Australian Journal of Basic and Applied Sciences, 9(2) February 2015, Pages: 216-223

AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

Detection of Pedestrian from Moving Object Using Histogram of Gradient and Optical Flow Method

[1]K. Thulasimani, [2]Dr. K.G. Srinivasagan, [1]T.K. Renugha

[1]Department of Computer Science and Engineering, Government College of Engineering, Tirunelveli, Tamil Nadu India
[2]Department of Computer Science and Engineering, National Engineering College, Kovilpatti, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 25 October 2014
Received in revised form 26 November 2014
Accepted 29 December 2014
Available online 15 January 2015

Keywords:
Histogram of Optical Flow, Histogram of Oriented Gradient , Support Vector Machine, pedestrian detection

ABSTRACT

Pedestrian detection is the process of identifying and detecting a human from an image or video stream. In computer vision, it is a challenging process but an important subject in video surveillance system. The proposed method for pedestrian detection consists of Histogram of Optical Flow (HOF), the features of Histogram of Oriented Gradient (HOG) and a simple linear Support Vector Machine (SVM) classifier. Using background subtraction algorithm, the foreground and background of the video frame are segmented from which the moving object is extracted from a video frame. Two consecutive frame are split into 16 x 16 pixels grid cells. In each , affine transformation is applied and the corresponding cell in the consecutive of atleast three corresponding frames. Finally the confirmed optical flows are extracted from the consecutive frames. The histogram of gradient features are extracted from the local region which are given as a input of support vector machine. The SVM classify the features of histogram of gradient and detect the pedestrian from the moving object. The experimental results were obtained from real time video are provided which demonstrate the effectiveness of the method.

To Cite This Article: K. Thulasimani, Dr. K.G. Srinivasagan, T.K. Renugha, Detection of Pedestrian from Moving Object Using Histogram of Gradient and Optical Flow Method. ***Aust. J. Basic & Appl. Sci.,*** *9(2): 216-223, 2015*

INTRODUCTION

Pedestrian detection is one of the essential task for understanding the environment. In robot vision, vision based environment detection methods have been actively developed. The detection of moving object or pedestrian is more challenging task to keep away obstruction and control locomotion of the vehicle in the real world environment. In earlier period moving object or pedestrian detection methods for moving vehicle and a mobile robot have been developed actively. Hierarchical shape matching (Gavrila and Munder, 2007) employed a method to find pedestrian candidates from moving vehicle. Multi cue vision system method is used for real time detection and tracking of pedestrian. Support Vector Machine with automated selection of the components proposed by (K. Nishida and T. Kurita, 2005) using Ada Boost is applied, it shows a good performance for pedestrian detector. Object recognition and retrieval (Mikolajczyk and Schmid, 2005) proposed the local descriptors. Various local descriptors performance are compared and analyzed, among which the Scale Invariant Feature Transform (SIFT) descriptor returns the top matching result. Human detection algorithm using histograms of oriented gradients (HOG) (Dalal *et al*., 2006) are similar to the features used in the SIFT desriptor. The HOG features are extracted by orientation histograms of edge intensity in a local region. Selected features of HOG using PCA (Kobayasi *et al*, 2008) to decrease the number of features. Optical flow for a mobile robot was developed by motion detection methods and moving object detection methods. The directional divergence of the motion field is used in a qualitative obstacle detection method that was proposed by (Talukder *et al*, 2003). Moving object was detected by using optical flow which was offered during the translational of robot motion in real time. (R. F. Vassallo *et al*, 2002 and H. Liu *et al*, 2005) developed method for ego motion estimation from mobile robot using omini directional camera. Lucas Kanade optical flow tracker was used to obtain the corresponding features of the background in the successive two omini directional images. The camera ego motion was obtained by the compensated frame difference based on an affine transformation of the two successive frames. The corner features detection and tracking (Kanade-Lucas-Tomasi (KLT), 1991) were tracked by optical flow tracker. Only one affine transform model could not represent the whole background changes for detecting the moving objects because of the panoramic images has a local changes such as scaling,

Corresponding Author: K. Thulasimani, Department of Computer Science and Engineering, Government College of Engineering, Tirunelveli, India.
E-mail: thulasimani@gcetly.ac.in

translation and rotation of pixel group. Affine transform of the each local pixel groups are not a type of image features such as corner or edge which should be tracked by KLT tracker by (J. Hariyono *et al*, 2014). Each local sector of panoramic image was tracked by grid window based KLT tracker while the other methods used by sparse features based KLT tracker.

MATERIALS AND METHODS

The proposed method is motivated by the mechanism on pedestrian detection from moving vehicle by (D. M. Gavrila and S. Munder, 2007 and A. Talukder *et al*, 2003). Using optical flow method was proposed by (C. Tomasi and T. Kanade, 1991) and egomotion estimation defined by (R. F. Vassallo *et al*, 2002). Detection of pedestrian from a moving object is extracted from the relative motion represented by the same optical flow of the segmented region. The optical flow was obtained by the input image is divided into number of grid windows. But the confirmed optical flows are extracted from each window after performing affine transformation. The regions of moving objects and transformed objects are detected and are different from the previously registered background. Morphological operation is applied on the candidate region, the shape of the human an object is obtained. HOG features are extracted from the candidate region to recognize an object. The linear support vector machine classify the input of HOG feature vectors and detect the pedestrian from an video frame.

The framework of pedestrian detection from moving object using histogram of gradient and optical flow method is shown in Figure1. The input images are normalized into size of 64*128. For each image, histogram of gradient feature was calculated. All the HOG feature vectors are used as a input to the support vector machine to classify pedestrian or non-pedestrian.

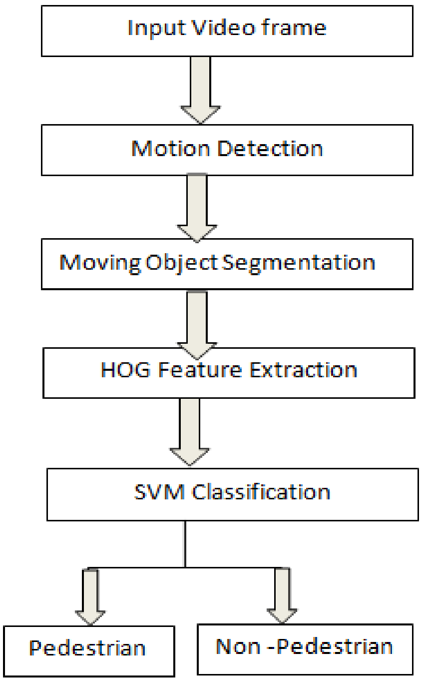


Fig. 1: The overview of the Pedestrian Detection system.

Moving Object Segmentation:

Moving object area are detected from a video or sequence of images. Because of camera egomotion caused from moving a camera, it is not easy to segment the moving area of an object. J. Hariyono *et al*, (2014) proposed a method to solve such a situation. The optical flow method was used to segment the unique motion of the moving object from the compensated egomotion caused by the camera. As compared the optical flow caused by egomotion from a camera with optical flow caused by independent motion of moving object have different pattern as a region. To detect pedestrian ,the HOG was applied on the different pattern as a region of moving object.

Frame Difference:

The process of checking the difference between one video to another in frames is called frame difference. The frame difference represents all motions caused by the moving object in the scene and camera egomotion.

Frame difference method was used to segment only the independent motion of the moving object to compensate this effect. So how much, the background image has been transformed in two sequences of images. Equation(1) represents the Affine transformation used to represent the pixel movement between two sequence images:

$$P' = AP + 1 \tag{1}$$

Where A is a transformation matrix. t is translation vector. P and P' are pixel location of the first and second frame. The Affine parameters are calculated by least square method using at least three corresponding features in two images.

Camera Egomotion:

The camera egomotion compensated was obtain by the frame difference method which is applied on the two consecutive input images. Equation(2) represents by calculating based on the tracked corresponding pixel cells :

$$I_d(x, y) = |I_{t-1}(x, y) - I_t(x, y)| \tag{2}$$

In this paper the input images are converted into gray scale images and one channel intensity pixel value was obtained from the input image. The two consecutive images $I_{t-1}(x, y)$ and $I(x, y)$ are divided into 14 x 14 pixels. Find and Compare the intensity value of each cell in the current frame to corresponding cell in the next frame. Similar intensity value are grouped together and will be selected as a corresponding value. Equation (3) represents to find the motion distance (C. Tomasi and T. Kanade , 1991) of each pixel in a group of cells.

$$g_{t-1}(i, j) = g_t(i + d_x, j + d_y) \tag{3}$$

The motion d in the $x-axis$ and $y-axis$of each cell $g_{t-1}(i, j)$, by finding most similar cell $g_{t-1}(i, j)$ in the next frame. where d_x and d_y are motion distances in $x-axis$ and $y-axis$respectively. Using Least square method, three consecutive features are atleast used to estimate the affine parameters.The affine transformation of each pixel in the same cell is rewritten from equation(4) is :

$$I_t(x, y) = AI_{t-1}(x, y) + d \tag{4}$$

where $I_t(x, y)$ and $I_{t-1}(x, y)$ are vector 2 × 1 which represent pixel location in the current and previous frame, respectively. A is 2 × 2 projection matrix and d is 2 × 1 translation vector.

Moving Object Localization:

The output of each pixel from frame difference method using egomotion compensated not clearly shown as silhouette. Only it returns the information about the motion areas from moving objects. Morphological operation is applied on the moving areas, the noise are removed and the region of moving objects are obtained. The region segmentation algorithm accurately locate the motion regions in the difference image that are represented by the bounding boxes. In this paper, detection of moving objects are represented by the position in width in $x-axis$. The projection histogram h_x is used by pixel voting vertically projects the image into $x\_coordinate$. The region segmentation technique proposed by (A. Ess *et al*, 2007) , define the region using boundary saliency from which the horizontal difference of data density in the local neighbourhood is measured. In moving object detection, the local maxima keeps to track where the maximal change in data density occurs are candidates for region boundaries of pedestrian.

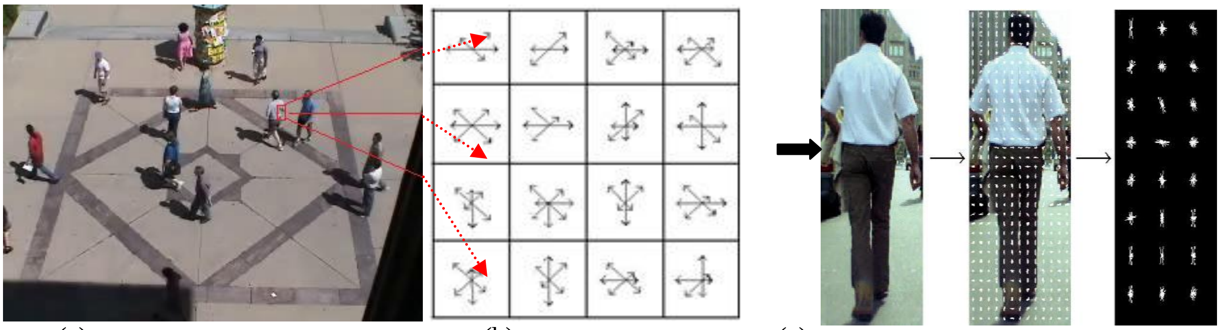
Histogram Of Gradient Feature Extraction:

The histogram of gradient features are extracted from 16 × 16 local regions. The edge orientation, gradients value are calculated from each pixel from the local region using Sobel filters. The edge orientation $\theta(x, y)$ and gradient magnitude $m(x, y)$ are calculated equation (5) & (6) using the directional gradients $dx(x, y)$ and $dy(x, y)$ are computed as follows:

$$\theta(x,y) = \begin{Bmatrix} tan^{-1}\left(\frac{dx(x,y)}{dy(x,y)}\right) - \pi, if\ dx(x,y) < 0, dy(x,y) < 0 \\ tan^{-1}\left(\frac{dx(x,y)}{dy(x,y)}\right) + \pi, if\ dx(x,y) < 0, dy(x,y) > 0 \\ tan^{-1}\left(\frac{dx(x,y)}{dy(x,y)}\right), Otherwise \end{Bmatrix} \tag{5}$$

$$m(x,y) = \sqrt{dx(x,y)^2 - dy(x,y)^2} \tag{6}$$

The image of the local region is segmented into small spatial or cell. Each cell size is 4 × 4 pixels. The histograms of edge gradients are calculated based on eight orientations from every local cells. Therefore the total number of histogram of gradient features are 128 = 8 × (4 × 4) which form HOG feature vector. Sudden changes in the descriptor will be avoided with small changes in the position of the window and less emphasis to gradients are given that are far from the center of the descriptor. A Gaussian function σ equal to one-half of the width of the descriptor window. It is used to assign a weight to the magnitude of each pixel. A vector of histogram of gradient features represent the information about edge cells and local shape of an object. The HOG has a flatter distribution in flatter regions like a wall of a building or a ground and only one of the edge cell in the histogram has highest element which represents the direction of the edge. The local geometric and photometric transformations of histogram of gradient features are very robust. If the smaller object of translations, rotations are smaller than the local spatial bin size and result of the effect is very low. A set of HOG feature vectors from all locations in an image grid are extracted by (Dalal and Triggs,2005). Figure 2 represents the extraction of HOG features from all the locations on the local region from an input image.



(a) (b) (c)

Fig. 2: (a). Input Image. (b). Histograms of edge gradients with 8 orientations are calculated from each of the 4 ×4 local cells. (c). HOG features are extracted from all locations of the local regions with 16 × 16 pixels.

Support Vector Machine:

A support vector machine constructs a hyper plane in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks. A good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class.

Given some training data is a set of points of the form:

$$D = \{(x_i, y_i) \mid x_i \in R^p, y_i \in \{-1,+1\}\}_{i=1}^{n} \quad (7)$$

where the y_i is either 1 or −1, indicating the class to which the point x_i belongs. Each x_i is a p -dimensional real vector. Find the maximum-margin hyper plane that divides the points having $y_i = 1$ from those having $y_i = -1$. Any hyper plane can be written as the set of points x satisfying $w.x - b = 0$. The parameter $\frac{b}{||w||}$ determines the offset of the hyper plane from the origin along the normal vector W.

Experimental Results:

Support Vector Machine is mainly preferable for human classification. This classification method can be used to real time video. The resolution of the real time image is 640 x 480 and 480 x 320. According to the classification system, the positive and negative example of person training set of images are trained using SVM classifier in MATLAB as shown in figure 3. The size of the sample images are 128×64 and each 100 samples of human and non-human are prepared as show in Figure 3. The HOG features of every sample image should be extracted. Each 100 test images of human and background from Person training sets are prepared. Table I shows the detection rates and miss rates of test image samples. For illustration, there are 96% human samples are classified as human and miss rate is 4%. While 7% Non-human samples are recognized as human. So the detection rate of Non-human sample is 93%. Table I shows the classification result on image using SVM and Neural Network classifier.

(a) Positive sample images

(b) Negative sample images

Fig. 3: Positive and negative samples images.

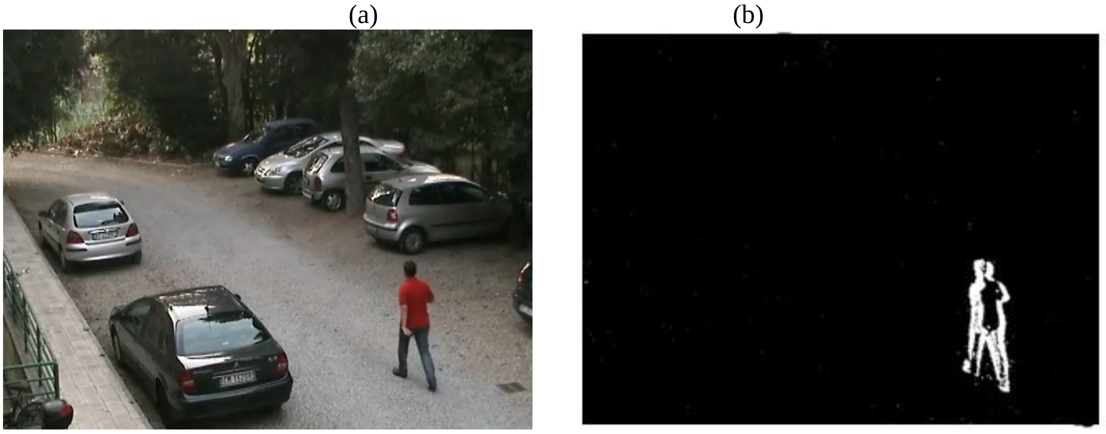
Comparison of SVM classifiers with Back Propagation neural network method is used to show its classification ability. The same with SVM, each 100 samples of human and non-human are prepared, the HOG features are input vector of neural network, and the training signal is defined as 1 and 0. New data is prepared to test the classification result after training. Table I shows the result.

Table I : Svm and Nn Classification Results For Images

Test input	SVM classification		Neural Network Classification	
	Human (%)	Non-Human(%)	Human (%)	Non-Human(%)
Human	96	4	81	19
Non-Human	7	93	28	72

Our classification method can be used in real time video, when the resolution of scene is 640 x 480 and 480 x 320. Comparison of SVM is preferable to human classification. In our experiment , we are implemented three real time video. Figure 4 represent output of moving object segmentation for the given the input frame. The histogram of optical flow , histogram of gradient and tracking of moving object of the three real time video are shown in figure 5, 6, 7 respectively.

(a) (b)

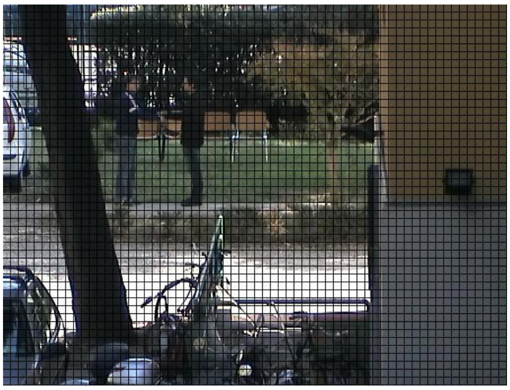


(c) (d)


(e)


(f)

Fig. 4: (a), (c), (e) - input frames of real time videos. (b), (d), (f) - Segmentation of the input frames.

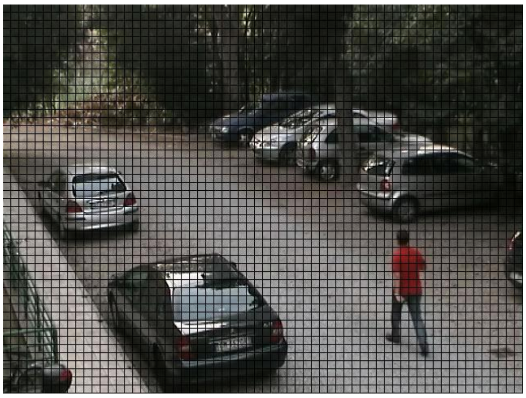

(a)

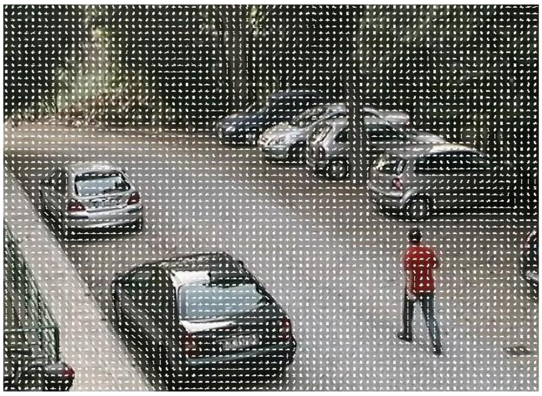

(b)


(c)


(d)

Fig. 5: Video:(a).Grid View (b). Histogram of Gradient (c). Histogram of Optical flow (d). Object Tracking


(a)

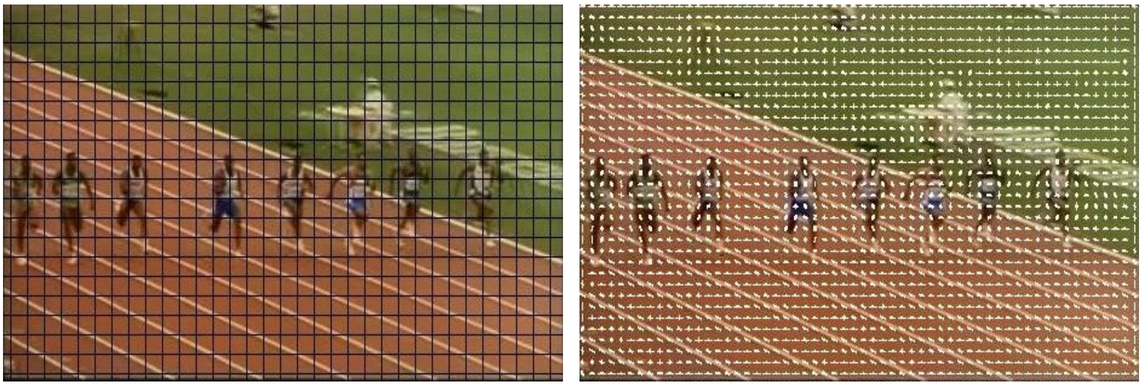

(b)

Australian Journal of Basic and Applied Sciences, 9(2) February 2015, Pages: 216-223

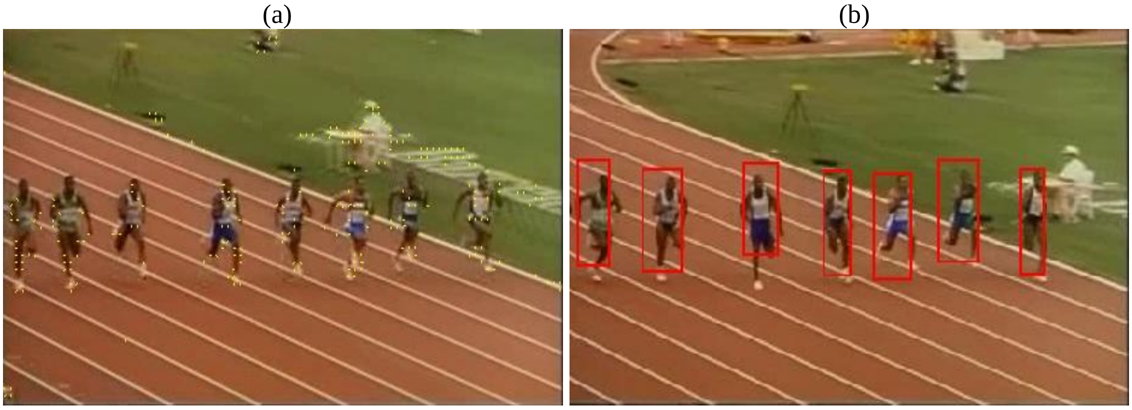
(c) (d)

Fig. 6: Video2:(a).Grid View (b). Histogram of Gradient (c). Histogram of Optical flow (d). Object Tracking

The recognition rate for test dataset is 98.3% by Dalal *et al*.,(2006). The combination of methods based on optical flow and HOG feature are extracted from all locations of the grid for each training sample. The selected feature vectors were used as input of the SVM. Recognition rates of the constructed classifier using test samples are evaluated. The relation between the detection rates and the number of false positive rate is calculated. The best recognition rate 99.3% was obtained using test samples. Thus the higher detection rate was achieved using the combination methods of optical flow and HOG.



(a) (b)



(c) (d)

Fig. 7: Video3 :(a).Grid View (b). Histogram of Gradient (c). Histogram of Optical flow(d). Object Tracking

Conclusion:

In this paper, a pedestrian detection approach is based on histogram of gradient and histogram of optical flow. The moving object is extracted from the relative estimation of optical flow. The candidate region of the pedestrian is obtained from morphological operation. The candidate region of the histogram of gradient features were extracted to recognize an object. The HOG features are used as an input of the support vector machine.

SVM classify the features of HOG into pedestrian or non-pedestrian. Experimental results from real-time video are provided the effectiveness of the method.

REFERENCES

Dalal, N., B. Triggs, and C. Schmid, 2006. Human detection using oriented histograms of flow and appearance. Proceedings of the 9th IEEE European Conference on Computer Vision, Graz, Austria.

Gavrila, D.M. and S. Munder, 2007. Multi-cue pedestrian detection and tracking from a moving vehicle. International Journal of Computer Vision, 73(1): 41–59.

Hariyono, J., V.D. Hoang and K.H. Jo, 2014. Human detection from mobile omnidirectional camera using ego-motion compensated. Intelligent Information and Database Systems, vol. 8397 of Lecture Notes in Computer Science, pp: 553–560.

Kobayasi, T., A. Hidaka and T. Kurita, 2008. Selection of histograms of oriented gradients features for pedestrian detection. Proceedings of the 14th International Conference on Neural Information Processing (ICONIP '08), vol. 4985 of Lecture Notes in Computer Science, pp. 598–607, Kitakyushu, Japan.

Mikolajczyk, K. and C. Schmid, 2005. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10): 1615–1630.

Nishida, K. and T. Kurita, 2005. Boosting soft-margin SVM with feature selection for pedestrian detection. Proceeding of International Workshop on Multiple Classifier Systems, 13: 22–31.

Talukder, A., S. Goldberg, L. Matthies and A. Ansar, 2003. Real-time detection of moving objects in a dynamic scene from moving robotic vehicles," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp: 1308–1313, Las Vegas, Nev, USA.

Tomasi, C. and T. Kanade, 1991. Detection and tracking of point features," International Journal of Computer Vision, 9: 137–154.

Vassallo, R.F., S. Victor, and H. Schneebeli, 2002. A general approach for ego motion estimation with omnidirectional images. Proceedings of the 3rd Workshop on Omnidirectional Vision, pp: 97–103, Copenhagen, Denmark.