A Study on Automatic Pedestrian Detection Using Computer Vision

Najila. A.L[1], Shijin Knox. G.U[2]

[1]*PG Scholar, Network Engineering, Govt. Engineering College, Barton Hill, Thiruvananthapuram, India*
[2]*Asst. Professor, Dept. of Information Technology, Govt. Engineering College, Barton Hill, Thiruvananthapuram, India*

---***---

Abstract - *Pedestrian Detection in images or videos plays a vital role in ADAS (Advanced Driver Assistance System) and AV (Autonomous Vehicle). The problem of pedestrian detection has been considered as a vital one in the domain of Computer Vision and Pattern Recognition. It is formulated as the problem of detecting and tracking the pedestrians in images or videos. Many techniques have been proposed in the past to solve this problem accurately and efficiently. However, the challenges occur at three different levels of pedestrian detection, such as video acquisition, human detection and its tracking. The challenges in video are illumination variation, abrupt motion, clutter present in the background, etc. Pedestrian detection and tracking challenges such as body attire and pose, occlusion, different illumination parameters in different scenarios, etc. These results in reducing the rate of attaining high accuracy. This paper provides a study of some well-known existing techniques for pedestrian detection and tracking in images. Some benchmark datasets are discussed, performance evaluation metrics are also presented.*

***Key Words*: Autonomous vehicle, Pedestrian detection, Pedestrian tracking, Pattern recognition, etc.**

1.INTRODUCTION

Pedestrian detection refers to the task of detecting the pedestrians and tracking them with the help of the rectangular boxes known as bounding boxes. Fig. 1 shows an example of pedestrian detection. Potential applications include pedestrian detection for smart vehicles, visual surveillance, robotics, autonomous driving etc. There is growing demand in the development of intelligent video surveillance system. Pedestrians are among the most vulnerable road user in India. Data given in Table I represent the pedestrians are among the major road users who are killed in the road accidents [1]. Most of the accidents occur during at night or at poor visibility. Some other reasons for accidents are inefficient driving, defective vehicles, lack of education, roads are occupied by the parked vehicles etc. In automotive safety, driver assistance domain system provides safety of pedestrians while using autonomous vehicles. Among all of these, one of the important computer vision problems is to detect the pedestrians. Detecting the pedestrians is difficult due to body attire and pose, occlusion, different illumination parameters in different scenarios, etc.

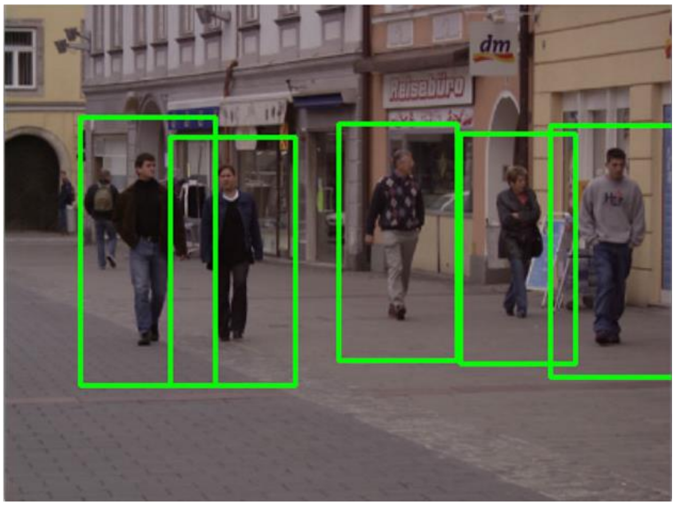

Fig -1: Example of detecting pedestrians.

2. CHALLENGES IN PEDESTRIAN DETECTION

Many reasons that are responsible for making pedestrian detection difficult which include variation in pedestrian's body attire and pose, occlusion, different illumination parameters in different scenarios and clutter present in the background poses challenges in attaining high accuracy. And also different appearances such as color, textures, carry different type of the objects such as bags, bicycles, variety of different activities such as standing, sitting, handshake position, different environmental conditions, low resolution images.

Behaviour and characteristics of pedestrians are always varying. In [2] earlier studies provide significant facts about pedestrian demographic characteristics (such as age, gender) and how these characteristics influence road crossing behavior. Children are found to cross at higher speeds than others. There is not much difference between the average crossing speeds for adults and older people. Males have a tendency to show more hazardous road crossing behavior than females due to less waiting time [2]. It also depends on crossing time, pattern, speed and waiting time. Majority of pedestrians, crossing time varies between 4 seconds to 10 seconds and their waiting time varies 1 second to 6 seconds. Mainly two types of crossing patterns are there- perpendicular movement (average crossing speed 1.36 m/sec) and Oblique movement (average crossing speed 1.98 m/sec). Crossing speed of males is found to be 1.85 m/sec and female was 1.67 m/sec.

Road-user category	No.of people killed during 2018	No.of people killed during 2019
Pedestrian	22,656	25,858
Bicycles	3,673	4,196
Two-wheelers	55,336	56,136
Auto-Rickshaws	6,629	6,655
Cars, Taxis, Vans, LMVs	25,115	23,900
Trucks/Lorries	15,150	13,532
Buses	8,164	6,529
Other Motor Vehicles(including e-rickshaw)	11,109	3,065
Others(Animals drawn vehicle, cycle rick-shaws, hand carts, other persons)	3,585	11,242
Total	1,51,417	1,51,113

Table -1: Total number of people killed in India due to road accidents during 2018-2019.

3. SENSORS USED FOR PEDESTRIAN DETECTION

Pedestrian detection in real time is one of the most crucial parts. The current researches are mainly focused on the application of visual sensors [5], infrared IR imaging sensors [3] [4], and radar sensors to aware of pedestrians and obtain their safety state information for realizing active pedestrian protection. Visual sensors capture detailed descriptions of a road scene but does not depend temperature variation and thus it remains constant in all weather conditions. They depend on light so their performance is usually unacceptable at night. On the other hand, infrared sensors do not depend on illumination and can be used both day and night with little difference. Normal infrared sensors are not effective under high temperature conditions and clothes affect the pedestrians' thermal footprint [4]. LiDAR and radar have caused serious accidents because they fail to recognize objects accurately in bad weather and at night.

4. PEDESTRIAN DETECTION FEATURES

Pedestrian detection can be considered as a two step process which includes feature extraction and its classification. There are different types of the features that have been used for detecting the pedestrians and are discussed in this section.

a) Gradient-based features: Gradient based features use gradients as features for extracting the useful information from the images. They are better to capture the shape of an object. In [5] the author Dalal et al. proposed a shape descriptor called Histogram of oriented gradients which is based on the concept that the local appearance and shape of the object can be characterized by the distribution of the gradients and edge directions. First step to calculate the gradients with the help of masks. Second step, each pixel computes a weighted vote for an edge orientation histogram channel based on the orientation of the gradient element centered on it, and the votes are accumulated into orientation bins cells. In the next step the cells are organized into blocks and normalization is performed so that the effects of changes in the illumination and contrast differences between the foreground and the background are minimized. The feature descriptor is then calculated as combined vector formed from values from all the blocks. The combined histograms represent a feature vector which is given as input to the Support Vector Machine (SVM) classifier which classifies it as either a pedestrian or non-pedestrian. The dense grids, normalized histograms and non-smoothing gradients improved the performance of HOG descriptor. However, the performance of the HOG-feature extraction degrades in cluttered images.

b) Motion-based features: Motion based features are also helpful in detection of pedestrians as the motion of pedestrians is different from other objects. In [6] built a detector based the rectangular filters that they defined earlier [7]. The technique combines motion information with the appearance information. They defined different motion filters and Adaboost was used for feature selection and building the classifier. Combined motion-based descriptors extracted from optical flow with Histogram of Oriented Gradient descriptors [8]. Then the feature vector is given as input to Linear SVM Classifier.

c) Texture-based features: Texture based features are used for capturing the presence of pedestrians. There has been an increase in the performance accuracy, when the texture is combined with other approaches. Local Binary Patterns (LBP) is one of the texture-based approaches for classifying the textures [9]. LBP was used for describing the pedestrian appearance [10]. Performances of these features are not affected by the noise. Due to large number of bins the dimensionality of the feature vector is very high. Different variants of LBP exist in literature. Center Symmetric Local Binary Pattern (CSLBP) is good for capturing both texture and shape information [11]. These features have orientation information of the pixel intensity, so they are sensitive to the lightning conditions. Local Trinary Patterns (LTP) is another modified version of LBP which is less sensitive to noise.

d) Shape based features: Shape based features plays a vital role in detecting the pedestrians. Shape-based object detection which used for distance comparing the feature image with the template image. Lesser the distance get better match. In [12] described a technique used for shapelet features were learnt from information provided by gradients that differentiates pedestrian class from non– pedestrian class. Adaboost algorithm was used for selection of the subset of the features and for classification. Another technique presented in [13] based on improved Shape

Context feature which is robust to image deformation for detecting pedestrians present in the infrared images.

e) Part-based features: Pedestrians are modeled as a collection of parts. Part based models refers to a broad class of detection algorithms and are used on images. In various parts of the image are used separately in order to determine if and where an object of interest exists. In [14] modeled the humans as components and combined them using the joint probability body model. These features are better than Haar wavelets and perform better in handling occlusion. The likelihood score is attached to the assembly of the parts that improves the human detection. Adaboost is used for learning and selecting the best features.

5. PEDESTRIAN DETECTION APPROACH

In general pedestrian detection can be performed as a two step process which includes feature extraction and its classification to check whether the image or video contains pedestrians or not. This process is defined as a flow chart in fig.2. Preprocessing the input image is a process that is useful for removing the unwanted noise, brightness improving, removing all blur etc. Then the ROI is found where the probability of finding pedestrians is high. Using different techniques, the features that are relevant to the pedestrians are found. Then feature vector obtained is given as input to the classification algorithm which has been provided training for the learning different features for prediction regarding presence or absence of pedestrians. If the pedestrian is present in the image or video then a bounding box which is generally a rectangular box that specifies the location of pedestrian is formed around pedestrian.

Earlier techniques are used the sliding window approach. Here image is scanned across on several windows. These windows are extracted at various positions and scales. Pedestrians are present only in some of the candidate windows. This approach is mostly useful in static images. For video sequence, motion is associated so background removing techniques are useful. In this technique the moving objects are identified by computing the difference between the current image and the reference background image. Different techniques are used which depending on the applications.

6. APPLICATION AREAS

Pedestrian detection plays a vital role in many application areas. Currently it is used for surveillance because several cameras are installed at the parking areas, railway stations,

Detection process Training process

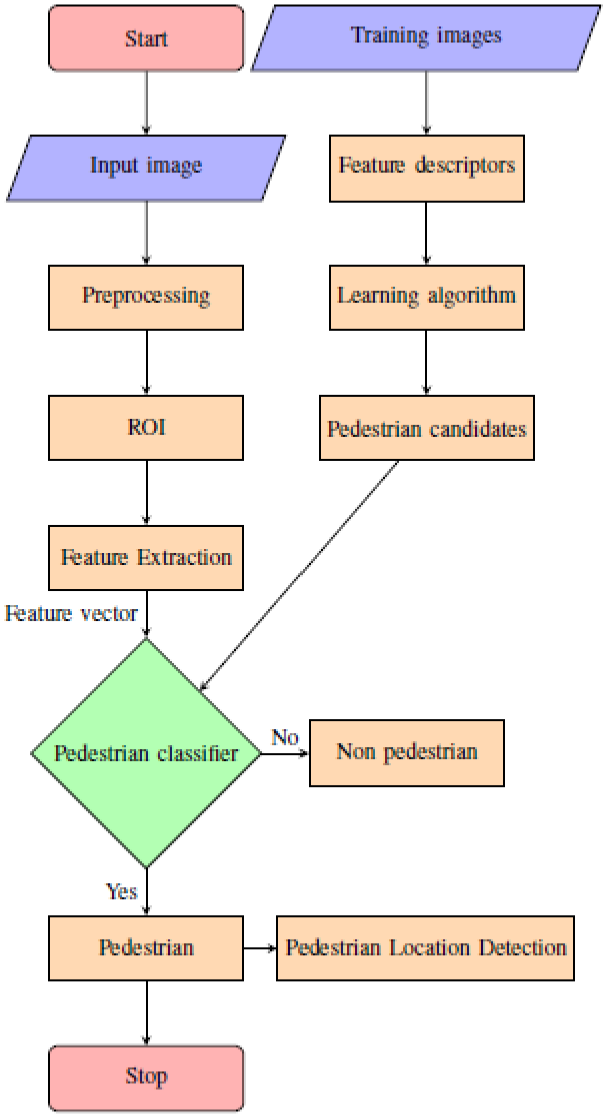

Fig -2: Flowchart of pedestrian detection.

airports, malls etc. Now it is also used in autonomous vehicles. Fig.3 shows an example view of pedestrians from autonomous vehicle. Early detection of the pedestrians can reduce the injuries and death of pedestrians and also the passengers in the vehicle. For these purpose cameras can be mounted on the vehicle by applying intelligent image processing techniques to identify the presence of pedestrians. The improved and efficient use of pedestrian detection can help to detect the pedestrians in the video streams. The pedestrian systems can be designed for applications that involve detecting as well as counting of pedestrians in different environments. For traffic safety, as the number of the vehicles increased by number of folds the safety of the pedestrians, cyclists and other vehicles can be achieved using these systems.

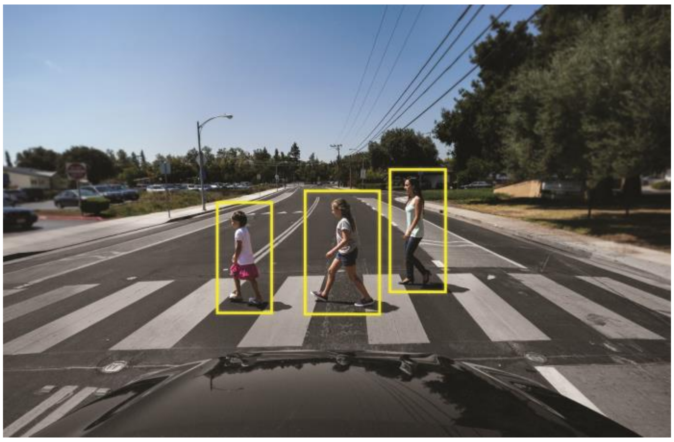

Fig -3: Example of Pedestrian detection view from autonomous vehicle.

7. DATASETS AND PERFORMANCE EVALUATION MEASURES

A. Datasets

Several benchmark datasets that are available publicly for evaluating the different pedestrian detection techniques. These datasets contain different images captured from different scenarios. It has a wide range of the applications such as automotive safety, surveillance etc. where they are used. Table II shows some pedestrian detection datasets that are commonly used in the experiments such as Caltech, INRIA pedestrian test dataset, ETH Pedestrian Dataset, TUD-Brussels Pedestrian Dataset, Daimler Pedestrian Detection [15]. In Fig.4 example images (cropped) and annotations from six pedestrian detection datasets.

Dataset	No. of negative images		No. of positive images	
	Training	Testing	Training	Testing
Caltech	61k	56k	67k	65k
INRIA	1218	453	614	288
ETH			499	1804
TUD-Brussels	218			508
Daimler	15.6			21.8

Table -2: Commonly used pedestrian detection dataset.

B. Performance Evaluation Measures

Pedestrian classification can be framed as binary classification problem in which the output is either pedestrian or no pedestrian. Detection performance of the algorithm can be represented with the help of Receiver Operating Characteristic (ROC) curve which represents a graph plotted between the True Positive Rate (TPR) and False Positive Rate (FPR). TPR and FPR are computed by the following formulas 1 and 2 respectively.

$$TPR = \text{No: of true detections/No: of positive images} \quad (1)$$
$$FPR = \text{No: of false detections/No: of negative images} \quad (2)$$

In some methods given [5] True Positive Rate is substituted by Miss Rate (MR) defined as

$$MR = 1.0 - TPR$$

False Positive Rate as False Positive per Window (FPPW). In case of the datasets do not provide the samples explicitly in that case instead of ROC curve, Precision-Recall (PR) curve is used as performance measure where Precision Recall is computed by the following formulas 3 and 4 respectively.

$$Recall = \text{No: of true positives/ (No: of true positives + No: of false negatives)} \quad (3)$$
$$Precision = \text{No: of true detections/ (No: of true positives + No: of false positives)} \quad (4)$$

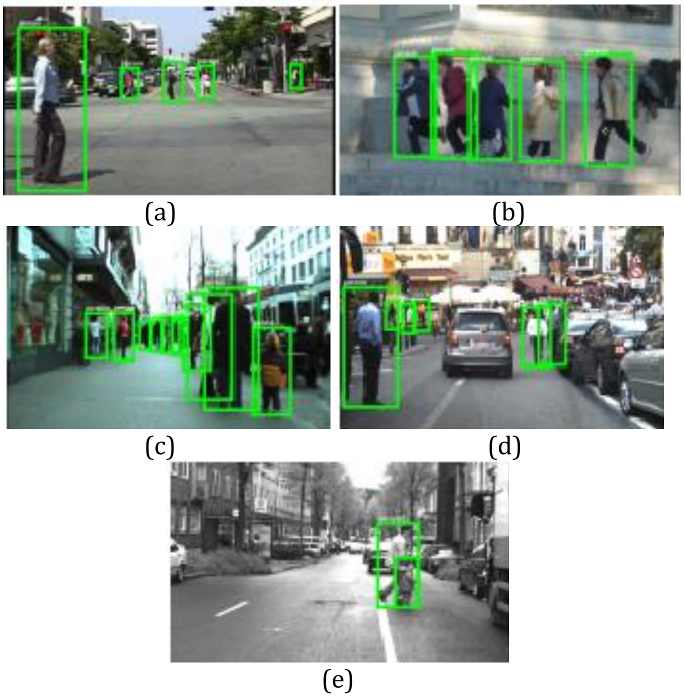

Fig -4: Example images from six pedestrian detection datasets: (a) Caltech, (b) INRIA (c) ETH (d) TUD-Brussels (e) Dailmer

8. CONCLUSION AND FUTURE SCOPE

Due to the wide range of applications the problem of pedestrian detection gained a lot of attention. Many papers have been published based on pedestrian detection but still there are certain challenges that need to be addressed. For example, most of the techniques can detect the pedestrians in the images containing pedestrians in the upright standing position. For other poses such as when they carry some baggage, sitting position, etc. then the detection algorithm's accuracy reduces. Similarly, part based models work well against the deformation, but they require prior information of the number of the parts. The count of different parts varies with the number of poses and viewing angles. This challenge becomes more complicated in complex scenarios. Furthermore, occlusion is difficult to be modeled. Self occlusion is major challenge that needs to be handled. The low resolution causes the performance to become progressively worse.

As our future work, collect images or videos using a thermal infrared camera rather than visible camera. Performance of visible cameras does not depend temperature variation and thus it remains constant in all weather conditions. They depend on light so their performance is usually unacceptable at night. But the thermal infrared camera works on all weather conditions and also during both day and night. Due to the temperature variation between backgrounds and pedestrians observed, thermal infrared camera assuming warmer pedestrians against colder background. However the visible imaging like techniques results in efficient segmentation in infrared images even in all weather, due to the high contrast between pedestrians and backgrounds.

REFERENCES

[1] https://morth.nic.in/road-accident-in-india.

[2] Akash Jain, Ankit Gupta, and Rajat Ratogi, "Pedestrian crossing behaviour analysis at intersections," International Journal for Traffic and Transport Engineering, 2014, 4(1): 103-116, January 2014.

[3] M. Bertozzi, A. Broggi, P. Grisleri, T. Graf, M. Meinecke, "Pedestrian Detection in Infrared Images," IEEE IV2003 Intelligent Vehicles Symposium.

[4] M. Bertozzi, A. Broggi, M. Felisa, G. Vezzoni, and M. D. Rose, "Lowlevel pedestrian detection by means of visible and far infra-red tetravision," in IEEE Intelligent Vehicles Symposium, Tokyo, Japan, June 2006.

[5] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," In Computer Vision and Pattern Recognition, IEEE Computer Society Conference on Vol. 1, pp. 886-893, 2005.

[6] P. Viola, M.J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," International Journal of Computer Vision, 63(2), 153-161, 2005.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," In Computer Vision and Pattern Recognition, IEEE Computer Society Conference on Vol. 1, pp. I-I, 2001.

[8] N. Dalal, B. Triggs and C. Schmid, "Human detection using oriented histograms of flow and appearance," In European conference on computer vision pp. 428-441, May 2006.

[9] T. Ojala, M. Pietik¨ainen and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern recognition, 29(1), 51-59, 1996.

[10] X. Wang, T. X. Han and S. Yan, "An HOG-LBP human detector with partial occlusion handling," IEEE 12th International Conference pp. 32- 39, September 2009.

[11] Y. Zheng, C. Shen, R. Hartley and X. Huang, "Effective pedestrian detection using center-symmetric local binary/trinary patterns," 2010.

[12] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," In Computer Vision and Pattern Recognition, IEEE Conference on pp. 1- 8, June 2007.

[13] L. Chen, W. Li, Z. Xu and L. Tang, "Pedestrian Detection based on ISC in Infrared images," IEEE Third International Conference on pp.166- 169, October 2012.

[14] K. Mikolajczyk, C.Schmid and A. Zisserman, "Human detection based on a probabilistic assembly of robust part detectors," In European Conference on Computer Vision pp. 69-82, May 2004.

[15] 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, 34(4), 743-761, 2012.