Journal of Computer Science

Original Research Paper

Pedestrian Recognition Based on Multi-Scale Weighted HOG

Monther Hussein Al-Bsool

Department of Information Technology,
AL-BALQA Applied University, AL-HUSON University College, Jordan

Article history
Received: 18-08-2018
Revised: 25-09-2018
Accepted: 05-11-2018

Email: monther.bsool@bau.edu.jo

Abstract: Pedestrian recognition receives a great attention in recent years due to its importance in traffic accidents identification. Traffic surveillance systems can provide valuable information for pedestrian recognition using computer vision and image processing techniques. Most techniques exploit simple feature extraction and multi-stage feature matching to train classifiers. In this study, Canny edge information and Histogram of Oriented Gradient (HOG) has been integrated into multi-scale coarse-to-fine feature extraction. Edge distribution provides a variable weight to highlight distinctive gradients in a Multi-Scale Weighted HOG (MS-WHOG) to identify pedestrian. As a result, the pedestrian distinctive features are highlighted and expanded along the edges to improve the recognition process. The proposed technique is also scale and orientation invariant, due to the use of multi-scale and edge information.

Keywords: Traffic Surveillance Systems, Pedestrian Recognition, MS-WHOG, Canny Edge Detection

Introduction

Surveillance cameras play a major role in intelligent transportation systems, security and other smart applications. Object recognition in surveillance videos has been an active research area in recent years (Su *et al*., 2015; Al-Hazaimeh *et al*., 2017; Al-Smadi *et al*., 2016a). Pedestrian is an important target in traffic surveillance, thus identifying fine details of a pedestrian will improve pedestrian recognition and behavior understanding.

There exists extensive work on pedestrian recognition, most of which assume that the pedestrian object has been successfully detected and consequently cropped from the background scene. Pedestrians are people moving or standing in an upright manner, thus their configurations are very limited. In (Dalal and Triggs, 2005; Tuzel *et al*., 2008), template-based approaches were used with a sliding window classifier, which provide a favorable result. Moreover, the presence of background and other rigid objects such as vehicles can be utilized in the detection process to improve the recognition performance (Dikmen *et al*., 2010; Al-Hazaimeh *et al*., 2018).

The recognition problem has been tackled in several researches by matching pedestrians with their appearance only. Color histograms matching was used in (Park *et al*., 2006; Gharaibeh *et al*., 2018), they extracted the color histograms from three horizontal partitions of the human image. Color appearance was modeled over the principle axis of the human body in (Luo *et al*., 2014). Nevertheless, crowded scenes and cluttered background affects finding the principal axis. Deformable part-based detectors (Felzenszwalb *et al*., 2010; Al-Nawashi *et al*., 2017) represents pedestrian as a set of deformable parts. Many researchers explore the variation in part modeling and deformation.

Gray and Tao (2008) transform the pedestrian recognition from simple matching problem into higher level classification problem. Thus, if a pair of images is matched they are assigned a positive label, otherwise negative label is assigned for unmatched pair.

Feature extraction techniques improve recognition quality by increasing the diversity of the features extracted from the pedestrian image. The recognition task become easier and provide better results with richer and higher dimensional representations of the pedestrian object. A large set of representative features have been proposed in the literature, which include: Edge information (Luo *et al*., 2014; Lim *et al*., 2013), color information (Walk *et al*., 2010; Dollár *et al*., 2009), texture information (Wang *et al*., 2009), local shape features (Daniel Costea and Nedevschi, 2014), covariance features (Paisitkriangkrai *et al*., 2013) and many others (Obaida, 2015). The more increase in

DOI: 10.3844/jcssp.2018.1431.1439

feature diversity the more systematically improvement in recognition performance.

In this study, pedestrian recognition using support vector machine with multi-scale weighted Histogram of Oriented Gradient is proposed. First images are resized into multi-scale. Next, edge detection is applied for each scale and used to compute a variable weight. Then instead of using the standard HOG, a Variable weight is adopted to generate a higher diversity HOG features around pedestrian edges. Finally, support vector machine is used to train the extracted features and perform pedestrian recognition.

The remainder of the paper is organized as follows: Section 2 presents the proposed Multi-Scale weighted HOG feature extraction technique. Experimental results are discussed in section 3. Finally, Section 4presents the conclusion.

Multi-Scale Weighted HOG Feature Extraction

This section gives an overview of proposed multi-scale weighted Histogram of oriented Gradient feature extraction method as shown in Fig. 1. The method is based on evaluating weighted HOG using multiple scales. First, the input image is rescaled to generate two scales 32 by 16 pixels and 16 by 8 pixels. Next, canny edge detector is applied on both scales. Then, the edge image is used to compute a variable weight to sale the HOG features of each scale. In the next step the generated features are concatenated and used to train a medium Gaussian SVM. Finally, Pedestrian recognition is achieved.

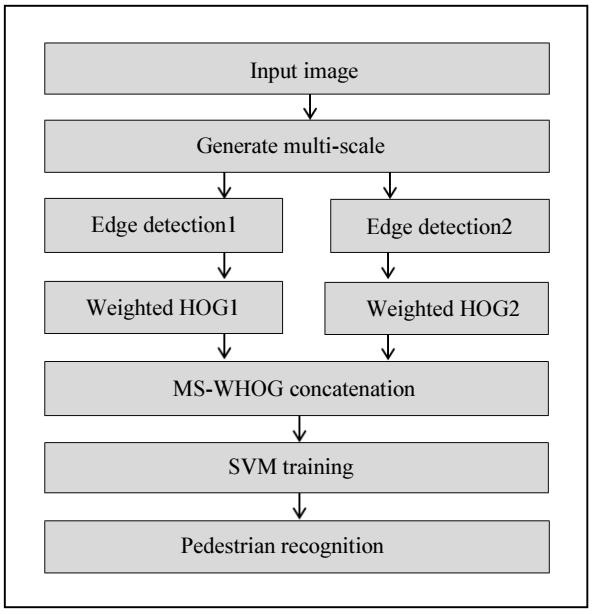


Fig. 1: Flow diagram of MS-WHOG

Edge Detection

Image edge is defined as a set of pixels with large variation in the neighboring gray scale pixels. It occurs at the border of object to distinguish between regions. Edge detection can be used in image segmentation, texture analysis and object recognition. There are several operators used to detect image edge, most of which use differential operators to detect to the step variation in grey scale level (Al-Hazaimeh *et al*., 2018; Binelli *et al*., 2005; Al-Smadi *et al*., 2016b). The first order derivative use either 2×2 or 3×3 directional derivative mask as an edge operator like Sobel, Prewitt and Roberts edge operators. The operators are convolved with each pixel to extract edge. On the other hand, canny operator assumes certain constraint conditions, which make it an optimal edge detector.

Sobeledge operator used two convolutional masks as shown in Fig. 2. Both kernels are convolved with each pixel in the image and the maximum value is the edge magnitude of either vertical or horizontal edge.

Prewitt edge operator also use two convolutional masks as shown in Fig. 3. The convolution is performed like Sobel operator and the result also provide either vertical or horizontal edge according to the maximum convolutional result.

Roberts edge operator use a single local differential operator as follow:

$$g(x,y)=\sqrt{\left[\sqrt{f(x,y)}-\sqrt{f(x+1,y+1)}\right]^2+\left[\sqrt{f(x+1,y)}-\sqrt{f(x,y+1)}\right]^2}$$

The Laplacian of Gaussian filter (LoG) use the second order derivative to compute gradient as:

$$\nabla^2 f=\frac{d^2 f}{dx^2}+\frac{d^2 f}{dy^2}$$

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Fig. 2: Sobel Edge operators

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
-1	0	1

Fig. 3: Prewitt Edge operators

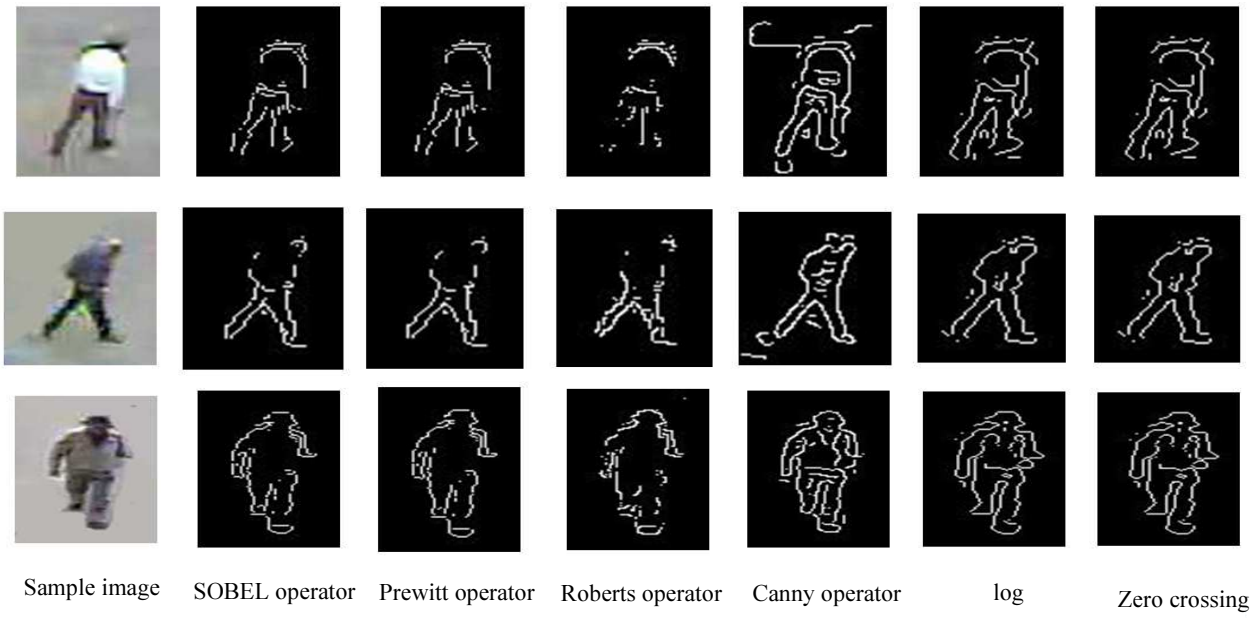


Fig. 4: Edge detection using Sobel, Prewitt, Roberts Canny, log and zero crossing operators for various pedestrian images

Zero crossing edge detector is based on the Laplacian of Gaussian filter. Where image edge give rise to the zero crossing in the Laplacian of Gaussian output. Figure 4 shows some pedestrian edge detection samples processed by Sobel, Prewitt, Roberts Canny, log and zero crossing edge operators.

Canny Edge Detection

Canny edge detector implements calculus of variation to find the optimal edges. It consists of four main steps: Gaussian filter for noise reduction, magnitude and angle gradient computation, non-maxima suppression and hysteresis threshold. Noise reduction or image smoothing is performed by using a Gaussian function as follow:

$$F_s(x,y) = F_s(x,y) \times G(x,y)$$

$$G(x,y) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} = \frac{1}{\pi\sigma^4}\left(\frac{x^2+y^2}{\sigma^2} - 1\right)e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)}$$

where, x,y are pixel coordinate, s is the mean square deviation of Gaussian distribution. Gaussian filter with small kernel size is used to detect explicit and small edges, while large kernel size is used to discover smoothed and thick edges.

In the second step, the maximum derivative values are calculated using the first differential operator and the gradient of $F_s(x,y)$ is calculated as follow:

$$P(i,j) \approx \frac{\left(F_s(i,j+1) - F_s(i,j)\right) + \left(F_s(i+1,j+1) - F_s(i+1,j)\right)}{2}$$

$$Q(i,j) \approx \frac{\left((F_s(i,j) - F_s(i+1,j)\right) + \left(F_s(i,j+1) - F_s(i+1,j+1)\right)}{2}$$

Then, orthogonal coordinates of each pixel are transformed into polar coordinates to find the magnitude and orientation of each pixel as follow:

$$M(i,j) = \sqrt{P(i,j)^2 + Q(i,j)^2}$$

$$\theta(i,j) = \tan^{-1}\left(\frac{Q(i,j)}{P(i,j)}\right)$$

where, $M(i,j)$ is the edge magnitude, $\theta(i,j)$ is the edge orientation. The overall edge orientation is expressed by $\theta(i,j)$ for which $M(i,j)$ reach the local maximum.

Canny operator is an accurate, robust, powerful and noise invariance edge detection technique that differs from other edge detection techniques in that it utilizes two different thresholds (upper and lower) to detect strong and weak edges. The output will include weak edges only if they are connected to strong ones. Thus, it can detect all edges as close as possible to the real edges existing in the image with minimal response.

The edge image of each scale is divided into cells similar to HOG cells. For each cell the weight is calculated as the ration of edge pixels to the total cell size, which is always less than on:

$$W_{cell} = \frac{\sum_{cell} E_{x,y}}{m \times n}$$

where, $E_{x,y}$ is edge pixels, m, n are cell dimensions.

Multi-Scale weighted HOG Feature

The Histogram of Oriented Gradient features HOG are defined as the orientation of gradients for each pixel within the region of interest. After that, the orientations are quantized into a predefined number of angles and their histograms are calculated in a feature vector. It was first proposed by Dalal and Triggs (2005) for object recognition. As a spatial shape descriptor, it uses statistical information to represent global and local shape regions to recognize objects. The horizontal (dx) and vertical (dy) a gradient for each pixel is given by the following equation:

$$dx = F(x+1, y) - F(x-1, y)$$
$$dy = F(x, y+1) - F(x, y-1)$$

The unsigned gradient orientation $\theta(i,j)$ calculated by the following equation, will be in the range (0°-180°):

$$\theta(x, y) = \tan^{-1}\left(\frac{dy}{dx}\right)$$

Alternatively, signed gradient orientation $\theta(i,j)$ calculated by the following equation, will be in the range (0°-360°):

$$\theta(x, y) = \tan^{-1}\left(\frac{dy}{dx}\right)$$

The resulting orientations are accumulated into n histogram bins. Then, the histogram values from all blocks are normalized and concatenated to form the feature vector. The size of the HOG feature vector is computed as:

$$N = \left(\frac{R_w}{C_w} - 1\right) \times \left(\frac{R_h}{C_h} - 1\right) \times B \times H$$

where, R is the region, C is the cell size, B is the number of cells per block and H is the number of histograms bins.

For pedestrian recognition, the image is smoothed with a Gaussian filter and scaled to 32×16 pixels and 16×8 images. The first image scale is divided into 8 cells of 8×8 pixel each. A block contains a group of 2×2 adjacent cells is formed to compute the WHOG for the first scale. The second scale is also divided into 8 cells with 4×4 pixel each. The block is also formed of 2×2 adjacent cells. The number of orientation bins is 8 for both scales, which are equally spaced between 0 and 360 degrees. The generated feature vector size is 12×8×2 = 192.

For each HOG cell the normalized histogram is multiplied by the weighting factor (1+ W_{cell}), which will increase histogram bins over the edge pixel depending on the edge content of cell. This is means, that the cell containing higher edge pixels will have higher histogram bins in the feature vector. On the other hand, the pedestrian surrounding cells with lower or no edge pixels will be represented by the standard HOG bins.

Experimental Results

The experimental setup aims to compare pedestrian recognition methods based on Support Vector Machine SVM, using HOG, MS-HOG and MS-WHOG, to find the best recognition algorithm.

The experimental data was taken from MIO-TCD classification dataset (Luo *et al*., 2018) contains 648,959 images of different categories, 6156 of them are pedestrian images. A total of 2150 pedestrian and non-pedestrian image were selected as experimental data, half the experimental data were used for training and the other half for testing. Some samples are shown in Fig. 5a and 5b show a pedestrian image and Fig. 5c and 5d show non-pedestrian image.

Table 1, provide a comparison of the overall accuracy and error results between Linear, Quadratic, Cubic, Medium Gaussian and Course Gaussian SVM using Three different features: HOG, MS-HOG and the proposed MS-WHOG. The results are the median of the 5-fold cross validation. The table shows that the Medium Gaussian SVM outperforms other SVM's, using HOG or MS-WHOG feature sets individually to recognize pedestrian. In this case, MS-WHOG outperforms the standard HOG in 5.5% increase in the accuracy. Although, Cubic SVM is the best SVM for MS-HOG, the accuracy is 4.9% lower than the proposed MS-WHOG.

The accuracy of the proposed MS-WHOG approach using different edge detection operators is shown in Table 2. It is clear that canny edge operator provide the best performance with the highest accuracy. While the second-best performance is achieved by Prewitt, Roberts, Log of Gaussian or Zero crossing depending on the SVM training technique. The worst performance of the proposed MS-WHOG was obtained using Sobel edge operator, which is still slightly better than HOG and MS-HOG.

DOI: 10.3844/jcssp.2018.1431.1439

(a) (b) (c) (d)

Fig. 5: (a, b) pedestrian image and (c, d) non-pedestrian image

Table 1: Accuracy and error results for HOG, MS-HOG and MS-WHOG using various SVM training techniques

	HOG		MS-HOG		MS-WHOG	
	Accuracy	Error	Accuracy	Error	Accuracy	Error
LSVM	75.9	24.1	80.3	19.7	83.2	16.8
QSVM	82.4	17.6	84.4	15.6	88.1	11.9
CSVM	82.6	17.4	84.7	15.3	89.2	10.8
MGSVM	83.8	16.2	83.4	16.6	89.6	10.4
CGSVM	70.5	29.5	78.7	21.3	82.6	17.4

Table 2: Accuracy results for MS-WHOG using various SVM training techniques for different edge operators

MS-WHOG Accuracy						
Edge Operator	Sobel	Prewitt	Roberts	Log	Zero crossing	Canny
LSVM	81.7	81.7	82.0	82.2	82.0	83.2
QSVM	86.2	86.3	86.8	86.2	86.5	88.1
CSVM	86.9	86.5	87.4	86.5	87.0	89.2
MGSVM	86.9	86.4	87.0	86.5	86.8	89.6
CGSVM	79.8	80.2	80.0	79.7	80.1	82.6

Table 3: Confusion matrix for Medium Gaussian SVM with HOG features

Predicted class			
		Pedestrian	Non-Pedestrian
True class	Pedestrian	901.0174.0	
		83.8%	16.2%
	Non-pedestrian	145.0930.0	
		13.5%	86.5%

Table 4: The Confusion matrix for Cubic SVM with MS-HOG features

Predicted class			
		Pedestrian	Non-pedestrian
True class	Pedestrian	910.0	165.0
		84.7%	15.5%
	Non-pedestrian	186.0	889.0
		17.3%	82.7%

Table 5: Confusion matrix for Medium Gaussian SVM with MS-WHOG features

Predicted class			
		Pedestrian	Non-pedestrian
True class	Pedestrian	964.0	111.0
		89.7%	10.3%
	Non-Pedestrian	113.0	962.0
		10.5%	89.5%

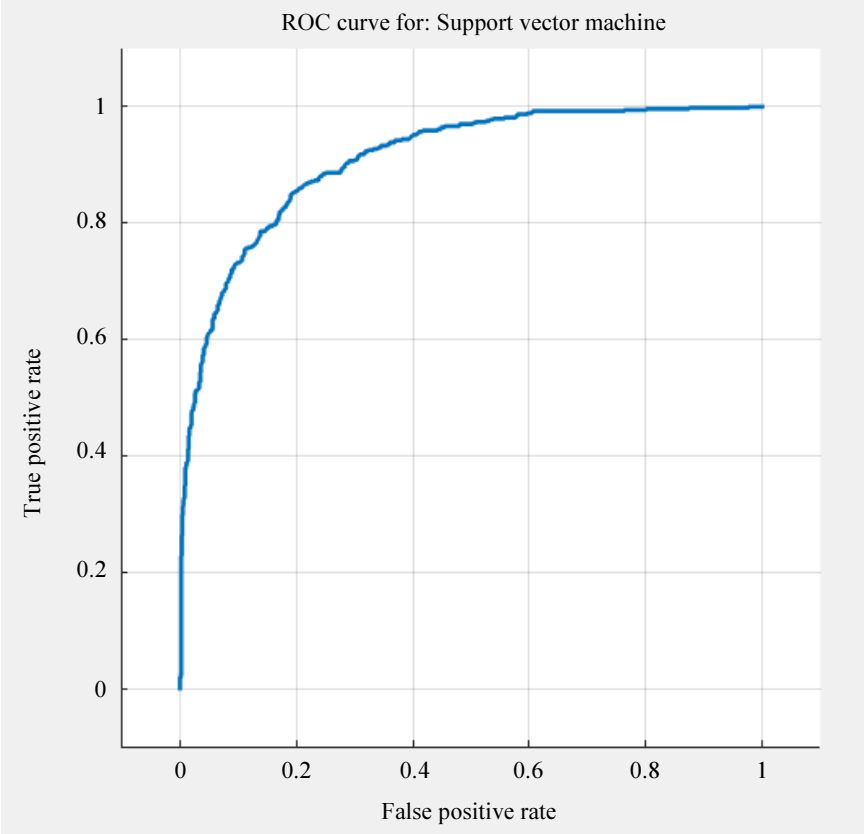
In order to evaluate the pedestrian recognition, the true positive rate (TPR) is defined as:

$$TPR = \frac{Number\ of\ correctly\ recognized\ pedestrian}{Total\ Number\ of\ Pedestrian}$$

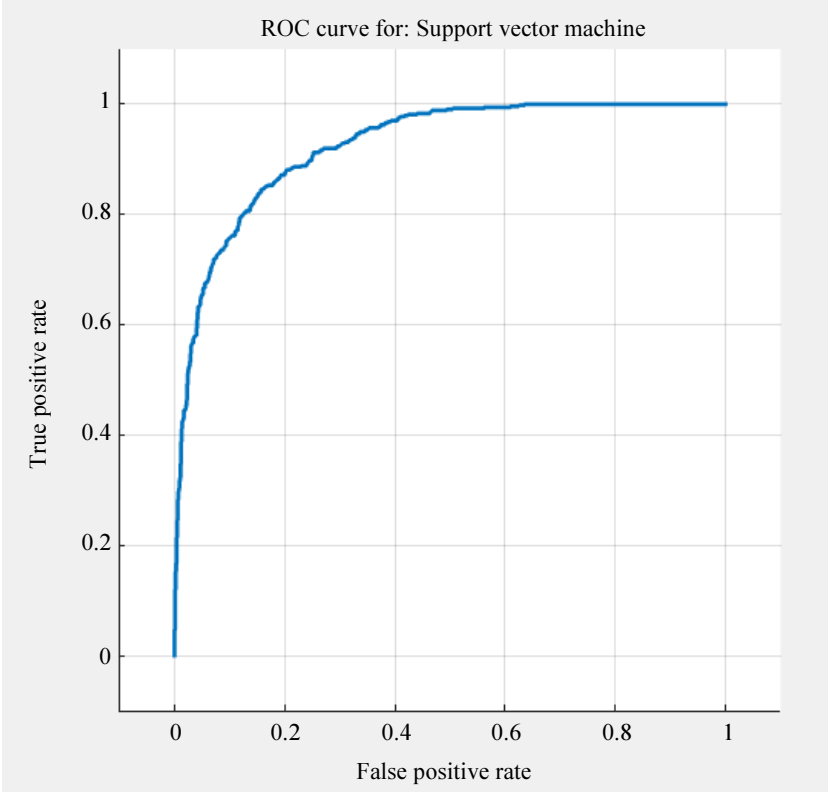
The confusion matrix associated with medium Gaussian SVM and standard HOG feature is shown in Table 3. The TPR for pedestrian recognition is 0.84.

The Confusion matrix for Cubic SVM with MS-HOG features is shown in Table 4. The TPR for pedestrian recognition is 0.89.

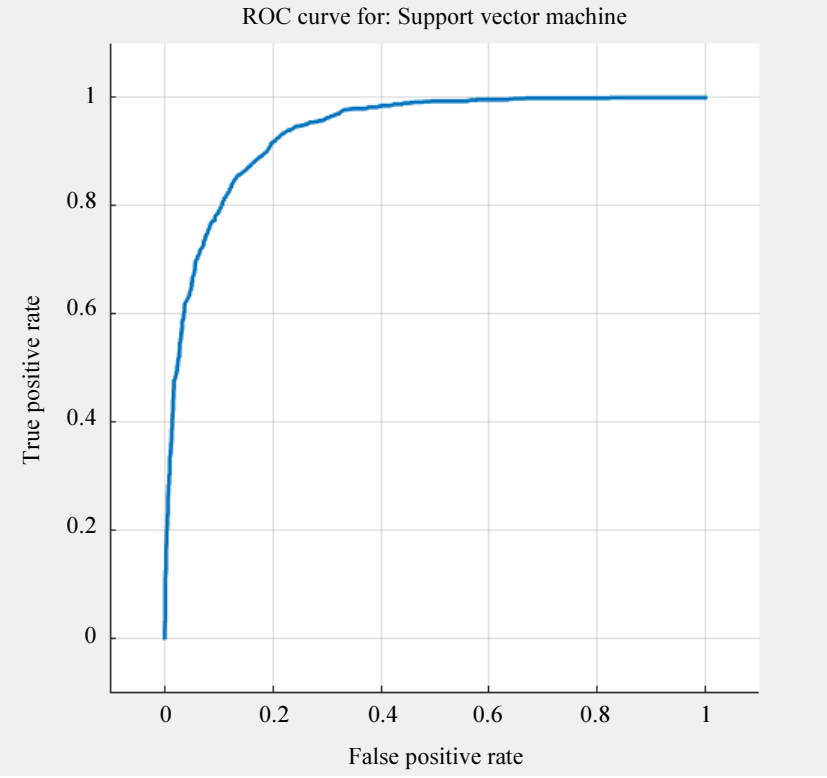
DOI: 10.3844/jcssp.2018.1431.1439



(a)



(b)



(c)

Fig. 6: ROC curve for pedestrian recognition using the best SVM for (a) HOG, (b) MS-HOG and (c) MS-WHOG features

In this study, Medium Gaussian SVM was used to perform pedestrian recognition for the entire data set, which provides the best performance. The associated confusion matrix is shown in Table 5, with overall TPR of 0.897.

Figure 6 show the ROC curve for pedestrian recognition using the best SVM HOG, MS-HOG and MS-WHOG features. The lower false positive rate is more interesting, for example if it is less than 0.1 the true positive rate will be less than 0.8 for both Medium Gaussian SVM with HOG and Cubic SVM with MS-HOG. On the other hand, the true positive rate for MS-WHOG will be greater than 0.8 for the same false positive rate.

Conclusion

This paper proposed a detection algorithm using a Multi-scale weighted HOG (MS-WHOG) model based on Median Gaussian SVM for pedestrian recognition. Because pedestrians have deformable representation with complex details, based on Canny edge operator, weighted feature extraction was applied on multiple scale images. Unlike the original HOG, edge content was used as a weighting factor to increase the representation of distinguished features. Experimental results show that the proposed MS-WHOG using Median Gaussian SVM outperforms other SVM's with accuracy of 89.6. Moreover, it outperforms the standard HOG and the multi-scale HOG.

Acknowledgment

The author would like to thank all the people who have supported this work, as well as special thanks to Dr. Obaida M. Al-Hazaimen and Eng. Ma'moun Al-Smadi from Al-Balqa Applied University.

Ethics

This article is original and contains unpublished Material, the corresponding author confirms that no ethical issues involved.

References

Al-Hazaimeh, O.M., M. Al-Nawashi and M. Saraee, 2018. Geometrical-based approach for robust human image detection. Multimedia Tools Applic. DOI: 10.1007/s11042-018-6401-y

Al-Hazaimeh, O.M., M.F. Al-Jamal, N. Alhindawi and A. Omari, 2017. Image encryption algorithm based on Lorenz chaotic map with dynamic secret keys. Neural Comput. Applic. DOI: 10.1007/s00521-017-3195-1

Al-Nawashi, M., O.M. Al-Hazaimeh and M. Saraee, 2017. A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. Neural Comput. Applic., 28: 565-572. DOI: 10.1007/s00521-016-2363-z

Al-Smadi, M., K. Abdulrahim and R.A. Salam, 2016a. Traffic surveillance: A review of vision based vehicle detection, recognition and tracking. Int. J. Applied Eng. Res., 11: 713-726.

Al-Smadi, M., K. Abdulrahim and R.A. Salam, 2016b. Cumulative frame differencing for urban vehicle detection. Proceedings of the 1st International Workshop on Pattern Recognition, (WPR' 16), SPIE, pp: 100110G-100110G. DOI: 10.1117/12.2242959

Binelli, E., A. Broggi, A. Fascioli, S. Ghidoni and P. Grisleri *et al*., 2005. A modular tracking system for far infrared pedestrian recognition. Proceedings of the IEEE Intelligent Vehicles Symposium, Jun. 6-8, IEEE Xplore Press, Las Vegas, NV, USA, pp: 759-764. DOI: 10.1109/IVS.2005.1505196

Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 20-25, IEEE Xplore Press, San Diego, CA, USA, pp: 886-893. DOI: 10.1109/CVPR.2005.177

Daniel Costea, A. and S. Nedevschi, 2014. Word channel based multiscale pedestrian detection without image resizing and using only one classifier. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 23-28, IEEE Xplore Press, Columbus, OH, USA, pp: 2393-2400. DOI: 10.1109/CVPR.2014.307

Dikmen, M., E. Akbas, T.S. Huang and N. Ahuja, 2010. Pedestrian recognition with a learned metric. Proceedings of the Asian Proceedings of the 10th Asian Conference on Computer Vision, Nov. 08-12, Springer, Queenstown, New Zealand, pp: 501-512. DOI: 10.1007/978-3-642-19282-1_40

Dollár, P., Z. Tu, P. Perona and S. Belongie, 2009. Integral channel features.

Felzenszwalb, P.F., R.B. Girshick, D. McAllester and D. Ramanan, 2010. Object detection with discriminatively trained part-based models. IEEE Trans. Patt. Anal. Mach. Intell., 32: 1627-1645. DOI: 10.1109/TPAMI.2009.167

Gharaibeh, N., O.M. Al-Hazaimeh, B. Al-Naami and K.M. Nahar, 2018. An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images. Int. J. Signal Imag. Syst. Eng., 11: 206-216. DOI: 10.1504/IJSISE.2018.093825

Gray, D. and H. Tao, 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features. Proceedings of the 10th European Conference on Computer Vision, (CCV' 08), Springer, Marseilles, France, pp: 262-275. DOI: 10.1007/978-3-540-88682-2_21

Lim, J.J., C.L. Zitnick and P. Dollár, 2013. Sketch tokens: A learned mid-level representation for contour and object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 23-28, IEEE Xpllore Press, Portland, OR, USA, pp: 3158-3165. DOI: 10.1109/CVPR.2013.406

Luo, P., Y. Tian, X. Wang and X. Tang, 2014. Switchable deep network for pedestrian detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 23-28, IEEE Xplore Press, Columbus, OH, USA, pp: 899-906. DOI: 10.1109/CVPR.2014.120

Luo, Z., B. Frederic, C. Lemaire, J. Konrad and S. Li *et al*., 2018. MIO-TCD: A new benchmark dataset for vehicle classification and localization. IEEE Trans. Image Process., 27: 5129-5141. DOI: 10.1109/TIP.2018.2848705

Obaida, M.A.H., 2015. Combining audio samples and image frames for enhancing video security. Ind. J. Sci. Technol., 8: 940-949. DOI: 10.17485/ijst/2015/v8i10/53149

Paisitkriangkrai, S., C. Shen and A. Van Den Hengel, 2013. Efficient pedestrian detection by directly optimizing the partial area under the ROC curve. Proceedings of the IEEE International Conference on Computer Vision, Dec. 1-8, IEEE Xplore Press, Sydney, NSW, Australia, pp: 1057-1064. DOI: 10.1109/ICCV.2013.135

Park, U., A.K. Jain, I. Kitahara, K. Kogure and N. Hagita, 2006. VISE: Visual search engine using multiple networked cameras. Proceedings of the 18th International Conference on Pattern Recognition, Aug. 20-24, IEEE Xplore Press, Hong Kong, China, pp: 1204-1207. DOI: 10.1109/ICPR.2006.1176

Su, C.L., Y.H. Chang, K.P. Chen and J.H. Wu, 2015. Pedestrian recognition using feature extraction. Proceedings of the IEEE International Conference on Consumer Electronics-Taiwan, Jun. 6-8, IEEE Xplore Press, Taipei, Taiwan pp: 274-275. DOI: 10.1109/ICCE-TW.2015.7216895

Tuzel, O., F. Porikli and P. Meer, 2008. Pedestrian detection via classification on riemannian manifolds. IEEE Trans. Patt. Anal. Mach. Intell., 30: 1713-1727. DOI: 10.1109/TPAMI.2008.75

Walk, S., N. Majer, K. Schindler and B. Schiele, 2010. New features and insights for pedestrian detection. Proceedings of the IEEE Conference on in Computer Vision and Pattern Recognition, Jun. 13-18, IEEE Xplore Press, San Francisco, CA, USA, pp: 1030-1037. DOI: 10.1109/CVPR.2010.5540102

Wang, X., T.X. Han and S. Yan, 2009. An HOG-LBP human detector with partial occlusion handling. Proceedings of the IEEE 12th International Conference on in Computer Vision, Sept. 29-Oct. 2, IEEE Xplore Press, Kyoto, Japan, pp: 32-39. DOI: 10.1109/ICCV.2009.5459207