iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VII **Month of publication:** July 2018

DOI: http://doi.org/10.22214/ijraset.2018.7027

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VII, July 2018- Available at www.ijraset.com

Tracking Human Motion Based on Bounding Box Concept

Ms. Rajshree B. Lande[1], Mr. R.M. Mulajkar[2]
[1]*PG Student, Jaihind COE,* [2]*Assistant Professor, Jaihind COE, Kuran, Department of Electronics & Telecommunication, Savitribai Phule Pune University*

Abstract: ***This paper presents moving object detection based on background subtraction for video surveillance system. In all computer vision system, the important step is to separate moving object from background and thus detecting all the objects from video images. The main aim of this paper is to design a bounding box concept and alarm system for the human detection and tracking system in the presence of crowd. The bounding box around each object can track the moving objects in each frame and alarm will fire when object is detected and it can be used to detect crowd or theft and the estimation of crowd. This paper gives the implementation results of bounding box and alarm system for detecting objects and its tracking. The evaluation test performed on proposed system like PSNR, MSE was found.***
Keywords: Background Subtraction, Alarm system, Bounding box, Video surveillance system, Moving object

I. INTRODUCTION

In the various applications of computer vision systems, human detection and human activity recognition are the popular areas in which the researchers focused. These systems are now used for collecting, analyzing and recording of all the moving objects in a scene and it can detect some complex work such as human interactions, vehicle classification, monitoring of intersections like occlusion and comparison of two images. Extracting complex information from the video with the use of computer has become essential in cyber age. A community network consists of a collection of citizens with different pattern of interactions between them and the understanding of such interactions in community venues has great interest in low enforcement and motherland security. The human cannot be detected using radar because of the reflection of radar signal from human body parts is very low. So computer vision techniques are used to detect the human. In the early stages, the vision system works with the help of a human operator to detect the doubtful activity. In this type of semi-automated system, the operator may give false alerts and he may not have interest to identify the genuine.

The frame subtraction method, the background subtraction method and the optical flow method are widely used moving object detection methods. In the frame difference, or Frame Subtraction method detect moving objects by calculating the differences between pixels in consecutive frames of a video sequence, as well as extracts the motion regions by the threshold of time difference in adjacent frames pixels. Frame Subtraction method is adaptive to environments with sudden illumination change. This results in holes inside moving entities. This method uses pixel-based difference to find the moving object.

Perform clustering process after calculating image optical flow field according to the optical flow distribution characteristics of image and scene. It is very easy to detect moving object from background and can get complete movement information of moving object. There are some disadvantages of this system also like more number of calculation, sensitive to noise, poor anti-noise performance and not suitable for real time processing.

II. LITERATURE SURVEY

In 2014, [1] presented a paper on motion human detection based on background Subtraction using bounding box concept which proposes a new method to detect moving object based on background subtraction. Moving object can obtain using dynamic threshold method and background updating model and detect the object by using bounding box. In all computer vision system, the important step is to separate moving object from background and thus detecting all the objects from video images. The main target is to design a bounding box concept for the moving object detection and tracking system in the presence of crowd. The images are captured by using monocular cameras. Moving object in each frame track by bounding box around it and it can be used to detect crowd and the estimation of crowd. For detection of crowd and estimation of crowd the bounding box around each moving object can track the moving objects in each frame. This gives the implementation results of bounding box for detecting objects and its tracking.

©IJRASET: All Rights are Reserved

In April 2013, [2] presented a paper on Estimating the density of people and counting the number of people in a crowded environment for Human Safety. For estimating the density of people in crowd proposed system divided into two fold. First, density estimation of the crowd size and second, count the number of people in the crowd. As the crowd density increases, occlusion between the people increases .We can use Adaptive K-GMM Background subtraction algorithm to extract foreground. By applying boundary detection algorithm, we can estimate the size of the crowd. By using "Canny edge detector" algorithm and centroid method using bounding box concept the number of people can count from crowd.

In 2014, [3] presented a paper on event detection technique for detecting abnormal events in traffic video surveillance. The main objective of this work is to detect the abnormal events which normally occur at junction, in video surveillance. Our work consists of two phases 1) Training Phase 2) Testing Phase. Initially, the frames are divided into grid regions at the junction and labels are assigned. Blob detection and tracking, conversion of object location to data stream and pattern matching are flow of algorithm. In the training phase, separating the modelled static background frame using Gaussian mixture models (GMM) for blob detection and this will be carried out for every frame for tracking purpose.

In April 2013, [4] presented a paper we propose the use of Distance Metric Learning (DML) in combination with Nearest Neighbour (NN) classification for object tracking. Initially a video file is read and the frames in the video are accessed individually. The process is repeated for all the frames. Then the object is tracked using the Distance Metric Learning algorithm. Using normalized correlation between the frames. The human appearance model is identified using the Blob detector which uses the skin color to identify the object. Then the bounding box is fixed for the object in that frame. Then the video is reconstructed with the processed frames. Feature extraction is done using Region Props which threshold the image and extract the features.

In 1995, [5] presented a paper on some solid modelling and robotics algorithms need also to test quickly the possible interferences of the different parts of a system such as the arms of different robots. To speed-up the test, the free form objects to be tested are encapsulated into boxes and the test starts by performing a check on overlapping of the bounding boxes, thus producing a section of the space where an intersection can occur. Then, only if the boxes overlap the test continues on the original free form objects only in the portion of the 3D-space resulting from the first step. This paper present a simple, fast and efficient algorithm (called the Bounding Bozes Group method) to compute a tight bounding entity of a complex object modelled in the Constructive Solid Geometry (CSG). If no upper limit is set to the number of elements of each Bounding Box Group, we demonstrate that our algorithm achieves the tightest possible solution.

In 2007, [6] presented a paper method to identify frames with significant segmentation errors in an individual's track by analyzing the changes in appearance and size features along the frame sequence. The features used and compared include global Colour histograms, local histograms and the bounding box' size. Experiments were carried out on 26 tracks from 4 different people across two cameras with differing illumination conditions. This indicates that the analysis of such features along a track can be useful in the automatic detection of significant segmentation errors. This can improve the results of many applications which requires robust segmentation results from a tracked person.

In 2008, [7] presented a paper on a novel and effective approach for real-time analysis of crowd congestion (density) in a physical space monitored by surveillance cameras. A region of interest (ROI) is specified in the space and partition of the ROI into an irregular array of sub-regions (blobs) automatically carried out, to each of which a congestion contributor is computed. For blob-based dynamic congestion detection uses the short-term responsive background (STRB) model and for blob-based 'zero-motion' (static congestion) detection uses the long-term stationary background (LTSB). And finally, the combination of the local and global analysis gives the accurate congestion rating. Besides, this scheme is adapted to perform the task of moving object presence detection with success. Extensive tests and field trials validate both the accuracy and robustness of the approach.

III.PROPOSED METHOD

In video analysis system the detection of object, tracking and analyzing the object in each frame to detect the behaviour of object are basic steps involved. Background subtraction is a general method of separating moving object from static object. It gets moving region from current image pixel-by-pixel from background image. This is done by equating multiple images from the start. The main aim of background subtraction is the detection of objects in motion. Background subtraction used after initialization of background and then subtract current frame. This method is easy and it perform in accordance to a simple protocol. So now we propose a simple system to detect human in each frame using simple object detection method and bounding box creation from monocular videos and developed alarm system when moving object is detected. This work is implemented in processor using MATLAB2014b.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VII, July 2018- Available at www.ijraset.com

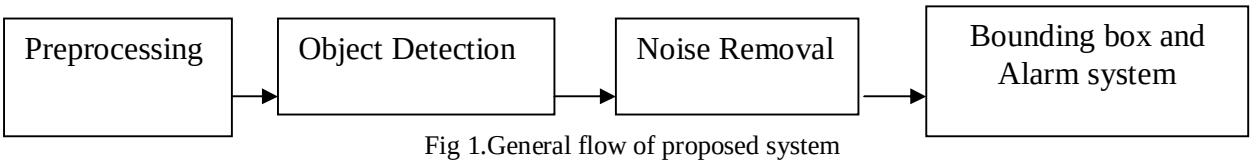


Fig 1.General flow of proposed system

A. Pre-processing and Transformation

From the captured scene Frame separation is the first step in the background subtraction algorithm, the purpose of this step is to prepare the compatible and modified video frames by removing noise from the captured scene and unwanted objects in the frame in order to increase the amount of information gained from the frame. Pre-processing of image is a process of collecting simple image processing tasks that change the raw input video info into a required format. This can be processed by subsequent steps. Pre-processing module that module contains the entire basic step required to compute data. The input to the algorithm is a sequence of video frames which convert RGB to gray-level format. The algorithm produces a binary mask for each video frame. To improve the detection of moving object's Pre-processing is necessary, For example; by spatial and temporal smoothing, snow as moving, leaves of a tree, can be removed by morphological processing of the frames.This is done after the identification of the moving object. Frames of background and original scene's frames are decomposed and reconstructed by wavelet transform. Frames are separated into different frequencies like low frequency image and some high frequency images. Finally removes low frequency image and high frequency images are kept as it is.

Wavelet transform used after frame separation to get time-frequency representation of each frame. Wavelet transform convert each frame into different time- Frequency component like LL, LH, HL, HH.

B. Background Image Initialization

There are many methods to obtain the initial background image. For example, the average pixel brightness can obtain from first few frames as the background or assume first frame as the background directly. From these methods, the time average method is the mostly used method of initialization of background However, this method cannot used with the background image (especially the region of frequent movement) which has the shadow problems. Median of multi-frame can avoid shadow problem simply. So the median method is selected to setting background initial image in this paper. Expression is as follows:

$\beta init\ (X, y) = \text{median}\ fk\ (x, y)\ k = 1, 2 \ldots n$ (1)

Where βinit is the initial background, n is the total number of frames selected.

C. Background Update

For the background model can better adapt to light changes, the background image needs to be updated in real time so as to accurately extract the object in motion. In this paper, the update algorithm is as follows:

In detection of the object in motion, the pixels judged with respect to moving object maintain the original background gray values, not be updated. We update the background model in which pixels are part of background model according to following rules:

$Bk+1(x, y) = \beta Bk(x, y) + (1-\beta)\ Fk(x, y)$ (2)

Where β belongs to (0, 1) is update coefficient. Fk(x, y) is the pixel gray value in the current frame. Bk(x, y) and Bk+1(x, y) are respectively the value of the current frame and the next frame.

D. Moving Object Extraction

Background subtraction is a popular technique to detect the interested moving objects in a frame. This technique involves subtracting an image that contains the object, with the initialised background image that has no moving objects of interest is present. If there is a difference between current frame and initialised background frame then current frame consider a part of moving object .Detected moving object which are represented by groups of pixel, are then separated from the background image by using threshold technique.

After the background image B(x, y) is initialized, subtraction performed between the background image B(x, y) and current frame Fk (x, y). Pixel is a part of moving object, if the resulting pixel difference is greater than the set threshold T. otherwise, it is background pixels. After threshold operation moving object determined. Its expression is as follows:

$Dk(x, y) = 1$ if $|Fk(x, y)-Bk(x, y)|>T$

0 other

©IJRASET: All Rights are Reserved

Where Dk (x, y) is the binary image of resultant difference, its size determines the accuracy of object identification.

E. Noise Removal

Camera noise and irregular object motion noises can be clean out with the use of morphological operations. The noises are normally appeared as white pixels in background and black pixels in foreground. After object detection, image restoration Process is used to fill the missing spaces in moving objects. The morphological operations are to search the image with a Simple pre- defined shape, illustrating conclusions on how this shape fits or misses the shapes in the image. Morphological operations consists of some structure elements like disk, square and cross shaped elements of 3x3 sizes. The essential Operations of binary morphology operations are dilation, erosion, closing and opening. Dilation operation enlarges the Region, while erosion makes the region small. Closing operation is defined as performing erosion after dilation and it can fill the internal holes in the region. In opening operation dilation is performed after erosion and it can clear small Portions that are just out from the boundary in to background. The mathematical operators "bwmorph" is used on each image Frames perform erosion and dilation.

F. Bounding Box and alarm system

The bounding box is a single rectangular box that capture the entire objects in a scene. It gives information of coordinates of upper left and lower right corners of the bounding box. It is computationally efficient, implementation is easy and it used to detect the motions associated with human body. From height and width of bounding box can be used detect the human activity. If the dimension of the bounding box is very small, then some part of the objects will not be covered in the bounding boxes. The complete human body can be bounded using single rectangular bounding box and motion object has to be detected using this single box. The exact location of the pedestrian can also be found using bounding boxes. The bounding box used shapes are sphere, parabolic, square and rectangular. The cylindrical shape also used as bounding box. When object is detected installed alarm system will fire alarm to give indication that moving object is detected. This alarm system will protect and keep safe area from theft or unknown person in which this system has been used.

IV. RESULT

PSNR (Peak Signal-to-Noise Ratio)

The PSNR calculate the peak signal-to-noise ratio, in decibels, between background model and captured video frames. This ratio is often used as a quality measurement between the background model and captured image. The better the quality of the output image has the higher the PSNR, to compute the PSNR, we first calculate the mean-squared error using the following equation:

$$mse = \Sigma ([I1(m,n) - I2(m,n)]2 / \text{prod. of rows, columns}) \quad (3)$$

$$PSNR = abs(20 * log10(255/sqrt(mse))) \quad (4)$$

In the above equation, m and n are the number of rows and columns in the input images, respectively where **Mean Square Error** (mse) (3) indicates the average difference of the pixels throughout the image. A higher mse indicates a greater difference between the original and processed image. The different observed mse and PSNR of captured video as per below table.

Input Video	mse	PSNR
Input Video 1	0.0092	68.4709
Input Video 2	0.0300	63.3620
Input Video 3	0.0190	65.3439
Input Video 4	0.0288	63.5425

Table 1.mse and PSNR result of different input video

A. Video Capturing

Camera coordination and camera calibration are time consuming process, if we are using multiple cameras. So in this work, we use single stationary camera to capture the video.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VII, July 2018- Available at www.ijraset.com



Fig 2. Captured Video

B. Video Framing

Different image processing techniques are used for video framing. Fig 3 shows some of frames from the video clips. To reduce the complexity, the objects come closer are taken to shown as result because the main event going to occur in almost all visual videos when objects come closer.



Fig 3. Result of video framing

C. Background Reference Image

To perform background subtraction, we need a static reference image. Fig.4 shows the result of reference background image. Reference image contains image which has static information.



Fig 4. Reference Image

D. Wavelet Transformed image

Wavelet transform convert preprocessed image into time-frequency component. It convert each frame into different time- Frequency component like LL, LH, HL, and HH. High frequency component contains noisy information. Fig 5 .shows transformed output

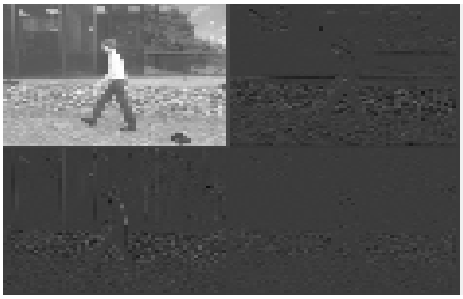


Fig 5. Wavelet Transformed image

E. Background Subtraction

Object detection is the isolation of objects from back ground. After video framing the videos, background subtraction is performed for object detection. The back ground subtraction method is good for static scenes. In background subtraction each current frame is subtracted from the reference background and thus the foreground object can be detected. Fig.6 shows the result of background subtraction

©IJRASET: All Rights are Reserved



Fig 6 .Result of Background Subtraction

F. Moving Object Detected with Bounding Box

For getting the coordinate information of object bounding box is around the object. To track the object it is mostly used.
It has rectangular shape .There are many more shapes also available. Fig 7 shows result of moving object with bounding box.



Fig 7 .Result of bounding box concept

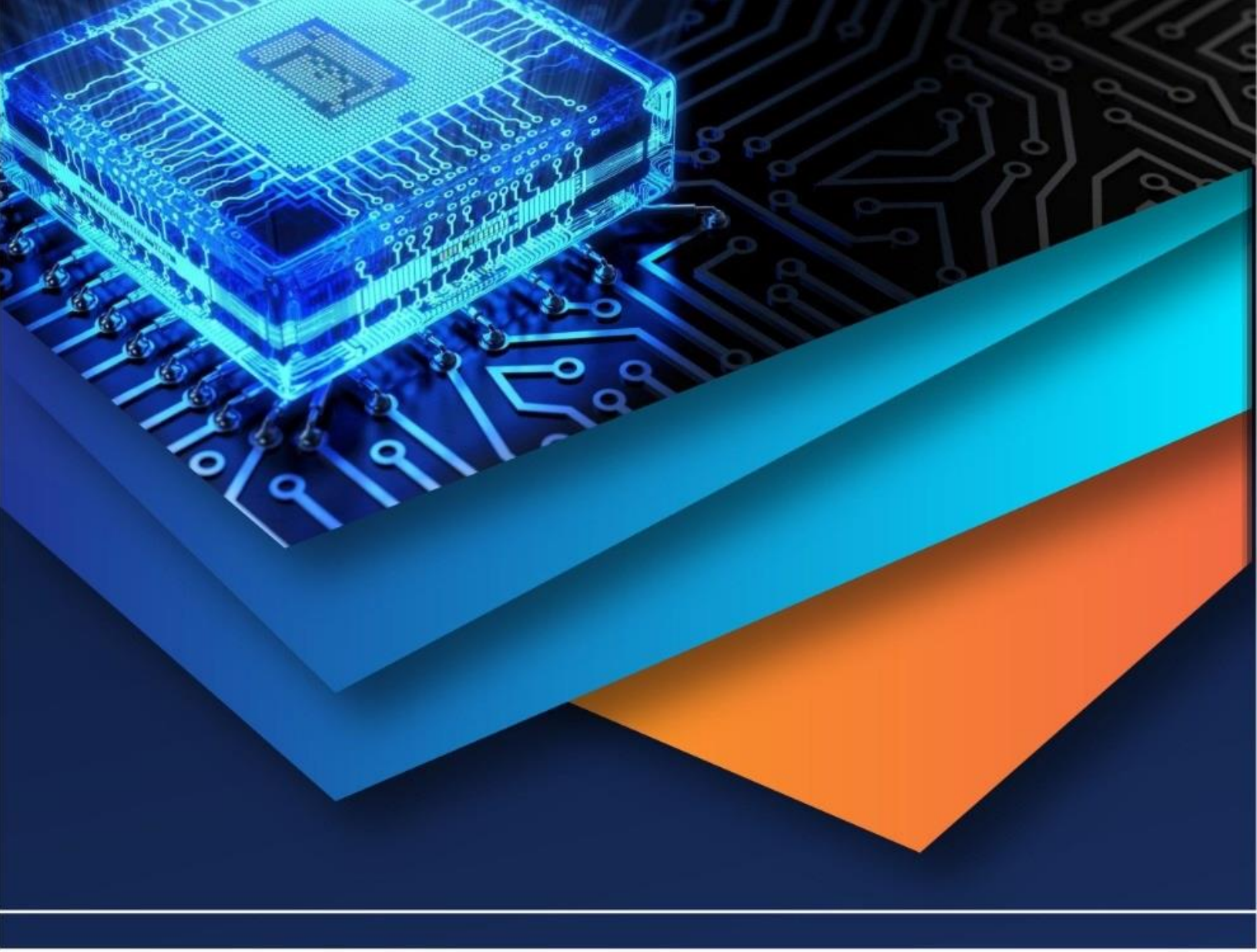
V. CONCLUSION

Nowadays visual surveillance system has been used in many areas. All video surveillance system used in this field can be treated as adjusting systems because it can be used for many of the applications without changing hardware but only slight changes in the algorithms. Our system can detect human using background subtraction algorithm and used wavelet transform to get accurate result. It is used in many applications such as occlusion detection, crowd detection, and crowd density estimation etc. In the presence of dense crowd, accurate counting of people is troublesome if single camera is used. To deal with this, we proposed advance crowd detection in a frame before it actually happened.

REFERENCES

[1] Deepak P. and S. Krishnakumar and Smitha Suresh “Human Recognition for Surveillance Systems Using Bounding Box” International conference on Communication and Signal Processing, IEEE 2014.
[2] P Karpagavalli and A.V. Ramprasad, “Estimating the density of people and counting the number of people in a crowded Environment for Human Safety, International Conference on Communication and Signal processing, IEEE 2013.
[3] P.M.Ashok Kumar, Dr.V.Vaidehi and E.Chandralekha, “Video Traffic Analysis for Abnormal Event Detection Using Frequent Item Set Mining”, International Conference on Recent Trends in Information Technology, IEEE 2013.
[4] S. Mahendran, D. Vaithiyanathan and R. Seshasayanan, “Object Tracking System Based on Invariant Features”, International conference on Communication and Signal Processing, IEEE 2013.
[5] Andrea Sanna and Paolo Montuschi,” On the computation of groups of bounding boxes for Fast test of objects Intersection”, IEEE, 1995.
[6] Christopher Madden and Massimo Piccardi, “Detecting Major Segmentation Errors for a Tracked Person Using Colour Feature Analysis”, 14th International Conference on Image Analysis and Processing (ICIAP 2007), 2007.
[7] Li-Qun Xu and Arasanathan Anjulan” Crowd Behaviours Analysis in Dynamic Visual Scenes of Complex Environment”, IEEE2008.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)