Background Modelling, Detection and Tracking of Human in Video Surveillance System

Rajvir Kaur
Discipline of ECE
Lovely Professional University
Phagwara, Punjab (INDIA)
er.rajvir.seehra@gmail.com

Sonit Singh
Discipline of ECE
Lovely Professional University
Phagwara, Punjab (INDIA)
enggsonit7@gmail.com

***Abstract*—Video Surveillance System is a powerful tool used for monitoring people and their activities for public security. The motive of having surveillance system is not only to put cameras in place of human eyes, but also making it capable for recognizing activities automatically. In this paper, human detection and tracking is performed on Weizmann dataset having various activities like run, bend, hand wave, skip, etc. First background modelling is done by taking mean of first n frames. After this, human detection is done using background subtraction algorithm and then tracking is done using Kalman filter. Result of each stage has been discussed. The proposed methodology shows promising results which can further be used for activity recognition.**

Keywords-Video surveillance systems, Background subtraction, Gaussian Mixture Model, Kalman filter.

I. INTRODUCTION

Human action detection and tracking in a video surveillance system is an active research area in image processing and computer vision system. Due to increase in terrorist activities and many general social problems, security has become the top most priorities of all nations [3]. So, there is a need for effective monitoring of public places for security at airports, railway stations, shopping malls, banks, etc. To monitor the activities of a human, surveillance cameras are used. Surveillance cameras are used for monitoring banks, department stores, museums, patrolling of highways and railways for accident detection, for fire detection, patrolling national borders, monitoring peace treaties and so many other applications[11]. They are also used for observing the activities of elderly and infirm people for early alarms. In traditional surveillance system [3], the video is captured by the camera and is displayed on the monitor in a control room. To monitor the videos, human resources are presents in a control room and continuously monitor the video for recognizing the activities. In many situations [1], it is common to find poor monitoring due to human factor like fatigue because it is very tedious or boring task to continuously monitor the scene because sometimes nothing strange or uncommon thing happens in a scene that catches the attention. So, there is a need to design an intelligent surveillance system that can automatically detect and track the object (i.e. Human in our case) in a video. This paper is divided into three stages: Background Modelling stage, Human Detection stage and Human Tracking stage. First Background Modelling is done then Human Detection and at last Human Tracking. The very first step is to model the background. After Background modelling Human Detection is done. Human detection can be done by using algorithms like Background subtraction, Optical flow, Gaussian mixture model, Temporal differencing and so on. The commonly used algorithm is Background subtraction because of its simplicity. After detection, tracking is done. For tracking, we can use Kalman filter, Camshift, Meanshift, Particle filter, etc and the most commonly used is Kalman filter because it is recursive, adaptive filter that is well known for its ability to track the object in a timely and accurate manner. Section II gives the overview of video surveillance system. Section III gives related work followed by proposed methodology in section IV. In section V results are discussed, following which conclusion and future work is given in section VI.

II. OVERVIEW OF VIDEO SURVEILLANCE SYSTEM

The video is continuously captured by the surveillance camera deployed at the suitable locations at different places. The flow of work can be explained well by the step by step approach as given below:

Step 1: Video: Video is captured by the surveillance camera deployed in public spaces.

Step 2: Background Modelling: Background modelling is a technique used to detect moving objects in video acquired from cameras. The background modelling determines how the background looks and it also determines which areas of the image may contain interesting information and it is an essential step in any detection system. To model the background in video frames, first a background is constructed and then the current frame is subtracted from the background and the difference determines the moving objects. A good background modelling method should be able to adapt the following scenery changes [5]. These are gradual variations of lighting condition in the scene, small movements of the non-static objects such as tree branches and bushes blowing in the wind, sudden changes in the light condition, e.g. raining, sudden change from daylight to lights in the evening, shadow regions and multiple objects moving in the scene both for long and short periods.

Step 3: Human Detection: Human detection means to identify the presence of humans in a video sequence and differentiating

©2014 CIPECH14.

them from non-human objects. In detection process, the foreground is separated from the background in order to detect the moving object. Due to the dynamic changes of the background image, such as weather, light, shadow and other interfering factors, motion detection becomes difficult task.

Step 4: Human Tracking: The goal of human tracking is to derive a correspondence of the human detected in one frame with the human detected in the next frame. If the features are matched, then the human detected in the current and the previous frame is said to be the same. Features can be color, orientation, speed, posture, intensity or any other information that can be obtained from a pixel. The main purpose of the tracking is to carry out real-time tracking on the detected target and to calculate the tracking target in the exact location of the image scene, moving speed and other information. The primary purpose is to track the moving object when the monitoring environment is dynamic.

Fig 1. Flowchart of video surveillance system.

III. RELATED WORK

A. Background Modelling

Background modelling is a technique for extracting the moving object in video frames. There are various methods to do Background modelling. During converting the video into frames, sometimes we can get a frame which we can use as a background. If there is no single frame which we can use as a background, then we need to model the background by using some techniques such as taking mean or median of n number of frames, Adaptive Gaussian mixture model [4].

B. Human Detection

In Human Detection, the human is detected in the area under surveillance. There are various algorithms [10] that are used for human detection like Optical Flow algorithm, Background Subtraction algorithm, Temporal Differencing, Gaussian mixture model, Min-Max method, Kernel density estimation (KDE), Eigen backgrounds, Codebook (CB_{RGB}), and so any others. Optical flow algorithm [3] can be used to detect independently moving targets in the presence of camera motion, however Optical flow method is very complex and very sensitive to noise and is not applicable to real-time algorithms. Temporal differencing do the pixel-wise difference between two or three consecutive frames in an image sequence to extract moving regions. The advantage of this technique is that it is very adaptive to dynamic environments, but does a poor job of extracting all relevant feature pixels. Background subtraction [3] is a popular method for human detection where the background is static in nature and it attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel manner. But it is extremely sensitive to changes of dynamic scenes due to lightening and extraneous events. Therefore, it is highly dependent on a good background model to reduce the influence scenery changes.

C. Human Tracking

Human Tracking means deriving a correspondence of the human detected in one frame with the human detected in the next frame. If the features are matched, then the human detected in the current and the previous frame is said to be the same. Useful mathematical tools for tracking include the Kalman filter, Condensation algorithm, dynamic Bayesian network, Camshift, Meanshift, Particle filter, etc [9]. Tracking methods [6] are divided into four major categories: region-based tracking, active-contour based tracking, feature based tracking and model-based tracking. In region based tracking, the features of the blob, detected in one image frame are matched to the blob detected in the other frame. If there is a match then the detected image is linked with the image in the previous frame. Region based tracking works well in scenes containing only few objects, but cannot reliably handle occlusion between objects. Active contour-based tracking algorithms track objects by representing their outlines as bounding contours. The advantage of active contour-based algorithms is that they describe objects more simply and more effectively and reduce computational complexity, but they are highly sensitive to the initialization of tracking, making it difficult to start tracking automatically. Feature-based tracking algorithms perform tracking of objects based on extracting relevant features based on various feature extraction techniques which are invariant of various affine transformations, occlusion, etc [12]. These algorithms can adapt successfully with changing background or other dynamic changes in environment, allow fast real-time processing and tracking of multiple objects, handle partial occlusion by using information on object motion, local features and dependence graphs.

IV. PROPOSED METHODOLOGY

A. Background modelling using mean

To estimate the background we have taken mean of n number of frames. First convert the video into frames. If the background is static and the object is moving then we can easily estimate the background by taking mean of all the frames and it is given by:

$$B(x, y, t) = \frac{1}{n}\sum_{i=0}^{n-1} I(x, y, t-i) \qquad (1)$$

where, $B(x, y, t)$ = Background estimated at time t,

n = Total number of frames,

$I(x, y, t-i)$ = image at time $t-i$.

If the background is varying in nature or there is change in illumination or lightening effect then mean does not give accurate background. For this we can use Gaussian Mixture Model.

B. Human detection using Background Subtraction algorithm

After background modelling, detection is done. By using Background subtraction algorithm we can separate the Foreground and Background and then we will be able to detect the human. Background subtraction detects the moving region in an image by differencing the current image and a reference image in a pixel-by-pixel form as shown by equation:

$$\left|I_c - I_{bk}\right| > T \tag{2}$$

Where, I_c = current image

I_{bk} = background or reference image

T =Thresholdvalue.

When the image difference is above the threshold value, it is considered as the foreground image. After background subtraction, morphological operation is applied if the output is not coming proper. Morphological operation is used to remove the erroneous blobs present in the image.

C. Human tracking using Kalman filter

For human tracking we are using Kalman filter though there are many others like meanshift, camshift, particle filter. Kalman filter is a recursive, adaptive filter that is well known for its ability to track the object in a timely and accurate manner. The kalman filter is capable to track the person correctly. The Kalman filter[2] estimates the position of the object in each frame of sequence. The input parameters of the kalman filter are the position of the object in the image, the size of the object and the width and the length of the search window of the object.

The variable parameters of the Kalman filter are the state vector and the measurement vector. The state vector is composed of the initial position, width and length of the search window and the center of the mass of the object [8]. And the measurement vector is composed of the initial position, width and length of the search window.

The kalman filter estimates the state of a discrete time controlled process that is modeled by the linear equation:

$$x_k = Ax_{k+1} + Bu_k + w_{k-1} \tag{3}$$

with a measurement z$\in \Re^m$ that is

$$z_k = Hx_k + v_k \tag{4}$$

The random variables w_k represents the process noise and v_k represents the measurement noise. They are assumed to be independent, white and with normal probability distributions

$$p(w) \sim N(0, \mathrm{Q}), \tag{5}$$

$$p(v) \sim N(0, \mathrm{R}). \tag{6}$$

The process noise covariance Q and measurement noise covariance R matrices might change with each time step or measurement, however here we assume they are constant.

The $n \times n$ matrix A in the difference equation relates the state at the previous time step $k-1$ to state at the current step k. The Kalman filter estimates a process by using a form of feedback control: the filter estimates the process state at some time and then obtains feedback in the form of measurements. The equations for kalman filter fall into two groups: time update equation and measurement update equation.

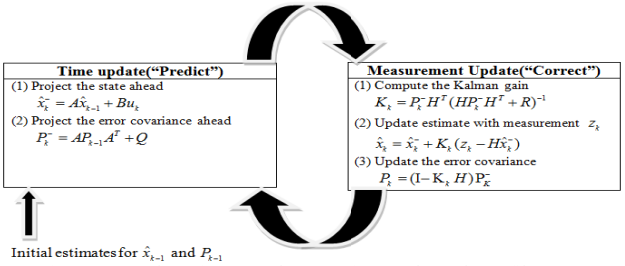


Fig 2. A complete picture of the operation of the Kalman filter [8].

V. RESULTS AND DISCUSSION

There are various databases [7] available like KTH, WEIZMANN, UT-Interaction, MSR Action, etc which contain different number of action classes like running, bending, jumping, walking, hand-waving, skipping, etc. And we using WEIZMANN database which contains 10 action classes likes walk, run, jump, gallop sideways, bend, one hand wave, two hand waves, jump in place, jumping jack, skip.

Fig.3. Selected frames of the videos of various human activities from Weizmann Dataset [7] a). Bend (b). Gallop Sideway, (c). One-hand wave, (d). Two-hand wave, (e).Jump, (f). Skip (g). Walk (h). Jumping Jack & (i). Run .

A. Results of background modelling

There are different action classes and we have done our experiments on different activities done by a single person like running, bending, one hand-waving, two hand-waving and skipping activity. The results of background modeling are:

For running activity: Fig 4(a) is a single frame taken from a video from Weizmann dataset for running activity and figure 4(b) is the estimated background. Similarly, for other activities like bending activity, one hand waving activity, two hand waving activities, skipping activity background is estimated and the results are shown below:

(a) (b)

Fig 4.(a) Single frame taken from video from Weizmann dataset.
(b) Background estimated from video frame.

Bend Activity:

(a) (b)

Fig 5. (a) Single frame taken from video from Weizmann dataset.
(b) Background estimated from video frames.

One Hand Wave Activity:

(a) (b)

Fig 6. (a) Single frame taken from video from Weizmann dataset.
(b) Background estimated from video frames.

Two Hand Wave Activity:

(a) (b)

Fig 7. (a) Single frame taken from video from Weizmann dataset.
(b) Background estimated from video frames.

Skip Activity:

(a) (b)

Fig 8. (a) Single frame taken from video from Weizmann dataset.
(b) Background modelling from video frames.

Gallop Sidewalk Activity:

(a) (b)

Fig 9. (a) Single frame taken from video from Weizmann dataset.
(b) Background modelling from video frames.

Jumping Jack Activity:

(a) (b)

Fig 10. (a) Single frame taken from video from Weizmann dataset.
(b) Background modelling from video frames.

B. Results of Human Detection

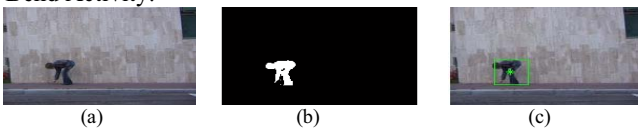
Results of human detection are obtained by using Background subtraction algorithm. The output will be a binary image. If the foreground is not coming properly, then we can apply morphological operations in order to make the foreground proper.

For running activity: Fig 11 (a) is the single frame taken from video from Weizmann dataset for running activity and Fig 11(b) is the foreground detected using Background subtraction algorithm and Fig 11 (c) is the detected human surrounded by the green color bounding box and "*" points the centroid. Similarily for other activities like bending, one hand waving, two hand waving, skipping activity results are shown below:

(a) (b) (c)

Fig 11. (a) Single frame taken from video from Weizmann dataset. (b) Foreground detection after background subtraction algorithm. (c) Detected Human shown as single frame from the Video.

Bend Activity:



(a) (b) (c)

Fig 12. (a) Single frame taken from video from Weizmann dataset. (b) Foreground detection after background subtraction algorithm. (c) Detected Human shown as single frame from the Video.

One Hand Wave Activity:

(a) (b) (c)

Fig 13. (a) Single frame taken from video from Weizmann dataset. (b) Foreground detection after background subtraction algorithm. (c) Detected Object Human shown as single frame from the Video.

Two Hand Wave Activity:

(a) (b) (c)

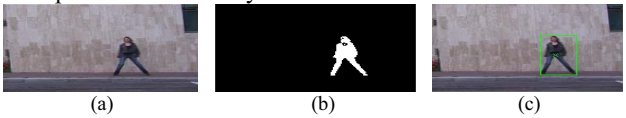
Fig 14. (a) Single frame taken from video from Weizmann dataset. (b) Foreground detection after background subtraction algorithm. (c) Detected Object Human shown as single frame from the Video.

Skip Activity:

(a) (b) (c)

Fig 15. (a) Single frame taken from video from Weizmann dataset. (b) Foreground detection after background subtraction algorithm. (c) Detected Object Human shown as single frame from the Video.

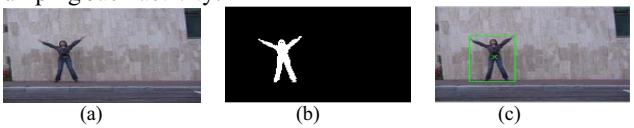
Gallop Sidewalk Activity:



(a) (b) (c)

Fig 16. (a) Single frame taken from video from Weizmann dataset. (b) Foreground detection after background subtraction algorithm. (c) Detected Object Human shown as single frame from the Video.

Jumping Jack activity:



(a) (b) (c)

Fig 17. (a) Single frame taken from video from Weizmann dataset. (b) Foreground detection after background subtraction algorithm. (c) Detected Human shown as single frame from the Video.

C. Results of Human Tracking

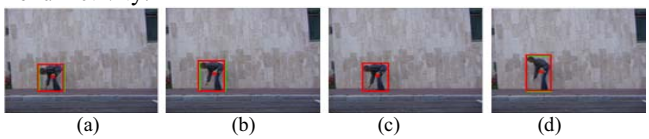
Human Tracking results are obtained by using Kalman filter. The output of the Kalman filter is shown by the red color bounding box and "*" is the Centroid. The Human tracking results for different activities are shown below:

Run Activity:

(a) (b) (c) (d)

Fig 18: (a), (b), (c), (d) Tracking output for Run activity.

Bend Activity:



(a) (b) (c) (d)

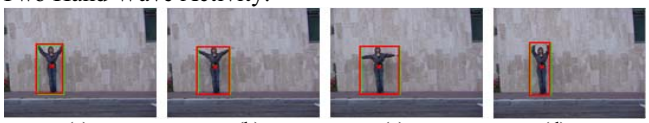
Fig 19: (a), (b), (c), (d) Tracking output for bend activity.

One Hand Wave Activity:

(a) (b) (c) (d)

Fig 20: (a), (b), (c), (d) Tracking output for One hand wave activity.

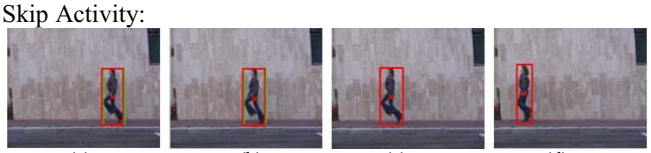
Two Hand Wave Activity:



(a) (b) (c) (d)

Fig 21: (a), (b), (c), (d) Tracking output for Two hand wave activity.

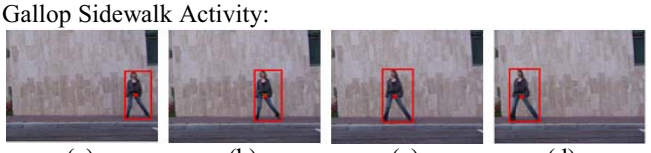
Skip Activity:



(a) (b) (c) (d)

Fig 22: (a), (b), (c), (d) Tracking output for Skip activity.

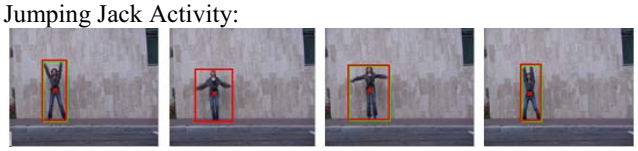
Gallop Sidewalk Activity:



(a) (b) (c) (d)

Fig 23: (a), (b), (c), (d) Tracking output for Gallop sidewalk activity.

Jumping Jack Activity:



(a) (b) (c) (d)

Fig 24: (a), (b), (c), (d) Tracking output for Jumping Jack activity.

VI. CONCLUSION AND FUTURE WORK

Video surveillance system is an active research area in computer vision system because of its various applications in public security, in military security, bank security, sports,etc. So, there is a need to design an intelligent surveillance system that should be capable to detect, track end recognize the activities of humans automatically. In this work, first of all background estimation/modelling has been done on the video captured by the camera by taking mean of n frames. After this, human detection has been done using background subtraction algorithm and then morphology operation is applied. At last, tracking is done using Kalman filter. Further, results of each stage has been discussed in detail. In future, the proposed work will be used for activity recognition in video surveillance system.

REFERENCES

[1] Oluwatoyin P. Popoola, Kejun Wang, "Video-Based Abnormal Human Behavior Recognition- A Review," IEEE Transactions on systems, man and cybernetics, vol.42, no.6, pp.865-878, Nov.2012.

[2] Afef SALHI and Ameni YENGUI JAMMOUSSI, "Object Tracking using Camshift, Meanshift and Kalman filter", World Academy of Science, Engineering and Technology 64 2012.

[3] Helly Patel, Maheah P.Wankhade, "Human Tracking in Video Surveillance", International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459), volume 1, December 2011.

[4] Chris Stauffer, AW.E.L Grimson, " Adaptive Background Mixture Models for real time tracking", The Artificial Intelligence Laboratory Massachuestts Institute of Technology Cambridge, MA 02139.

[5] Rupali S. Rakibe, Bharati D. Patil, "Background Subtraction algorithm based Human Motion Detection", International Journal of Scientific and Research (ISSN 2250-3153), volume 3, issue 5, May 2013.

[6] Weiming Hu, Teiniu Tan, Liang Wang and Steve Maybank, "A Survey on the Visual Surveillance of Object motion and Behaviors", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Volume. 34, No. 3, August 2004.

[7]. www.cs.utexax.edu/~chaoyeh/web-action-data/datasetlist.html.

[8]. G. Welch, G Bishop, "An Introduction to the Kalman Filter", Technical report : TR95-041, University of North Carolina, 2006.

[9]. Caius Suliman, Cristina cruceru, Florin Moldoveanu, "Kalman filter based tracking in a video surveillance system", 10th International conference on development and applications systems, May 27-29, 2010.

[10]. Aravinda S Rao, Jayavardhana Gubbi, Slaven Marusic, Marimuthu Palammmiswami, "A robust algorithm for foreground extraction in crowded scenes", ISCIT, IEEE, 2012.

[11]. Sarvesh Vishwakarma, Anupam Aggarwal, "A survey on activity recognition and behaviour understanding in video surveillance", Springer Verlag, 2012.

[12]. Ahlem Wehla, Ali weli, Adel Al alimi, "Video stabilization with moving object detection and tracking for aerial video surveillance", Multimedia Tools & Applications, Springer Science Business Media NY 2014.