Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 321 – 328

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

Vision based pedestrian detection for advanced driver assistance

Vijay Gaikwad[a,*], Shashikant Lokhande[a]

[a]*Sinhgad College of Engineering Research center, Savitribai Phule PuneUniversity, Pune 411 041, Maharashtra, India*

Abstract

In pedestrian detection intricate feature descriptors are used to improve the detection rate at the cost of computational complexity. In this paper, we propose a detector based on simple, robust edgelet features to enhance the detection rate and classifier based on k-means clustering approach to reduce computational complexity. The proposed framework consists of extraction of candidate features of pedestrian detection using edgelet features and use of the cascade structure of k-means clustering for classification enabling high detection accuracy at low false positives. Experimental results show that the proposed method requires less processing time per frame, making it suitable for real-time systems.

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

*Keywords:*K-means classifier; edgelet; pedestrian detection; cascade structure; advanced driver assistance

1. Introduction

Pedestrian detection is one of the most exigent issues in computer vision, and it has diverse practical applications such as advanced driving assistance, autonomous driving, cruise control and surveillance. Over the past decade, the essential role of the computer as well as machine vision modules to realize active safety systems for accident prevention using pedestrian detection is clearly established not only in innovative systems introduced by industry but also in academic research. Pedestrian detection has multiple uses, with the most prominent being advanced driver assistance systems (ADASs). An important facet is to equip vehicles with recognizing capabilities to detect and act on pedestrians in dangerous situations, where the driver would not be able to avoid a collision. A full ADAS with regard to pedestrians would as such not only include detection, but also tracking, orientation, intent analysis, and collision prediction. Pedestrian detection brings many challenges such as changing appearances among pedestrians,

* Corresponding author. Tel.:+91 095 4545 6608.
E-mail address: vijay.gaikwad@vit.edu

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.027

chaotic backgrounds, scenes with moving pedestrians as well as camera, and need of high speed operation and greater reliability that makes detection issues more complex. Many researchers proposed various methods in order to overcome these issues, but still detection accuracy and speed remains on top priority.

For pedestrian detection, Bo Wu[1] introduced edgelet features. These are shorter segments of the line. The detector is designed based on these edgelet features to detect various body parts. The intermediate results from these body part detectors are pooled together to form a likelihood model which enables pedestrian detection in the complex scenes. The results show that this method shows better pedestrian detection as compared to SIFT and SVM. A combination of Adaboost and SVM is used by Wang[2] that shows better accuracy, using a variety of base classifiers. Experimental results show the efficiency of the proposed algorithm and achieve the accuracy up to 94%. M. Bertozzi[3] shows Pedestrian detection using of far-infrared stereo vision. Such methodology shows limitations in high detection accuracy in the presence of poor lighting conditions. It uses temperate area detection, detection by identifying edges, and computational disparity to improve the detection performance. Final detection is carried out using head morphology and thermal characteristics. This method is able to mitigate false negatives caused due to occlusions but still the question of false positives remains unanswered.

Krotosky[4] shows the pedestrian detection with the help of color and infrared approach. The cross spectral approach is highlighted during multimodal and multi-perspective pedestrian analysis. Liping Yu[5] highlights pedestrian detection based on monocular vision for smart vehicles. Pedestrian detection is based on the two-stage framework. A full body pedestrian detector is built using robust Haar-like features. Adaboost classifier is used to discriminate some pedestrian candidates on the image. Experimental results show the robust performance. The pedestrians in different poses, clothing structure, diverse lighting conditions, and occlusion are detected. Fen XU[6] proposed a pedestrian classifier based on the histogram of oriented gradients as well as support vector machine. They showed that to handle occluded pedestrians, the particle filter based method shows high detection rate. The results of particle filter are better than Kalman-filter based tracking method. Yanwei Pang[7] shows pedestrian detection algorithm can be based on HOG and SVM for high detection accuracy but it suffers from time complexity. A´ kos Utasi[8] used a Bayesian approach on finding pedestrians in stereo image based system using the density of cylindrical objects.

The cascade classifier based on Haarlike features is an algorithm that detects pedestrians at a faster rate. A drawback of this approach[9] is the close link with the appearance of pedestrians and the resulting lack of robustness. A substitute is the solution using HOG and support vector machines (SVMs) [10]. This algorithm is much more robust and detects pedestrians in difficult scenes, but compromising computational speed. The hierarchical SVM classifier[11] and the star cascade[12] cannot be applied to general SVM classifiers because they depend on their specialized classifiers. The HOG-based cascade classifier[13] cannot maintain the detection accuracy of its original classifier because it is an entirely new boosting classifier. The HOG-LBP-based cascade classifier[14] and two CoHOG-based cascade classifiers[15, 16] show equivalent detection accuracy to their original classifiers, but their processing capabilities are simply not sufficient. Human detection algorithm and its applications[18, 19] are discussed in defence application and need of high detection accuracy at low computational cost was stated.

Our approach is based on the edgelet feature based pedestrian detection taking the advantages from Bo Wu[1], providing our model robustness to occlusion and pose variation. Although, Wu had better performance with mostly single human, we train our algorithm to work more efficiently for multiple humans in an image. We also propose the novel k-means cascade framework designed to provide higher detection rate than Wu with few false alarms. This cascade framework has multiple stages and each stage has two level to improve detection rate. As we use edgelet features of the leg and torso along with the head-shoulder unlike Wu, the proposed framework detects pedestrians efficiently. Also, some object classes have lots of within-class variation and key point based descriptors have limited performance in this case. In particular, for humans, it is unclear which key point detector to use. All the standard ones return either points in textured regions or blob like structures. In human detection process, texture is seen mainly on clothing so detections are not repeatable across changes of clothing. Variations in pose tend to confuse the blob based detectors. Thus we use the part-based, multi-feature with trained classifier to locate humans.

The rest of this paper is structured as follows. Section 2 describes edgelet features. These features are used for extraction of candidate features in the pedestrian detection process. Section 3 explains k-means clustering approach used in the proposed framework. Section 4 presents the cascading structure of k-means weak classifiers to enhance detection accuracy without increasing computational complexity. Section 5 elaborates experimental results, and finally conclusions are presented in Section 6.

Nomenclature

$A(\bullet)$	affinity between feature set and the test image
f_{ij}	edgelet
$L(i,j)$	labelling function
$E(\bullet)$	energy of the element
k	cluster representing the separate data points

2. Feature descriptor

2.1. Feature set preparation using edgelet

Bo Wu has introduced the edgelet features[1] which are the short segments of human contour. Jie Xu[17] used edgelet features in the context of moving pedestrian detection. An edgelet feature can be defined in the form of a line, arc and symmetric pair. The position w in the edge intensity $M^I(w)$ and normal at position w, $n^I(w)$ an affinity is estimated. It is calculated between the edgelet and the image I at the position w and is given by equation (1).

$$A(w) = (1/\tau)\sum_{i=1}^{\tau} M^I(p_i + w)\left| < n^I(p_i + w), n_i^E > \right| \tag{1}$$

where $\{p_i\}_{i=1}^{\tau}$ is the position of an edgelet and $\{n_i^E\}_{i=1}^{\tau}$ is the vector of the points in a given edgelet, τ is the length of the edgelet. $M^I(w)$ and $n^I(w)$ are calculated by 3×3 Sobel kernel convolution, and are used for the quantization of the normal vectors into the discrete values. In this paper, edgelet feature set is used for detection of head as well as shoulder, torso-the middle body part and leg as shown in Fig. 1. The parameter described in equation (1) maps the intensity of the edge, along with its shape information. Sobel operator is used in the proposed method for edge image extraction because of its low computational complexity. The possible length of one single edgelet is from 2 pixels to 16 pixels. The edgelet features used consist of single edgelet, or parts of the circumference of the circle with variable lengths.

2.2. Learning of edgelet feature set

Various edgelet features are extracted for detection of pedestrians. The feature set is prepared using training images from the INRIA dataset which is publicly available. Multiple edgelet features are extracted at random using a program for various human body parts. Various images with complex scenes, occluded pedestrians, diverse backgrounds, different illuminating conditions are used while preparing the feature set. Extraction of edgelet f_{ij} is carried out for image I by a binary labeling function $L(i,j)$. Total area of the edgelet is composed of edgelet region I^k and a neighborhood region I_r.

$$f_{ij} = \sum_{i=1}^{m}\sum_{j-1}^{n} I^{\tau}(ij, L(i,j)) + \lambda\sum_{i=1}^{m}\sum_{j-1}^{n} I_r(ij, L(i), L(j)) \tag{2}$$

An energy minimization framework is adopted to represent the objectives of the edgelet labeling problem. The size of the edgelet is kept fixed to reduce the computation time required to perform convolution.
For the selection of the edgelet feature the strength of the edgelet is considered.

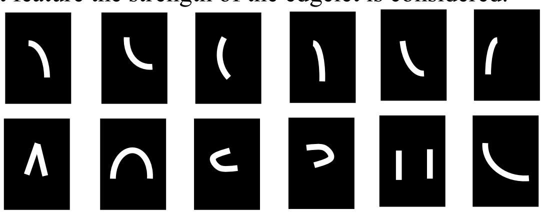


Fig. 1. Sample edgelet features

Thus the total energy of the edgelet represents the two terms:

$$E(ij, L(i,j)) = E^e(i,j) + \nu E^b(i,j) \tag{3}$$

where $E^e(i,j)$ represents the strength of the edgelet and $E^b(i,j)$ represents the energy of the pixels outside of edgelet pixels. Fig. 2 shows the graph for detection of pedestrians by each feature in the feature set. A threshold of 10 is set in the program. Number of features providing detections equal to or more than 10 is observed. It is also taken into account that the feature selected giving fewer false alarms, which is less than 1.1. Fig. 3 presents the steps involved in the learning of feature set. Fig. 3 (a) shows the original image and (b) shows the edges of the original image. In (c) matching region is highlighted when a particular feature detects the pedestrian body part while operating on training images of the data set.

3. Classifier design using k-means clustering

The proposed classifier is based on the k-means clustering, which is a method widely used for quantization of vectors. This method aims at classifying multiple data points or vectors into the few clusters in which each data point belongs to the cluster with the nearest average value which can be considered as a parameter for classification. This results in separation of candidate features of the data point into specific cells or clusters. The proposed framework of "k-means" classifier is designed based on the Euclidean distance as a metric for cluster scatter. Variance as a measure for cluster scatter is taken into consideration.

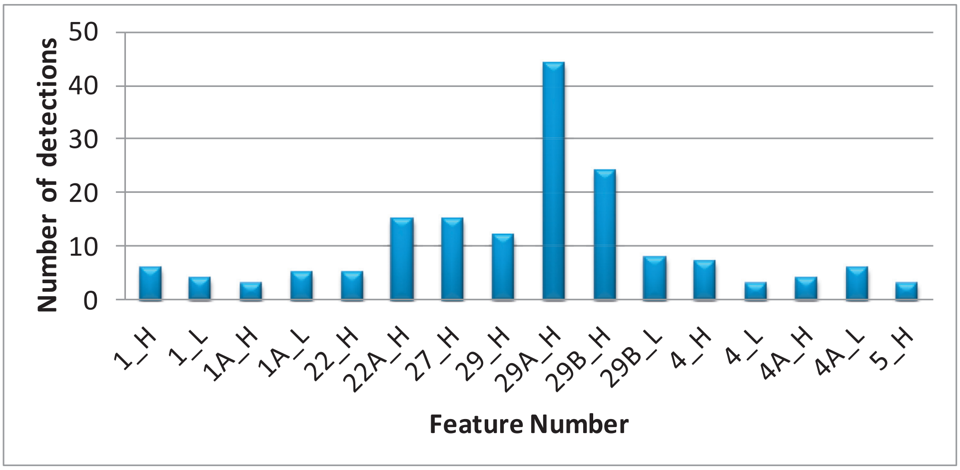


Fig. 2. Number of detections by various edgelet features

(a) (b) (c)

Fig. 3. Learning of the feature set; (a) original image from the INRIA data set; (b) edge map of the original image; (c) selection of edgelet features from the edge map, blue markings show the detection of pedestrian body part by edgelet features.

Multiple iterations are carried out to achieve the end results of pedestrian detection based on collection of multiple candidate data points obtained from the edgelet features. Initial set of k means is assumed as $m_1(1), m_2(1),.....,m_k(1)$. Cluster assignment for each observation is carried out using the least within-cluster sum of squares (WCSS). The sum of squares S is the squared Euclidean distance and is described as

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \le \left\| x_p - m_j^{(t)} \right\|^2 \quad \forall j,\ 1 \le j \le k \right\} \tag{4}$$

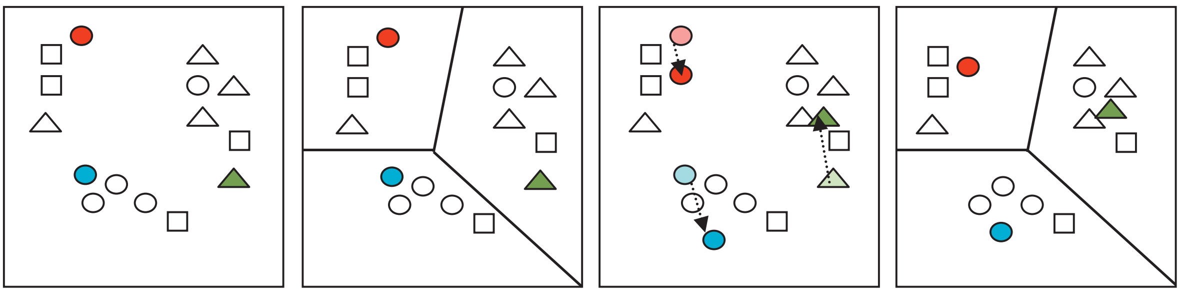
where x_p is the test element, m_i is the mean of the formed cluster, m_j is mean of the j^{th} cluster. In the next iteration, new mean values are calculated as,

$$m_i^{(t+1)} = 1 \Big/ \left| S_i^{(t)} \right| \sum_{x_j \in S_i^{(t)}} x_j \tag{5}$$

This process is executed in an iterative manner, leading to separation of data points or candidate features by reducing the sum for all the available clusters of the intra-cluster addition of point-to-cluster-centroid distances. The output of the multiple iterations results in the formation of various clusters as shown in Fig. 4.

4. Cascading structure of the classifier

The matching function is estimated based on the edgelet feature descriptors. If this matching function is above the threshold limit, then correspondent candidate features are classified with the help of k-means classier at stage one. Upon completing multiple iterations, score function is passed to the second stage of the cascaded structure that computes the sum of all the distances in the given cluster and maps it to the pre-defined values. The existing clusters are reformed into the new set of clusters based on the estimated scoring function. Various values of the score function, classifies the candidate features of the pedestrians into multiple clusters as shown in Fig. 5. Cascading structure implementation using k-means clustering approach is illustrated in Fig. 6. In the initial and intermediate phase of the classification process, candidate features are acquired and shown in Fig. 6 (a). After convergence is completed final clusters are formed with new centroids and the result in shown in Fig. 6 (b). Depending upon the matching core, either of k-means classifier stage is selected to form the new cluster. Various thresholds are set up for the matching scores during training phase of the classifier. INRIA dataset is used to train the classifier. Various thresholds are designed in such a way that single pedestrian as well as occlusions are easily detected.



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

Fig. 4. Design of the classifier; (a) initial distribution of data points; (b) formation of *k* clusters by based on the nearest distance; (c) Formation of new centroids after multiple iterations, dotted arrow indicates change of cluster position from previous location to the new location; (d) formation of final set of clusters with new centroids shown by filled objects after the convergence .

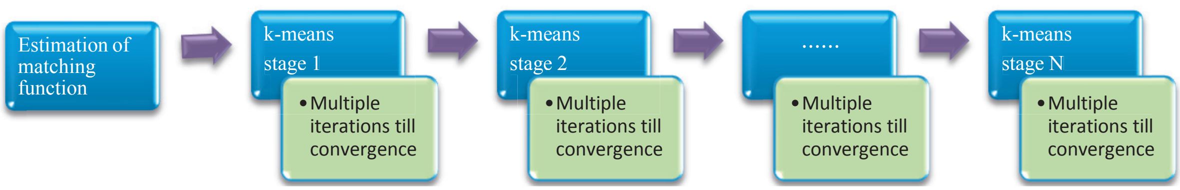
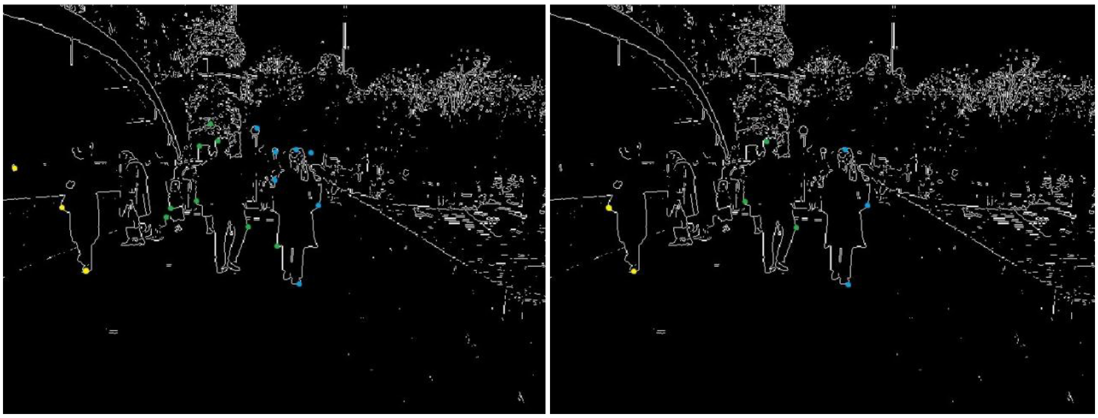


Fig. 5. The cascade structure of the weak classifiers using *k*-means clustering approach.



(a) (b)

Fig. 6. Results of the classifier; (a) extraction of the initial candidate features shown by circles filled with different colour; (b) result of the complete cascaded framework

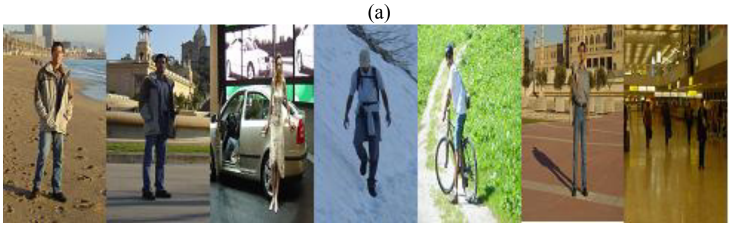
These thresholds are tested using training images of the INRIA dataset and successful results are obtained. Each stage of the classifier has different set of thresholds depending upon the number of pedestrians to be identified.

5. Experimental results

In the proposed framework for pedestrian detection, the INRIA image data set is used as shown in Fig. 7. The training images of the data set are used to learn the classifier and test images are used to find the detection accuracy of the proposed algorithm. The training data set is resized to 128 x 96 for reducing computational time. The training data set is divided into 3 sets of images based on the depth of humans in the given image – near, medium and far distance human objects from the camera. The size of the bounding box for the pedestrian object is set to the average size of all the features which detects the human objects in the given image.

Fig. 8 shows the results for pedestrian detection using INRIA data set. The detection accuracy achieved using proposed framework is 95% at false alarm rate of 1.1. The average computational time per frame is less than 50ms; this allows processing of maximum 20 frames per second. Use of robust edgelet feature set ensures the false alarm rate is at its minimum without compromising the detection accuracy. This indicates that the proposed system shows suitability for real-time implementation. The performance of the proposed vision based pedestrian detection system

(a)



(b)

Fig. 7. INRIA dataset; (a) training and (b) testing sample data set

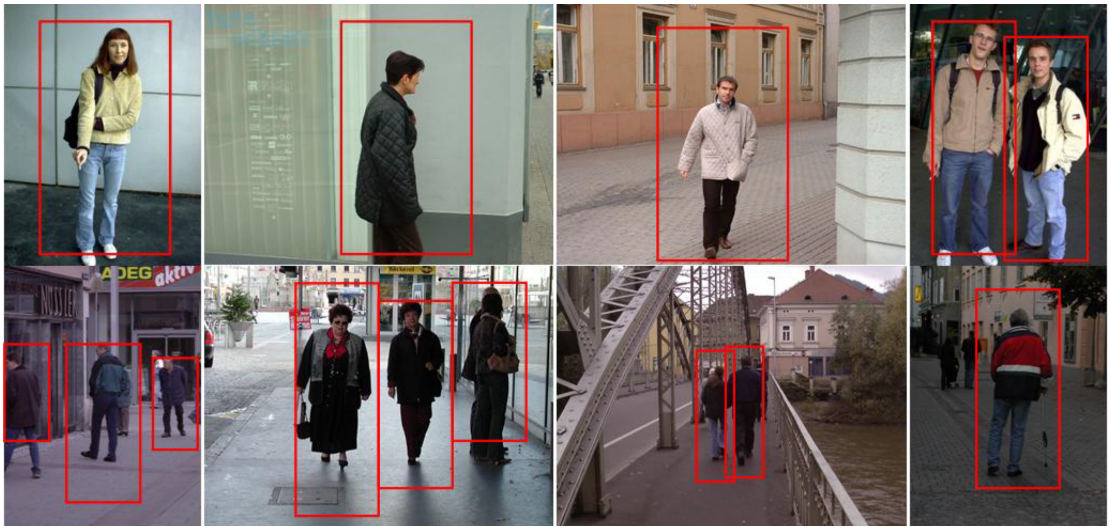


Fig. 8. Results of the pedestrian detections using INRIA dataset

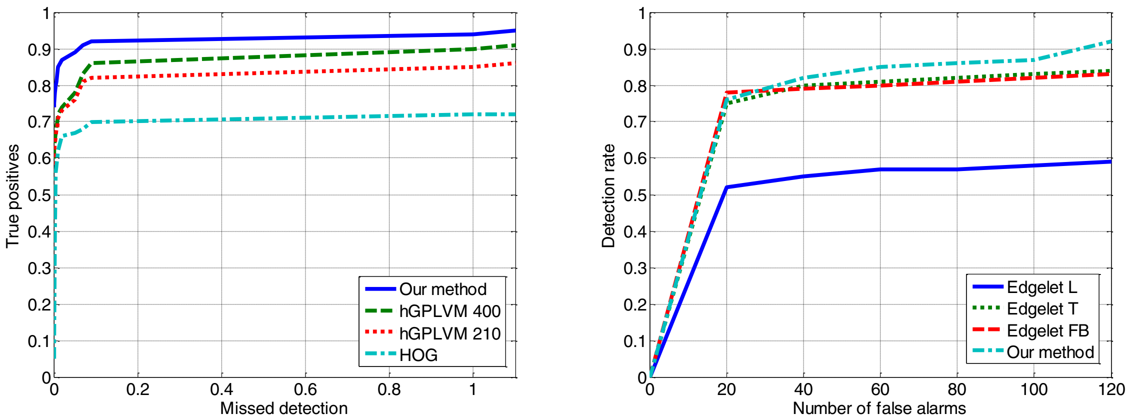


Fig. 9. ROC curves comparison with state-of-the-art detectors

is examined with the help of ROC curves, i.e. true positive rate versus missed detection rate as well as detection rate versus number of false alarms as shown in Fig. 9 and compared with the recent techniques in [20, 1]. It is observed that the proposed method shows better performance than the state of the arts due to the use of robust edgelet based pedestrian detection and the use of k-means classification approach.

6. Conclusion

Vision based pedestrian detection system is proposed. High detection accuracy and low computational time per frame are the essential elements of the pedestrian detection system. Use of complex feature descriptors may enhance the detection accuracy, but its processing speed becomes an important issue. In this paper, efficient framework based on edgelet feature set and k-means classifier is introduced for pedestrian detection. Using new set of edgelet features used, the detection accuracy is improved while k-means based clustering approach used for classification, reduces the computational time per frame. Experimental results show that the proposed framework for pedestrian detection is well-suited for real-time applications due to its high detection rate and minimum computational complexity.

References

1. Bo Wu, Ram Nevatia. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of edgelet Part Detectors. *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)* 2005.
2. Xiaodan Wang, Chongming Wu, Chunying Zheng, Wei Wang. Improved Algorithm for Adaboost with SVM Base Classifiers. *Proc. 5th IEEE Int. Conf. on Cognitive Informatics (ICCI'06)* 2006; 948-952.
3. M. Bertozzi, A. Broggi, C. Caraffi, M. Del Rose, M. Felisa, G. Vezzoni. Pedestrian detection by means of far-infrared stereo vision. *Elsevier Computer Vision and Image Understanding* 2007; **106**:194–204.
4. Stephen J. Krotosky, Mohan Manubhai Trivedi. On Color-, Infrared-, and Multimodal-Stereo Approaches to Pedestrian Detection. *IEEE transactions on Intelligent Transportation Systems* 2007; **8-4**: 619-630.
5. Liping Yu, Wentao Yao, Huaping Liu, Fasheng Liu. A Monocular Vision Based Pedestrian Detection System for Intelligent Vehicles. *IEEE Intelligent Vehicles Symposium* 2008; 524-529.
6. Fen XU, Ming Gao. Human Detection and Tracking based on HOG and Particle Filter. *3rd International Congress on Image and Signal Processing (CISP2010)* 2010; 1503-1507.
7. Yanwei Pang, YuanYuan, XuelongLi, JingPan. Efficient HOG human detection. *Elsevier-Signal Processing* 2011; **91**:773–781.
8. A´ kos Utasi, Csaba Benedek. A Bayesian Approach on People Localization in Multicamera Systems. *IEEE transactions on circuits and systems for video technology* 2013; **23-1**:105-115.
9. P. Viola, M. J. Jones, D. Snow. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision* 2005; **63-2**: 153–161
10. N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. *Proc. IEEE Comput. Soc. Conf. CVPR* 2005; **1**: 886–893.
11. B. Heisele, T. Serre, S. Prentice, T. Poggio. Hierarchical classification and feature reduction for fast face detection with support vector machines. *Pattern Recognit* 2003; **36**:2000–2017.
12. P. Felzenszwalb, R. Girshick, D. McAllester. Cascade object detection with deformable part models. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit* 2010; 2241–2248.
13. Q. Zhu, M. Yeh, K. Cheng, S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit* 2006; **2**: 1491–1498.
14. M. Hiromoto, R. Miyamoto. Cascade classifier using divided CoHOG features for rapid pedestrian detection. *Proc. Int. Conf. Comput. Vis. Syst* 2009; 53–62.
15. H. Ren, C. K. Heng, W. Zheng, L. Liang, X. Chen. Fast object detection using boosted co-occurrence histograms of oriented gradients. *Proc. 17th IEEE Int. Conf. Image Process* 2010; 2705–2708.
16. C. Zeng, H. Ma, A. Ming. Fast human detection using mi-sVM and a cascade of HOG-LBP features. *Proc. 17th IEEE Int. Conf. Image Process* 2010; 3845–3848.
17. Jie XU, Getian Ye, Gunawan Herman, Bang Zhang. An Efficient Approach to Detecting Pedestrians in Video. *Proceeding of the 16th ACM international conference on Multimedia* 2008;789-792.
18. Vijay Gaikwad, Akshara Jorigal, Deepali Sanap. Human Detection and Tracking: Algorithms, Systems and Future scope. *3rd International Conference on Competency Building Strategies in Business and Technology for Sustainable Development* 2014.
19. Vijay Gaikwad, Pranoti Joshi, Jayshree Waykole. Human Detection, Tracking and Its Applications in Defense. *3rd International Conference on Competency Building Strategies in Business and Technology for Sustainable Development* 2014.
20. Mykhaylo Andriluka, Stefan Roth, Bernt Schiele. People-Tracking-by-Detection and People-Detection-by-Tracking. *Proc. of IEEE* 978-1-4244-2243-2/08 2008.