

Weighted CoHOG (W-CoHOG) Feature Extraction for Human Detection

Nagaraju Andavarapu and Valli Kumari Vatsavayi

Abstract Human recognition techniques are used in many areas such as video surveillance, human action recognition, automobile industry for pedestrian detection, etc. The research on human recognition is widely going on and is open due to typical challenges in human detection. Histogram-based human detection methods are popular because of its better detection rate than other approaches. Histograms of oriented gradients (HOG) and co-occurrence of histogram-oriented gradients (CoHOG) are used widely for human recognition. A CoHOG is an extension of HOG and it takes a pair of orientations instead of one. Co-occurrence matrix is computed and histograms are calculated. In CoHOG, gradient directions alone are considered and magnitude is ignored. In this paper magnitude details are considered to improve detection rate. Magnitude is included to influence the feature vector to achieve better performance than the existing method. In this paper, weighted co-occurrence histograms of oriented gradients (W-CoHOG) is introduced by calculating weighted co-occurrence matrix to include magnitude factor for feature vector. Experiments are conducted on two benchmark datasets, INRIA and Chrysler pedestrian datasets. The experiment results support our approach and shows that our approach has better detection rate.

Keywords Histograms of oriented gradients (HOG) • Co-occurrence histogram of oriented gradients (CoHOG) • Weighted co-occurrence histogram of oriented gradients (W-CoHOG) • Human detection • Pedestrian detection

N. Andavarapu (✉) · V.K. Vatsavayi
Department of Computer Science and Systems Engineering,
Andhra University, Visakhapatnam, India
e-mail: nagraz.a@gmail.com

V.K. Vatsavayi
e-mail: vallikumari@gmail.com

1 Introduction

Computer vision is a wide and emerging area over the past few years. The analysis of images involving humans comes under computer vision problem. Human detection techniques are used in many areas such as people abnormal behavior monitoring, robots, automobile safety systems, and gait recognition. The main goal of a human detector is to check whether humans are present in the image or not. If human is identified in the particular image then it can be used for further analysis. Human detection is still an open problem. Human detection is one of the active and challenging problems in computer vision, due to different articulations and poses, different types of appearances of clothes and accessories acting as occlusions. In this paper humans are identified in a static image. Identifying humans in a static image is more difficult than in a video sequence because no motion and background information is available to provide clues to approximate human position. In our approach, input of the human detector is an image and output is a decision value finding whether there is a human in a given image or not. In this paper static images are considered to detect humans.

2 Related Works

Many human detection techniques have been proposed so far in different approaches. The implicit shape model (ISM) [1], a part-based object detection algorithm proposed by Leibe et al., uses local features derived from a visual vocabulary or codebook as object parts. Codebooks are generated using SIFT [2] or shape context local feature descriptor. Lu et al. proposed image depth based algorithm [3] to detect humans by taking depth information of given image. Jiaolong et al. proposed a part-based classifier technique [4] to detect humans in a given image window. In this method mixture of parts technique was used for part sharing among different aspects. Andriluka et al. proposed a generic approach for nonrigid object detection and articulated pose estimation based on the pictorial structures framework [5]. Gavrilu et al. introduced a template matching approach for pedestrian detection [6]. Template hierarchy of pedestrian silhouettes is built to capture the variety of pedestrian shapes. For identifying shapes, canny edge detector [7] is used.

Gradient orientation based feature descriptors such as SIFT [2], HOG [8], CoHOG [9], etc., are recent trends in human detection. SIFT [2] (scale-invariant feature transform) features proposed by Lowe et al. are used in human body parts detection in [10]. Histogram-based features are popularly used in human recognition and object detection because of their robustness. Histograms of oriented gradients (HOG) [8] is a famous and effective method for human detection. It uses histograms of oriented gradients as a feature descriptor. HOG features are robust towards illumination variances and deformations in objects. Co-occurrence histograms of oriented gradients (CoHOG) [9] is an extensive work of HOG which has

more detection rate and lesser miss rate. In recent days, CoHOG used in many computer vision applications such as object recognition [11], image classification [12], and character recognition [13]. In CoHOG, co-occurrence matrices calculated for oriented gradients for making feature descriptor strong. In CoHOG only gradient direction details are considered and gradient magnitude details are ignored. In the proposed method, gradient magnitude components are also considered to bring more accuracy to the existing CoHOG.

The rest of the paper is organized as follows: Sect. 2 gives a brief overview of HOG and CoHOG. Proposed method W-CoHOG is discussed in Sect. 3 in detail. Section 4 contains experimental results and comparison with existing methods. Finally the work concluded in Sect. 5.

3 Background: HOG and CoHOG

3.1 HOG

In HOG, initially gradients are computed on each pixel in a given image and are divided into nine orientations. Next the image is divided into small nonoverlapping regions. Typical regions are of size 8×8 or 16×16 . Then HOGs are calculated for each and every small region. Finally histograms of each region are concatenated using vectorization.

3.2 CoHOG

Co-occurrence histograms of oriented gradients (CoHOG) is an extension to HOG and more robust than HOG. In CoHOG pair of oriented gradients is used instead of single gradient orientation. Co-occurrence matrix is calculated for pair of gradient orientation with different offsets.

$$C_{\Delta x, \Delta y}(p, q) = \sum_{i=1}^n \sum_{j=1}^m \begin{cases} 1 & \text{if } I(i, j) = p \text{ and } I(i+x, j+y) = q \\ \text{None} & \text{Otherwise} \end{cases} \quad (1)$$

Equation (1) shows the calculation of co-occurrence matrix. Figure 1a shows typical co-occurrence matrix histograms of oriented gradients and Fig. 1b shows possible offsets for CoHOG.

In CoHOG, orientation values of gradient are alone considered and magnitude is ignored. In the proposed method magnitude is also considered, as magnitude also contains discriminative information for human detection. Let us consider the following example: Fig. 2a is quite different from Fig. 2b because of different magnitude values even though it has same gradient orientation. Hence magnitude also

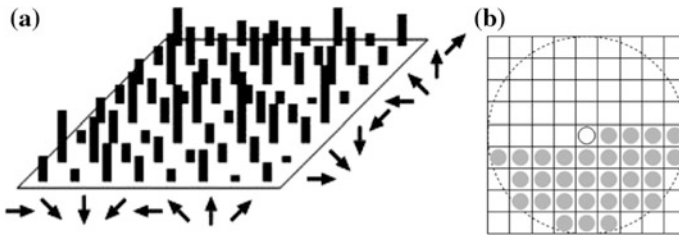


Fig. 1 a Typical co-occurrence matrix histogram. b Possible offsets

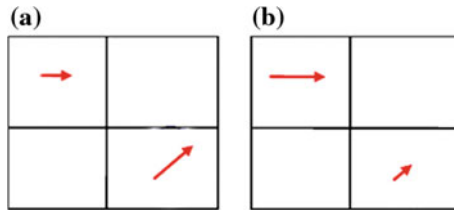


Fig. 2 Magnitudes of two gradients having same orientation. a and b are in same orientations but they are not the same

describes about what image contains. Existing feature descriptors does not consider magnitude details.

4 Proposed Method (W-CoHOG)

4.1 Overview

In the CoHOG method gradient directions are alone considered and magnitude is ignored. In the proposed method magnitude is also considered to extract more robust feature. Magnitude weighted co-occurrence histograms of oriented gradients (W-CoHOG) is proposed for better feature descriptor. Figure 3 briefly explains the classification process for human detection using W-CoHOG extraction method.

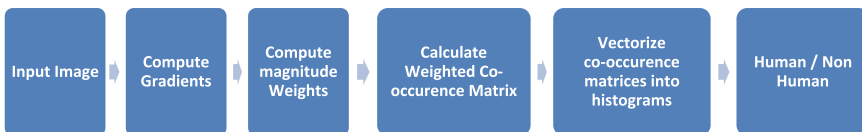


Fig. 3 Our proposed classification process

Initially, gradients of image are computed in magnitude and direction form and converted into oriented gradients. Next, image is divided into 3×6 or 6×12 sized non-overlapping cells. Then, weighted co-occurrence matrices are computed for each region. After that, all co-occurrence matrices of all regions are combined.

4.2 Feature Extraction

For a given input image, gradients are computed for each pixel. In this method, Sobel and Robert’s filters are used to compute gradients of a given input image. Equations (2) and (3) show gradient calculation using Sobel and Robert’s filters, respectively, for a given input image I , as shown in below.

Sobel gradient operator

$$(a) G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \quad (b) G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I \quad (2)$$

Robert’s gradient operator

$$(a) G_x = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} * I \quad (b) G_y = \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix} * I \quad (3)$$

Then, gradients are converted into magnitude and direction using Eq. (4). The gradients directions are converted into eight equal bins with 45^0 intervals.

$$(a) \theta = \tan^{-1} \frac{g_x}{g_y} \quad (b) m = \sqrt{g_x^2 + g_y^2} \quad (4)$$

After that, magnitude matrix is convoluted with mean mask to eliminate noise which may cause aliasing effect. Equation (5) shows the 7×7 mean mask used in the proposed method. Figure 4 shows the overview of W-CoHOG feature calculation process.

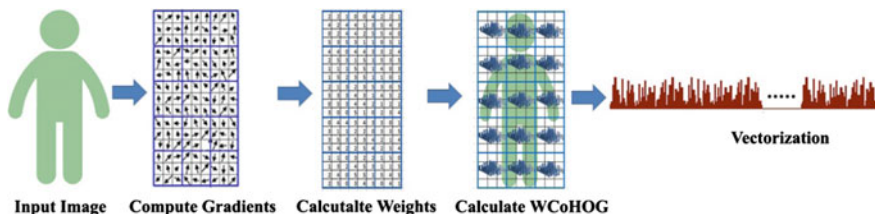


Fig. 4 Overview of W-CoHOG calculation

$$\text{Conv}_{7 \times 7} = \frac{1}{49} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{5}$$

Weight Function

In this proposed method magnitude component of a gradient is used as weight function to calculate weighted co-occurrence matrix. In order to calculate magnitude weighted co-occurrence matrix, the magnitude weights of each pixel are calculated. Weight function is applied to co-occurrence matrix to influence the co-occurrence matrix using gradient magnitude of each pixel. The weight functions used in this method are described in following paragraph.

Let I be a given input image. i, j are the any two orientations in the given eight orientations and $\Delta x, \Delta y$ are offset for co-occurrence. $C_{\Delta x, \Delta y}(i, j)$ is weighted co-occurrence matrix for a given offset $\Delta x, \Delta y$ and orientation i, j . The Eqs. (6 and 7) describes the calculation of the weighted co-occurrence matrix.

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \{ W_{(p,q),(p+\Delta x,p+\Delta y)} * \alpha \} \tag{6}$$

where

$$\alpha = \begin{cases} 1 & \text{if } O(p, q) = i \text{ and } O(p + \Delta x, q + \Delta y) = j \\ 0 & \text{Otherwise} \end{cases} \tag{7}$$

Let $m_{p,q}$ be a gradient at a given pixel p, q for a given input image I . \bar{M} and M_{\max} are mean and maximum gradient values in I . The weight calculation was performed with simple operations like mean and division operations. Equations (8 and 9) show two possible weight functions to calculate weight for a given pixel (p, q) and $(p + \Delta x, q + \Delta y)$. Any of the two functions is preferable to calculate weights for calculating weighted co-occurrence matrix. In this proposed method, Eq. (8) is used to calculate weights for experimental results.

$$W_{(p,q),(p+\Delta x,p+\Delta y)} = \left(\frac{m_{p,q}}{\bar{M}} * \frac{m_{p+\Delta x,p+\Delta y}}{\bar{M}} \right) + \mu \tag{8}$$

$$W_{(p,q),(p+\Delta x,p+\Delta y)} = \left(\frac{m_{p,q}}{M_{\max}} * \frac{m_{p+\Delta x,p+\Delta y}}{M_{\max}} \right) + \mu \tag{9}$$

where, μ is constant and $\mu = 1$.

After computing magnitude weighted co-occurrence matrices for all regions, the matrices are vectorized by simple concatenation of all matrix rows into a single row. There are 31 offsets possible for calculating co-occurrence matrix shown in Fig. 1b. Co-occurrence matrices need not be calculated for all offsets. In calculation of W-CoHOG, two offsets are good enough for pedestrian detection problem.

The size of feature vector is very large in histogram-based feature descriptors. For these types of features linear SVM [04] classifier is suitable. In this proposed method LIBLINEAR classifier [14] is used. LIBLINEAR classifier is an SVM-based classifier which works faster than SVM classifier [15] even for million instances of data. HOG and CoHOG also used SVM classifier for classification.

Algorithm 1: Weighted-CoHOG Calculation

```

1. Given Image: I
2. Given offset: x,y
3. Begin
4. Calculate O,M of I
5. Divide image into cells
6. For each cell in the image
7.   Initialize co-occurrence matrix C←0
8.   For each pair of orientation (i,j)
9.     For each pixel p,q in the cell
10.      Calculate  $W_{(p,q), (p+x, q+y)}$ 
11.      If  $O_{p,q} = i$  and  $O_{p+x, q+y} = j$  then
12.         $C_{p,q} = C_{p,q} + 1 * W_{(p,q), (p+x, q+y)}$ 
13.      end
14.      else
15.         $C_{p,q} = C_{p,q} + 0 * W_{(p,q), (p+x, q+y)}$ 
16.      end
17.    end
18.  end
19. end
20. H ← vectorize all co-occurrence matrices
21. return H
22. end

```

5 Experimental Results

Experiments are conducted on two datasets Daimler Chrysler [11] and INRIA dataset [8]. These are the familiar benchmark datasets for human detection. In Daimler Chrysler dataset, 4,800 human images and 5,000 nonhuman images are taken for training; and another 4,800 human images for training and 5,000 images



Fig. 5 INRIA sample images in dataset



Fig. 6 Chrysler sample images in dataset

are taken for testing. Each image size in Daimler Chrysler dataset is 48×96 pixels. In INRIA dataset, 1208 positive images and 12,180 patches are randomly sampled from person-free image for training and testing.

Figures 5 and 6 show that sample positive and negative examples of INRIA dataset and Chrysler dataset, respectively. Negative images are generated by taking 64×128 patches from no person images in INRIA dataset. Simple Sobel filter and Roberts filter were used to calculate the gradients of input image.

ROC curves are used for performance evaluation of binary classification like object detection problems. Sliding window technique is used to detect the humans in the image. A typical scanning window size for INRIA dataset is equal to the same as positive image size 64×128 . In this paper, true positive rate versus false positive per window (FPPW) was plotted to evaluate the performance of proposed method and to compare with state-of-the-art methods. An ROC curves towards the top-left of the graph means better performance for classification problem. Figures 7 and 8 clearly show that curves obtained by proposed method achieved better

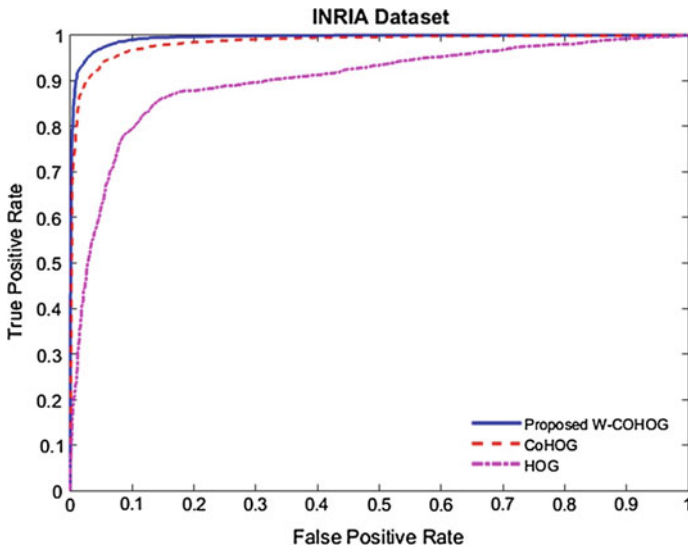


Fig. 7 Comparison of proposed method with other state-of-the-art methods (INRIA Dataset)

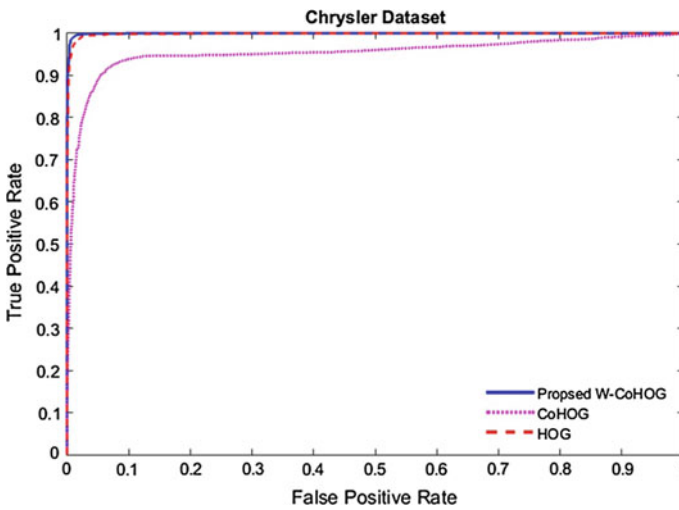


Fig. 8 Comparison of proposed method with other state-of-the-art methods (Chrysler Dataset)

detection rate for all false positive rates than other existing methods or at least comparable. The results clearly show that our method reduced miss rate around 20 % compared with CoHOG. The accuracy of the classifier is also better than other state-of-the-art methods shown in the figure. In the proposed method only two offsets are used instead of all 31 possible offsets, even though good results are acquired by adding gradient magnitude component.

6 Conclusion

In this paper a new method called weighted CoHOG is proposed which is an extension work to CoHOG. Magnitude component is also added to feature vector to improve the classification. The proposed method achieved improvement in accuracy on two benchmark datasets. Experimental results prove that performance of the proposed method is better than the other state-of-the-art methods. Even though calculation of weights adds additional computational complexity, the overall feature vector generation time decreased by reducing the number of offsets to two. Future work involves proposed feature descriptor to be used in other applications such as person tracking.

References

1. Leibe, B., Leonardis, A., Schiele, B.: Combined object categorization and segmentation with an implicit shape model. In: Workshop on statistical learning in computer vision, ECCV, vol. 2. no. 5 (2004)
2. Lowe, David G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60**(2), 91–110 (2004)
3. Lu, X., Chen, C.-C., Aggarwal, J.K.: Human detection using depth information by Kinect. In: 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE (2011)
4. Xu, J., et al.: Learning a part-based pedestrian detector in a virtual world. *IEEE Trans. Intell. Transp. Syst.* **15**(5), 2121–2131 (2014)
5. Andriluka, M., Roth, S., Schiele, B.: Pictorial structures revisited: people detection and articulated pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE (2009)
6. Gavrila, D., Philomin, V.: Real-time object detection for “smart” vehicles. In: The Seventh IEEE International Conference on Computer Vision, vol. 1, pp. 87–93. IEEE Computer Society Press, Los Alamitos (1999)
7. Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 679–698 (1986)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1. IEEE (2005)
9. Watanabe, T., Ito, S., Yokoi, K.: Co-occurrence histograms of oriented gradients for pedestrian detection. In: *Advances in Image and Video Technology*, pp. 37–47. Springer, Berlin (2009)
10. Shashua, A., Gdalyahu, Y., Hayun, G.: Pedestrian detection for driving assistance systems: single-frame classification and system level performance. In: *Intelligent Vehicles Symposium*, IEEE (2004)
11. Iwata, S., Enokida, S.: Object detection based on multiresolution CoHOG. *Advances in visual computing*. In: Springer International Publishing, pp. 427–437 (2014)
12. Kawahara, T., et al.: Automatic ship recognition robust against aspect angle changes and occlusions. In: *Radar Conference (RADAR)*, IEEE (2012)
13. Su, B., et al.: Character Recognition in natural scenes using convolutional co-occurrence HOG. In: 22nd International Conference on Pattern Recognition (ICPR), IEEE (2014)

14. Fan, R.-E., et al.: Liblinear: A library for large linear classification. *J. Mach. Learn. Res.* **9**, 1871–1874
15. Hearst, M.A., et al.: Support vector machines. In: *Intelligent Systems and their Applications*, IEEE **13**(4), 18–28 (1998)