Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 45 – 52

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

ACM based ROI Extraction for Pedestrian Detection with Partial Occlusion Handling

Viswajith P. Viswanath[a], Ragesh N. K.[b], Madhu S. Nair[a,*]

[a]*Department of Computer Science, University of Kerala, Kariavattom, Thiruvananthapuram-695581, Kerala, India.*
[b]*Transportation Business Unit, Network Systems & Technologies (P) Ltd., Technopark Campus, Kariavattom, Thiruvananthapuram -695581, Kerala, India.*

Abstract

Pedestrian detection in video surveillance systems is an integral part of Advanced Driver Assistance Systems (ADAS). In this paper, a new method for efficient pedestrian detection is proposed. The proposed method uses ACM (Active Contour Model) for efficiently locating pedestrian position in each video frame and thereby speeding up the detection time. This method uses a combination of HOG (Histogram of Oriented Gradients) and LBP (Local Binary Patterns) as features for training a two level linear SVM (Support Vector Machine). The proposed method handles partial occlusion using a two-level SVM classifier and eliminates multiple detection using Non Maximum Suppression (NMS) algorithm. The performance analysis is done using INRIA Person dataset and CVC Partial Occlusion dataset; and it is found that the proposed method gives promising results in terms of detection accuracy and detection speed.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Pedestrian detection;region of interest; histogram of oriented gradients; active contour model; local binary pattern;

1. Introduction

With the increasing rate of growth of automobile industry over the last decade, the rate of road accidents also rises. Millions of road accidents are happening around the world each year. About 26 percent of these accidents

* Corresponding author. Tel.: +91-9447364158.
E-mail address: madhu_s_nair2001@yahoo.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.01.049

cause serious injuries to people[1]. So the automobile industry is trying to develop new technologies to improve traffic safety. Now the research is concentrating on making the vehicle intelligent in order to avoid accidents. These intelligent systems are referred to as Advanced Driver Assistance Systems (ADAS). An integral part of ADAS is Pedestrian Protection System (PPS). PPS tracks people in front of a moving vehicle and warns the driver about a possible collision[2]. But detecting and tracking pedestrians is a challenging task primarily because of large intra-class variance (pose, clothes, height etc.). Another challenge is that the PPS must respond in real time while maintaining the robustness.

Pedestrian detection systems are of two types, model-based and feature-classifier based. In model-based pedestrian detection, we search the image or video frame to find matched locations with a predefined pedestrian model or template for finding the pedestrians[3]. But we need a large number of templates to efficiently represent various poses of pedestrians, which in turns increases the matching time. Z.Lin and L.S.Davis[4] uses a shape-based, hierarchical part-template matching approach combining local part-based and global shape-based schemes that matches a part-template tree to images (hierarchically) to detect humans. D.M Gavrila[5] uses a template tree to efficiently represent and match a variety of shape exemplars. Marco Pedersoli et al.[6] proposed a system for pedestrian detection based on a hierarchical multi-resolution part-based model. These approaches show a reasonable detection performance, but they are computationally too expensive for real-time performance.

Feature-classifier based approach shows more promising results in terms of computational complexity. Most of the pedestrian detection algorithms use shape features[3] such as Edgelet feature[7], Shapelet feature[8] etc. N. Dalal and B. Triggs[9] proposed Histogram of Oriented Gradient (HOG) as a shape descriptor for detecting the pedestrians. It can describe the local shape of objects effectively. But it performs poor when the background of the image is cluttered with edges. These feature-classifier based approaches uses blind sliding-window search for detecting the pedestrians from a single image. It is a simple method but it is time consuming as a large number of non-pedestrian windows need to be processed. For an efficient PPS, detecting pedestrians in presence of partial occlusions is of great importance. But in all these approaches, the occurrence of partial occlusions causes degradation in performance.

In order to avoid the above mentioned problems our proposed method (i) uses a combination of LBP[13] (Local Binary Pattern), which captures texture information and HOG[9] as feature vector (ii) employs ACM based region of interest (ROI) segmentation method for reducing unwanted area to be processed (iii) uses a two-level classification for handling partial occlusions and (iv) uses a new Non Maximum Suppression (NMS) method for avoiding the multiple detections.

The rest of the paper is organized as follows. Section 2 explains the proposed method. Section 3 deal with the experimental analysis and section 4 concludes the work.

2. Proposed Method

The framework of the proposed method is shown in Fig. 1. It consists of two phases, training phase and detection phase. During the training phase we train the two-level linear Support Vector Machine (SVM) using training vectors from INRIA Person dataset and CVC Virtual Pedestrian dataset. In detection phase the system identifies the pedestrians present in the real time input image.

2.1 Training Phase

We use the HOG-LBP feature vectors extracted from training images in INRIA Person dataset for training the first level SVM classifier. We also use partially occluded pedestrian images from CVC Virtual Pedestrian dataset for training the 2nd level SVM classifier.

2.2 Detection Phase

The detection phase is comprised of five modules, (i) ROI extraction, (ii) image scaling, (iii) feature extraction, (iv) classification and (v) NMS. The ROI extraction module locates the region of the image containing pedestrians and passes it to image scaling module. The image scaling module creates a pyramid of scaled images for finding pedestrians with different heights. The feature extraction module employs a sliding-window approach for extracting the HOG-LBP feature vector for each window. After that the classification module classifies each window as

pedestrian or non-pedestrian using a two-level linear SVM classifier. Finally, the NMS algorithm will be applied on the classification results to avoid multiple detections.

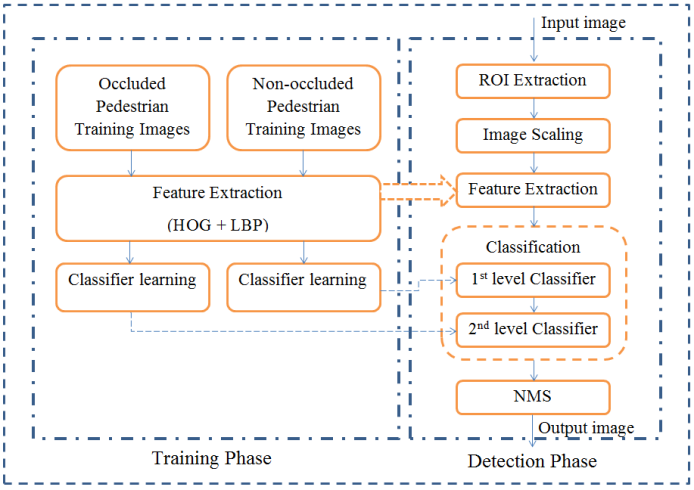
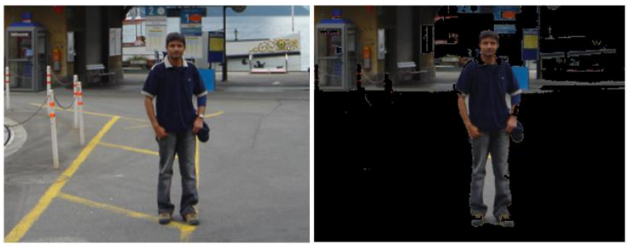


Fig. 1. The framework of proposed method

2.2.1 ROI Extraction

In traditional pedestrian detection methods, detection windows are extracted using a simple sliding window search. This method is simple but it adds a lot of overhead for the algorithm such as processing unwanted regions. So, instead of using a blind search mechanism this method uses an ROI based detection window extraction method. But while segmenting the ROI we must make sure that the algorithm is not altering the shape information of pedestrian. The detection process will be more accurate if the segmentation algorithm highlights the shape information of the pedestrian. A segmentation algorithm that best suits this purpose is Active Contour Model (ACM)[10].



(a) (b)

Fig. 2. (a) Input image (*I*) (b) Output of ROI Extraction module (I_o)

ACM tries to minimize the total energy associated with the current contour, which is defined as the sum of internal and external energy. When the contour is at the desired object boundary location both the internal energy and external energy becomes minimum. Let I be the image currently being processed, C be the desired object boundary, and c_1, c_2 be average of I inside C and outside C, respectively. Then the energy function is defined as:

$$F(C, c_1, c_2) = \mu \,.\, Length(C) + \nu \,.\, Area(C) + \lambda_1 \int_{Inside(C)} |I(x,y) - c_1|\, dx\, dy + \lambda_2 \int_{Outside(C)} |I(x,y) - c_2|\, dx\, dy, \quad (1)$$

where $\mu \geq 0, \nu \geq 0, \lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are fixed parameters. In our method, we empirically fix the values of $\lambda_1 = \lambda_2 = 1$, $\nu = 0$ and $\mu = 0.3 \times 255^2$. Now we solve this minimization problem to find the set of points in I that defines C. After finding C the segmented image I_o is obtained as follows:

$$I_o(x,y) = I_{ACM}(x,y) \times I(x,y),$$
$$where\ I_{ACM}(x,y) = \begin{cases} 0, if\ (x,y) is\ outside\ C \\ 1, \qquad otherwise \end{cases} \tag{2}$$

The image I_o thus obtained eliminates most of the unwanted region in the image by replacing the grey level values of pixels in that region with zeros (black colour). For example, in Fig. 2(b), the black coloured (zero valued) pixels shows the unwanted area. This eliminates the burden of processing a lot of unwanted area and thereby reduces the computational complexity of the detection process.

2.2.2 Image Scaling

After finding ROI the original high resolution image will be scaled to different dimensions. It helps in finding pedestrians with different heights. It is assumed that pedestrians closer to the vehicle will be present in coarse scale image and those who are far away from the vehicle can be found by scanning the fine scale image. The following equation shows scaling function:

$$Scale(I_o) = \{(I_o,\sigma_1),(I_o,\sigma_2),(I_o,\sigma_3),\dots,(I_o,\sigma_n)\} \tag{3}$$

where $Scale(I_o)$ is the scaling function which creates n scaled versions of I_o with scaling factor σ, using bi-cubic interpolation. For high quality images $\sigma \in [1,0)$.

2.2.3 Feature Extraction

Histogram of Oriented Gradients (HOG) is a feature descriptor used in object detection systems. The HOG descriptor describes the shape and appearance of the object in an image. But its performance is not promising when the background of the image is cluttered with edges. So if we use some other feature descriptor which captures texture information (like LBP) along with HOG, the detection accuracy will get improved.

Here, we divide the scaled image (I_o,σ_n) into 64×128 sized windows with 90% overlap. We use the method specified by N. Dalal and B. Triggs[9] for extracting the HOG features. For finding LBP feature we divide the examined window say w into non overlapping cells of size 16×16. Let c_i be the i[th] cell in w, $c_i(x,y)$ be any pixel in c_i, and $p_1, p_2, \dots\dots, p_8$ be the 8 neighboring pixels of $c_i(x,y)$; then for each pixel in c_i, we compute:

$$c_i(x,y) = \sum_{S=1}^{8} f\big(p_S - c_i(x,y)\big) \times 2^{s-1},$$
$$where\ f(z) = \begin{cases} 1, & if\ z < 0 \\ 0, & otherwise \end{cases}, \tag{4}$$

Then we compute the 128 bin histogram for each cell. Now the LBP feature vector for the window w is found by concatenating the histograms of all cells. Finally, we concatenate the 4096 features of LBP with 3780 HOG features to form the final feature vector F_w. That is:

$$F_w = [HOG_1, HOG_2, \dots, HOG_{3780}, LBP_1, LBP_2, \dots, LBP_{4096}\ ] \tag{5}$$

2.2.4 Classification

After extracting feature vector F_w for all windows w, the classification function classifies F_w as pedestrians or non-pedestrians based on detection scores returned by a trained classifier (SVM). The following equation shows the process:

$$Classify(F_w) = \begin{cases} pedestrians, & if\ SVM_1(F_w) \geq \alpha \\ non\text{-}pedestrians, & if\ SVM_1(F_w) \leq -\alpha \\ not\text{-}sure, & otherwise \end{cases} \tag{6}$$

where $SVM_1(F_w)$ represents the 1[st] level classifier. If the detection score returned by $SVM_1(F_w)$ for a feature vector F_w is greater than or equal to the threshold α, then the window w may contain a pedestrian. In our proposed method, the value of α is empirically set as 0.3.

It is seen that when the SVM is confused about the class of a feature vector, it returns detection score near to zero. There is a huge possibility that this class of windows contains partially occluded pedestrians. So in order to detect them we use a second level classifier (trained with partially occluded pedestrian training vectors) that classifies 'not-sure' class of windows to partially occluded pedestrian and non-pedestrians. Let F_{NS_i} be the i[th] feature vector in the set F_{NS}, where $F_{NS} = \{F_w | \, F_w \in not\text{-}sure \}$. Now the classification process can be defined as:

$$Classify(F_{NS_i}) = \begin{cases} pedestrians, & if\ SVM_2(F_{NS_i}) > \beta \\ non\text{-}pedestrians, & otherwise \end{cases} \quad (7)$$

where $SVM_2(F_{NS_i})$ is the 2[nd] level classifier which returns the detection score and β is the threshold. If the detection score is greater than or equal to the threshold β, then the window w may contain a pedestrian. In our experiments, we set the value of $\beta = 0$.

2.2.5 Non Maximum Suppression

Since we are using sliding window approach with 90% window overlap, the overlapped windows might detect the same pedestrians over and over again. Also there is a high chance of detecting the same pedestrian in more than one scaled version of the image I_o. For instance, a single pedestrian may be detected twice in images (I_o, σ_i) and (I_o, σ_{i+1}). So we need to apply NMS to avoid multiple detections. Fig. 3(b) shows the output just before the NMS step.

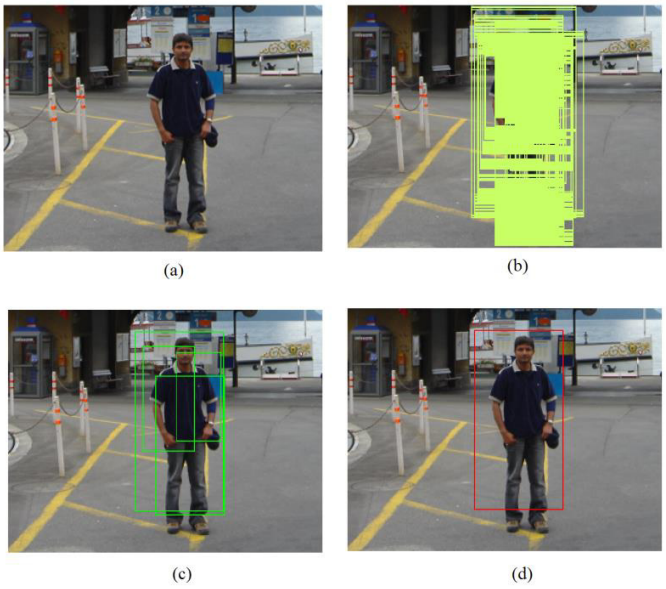


Fig. 3. (a). Input image (b) Detected windows (without applying NMS) with 99% overlap. (c). output of applying clustering the detection windows based on corner points. (d). Output of NMS

Traditional approach for avoiding multiple detection of a single pedestrian is to cluster the detection windows based on corner points. Fig. 3(c) shows the result of applying this idea. But we can see that some unwanted detections still remain there. Another approach is to cluster the detection windows based on their center points. But this method won't work when two more pedestrian stand close to each other. To overcome these problems the proposed method adopts a novel approach. The idea behind this approach is to cluster the detection windows based on their area of intersection. Let W_D be the set of all windows classified as pedestrians (including multiple detections) and W_{DE} be the set of all windows without multiple detection. Suppose there are r windows detected as pedestrians. That is:

$$W_D = \{ws_1, ws_2, ws_3, \dots, ws_i, \dots, ws_r\}, \qquad and \quad W_{DE} = \emptyset \quad (8)$$

where ws_i is the i[th] detected window scaled back to its original size. Then for each window ws_i we find a threshold $\theta_{i,k}$ as follows:

$$\theta_{i,k} = \frac{S(ws_i, ws_k)}{height(ws_k) \times width(ws_k)}, \tag{9}$$

where $i, k \in [1, r], i \neq k, k > i$, and the function $S(ws_i, ws_k)$ calculates the total number of overlapped pixels in ws_i and ws_k. After computing $\theta_{i,k}$ we remove all windows from W_D for which $\theta_{i,k} > \tau$, where τ is empirically set as 0.7. Then we add the window with highest detection score which satisfies the condition $\theta_{i,k} > \tau$ to the set W_{DE}. We repeat these steps until the set W_D becomes empty. The resulting set W_{DE} is the new set of detected windows. The result of applying NMS algorithm is shown in Fig. 3(d).

Fig. 4. (a) Images used for the comparison (b) Bar-diagram showing the total number of windows to be processed; with ACM and without ACM

3. Experimental Analysis

The algorithm is tested over various test images in INRIA Person dataset and CVC Partial Occlusion dataset. The experiments were conducted on a 2.20 GHz Windows7 PC with 4 GB RAM. Results show that the proposed method performs better than other state-of-the-art methods reported in the literature. Fig. 4(b) compares the computational complexity of our method with other methods in terms of number of windows being processed, using the images shown in Fig. 4(a). From Fig. 4(b) it is evident that our method outperforms the other state-of-the-art methods in terms of computational complexity by eliminating the overhead of processing large number of unwanted windows. Fig. 5(d)(ii) shows the robustness of proposed method in detecting pedestrians when the background of the image is cluttered with edges. In Fig. 5(d) (iii),(iv) we show that the proposed method outperforms the other state-of-the-art methods in detecting partially occluded pedestrians.

3.1 Training the Classifier

We have used 2416 positive training vectors and 2500 negative training vectors from INRIA Person dataset for training the first level SVM classifier. We also used 1000 positive training vectors from CVC Virtual Pedestrian dataset for training the 2nd level SVM classifier.

3.2 Results

We used three evaluation metrics for the performance analysis of the proposed method (i) detection accuracy, (ii) False Positives Per Image (FPPI), and (iii) detection speed. Let TP be the number of 'true positive' detections, TN be the number of 'true negative' detections, and FP be the number of 'false positive' detections. Then the detection accuracy and FPPI are defined as:

$$Detection\ Accuracy = \frac{TP + TN}{Total\ number\ of\ windows} \tag{10}$$

$$FPPI = \frac{FP}{Total\ Number\ of\ Images} \tag{11}$$

Fig. 5. (a) set of input images, (b),(c),(d) shows corresponding outputs of HOG-SVM system, HOG-LBP-SVM system, and proposed system, respectively.

We used 1126 positive images and 1000 negative images from INRIA Person dataset for finding the detection accuracy. As shown in Table I our method has 96.1% detection accuracy and it outperforms the other two methods. This method performs 50% faster than the HOG-SVM. The HOG-SVM system which takes 60 sec detection time can perform at 30fps on ADAS framework (using dedicated hardware). It implies that our method could achieve real-time video processing speed (ie. greater than 30fps) on ADAS framework. It is clear from Table I that our method has a good improvement in terms of FPPI. The proposed method achieves a minimum FPPI while maintaining high detection accuracy.

Table I. Comparison of Detection Accuracy and FPPI

Methods	Detection Accuracy	FPPI	Average Detection Time (in sec)
Proposed Method	96.1	0.1	30
HOG-LBP-SVM	94.12	0.27	65
HOG-SVM	90.40	0.33	60

4. Conclusion

In this paper, a new method for efficient pedestrian detection is proposed. The proposed method uses ACM based ROI extraction module for eliminating the unwanted areas in the image and thereby reducing the computational complexity of the detection process. This method uses a two-level SVM classifier for handling the partial occlusions. The proposed method also eliminates multiple detections using an NMS algorithm. As shown in experimental analysis, the proposed method reduces the computational complexity, improves detection accuracy and handles partial occlusions.

Acknowledgements

We would like to express our heartfelt thanks to Network Systems and Technologies (P) Ltd. for the kind help and support delivered.

References

1. Jin Nie, Jikuang Yang, Fan Li. A Study on Pedestrian Injuries based on Minivan and Sedan Real-World Accidents. *International Conference on Optoelectronics and Image Processing* 2010; 1(): 160-163.
2. David Geronimo, Antonio M. Lopez, Angel D. Sappa, Thorsten Graf. Survey of Pedestrian Detection for Advanced Driver Assistance Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2010; 32(7): 1239-1258.
3. Bo Li, Qingming Yao, Kunfeng Wang. A Review on Vision-based Pedestrian Detection in Intelligent Transportation Systems. *IEEE International Conference on Networking, Sensing and Control (ICNSC)* 2012; (): 393-398.
4. Z.Lin, L. S.Davis. Shape-based Human Detection and Segmentation via Hierarchical Part-Template Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2010; 32(4): 604-618.
5. D. M. Gavrila. A Bayesian, Exemplar-Based Approach to Hierarchical Shape Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2007; 29(8): 1408-1421.
6. Marco Pedersoli, Jordi Gonzàlez, Xu Hu, Xavier Roca. Toward Real-Time Pedestrian Detection Based on a Deformable Template Model. *IEEE Transactions on Intelligent Transportation Systems* 2014; 15(1): 355-364.
7. B. Wu, R. Nevatia. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. *IEEE International Conference on Computer Vision (ICCV)* 2005; 1(): 90-97.
8. P. Sabzmeydani, G. Mori. Detecting Pedestrians by Learning Shapelet Features. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2007; (): 1-8.
9. N. Dalal, B. Triggs. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* 2005; 1(): 886-893.
10. T. F. Chan, L. A. Vese. Active Contours Without Edges. *IEEE Transactions on Image Processing* 2001; 10(2): 266-277.
11. Jianxin Wu, Nini Liu, Christopher Geyer, James M. Rehg. C4: A Real-Time Object Detection Framework. *IEEE Transactions on Image Processing* 2013; 22(10): 4096 - 4107.
12. J. Marín, D. Vázquez, A.M. López, J. Amores, L. I. Kuncheva. Occlusion Handling via Random Subspace Classifiers for Human Detection. *IEEE Transactions on Cybernetics* 2014; 44(3): 342-354.
13. T. Ojala, M. Pietikainen, D. Harwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. *Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR)* 1994; 1(): 582-585.