New Improved Methodology for Pedestrian Detection in Advanced Driver Assistance System

Vijay Gaikwad
Department of Electronics
Vishwakarma Institute of Technology
Pune

Shashikant Lokhande
Department of E&TC
Sinhgad College of Engineering
Pune

Sanket, Manthan, Pravin
Department of Electronics
Vishwakarma Institute of Technology
Pune

ABSTRACT
In recent years, pedestrian detection (PD) plays a vital role in a variety of applications such as security cameras, automotive control and so forth. These applications require two essential features, i.e. high speed performance and high accuracy. Firstly, the accuracy is determined by how the image features are described. The image feature description must be robust against occlusion, rotation, and the change in object shapes and illumination conditions. A number of feature descriptors have been proposed. Previously histogram of oriented gradients(HOG) features were extensively used along with support vector machine (SVM) classifier for PD. HOG features and SVM classifier can achieve good performance for PD, but they are time consuming. To achieve high detection speed with good detection performance, a Two-step framework method was proposed by Zhen Li which was the fusion of Haar-like and HOG features to get better performance. Edgelet features were used for classification and detection. But, the detection rate was poor and computation speed was less. In order to alleviate these limitations, we propose here a new methodology for improving the detection rate and speed. The performance and accuracy of the detection can be improved by the combination of Haar-like and Triangular features for FBD and Edgelet and Shapelet for HSD. We expect an average 95% detection rate and 60% faster speed for the proposed method.

General Terms
Algorithms, Detection Rate, Image, Pedestrian, Feature-set, Learning method, Performance, Process, Experimentation, Accuracy, Processor.

Keywords
HOG, Haar-like, Triangular, Edgelet, Shapelet, Adaboost Classifier, and Video Processor

1. INTRODUCTION
Automatic PD is a rapidly evolving area in image analysis and surveillance, it is essential in many Applications such as advanced robotics and intelligent vehicles. It is a challenging problem due to changing articulated pose, style and color of clothing, background, illumination and weather conditions. More importantly, robust pedestrian identification and tracking are highly dependent on reliable detection and segmentation in each frame. There are many different techniques which have been proposed to address the problem of people detection. Most popular methods are based on feature extraction, which was developed in recent years. A good feature descriptor is crucial for accuracy of PD. Voila built an efficient moving person detector, which is based on Haar-like features. Takuya Kobayash improved the histogram of oriented gradients (HOG) features for pedestrian and obtained a high recognition rate. Bo Wu has used local edge pieces called edgelet for people detection. And Supriya Rao proposed a new method to detect individual body parts of people. However, these features are not robust enough to detect many unpredictable problems, so the fusion of these features is needed to realize a higher detection rate. Haar-like features can capture different image details and fast algorithms which use integral images can calculate such features very quickly. But they are not very efficient for PD due to the high variety of pedestrian appearances. HOG features and Shapelet features can achieve good performance for PD, but they are time consuming. To achieve high detection speed with good detection performance, a new methodology is proposed. A new 2-step framework detecting pedestrian through video image is introduced. This framework contains two steps, the first of which is full body detection (FBD). In this step, a detector is trained by the fusion of Haar-like and Triangular features to detect candidate areas. In the second step, called head–shoulder detection (HSD), another detector trained by edgelet/Shapelet features is applied to only detect the head–shoulder part of people in the FBD output area. Experiment results show that this method can take advantage of the speed merit of Haar-like features and good performance of edgelet/Shapelet features. The detection process includes two stages: In the first stage, a detector using Haar-like features scans the whole image over different positions and scales. So we get several rectangular regions which may contain pedestrians. In the second stage, a classifier using Shapelet features classifies them into pedestrian or non-pedestrian. The general architecture is shown in Fig.1 Viola and Jones proposed Haar-like features for rapid object detection and then applied to the pedestrian detection.

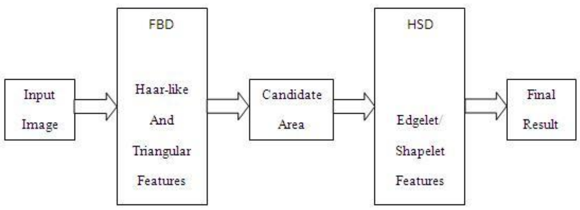

Fig1 : General Architecture

With the simple Haar-like features which can be calculated efficiently by using integral images and AdaBoost classifiers in a cascade structure, their detector has high detection speed. Lienhart add some other similar features to extend the Haar-like feature sets. In our method we use the Haar-like features used in and some extended Haar-like features proposed by Lienhart. These features are shown in Fig. 2.

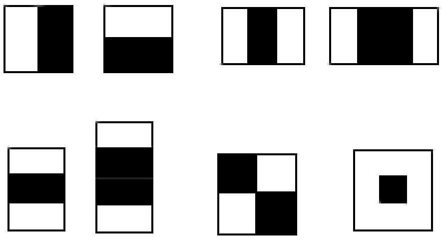

Fig 2 : Haar-like features

A Haar-like feature is composed of several white or black areas. The intensity values of pixels in the white or black areas are separately accumulated. Then the feature value is computed with a weighted combination of these two sums. The detector is trained through AdaBoost algorithm and finally composed of a cascade of classifiers. The first several classifiers are used to reject a large number of negative examples with simple computation. Then subsequent classifiers are used to achieve low false positive rates with more complex processing. So this kind of cascaded structure can achieve increased detection performance while radically reducing computation time. The detector should have a high detection rate to generate as many ROI as possible. If the pedestrian region is rejected in this stage, it would be judged as non-pedestrian directly, while the false positives can be further classified in the second stage. So the focus of the first stage is to keep high detection rate even with many false positives.

2. EDGELET FEATURES

Improved edgelet features were proposed by Bo Wu and Jie Xu for moving pedestrian detection. Edgelet is a short segment of line or curve, which represents for a little part of human contour. Bo Wu has defined three kinds of edgelet, line, arc and symmetric pair. In order to detect whether position w in an image is similar with above kinds of edgelet features, define $M^I(w)$ and $N^I(w)$ to be the edge intensity and normal at position w. $I(w)$ and $N^I(w)$ are calculated by 3 x 3 sober kernel convolution, and quantize the orientation of the normal vector into six discrete values. Because edgelet features is only used for head-shoulder detection in our method, we select 3 edgelet features shown in Fig.3 for detection, all of which give an accurate description of the head-shoulder part.

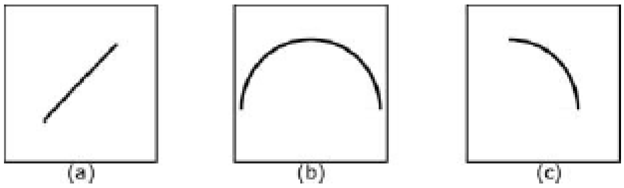

Fig 3: Edgelet Features

3. ADABOOST ALGORITHM

Adaboost is an iterated learning method, a strong learning algorithm upgraded from a group of weak learning algorithms. And the algorithm is implemented by changing the distribution of data, via training set result to determine the correctness of the classification of samples and the last total accuracy of classification, then determine the weight of every single sample, and combine the weak classifiers obtained by every training process together as the final decision classifier. In Adaboost, every training sample is assigned with a weight, which indicates the probability of some kind of weak classifier to be selected into the training set. If one sample is not classified correctly, when construct the next training set, the probability to be chosen will increase. In contrast, the probability will decrease. In this way, Adaboost can focus on those more difficult samples which have more extra information. However, when training classifier by using AdaBoost, it is a great probability to encounter such situation that if one or several samples of the training set is difficult to be classified correctly, after several rounds of updating, these “difficult samples” will be assigned by a high weight. And during the subsequent training process, the selected weak classifier and weight will be the small portion of the total samples, in other words, training samples are simply over fitting on these samples. For the purpose of reducing over fitting, there are two ways to achieve: pick out the high weight difficult samples, it’s easy to implement but also restrict the classification capacity; or adjust the weight of difficult samples, and this way is adopted by this paper.

4. CLASSIFIER USING SHAPELET FEATURE

Sabzmeydani and Mori proposed Shapelet features for pedestrian detection. These features are a set of mid–level features focused on local regions of the image and built from low–level gradient information. The training phase of their algorithm consists of

of three steps: Firstly, for a training sample, the gradient responses in different directions are extracted. Then the local average of these responses around each pixel is used as low-level features to build mid-level features (Shapelet features). Secondly, the training sample is divided into a number of small sub-windows. Then for each sub-window a subset of its low-level features is selected to construct a mid-level Shapelet feature through AdaBoost algorithm. So each Shapelet feature is a combination of gradients with different orientations and strengths at different locations within the sub-window. We can see that the Shapelet features describe local neighborhoods of the image. To combine the information from different parts of the image, the final step is to train the final classifier using Shapelet features through AdaBoost algorithm.

In Shapelet classifier one important parameter is the size of sub-window. Sabzmeydani and Mori define three sets of Shapelet features with different sizes and investigate the influence of Shapelet size on the detector performance. The six types of sub-windows with size from 5×5 to 18×18 are used. In our classifier we use all the six types of sub windows. For each Shapelet feature, we choose $m_i = \sqrt{n_i}$ weak classifiers, where n_i is the number of features inside the sub-window. For example, the 5×5 type has 100(=5×5×4) low level features, so the number of weak classifiers in the Shapelet feature is 10. The six types of sub windows are shown in table 1. For each type of Shapelet feature, we scan the detection window (of size 32×64) with that Shapelet's sub-window size, with strides of 4 pixels between sub-windows.

Table 1. Six types of sub-windows

Size	Number of weak classifiers
5×5	10
8×8	16
10×10	20
12×12	24
15×15	30
18×18	36

5. EXPERIMENTAL DATA

5.1 Software and Hardware Platform

MATLAB has been the ready tool for simulation purpose of various papers. OPEN CV too forms a tool to test the various algorithms. These platforms are generally used in initial stages of projects. Various DSP processors are used in the real time implementation of the Algorithms.

Blackfin 16/32-bit embedded processors offer software flexibility and scalability for convergent applications: multi-format audio, video, voice and image processing, multimode baseband and packet processing, control processing, and real-time security.

DaVinci Digital Video Processors – Optimized for digital video systems, DaVinci digital media processor solutions are tailored for digital audio, video, imaging, and vision applications. The DaVinci platform includes a general purpose processor, video accelerators, an optional DSP, and related peripherals.

5.2 Database

The training dataset used in experiments contains more than 3000 images, most positive samples consist of two well-known pedestrian datasets as shown in Fig. 6 , the MIT dataset and the INRIA dataset with the size of 32 × 64 pixels, and others come from surveillance videos. Also, the head-shoulder training samples are obtained by cutting the image from these two datasets and zoom out to 20 × 20 pixels. Some negative samples are selected from the INRIA dataset, and others are from the Internet. And the experiments are carried out on videos from CAVIAR sequences and campus, which are down sampled to 320 × 240 pixels in our framework.

6. RESULTS

After experimentally testing various images from the MIT and INRIA datasets, following results as shown in Fig.4 and Fig.5 were obtained. Fig.4 shows detection rate for head-shoulder detection for various methods like Haar- like, HOG and edgelet. From the obtained results it can be clearly concluded that Edgelet provides better Detection Rate than the Haar- like and HOG methods. Thus from the obtained results it can be proposed that the performance and accuracy of the detection can be improved by the combination of Edgelet/Shapelet for HSD and Haar-like and Triangular features for FBD. Thus MATLAB can be used as tool for simulation purpose and OPEN CV as a tool to test our proposed method. Finally, our proposed method can be practically tested in the real world by developing an embedded system for it using either Blackfin 16/32-bit embedded processors or DaVinci Digital Video Processor and we expect our proposed method will provide a very high average detection rate and less time for the computation.

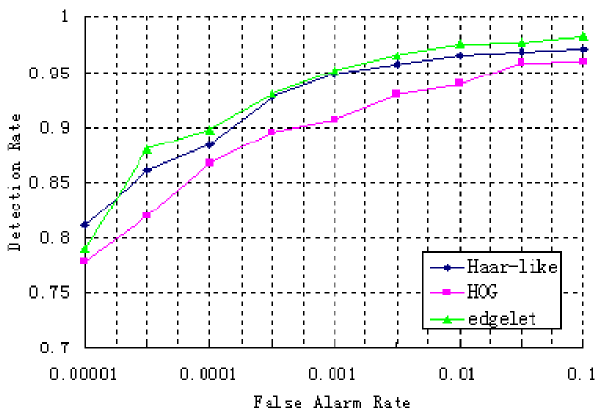

Fig 4 : Result from head-shoulder part detection using Haar-like, HOG, and edgelet

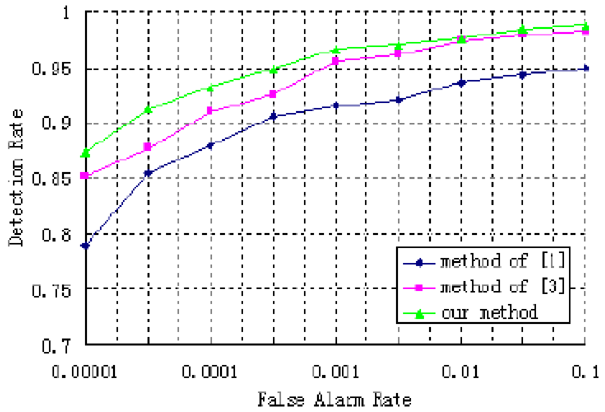

Fig 5 : Comparison with other methods

Fig 6 : MIT dataset and the INRIA dataset used for training and testing purpose

7. CONCLUSION

A number of feature descriptors have been proposed. Previously HOG features were extensively used along with SVM classifier for PD. HOG features and SVM classifier can achieve good performance for PD, but they are time consuming. To achieve high detection speed with good detection performance, a Two-step framework method was proposed. The Two-step framework consists of a full-body detection (FBD) step and a head-shoulder detection (HSD) step. The fusion of Haar-like and HOG features was used earlier in which HSD step utilizes edgelet features for classification and detection. But, this combination gives low detection rate. Also due to use of HOG which is computationally complex, the computation speed obtained was very less. In order to alleviate these limitations we propose here a new methodology for improving the detection rate and speed. The performance and accuracy of the detection can be improved by the combination of Haar-like and Triangular features for FBD and Edgelet/Shapelet for HSD. We expect an average 95% detection rate and 60% faster speed for the proposed method.

8. ACKNOWLEDGMENTS

Our thanks to Dr. R. M. Jalnekar who have given support to this research work.

9. REFERENCES

[1] P. Viola, M. Jones and D. Snow, 2003. Detecting Pedestrians Using Pattern of Motion and Appearance, The 9th ICCV, France, 734-741.

[2] Takuya Kobayashi, Akinori Hidaka, Takio Kurita, 2007. Selection of Histograms of Oriented Gradients Features for Pedestrian Detection, Neural Information Processing: 14th International Conference. Japan, 598-607.

[3] B. Wu and R. Nevatia, 2005. Detection of multiple, Partially occluded humans in a single image by bayesian combination of edgelet part detectors. IEEE International Conference on Computer Vision , China, 90-97.

[4] Supriya Rao,N. C. Pramod, Chaitanya Krishna Paturu, 2008. People Detection in Image and Video Data.Proceeding of the 1st ACM workshop on Vision networks for behavior analysis, Canada, 85-92

[5] J. Janta P. Kumsawat, K. Attakitmaongkol and A. Srikaew, 2007. A Pedestrian Detection System Using Applied Log-Gabor Filters. Proceedings of the 7th WSEAS International Conf. on Signal, Speech & Image Processing, China, 55-60

[6] Navneet Dalal and Bill Triggs, 2005. Histograms of Oriented Gradients for Human Detection. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. USA, Vol 1, 886-893.

[7] MIT Pedestrian Dataset http://cbcl.mit.edu/cbcl/softwaredatasets/PedestrianData.html

[8] INRIA Person Dataset http://pascal.inrialpes.fr/data/human/