International Journal of Security and Its Applications
Vol.8, No.1 (2014), pp.93-102
http://dx.doi.org/10.14257/ijsia.2014.8.1.09

A Novel Approach to Design the Fast Pedestrian Detection for Video Surveillance System

Shuoping Wang[1], *Zhike Han[1], Li Zhu[2] and Qi Chen[2]

[1]*Zhejiang University City College, Hangzhou, P.R. China, 310015*
[2]*College of Computer Science and Engineering, Zhejiang University, Hangzhou, P.R. China, 310027*

Abstract

The pedestrian detection is a hot research topic in computer recognition. It involves not only the pedestrian location information but also the intrusion detection function, which has wide prospects in the application of vehicle traffic, campus monitoring, and building guard. However, the identification accuracy and recognition speed play an important role in the pedestrian detection, which calls for a fast pedestrian detection approach. The general pedestrian detection implementation, based on the integral channel features method and soft cascade classifier, is the popular technique in the current business application since its better speed and accuracy. Thus, this method uses the feature approximation technique and multiple classifiers to achieve the feature computing, which speeds up the detection without resizing image. To this end, this paper is motived to propose a multi-scale handling method for the fast pedestrian detection, using the tactics detection from sparse to dense. Our pedestrian detection method consists of four parts functions, mainly pedestrian statistics and intrusion detection, pedestrian tracking and pedestrian flow statistics. All these modules are introduced with its details about design and implementation. In Addition, the proposed multi-scale handling method can be applied into most of object detectors to improve their recognition speed. In conclusion, our proposed approach has a good potential application prospect in the video surveillance system.

Keywords: *Pedestrian Detection, Integral Channel Features, Soft Cascaded Classifier*

1. Introduction

Pedestrian detection is a key problem in computer vision, which has promising applications in various domains such as surveillance video, robotics, assistant driving, and smart cameras. Applying pedestrian detection technique into vehicle-mounted monitoring system helps driver to timely stop the vehicle avoiding the traffic collision, by which it can identify pedestrians in the front of vehicle. It also can be applied into apartment, supermarket, museum and scenic spots to count the number of pedestrians and figure out the pedestrian flow statistics for business optimization. However, with the development of society and technique, the emerging complex application scenarios are demanded to be handled. For example, the speed requirement on embedded device of vehicle should be fast, and the detection accuracy in crowded area of supermarket should be high.

Various methods have been proposed to the pedestrian detection for last decades, such as Integral Channel Feature (ICF), Histograms of Oriented Gradients (HOG) Feature, Soft Cascaded Classifier (SCC), and Local Binary Patterns (LBP). For integral channel feature (ICF), the core concept about channel can be traced back to the earlier computer vision research. It has been used to efficiently compute the histograms of oriented gradients. As in

ISSN: 1738-9976 IJSIA

paper [1] described, the rectangular histogram can be computed by quantizing an image into multiple channels. The gradient magnitude and color features are unified together, which help detecting for getting a better effect. For histograms of oriented gradients (HOG), it is first proposed by Dalal and Triggs [2]. This method divides the original image into dependence blocks where each block is further divided into smaller cells. In that each cell, the histograms of oriented gradients are computed respectively. From the cell to block, all histograms of oriented gradients are connected as a total histograms of oriented gradient for the original image. For the soft cascaded classifier (SCC), Viola and Jones [6] initially proposed it to the fast facial recognition. Zhu *et.al.*,[3] combined HOG and soft cascaded classifier to the pedestrian detection. The soft cascaded classifier trains a set of classifiers from simple and complex. The detection windows beyond to target are excused, and the complex and hard detection windows are delivered to complex classifier. This approach saves the computing consume and sorting time. Note that, only the positive sample needs to compute its feature which improves the computational efficiency. Finally, for the local binary patterns (LBP), it is used to describe the texture of image feature. Ojala [4] proposed LBP in 1996. The 8 pixels around the selected pixel are considered as the threshold value which is assigned to 1 or 0. The probability value for these 8 pixels is 256. After that, it counts the value of histograms of oriented gradients of each image area when using LBP method. The result is employed as the texture description. Based on above existing researches, this paper proposes a novel approach to design a fast pedestrian detection for video surveillance system. The design mainly includes pedestrian statistics, intrusion detection, pedestrian tracking and pedestrian flow statistics functions.

The remainder of this paper is organized as follows. Section 2 summarizes the related works. Section 3 shows requirements about the fast pedestrian detection. Section 4 presents the each module and it designs. Section 5 discusses conclusions and future works.

2. Related Work

Various methods have been proposed to pedestrians detecting. They have a good performance in the upright holistic pedestrian detection. Papageorgiou *et. al.*, [5] proposed a sliding window-based target detector, where the SVM classifier [9] is used to the object identification of multi-scale Harr feature. Viola and Jones *et. al.*, [6] improved the real-time face recognition system at the running speed level. Their approach mainly computes the picture integral to reduce the redundancy. It only calculates the minor features for simple classifier which can rapidly remove most of negative samples for accelerating the target task checking speed. This method is regarded as a basis research for successor. Most researches are contributed to the new features extraction and its utilization. Dalal and Triggs *et. al.*, [2] given a feature using histogram of oriented gradient (HOG), and applied it into pedestrian detection. The method has a good effectiveness since it has reduced the miss-detection ratio at least one magnitude, comparing to the Harr-based detector. Zhu [3] combined the HOG feature and Viola-Jones method via calculating the gradient integral and adopting the cascaded classifier. This approach achieved pedestrian detection from the 320*240 px image which helps reducing the detection time consumption to 0.1 second. In latter, Wang *et. al.*, [7] added LBP features of image texture descriptions to HOG features to increase detection rate to 97.9% when the miss-detection ratio of negative windows is 0.0001 using INRIA dataset.

Other features also are researched and applied into specific pedestrian detection systems, such as color features and running features. Dollar *et. al.*, [8] proposed the concept about integral channel features which extracted gradient and color channels from the transformed image. After that, the accumulative integral value of special channels area is selected as image feature, which has a better effectiveness during its applications. In the improvement

aspect of learning and classification algorithms, Maji *et. al.*, [9] proposed a IKSVM classifier in which IKSVM classifier had a better effectiveness than the liner SVM classifier in general. Turnel *et. al.*, [10] proposed a Riemann flow-based classifier for pedestrian detection. To handle various kinds of problems in Viola-Jones-based cascaded classifier, soft-cascaded classifier is proposed in papers [11, 12]. The strong classifier consists of some week classifiers trained by Boosting algorithm. The classifier is sorted with a threshold value which is compared to the total response value computed by all current classifiers. If it is less than the threshold value, the classifier immediately outputs negative value.

To improve the detection speed of pedestrian detection, these are mainly concerning on reducing the feature computing time or classifier computing time. For the former, integral feature is used, for example, Zhu *et. al.*, adopted HOG feature, Dollar *et. al.*, adopted integral channel feature. For the latter, the approximate calculation of adjacent scale features is used. Due to the different pedestrian has different size in scenarios, the image should be zoomed for several times. Thus, the zoomed image needed to be recomputed again. It made the computing task increased rapidly. They also found that the channel value is changeable with exponential when the image is zoomed. The approximately calculating the feature can reduce the computing time. Thus, Benenson *et. al.*, introduced this ideal to channel computing that multi-image scales were replaced by multi-classifiers. It made feature computing only execute at only once [31].

Adopting GPU helps to reduce the classifier's computing time. Due to the detection computing is usually independent, sliding window detection of multi-image scale intensions can be computed in parallel. Therefore, GPU accelerates the speed of recognition. Beneson *et. al.*, employed GPU to increase the pedestrian detection speed among 20FPS and 100FPS.

The current pedestrian detection is often to distinguish the upright holistic pedestrian detection. But when the pedestrian flow is intensive the algorithm performance will be suddenly drop. To handle this problem, many researchers proposed some solutions. For example, Wang *et. al.*, proposed a HOG features-based algorithm for overlap detection method in SVM response. Zeng *et. al.*, applied PCA-based mutil-scales HOGLBP features to body's head and shoulder detection. Felzenszwald *et. al.*, proposed a body model-based LatSVM algorithm to pedestrian detection, which wined a good score in PASCAL VOC target recognition content.

3. The System Requirements of Pedestrian Detection

In following paragraphs, we will briefly introduce the pedestrian detection requirement of video surveillance system.

3.1. Video Monitoring

The camera used for video monitoring plays a vital role in pedestrian detection. It is considered as the image source. However, camera only records the video stream of monitoring area. Thus, the video should be readable and replayed, and the camera parameters should be settable. To this purpose, multiple video cameras need to be connectable and they can be encode/decoded for replay. For the camera parameters setting, it should consist of two type parameters. The first parameter acts on camera which includes video format, resolution ratio and code rate. The other parameter acts on pedestrian detection which includes the parameter about detection area.

3.2. Pedestrian Statistics and Intrusion Detection

The statistics function counts the pedestrian number in a specific image, which is picked by cameras. It can determine the location of pedestrians in that image. The interface of Video Player will display the statistics results. After that, the result and camera ID are saved into the log database, as well as the information about the current time.

Figure 1. The Example of Statistics Result

As Figure 1 shown, it shows an example of statistics result. The specific area of images displays the pedestrian number in real-time, showing the entering and exiting people. Note that the specific size and location of areas can be freely setup according to the user requirement. Moreover, the intrusion detection requires recognizing the suspects of invasion in video surveillance system. The pedestrian recognition-based intrusion detection needs not only the low ration of miss-detection but also the high speed of detection. It supports sending alarms and saving suspects into database if the invasion occurs.

3.3. Historical Data and Statistical Chart Display

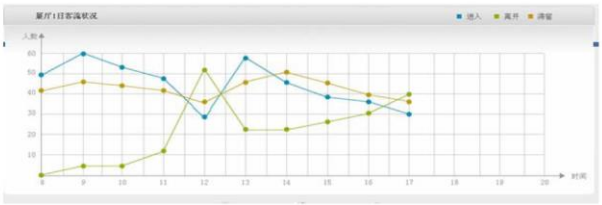

Figure 2. The Example of Statistical Chart Display

The pedestrian statistics and intrusion detection need to be saved into the historical database, in which information can be queried and displayed in chart. For example, the chart displays the number change of pedestrian within a month or the number of pedestrian that they cross the specific detection line from Monday to Friday. As Figure 2 shown, in the intrusion detection, the information about the suspicious, time and image information are queried and displayed showing the pedestrian who enter the monitoring area at a certain time.

4. The Design of Pedestrian Detection

The design consists of four modules, mainly the pedestrian detection and tracking module, camera control module, video play module and chart display module. These modules communicate with each other, invoking or transmitting messages to implement system

functions. The camera control module and video play module are used to connect multi-cameras in order to capture the video data and achieve video stream decoding and display. The pedestrian detection and tracking module is to locate the pedestrian for each frame. The chart display module saves information into database, such as location, current time and camera ID. When querying these data, the figure and table will be used to display.

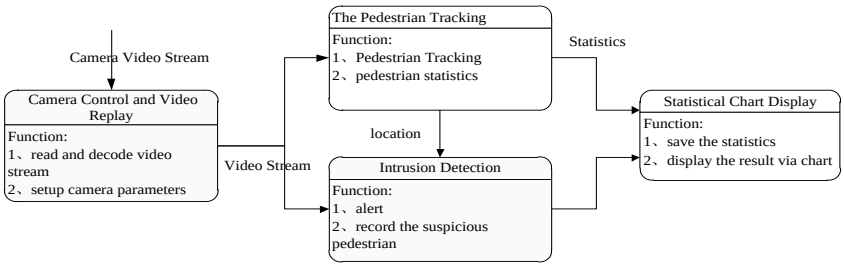

Figure 3. The System Design

As Figure 3 shown, it receives the pedestrian detection, tracking module, and the alarm generated from the pedestrian statistics and intrusion detection. Then it saves all received information into database. The traditional pedestrian detection, such as HOG feature, LBP feature and SVM classifier, used for analyzing images mainly usually expense at lease seconds for multi-scales pedestrian recognition. Thus, it is hard to satisfy the application requirement. In our paper, the integral channel features (ICF) and soft cascaded classifier (SCC) are employed to pedestrian detection in order to meet the accuracy and speed.

4.1. The Overall Workflow of Pedestrian Detection

We propose a sliding window-based detection approach for the fast pedestrian detection. The workflow of sliding window-based target detection is as follows.

1) The image is zoomed for multi-scales,
2) The movement is intensively executed in each zoomed image via sliding window,
3) Using a classifier sorting current window to judge which is window or which is pedestrian,
4) For all discovered windows of pedestrians, non-maxima suppression method is used to find the location of pedestrians.

It mainly consists of feature extraction and classifier training. According to the number of zoomed image scales, the size of sliding window and the number of classifier may be different. Thus, it is suitable to be divided into different sliding windows for detections. The original sliding window detection doesn't magnify the image. Thus, we adopt different sliding window to move at the original image. Each sliding window corresponds to a windows classifier in which the image shouldn't be zoomed. But in order to distinguish N kinds of pedestrian in different size, it is to train the N classifier. However, the other plan is to train N classifier. Then the image is zoomed for N times where each zoomed scales adopt different size of sliding windows. This method computes the feature of N images. But it only should train one classifier, based on which the classifier can be suitable to the different scales of zoomed image. To this end, this paper combines these two method's advantages. One reason is that it doesn't magnify the captured images. The integral channel feature has the invariance property during movement. The other reason is that it doesn't train N classifier. In general, the N/K classifiers can be satisfied. The integral image channel under small zoom can be figured by the original feature result approximate calculation.

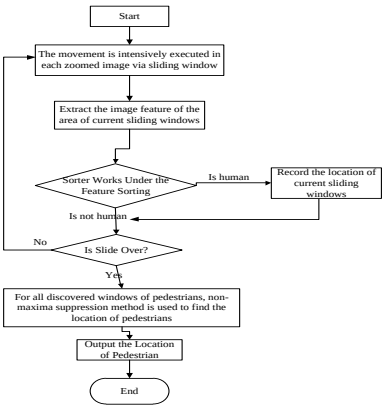

Figure 4. The Overall Workflow

4.1.1. Integral Channel Features

For Integral Channel Features, we adopt the Dollar and Beneson's channel features method, including 6 gradient direction channels, 3 LUV color channels and 1 gradient amplitude channels. And it computes the feature only for once at the original image.

In gradient direction and gradient amplitude channels computing, the first work is to change the original image into image gray processing. Then, each pixel's gradient direction and amplitude is computed. The amplitude matrix is gradient angle channel. The gradient direction is discretized to 6 block areas. If one pixel's gradient direction belongs to No.i block area, then the amplitude value is voted to the location of the gradient amplitude channels. After that, the other value for 5 locations of the gradient amplitude channels is setup as 0. And 3 color channels correspond to channels of LUV. After completely computing channels features, each channel needs to be calculated for integral. Note that the LUV channel and gradient channel computing are independent. They can be operated in concurrent.

During reading the integral channel computing, it needs approximate calculation. The current window is zoomed to the detection window size of the all matching adjacent classifiers. The approximate estimate formula figure out the feature value after zooming the image via integral channel computing. For example, if the size of the current sliding window is 40*80, it needs to train 3 classifiers that the training window size is 32*64, 64*128, and 128*256, respectively. Then the image of current sliding window should be zoomed to the window scales of adjacent classifiers, mainly 32*64. The new computing formula is as follow

$$r(s) = \begin{cases} a_u \cdot s^{b_u} (s > 1) \\ a_d \cdot s^{b_d} (s < 1) \end{cases}$$

The parameter s is the zoom rate of window. The parameter $r(s)$ is the zoom rate of feature. The parameters a_u, b_u, a_d, and b_d are used during image zoom. Usually, each channel feature parameters are different.

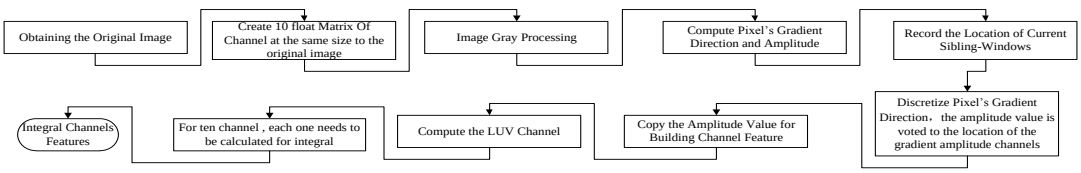

Figure 5. The Workflow of Computing Integral Channel Features

4.1.2. The Soft Cascaded Classifier Design

According to Dollar's researches [13, 14], this section designs a soft cascaded classifier. The week classifier adopt two level decision tree classifier, using Ada boost learning algorithm. The threshold value of week classifier is setup as follow. The trained strong classifier is used at the tested or verified dataset. The minimum value at each stage accumulation from the all sorted samples is considered as the threshold value of week classifier. The strong classifier has the same sort value before and after setting. It can improve the detection speed.

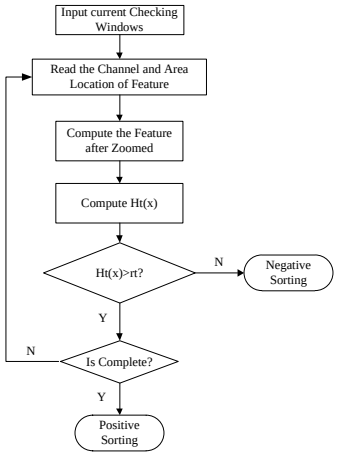

Figure 5. The Workflow of Soft Cascaded Classifier

Considering the multi-scales pedestrian detection, it needs to detect the pedestrian at multi-scales. In pedestrian statistics and intrusion detection, the monitoring cameras usually take only a small area of pedestrians. Usually, in the small piece of pedestrian video images, the pixel size is not big. The major work of zoom scale can't detect pedestrians because the pedestrians are often concentrated at small scale of zoomed image. Thus, the multi-scale detection method is starting from sparse to dense. The method starts from detecting sparse zoom scale, then to scanning the pedestrian existing level of adjacent level. After that, no pedestrian level of adjacent level can be given up. It reduces the useless computing to improve the detection speed.

4.2. The Multi-scales Detection from Sparse to Dense

For the sliding window-based detection method, the detection can only distinguish the pedestrian who has the same size to the sliding window since the sliding window is fixed. In the real scenario of vehicle, the pixel size is usually lager because there has a big distance between pedestrian and camera. In order to detect different size of pedestrians, there are two way to implement. The first method is that the original image is zoomed at multi-scales. Then, the same size of sliding window is used to checking each zoomed image. The second method is that using different sliding windows to check the original image consecutively. We adopt the later method to implement the multi-scales detection from sparse to dense.

Here, the zoom scale is used as the image or window zoom rate. For example, if zoom scale is 0.5, the original image is magnified for twice in the former method, while in the latter method it means that the current detection window of scale is a half of the initial window.

The multi-scales detection merges detection windows into a final pedestrian window. The location near to the pedestrian and the same in the adjacent scales are detected as the positive

samples. The Non-Maxima Suppression (NMS) method is used to merge these detection windows, in which the mainly method include Mean-Shif state estimation and Parirwise Max Suppression.

At present, the multi-scales detection algorithm has been widely proposed. Beneson adopted 55 detection scales. This approach has been used in moving vehicle for capturing database, such as, Caltech database. These cameras usually shoot at horizontal level because the pedestrian is usually far from the cameras. Given 640*480 pixels, the nearby pedestrian may occupy the total high of video while the forane pedestrian may only have 50 pixels. Given 64*128 pixels that the zoom scales is changed from 0.4 to 4, it adopts multi-zoom scales in order to avoiding the omitting the different size pedestrian.

For text application scenario for pedestrian statistics and intrusion detection, especially the image form monitoring camera, the camera is far from the ground and the monitoring area is small. According to requirements, it should handle multi-camera with different resolution ratios and different locations. But it is complex to setup the parameter for cameras. Thus, we design a multi-scales detection from sparse to dense. For the N scales between the min zoom scale S_{min} and the max zoom scale S_{max}, 1) the first work is to select a scale to distinguish with intervals M scales, 2) if the pedestrian is detected then the intervals M scales is changed as M/2 scales.

4.3. The Pedestrian Tracking based Statistics

The target tracking finds the location of target or multi-targets in each frame of video. The track of each running target in video can be got via target video, which has application values in the aspect of the human-computer interaction, video compression, video monitoring, and video editing. As a result, it can figure out the pedestrian number about the entering and leaving information.

The computation of traditional particle filter tracking algorithm is large. It is hard to meet the real-time requirement. This paper first uses the fast pedestrian detection algorithm. Then it compares the similarity for the checking result of adjacent frames. Finally, for each detected pedestrian it confirms the location for the next frame.

The similarity formula between detection results is as follow.

$$Similarity(BBa, BBb) = \left(\frac{Fa \square Fb}{\|Fa\|\|Fb\|}\right)^{M} \frac{D}{\|Ca - Cb\|^{N}}$$

The parameters F_a and F_b are the vector of bounding box a and b's of the integral channel features. This vector can be implemented by 6 dimension gradient direction integration or 3 dimension LUV color integration. Due to the integral channel features of each frame has been figured out, the feature vector can be quickly computed via multi-memory read and additions/subtractions. The parameters C_a and C_b are the center of each bounding box respectively. M is the impact parameter about how the control feature similarity acts on the final similarity. D and N are the impact parameter about how the distance of bounding box acts on the final similarity. The parameter T is used to compare whether the similarity of bounding boxs is great than or not. If it is great than T, the two bounding boxes are considered as one person. In general, the M, D, N and T are obtained by experiments.

The process of pedestrian tracking algorithm is as follow. First, it reads the next frame of video stream. Second, it searches the pedestrian who has highest similarity for the current frames according to the former frames. These two pedestrians are marked as the same pedestrian. Third, it assigns new pedestrian ID for the un-matching pedestrian in current frame. This process is executed for repeat.

In order to reduce the impact of the miss-detection and omit-detection in pedestrian detection, the new pedestrian only displayed in continuous frames and contained in bounding box with success matching is considered to add the pedestrian list. It avoids miss-detecting the new pedestrian. Only continuous 3 times miss-matching can be deleted. For the fixed monitoring camera, the detection line and detection area can be setup to judge the direction of pedestrian for pedestrian statistics.

4.4. The Camera Control and Video Replay

This module is used to interact with monitoring camera, in which the parameters includes the camera, video stream reading and video stream encoding. The parameter has two types. The first type consists of resolution ratio, code rate and video encode format, which are related to the camera. The second type consists of the detection line and area, which is related to the pedestrian detection.

In most video stream, the compression algorithm adopts MPEG4 or H.264, and the video is sent to client via HTTP or RTSP protocol. For the client, it decodes the video stream. FFMPEG is open source for decoding the compressed video stream. The workflow is as follow.

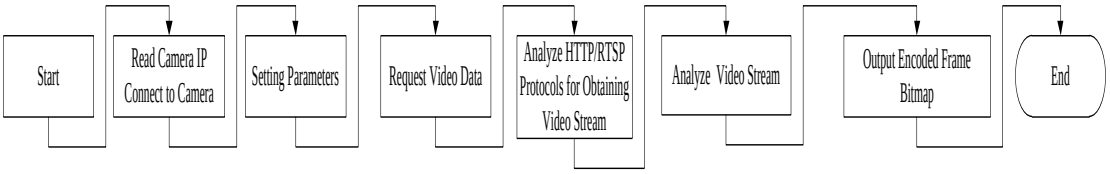

Figure 6. The Workflow about Camera Control and Video Replay

4.5. The Chart Display

The chart display module shows the pedestrian number and pedestrian flow statistics in visualization, supporting the chart display at time and space dimension. The time dimension shows the pedestrian flaw change in the line chart. The space dimension compares the number of pedestrian and pedestrian flow.

This function is implemented by database and GUI technique. The number of pedestrian and the pedestrian flow statistics are saved into database after computation. The end user can query the camera ID and time interval. And then the query parameters are translated into SQL. The result about the target SQL are executed in database and the chart including line chart and table chart are displayed to user as a feedback.

5. Summary and Future Works.

In order to implement the fast pedestrian detection, this paper gives an overall design and its details. To meet real requirements, the fast pedestrian detection system consists of four part functional functions. It first introduces the overall design of system architecture and analyzes the relation between each module. For the detail function of each module, the target function is to be implemented under comparing previous studies in order to show the advantages and disadvantages. Based on this strategy, the detail plan of implementation for the fast pedestrian detection is discussed. For the future work, we will give a system prototype and execute some experiments to demonstrate the feasibility of our proposed.

Acknowledgment

This paper is supported by Hangzhou Key Laboratory for IoT Technology & Application, 2011 Open Fund Projects of Mobile Network Application Technology Key Lab of Zhejiang Province No.MNATKL2011004 and No.MNATKL2011005, the Zhejiang Scientific and Technical Key Innovation Team of New Generation Mobile Internet Client Software (2010R50009).

References

[1] F. Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. Proceedings of the International Conference on Computer Vision and Pattern Recognition.(**2005**) June 20-25,Vol.1:829-836; San Diego, CA, USA.
[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. Proceedings of the International Conference on Computer Vision and Pattern Recognition.(**2005**) June 20-25,Vol.1:886-893; San Diego, CA, USA.
[3] Q. Zhu, S. Avidan, M. Yeh, and K. Cheng. Fast human detection using a cascade of histograms of oriented gradients. Proceedings of the International Conference on Computer Vision and Pattern Recognition.(**2006**) June 20-25, Vol.2:1491-1498; New York, NY, USA.
[4] Ojala, T., Pietikäinen, M. and Harwood, D. , A Comparative Study of Texture Measures with Classification Based on Feature Distributions. Pattern Recognition.19(3):51-59 (**1996**).
[5] C. Papageorgiou and T. Poggio. A trainable system for object detection. Intl. Journal of Computer Vision. 38(1):15-33(**2000**).
[6] P. A. Viola and M. J. Jones, "Robust real-time face det. Intl. Journal of Computer Vision. 57(2):137-154(**2004**).
[7] X. Wang, T. X. Han, and S. Yan.An HOG-LBP human detector with partial occlusion handling. Proceedings of the 2th International Conference on Computer Vision.(**2009**) September 27 - October 4,32039,Kyoto, Japan.
[8] P. Doll´ ar, Z. Tu, P. Perona, and S. Belongie. Integral channel features. Proceedings of the British Machine Conference (**2009**) 7 September,1-11 London, UK.
[9] S. Maji, A. Berg, and J. Malik. Classification using intersection kernel SVMs is efficient. Proceedings of the Conf. Computer Vision and Pattern Recognition(**2008**). 24-26 June,1-11, Anchorage, Alaska, USA.
[10] O. Tuzel, F. Porikli, and P. Meer.Pedestrian detection via classification on riemannian manifolds. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30(10),pp:1713-1727(**2008**).
[11] Bourdev, L., Brandt, J.: Robust object detection via soft cascade. Proceedings of the International Conference on Computer Vision and Pattern Recognition.(2005) June 20-26, 236-243; San Diego, California , USA.
[12] Zhang, C., Viola, P.: Multiple-instance pruning for learning efficient cascade detectors. U.S. Patent US20090018980 A1, July 13 **(2007)**.
[13] Dollar, P., Belongie, S., Perona, P.: The fastest pedestrian detector in the west. Proceedings of the British Machine Conference (**2010**) 31 August- 3 September ,1-11 Welsh, UK.
[14] P. Dollár, R. Appel and W. Kienzle. Crosstalk Cascades for Frame-Rate Pedestrian Detection. Proceedings of the European Conference on Computer Vision (**2012**) 7-13 October , 645-659, Florence, Italy

