

# Intelligent Pedestrian Detection using Optical Flow and HOG

Huma Ramzan, Bahjat Fatima, Ahmad R. Shahid, Sheikh Ziauddin and Asad Ali Safi  
Department of Computer Science  
COMSATS Institute of Information Technology  
Park Road, Islamabad

**Abstract**—Pedestrian detection is an important aspect of autonomous vehicle driving as recognizing pedestrians helps in reducing accidents between the vehicles and the pedestrians. In literature, feature based approaches have been mostly used for pedestrian detection. Features from different body portions are extracted and analyzed for interpreting the presence or absence of a person in a particular region in front of car. But these approaches alone are not enough to differentiate humans from non-humans in dynamic environments, where background is continuously changing. We present an automated pedestrian detection system by finding pedestrians' motion patterns and combing them with HOG features. The proposed scheme achieved 17.7% and 14.22% average miss rate on ETHZ and Caltech datasets, respectively.

**Keywords**—Pedestrian detection, pedestrian protection system, HOG descriptor, optical flow, motion vectors, FPPI, miss-rate

## I. INTRODUCTION

According to National Highway Traffic Safety Administration (NHTSA), traffic fatality rate has been increased by 6% in 2012 and on an average nearly 4,743 pedestrians were killed which accounted for 14% of the total traffic related fatalities along with 76, 000 ended up injured in USA [1]. In countries of Asia and Europe due to high population, the rate of road user deaths is much higher. This rate of pedestrian deaths and injuries could be reduced by employing intelligent frameworks for detecting people on road. But to give viable safety, such frameworks need to recognize people on foot in changing ecological conditions, as well as anticipate the probability of impact. Road users particularly pedestrians are more susceptible to serious injuries in contrast with drivers in such collisions. Pedestrian Safety has therefore gain the attention of many researchers now a days. Safety components are intended to avoid the impacts and resulting casualties and injuries by offering advancements that alert the driver to potential issues before time. These safety advancements might lit up the car light, automatic braking, consolidate GPS traffic flow notifications, interface with cell phones, alert driver about other cars and dangers, keep the driver in the right lane, or show what is in blind corners. For effective road user protection, frameworks like Advanced Driver Assistance System (ADAS) and Intelligent Vehicles (IV) are built up that lessened the auto collisions by giving knowledge to drivers [2][3]. Pedestrian recognition is a critical and all the more difficult problem in the field of machine vision. The essential target of road user detection based vision frameworks is to avoid impact of vehicles with people while driving.

Although, several pedestrian detection algorithms have

been proposed so far, but still there is a need for an automatic system that can detect the human in urban environments that are more challenging as compared to highway traffic. Therefore, the need is to develop a system capable of discriminating humans from a captured image or video with high accuracy, precision, and low miss rate. In this paper, we have presented a motion vector based pedestrian detection system. We will discuss and analyze the experimental results of our proposed technique on two publicly available ETHZ and Caltech pedestrian benchmarks. Novel contributions of this work include:

- Development of a representation of human motion which is extremely efficient for detecting people.
- Implementation of proposed technique on two pedestrian benchmarks.
- Accurate human detection with smaller FPPI (false positive rate per image).
- A system for automatic pedestrian detection with Low average miss rate.

The rest of the paper is organized as follows: A comprehensive survey of related schemes can be found in Section II., Section III presents the proposed methodology employed for effective pedestrian detection, Section IV includes performance evaluation of our technique and Section V concludes the paper.

## II. LITERATURE REVIEW

In spite of advancements in pedestrian protection, numerous road accidents still happen all around the globe because of poor driving conditions (e.g. low light or mist) or a transient diversion of either the driver or pedestrian. A programmed framework to identify people in the surroundings of a vehicle is exceedingly desirable and is one of the fundamental concerns for both auto makers and researchers today. Related literature for pedestrian detection and recognition framework are reviewed from both the application and also, from hardware point of view.

From the application's point of view, pedestrians location can be utilized by intelligent vehicles and for surveillance videos and aided driving systems [4][5]. From technology perspective, detection and recognition of on foot people in front of the vehicle can be done by utilizing the perceptible and imperceptible light range, for example, visible, ultrasonic, infrared, sensors and radars [6][7].

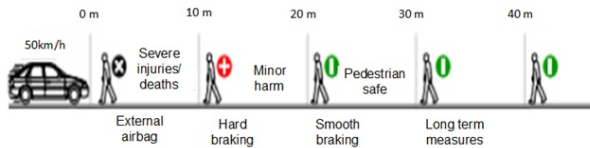


Fig. 1. PPS driving scenario [8]



Fig. 2. Difference in pedestrian appearances [9]

#### A. Pedestrian Protection Systems (PPS)

PPS are a special kind of intelligent frameworks concerned to pedestrian safety. It is a safety framework that typically recognizes moving as well as stationary pedestrians in front of the vehicle in order to perform braking actions or to provide knowledge to driver [2]. Any decline in the speed of the vehicle can intensely lower the fatality rate due to reduced kinetic energy of the approaching vehicle in case of impact. Pedestrians are having a 90% likelihood of lasting the accident due to vehicle impact coming at small speeds like 30km/h or below, however if the vehicle is approaching at the speed of 45km/h or above then in such cases there is less than half likelihood of surviving [2] as shown in Fig. 1.

#### B. Challenges to the Task of Pedestrian Recognition

These are the principle elements and difficulties that influence the performance of pedestrian detection frameworks:

1) *Appearance changeability*: Pedestrian constantly changes in appearances because of the color, shading, texture, size of the garments, and they move with different items and articles (boxes, bags, umbrellas) as appeared in Fig. 2.

2) *Inconsistency of the surroundings in which people appear*: Different type of environments include urban and congested city zones which are more complex to handle as compared to highways under different climate conditions and changes in light add irregularity in the information. Fig. 3 indicates distinctive surrounding environmental conditions.

3) *Variability in pedestrian shapes and postures*: Pedestrians may have diverse weights, postures, poses and tallness figures. Fig.4 shows such differences.

4) *Variability of the activities*: Different kind of actions which they may perform positions which they may have (stand,



Fig. 3. Inconsistency of the surroundings in which people on foot appear [2]



Fig. 4. Variable Body Postures [2]

run, walk, sit, and hands shake etc.) Besides, they can show up under different observing angles (longitudinal or sidelong positions).

5) *Motion of camera and Road user*: When both (walker and camera) are in moving condition, this marks detection and tracking more troublesome. The vast majority of the frameworks are focused on the high-hazard zone i.e. separation from 5m-25m to the camera [9]. However, 50m identification range speaks to a generally safe area that demonstrates an extraordinary support for PPSs in the long period crashes.

The problem of pedestrian detection has been approached both from hardware and software perspective, utilizing different sensors and creating numerous different algorithms for detection.

#### C. Hardware Based Solutions

Finding which sensor is most appropriate for pedestrian identification and detection is a major question especially for cluttered urban traffic environments. Today the most appealing sensors include Radar and Laser. Sensors that use visible light are efficient during day but do not work at night. In addition, there is an emerging attention towards infrared frameworks that are guaranteed to be extremely useful being less expensive as compared to other sensing mediums like Laser, Lidars and Radars and are not affected by the weather conditions or time of the day. From the hardware aspect, there are passive or active sensors [10].

Passive sensors capture light by using scan chips like CMOS or CCD cameras and can perform in infrared or visible spectra, consisting of all camera-based systems. Active sensors that recently have provided the good results are the laser sensors, particularly LIDARs, but these are too expensive for intelligent systems [11][12]. For that reason, mostly passive sensors are used in intelligent vehicles to detect pedestrians because they are cheaper and provide useful information from different clues like color and texture information [2].

#### D. Software Based Machine vision for Pedestrian Detection

Based on humans owned knowledge, vision-based pedestrian detection is a preferred choice. It can acquire much richer

information about the environment than laser scanner or radar [7]. For this purpose, various types of cameras have been used for the detection of pedestrians. Camera may be either still or moving that is installed on the vehicle. According to their working range, cameras can be divided in the electromagnetic spectrum. The range of visible spectrum (VS) is in 0.4-0.74 ( $\mu m$ ), near infrared is in 0.75-1.4 ( $\mu m$ ) and thermal infrared (TIR) covers 6-15 ( $\mu m$ ) [6], [7], [13]. Visible cameras are more commonly used because pedestrian detection is mainly focused on day time as compared to thermal infrared cameras that are used on night time [8].

The problem of pedestrian detection is to determine whether a local image area/section represents people or not and thus it is a typical two class classification problem. The detection process can be applied in two steps: feature extraction and classification. In the learning-based discriminative framework, various feature descriptors and classification approaches have been proposed for use in visible images [2], [3]. Some of the computer vision based approaches for pedestrian detection are:

1) *Feature Based Detection*: Detectors are developed that use feature information by extracting gradient features from the image to detect pedestrian in front of the vehicle within an unsafe range. The two popular gradient based feature detection methods are Covariance matrix (COV) and HOG descriptors [4], [14], [15], [16].

Such local gradient feature based detector was first presented by Dalal and Triggs [17]. They represented a human in the image as a thick web of Histogram of oriented Gradients then these feature maps were given to an SVM classifier to detect human. The image is first segmented into equal sized 16x16 blocks and each block further consists of 8x8 cells. These cells combined together to form a dense grid of with each cell representing edge orientation. Orientation gradients for each individual cell is quantized into 9 equal sized bins defined for 0-180 unsigned orientation each comprising of 20. The combined histograms represent a feature vector of normalized 1-D histograms from each block it is then given to an SVM classifier which classifies it as either a pedestrian or not. The dense overlapping grids, normalized histograms and non-smoothing gradients made HOG descriptor a better detector by decreasing the count of false positives. However HOG-feature extraction performs poorly for the cluttered images where it becomes difficult to detect images to create histograms.

Costea et al. [4] proposed a novel method for image recognition using word channels because of their high discriminating power. This approach uses image with a single size and a single classifier for each pedestrian sliding window scale. For extracting features this approach uses high level word channels inspired from codebook based techniques instead of low level pixel gradients used in HOG techniques. At each pixel level three descriptors HOG, LRP and LUV color channel values are computed. The authors use LUV color channel as the descriptor for computing feature vector. The computed results are matched to the visual codebooks. Three different word maps are generated with multiple word channels one channel per descriptor. Approach was found to have promising results on both INRIA and Caltech pedestrian datasets.

Richer image representation provides a good chance for improving the detector performance in image analysis. But

for richer image representations detectors require more computational time to process an image that is the improvement in recognition comes at the cost of more computational time. Dollar et al. [18] proposed that for richer image representation such finely sampled pyramids could be obtained by estimation. Experiments indicate that by extrapolating the features for coarsely sampled pyramids, we could estimate the features at any given scale and hence get a rich representation of the image inexpensively. The effectiveness of the proposed scheme is demonstrated with different detection architectures including integral channel features, aggregate channel features and deformable part models. Results indicate that the proposed method has the same detection rate as the current state-of-the-art but has decreased computational cost. However, this strategy could not be applied in cases where the image contains texture or white noise.

Dollar et al. [19] proposed a multi scale pedestrian detection technique. This technique uses feature approximation for the features at nearby scales by computing the feature at one scale. This approximation is shown accurate within entire scale octave. Algorithm thus computes features in the image once every half octaves and approximate features for the rest of scales resulting an overall speed up in the detection process with a little loss in detection accuracy.

Dollar et al. [20] proposed a novel method for pedestrian detection using Integral channel features. Using linear and non-linear transformations multiple image channels are computed. Features from these multiple channels are computed by summing the local rectangles. Using integral images features such as Haar wavelets, local histograms are computed. A feature is defined as the weighted sum of the integral channels. Much time of this algorithm is spent on constructing these channel features, making it a fast detector. The authors combined Histogram of Oriented gradients with LUV color channels coupled with a boosting classifier. Integral channel features when combined with a boosting classifier are proved as a fastest object detector. On Cal Tech dataset the detection rate of integral channel features was found to be 60% while that of its competent HOG is 50%. Results indicate that channel filters outperform other feature extraction techniques including HOG.

[21], [22], [23], [24] are the other feature based methods proposed by the researchers for detecting pedestrian in order to avoid or anticipate the likely hood of collision.

2) *Texture Based Detection*: Various texture based approaches like Local binary patterns (LBP) [25], [26] also provide good outcomes to the Pedestrian detection. The LBP advantages are its less computational complexity and multi-scaling [25]. Multi-block local binary patterns (MBLBP) [27], have been presented as efficient applications of traditional LBP for detection. Zhang et al. [28] propose that for accurate and fast detection of a human in an image scene or a video sequence, a robust detector is needed which can compute promising features in least computational time. They presented a set of effective features that can be computed easily and are robust to external noise. This set includes dense center-symmetric local binary patterns (CS-LBP) which captures the gradient information combined with texture details and pyramid center-symmetric local binary/ternary patterns (CS-LBP/LTP) which is more descriptive and computationally

efficient for real life applications. Experiments on INRIA pedestrian dataset indicate that the proposed features of CS-LBP when coupled with linear SVM give comparable results as HOG/SVM and pyramid CS-LBP when coupled with HIKSVMS outperforms the previous PHOG. The authors also suggest that by combining pyramid C-LBP with PHOG feature, detection performance could further be improved.

For the detection of a pedestrian in a still image Jiu Xu et al. [29] proposes a novel feature named Bidirectional Local Template Patterns (B-LTP). B-LTP is inherited from CS-LBP and HOT and thus combines their desirable properties. It takes texture properties from CS-LBP feature and gradient based properties from HOT. Moreover, B-LTP is a short length feature and thus it is cheaper to implement as well as cost less memory making it suitable for real time applications. This technique proposed a two directional template in which for each pixel, four templates are defined containing the pixel itself and its two center-symmetric neighbors. Results on INRIA dataset shows that B-LTP performs better than its competent features like HOG, HOT and COV in both speed and detection rate.

3) *Deep Learning for effective pedestrian Detection:* In the recent past, Deep learning has been applied to the domain of pedestrian detection problem which learns features in a supervised or unsupervised fashion and has shown very promising results. The input data moves from lower layers and is gradually transformed into higher level representations. The output features from the top layer is then given to classifier and the network is fine tuned with back propagation algorithm.

In his paper, Luo et al. [30] proposed a deep learning architecture named "Switchable Deep Network (SDN)" for pedestrian detection. His work focuses on using deep networks to model hierarchical features, stressed locations from multiple feature maps called saliency map and a mixed representation of body parts. SDN is an extension of the traditional convolutional neural network with the addition of multiple switchable layers. In order to model the complex visual postures the paper introduces a Switchable Restricted Boltzmann Machine (SRBM) that explicitly develops saliency maps at each level indicating if the pixel belongs to the background or a pedestrian and hence suppressing background clutters from discriminative regions containing pedestrian. Results indicate the state-of-the-art performance on the public pedestrian detection datasets.

4) *Template Based Detection:* This technique is used to find the features in a particular image region and then compare it with a standard template. Image is scanned to find a set of features representing pedestrian and then are compared with a template image. But these methods fail to handle articulations and occlusions in the scene and are computationally expensive [31]. The disadvantage of template based detection is that object occlusion is difficult to compute and high computational complexity.

5) *Deformable part based Detection:* In his paper, Yan et al. [32] presented an extension of the prevalent deformable part model [33] (DPM) called Multi task DPM (MT-DPM) which aims to explore the relation among multiple resolutions by combining an optimal DPM detector and resolution aware transformations. It takes the pedestrians from multiple resolutions and determines their commonness and differences

jointly. To map the pedestrians from different resolutions the model transforms them into a common space where a detector separates the pedestrians from the background. The global spatial assembly for example part configurations remains the same while the differences exist in the local features. The differences among these local features are reduced by mapping them from multiple resolutions to a common subspace and a detector is learned on these locally mapped features. The authors further develops a context model depending on the vehicle-pedestrian relation to improve pedestrian detection by reducing the false positive rates especially in crowded scenes. Results indicate a reduction of miss rate to 60% for Caltech dataset which outperforms the recent state of the art.

6) *Infrared Thermal Imaging for Pedestrian Detection:* Effective pedestrian detection and tracking algorithms in visible spectrum have found many important applications from video surveillance to intelligent vehicles. However, under certain circumstances (e.g., in nights or bad weathers), sensing in visible spectrum becomes infeasible or severely impaired, which calls for the imaging modalities beyond visible spectrum. In particular, the cost of thermal sensors has reduced dramatically in the past decades.

Infrared imaging is here to rescue the environments in which we have little or no light. Dai et al. [11] proposed a Generalized expected maximization (EM) algorithm using IR imagery. The image is first segmented into a layered structure consisting of foreground and background layer. In the second pass using the shape and appearance details a pedestrian is traced from the foreground layer. Shape based classification is performed by SVM and appearance based localization is done through principal component analysis(PCA). Similarly for a video , the sequence is first divided into segments called shots and pedestrian detection through EM algorithm is then applied within each shot. The pedestrians present in the same shot are identified through a graph matching technique. The algorithm performs quiet well in case of crowded scenes and does not require prior assumptions about the motion trajectory. Experimental results showed the overall accuracy of 88%.

7) *Machine Learning Based Methods:* The accuracy of the detection system usually comes at the cost of high false positive rate. In order to reduce this false detection rate Z. Wang et al. [34] presented a two stage machine learning algorithms based approach for efficient and accurate pedestrian detection. This approach is based on highly efficient combination of cascade AdaBoost detector and vector function link net derived from machine learning domain. By using multi-scale sliding window detectors; all sub windows extracted from a still image and are normalized and resized. Then the two detectors cascade AdaBoost [35] deetctor and random vector functional-link net [36][37] are applied simultaneously on this candidate feature set to check if it is a pedestrian or not. Experiments with four datasets have shown that this technique outperforms other methods in terms of detection accuracy and false positive rates.

Behera et al. [38] presented a real time vision based image segmentation algorithm for accurate pedestrian detection in day time. The image is scanned in all directions for finding the edges. Before segmentation, the edges are first linked by an edge linking algorithm. The correctness of the edge map is required for accurate segmentation. After segmentation the image is divided into foreground and background segments. In

order to boost the probability to find the presence of pedestrian, a combination of head and leg edges are used. Using these head and leg edge patterns the whole pedestrian image is reconstructed. Accuracy could be further improved by applying a classifier on the extracted segments. Results indicate that this algorithm performs well on real images for accurate pedestrian detection. In order to improve the accuracy of a pedestrian detector much work has already been done in the current state-of-the-art, the paper by Smedt et al. [39] in this regard presented a generic framework to combine multiple pedestrian detectors in an optimal and efficient manner. Each pedestrian detection technique uses a different set of candidate features. Highly accurate results could be achieved by an intelligent combination of these features from multiple detectors. The authors used the simple AND OR combinations of multiple detectors and by using performance measures determine the best combination which has the highest yield in terms of detection accuracy. However combining multiple detectors; results into long computational time but results showed that an improved accuracy is obtained by hiring this optimal combination approach as compared to the current state-of-the-art detection methods.

Image feature description can be improved significantly by using HOG features based on variant scale blocks. This idea was presented by Hoang et al. [40] who suggested that without restricting HOG blocks a comprehensive feature space is obtained with the help of which highly distinguished features can be obtained for classification in the next step. Image is first segmented into grid windows and affine transformations from each window are obtained after which optical flow from each transformed window are extracted. After morphological processing, correlated features of human shape are obtained as candidate regions within each window. In the next step HOG features from each segmented window are obtained in order to detect a human by using SVM classification. Experiments showed that the proposed detector gives 5% improved results as compared to standard HOG using SVM.

Hongyan Li [41] proposed a new method of segmentation and detection of objects in a video. Their algorithm uses mean, variance and standard deviation as features calculated from gray scale multi-frame images. These features are used to train SVM that views the categorization of pixels as a binary classification task. Trained SVM classifies a pixel given to it as either a static background pixel or a foreground moving object pixel. Accuracy of this SVM classifier could be significantly improved by customizing its kernel function and other parameter values.

8) *Motion Based Detection:* Use of the object motion provides extra information but the change of position make the process of detection more complex and problematic [31]. One of the popular method in this regard is Background subtraction in which the image is segmented into foreground and background layers but this is only possible in the cases in which video is captured by a fixed camera. However, this has an apparent disadvantage for pedestrian detection from the automobile because the moving vehicle provides a continually dynamic background. Therefore, motion-based pedestrian detection algorithms do not work as a primary detection approach for a camera that is mounted on a moving vehicle [9]. Sparse scene flow information is used to detect

objects using stereo cameras and optical flow information in. [42] has extracted interest points from consecutive video frames; using flow information the complete scene flow is constructed to model the movement of the background. Scene elements whose motion pattern varies from this background flow model are considered as distinct objects. Interest points belonging to adjacent segments represent a single rigid object. The proposed method employs a class independent approach for object detection using stereo cameras and optic flow. Experiments indicate that the proposed method outperforms the previously known techniques that only use optical flow information. Moreover, these solutions can work only for those classes on which the detector was trained during training phase.

Hariyono et al. [43] presented a novel method for moving pedestrian detection through moving camera using motion information and HOG features. After segmenting the regions that represent same motion vectors different moving objects are extracted. In order to differentiate a pedestrian from non-pedestrian HOG features are extracted from candidate segments. This feature vector is given as an input to the Linear SVM classifier that classifies the given segment in an image or video frame as pedestrian or non-pedestrian. Experiments reveal an outstanding performance on ETHZ pedestrian dataset as compared to the original HOG approach. Detection rate obtained was 99.3% with 0.09 false positive rate. Another method is CodeBook method [31]. A codebook collects a series of code words or color values for every pixel of background. After that, codewords will found out what color each pixel have and from that it determine the pixels of background. Advantage of this method is that it can handle dynamic scenes [31].

From the cited literature, the issues pertaining to the real-time pedestrian detection problem are as follows:

- Occluded and continuously changing backgrounds put a limit on the detection rate.
- Feature extraction techniques alone perform poorly for the cluttered images.
- Deep learning based methods have long learning and detection times.
- Template based approaches are affected by occluded environments.
- Relative motion between camera and pedestrian marks detection and tracking more troublesome.

A system with the following properties is therefore desired:

- Detect pedestrian in real time accurately and reliably with lower false positives.
- Detect people regardless of the differences in their postures and appearances.
- Improved system accuracy with small miss rates.
- Detect pedestrians in all type of environmental conditions.
- Detection of people even if there is relative motion between pedestrian and camera.



This research will combine the motion vectors method with HOG feature extraction technique for improved detection results in images and videos. Datasets used are ETHZ pedestrian dataset and Caltech pedestrian dataset.

### III. PROPOSED METHODOLOGY

Motion is a natural property of the world and is a rich foundation of data that supports a wide range of visual responsibilities. Identification of moving objects in videos is vital for numerous computer vision based applications, including action acknowledgment, activity recognition, and car safety. The issue of motion based object detection can be separated into two sections:

- 1) Identifying moving subjects in each video frame.
- 2) Associating the recognitions relating to the same subject over time.

#### A. Work Flow

The proposed methodology to detect the pedestrian in videos as indicated in systems block diagram below in Fig. 5 is composed of the following sequences of steps.

1) *Acquire images and videos:* The first step to create a successful system for pedestrian detection is the selection of right dataset. There are many standard publicly available datasets for solving the problem of moving person detection. We have used ETHZ and Caltech datasets for our research. The challenging part of this research is to compensate the motion of both camera and pedestrian while reducing the miss rate in order to achieve accurate detections with small false positives.

2) *Compute motion vectors to extract moving objects in a scene :* In order to make the task of detection easy we have opt for detection through motion vectors because it separates the dynamic objects in the foreground from static background in the very first step. Here in our research, first of all we have computed motion vectors from the each video frame and by associating these detections same object over time is identified. This gives a set of bounding boxes around each of the moving object in the scene as the output.

3) *Extract features from the computed flow vector:* Feature extraction is the process of selecting relevant attributes in the data that are used for the construction of the model. In the next step, by computing HOG features from each bounding box we are able to learn a single feature vector which is the combination of motion information from the previous step and HOG features. This feature set is given to the classifier in the next step that will learn the features of Pedestrians with the help of the input feature vector.

4) *Feed this feature set to a classification module:* This feature set is given to SVM/AdaBoost classifier one by one to train the classifier and learn the features of a human so as to discriminate human from non-human objects in the test videos.

5) *Detect pedestrians in the test set:* Detection is performed on test videos and performance of the detector for each classifier in terms of accuracy and average miss rate is computed and compared with state-of-the-art methods of pedestrian detection.

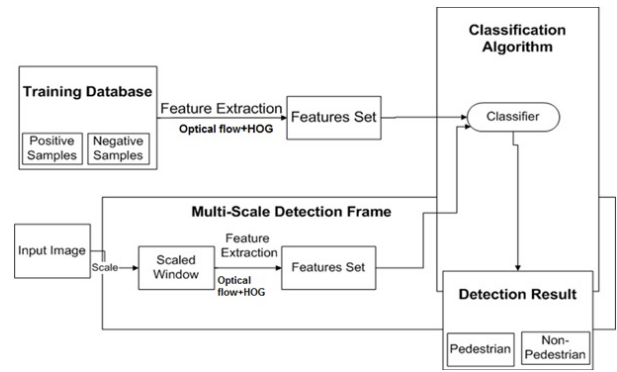


Fig. 5. Block diagram of proposed system



Fig. 6. Images from ETHZ and Caltech Datasets

### IV. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

We ran the experiments on Intel core i3 with 4GB of memory using MATLAB Release R2014 b as integrated development environment and runtime platform. We have evaluated our proposed methodology on two benchmark pedestrian datasets namely ETHZ and Caltech pedestrian benchmarks.

ETHZ is a popular and most commonly used pedestrian dataset. It consists of videos made from AVT Marlins F033C camera, it contains video frames of the size 640x480 each with a frame rate of approximately thirteen to fourteen frames per second. The dataset contains three setups consisting of three set of videos each. Second dataset is the Caltech which is the largest dataset amongst other datasets. It consists of 10 hours of 640 x 480 videos in an urban environment. These videos are taken from a CCD camera mounted on a vehicle. All these datasets are publicly available and some of the images of these datasets are shown in Fig. 6.

#### A. Evaluation Metrics

Evaluation of our work is done based upon the following evaluation metrics. The terms are defined as follows:

We trained our detector by using positive and negative samples once from ETHZ and next time with samples from Caltech training sets. First the training of SVM is performed and a model is created, one for each dataset. Similarly, the procedure is repeated to create AdaBoost training models. For the rest of our experiments, we test our pedestrian detectors on the reasonable subset of videos from both ETHZ and Caltech test sets.

TABLE I. METRICS FOR EVALUATING CLASSIFIERS PERFORMANCE

Metrics	Definition/Formula
Miss Rate	= 1- Recall = No. of FP in positive samples/ Total no. of positive samples
FPPI	= False positives/ Total images
Detection Rate	= TPR= Number of correctly classified pedestrians
Precision	= Fraction of true positives/ All positives
F1 Measure	= $2 * TP / (2*TP + FP + FN)$
False Detection Rate	= 1- Precision

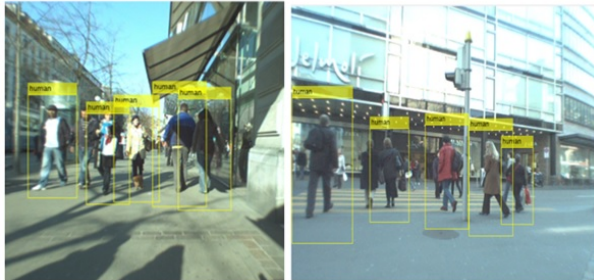


Fig. 7. Detectors Detection Results

### B. Testing the Detector Performance

When training process is completed, the next step is to test the performance of classifiers by giving any unknown images to the detector. The input image can be either a single image or a cell array of images. For detection, we slide a window over the whole image and consider the multiple window strides. For each video from the test set the experiment is performed with 8x8 (default) and 6x6 window stride. Detector uses this measure to slide the window over the image, smaller strides produce better detection rate. After this, the detector returns different bounding boxes in the form of [x, y, w, h, score]. Confidence of detection is measured by its score. Only detections with the score above threshold are returned. A few detection results are shown below in Fig. 7:

Confidence is high for the pedestrians in front of the camera or within a suitable range. Those already standing in the safe range are given small scores by the detector as shown in Fig 8.

### C. Experimental Results

1) *Detection:* Detection results on both datasets for each classifier are shown below. Detector creates a bounding box



Fig. 8. Testing the detector (a)Input test image (b)Detection scores

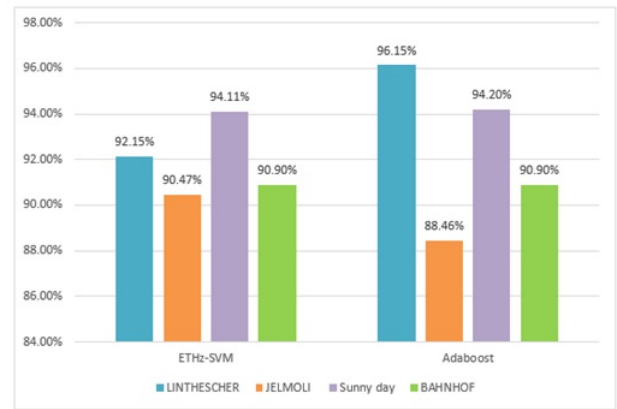


Fig. 9. Detection Results on ETHZ Pedestrian benchmark

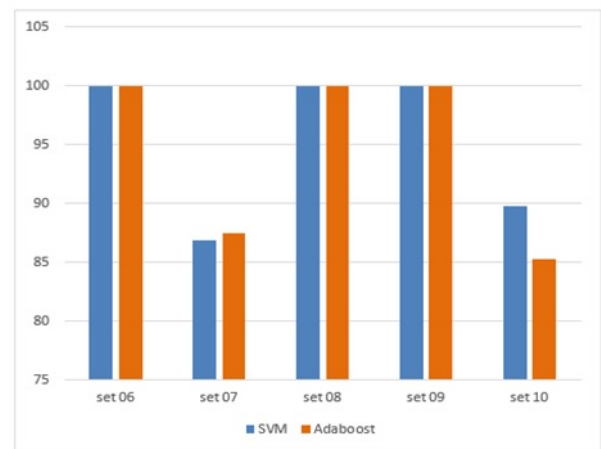


Fig. 10. Detection Results on Caltech Pedestrian benchmark

around a human. HOG and flow features from these bounding boxes are computed. Detector marks each bounding box with a score that represent the detection confidence. Confidence is high for the pedestrians in front of the camera or within a suitable range. Those already standing in the safe range are given small scores by the detector.

Detection results for each classifier on ETHZ benchmark are shown in Fig. 9. On average we have achieved 91.82% detection accuracy with SVM while AdaBoost provided 92.96% accuracy.

Detection results for each classifier on Caltech benchmark are shown in Fig. 10. On average we have achieved SVM provided 95.33% and AdaBoost provided 94.5% accuracy.

2) *Evaluation of Results:* In the second phase the detector performance is evaluated for both datasets by comparing the number of detections with the ground truth annotation which is provided with each test video. These annotation files contain information in the form of bounding boxes coordinates for each person present in the scene for each frame. However, for ETHZ, ground truth annotation is provided for every 4th frame in the video. Miss rate and the detection rates are computed here with the possible number of false positives per image and overall accuracy, recall and precision are computed for both ETHZ and Caltech datasets.

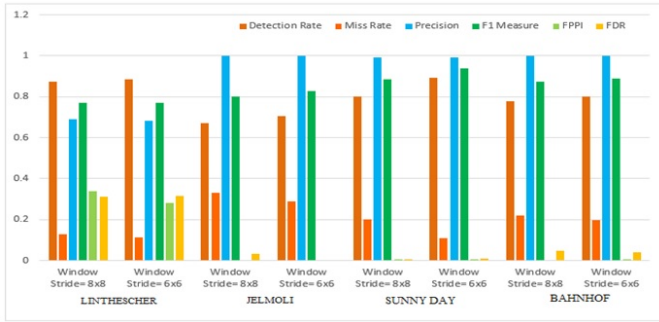


Fig. 11. Evaluation Results on ETHZ Pedestrian benchmark

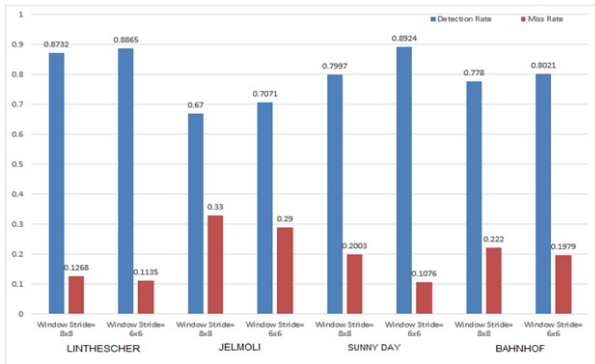


Fig. 12. Detection Rate VS Miss Rate on ETHZ Dataset

valuation with respect to the ground truth data is performed and results of evaluation are shown in Fig. 11. We have achieved an average miss rate of 17.7% on ETHZ Test set. Miss rate is calculated for each video with respect to detection rate as shown in Fig. 12.

Results of evaluation on Caltech are shown in Fig. 13. We have achieved an average miss rate of 14.22% on Caltech Test set. Miss rate is calculated for each video with respect to detection rate for each test set from Caltech benchmark as shown in Fig. 14.

### V. PERFORMANCE COMPARISON

In our proposed work we have developed a pedestrian detection system by combining two features that is HOG and motion vectors. The two features are concatenated to get a

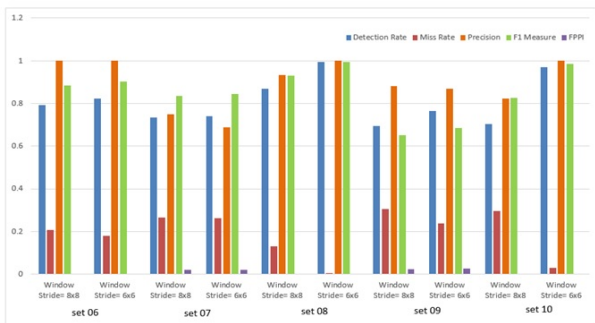


Fig. 13. Evaluation Results on Caltech Pedestrian benchmark

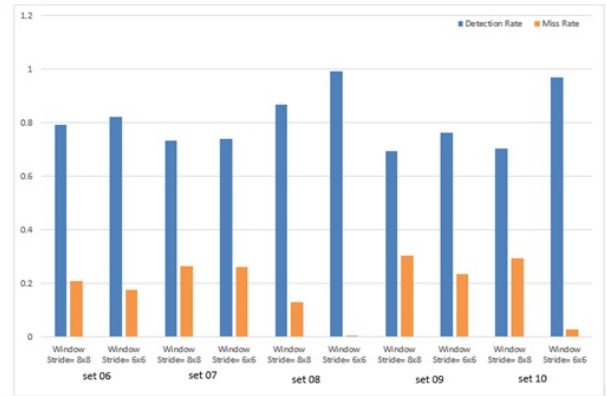


Fig. 14. Detection Rate VS Miss Rate on Caltech Dataset

single feature vector for the training of the classifier in the next step. Two classification models created during training phase for each classifier (SVM and AdaBoost) are used to discriminate human from non-human in the test videos and the detection results and accuracy is computed. Average miss rate computed from our proposed methodology is compared with those present in existing literature as indicated in Table 4.2. Our detector yields significant performance improvement as compared to the baseline HOG detector.

### VI. CONCLUSION

The problem of efficient pedestrian detection is studied in this research. This work has presented a novel technique for underlying problem of pedestrian detection by incorporation of motion information with feature extraction technique of HOG. Based on our experiments, we observe that the performance of pedestrian detection yields significant improvement with the use of motion vectors. Furthermore, implementation parameters also play an important role to achieve the best detection performance. We have achieved 17.7% miss detections on ETHZ and 14.22% miss detections on Caltech with a window stride of 6x6. Our future research objective include pedestrian tracking over multiple frames by using the optical flow based motion estimation. Currently, we have used datasets made in different lightening conditions during day time only and the pedestrian detection at night time is not included in the current scope of work. In future, we will also include the datasets that contain videos made at night. Moreover, the proposed scheme can also be extended to work in different weather conditions including rain, snow etc. Similarly, by combining more features with currently used features can help further improve the detector performance.

### REFERENCES

- [1] National Highway Transportation Safety Administration, "TRAFFIC SAFETY FACTS 2012 Data," <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812018>, 2014, online; accessed 27 July 2016.
- [2] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 4, pp. 743–761, 2012.
- [3] Y. Yang, R. Mingwu, and Y. Jingyu, "Obstacles and pedestrian detection on a moving vehicle," *International Journal of Advanced Robotic Systems*, vol. 11, 2014.



TABLE II. PERFORMANCE COMPARISON WITH EXISTING LITERATURE

	Paper	Year	Techniques used	Average Miss Rate		
				INRIA	CalTech	ETHZ
1	Dalal [17]	2005	HOG, SVM	46%	68.46%	–
2	Dollár [20]	2009	HOG + LUV Channels, AdaBoost	22.18%	56.34%	–
3	Wang [26]	2009	HOG + LBP, SVM	39.10%	67.77%	–
4	Dollar [19]	2010	Channels, AdaBoost	21%	–	–
5	Yan [34]	2013	MT-DPM, HOG, Latent SVM	–	37.64%	–
6	Benenson [35]	2013	Channels, SVM	13.53%	48.35%	–
7	J. J. Lim [36]	2013	Channels, Random Forest	13.32%	–	–
8	P. Luo [30]	2014	SDN, HOG, SVM	–	37.87%	–
9	Zhang [24]	2014	HoG + LUV, AdaBoost	14.30%	34.60%	–
10	Katamari [37]	2014	Square Chns + DCT + Optical Flow, AdaBoost	–	22%	–
11	Proposed Scheme		HOG + Optical Flow, SVM, AdaBoost	–	14.22%	17.70%

- [4] A. Daniel Costea and S. Nedeveschi, "Word channel based multiscale pedestrian detection without image resizing and using only one classifier," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2393–2400.
- [5] Y. Wei, Q. Tian, and T. Guo, "An improved pedestrian detection algorithm integrating haar-like features and hog descriptors," *Advances in Mechanical Engineering*, vol. 5, p. 546206, 2013.
- [6] C.-Y. Chan and F. Bu, "Vehicle-infrastructure integrated approach for pedestrian detection: feasibility study based on experimental transit vehicle platforms," in *Transportation Research Board 85th Annual Meeting*, no. 06-0118, 2006.
- [7] C.-Y. Chan, F. Bu, and S. Shladover, *Experimental vehicle platform for pedestrian detection*. California PATH Program, Institute of Transportation Studies, University of California at Berkeley, 2006.
- [8] D. G. Gomez, *A global approach to vision-based pedestrian detection for advanced driver assistance systems*. Universitat Autònoma de Barcelona, 2010.
- [9] D. Gerónimo and A. M. López, *Vision-based pedestrian protection systems for intelligent vehicles*. Springer, 2014.
- [10] A. Miron, A. Bensrhair, R. I. Fedriga, and A. Broggi, "Swir images evaluation for pedestrian detection in clear visibility conditions," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, 2013, pp. 354–359.
- [11] C. Dai, Y. Zheng, and X. Li, "Pedestrian detection and tracking in infrared imagery using shape and appearance," *Computer Vision and Image Understanding*, vol. 106, no. 2, pp. 288–299, 2007.
- [12] C. Premebida, O. Ludwig, and U. Nunes, "Lidar and vision-based pedestrian detection system," *Journal of Field Robotics*, vol. 26, no. 9, pp. 696–711, 2009.
- [13] A. Bartsch, F. Fitzek, and R. Rasshofer, "Pedestrian recognition using automotive radar sensors," *Advances in Radio Science*, vol. 10, no. B. 2, pp. 45–55, 2012.
- [14] B. Besbes, A. Rogozan, A.-M. Rus, A. Bensrhair, and A. Broggi, "Pedestrian detection in far-infrared daytime images using a hierarchical codebook of surf," *Sensors*, vol. 15, no. 4, pp. 8570–8594, 2015.
- [15] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *European conference on computer vision*. Springer, 2006, pp. 428–441.
- [16] R. Ranjan, V. M. Patel, and R. Chellappa, "A deep pyramid deformable part model for face detection," in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE, 2015, pp. 1–8.
- [17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.
- [18] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532–1545, 2014.
- [19] P. Dollár, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *BMVC*, vol. 2, no. 3. Citeseer, 2010, p. 7.
- [20] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," 2009.
- [21] S. Wang, Z. Han, L. Zhu, and Q. Chen, "A novel approach to design the fast pedestrian detection for video surveillance system," *International Journal of Security and Its Applications*, vol. 8, no. 1, pp. 93–102, 2014.
- [22] C.-H. Chuang, S.-S. Huang, L.-C. Fu, and P.-Y. Hsiao, "Monocular multi-human detection using augmented histograms of oriented gradients," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.
- [23] W. Xing, Y. Zhao, R. Cheng, J. Xu, S. Lv, and X. Wang, "Fast pedestrian detection based on haar pre-detection," *International Journal of Computer and Communication Engineering*, vol. 1, no. 3, p. 207, 2012.
- [24] S. Zhang, C. Bauckhage, and A. B. Cremers, "Informed haar-like features improve pedestrian detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 947–954.
- [25] S. Yang, X. Liao, and U. Borasy, "A pedestrian detection method based on the hog-lbp feature and gentle adaboost," *International Journal of Advancements in Computing Technology*, vol. 4, no. 19, 2012.
- [26] X. Wang, T. X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 32–39.
- [27] A. Halidou, X. You, and B. Bogno, "Pedestrian detection based on multi-block local binary pattern and biologically inspired feature," *Computer and Information Science*, vol. 7, no. 1, p. 125, 2014.
- [28] Y. Zheng, C. Shen, R. Hartley, and X. Huang, "Effective pedestrian

- detection using center-symmetric local binary/trinary patterns,” *arXiv preprint arXiv:1009.0892*, 2010.
- [29] J. Xu, N. Jiang, and S. Goto, “Pedestrian detection based on bidirectional local template patterns,” in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE, 2012, pp. 400–404.
- [30] P. Luo, Y. Tian, X. Wang, and X. Tang, “Switchable deep network for pedestrian detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 899–906.
- [31] V. Bandarupalli, “Evaluation of video based pedestrian and vehicle detection algorithms,” 2010.
- [32] J. Yan, X. Zhang, Z. Lei, S. Liao, and S. Z. Li, “Robust multi-resolution pedestrian detection in traffic scenes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3033–3040.
- [33] H. Cho, P. E. Rybski, A. Bar-Hillel, and W. Zhang, “Real-time pedestrian detection with deformable part models,” in *Intelligent Vehicles Symposium (IV), 2012 IEEE*. IEEE, 2012, pp. 1035–1042.
- [34] Z. Wang, S. Yoon, S. J. Xie, Y. Lu, and D. S. Park, “A high accuracy pedestrian detection system combining a cascade adaboost detector and random vector functional-link net,” *The Scientific World Journal*, vol. 2014, 2014.
- [35] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. 1–511.
- [36] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, “Learning and generalization characteristics of the random vector functional-link net,” *Neurocomputing*, vol. 6, no. 2, pp. 163–180, 1994.
- [37] G. H. Park and Y. H. Pao, “Unconstrained word-based approach for off-line script recognition using density-based random-vector functional-link net,” *Neurocomputing*, vol. 31, no. 1, pp. 45–65, 2000.
- [38] J. Xu, L. Yao, and L. Li, “Argumentation based joint learning: A novel ensemble learning approach,” *PloS one*, vol. 10, no. 5, p. e0127281, 2015.
- [39] F. De Smedt, K. Van Beeck, T. Tuytelaars, and T. Goedemé, “The combinator: Optimal combination of multiple pedestrian detectors,” in *ICPR, 2014*, pp. 3522–3527.
- [40] V.-D. Hoang, M.-H. Le, and K.-H. Jo, “Hybrid cascade boosting machine using variant scale blocks based hog features for pedestrian detection,” *Neurocomputing*, vol. 135, pp. 357–366, 2014.
- [41] H. Li and J. Cao, “Detection and segmentation of moving objects based on support vector machine,” in *Information Processing (ISIP), 2010 Third International Symposium on*. IEEE, 2010, pp. 193–197.
- [42] P. Lenz, J. Ziegler, A. Geiger, and M. Roser, “Sparse scene flow segmentation for moving object detection in urban environments,” in *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE, 2011, pp. 926–932.
- [43] J. Hariyono, V.-D. Hoang, and K.-H. Jo, “Moving object localization using optical flow for pedestrian detection from a moving vehicle,” *The Scientific World Journal*, vol. 2014, 2014.