ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

A Review of Segmentation and Object Detection Methods for Real-Time Image Processing for Detecting Pedestrians

Mitali Mehta[1], Vasvi Desai[2], Rohan Gupta[3], Avani Shah[4]

B.Tech. Student, Department of Computer Engineering, MPSTME College, Maharashtra, India[1,2,3]

Assistant Professor, Department of Computer Engineering, MPSTME College, Maharashtra, India[4]

ABSTRACT: In today's world, with the constant increase in traffic and the alarming increase in vehicles on the roads, the need for pedestrian safety is an important issue that needs to be addressed. Even though there are many pre-existing systems for this purpose, the research for better real-time systems is still going on. This paper reviews the steps involved in digital image processing for a pedestrian detection system, namely Image Segmentation and Object Recognitionalong with some of their related techniques.

KEYWORDS: Digital Image Processing, Image Segmentation, Object Recognition, Edge based Segmentation, Template matching, K-Means Clustering, HOG, SVM, Cascade-AdaBoost.

I. INTRODUCTION

More than 140,000 people were killed on India's roads in the year 2015, according to figures released by the government. Nearly eight in ten accidents were caused by drivers, with 62% of those blamed on speeding [1]. These statistics and various other data from previous years indicate that there is an urgent need for pedestrian detection systems, as this can be used to safeguard the pedestrians.

Research indicates that there are pre-existing Pedestrian Detection Systems that use various combinations of the fundamental image processing steps for the implementation of their systems. Before studying pedestrian detection systems in detail, one must understand the steps involved in the actual process of detecting a pedestrian. This paper aims at identifying the fundamental image processing steps involved in the process of designing a pedestrian detection system, but does not study the different pedestrian detection systems. It studies the techniques within each step that are used in pedestrian detection systems.

This review paper is divided into four sections. In the subsequent sections, we aim to give a detailed overview of digital image processing steps. Section I is the introduction, Section II explains the concept of digital image processing in brief and gives an overview of its steps. Section III further elaborates its steps, with each of their techniques explained. Section IV concludes this paper.

II.INTRODUCTION TO DIGITAL IMAGE PROCESSING

Image Processing refers to the study of an algorithm that takes an image as an input and produces an image as an output, after performing some computations on the input image. Image processing is widely used today in many real life applications, as its backend processing logic. Some of these include Medicine, Security, Biometrics, Satellite Imagery, Personal photography, Remote Sensing, Pattern Recognition and many more. Digital Image Processing refers to the process of extracting information from a digital image.

Image Processing contains a large number of steps, but this paper only takes a close look at some of these steps that are often used in pedestrian detection. The first step this paper looks at is Image Segmentation. Image Segmentationaims at extracting required features of the image needed for further processing. This extraction is done with the help of certain algorithms. The extracted features can include many things like edges, corners or interest

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

points, blobs or regions of interest points and ridges. After this, object recognition is done, by matching the features previously extracted with the algorithmic specifications of the object.

Under the area of Pedestrian Detection, we are interested in the Region of Interest. Once this Region of Interest is obtained, further processing can be done to extract the features of the objects, which aids in classifying them as pedestrians or otherwise. Some of these features include colour, head orientation, body orientation, walking trajectory and height.

III.LITERATURE SURVEY

A. LITERATURE SURVEY OF IMAGE SEGMENTATION TECHNIQUES

Image Segmentation is the process of partitioning an image into sub-images based on certain features. Some of these features include intensity, colour and texture, among others. In pedestrian detection, image segmentation is mainly done to obtain a part of the image that could possibly contain a pedestrian. This segmentation is done against a background that might include buildings, traffic lights, vehicles, trees, among others.The background is later eliminated to single out the pedestrian. This sub-image is then sent for further classification, where the object is finally determined to be a pedestrian or not.

Pedestrian detection systems performing image segmentation before object recognition have a greater level of accuracy as opposed to a system that performs only object detection on the entire input image with the background. Due to the partitioning of an image into smaller parts, the object recognition technique needs to process fewer pixels thereby reducing the computational cost. There are three methods under image segmentation, namely region growing methods, thresholding methods and edge based segmentation methods, from which, edge based segmentation is used for pedestrian detection and has been elaborated below. The other two methods are used for various other applications in the field of image processing but, they do not give the required results in a pedestrian detection system.

1. *Edge based segmentation*

Edges are the boundaries of different objects present in a scene. They represent the high frequency components in an image. Edge based segmentation methods aim to partition the image based on the edges. After this, edge linking is performed, wherein the edges with similar intensity values are linked together to form different objects. The different types of Edge operators used for detecting edges are as follows.

i. *Roberts Operator*

Roberts Cross Operator is a 2D spatial gradient measurement on an image. It is an even length mask, and is used to detect diagonal edges. It consists of a 2x2 kernel, as shown in figure 1. The disadvantage of this mask is that it is difficult to find its centre pixel since it is an even sized mask. Hence, other masks were later made as odd length masks to overcome this drawback.

1	0
0	-1

Gx

0	1
-1	0

Gy

Fig 1: Roberts masks in x and y directions[11]

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

ii. *Prewitt Operator*

Prewitt Operator is a discrete differentiation operator. It is an odd length operator with an integer sized kernel. It computes the gradient in the horizontal and vertical directions separately. It can also be modified to calculate the diagonal gradients. The horizontal and vertical 3x3 kernels are shown in figure 2.

1	1	1
0	0	0
-1	-1	-1

-1	0	1
-1	0	1
-1	0	1

Fig 2: Prewitt masks in x and y directions[11]

iii. *Sobel Operator*

The difference between Sobel and Prewitt Operator is that Sobel does not have fixed coefficients in its operator and can be adjusted as long as the sum of the coefficients equals the centre pixel value of the kernel. Sobel is computationally expensive as compared to Prewitt Operator.Sobel Operator can also be modified to calculate the diagonal gradients. The horizontal and vertical 3x3 kernels are shown in figure 3.

1	2	1
0	0	0
-1	-2	-1

-1	0	1
-2	0	2
-1	0	1

Fig 3: Sobel masks in x and y directions[11]

iv. *Canny Operator*

The Canny operator is one step better than the Sobel operator. It takes Sobel's output as the input and further processes it for eliminating unnecessary edges. This edge thinning is done by moving the operator along the edge and checking if it is the local maxima as compared to its immediate neighbours in the same direction as the edge. After the edges are thinned, a process called hysteresis thresholding is performed, wherein the unnecessary edges are removed by a dual threshold. The edges that fall above the upper threshold are acceptedand the ones that fall below the lower threshold are trivially rejected. The edges in the middle are accepted only if they are connected to any edge above the upper threshold. This operator is better as compared to the aforementioned operators by giving only those details that maintain an accurate system without increasing its computational cost.

v. *Laplacian of Gaussian (LoG) Operator*

LoG is a second order derivative. Only applying the Laplacian mask provides erroneous result, since the second order derivative is highly sensitive to noise and will detect even the slightest noise in an image. Hence, a Gaussian filter is applied to the image before the Laplacian mask is applied to remove the noise and get better results. The Laplacian kernel can have many variations, and one of them is shown in figure4.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

0	-1	0
-1	4	-1
0	-1	0

Fig 4: Laplacian of Gaussian mask[11]

B. LITERATURE SURVEY OF OBJECT DETECTION AND CLASSIFICATION TECHNIQUES

Object detection is the process of finding instances of real-world objects such as faces, bicycles, and buildings in images or videos. Object detectionalgorithms typically use extracted features, such as shape and size, and learning algorithms to recognize instances of anobject category [2].After an image is subdivided by the segmentation algorithm, the pedestrian is located in that image using a learning algorithm. The output of this stage can be classified in four categories: positive, negative, false positive and false negative. A positive result is obtained if a pedestrian is correctly detected. The system gives a negative output if there is no pedestrian present in the segmented image. A false positive refers to that output which is generated when a pedestrian is detected without actually being present in the image and a false negative is obtained when a pedestrian is not detected by the system, even though he/she is present in the image.

The formulas of the pedestrian detection rate (PDR) and false positives per frame (FPPF) are demonstrated as follows:

$$\mathrm{PDR} = \frac{TP}{TP + FN} \times 100\%,$$

$$\mathrm{FPPF} = \frac{FP}{N_{\mathrm{frame}}} \times 100\%,$$

Fig 5: PDR and FPPF formulas[3]

where TP is the number of pedestrian samples correctly predicted to be pedestrians; FP is the number of non-pedestrian samples incorrectly predicted to be pedestrians; FN is the number of pedestrian samples incorrectly predicted to be non-pedestrians; N_{frame} is the number of total frames of the corresponding dataset sequences. For an ideal system, PDR should be maximum and FPPF should be minimum [3].

1. *Object detection*

1.1 *Template matching*

Template matching is an object detection technique in which a template made by a learning algorithm is compared to an input image. The matching is done on the basis of distance between the template and the image, namely Euclidean distance.

The process involves moving the template all over the image like a mask, and calculating the sum of products between the template and the pixels of the image. At the end, the point where the largest sum of products is obtained is selected as the region similar to the template. That region is then clipped out from the rest of the background, and a positive result is displayed on successful detection of the object.

As explained in [4], single template matching refers to the process where there is one template against which all the input images are compared. Another template matching approach is multiple-template matching in which multiple templates are compared against the input image and their individual distances are calculated. The template with the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

least distance is then chosen. This is an improvement over single template matching technique as there are a few problemsassociated with single template matching. They include the following:

(1) Different templates are needed for the pedestrian based on their **body orientation**. If a pedestrian is facing in the forward or rear direction, a single template can be used for the detection, but if the pedestrian is half turned to the left or the right, then that side body orientation also needs to be accounted for when the pedestrian gets detected, else there are possibilities that the system would result in a false negative.
(2) Different templates are needed for the pedestrian based on their **distance from the camera**. When a single template is used, it assumes a standard depth while processing input images in which a pedestrian may be very far or too close to the camera which would lead to inaccurate results.

In some instances, template matching may not give a direct result, even with multiple templates. Such cases include when the object is partly occluded. In those cases, one possible solution for identifying the image is to subdivide the image and identify the different parts.

1.2 *Clustering*

Clustering is an object recognition technique in which similar types of objects are grouped together and put in one cluster. This similarity depends upon intensity values or texture to name a few. One of the algorithms under clustering is called K-means clustering.

In K-means clustering, K is a variable that stands for the number of clusters that the image is getting divided into. Since there are K different clusters, first K arbitrary points are chosen in an image and these are taken as the initial mean positions of the K clusters. Based on these, the objects close to the K_i mean are placed under thei^{th} cluster. Then, the means of the clusters are recomputed, and the objects are again placed under different clusters based on which object is closest to which mean. This is an iterative process, and is performed until all the clusters are well-defined and separated.

One point of concern is that the initial assignment of the mean values must be done properly, else it could lead to improper cluster formation. If the points are placed at a large distance from the objects, then the computation of the mean has to be done more number of times. If the clusters are actually supposed to be formed by dividing an image horizontally, but due to the initial mean positions, on vertical division, the points still get evenly distributed, then the clustering halts. But, this results in an improper cluster since it does not give the desired output. One possible solution for this would be that if the clusters are formed by the initial positions itself, new initialization points are chosen and the computation is started again.

1.3 *Histogram of Oriented Gradients and Support Vector Machine*

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image.A feature descriptor is an algorithm which is generally applied to an image or a patch of image. In this process the useful information from the image is extracted thus throwing away the irrelevant information. Feature descriptors unlike feature detectors give information so that we can use this information to distinguish one feature from another.

An image is initially divided into small spatial regions called cells. HOG firstly computes the histograms for cells and later these are combined to form a normalised histogram for larger spatial regions called blocks.The histograms are formed by placing votes into bins. These votes are based on magnitude values. There are 9 bins, equally dividing the angles between 0-180 degrees. For example, if a pixel has an orientation of 20 degrees and a magnitude of 5 then the 1^{st} bin gets a value of 5 added to it. If an orientation falls halfway between two bins, the vote is equally distributed and given to the two bins [9].

HOG descriptors are computed for all locations or the ROI in the image. Thus it is a dense extraction method. After the histogram for the oriented gradients is formed, this information is passed on to a learning algorithm like Support

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

Vector Machine (SVM) for further classification. The combination of HOG and SVM for object detection works as follows:

First the HOG descriptors of positive samples (those that contain the object that you want to detect) from the training dataset are extracted, the same is done for the negative samples (those that do not contain the object). Usually in practice, the feature descriptors extracted for negative samples should be much greater than that for positive samples. Then the Support Vector Machine is trained on the positive and negative samples. Then the false positives for each negative sample are found by applying the sliding window technique. For each window, the HOG features are extracted and then the classifier is applied. This is called hard negative mining. Then each of these false positive patches are recorded along with their probability. After this, the samples obtained from the hard negative mining technique are sorted out and the classifier is retrained on these samples. Now after the classifier has been trained, it is applied on the test dataset. For each image on the test set first the sliding window is applied, then the HOG features are extracted for each window and lastly the classifier is applied to them. When the object with high probability is detected a bounding box of the window is recorded.

2. *Classification techniques*

2.1*Support Vector Machine*

A Support Vector Machine (**SVM**) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples [6].

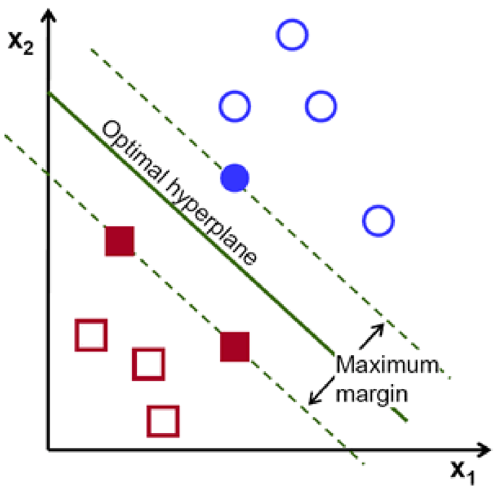

Fig 6: SVM[6]

This means that if we want to detect the class of an object, that is, if we want to discriminate between two types of objects, SVM uses a hyperplane for the separation. This hyperplane divides the two types of objects and has some distance from the object closest to it from each class. There may be more than one hyperplanes for dividing the two object classes, but the one that has maximum margin from the two classes is called the optimal hyperplane and is hence chosen.

SVM can also be used in combination with other descriptors and not only with HOG for object detection.Some of these descriptors include SIFT (Scale-Invariant Feature Transform) and GIST.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

2.2 *Cascade-AdaBoost*

AdaBoost refers to Adaptive Boosting. It is a technique in which a number of weak learners are grouped together to form a strong classifier.

These weak learners are called 'weak' learners because they only focus on one feature while performing their classification. Since classification of data can depend and vary according to multiple factors, weak learners cannot suffice. So, to increase their performance, they are combined together and each of them classify according to their own single feature, and this combination would lead to a strong classifier, because, when the data is given to this combination of weak learners, all their different features will be considered while classifying this data.

The weak learners that are used in AdaBoost are often decision tree classifiers.The training of the decision tree classifier is to construct a tree on the basis of the training samples. It aims to minimize entropy and maximize gain. At the root node, all the training samples are input. Then, according to one feature, the samples travel forward along different branches. The samples from the same class should walk together as long as possible along the same branch. When it comes to the leaf node, all the samples are from the same class. At one node, the feature X of the maximal gain ratio is chosen to split the training samples reaching the node into m branches. When one of the following conditions is satisfied, it stops to yield child nodes: The samples reaching at the nodes are from the same class or if there are a few number of samples at the node [5].

Some factors must be considered while using AdaBoost. It can be sensitive to noise in the data and outliers. While AdaBoost considers various features for classification, it does not give each of them equal priority. Some features have a higher priority for classifying an object as opposed to others. For example, if the object that has to be detected is a walking pedestrian, some of the features of that image would be the detection of a head, the torso and the legs. Here, whether a pedestrian is standing or moving, the head position would have the same importance, but the real difference would be seen in the pedestrian's legs. If the legs are open, the pedestrian is more likely to be walking and if the legs are closed, he is more likely to be standing. Hence, the AdaBoost classifier would give a higher priority to the legs' classification and the head classification would get a lower priority. These priorities are called as 'weights' and hence an AdaBoost classifier is a weighted sum of the weak classifiers. AdaBoost can be used with other learning algorithms in conjunction, so that it leads to better performance.

A Cascade-AdaBoost classifier uses a number of AdaBoost classifiers linked together sequentially. An input is sent to the first classifier and if it travels through all the classifiers without being discarded, it is accepted as a positive sample. If, at any stage, it gets termed as a negative sample, it is rejected from the training sample set and further classification halts with a final value as negative.

IV. **CONCLUSION**

In this paper, we have presented the steps involved in digital image processing. This paper explains some of the predominant techniques under each of these steps that are only related to pedestrian detection. The steps reviewed here include Image Segmentation and Object Recognition. A lot of systems can be implemented combining some of these techniques, whose hybrid may prove to be better in terms of computational cost or accuracy. One of the classification techniques not mentioned in this paper is Neural Networks. A lot of research is being conducted currently on Neural Networks and may yield a very accurate object classification technique. This technique shows a prospect for future research.

REFERENCES

1. [Online] Available: http://www.bbc.com/news/world-asia-india-36496375. [Accessed: 26- Feb- 2017]
2. "Object Detection - MATLAB & Simulink". *Mathworks.com*. N.p., 2017. Web. 7 Apr. 2017.
3. Zhihui Wang et al., "A High Accuracy Pedestrian Detection System Combining a Cascade AdaBoost Detector and Random Vector Functional-Link Net," in*The Scientific World Journal Volume 2014*.
4. OmidHosseiniJafari and Michael Ying Yang,"Real-Time RGB-D based Template Matching Pedestrian Detection," in *IEEE International Conference on Robotics and Automation (ICRA),2016*.IEEE,2016.
5. Yiting Zhang and Jianning Liang, "A Vision Based Method for Object Recognition," in *3rd International Conference on Information Science and Control Engineering,2016*. IEEE,2016.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

6. "Introduction To Support Vector Machines — Opencv 2.4.13.2 Documentation". *Docs.opencv.org*. N.p., 2017. Web. 15 Apr. 2017.
7. Rohan Kandwal,Ashok Kumar and Sanjay Bhargava, "Review: Existing Image Segmentation Techniques," in *International Journal of Advanced Research in Computer Science and Software Engineering,2014*.
8. David Gero´ nimo, Antonio M. Lo´ pez, Angel D. Sappa, Member, IEEE, and Thorsten Graf, "Survey of Pedestrian Detection forAdvanced Driver Assistance Systems," in *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 7, JULY 2010*. IEEE,2010.
9. NavneetDalal and Bill Triggs, "Histograms of Oriented Gradients for Human Detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05),2005*. IEEE,2005.
10. Computerphile, Dr Mike Pound: Computer Vision: Filters (Blur, Edge Detection etc). 2015 [Online]Available:https://www.youtube.com/playlist?list=PLzH6n4zXuckoRdljSlM2k35BufTYXNNeF. [Accessed: 01- Mar- 2017]
11. SonamSaluja ,AradhanaKumari Singh2 and Sonu Agrawal, "A Study of Edge-Detection Methods," in *International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013*.