International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 9, Issue 6, November - December 2020 ISSN 2278-6856

A Histogram-Refinement Histogram-of-Oriented-GradientObject and Pedestrian Detector

Daniel Asher Mitchell[1] , **Harvey Barnett Mitchell** [2]

[1]Rehov Brosh, MazkeretBatya, Israel

[2]Rehov Brosh, MazkeretBatya, Israel

Abstract:*The Histogram-of-Oriented-Gradient (HOG) is widelyused forobject and pedestrian detection. The basic idea of HOG is that theappearance of an object can be characterized by the spatial distribution of the intensity of the local gradients and their directions. In this article we describe a new HOG detectorin which we use the method of histogram refinement to split the traditional HOG vector into two complementary vectors which we subsequently fuse together. The new HOG has an enhanced performance but at the same time is fully compatible with the traditional HOG. Experimental results on the INRIA pedestrian detection database shows that for noisy input images the histogram-refined HOG significantly outperforms the traditional HOG detector.*

Keywords: Histogram-of-oriented-gradient, Object detection, Pedestrian detection, histogram refinement, HOG, image processing

1. INTRODUCTION

Objectdetection, and in particular pedestrian detection, is an important branch of patternrecognition which is required in many applications ranging from videosurveillance [1-6] andbiometrics [7,8] to re-identification [9] and robotics [10]. Detecting objects in a static imageis oftenchallenging especially when the object has a wide variability in appearance. Part of this variability is a result of projecting a 3D object onto a 2D image plane. However, there is often someintrinsic variability. This is true for pedestrian detection in which the intrinsic variability is often very large. In addition,there are changes in theillumination and in the background. In spite of these difficulties, objectand pedestrian detection hasattracted a sustained researcheffort and,in recent years, great progresshas been made in pedestrian detection [11-14].

A basic step inboth objectand pedestrian detection is the extraction of characteristic features from the input image. The many different features which have been employed for thispurpose can be broadly divided into two groups: hand-crafted features [15, 16, 17] and deep convolutionalneural network (CNN) features [18, 19, 20, 21].

In general, the CNN features have the best performance. However, the CNN has several drawbacks: they require a very large amount of training data and have a long training time. In addition, the CNN's are usually complex with a very high computational load. On the other hand, the traditional methods require much less training data, are much simpler to train and have a much lower computational load. They often employ a sliding window paradigm with hand-crafted features and a traditional classifier. Among the hand-crafted features, the histogram of oriented gradient (HOG) [22] descriptor is the most well-known and may be used with any convenient classifier, e. g. the K-nearest neighbor or linear Support Vector Machine (SVM) [23]. Since its introduction by [22], HOG has been intensively researched [24, 25] and widely used for real-time, or near real-time, object detection applicationswith limited computational resources [24, 26- 28].

Recently [29] described a new HOG detector in which twocomplementary feature vectors are fused together [30]. The first feature vector is the traditional HOG vector as calculated on the input image. The second feature vector is a HOG vector calculated on the input image after it has been segmented into superpixels. In this article, we use instead the method of histogram-refinement [31,32] to generate two complementary feature vectors. The advantages of using histogram-refinement are that it is both simpler and computationally faster than the superpixel segmentation used in [29]. By fusing the together the two complementary vectors we ensure the new HOG has the samesize as the traditional HOG and may therefore be used without modification, wherever the traditional HOG detector is used.

By using two complementary feature vectors, the new HOG detector has an improved performance combined with robustness against additive noise. This is verified in a series of pedestrian detection experiments.

2.HISTOGRAM OF ORIENTEDGRADIENTS (HOG)

The basic idea of a HOG feature is that the local object appearance and shape can be characterized by the distribution of local intensity gradients or edge directions.Suppose $I(x, y)$ denotes the intensity of a pixel (x, y) in the image I . To establish our notationwe briefly review the principal steps in extracting the HOG descriptor [22, 26]:

1. **Gradient Calculation**. At each pixel (x, y) we

calculate the intensity gradient $G(x,y)$ and the edge direction $\theta(x,y)$ as follows:

$$G(x,y)=\sqrt{G_x^2(x,y)+G_y^2(x,y)}$$
$$\theta(x,y)=\tan^{-1}(G_y(x,y)/G_x(x,y))$$

where

$$G_x(x,y)=I(x+1,y)-I(x-1,y),$$
$$G_y(x,y)=I(x,y+1)-I(x,y-1),$$

represent, respectively, the horizontal gradient and vertical gradient at the pixel (x,y).

2. **Histogram**. The sliding window is divided into large partially overlapping spatial regions, or "blocks". Each block is then divided into 2×2 small square regions, or "cells". For each cell $C_m, m\in\{1,2,3,4\}$, we quantize $\theta(x,y)$ using 9 equal-width bins $\Delta\theta_h, h\in\{1,2,\ldots,9\}$. Thenthe histogram of the m th cell is computed as follows:

$$H_m(\theta_h)=\sum_{(x,y)\in C_m} v_h(x,y)$$

where

$$v_h(x,y)=\begin{cases}G(x,y) & \text{if } \theta(x,y)\in\Delta\Theta_h,\\ 0 & \text{otherwise}.\end{cases}$$

In each block, the four histograms $H_m, m\in\{1,2,3,4\}$,are concatenated to produce a $36-D$ feature vector. We denote the feature vector for the i th block as $[f(i,1)f(i,2)\ldots f(i,36)]$.

3. **L2-Hys Normalization**. In each block we normalize the feature vector using the L2-Hys norm. This isdefined as L2-normalization followed by clipping maximum values to 0.2 and then re-normalization:

$$f_n(i,j)=g_n(i,j)/\sum_{j=1}^{36} g_n^2(i,j)),$$

where

$$g(i,j)=f(i,j)/\sum_{j=1}^{36} f^2(i,j),$$

and

$$g_n(I,j)=min(g(i,j),0.2).$$

4. **Concatenation**. The normalized $36-D$ feature vectors of all blocks in the sliding window are concatenated. This is the *normalized* HOG vector F_n :

$$F_n=[f_n(1,1)\ldots f_n(1,36)\,f_n(2,1)\ldots].$$

3.HISTOGRAM REFINEMENT

Histogram refinement is a method for enhancing histogram features. These are often used in image classification applications andwas first used in image retrieval systems by Pass and Zabih [31]. In an image retrieval system global histograms are used as features for retrieval purposes. In histogram refinement we first classify the pixels in an image using their local characteristic. We subsequently generate multiple histogram features which correspond to each category of pixels. In [32] each pixel is classified into one of two categories based on the skewness of the local distribution of pixel values. Mathematically, each pixel (x,y) is given an index, or *local skew property*, $LSP(x,y)$:

$$LSP(x,y)=\begin{cases}1 & \text{if } m_S>\mu_S,\\ 0 & \text{otherwise},\end{cases}$$

where m_S and μ_S are, respectively, the median and mean of the gray-levels of the pixels contained in the set $S(x,y)$.

Until now histogram-refinement has not been used in HOG detectors. One possible reason for this is that the multiple histograms generated in histogram-refinement will substantially increase the size of the HOG vector. In the proposed histogram-refined HOG we avoid this problem by fusingtogether the multiple histograms.

4.HISTOGRAM-REFINED HOG

In the new histogram-refined HOG detector we calculate two *non-normalized* HOG vectors as follows. Initially, we calculate the classify each pixel using the local skew property $LSP(x,y)$. Then instead of calculating a single histogram $H_m(\theta_h)$ for each image cell C_m we split the histogram $H_m(\theta_h)$ into two histograms $H'_m(\theta_h)$ and $H''_m(\theta_h))$, where

$$H'_m(\theta_h)=\sum_{(x,y)\in C_m} v_h(x,y)\times LSP(x,y)$$
$$H''_m(\theta_h)=\sum_{(x,y)\in C_m} v_h(x,y)\times(1-LSP(x,y))$$

We then use $H'_m(\theta_h)$ and $H''_m(\theta_h)$ to calculate two *non-normalized* HOG vectors:

$$F'=[f'(1,1)f'(1,2)\ldots f'(1,36)\,f'(2,1)\ldots],$$

$F'' = [\, f'(1,1)\, f'(1,2) \ldots f'(1,36)\, f'(2,1) \ldots]$.

We fuse F' and F'' together by multiplying them together element-by-element:

$\hat{f}(i,j) = f'(i,j) \times f''(i,j)$.

The final HOG vector is obtained by L2-Hys normalizing $\hat{f}(i,j)$. We denote the normalized dual HOG vector as:

$\hat{F}_n = \left[\, \hat{f}_n(1,1)\ \hat{f}_n(1,2) \ldots \hat{f}_n(1,36)\ \hat{f}_n(2,36) \ldots \right]$.

5.EXPERIMENTS

We tested the histogram-refined HOG detector on the standard INRIA pedestrian database. This contains gray-scale images (size 64×128) of humans cropped from a varied set of personal photographs. We randomly selected 1218 of the images as positive training examples, together with their left-right reflections (2436 images in all). A fixed set of $12180 = 1218 \times 10$ patches sampled randomly from 1218 person-free training photos provided the negative set. A separate set of 352 positive images and 3520 negative images were selected from the INRIA database to be used as test samples. A simple linear SVM classifier was then used to classify the test images.

To evaluate the robustness with respect to distortions and noise, we considered additive Gaussian noise (standard deviation σ) and salt-and-pepper noise (noise density ρ) andGaussian image blurring (standard deviation σ).We use the original noise-free INRIA images for training while testing on the noisy images. The results shown are the average of 10 cross-validated independent runs.

Operating parameters for the HOG algorithms are given in Table 1. We give the experimental results obtained with

Table 1. HOG operating parameters

Parameter	Value
Cell size	8 by 8 pixels
Block Size	2 by 2 cells
Block Overlap	50%

the traditional HOG and the new histogram-refinedHOG using the F value. This is defined as

$$F = 2\frac{P \times R}{P + R}$$

where P and R are, respectively, the precision and the recall which are defined as

$$P = \frac{TP}{TP + FP},$$

$$R = \frac{TP}{TP + FN},$$

where TN(true negative) and TP(true positive) are, respectively, the number of non-pedestrians and pedestrians correctly identified as non-pedestrians and pedestrians; FP(false positive) is the number of non-pedestrians incorrectly identified as pedestrians and FN(false negative) is the number of pedestrians incorrectly identified as non-pedestrians.

The precision and recall are not independent: We may increase the recall by reducing the precision and vice versa. For this reason it is common practice to combine P and R in a single F value.

In Tables 2-3we give the F values as measured on the test data with additive Gaussian and salt-and-pepper noise. For completeness we also give the P and R values.

We see that for additiveGaussian and salt-and-pepper noise, the performance of the histogram-refined HOG (as measured by the average F number) always exceeds that of the traditional HOG. For Gaussian noise, increases in both P and R contribute to the increase in F. In the case of salt-and-pepper noise, the increase in F is mainly due to an increase in R.

Table 2. Traditional vs Histogram-Refined HOG: Additive Gaussian Noise

σ	Traditional HOG			Histogram-refined HOG		
	R	*P*	*F*	*R*	*P*	*F*
0	0.910	0.919	0.910	0.916	0.913	0.915
1	0.911	0.898	0.905	0.912	0.911	0.912
2	0.893	0.909	0.901	0.898	0.922	0.910
3	0.861	0.913	0.886	0.800	0.920	0.900
5	0.788	0.915	0.846	0.828	0.927	0.875
10	0.594	0.943	0.729	0.642	0.937	0.762

Table 3. Traditional vs Histogram-Refined HOG: Salt-and-Pepper

ρ	Traditional HOG			Histogram-refined HOG		
	R	*P*	*F*	*R*	*P*	*F*
1	0.674	0.972	0.796	0.809	0.970	0.877
2	0.446	0.984	0.614	0.667	0.981	0.790
3	0.257	0.985	0.408	0.520	0.975	0.680
5	0.138	0.993	0.243	0.384	0.960	0.551
10	0.063	1.000	0.118	0.255	0.973	0.404

6.CONCLUSION

We have described an enhanced histogram-refined HOG object detector. The new HOG detector has the same input and output as the traditional HOG and may thus be used as a plug-in replacement for the traditional HOG. Detection results obtained on the standard INRIA pedestrian database show the new HOG detector is very successful in the case of additive Gaussian and salt-and pepper noise: the average recall and average precision always exceed that of the traditional HOG. For Gaussian blur, average recall of the new HOG exceeds that of the traditional HOG, while the average precision of the dual HOG is less than that of the traditional HOG.

References

[1] V. Gajjar, Y. Khandhediya, and A. Gurnani, "Human detection and tracking for video surveillance: a cognitive science approach", IEEE Int. Conf. Comp. Vis. Workshops (ICCVW), pp. 2805-2809, 2017

[2] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: a survey," Comp. Sci. Rev, Elsevier, 2018

[3] M. Elhoseiny, A. Bakry and A. Elgammal, "Multiclass object classification in video surveillance systems: experimental study,", Proc Comp. Vis. Patt. Recog., 2013

[4] K. Mizuno, Y. Terachi, K. Takagi, S. Izumi, H. Kawaguchi and M. Yoshimoto, "Architectural Study of HOG Feature Extraction Processor for Real-Time Object Detection," IEEE Workshop Sig. Proc. Sys. Quebec City, pp. 197-202, 2012.

[5] J.-L. Chua, Y. C. Chang, and W. K. Lim, "A simple vision-based fall detection technique for indoor video surveillance," Sig. Imag. Video Proc., 9, pp. 623-633, 2015.

[6] M. Bilal, A. Khan, M. U. K. Khan, and C.-M. Chong-Min Kyung, "A low complexity pedestrian detection framework for smart video surveillance systems," IEEE Trans. Circ. Syst. Video Tech., 27, pp. 2260-2273, 2017.

[7] J. Neves, S Narducci, F. Narducci, S. Barra, and H. Proenca, "Biometric recognition in surveillance scenarios: a survey," Art. Intell. Rev., 46, pp. 515-541, 2016.

[8] I. Bouchrika, "A survey of using biometrics for smart visual surveillance: Gait recognition," In Surveillance in Action. Advanced Science and Technologies for Security Applications, P. Karampelas and T. Bourlai (eds), Springer, 2018.

[9] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang and Q. Tian, "Person re-identification in the wild," In Proc. IEEE Conf. Comp. Vis.Patt.Recogn. (CVPR), pp. 1367-1376, 2017.

[10] L. Dong, X. Yu, L. Li, and K. K. E. Hoe, "HOG based multi-stage object detection and pose recognition for service robot", Int. Conf. Contr. Automat. Robot Vis., pp. 2495-2500, 2010.

[11] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: an evaluation of the state of the art," IEEE Trans.Patt. Anal. Mach.Intell., 34, pp. 743-761, 2012.

[12] R. Trichet, and F. Bremond, "Dataset optimization for real-time pedestrian detection," IEEE Access, 6, pp. 7719-7727, 2018.

[13] R. Benenson, M. Omran, J. Hosang, and B. Schiele, "Ten years of pedestrian detection, what have we learned," In EECV, CVRSUAD Workshop, 2014.

[14] M. Bilal and M. S. Hanif, "Benchmark revision for hog-svm pedestrian detector through reinvigorated training and evaluation methodologies," IEEE Trans.Intell. Transport, 21, pp. 1277-1287, 2020.

[15] C. Q. Lai and S. S. Teoh, "A survey of traditional and deep learning-based histogram of oriented gradient feature," In IEEE Res. and Develop. (SCOReD), pp. 1-6, 2014.

[16] T. Georgiou, Y. Liu, W. Chen, and M. Lew, "A survey of traditional and deep learning-based features for high dimensional data in computer vision," Int. Journ. Multimedia Inform. Retrieval, 9, pp. 135-170, 2020.

[17] B. Fan, Z. Wang, and F. Wu, Local Image Descriptor: Modern Approaches, Springer, 2015.

[18] Y.Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: a review,"Neurocomp., 187, pp. 27-48, 2016.

[19] A. Angelova, A. Krizhevsky, V. Vanhoucke, A. S. Ogale, and D. Ferguson, "Real-time pedestrian detection with deep network cascades," In Proc. BMVC, pp. 1-12, 2015.

[20] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast r-cnn for pedestrian detection," IEEE Trans. Multimedia, 20, pp. 985-996, 2018.

[21] W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian detection based on yolo network model," In IEEE Int. Conf.Mechantron. and Automat. (ICMA), pp. 1547-1551, 2018.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In Int. Conf. Comp. Vis.Patt.Recogn. (CVPR '05), pp. 886-893, 2005.

[23] H. Bristow and S. Lucey, "Why do linear svm's trained on hog features perform so well?," ArXiv, 1406.2419, 2014.

[24] V.-D. Hoang, M.-H. Le, K.-H. Jo, "Hybrid cascade boosting machine using variant scale blocks based hog features for pedestrian detection," Neurocomp., 135, pp. 357-366, 2014.

[25] R. Trichet and F. Bremond, "How to train your dragon: Best practices in pedestrian classifier training," IEEE Access, 8, pp. 3527-3538, 2020.

[26] K. Piniarski, P. Pawlowski, and A. Dabrowski, "Tuning of classifiers to speed-up detection of pedestrians in infrared images," Sensors, 18, pp. 4363, 2020.

[27] J. H. Luo and C. H. Lin, "Pure fpga implementation of an hog based real-time pedestrian detection system," Sensors, 18, pp. 1174, 2018.

[28] A. Helali, H. Ameur, J. M. Gorritz, J. Ramirez, and H. Maaref, "Hardware implementations of real-time pedestrian detection systems," Neural Comp. App., 32, pp. 12859-12871, 2020.

[29] H. B. Mitchell and D. A. Mitchell, "Dual Superpixel HOG pedestrian detector," Int. J. Emerging Trends Tech Comp. Sci., 9, No. 5, 2020

[30] H. B. Mitchell, Data Fusion: Concepts and Ideas,Springer, 2012.

[31] G. Pass, and R. Zabih, "Histogram refinement for content based image retrieval,", Proc IEEE Workshop App Comp Vis, pp. 96-102, 1996

[32] A. K. Tiwari, V. Kanhanangad, R. B. Pachori, "Histogram refinement for texture descriptor based image retrieval," Sig Proc: Imag Comm, 53, pp. 73-85, 2017

AUTHORS

D. A. Mitchell was born and educated in Israel. He is currently an undergraduate studying in the department of Electrical Engineering at Ben-Gurion University of the Negev. His research interests are signal processing, computer vision, deep learning, virtual reality and telepresence.

H. B. Mitchell was born and educated in UK receiving his BSc, MSc and PhD in Physics in 1972, 1974 and 1977. Since 1979 he has worked in Israel on fuzzy logic, computer vision and data fusion. He has published widely in these fields including three graduate textbooks: Multi-sensor data fusion (Springer 2007), Image Fusion (Springer 2010) and Data Fusion (Springer 2012).