Fast Pedestrian Detection using Smart ROI separation and Integral image based Feature Extraction

Bineesh T.R and Philomina Simon
Department of Computer Science, University of Kerala
Thiruvananthapuram, Kerala, India
bineeshtr@gmail.com
philomina.simon@gmail.com

***Abstract* – This paper discusses a fast pedestrian detection system for near infrared imaging system. The Advanced Driver Assistance Systems include pedestrian detection system to avoid accidents. Most pedestrian detection systems produce false alarms or they are not fast. To overcome these issues a new approach for pedestrian detection is presented here. Initially the foreground is segmented by a smart region detection method to generate candidates. Then a series of rejecters are integrated to filter out non-pedestrians. After filtering out typical non-pedestrian objects, the remaining number of region of interest (ROI) is verified using a Support Vector Machine (SVM) classifier with Histogram of Oriented Gradients (HOG) feature. A second level classification is performed with HAAR feature to reduce the False alarms. The integral image representation is used for extracting both features, which significantly improves the computation speed. Experimental result shows that the proposed pedestrian detection system is suitable in the real-time environment, as it gives high detection rate and very low false alarm rate.**

Keywords – **Pedestrian Detection, Near Infrared Camera, Histogram of Oriented Gradient, HAAR wavelet, Integral Image, Support Vector Machine, Smart ROI Detection, Feature Extraction.**

I. INTRODUCTION

In any traffic accident, the probabilities for severe injuries are higher in the case of pedestrians travelling at night. During night, inadequate illumination is the main reason for high fatality rate. Indiscriminant use of high-beam head lamps can dazzle drivers in approaching vehicles as well as blind pedestrians. Analysis shows that [2] nearly 70% of pedestrian fatalities happen at night, while nighttime driving is only 20% of the total traffic. In order to overcome these difficulties, the automotive industry is continuously trying to develop safety mechanisms in vehicles, initially they developed seat belts, airbags and Anti-lock Breaking System (ABS). Only a vision enhancement system can help drivers to see lane markings, roadside reflectors, pedestrians, signs, bicyclist, etc. Pedestrian Detection at night time for Advanced Driver Assistance Systems (ADAS) aims to warn the driver in advance about the risky situations or initiate some appropriate warning mechanisms in time. Now the researchers are concentrating more on on-board devices such as Pedestrian Detection Systems.

Detecting pedestrians at night time on a moving platform is challenging due to insufficient illumination, movement of camera (changing background), different human appearances and poses, or strict performance criteria and hard real-time constraints. In this paper, a real-time pedestrian detection system from video images captured from monocular near infrared camera during night time driving. The system is basically designed to eliminate the false alarms and to reduce the detection time.

The rest of the paper is organized as follows. Literature review about pedestrian detection is discussed in Section II; Section III focuses on the proposed system. Section IV discusses the experimental results and analysis. Conclusion and future scope is described in Section V.

II. RELATED WORKS

Generally a vision based pedestrian detection can be divided in to 3 modules: Image acquisition, ROI generation and Object classification. To detect humans at the night time, some Pedestrian Detection systems use the Far Infra Red (FIR) camera to detect the “bright spots” [3], objects with high temperature; or some systems use Near Infra Red (NIR) camera to detect warm areas. Since the price of NIR camera is lower, usually NIR camera is chosen for image acquisition. Also, it holds better signal to noise ratio and good resolution. A night vision system which utilizes infrared light for image acquisition does not dazzle road users. Thresholding based methods [2], Detection based methods [5, 6], and Adaptive thresholding methods [2][4] are commonly used methods for segmenting foreground objects. In [4], the algorithm cannot deal with situations like pedestrians walk shoulder to shoulder or pedestrians staying close to other obstacles like road signs or trees. A dual threshold method is used in [7], also fail in some situations like identical pixels at pedestrian area and adjacent background area. For NIR camera, Tian et al. [8] extract the target regions used adaptive threshold based segmentation method. In this work, a detection based segmentation method called smart region detector [1] is chosen to extract candidate region. This method overcomes the limitations of the existing methods; a series of rejecters is designed for candidate selection after the segmentation process. The candidate selection method used in Bertozzi et al. [9] is entirely different, which uses stereo FIR vision technique. The candidate selection method proposed by Alonso et al. [10] is based on the direct computation of the 3-D coordinates of relevant points in the image. Once the ROI is generation step is completed, the next phase is object classification. The performance of classification is purely depends on the extracted feature and the classifier used. The common features extracted for pedestrian detection are the shape, appearance and motion. Dalal and Triggs uses Histogram of Oriented Gradient (HOG) [1][12] feature for classification while P. Alonso et al. [10] discussed the Haar wavelet feature. Y. C. Lin et al. selected Contour based feature [1] because the edge based feature is not sensitive to contrast changing and texture of image. In [13], the pedestrian is segmented into 3 parts. As a prior knowledge for object detection in [14], they propose a generic shape model named Boundary Fragment Model (BFM). In the proposed method a faster algorithm for pedestrian detection is implemented with the help of integral image [15] for feature extraction; integral HAAR feature and integral HOG [16] features are extracted. The next step in object classification is classification based on the extracted feature. The typical classification techniques are template matching [17], Artificial Neural network (ANN) [18], Support Vector Machines (SVM) [10][3][11][12] and Adaboost [19],[20]. Obviously SVM is the most popular learning method which produces more accurate classification results. Existing object detection methods produce more false alarms and requires more time for execution. Here, in the proposed method we have implemented a two level classification scheme based on HOG and HAAR feature, which eliminates the false alarms to a greater extend. A prominent candidate selection method is implemented and integral image based feature extraction also results in very fast detection of pedestrians.

III. Proposed Pedestrian Detection System

The proposed system consists of two modules, Region of Interest (ROI) generation and Object classification. Fig1 shows the block diagram of the proposed system. The input videos are captured by a near infrared (NIR) camera fixed behind the rear view mirror of the car with illumination from full beam headlight. The ROI generation module consist two sub modules: Image segmentation and Candidate selection. The image segmentation is done based on the assumption that the foreground objects are regions usually brighter than the background in the image. This segmentation method adopts vertical projection and horizontal projection at the same time. Local thresholding method is applied to convert the input image into a binary image. This method works efficiently in non-uniform illumination conditions.

Connected component analysis is performed in the candidate selection step. To improve the speed and reduce the false positives, only a selected number of candidate windows are generated in ROI generation module. Based on the size, shape, position of the connected component, the candidates are selected.

In the Object classification module, there are two sub modules: feature extraction and classification. In real-time application, the system should execute in a high speed, so less time should be taken for feature extraction. In order to improve the speed of feature extraction, Integral image [21] representation is used. To ensure that the detected objects are pedestrians, a two level filtering approach is applied to filter out the non-pedestrians from the samples. In the first level filtering, Integral HOG feature is extracted and given to SVM classifier. In the second level filtering, Integral HAAR feature is extracted. The result obtained after the filtering contains only the true positive objects. This system overcomes the limitations of some existing systems such as more false alarm rate and less execution speed.

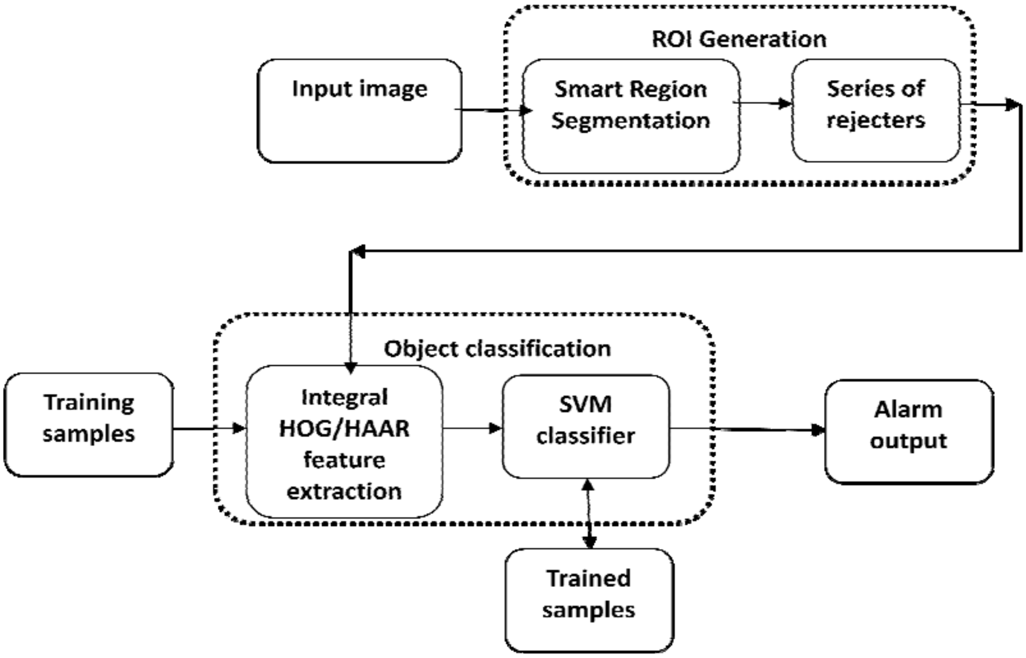


Fig1. Block diagram of the proposed system

The algorithm of the proposed method is given as follows:

(1) Extract the first frame from the NIR video.
(2) A smart region segmentation method is applied for input frame. Based on the local statistics of each segment, convert the image into a binary image.
(3) A series of rejecter filter is applied to all the connected components (or ROI) in the binary image to eliminate the non-pedestrians.
(4) Extract the integral HOG feature for each of the remaining connected components, and feed to SVM classifier.
(5) SVM determines whether the features are of pedestrians or not. If it detected as pedestrians, then go for next step else reject the corresponding connected component (or mark as non-pedestrians).
(6) If one connected component is detected as pedestrians from the first level classifier (step 5), then extract the integral HAAR feature from the same connected component. These features are then fed to SVM classifier.
(7) If again the SVM detects pedestrians, the system should generate a warning output.

These steps must be repeated for all the frames of the input video. The system will generate warning with minimum false alarms.

A. Smart Region detection

A robust local threshold approach in [22] is applied for foreground segmentation in night time. This method considers both vertical and horizontal projection at the same time. This segmentation is based on the assumption that foreground objects are usually brighter than background. An assumption is made that all the pedestrians are vertical in images. The steps for smart region segmentation [1] are:

Let *f*(x, y) is the original gray level image with size M by N, where (x, y) represents the coordinates of each pixel.

First calculate the vertical projection by

$$V(x) = \sum_{y=0}^{N-1} f(x,y) \qquad (1)$$

Here V(x) represents the vertical histogram.

In the gray level pedestrian image pixel values must be brighter than background pixel values. Thus the peaks and valleys of vertical histogram represent the pedestrian and non-pedestrians. The contrast of the night image is very poor and it may be affected by noise from the surroundings. A smoothing filter is applied to avoid these noises, thus we get $V_s(x)$ as smoothened vertical histogram.

$$V_s(x) = \frac{1}{n} \sum_{k=x-\left\lfloor\frac{n}{2}\right\rfloor}^{x+\left\lceil\frac{n}{2}\right\rceil-1} V(k) \qquad (2)$$

where *n* denotes the window size.

The next step is to find the boundary between pedestrian and non-pedestrian. Hence the distance between two boundary points is the width of detected block. The boundary point can be calculated by the maximum values in the first order derivative of $V_s(x)$.

$$V_s'(x) = \frac{dV_s(x)}{dx} = \frac{V_s(x+1) - V_s(x)}{(x+1) - x} = V_s(x+1) - V_s(x) \qquad (3)$$

The first order derivatives at some points are zero. Select those points (*X*) where the first order derivatives are zero.

$$X = \{X_j \mid 0 < j < m\} \qquad (4)$$

Here X_j values are the points where first order derivatives are zero. The next step is to take two points X_j and X_{j+1} , and select the maximum V_s' in between X_j and X_{j+1}, which will be a boundary point, P_j .

$$P_j = \max_{X_j \le x \le X_{j+1}} \{|V_s'|\} \qquad (5)$$

Then all boundary points are derived, the next step is to find the width of the detected block, which can be calculated by taking the difference of two consecutive boundary points.

$$D_j = P_{j+1} - P_j \qquad (6)$$

The distance D_j represents the width of pedestrian or background. Based on D_j, the input image is partitioned into *m*-1 vertical stripes from R_1 to R_{m-1}. Each vertical sub region consists of D_j by N pixels. These regions are calculated from the vertical histogram.

The next step is to divide each of these vertical stripes into different set of blocks. The horizontal histograms are calculated for each of these vertical sub regions and divide each vertical stripe into different blocks. Then apply the above steps again. i.e after finding horizontal histogram, apply smoothening technique, find the first order derivative, then find the border points. Finally divide each sub region into a set of blocks. Thus each block consists of $D_j \times D_i$ pixels. After image segmentation is performed, the binary image is obtained by thresholding the amount of blocks. Local mean μ and local standard deviation *s* is calculated for each block, to determine whether the adaptive threshold has to be evaluated for each block.

Fig. 3a.Input image

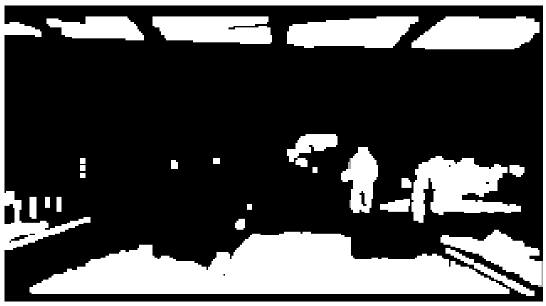


Fig. 3b. Segmented image

B. Candidate Selection

Even though the smart region detection gives better segmentation, it also produces a large number of non-pedestrians. There may be chances to segment other background regions such as noisy pixels, light sources, objects with high reflectance, traffic sign, head light, electric pole, other sign boards etc. These Non-pedestrians should be filtered out for a better classification. The prior knowledge of the scene information helps in filtering. Most of the non-pedestrians can be filtered out based on the size, shape characteristics of the image, such as width, height and aspect ratio. The width and height limits at various distances. The saturated objects are filtered out when the boundary box of the Region of Interest (ROI) meets the condition

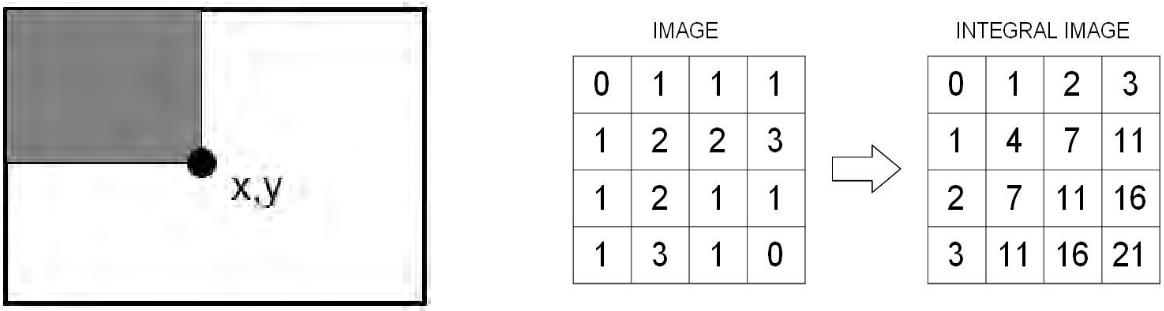
N_{sat} / (W × H) < *T* where W, H is the width and height of the ROI and N_{sat} is the number of saturated pixels in the ROI. $W < W_{max}$ *This* condition was introduced to avoid elimination a close pedestrian having high reflection. W_{max} can be defined as the width of the boundary box of a pedestrian at distance of 7m. Fig4. shows the output of the candidate selection step, in Fig.4(a) only the pedestrian area is successfully marked with a rectangle. Fig 4(b) is a pedestrian walking in the side of a road, here two areas are marked as region of interest, but only one is correct. The other marked region is a metallic reflection. In order to correctly classify such situations, two levels of SVM classification are adopted.

(a) (b)

Fig. 4 The output of ROI generation

C. Integral Image Representation

Using the integral image representation the value of any rectangular sum can be computed in constant time. The integral image at location (x, y) is the sum of the pixels above and to the left of (x, y), inclusive. The figure 5 shows the integral image representation.



(a) (b)

Fig. 5 Integral Image representation

Integral Image (x, y) = Sum of the values in the gray region.

Figure 5(b) shows a simple example for an integral image representation of a 4×4 image. Hence, the feature extraction takes only very less amount of time.

D Feature Extraction

Feature extraction is the first step in object classification. The simplest feature for any object classification is the image intensity, which can be directly taken from the input image. But this feature repeat several times in the same image. The gradient magnitude can be extracted from the intensity values, hence shape information is very sensitive to noise, it is not widely used for accurate classification. Since the classifier performance strongly depends on extracted features, the proposed method is implemented using two level classifications with Integral Histograms of oriented gradients (HOG) and Integral Haar feature extraction.

1). Integral HOG

In the proposed approach, the first level classification is performed using Integral HOG. Local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. Dalal and Trigs [11] and Zhu et al. [16] utilizes the use of histogram with a dense scan approach. HOG feature is first used for pedestrian detection by Shashua et al. [23]. Dalal and Triggs divide each detection window of 64 × 128 into cells of 8×8 pixels and a group of 2×2 cell is selected as a block with an overlap of one cell in both horizontal and vertical directions. A 9 bin histogram is calculated for each cell, thus we get a 36-D feature vector for each block. L2 normalization is used for all the 7×15 blocks. Integral histogram [21] suggested by Porikli [24] compute the histograms very efficiently.

The HOG feature is computed in a similar way but the ROI size is resized to 24×60. The cell size is taken as 4×4 pixel, thus a total of 6×15 cells. In the first step, the horizontal and vertical gradient of the image is computed by sobel filter. If the x direction sobel derivative is zero for a pixel, a small value is added to it, to avoid division by zero. The magnitude and orientation of the gradient is further computed. So we discretize each pixels orientation including its magnitude into 9 histogram bins. The integral image is computed and stored for each 9 bin of the HOG. The 9 bins are taken over 0^o - 180^o. 90 is added to each orientation, to shift the orientation values range from {-90 - 90} to {0 - 180}. This is just a matter of convention. {-90 - 90} values can also be used for the calculation. An array of 9 images (9 because the bin size 20 degrees and unsigned gradient (180/20 = 9) is created, one for each bin which have zeroes for all pixels, except for the pixels in the original image for which the gradient values correspond to the particular bin. These are referred to as bin images. These bin images are then used to calculate the integral histogram, which quicken the calculation of HOG descriptors.

The size of one block is selected as 2×2 cell. The length of the feature vector for a cell is 9 (since no. of bins is 9), for block it would be 9×(no. of cells in the block) = 9×4 = 36. The length of the feature vector for a window would be 36×(no. of blocks in the window. Dalal and Triggs use an L2 normalization step for each block. In the proposed method, L2 normalization is replaced with L1 normalization for faster computing. A 24×60 image consists of 5×14 overlapped blocks, each block consist of 36-D block descriptor, so finally a 2520-D HOG descriptor is obtained as feature vector.

2) Integral Haar-like feature

For a second level classification, Haar features are used. It improves the efficiency of the system with minimum false alarms. A fast algorithm is implemented with the help of integral image concept. 5 types of haar like features are used for feature extraction which is shown in Fig. 6. It contains horizontal, vertical, diagonal and edge features in [21, 25**]**. We apply the Haar feature extraction step to all the ROI generated. Here also the region of interest is resized to 24×60 and grouped 4×4 pixels as a block.

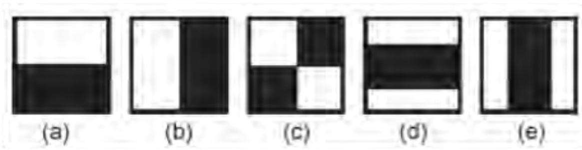


Fig. 6 The 5 Haar like features we used. (a) Horizontal edge feature (b) Vertical Edge Feature (c) Diagonal Feature (d) Horizontal-line feature (e) Vertical-line feature.

The value of a HAAR-like feature is the difference between the sum of the pixel gray level values within the black & white rectangular regions.

Value = ∑ *(pixels in white area)* - ∑ *(pixels in black area)*

First, the integral image is computed for the resized ROI and the Haar feature is calculated from it. A 24×60 image consists 6×15 blocks and for each block 5 features are extracted. So the final HAAR feature (450-D) is the concatenation of the entire five HAAR-like features for all blocks. *i.e.* 5× (6×15) = 450.

E. SVM classification

The success of any classifier is purely depends on the features we extract. The HOG and HAAR features are extracted in a very faster manner using integral image concept. Before classification, images are trained using HOG and HAAR features. A training dataset for both pedestrian and non-pedestrian is required. SVM constructs a hyper plane or set of hyper planes in a high or infinite dimensional space, which can be used for classification. Choose the hyper plane so that the distance from it to the nearest data point on each side is maximized. A good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin).In general, the larger the margin, the lower the generalization error of the classifier. Support vectors are the data points that lie closest to the decision surface.

IV. RESULTS AND DISCUSSION

Fig. 7 Sample dataset

The proposed system is implemented in OpenCV for faster execution. In OpenCV, the trained feature set is stored as an *.xml file*, which is loaded at the time of classification. To evaluate the performance of the proposed system, a training system is collected which consists of 5200 samples with size 24×60. Among these, 2040 are pedestrian samples and rest are non-pedestrian samples. Since there is no public database is available, a dataset is created from different NIR videos and using NIR camera. Fig. 7 shows some sample dataset for both pedestrian and non pedestrians. Fig.8 shows the classifier output of various kinds. Table 1 shows the processing time of proposed algorithm under 320×240 image size. Since the classification is faster, the number of candidates alone slightly affect the processing time. While testing the videos, our system runs at 12~25 frames per second, with a recall of 91.69% and 97.45% precision.

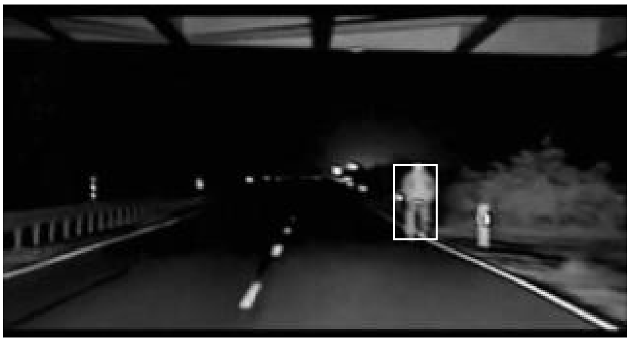


Fig.8 (a) Cyclist

(b) Walking pedestrian

Process	**Processing Time**
Segmentation / frame	35ms
HAAR classification/window	0.3ms
HOG classification/window	4ms
Entire Detection Algorithm/frame	39.3ms ~ 82.3ms (25~12 frames/sec)

Table 1. Time required evaluating a 320×240 image.

Video Name	Detected	Missed	False alarm	Detection rate %
Video1	63	6	2	91.3
Video2	46	4	2	92.0
Video3	33	3	1	91.67
Video4	56	5	0	91.8

Table 2. Performance and analysis on different video sequence of proposed system, a smart ROI generation with two levels of SVM classification with HOG and HAAR features.

SVM Classifier	Recall %	Precision %
HOG	80.12	93.87
HOG and Contour[1]	90.57	95.13
Proposed method	91.69	97.45

Table3. Comparison of Recall and Precision with other existing methods.

V. CONCLUSION

A novel approach to near infrared based pedestrian detection system is proposed with a series of rejecters and smart region segmentation is added to filter out the non-pedestrians. Integral HOG and HAAR features are computed for reducing the false alarms and improving the execution time. The proposed system satisfies the real time constraints, since it gives warning at proper time. Future scope of this work is to develop a candidate selection step, which itself gives the best classification and made the system available for public usage with low cost.

REFERENCES

[1] Yu-Chun Lin, Yi-Ming Chan, Luo-Chieh Chuang, Li-Chen Fu, Hsiao and Min-Fang Luo, "Near-Infrared Based Pedestrian Detection by Combining Multiple Features", *IEEE Conference on Intelligent Transportation Systems* , pp.*1549-1554, 2011.*

[2] J. Ge, Y. Luo, and G. Tei, "Real-Time Pedestrian Detection and Tracking at Nighttime for Driver-Assistance Systems," *IEEE Transactions on Intelligent Transportation Systems,* vol. 10, pp.283-298, 2009.

[3] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 63–71, Mar. 2005.

[4] Q. M. Tian, Y. P. Luo, and D. C. Hu, "Pedestrian detection in nighttime driving," in *International Conference on Image and Graphics*, pp. 116-119, 2004.

[5] M. Soga, S. Hiratsuka, H. Fukamachi, and Y. Ninomiya, "Pedestrian Detection for a Near Infrared maging System," in *IEEE Conference on Intelligent Transportation Systems*, pp. 1167-1172, 2008.

[6] M. Bertozzi, A. Broggi, A. Fascioli, T. Graf, and M. M. Meinecke, "Pedestrian detection for driver assistance using multiresolution infrared vision," *IEEE Transactions on Vehicular Technology,* vol. 53,
pp. 1666-1678, 2004.

[7] J. Dong, J. Ge, and Y. Luo, "Nighttime Pedestrian Detection with Near Infrared using Cascaded Classifiers," in *IEEE International Conference on Image Processing*, pp. VI - 185-VI - 188, 2007.

[8] Q. Tian, H. Sun, Y. Luo, and D. Hu, "Nighttime pedestrian detection with a normal camera using SVM classifier," in *Proc. 2nd. ISNN*, May 2005, vol. 3497, pp. 189–194.

[9] M. Bertozzi, A. Broggi, C. Caraffi, M. Del Rose, M. Felisa, and G. Vezzoni, "Pedestrian detection by means of far-infrared stereo vision," *Comput. Vis. Image Underst.*, vol. 106, no. 2/3, pp. 194–204, May 2007.

[10] P. Alonso, I. Llorca, D. Sotelo, M. Bergasa, L. de Toro, P. Nuevo, J. Ocana, and M. Garrido, "Combination of feature extraction methods for SVM pedestrian detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 292–307, Jun. 2007.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2005, pp. 886–893.

[12] M. Bertozzi, A. Broggi, M. Rose, M. Felisa, A. Rakotomamonjy, and F. Suard, "A pedestrian detector using histograms of oriented gradients and a support vector machine classifier," in *Proc. IEEE Conf. Intell. Transp. Syst.,* Sep. 2007, pp. 143–148.

[13] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors," in *IEEE International Conference on Computer Vision*, pp.90-97, 2005.

[14] A. Opelt, A. Pinz, and A. Zisserman, "Learning an Alphabet of Shape and Appearance for Multi-Class Object Detection," *International Journal on Computer Vision,* vol. 80, pp. 16-44, 2008.

[15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Dec. 2001, pp. 511–518.

[16] Q. Zhu, C. Yeh, T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2006, pp. 1491–1498.

[17] D. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *Int. J. Comput. Vis.*, vol. 73, no. 1, pp. 41–59, Jun. 2007.

[18] L. Zhao and C. Thorpe, "Stereo-and neural network-based pedestrian detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 3, pp. 148–154, Sep. 2000.

[19] X. Cao, H. Qiao, and J. Keane, “A low-cost pedestrian-detection system with a single optical camera,” *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 58–67, Mar. 2008.
[20] P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” *Int. J. Comput. Vis.*, vol. 63, no. 2, pp. 153–161,Jul. 2005.
[21] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Dec. 2001, pp. 511–518.
[22] Y.-T. Pai, Y.-F. Chang, and S.-J. Ruan, "Adaptive thresholding algorithm: Efficient computation technique based on intelligent block detection for degraded document images," *Pattern Recognition*, vol. 43, pp. 3177-3187, 2010.
[23] A. Shashua, Y. Gdalyahu, and G. Hayun, “Pedestrian detection for driving assistance systems: Single-frame classification and system level performance,” in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2004, pp. 1–6.
[24] F.Porikli, “Integral histogram: A fast way to extract histograms in Cartesian space”. *Conference of Computer Vision and Pattern Recognition (CVPR)*,2005.
[25] C. Papageorgiou, M. Oren, and T. Poggio, “A general framework for object detection,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Jan. 1998, pp. 555–562.