ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

Video Based Crowd Density analysis for Visual Surveillance

ONKAR S. KATAGI[1], JAYALAKSHMI D S[2], VEENA G S[3]

M.Tech, Department of Computer Science & Engineering, M.S.Ramaiah Institute of Technology, Bangalore, India[1]

Associate Professor, Department of Computer Science & Engineering, M.S.Ramaiah Institute of Technology, Bangalore, India[2]

Assistant Professor, Department of Computer Science & Engineering, M.S.Ramaiah Institute of Technology, Bangalore, India[3]

Abstract: Crowd density estimation is the challenging task for visual surveillance. Due to presence of high risk of degeneration, the safety of public events has large crowds and it has always been major concern. Video analysis techniques are becoming increasingly popular in the visual surveillance of public areas because of their great efficiency in gathering information and low cost in human resource. Its central topic is automatic analysis and detection of abnormal events. In monocular image sequences, the combined method of foreground based, feature based and group based methods is applied to extract crowd areas having motion. The specific number of persons and velocity of a crowd can be adequately estimated by the system from crowded areas. Having a camera network, crowd density can be predicted with the system. The system has lot of been applications such as in real applications, and experiments conducted in real scenes (station, college) demonstrate the effectiveness of this method.

Keywords: crowd density, human tracking, Haar based features.

I. INTRODUCTION

Video surveillance systems have long been in use to monitor security sensitive areas. The history of video surveillance can be divided in to three generations. The first generation surveillance systems (1GSS, 1960-1980) were based on analog sub systems for image acquisition, transmission and processing. 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. [1]The next 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. Due to dynamic environmental conditions such as illumination changes, shadows and waving trees branches in the wind object segmentation is a difficult and significant problem that needs to be handles well for a robust visual surveillance system. The steady population growth, along with the worldwide urbanization, has made the crowd phenomenon more frequent. It is not surprising therefore, that crowd analysis has received attention from technical and social research disciplines. The crowd phenomenon is of great interest in a large number of applications. Sociological and psychological studies on the crowd phenomenon make use of human observations. [2] MRF based method is used to estimate the density of the crowd .From the aspect of position of people in the crowd, the features meet the properties of Markov. The details of the processing of the crowd density estimation system based on MRF with input as features of the image. The algorithm can effectively count the number of people in the crowd. But then, as the number of people is not very large, its difficult make sure of its effectiveness in the case of crowded people.[3]

A foreground detection approach called OF-SMED which makes use of Lucas-Kanade OF (optical flow) and SMED (Separable Morphological Edge Detector). A perfect foreground cannot be obtained by using optical flow alone due to some brightness change. But, optimal foreground can be obtained by OF-SMED effectively. The approach OF-SMED combines the foreground together to eliminate noise. It has about eight times more computational power requirement, therefore, it is not suitable for real-time applications [4].

In [5] a method consists of a change detection algorithm that distinguishes the background from the foreground. This is done using a discontinuity preserving MRF-based approach where the information from different sources is combined with spatial constraints to provide a smooth motion detection map. An issue involves the formulation of a non stationary interaction model that accounts for the people size variations in depth. This method gives good results in subway scenes, but the minimization is very difficult and time-consuming.

The estimation of the number of people present in an area can be an extremely useful information both for security/safety reasons .The people counting present two conceptually different ways to face this task. In the direct approach (also called detection-based), each person in the scene is individually detected, using some form of segmentation and object detection; the number of people is then trivially obtainable. In the indirect approach (also called map-based or measurement-based), instead, counting is performed using the measurement of some features that do not require the separate detection of each person in the scene; these features then have to be put somehow in relation to the number of people. To consider this effect compute a rough estimate of the people density by measuring how close the interest points in the group, further problem are is the stability of the detected interest points. The points found are somewhat dependent on the perceived scale and orientation of the considered object: the same object will have different detected corners if its image is acquired from a different distance or when it has a different pose [6].

To estimate crowd velocity in a video sequence, perform KLT (Kanade Lucas Tomasi) [b] tracking on video frames, looking for interesting feature points in the scene and tracking them over time. Principal directions are simply computed from the direction histogram of crowd motion vectors. The speed is directly related to the length of crowd vectors. These methods cannot work for high density crowds. First, individual detection is impossible due to poor view angles. Another problem is that the gathering people may stay motionless for quite a while, thus foreground detection by background modelling or feature point tracking is very difficult [7]. We propose a method of detection of people in crowd by using Haar features, and estimating the crowd density and velocity. We also try for different video samples with different background and plot accuracy graph for it.

II. SYSTEM OVERVIEW

The paper is divided into following sections. Section I describe introduction and previous carried out on visual surveillance system. Section II describes overview of proposed system. Section III gives experimental results and evaluations. And conclude in section IV.

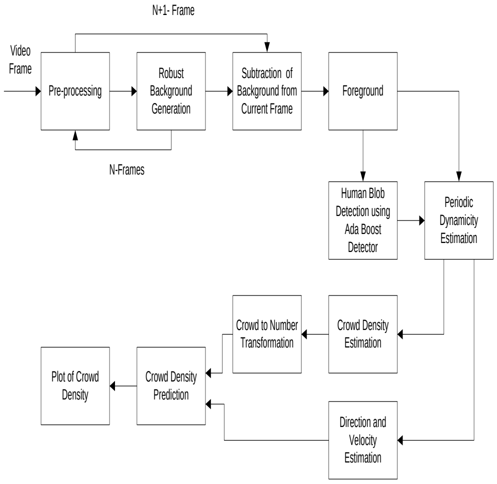

N+1- Frame
Video Frame
Pre-processing
Robust Background Generation
Subtraction of Background from Current Frame
Foreground
N-Frames
Human Blob Detection using Ada Boost Detector
Periodic Dynamicity Estimation
Crowd to Number Transformation
Crowd Density Estimation
Plot of Crowd Density
Crowd Density Prediction
Direction and Velocity Estimation

Fig 1. System Architecture

The figure 1 in above explains the architectural structure of the whole system. The system is initialized by feeding video imagery from a static camera monitoring a site. Most of the methods are able to work on both color and monochrome video imagery. The first step of method is distinguishing foreground objects from stationary background. To achieve this first N-frames are combined to generate robust background. This robust background is then subtracted from the input frame to extract foreground of the input. Human blobs are detected using AdaBoost detector and Periodic Dynamicity Estimation is performed to eliminate detected non-human movements if any. Then the crowd density is estimated. Also direction of movement of the crowd is estimated using motion vectors and the velocity is estimated using distance travelled by the crowd and time taken to travel that distance.. Hence the transformation from crowd density to number of people should be known. Subsequently the crowd density is estimated. This predicted density is represented as plot.

1. Pre-processing

The video of resolution 320×240 pixel is taken as input after that we extracting frames from video. After extracting frames from video, each frame undergoing pre-processing. In pre-processing step converting color image to gray image because time require for processing gray image is less compared to color image.

2. Background generation

In this algorithm, for the first N frames, number of columns and rows are computed for each frame and background image is generated by finding the mean of all the first N frames, then writing it to graphic file.

Step 1: Read the first N frames to an array after extraction of frames from video.

Step 2: Compute the differences between the frame elements such that,

Step 3: Determine the absolute value of each differences.

Step 4: Add the absolute values to form reference background frame.

Step 5: Store image to a graphic file as reference frame.

3. People Detection

This algorithm states that current frame is subtracted in reference frame to find the absolute difference. Difference image is converted to binary image also binary objects are removed. The image is validated for noise reduction and measuring the properties of the binary image where image properties are stored in a structure.

Now peoples in the image are detected by comparing the feature of the object detected with Haarcascade_fullbody.xml features. If the features match then corresponding people are detected. And name the people by numbering them.

Here morphological operation thin is used which removes pixel so that object without holes shrink to minimally connected stroke and objects with holes shrink to connected ring halfway between each hole and & outer boundary Properties such as area, centriod, orientation etc are measured and stored in a structure. Now the people objects in above context referred to the features of the image which are compared with standard Haar based full body features so if they match then people are detected.

Step 1: Generate frames from videos and read all the frames.

Step 2: Get reference frame and N+1 frame and convert both of them to gray scale

Step 3: Now find the Absolute differences between the two gray scale images.

Step 4: Convert images to binary image.

Step 5: Remove all the binary objects that have fewer pixel value than the threshold value in binary image.

Step 6: Validate the image by performing morphological operations on binary image and measure the properties of binary image.

Step 7: Detect people using Haarcascade AdaBoost Detector, which consist of .xml file of Haar features which are matched with the obtained features to detect people.

Step 8: Repeat the steps 2 – 7 till the end-frame.

Step 9: Identify the people by giving them a name. And store the number of people detected in a variable.

4. Crowd Density and Velocity Estimation

Crowd density is referred as number of people present in area under surveillance is given in equation 5.1. Its measured in per m^2.

$$\text{Crowd density} = \frac{\text{Number of people Detected}}{\text{Area under Survelliance}}. \qquad (1)$$

Area under Surveillance is provided manually. During people detection properties of the region are detected which includes centroid as a property in it. Let Centroid for a detected person at T frame be C($X_{1,}Y_1$). After Δt frames same person is detected at different position, its centroid is given by C(X_2,Y_2). Distance between the two positions is given by equation 2.

$$\text{Distance} = [(X_2 – X_1)^2 + (Y_2 – Y_1)^2]^{½} \qquad (2)$$

From Video properties we obtain number of frames generated per second. Therefore time taken for each frame is given by.

$$\text{time taken} = \frac{1}{\text{Frame rate}} \qquad (3)$$

Measured in seconds, now the velocity is obtained by

$$\text{Velocity} = \frac{\text{Distance}}{\text{time taken for each frame}} \qquad (4)$$

It’s measured in m/sec.

III. EXPERIMENTAL RESULT

This section represents snapshots of moving people detection using Haar features under video surveillance system.

Figures below shows 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.

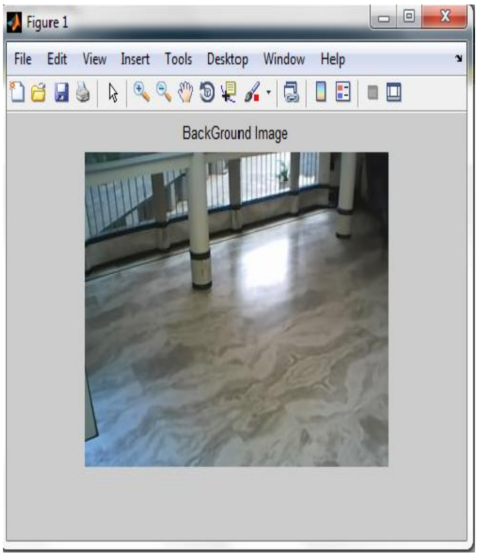


Figure 2. Generated background for video sample 1

The above figure 2 shows the background generation for a video sample 1. The frames from the video are obtained and each frame is processed. And the difference between the foreground and background image is obtained, after validation of the frame, properties like centroid, area, and only that part of the image is cropped and determine whether it is a human and numbering is done and its velocity is determined based on its centroid value by using distance formula and thus crowd density, velocity are estimated using equations 1 & 4. In figure 3 shows the number of detected people with their velocities.

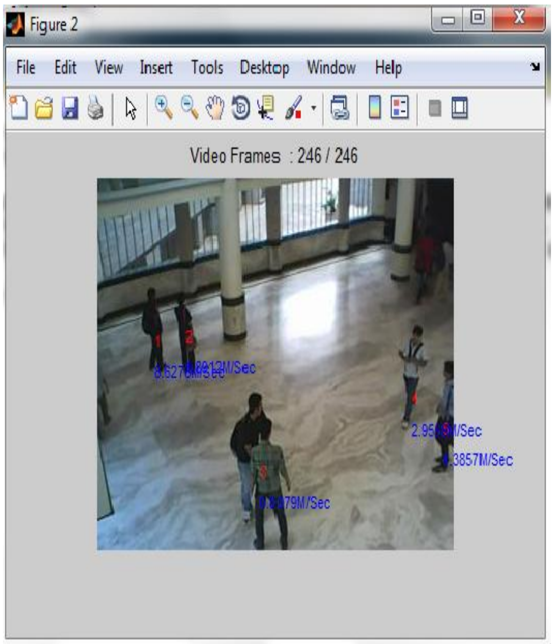


Figure 3 Number of people Detected along with their velocities.

Similarly for below figure 4 shows background generation for a video sample 2. The frames from the video are obtained and each frame is processed to detect people in the video and their velocities. In figure 5 shows the number of detected people with their velocities.

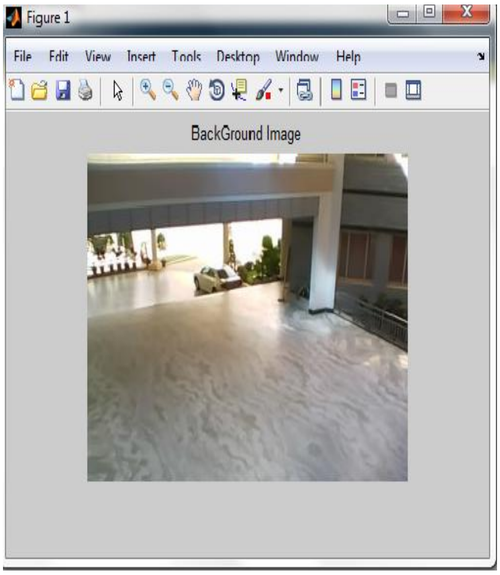


Figure 4 Generated background for video sample 2.

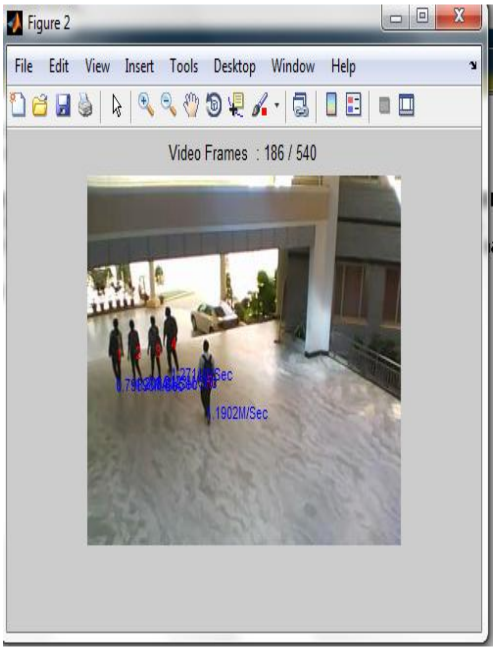


Figure 5. Number of people Detected along with their velocities

Frame number	Ground Truth	No of people Detected
20	3	2
40	5	4
60	6	4
80	7	5
100	7	5
120	6	6
140	7	6
160	7	5
180	5	5
200	7	5

Table I. Ground truth and detected people values for video sample 1

For above table I which contains number of people detected at frame number and also has ground truth which is refered as actual peole present at that moment for video sample 1 .A graph is plotted on the based on the number of people detected along y-axis vs no of frames at x-axis for above videos.

Here a blue line referred as number of people present in video irrespective of detection. Its known as Ground truth. Red line is reffered to number of people detected.

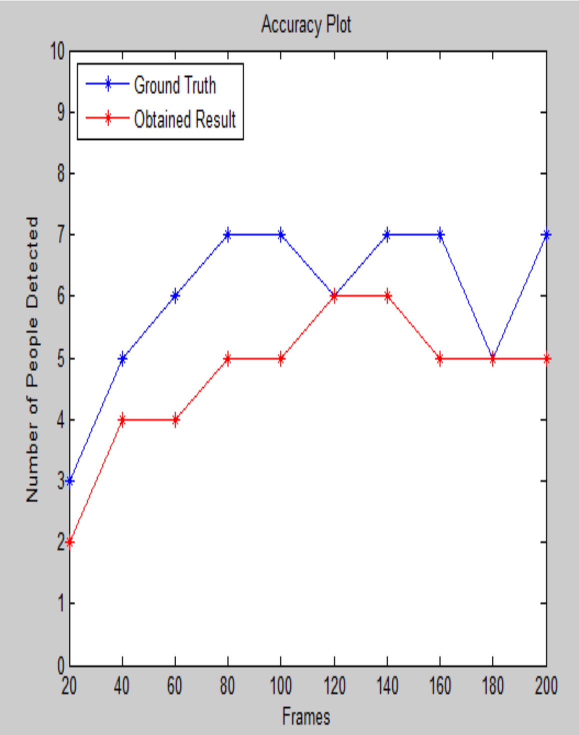


Figure 6.7 accuracy graph for video 1

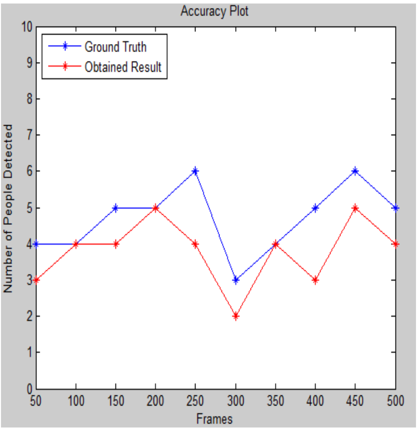
Frame number	Ground Truth	No of people Detected
50	4	3
100	4	4
150	5	4
200	5	5
250	6	4
300	3	2
350	4	4
400	5	3
450	6	5
500	5	4

Table II. Ground truth and detected people values for video sample 2

Similarly for above table II. which contains number of people detected at frame number and also has ground truth which is refered as actual peole present at that moment for video sample 2 .

A graph is plotted in similar way as the earlier

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940



(b)Figure 6.8 accuracy graph for video 2

Precision is Calculated using formula

$$\text{Precision} = \frac{Total\ number\ \ of\ people\ \ detected}{Ground\ \ trut\,h} \qquad (5)$$

For the video samples 1 & 2 the precision value is 78.3% and 80.8 % which are better computatinal result for detection of people and estimating crowd and velocity.

IV. CONCLUSION

Here crowd density estimated for wide-area security. Collective method of foreground detection feature detection and group based detection are used to detect crowded areas and also reduce perspective distortion. The number of people in a crowd is estimated by the AdaBoost detector which uses Haar features and the velocity is also obtained by finding the distance covered & time ratio and also direction by motion vectors. After crowd density and velocity are estimated. Compared to existing methods, the this method is a real time system for applications and the crowd density analysis algorithm can work properly in both low and high crowd density scenes. Experiments and real applications demonstrate the effectiveness and robustness of our method in real scenes although there are some aspects to be improved in the system.

REFERENCES

[1] Yigithan Dedeoglu, “Moving Object Detection, Tracking and Classification For Smart Video Surveillance”, Department of Computer Engineering, Bilkent University, 2008.

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

[3] ZHAN Beibei, MONEKOSSO D N, REMAGNINO P, et al. Crowd Analysis: A Survey[J]. Machine Vision and Applications, 2008, 19(5-6): 345-357

[4] HORN B K P, SCHUNCK B G. Determining Optical Flow[J]. Artificial Intelligence, 1981, 17(1-3):185-203.

[5] PARAGIOS N, RAMESH V. A MRF Based Approach for Real-Time Subway Monitoring[C] // Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition: December 8-14, 2001. Kauai, HI, USA, 2001: 1034-1040.

[6] CONTE D, FOGGIA P, PERCANNELLA G, et al. A Method for Counting People in Crowded Scenes[C] // Proceedings of 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS): August 29-September 1, 2010. Boston, MA, USA, 2010: 225-232.

[7] ANDRADE E L, BLUNSDEN S, FISHER R B. Modeling Crowd Scenes for Event Detection[C]// Proceedings of the 18th International Conference on Pattern Recognition, 2006 (ICPR 2006): August 20-24, 2006. Hong Kong, China, 2006: 175-178.