

2 Proposed Feature

Pedestrians show a various shape, e.g., standing, running, or shaking the hand. In addition, they wear clothes of various texture and color. Moreover, in the outdoor, illumination changes frequently and a lot of image noise occurs. CoHOG has solved these problems to some extent although it needs very large feature dimension. The proposed feature (SCHOG) improves CoHOG so that it can have better performance and lower feature dimension.

2.1 CoHOG

In this section, the outline of CoHOG is explained. CoHOG uses a two-dimensional histogram whose bin is a pair of edge gradient direction between the interest pixel and the offset pixel. The feature dimension becomes large since histograms are created for every combination of the interest pixel and the offset pixel. However, it can represent object shape finely and it is robust to the change of shape and illumination.

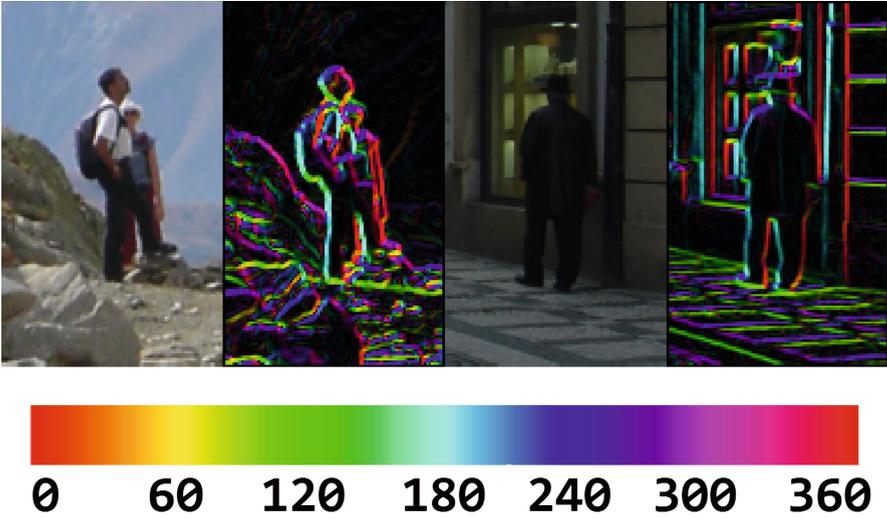


Fig. 1. Examples of gradient direction and gradient magnitude (Color figure online).

At first, the edge gradient magnitude (M) and the edge gradient direction (θ) are obtained from Eqs. (1) and (2).

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

$$M(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2}, \quad (2)$$

where $f_x(x, y)$ and $f_y(x, y)$ denote edge gradient magnitude of horizontal direction and that of vertical direction in the pixel (x, y) , which are calculated by Sobel operator. Gradient direction (θ) is quantized to eight directions by 45 degrees. Figure 1 shows the gradient direction image and the gradient magnitude image. The direction is represented by color and the magnitude is represented by the brightness. Same direction often appears around the contour of a pedestrian. CoHOG represents this characteristic by the co-occurrence of edge gradient direction between the interest pixel and the offset pixel.

31 offset pixels are set around the interest pixel as shown in Fig. 2. The interest pixel is included in offset pixels. Two dimensional histogram is created for every offset pixel. If the offset pixel corresponds the interest pixel, the histogram has eight bins because the gradient direction of each pixel is same. Except for this case, the histogram has $8 \times 8 = 64$ bins because the number of bins is a combination of gradient direction.

The input image is divided into several rectangular blocks as shown in Fig. 3. In each block, the 2D histogram is created for every offset pixels. Let (p, q) be

$C_{x,y}$ is incremented by

$$C_{x,y}(i, j) = \sum_{p=0}^{n-1} \sum_{q=0}^{m-1} \begin{cases} 1 & \text{if } I(p, q) == i \\ & \text{and } I(p+x, q+y) == j \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where I is the gradient-orientation image, n is the horizontal size of a block and m is the vertical size of a block. In each block, feature dimension is $8 + 64 \times 30 = 1928$. Figure 3 shows the example of the 2D histogram in CoHOG. In this example, an input image is divided into 2×8 blocks and in each block, the 2D histogram is created for every offset pixels. CoHOG does not perform the normalization of the histogram because CoHOG does not accumulate the gradient magnitude in the bin of the histogram, unlike HOG feature.

2.2 SCHOG

Since CoHOG uses only the relation between the gradient direction of the interest pixel and that of the offset pixel, other information acquired on the way, such as the pixel intensity or the gradient magnitude, is thrown away. SCHOG improves the performance by adding this information. SCHOG uses not only the co-occurrence of the gradient direction but also that of the similarity. In this paper, we evaluate the pixel intensity, the gradient magnitude and the gradient direction as the similarity although various features can be use as the similarity. The computing time does not increase because these features are obtained as the gradient direction is calculated.

The procedure of feature extraction is described below. At first, the gradient intensity and the gradient orientation are calculated by Eqs. (1) and (2) as well as CoHOG. Offset pixels around the interest pixel are set as the same position as CoHOG. Next, we create the 2D histogram representing the relation between the gradient direction of the interest pixel and that of the offset pixel. Unlike CoHOG, the gradient direction is quantized to four directions by 90 degrees. However, the gradient direction is quantized to eight directions by 45 degrees when the offset pixel corresponds the interest pixel because this hardly influences the number of feature dimension, as described later. The main difference between SCHOG and CoHOG is that SCHOG adds the similarity between features, such as the pixel intensity, the gradient magnitude or the gradient orientation, to the co-occurrence of the gradient direction. SCHOG can represent the shape of the object more finely than CoHOG since these features which CoHOG does not use directly are incorporated. The similarity between the interest pixel and the offset pixel is given by

$$F_{sim1}(V_o, V_i) = \begin{cases} 0 & \text{if } T_1 < \tan^{-1} \frac{V_i}{V_o} < T_2 \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

$$F_{sim2}(V_o, V_i) = \begin{cases} 0 & \text{if } T_3 < |V_o - V_i| \\ & \text{or } T_4 > |V_o - V_i| \\ 1 & \text{otherwise,} \end{cases} \quad (5)$$

where F_{sim1} is the similarity function for the pixel intensity or gradient magnitude, F_{sim2} is the similarity function for the gradient angle, V_i is the pixel intensity, the gradient magnitude or the gradient direction in the intensity pixel and V_o is the pixel intensity, the gradient magnitude or the gradient direction in the offset pixel. Thresholds T_1 , T_2 , T_3 and T_4 in Eqs. (4) and (5) were determined experimentally. The similarity returns 0 when features are similar and it returns 1 when features are different. The feature dimension is suppressed because the similarity is represented by the binary code.

Table 1. The name of SCHOG for each similarity.

Name	Similarity
SCHOG-pix	pixel intensity
SCHOG-gra	gradient magnitude
SCHOG-ang	gradient direction

We divide the input image into 6×12 blocks. Let (p, q) be the image coordinate system whose origin is at the upper left of each block, (x, y) be a offset coordinate system whose origin is at the interest pixel and $C_{x,y,s}$ be the histogram of an offset pixel (x, y) and similarity s . The bin $C_{x,y,s}(i, j, k)$ of histogram $C_{x,y,s}$ is incremented by

$$C_{x,y,s}(i, j, k) = \sum_{p=0}^{n-1} \sum_{q=0}^{m-1} \begin{cases} 1 & \text{if } I(p, q) == i \\ & \text{and } I(p+x, q+y) == j \\ & \text{and } F_{sim}(a, b) == k \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where I is gradient-orientation image, n and m represent the size of a block, a represents feature value at the offset pixel, b represents feature value at the interest pixel. $k(0$ or $1)$ represents the similarity. When the offset pixel corresponds the interest pixel, the dimension is 8. Since this case is not related to the co-occurrence, number of total feature dimension hardly increase even if the dimension is 8. The other offset pixel has 16 dimensions for a combination of 4 gradient directions and 2 dimensions for the similarity. Therefore, in each block, the total dimension of SCHOG is $8 + 16 \times 2 \times 30 = 968$. This is about a half of CoHOG. Figure 4 shows the example of the histogram representing the co-occurrence of the gradient direction and the gradient magnitude. There are two bins that represent the similarity for each combination of directions.

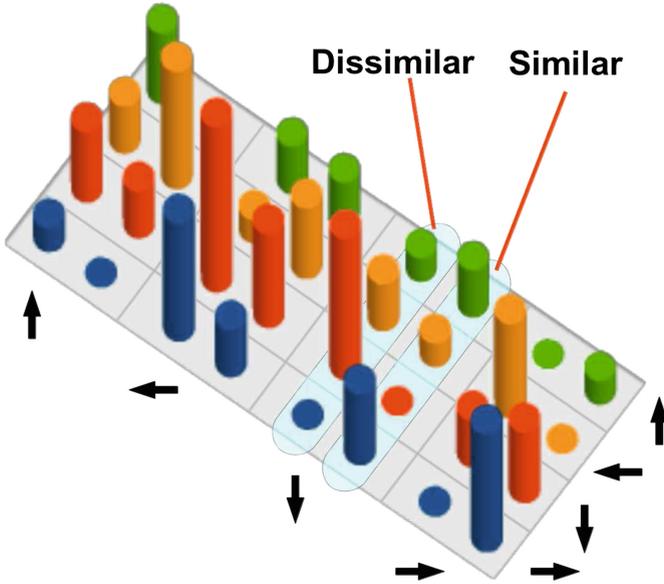


Fig. 4. Example of a histogram used in SCHOG.

In this paper, we use the pixel intensity, the gradient magnitude or the gradient direction as the feature for the similarity. However, the framework of SCHOG can easily introduce various features, avoiding the steep increase in a number of dimension because it uses the binary code to represent the similarity. The name described in Table 1 is attached for every kind of similarity. SCHOG-pix uses the pixel intensity as the similarity. SCHOG-gra uses the gradient magnitude as the similarity. Although this information is directly used in HOG, it is deleted in CoHOG. SCHOG-ang uses the gradient direction as the similarity. Since this similarity is calculated from the angle before quantization, it's expected that the finer relation between gradient directions can be expressed.

3 Experimental Results

We carried out experiments using a SVM classifier (SVMlight, Linear-kernel). The ROC curve, which shows the True Positive ratio for the vertical axis and shows the False Positive ratio for the horizontal axis, is used for evaluating the performance. It shows that performance is better, so that the curve goes to the upper left.

3.1 Dataset

We adopted INRIA Person Dataset that various previous paper have used for evaluation. Figure 5 shows some examples in this dataset. We used 2,416 positive



Fig. 5. Example of INRIA Person Dataset. (a) Person image (b) cropped negative.

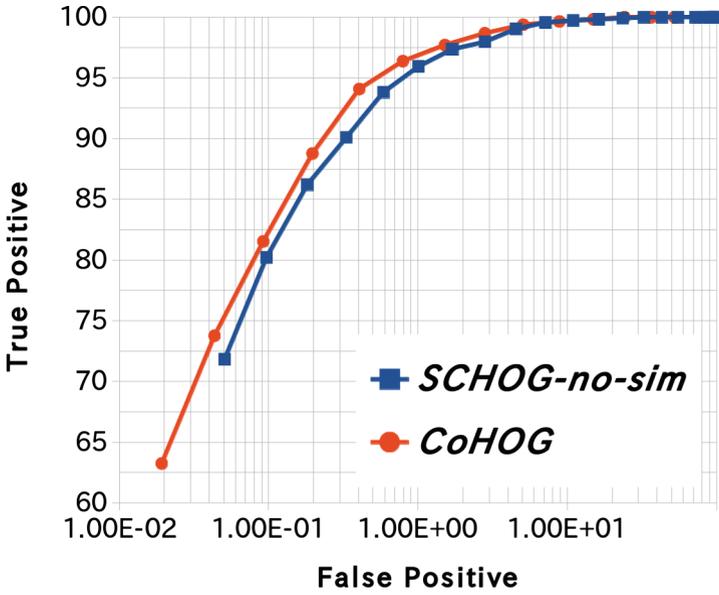
images and 12,180 negative images for training. Ten regions randomly extracted from an image were used as negative images. The size of a positive image is 64×128 pixels and the size of a negative image is from 214×320 to 648×486 . 1,132 positive images and 453 negative images are used for test. The size of a positive image is as same as an image for training and the size of a negative image is from 242×213 to 690×518 . The dataset used in experiments is summarized in Table 2.

Table 2. Details of INRIA Person Dataset.

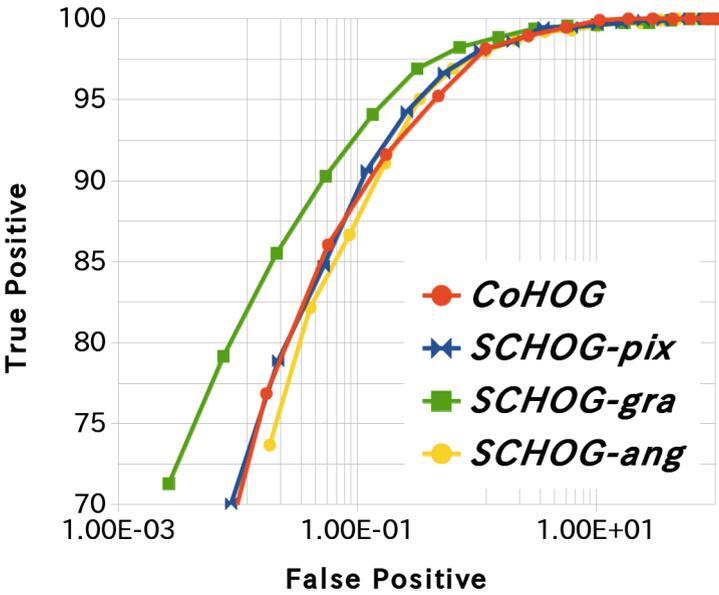
(a) Training data			(b) Test data		
image size	positive	64×128	image size	positive	64×128
	negative	$214 \times 320 - 648 \times 486$		negative	$242 \times 213 - 690 \times 518$
number	positive	2416	number	positive	1132
	negative	$1218 \times 10 = 12180$		negative	453

3.2 Effect of Similarity

Figure 6(b) shows the performance of CoHOG, SCHOG-pix, SCHOG-gra and SCHOG-ang. SCHOG-pix, SCHOG-gra and SCHOG-ang use the pixel intensity,



(a) Performance of CoHOG and SCHOG-no-sim.



(b) Comparison of the proposed method and CoHOG.

Fig. 6. Performance of the proposed method.

the gradient magnitude and the gradient direction as the similarity respectively. The dimension of these features is a half of CoHOG. In Fig. 6(b), the True Positive ratio of SCHOG-pix, SCHOG-gra, SCHOG-ang and CoHOG is 90.12 %, 93.13 %, 87.37 % and 88.07 % respectively when the False Positive ratio is 0.1 %. SCHOG-pix shows the almost same performance as CoHOG although it uses the simple feature like the pixel intensity as the similarity. SCHOG-gra, which uses the gradient magnitude as the similarity, shows quite better performance than CoHOG. This result shows that the gradient magnitude which CoHOG omitted

Table 3. Summary of proposed features.

Name	Dimension per block	Similarity	Performance	TP (FP=0.1[%])
SCHOG-pix	$8 + 16 \times 2 \times 30 = 968$	pixel intensity	fair	90.1066
SCHOG-gra	$8 + 16 \times 2 \times 30 = 968$	magnitude	good	93.1310
SCHOG-ang	$8 + 16 \times 2 \times 30 = 968$	gradient direction	bad	87.3676
CoHOG	$8 + 8 \times 8 \times 30 = 1928$		bad	88.0733



(a) Examples to which both CoHOG and SCHOG-gra failed in detection.



(b) Examples to which only CoHOG failed in detection.



(c) Examples only SCHOG failed in detection.

Fig. 7. Examples of failure detection. Left is the original image, and right is gradient direction image.

is effective to improve performance of pedestrian detection. The performance of SCHOG-ang is slightly inferior to CoHOG. This result shows that the similar feature does not contribute to improvement in performance. In this experiment, it was shown that SCHOG whose similarity is the gradient magnitude can obtain better performance than CoHOG although the resolution of the gradient direction is a half of CoHOG. The summary of features used in this experiment is shown in Table 3.

Figure 7 shows failure examples of SCHOG-gra and CoHOG. Figure 7(a) shows examples to which both SCHOG-gra and CoHOG failed in detection. Pedestrians with low contrast to a background were not detected. Figure 7(b) shows examples to which only CoHOG failed in detection. CoHOG failed because the gradient direction around a pedestrian's contour is scattering, but SCHOG succeeded using the difference in the gradient magnitude between a pedestrian and a background. Figure 7(c) shows examples to which only SCHOG failed in detection. In these examples, pedestrians were not detected because the gradient magnitude around the pedestrian's contour is similar.

4 Conclusions

In this paper, we proposed the novel feature named SCHOG which improved CoHOG feature so that the detection performance might improve and the feature dimension might decrease. SCHOG consists of the co-occurrence of edge gradient direction and the similarity. Although SCHOG quantizes edge gradient direction to the half of CoHOG, it can represent the shape of the object more finely than CoHOG by adding the similarity to the co-occurrence of edge gradient direction. Because the similarity is represented by the binary code, the dimension of SCHOG is a half of the conventional CoHOG in spite of adding the similarity. Experimental results using INRIA Person Dataset showed that reducing quantization of the gradient direction hardly causes the fall of performance and SCHOG whose similarity is gradient magnitude have quite better performance than CoHOG.

As the similarity, the pixel intensity, the gradient magnitude and the gradient direction were evaluated in this paper. However, since the similarity is simply represented by the binary code, other various features such as color information or a combination of features are allowed as the similarity. Therefore, SCHOG can be applied to various application of object recognition. Presently, our method uses the same arrangement of offset pixels as CoOG. However, this arrangement is important for improving performance and the number of dimension can be reduced greatly if the number of offset pixels is reduced. In the future, we will examine the optimal arrangement and optimal number of offset pixels. Then, we clarify the performance by experiments using different data sets.

References

1. Hattori, H., Seki, A., Nishiyama, M., Watanabe, T.: Stereo-based pedestrian detection using multiple patterns. In: BMVC British Machine Vision Association (2009)

2. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recogn.* **29**(1), 51–59 (1996)
3. Cao Yunyun, H.N.: Detecting pedestrians using an advanced local binary pattern histogram. In: 2011 Proceedings of 18th ITS World Congress, Orlando (2011)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Schmid, C., Soatto, S., Tomasi, C. (eds.) *International Conference on Computer Vision and Pattern Recognition*. vol. 2, pp. 886–893. INRIA Rhone-Alpes, ZIRST-655, av. de l’Europe, Montbonnot-38334, June 2005. <http://lear.inrialpes.fr/pubs/2005/DT05>
5. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3952, pp. 428–441. Springer, Heidelberg (2006). http://dx.doi.org/10.1007/11744047_33
6. Tomoki, W., Satoshi, I., Kentaro, Y.: Co-occurrence histograms of oriented gradients for pedestrian detection. In: Wada, T., Huang, F., Lin, S. (eds.) *PSIVT 2009*. LNCS, vol. 5414. Springer, Heidelberg (2009). doi:[10.1007/978-3-540-92957-4_4](https://doi.org/10.1007/978-3-540-92957-4_4)
7. Kobayashi, T., Otsu, N.: Image feature extraction using gradient local auto-correlations. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part I*. LNCS, vol. 5302, pp. 346–358. Springer, Heidelberg (2008)
8. Yamaguchi, K., Naito, T.: Two dimensional histograms of oriented gradients for pedestrian detection. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci. D, Inf. Syst.* **94**(1), 365–373 (2011). <http://ci.nii.ac.jp/naid/110008006578/>
9. Walk, S., Majer, N., Schindler, K., Schiele, B.: New features and insights for pedestrian detection. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, San Francisco, June 2010