Multiple Moving Object Detection and Tracking using Harr Features with Smart Video Surveillance System

Mr. Sunil Tiwari N.

4th sem (M.tech) Digital Electronics
Dept. Of Electronics & Communication Engineering
SJB Institute of Technology
Bangalore, Karnataka, India

Mr. Ravikumar A.V.

Assistant Professor
Dept. of Electronics & Communication Engineering
SJB Institute of Technology
Bangalore, Karnataka, India

Abstract-**Real-time human detection and tracking is a vast, vibrant yet inconclusive and complex area of computer vision. Due to its increased utilization in surveillance, tracking system used in security and many others applications have propelled researchers to continuously devise more efficient and competitive algorithms. Automatic visual human counting and video surveillance have important applications for home and business environments. Moving human detection and tracking is often the first step in applications such as video surveillance. The main aim of project is moving human detection and tracking system with a static camera has been developed to estimate velocity, distance parameters we propose a general moving human detection and tracking based on vision system using Harr based Human Mask Generation Method. This project focuses on detection of moving humans in a scene for example moving people meeting each other and tracking and detects people as long as they stay in the scene. This is done using Harr based Human Mask Generation algorithm and difference algorithms with Mat-lab software and we could calculate distance frame per time velocity.**

Index Terms-**Object detection, Haar features, Open CV, Frame difference, Block matching method.**

I. INTRODUCTION

Video surveillance systems have long been in use to monitor security sensitive areas. The history of video surveillance can be divided into three generations.

The first generation surveillance systems (1GSS, 1960-1980) were based on analog sub systems for image acquisition, transmission and processing. They extended human eye in spatial sense by transmitting the outputs of several cameras monitoring a set of sites to the displays in a central control room. They had major drawbacks like requiring high bandwidth, difficult archiving and retrieval of events due to large number of video tapes and it also needs human operators with limited attention span.

The second generation used both analog and digital subsystems to resolve some drawbacks of its predecessors. They made use of early video processing techniques to provide assistance to human operators by filtering spurious events in a video.

Third generations systems provide end-to-end digital systems. In third generation the image processing is distributed towards sensor level by the use of intelligent cameras that are able to digitize and compress acquired analog signals and perform image analysis algorithms like motion detection with the help of digital computing components.

The trivial approaches for human detection and tracking had high computational cost and require specialized and expensive hardware to work in real- time. Other algorithms use prior information about the geometry of the scene, such as the floor position and the camera calibration to restrict the data association and tracking problems. This approach makes more complex the system installation and setting, since it is necessary to compute the camera calibration and estimate the 3D plane of the floor which in turn depends on the camera location. Most of the techniques used for this problem deal with closed world representations which rely on a specific knowledge on the type of actions taking place. This more general case allows us to evaluate the proposed approach for processing video streams acquired in real world video surveillance situations.

There are various techniques for moving human detection like optical flow, low change of illumination, segmentation background subtraction, frame difference etc. Most available techniques for detecting moving humans resulted in occlusion and were leading to false detections. The proposed work of human detection will be based on Haar classifiers used as algorithm. The use of Haar classifiers has boosted to the upgrade of the system which is fast and more accurate. These methods allow segmenting each image into a set of regions representing the moving

humans by using a background differencing algorithm. More recently have proposed a local modeling of the background using a mixture of K-Gaussian allowing for processing video streams with time varying background. These methods give satisfactory results and can be implemented for real time processing without dedicated hardware.

In the proposed work the concepts of dynamic template matching and frame differencing have been used to implement a robust automated human tracking system. In this implementation a monochrome industrial camera has been used to grab the video frames and track a human. Using frame differencing on frame-by-frame basis a moving human has been detected it is tracked by employing an efficient Template Matching algorithm

This paper is organized as follows. In section II, the structure of the proposed project is presented where each module has to be implemented and interfaced. In section III, a brief approach with the help of flowchart has been demonstrated. In section IV, results are shown and in the final section V, conclusion of the paper is drawn.
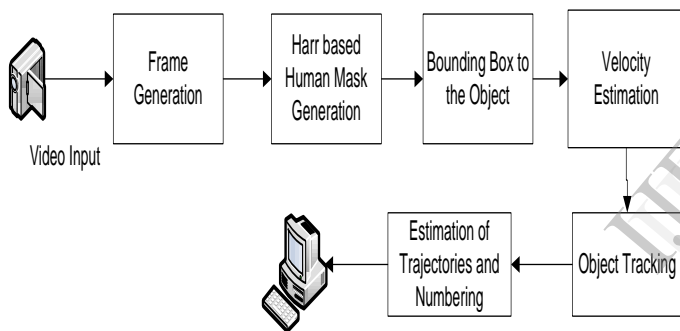
II.   METHODOLOGY

Figure 1: General Block diagram of the Proposed System

Fig. 1 shows the various stages in implementing the project, each block represent the module which has to be interfaced one another for obtaining the estimated output.

A.   *Human Detection Using Haar Like Features*

Haar-like features have scalar values that represent differences in average intensities between two rectangular regions. They capture the intensity gradient at different locations, spatial frequencies and directions by changing the position, size, shape and arrangement of rectangular regions exhaustively according to the base resolution of the detector.

The Haar basis functions are a set of rectangular 2D features derived from the Haar wavelet, see equation (1).

$$\psi(x) = \begin{cases} 1 & if \quad 0 \le x \le \frac{1}{2} \\ -1 & if \quad \frac{1}{2} \le x \le 1 \\ 0 & otherwise \end{cases} \quad (1)$$

Fig. 2 shows the example of Haar-Like feature sets. Haar-feature said to be present if subtracting the average dark-region pixel value from average light-region pixel value is above a threshold (set during learning). Ada-Boost combines many weak classifiers to create one strong classifier. Object detection via Haar-like Features with Cascade of Boosted Classifiers.
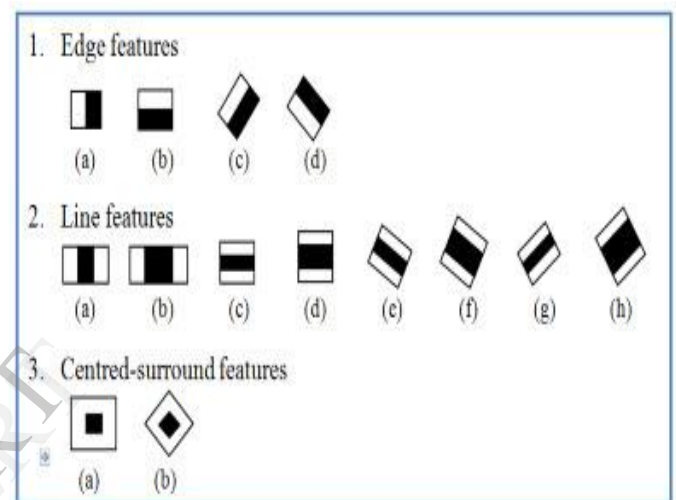

Figure 2 Examples of Haar-Like Feature Sets

The Haar basis functions forms the elemental part in extracting important features of an image i.e. a set of simple rules or weak classifiers is combined to enhance its performance. The role of weak classifier is to separate the given training set into positive and negative samples with more than 50% accuracy. To use Haar basis set as classifiers we compute the value of each feature i.e. computes the difference between the sum of pixels in black region and the white region. The feature value is then compared with a threshold to classify the samples.

The classifier discussed so far is implemented in Open CV, an open source library for image processing and computer vision. Open CV comes with sample database that are already trained to detect full body, upper body, frontal view. Figure 3 below shows Open CV Haar classifier cascade applied to detect pedestrians. To detect other objects the classifier can be trained with a few hundred labeled positive and negative samples of a target object, this process is highly time consuming. The output from Open CV process is .XML file of the Haar features and the thresholds used in the various stages. The output of each feature computation is compared with the threshold and the classifier output is decided as shown in equation (2). Likewise the classifier can be applied to whole image at various scales to localize a target image.

$$h_t(x) = \begin{cases} left\_val & if\ f(x) < threshold \\ right\_val & otherwise \end{cases} \quad (2)$$

Figure 3 Open CV Haar cascade applied to detect pedestrians

B. *Human Tracking Using SAD Algorithm.*

After the object detection is achieved, the problem of establishing a correspondence between object masks in consecutive frames should arise. Obtaining the correct track information is crucial for subsequent actions, such as object identification and activity recognition.

Sum of absolute differences (SAD) is a widely used e algorithm for measuring the similarity between image blocks. It works by taking the absolute difference between each pixel in the original block and the corresponding pixel in the block being used for comparison. These differences are summed to create a simple metric of block similarity.

In this algorithm the difference between two images taken at time t_i and t_j is defined using sum of absolute difference for each RGB color channel. The motion estimation is based on the calculation of the sum of absolute differences (SAD) according to following equation 3.

$$SAD = \sum_i \sum_j |I_k(\ i, j) - I_{k-1}(\ i, j)| \quad (3)$$

C. *Velocity of Objects in a Video*

The distance travelled by the object is determined by using the centroid. It is calculated by using the Euclidean distance formula as given in equation (4). The variables for this are the pixel positions of the moving object at initial stage to the final stage

$$Distance = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2} \quad (4)$$

Where
X1 = previous pixel position
X2 = present pixel position in width
Y1 = previous pixel position
Y2 = present pixel position in height

The velocity of moving oject is calculated by the distance it travelled with respect to the time. By using the value of distance with respect to frame rate, the velocity of the object is defined. The defined velocity is of two dimensions.

The velocity of moving object in the sequence frames is defined in pixels/second.

D. *Design Procedure*

The design procedure of the proposed filterbank can be described as:

1. For a video selected we detect the moving people by using trained Haar features i.e. the .xml file, which acts as database for detecting the people

2. After the objects have been detected, tracking of the detected object is done using SAD algorithm, Both detection and tracking has to be done simultaneously.

3. The detected objects has be indexed with a particular number so that determines the total number of people present at a particular point of time.

4. Once the number of people has been estimated, The velocity of people are determined in meter per second as given in equation (4).

III.   DESIGN AND IMPLEMENTATION

The entire process of object detection and tracking the moving object is illustrated in the following flowchart. The flowchart gives an idea about different techniques.

- Use of Haar features for object detection.

- Difference method used for tracking.

- Assigning values to detected objects.

- Morphological operations conducted on video frames

- Validating the image and determing the velocity

Figure 4 shows the process flow chart of the whole system

```
            ┌──────────────┐
            │    START     │
            └──────────────┘
                   │
    ┌──────────────────────────────┐
    │  Select bFrames as           │
    │  background frames           │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Read bFrames+1 upto         │
    │  end frames                  │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Convert background          │
    │  image to grayscale          │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Find Absolute difference of │
    │  background image and        │
    │  bFrames+1 image             │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Convert Image to Binary     │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Morphologically open        │
    │  binary image                │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Validate the image          │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Measure properties of       │
    │  image region                │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Detect the people in        │
    │  the image(pobjects)         │
    └──────────────────────────────┘
                   │
          ◇ If pobjects =0 ◇ ──────→  ┌──────────────┐
                   │                   │  No Human    │
                   │                   │  body found  │
    ┌──────────────────────────────┐  └──────────────┘
    │  Check for no. of pObjects   │
    │  in image                    │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Find the numbering by giving│
    │  person name                 │
    └──────────────────────────────┘
                   │
    ┌──────────────────────────────┐
    │  Find the total count        │
    └──────────────────────────────┘
                   │
            ┌──────────────┐
            │    STOP       │
            └──────────────┘
```

IV. EXPERIMENTAL RESULTS

Figure below shows the video snapshots of undetected humans in the video with different backgrounds which is the input and the output frames which is being tracked and also contain information about number of people in the scene along with their respective velocities represented in meter per second with which they are moving. The output also contains information about the total number of frames present in the video clip.
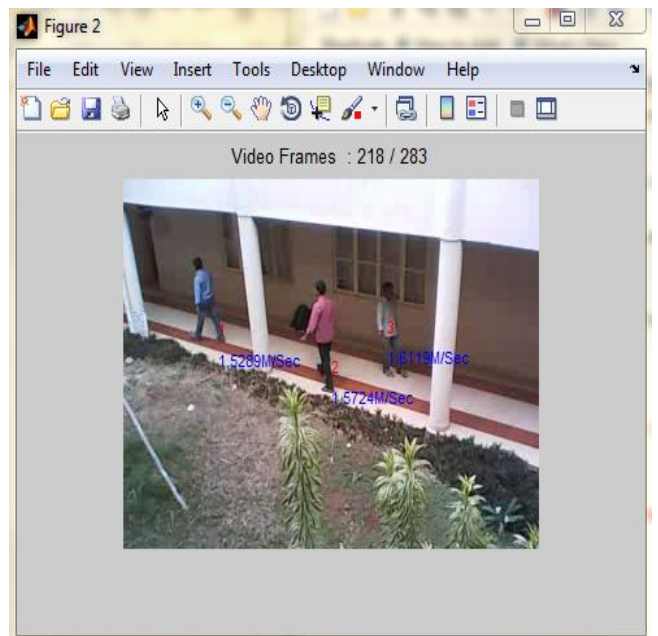

Figure 1 shows original video frame

Figure 1a shows the output after frame difference, which contains three people moving with respective velocities.

Figure 2 shows original video frame with different background as compared to Fig 1.

Figure 3 shows original video frame with different background as compared to Fig 1&2.

Figure 2a shows the output after frame difference, which contains five people moving with respective velocities.
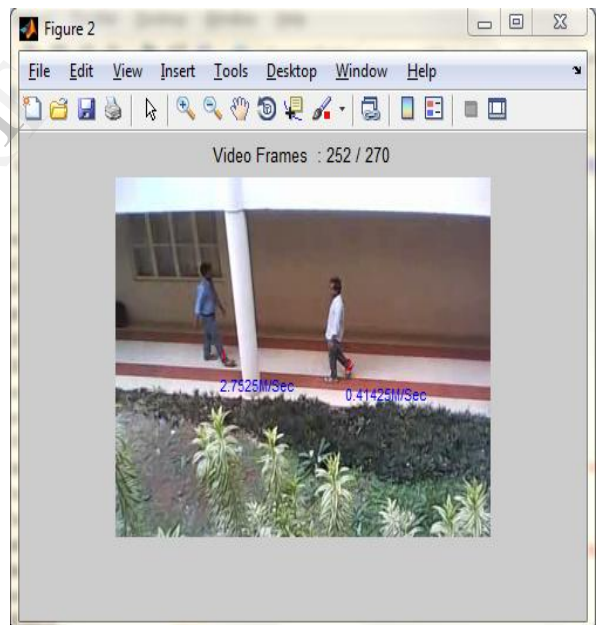

Figure 2a shows the output after frame difference, which contains two people moving with respective velocities.

V. CONCLUSION

This paper has proposed the methods to improve the performance of visual detection and tracking framework for surveillance with counting applications and determining the velocities and further works can be done on more challenging videos like the videwhich consist of varying background or the videos which contain noise present in it.

REFERENCES

[1] M. Piccardi, "Background subtraction techniques: a review", IEEE Proc. of International Conference on Systems, Man and Cybernetics, vol. 4, pp. 3099-3104, Oct. 2004.

[2] C. R. del Blanco, F. Jaureguizar, and N. García, "Visual tracking of multiple interacting objects through raoblackwellized data association particle filtering," IEEE Int. Conf. Image processing, pp. 821–824, 2010.

[3] Implementation of an Automated Single Camera Object Tracking System Using Frame Differencing and Dynamic Template Matching 'Karan Gupta1, Anjali V. Kulkarni2 Indian Institute of Technology, Kanpur, India

[4] Motion Object Detection of Video Based on Principal Component Analysis'Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008.

[5] Moving object detection tracking system: a real time implemented, SEIZIÈME COLLOQUE GRETSI — 15-19 SEPTEMBRE 1997.

[6] Detecting Moving People in Video Streams 'Department of computer science, university of Udine. Segmentation and tracking of multiple video objects Received 7 June 2005; received in revised

[7] Moving object detection using region tracking 'Eun Young Song · Ju-Jang Lee Received and accepted: November 25, 2003.

[8] O. Tuzel, F. Porikli, and P. Meer, "Learning on lie groups for invariant detection and tracking," IEEE Proc. of International Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.

[9] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," Pattern Recognition Letters, vol. 27, pp. 773–780, 2006.

[10] R, Y. Da Xu and M. Kemp, "Fitting multiple connected ellipses to an image silhouette hierarchically". IEEE Trans. on Image Processing. vol. 19, no. 7, 1673-1682, Jul 2010.

[11] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussianbayesian tracking," IEEE Trans. on Signal Processing, vol. 50, no. 2, pp. 174-188, 2002.

[12] A. Doucet, S. Godsill, and C. Andrieu, "On sequential montecarlo sampling methods for bayesian filtering," Statistics and Computing, vol. 10, no. 3, pp. 197-208, 2000.

[13] J. S. Kim, D. H. Yeom, Y. H. Joo, "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems", IEEE Trans. on Consumer Electronics, vol. 57, no. 3, pp. 1165-1170, Aug. 2011.

[14] Viola P., Jones M., "Rapid Object Detection using a Boosted Cascade of Simple Features", Mitsubishi Electrical Labs, Cambridge, USA, 2001.