

# HUMAN TRACKING AND ACTIVITY RECOGNITION FOR SURVEILLANCE APPLICATIONS

*A thesis submitted in partial fulfilment of the requirement for the degree of*

Master of Technology

In

Electronics and Communication Engineering

Specialization: Signal and Image Processing

By

**Suraj Prakash Sahoo**

Roll No: 213EC6268

Under the Guidance of

**Dr. Samit Ari**



Department of Electronics and Communication Engineering

National Institute of Technology Rourkela

Rourkela, Odisha, 769008, India

May 2015

# HUMAN TRACKING AND ACTIVITY RECOGNITION FOR SURVEILLANCE APPLICATIONS

*A thesis submitted in partial fulfilment of the requirement for the degree of*

Master of Technology

In

Electronics and Communication Engineering

Specialization: Signal and Image Processing

By

**Suraj Prakash Sahoo**

Roll No: 213EC6268

Under the Guidance of

**Dr. Samit Ari**



Department of Electronics and Communication Engineering

National Institute of Technology Rourkela

Rourkela, Odisha, 769008, India

May 2015



DEPARTMENT OF ELECTRONICS AND COMMUNICATION  
ENGINEERING,  
NATIONAL INSTITUTE OF TECHNOLOGY,  
ROURKELA, ODISHA -769008.

## CERTIFICATE

This is to certify that the work done in the thesis entitled **Efficient Human Tracking And Activity Recognition For Surveillance** by **Suraj Prakash Sahoo** is a record of an original research work carried out by him in National Institute of Technology, Rourkela under my supervision and guidance during 2014-2015 in partial fulfilment for the award of the degree in Master of Technology in Electronics and Communication Engineering (Signal and Image Processing), National Institute of Technology, Rourkela.

Place: NIT Rourkela

Date:

**Dr. Samit Ari**

Asst. Professor



DEPARTMENT OF ELECTRONICS AND COMMUNICATION  
ENGINEERING,  
NATIONAL INSTITUTE OF TECHNOLOGY,  
ROURKELA, ODISHA -769008.

## DECLARATION

I certify that,

- a. The work presented in this thesis is an original content of the research done by myself under the general supervision of my supervisor.
- b. The project work or any part of it has not been submitted to any other institute for any degree or diploma.
- c. I have followed the guidelines prescribed by the Institute in writing my thesis.
- d. I have given due credit to the materials (data, theoretical analysis and text) used by me from other sources by citing them wherever I used them and given their details in the references.
- e. I have given due credit to the sources (written material) used by quoting them where I used them and have cited those sources. Also their details are mentioned in the references.

**Suraj Prakash Sahoo**

# ACKNOWLEDGEMENT

This research work is one of the significant achievements in my life and is made possible because of the unending encouragement and motivation given by so many in every part of my life. The work is incomplete without mentioning about them.

Firstly, I would like to express my gratitude and sincere thanks to **Dr. Samit Ari**, Asst. Prof., Department of Electronics and Communication Engineering for his esteemed supervision and guidance during the tenure of my project work. His invaluable advices have motivated me a lot when I feel saturated in my work. His impartial feedback in every walk of the research has made me to approach a right way in excelling the work. It would also like to thank him for providing best facilities in the department. I would also like to thank all the faculty members of the EC department, NIT Rourkela for their support during the tenure spent here.

I would like to express my heartfelt wishes to my brothers, friends and classmates whose company and support made me feel much better than what I am. I would like to mention my special thanks to my close friends Sobhan, Satish, Anusha and Sowjanya. I would like to thank Chandra, Ujwal, Jitu, Sudipto, Rashmi, Abhijit to their help in creating my database.

Lastly, I would like to express my love and heartfelt respect to my parents, sister and brothers for their consistent support, encouragement in every walk of my life without whom I would be nothing.

**Suraj Prakash Sahoo,**

surajprakashsahoo@gmail.com

# ABSTRACT

Tracking and study of behavioural changes of human beings through vision is a challenging task. For surveillance, automated systems are important which can observe the traffic and can detect the abnormality. For tracking human or any kind of object, colour feature based mean shift technique is widely used. This technique uses Bhattacharya coefficient to locate the object based on the maximisation of the similarity function between object model and candidate model. Traditional mean shift algorithm fails when the object having large motion, occlusion, corrupted frames etc. In addition to that, the technique is not automatic to initiate the tracking. To overcome all these problems, this thesis work proposed a technique which uses three additional modules to the traditional method to make it more efficient. The proposed modules used human detection by modelling through star skeletonization, followed by block search algorithm and occlusion handling. Block search algorithm helped to supply an overlapping area to candidate model to continue the track when tracking fails due to fast motion. Occlusion handling helped in initiating the tracking after prolonged period of occlusion. The proposed method has been tested on real time data and it outperforms the conventional method effectively to overcome the mentioned problems up to large extent. Human activity recognition is a hierarchical procedure which confirms abnormality step by step. Low level activity recognition is a trajectory based application in which trajectory of tracks of a human being helps to detect the abnormal events like person fell down, illegal entry, abnormal loitering, line formation etc. At high level, human pose will be detected by the help of shape based human pose detection. The main aim of the system is to make a person independent real-time human activity recognition with decreased false alarm rates.

# INDEX

<b>Acknowledgement</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Index</b>	<b>iii</b>
<b>List of Figures &amp; Tables</b>	<b>v</b>
<b>Abbreviations</b>	<b>vii</b>
<b>Chapter 1: Introduction to the System</b>	<b>1</b>
1.1 Introduction	2
1.2 System Overview	2
1.3 Literature Survey	3
1.4 Traditional Tracking Methods	4
1.4.1 Background subtraction Based simple Tracking	5
1.4.1.1 Using frame differencing	5
1.4.1.2 Mean filter	6
1.4.1.3 Adaptive Background Subtraction	6
1.4.1.4 Background Modeling and Subtraction	7
1.4.2 Optical flow	7
1.4.2.1 Horn-Schunk method	9
1.4.2.2 Lucas & kanade method	9
1.4.3 Kalman Filter Based Tracking	10
1.4.4 Mean Shift Tracking	11
1.5 Thesis outline	13
<b>Chapter 2: Automated Human Tracking Using Advanced Mean shift Algorithm</b>	<b>14</b>
2.1 Introduction	15
2.2 Problem Description	15
2.2.1 Manual or Semi-Automatic Technique	15
2.2.2 Fast Motion of Object	15
2.2.3 Prolonged Occlusion	16
2.3 Proposed Framework	16
2.3.1 Simple Background Subtraction	18
2.3.2 Human Modelling	19
2.3.3 Block Matching Technique	21
2.3.4 Occlusion Handling	22

2.4 Results and Discussion	24
2.4.1 Block Matching	24
2.4.2 Occlusion Handling	25
2.5 Summary	28
<b>Chapter 3: Activity Recognition</b>	<b>29</b>
3.1 Introduction	30
3.2 Low level Activities	31
3.2.1 Person fell down	31
3.2.2 Loitering for large time and entering into Restricted area	32
3.2.3 Line formation	34
3.3 High level Activities	34
3.4 Results and Discussion	36
3.5 Summary	38
<b>Chapter 4: Conclusions and Future Work</b>	<b>39</b>
<b>Publication</b>	<b>41</b>



## LIST OF FIGURES & TABLES

Fig. 1.1	: Basic System Overview	3
Fig. 1.2	: Basic Tracking Procedure	5
Fig. 1.3	: Codeword for background modelling	7
Fig. 1.4	: Simple optical flow based tracking using Harish corner	8
Fig. 1.5	: (a)Reference frame (b)Optical flow (c)Tracking Result	8
Fig. 1.6	: Kalman Filter based Tracking	11
Fig. 1.7	: Traditional Mean Shift Tracking	13
Fig. 2.1	: Flow chart for Advanced automated mean shift tracking	17
Fig. 2.2	: Local Maxima Calculation Procedure	19
Fig. 2.3	: Contour Detection Procedure	20
Fig. 2.4	: Distance Function	20
Fig. 2.5	: Star Structure with Binary Human Object	21
Fig. 2.6	: Star Skeleton Analysis	21
Fig. 2.7	: Forward blocks for block matching in the motion direction	22
Fig. 2.8	: Prolonged Occlusion Handling	23
Fig. 2.9	: (a) Patch fragmentation (b) Result of traditional tracking (c) Result of patch fragmentation based tracking.	24
Fig. 2.10	: Similarity parameter vs. frame number curves for (a) normal mean shift tracking (b) mean shift with block matching	25
Fig. 2.11	: Similarity parameter vs. frame number curves for (a) tracking without Occlusion Handling (b) tracking with Occlusion Handling.	26
Fig. 2.12	: Human tracked frames of the video. (a) Human detected and start of tracking, (b) Tracked frame before occlusion, (c) Tracked frame after occlusion, (d) Tracked frame toward end of frame.	27
Fig. 3.1	: Basics of Activity Recognition	30
Fig. 3.2	: Event Object	30
Fig. 3.3	: Person Fell down	31
Fig. 3.4	: Flowchart for Human Fell down detection	32
Fig. 3.5	: Flowchart for detection of illegal entry	33
Fig. 3.6	: Loitering and Entering into restricted zone	33
Fig. 3.7	: Line Formation	34

Fig. 3.8	: (a) Original images and background subtraction results, (b) Collected training data, (c) Examples of human detection in different poses	35
Fig. 3.9	: Person Fell Down (a) Tracking of fell down event, (b) Explanation of Fell Down Event, (c) False Alarm due to Occlusion, (d) False Alarm Explanation	36
Fig. 3.10	: Entering to Restricted Area (a) Background frame (b) Polygonal Restricted area mask (c) Person Enters the Restricted Area (d) Final output with restricted area marker	37
Fig. 3.11	: Loitering (a) loitering at a place (b) loitering mark up	38
Table 1	: Performance of Proposed method	27
Table 2	: Objective result of Activity Recognition	38

# ABBREVIATIONS

- MOG : Mixture of Gaussian  
HMM : Hidden Markov Model  
ANN : Artificial Neural Network  
SVM : Support Vector Machine

# **CHAPTER 1**

## **INTRODUCTION**

## **1.1. Introduction**

Tracking and study of behavioural changes of human beings through vision is a challenging task. For surveillance, automated systems are important which can observe the traffic and can detect the abnormality. For this purpose, current surveillance systems are having cameras in place. But, the problem with it is, it needs efficient manpower to observe and understand the acquired data. However, image understanding and risk detection cannot be left totally to the human security personnel as it needs careful observation over long periods of time. So these types of problems lead the motivation towards an automated intelligent system for risk detection or abnormality detection.

So a clear motivation is to develop an efficient intelligent system by keeping an eye on following points,

- The system must be capable of real-time human activity recognition. If the system's responses are slow, it is tedious to use.
- The system must be person independent. For most applications it is desirable that many potential users can operate the system.
- The system should be operated in real complex crowd environment for human tracking and activity recognition.
- The system should be able to decrease false alarm rates.

## **1.2. System Overview**

Human tracking and activity recognition for surveillance should be an online procedure in which the video data acquisition and processing of that data should be done simultaneously. So the first step is video acquisition which is followed by tracking of human to get tracklets for further study. For this purpose an efficient tracking method should be used which can handle a real time environment containing problems like illumination change, occlusion, lightening, fast motion etc. After getting the tracklets, the system will move to low level recognition based on track information. If abnormality found at low level recognition, the system will proceed to high level recognition to ensure the abnormality. In this procedure background subtraction will be done to extract abnormal movable objects which are then put

into shape based activity recognition. The main goal of the system is to reduce the rate of false alarm.

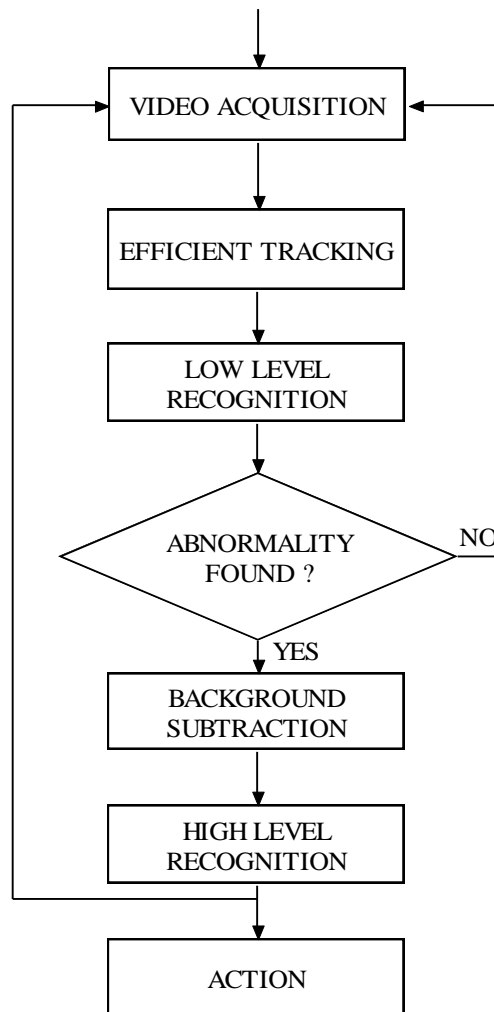


Fig. 1.1. Basic System Overview

### 1.3. Literature Survey

The first step of surveillance is tracking. The human should be tracked first before recognition. Various work has been done in this area, variety of methods have been attempted to track with maximum accuracy. Mean shift tracking, Kalman filter based tracking [1][2][3][4][5][6], optical flow are some of the varieties of tracking. The tracking may be of single camera tracking or can be multiple camera based tracking, it may of single pedestrian tracking [7][8][9][10][11], group tracking, or object detection.

Mean shift tracking is an easy method to work with and to increase the efficiency of traditional mean shift algorithm; various methods are reported from time to time. A robust object tracking approach using mean shift is presented in [12] using a weighted histogram and eliminating the background effect. Multi-fragment representations of the target and candidate models have been implemented to improve the robustness of tracking especially to partial occlusion in [13]. Using STAGE, object recovery can be implemented up to some extent as in [14]. An Adaptive Tracking window has been used to adopt scale and orientation change in [15].

After tracking the next step is activity recognition. The abnormal activities recognition is a challenging work. In this field a lot of work has been done. S. C. Lee and R. Nevatia have done hierarchical abnormal activity recognition [16] in which the activities are first recognised basing on tracklets and then at higher level by shape based human pose recognition. R. Bodor, B. Jackson and N. Papanikolopoulos have developed vision based low level human activity recognition [17] in which activities are recognised using human velocity, position, loitering for large time etc.

## **1.4. Traditional Tracking Methods**

Success of various available tracking algorithms mostly depends on the background features and movement of object in the video sequence. Problems which reduce the tracking efficiency are illumination changes, background clutter, and partial or full occlusions. A practical real time efficient tracking system should overcome all the problems and provide efficient performance by decreasing false alarm rate. Tracking is of two types: probabilistic and deterministic. In probabilistic model the system and its measurements remain uncertain (Kalman filtering, particle filter etc.) but, when occlusion is there, it works better than deterministic methods. Mean shift tracking belongs to the second category where a model will be compared with the current frame to find the most promising region. Various tracking methods are described briefly in following sections.

### **1.4.1. Background subtraction Based simple Tracking**

Basic tracking is a procedure which comprises of extracting the moving pixels and to track them in subsequent frames. For this purpose the background will be subtracted to extract the foreground moving object. The object can be detected by grouping connected

pixels in the background subtracted binary image. The centroid of the object will be marked and bounding box properties will be extracted. Then to show the object being tracked, rectangle will be drawn around it. Similar procedure will be done by calling next frames.

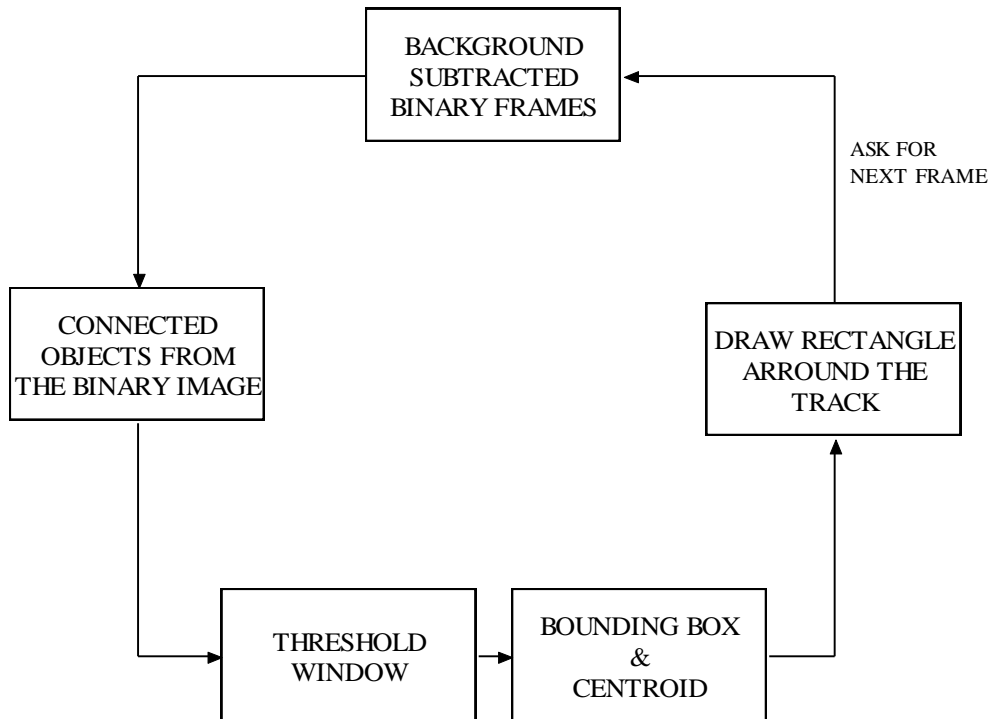


Fig. 1.2. Basic Tracking Procedure

#### 1.4.1.1. Using frame differencing

Frame differencing [18] is a simple arithmetic procedure in which the background frame will be subtracted from the current frame (Pixelwise) to get the movable coordinates. The first frame or previous frame of the current frame in the video will be used as the background frame and can be denoted as ' $B_g$ '. The current frame at time  $t$  can be denoted as  $I(t)$ . Mathematically it can be written as:

$$P[F_c(t)] = P[I(t)] - P[B_g]$$

If the background is assumed to be the frame at time  $t$  and will be subtracted from next frame then the subtracted image will give the information about the intensity change in between the consecutive two frames. This method is applicable when all the foreground pixels are moving and background pixels are static. Then the difference image or subtracted image will be thresholded by a threshold to improve the performance.



$$|P[F_c(t)] - P[F_c(t+1)]| > \textit{Threshold}$$

#### 1.4.1.2. Mean filter

Mean filtering [18] is a technique in which few preceding frames will be averaged to provide the background image.

$$B_g(i, j) = \frac{1}{N} \sum_{k=1}^N V(i, j, k-i)$$

Where  $N$  is the number of previous frames those have been taken for averaging. After constructing the background image  $B_g(i, j)$ , it can be then subtracted from the present frame  $V(x, y, t)$  at time  $t=k$  and thresholded. Thus the foreground is

$$|V(i, j, k) - B(i, j)| > T$$

Where  $T$  is the threshold.

#### 1.4.1.3. Adaptive Background Subtraction

If illumination varies rapidly and it becomes a noise to the motion data, then simple background subtraction fails. So the background needs to be updated during every current frame processing and should be adaptive to temporal change. This background subtraction technique is an adaptive procedure to choose the background which is a combination of both previous background and current frame.

$$B_g(t) = \alpha I_{\textit{current}}(t) + (1 - \alpha) B_g(t - 1)$$

Where  $\alpha$  is a learning rate. The binary motion detection mask  $D(x, y)$  can be calculated as follows:

$$D(t) = \begin{cases} 1, & \text{if } [f(t) - B_g(t)] > \textit{Threshold} \\ 0, & \text{Otherwise} \end{cases}$$

#### 1.4.1.4. Background Modelling and Subtraction by Codebook Construction

It is a method which quantizes background pixel values into codebooks [19]. So it becomes a compressed form for a long sequence data. It requires very less memory to save the background variations over long period

The reason why this methods is useful is that

- Generally single mode models cannot handle waving trees in the background.
- MOG cannot handle backgrounds which is having large variations.
- Illumination change can be handled upto large extent
- Does not require a separate training dataset.

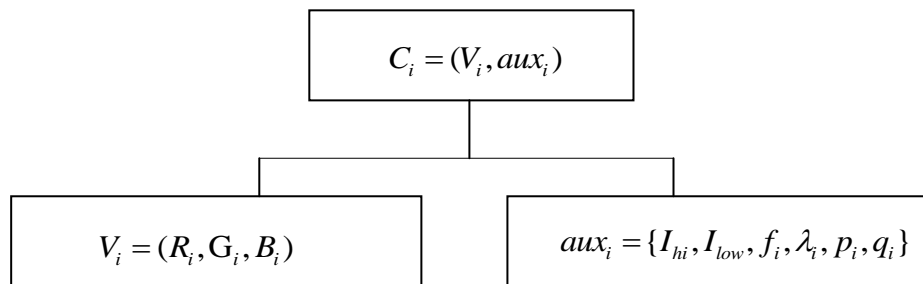


Fig. 1.3. Codeword for background modelling

Where,

- $I_{hi}, I_{low}$  : Highest and Lowest brightness levels for the codeword  
 $f_i$  : Frequency of codeword occurrence  
 $\lambda_i$  : Maximum negative run length  
 $p_i, q_i$  : First and last occurrence time respectively

#### 1.4.2. Optical flow

Motion between adjacent frames which has a time difference of dt can be calculated through Optical flow [20]. It is a Pixelwise operation which estimates the motion of moving pixels.

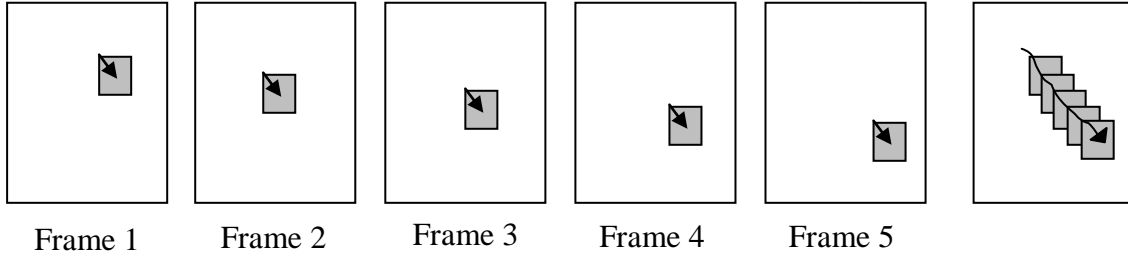


Fig 1.4. Simple optical flow based tracking using Harish corner

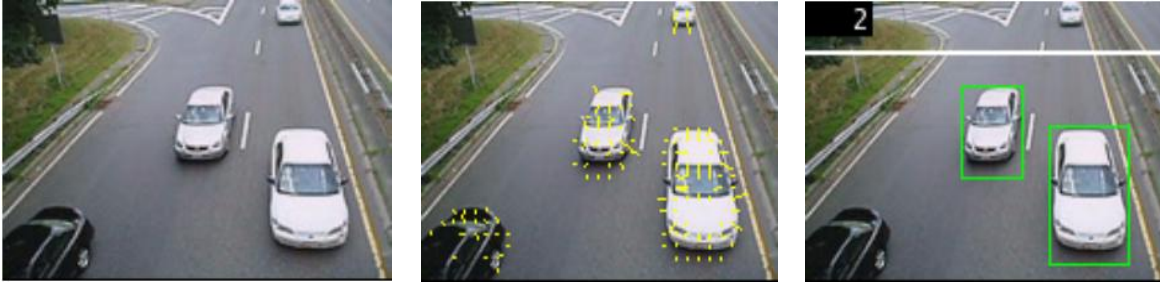


Fig 1.5. (a)Reference frame (b)Optical flow (c)Tracking Result [20]

The method uses local Taylor series approximations for the image signal. That means it uses partial derivatives with respect to the spatial and temporal coordinates. Let the point  $(x,y,t)$  has been moved to  $(x+dx,y+dy, t+dt)$  in the current frame. By taking Brightness constancy assumption it can be assumed that:

$$f(x, y, t) = f(x + dx, y + dy, t + dt) \quad (1.1)$$

Assuming the motion to be small, the Taylor series can be expanded as

$$f(x, y, t) = f(x, y, t) + \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial t} dt + \text{Higher Order Terms}$$

Neglecting higher order terms and taking  $f_x = \frac{\partial f}{\partial x}$ ,  $f_y = \frac{\partial f}{\partial y}$ ,  $f_t = \frac{\partial f}{\partial t}$ ,

$$f_x dx + f_y dy + f_t dt = 0 \quad (1.2)$$

Again let,  $u = \frac{dx}{dt}$ ,  $v = \frac{dy}{dt}$

Then,  $f_x u + f_y v + f_t = 0$

$$\boxed{v = -\frac{f_x}{f_y}u - \frac{f_t}{f_y}} \quad (1.3)$$

Here  $u$ ,  $v$  are the x and y components of velocity or optical flow of  $f(x,y,t)$ .

#### 1.4.2.1. Horn-Schunck method

The Horn-Schunck algorithm [20] assumes smoothness in the flow over the whole image. Thus, it tries to minimize distortions in flow and prefers solutions which show more smoothness. The flow is formulated as a global energy functional which is then sought to be minimized. This function is given for two-dimensional image streams as:

$$\int \left\{ \underbrace{(f_x u + f_y v + f_t)^2}_{\text{Brightness Constancy}} + \lambda \underbrace{(u_x^2 + u_y^2 + v_x^2 + v_y^2)}_{\text{Smoothness Consistency}} \right\} dx dy \quad (1.4)$$

Here,  $\lambda$  is the Regularization constant, larger value of which gives smoother flow of optical flow. Now, the x and y components of velocity ( $u$ ,  $v$ ) can be calculated as

$$u = u_{avg} - f_x \frac{P}{D} \quad v = v_{avg} - f_y \frac{P}{D}$$

Where,

$$P = f_x u_{avg} + f_y v_{avg} + f_t$$

$$D = \lambda + f_x^2 + f_y^2$$

$\lambda$ : Regularization constant

#### 1.4.2.2. Lucas & kanade method

This method [20] uses or combines information from neighbourhood pixels for the optical flow and thus, it is an over constrained system. The advantage of this method is that it is less sensitive to image noise. By assuming the displacement of the image contents between two nearby instants (frames) is small and approximately constant, it can be written as:

$$f_{x_1} u + f_{y_1} v = -f_{t_1}$$

$$f_{x_2} u + f_{y_2} v = -f_{t_2}$$

Similarly for  $n^{\text{th}}$  iterations it can be written as

$$f_{x_n} u + f_{y_n} v = -f_{t_n}$$

Now if these equations will be put into a matrix format it will be like:

$$\begin{bmatrix} f_{x_1} & f_{y_1} \\ f_{x_2} & f_{y_2} \\ \vdots & \vdots \\ f_{x_n} & f_{y_n} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -f_{t_1} \\ -f_{t_2} \\ \vdots \\ -f_{t_n} \end{bmatrix}$$

$$AU = B$$

$$A^T AU = A^T B$$

$$U = (A^T A)^{-1} A^T B$$

$A^T$  is the transpose of  $A$  and then it computes the final lucas kanade expression as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum f_{x_i}^2 & \sum f_{x_i} f_{y_i} \\ \sum f_{x_i} f_{y_i} & \sum f_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum f_{x_i} f_{t_i} \\ -\sum f_{y_i} f_{t_i} \end{bmatrix}$$

### 1.4.3. Kalman Filter Based Tracking

Kalman tracking is basically derived from Kalman filter theory. It is a probabilistic type of tracking which estimates the track positions in current frame by taking the information from some of the previous frames. At first the initial states for Kalman filter will be initiated by the help of the some of the previous frames. Then vectors will be constituted using position and motion of the object. The Kalman filter is a mathematical tool that can estimate the variables of a wide range of processes. It estimates the states of linear systems by using simple kinematic equations for updation. Velocity vector can be calculated as the change in location with respect to time difference between two frames. Location of the object can be chosen as the center of gravity of the object or in 2D the centroid.

The estimation of the states of linear systems can be done by following kinematic equation [21].  $K-1$  is the previous frame and  $k$  is the current frame.  $S_k$  is the position,  $v_k$  is the velocity and  $a_k$  is the acceleration of the current frame.

$$S_k = S_{k-1} + v_{k-1}t + \frac{1}{2}at^2, \quad v_k = v_{k-1} + a_{k-1}t, \quad a_k = a_{k-1}$$

$$z_k = H_k x_k + v_k \quad (1.5)$$

where,

- $z_k$  : Measurement Vector
- $v_k$  : Discrete white zero mean process noise with known covariance matrix  $R_k$
- $H_k$  : Relationship between the measurement vector
- $x_k$  : State Vector

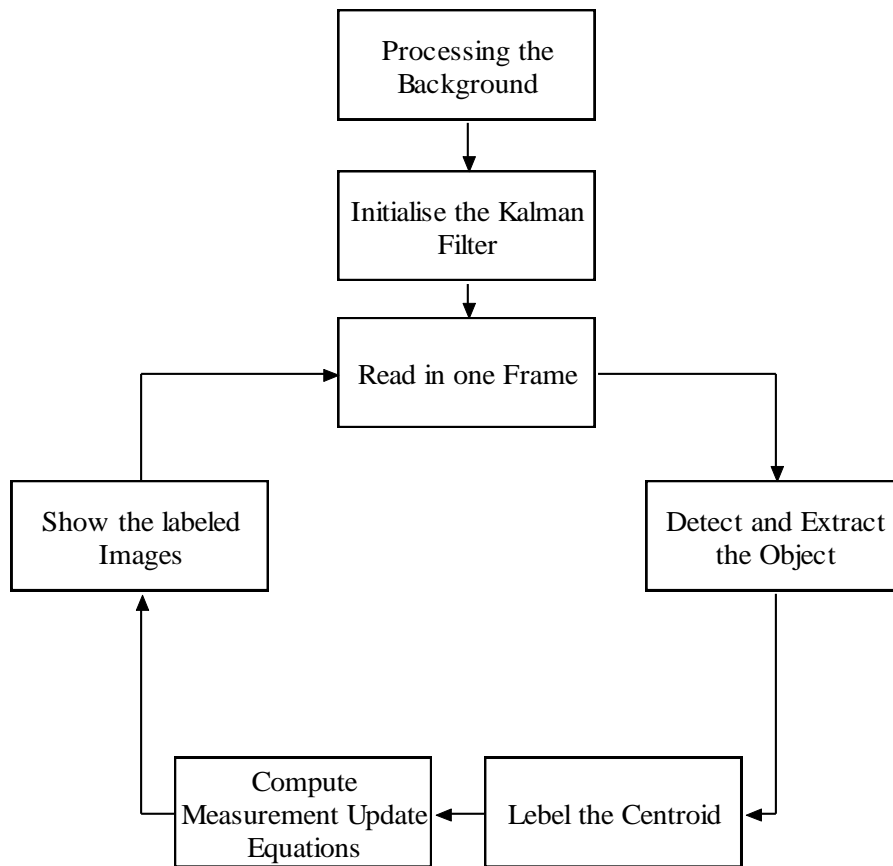


Fig 1.6. Kalman Filter based Tracking

#### 1.4.4. Mean Shift Tracking

In mean shift tracking method, the object which has to be tracked, will be selected first. A rectangular patch containing the object will be selected manually. The selected rectangular area is known as object model. Now to start with mean shift method, the colour histogram of the object model should be done. Colour histogram  $q_u$  [22] can be obtained by

(1.6)

$$q_u = C \sum_{i=1}^n k(\|x_i^*\|^2) \delta[b(x_i^*) - u] \quad (1.6)$$

where,

- $x_i^*, i=1, \dots, n$  : The pixel locations inside the selected region
- $b(x_i^*)$  : Bin number  $(1, \dots, m)$  associated with the colour at the pixel of normalized location  $x$
- $\delta$  : Kronecker delta function
- $C$  : Constant normalization

After getting the object model, the same procedure will be continued in the current frame. In current frame, again colour histogram will be calculated at the same position. But the object position will not be same as in previous frame due to motion. The new colour histogram is known as candidate model  $p_u(y)$  [22] and can be found by (1.7).

$$p_u(y) = C_h \sum_{i=1}^{n_h} k \left[ \left\| \frac{y - x_i}{h} \right\|^2 \right] \delta[b(x_i) - u] \quad (1.7)$$

where,

- $x_i^*, i=1, \dots, n_h$  : The pixel locations of the object's condition region, centred at  $y$  in same frame
- $k$  : Kernel function with radius  $h$

Now relation between these two colour probabilities can be calculated through similarity function  $\rho_1(y)$  [23] or Bhattacharyya coefficient [24][25]  $\rho_2(y)$  by following (1.8) and (1.9).

$$\rho_1(y) \equiv \rho[p(y), q] = \sum_{u=1}^m \sqrt{q_u / p_u(y)} \quad (1.8)$$

$$\rho_2(y) \equiv \rho[p(y), q] = \sum_{u=1}^m \sqrt{p_u(y) q_u} \quad (1.9)$$

Now, the patch for the candidate model will be shifted in the direction of mean shift in an iterative way to maximize the similarity parameter. When similarity parameter will be maximized, that patch will be the new position for the object in the current frame.

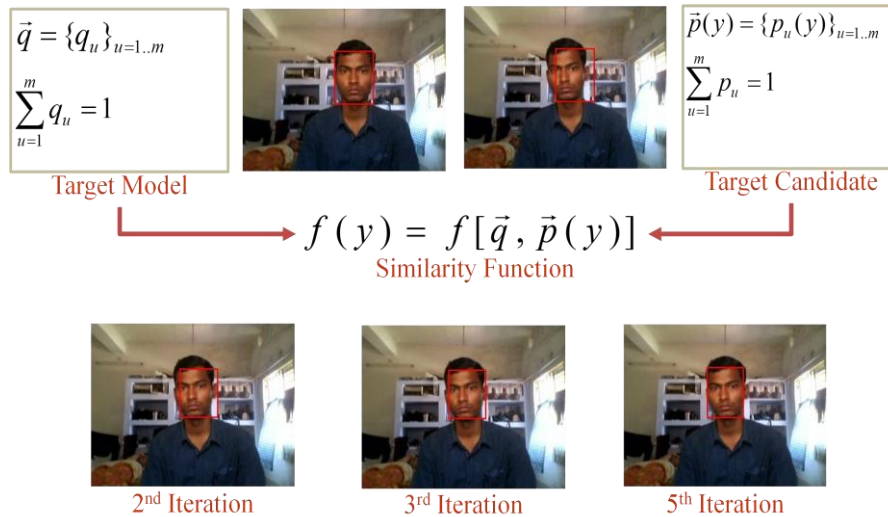


Fig 1.7. Traditional Mean Shift Tracking

## 1.5. Thesis outline

The remainder of this paper is organized as follows:

**Chapter 2** is mainly focussed on tracking and that to on mean shift tracking. This section explains about the advantages and disadvantages of mean shift tracking. What are the extensions that can be added to the mean shift procedure has been proposed and evaluated in this section.

**Chapter 3** is about activity recognition. It is divided into two portions: trajectory based low level activity recognition and shape based high level activity recognition.

**Chapter 4** concludes the thesis and discusses about the future works or extensions to the present work.



**CHAPTER 2**

**AUTOMATED HUMAN**

**TRACKING USING**

**ADVANCED MEAN SHIFT**

**ALGORITHM**

## **2.1. Introduction**

For tracking human or any kind of object, colour feature based [26][27][28][29][30] mean shift technique is widely used as it is simple and robust to target deformation and scale variance (shape and size). This technique uses Bhattacharya coefficient to locate the object based on the maximisation of the similarity function between object model and candidate model. To increase the efficiency of traditional mean shift algorithm, various methods are reported from time to time. When prolonged period of occlusion is there and in those period motional characteristics of the object (velocity, acceleration) changes abruptly, success rate of mean shift method decreases up to large extent. Similarly, automatic tracking of human is also a challenge in this field.

## **2.2. Problem Description**

Traditional mean shift tracker is unable to start the tracking automatically and could not handle the fast motion of the object. Similarly, while handling the occlusion, it shows poor result [16] in comparison to probabilistic method.

### **2.2.1. Manual or Semi-Automatic Technique**

The traditional mean shift algorithm is a manual procedure in which a moving patch, which has to be tracked in successive frames, is selected manually. In case of surveillance the tracking should be automatic and target oriented. The moving pixels in a frame are detected by simple background subtraction techniques but, we can't say whether that moving object is a human or not. Therefore, traditional method is not a fully automated human tracking technique.

### **2.2.2. Fast Motion of Object**

To work with mean shift algorithm it should be kept in mind that the motion of the object should not be very large. Because if the patch containing object will not have any overlapping area in the current frame with respect to previous frame, mean shift algorithm fails.

### **2.2.3. Prolonged Occlusion**

In addition to the above problem if the object is having occlusion or there is sudden change of illumination, the algorithm fails as the similarity function becomes considerably low. Probabilistic approach can solve the occlusion to some extent [31][32]. But, if the object is going for a prolonged period of occlusion and in that period the information about the object can't be estimated, tracking fails.

### **2.3. Proposed Framework**

Flow chart of the proposed automated human tracking using advanced mean shift algorithm is shown in Fig.2.1. The proposed method helps the traditional method to start the tracking automatically by providing a patch containing human. Background of the frame will be ignored to get the moving objects by background subtraction. Human detection procedure detects the movable human and selects the patch for mean shift tracking. During the tracking procedure, if similarity parameter goes below 0.05, block matching technique will be initiated which helps to continue the tracking if it has been stopped due to fast motion. If block matching fails to make the tracking continue, it will be concluded that occlusion is there and so, occlusion handling will be continued. In this technique, background subtraction and human detection will be done in an iterative way till the human is not redetected after occlusion. After detection of human, the patch will be compared with the previous patch containing human before occlusion through similarity parameter. If similarity parameter will be found more than 0.1, tracking restarts from that patch by mean shift algorithm, otherwise process ends.

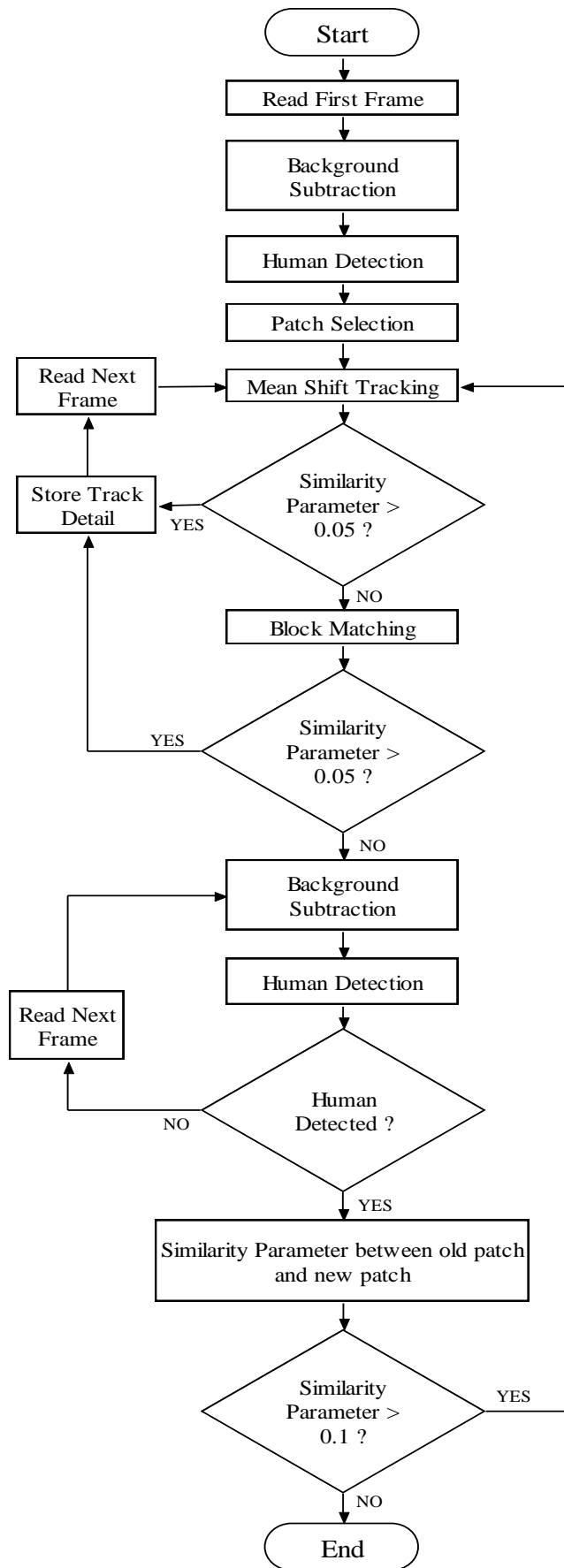


Fig.2.1. Flow chart for Advanced automated mean shift tracking

Details about the proposed techniques are explained as follows:

### 2.3.1. Simple Background Subtraction

Simple background subtraction [33] is used for the proposed method. In this technique the background frame remains fixed. Now, the current gray frames will be subtracted from fixed background frame to get the movable pixels. This can be done by taking a threshold and comparing the subtracted value with it. In our proposed work, the first frame has been taken as fixed frame. The following function explains the background subtraction.

$$BS(i, j) = \begin{cases} 1, & \text{if } [f(i,j)-B(i,j)] > T \\ 0, & \text{else} \end{cases}$$

where,

- $BS$  : Background subtracted image
- $f$  : Current Frame
- $B$  : Fixed Background Frame
- $T$  : Threshold

A threshold value is set to subtract background from the moving object based on the binary hypothesis. To discard noise from the background subtracted binary frame, morphological operations like erosion, dilation, and hole filling are used.

#### ***Erosion***

It is a morphological operation [34] in which the area of the object get eroded at the boundary points according to a structural element. The following equation (2.1) explains the operation through set theory.

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (2.1)$$

where,

- $A$  : Binary image
- $B$  : Structural element
- $z$  : Center of structural elements

#### ***Dilation***

It is a morphological operation [34] in which the area of the object get expanded at the boundary points according to a structural element. The following equation (2.2) explains the

operation through set theory.

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \phi\} \quad (2.2)$$

where,

- $A$  : Binary image
- $B$  : Structural element
- $z$  : Centre of structural elements

### **Hole Filling**

Sometimes the object (group of white pixels in a binary image) may contain holes (group of black pixels). So, it is needed to identify those regions and fill that with white pixels [34]. Otherwise, during erosion that will expand which leads to information loss and bias of the object.

### **2.3.2. Human Modeling**

Human modeling is required for human detection and it is necessary to choose the patch automatically to start the tracking procedure. It can be done by star skeletonisation method as in [35] to detect the human and separate it from all other moving objects. The process can be done by calculating local maxima of the Euclidean distance function as shown in Fig.2.2. Euclidean distance is the distance of centroid from the boundary points in clockwise manner. The median filter is used to preprocess the distance function before calculating maxima to avoid noisy peaks.

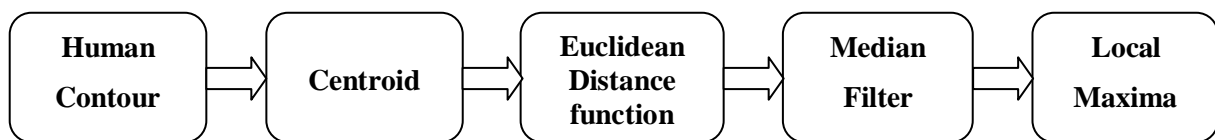


Fig. 2.2. Local Maxima Calculation Procedure

The procedure starts with detecting the boundary points of the object as in Fig.2.3. By thresholding, the binary image of the frame will be obtained. Dilation procedure will expand the object boundary area and similarly erosion will erode the boundary area. When the eroded image will be subtracted from the dilated one, boundary points will be obtained which is known as contour detection. Here also in pre processing stage, hole filling should be done prior to contour detection so as to remove unnecessary inner boundaries inside the object.

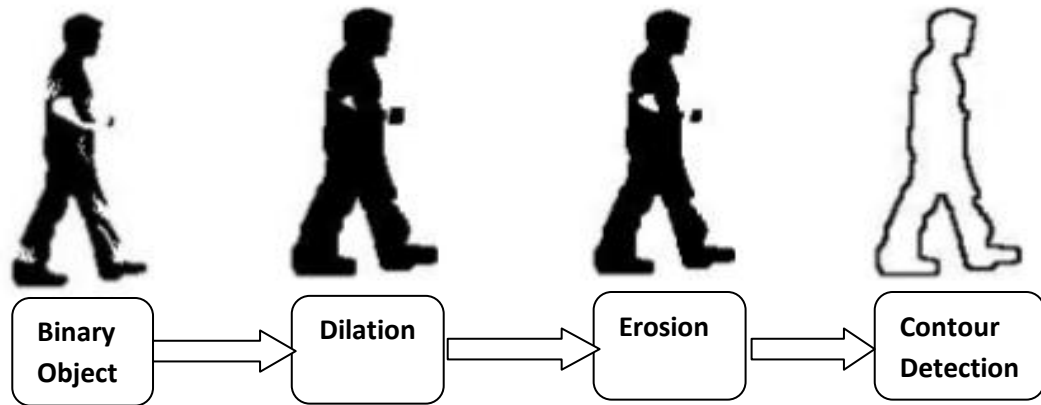


Fig. 2.3. Contour Detection Procedure

Then, a distance function will be calculated as shown in Fig.2.4. The Euclidian distance from the centroid will be calculated for every boundary point in a clock wise direction. After getting the distance function, the curve will be smoothed by median filtering and then local maxima will be calculated.

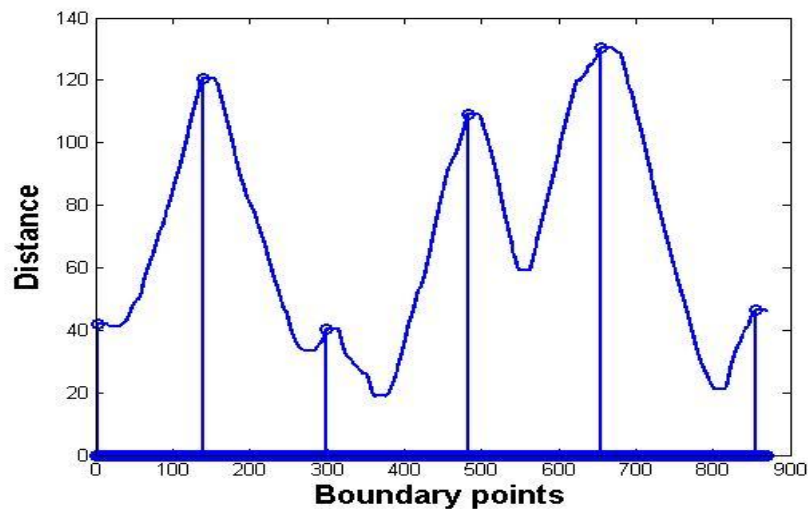


Fig. 2.4. Distance Function

The points, where the maxima will be found, will be joined with the centroid to get a star like structure. The concept of this method is to get distant points which are tips of legs, hands, and head from the centroid. So this helps to distinguish human from other objects. By analyzing the star skeleton, it can be concluded whether that moving body is human or not. Example of star skeleton is shown in Fig.2.5.

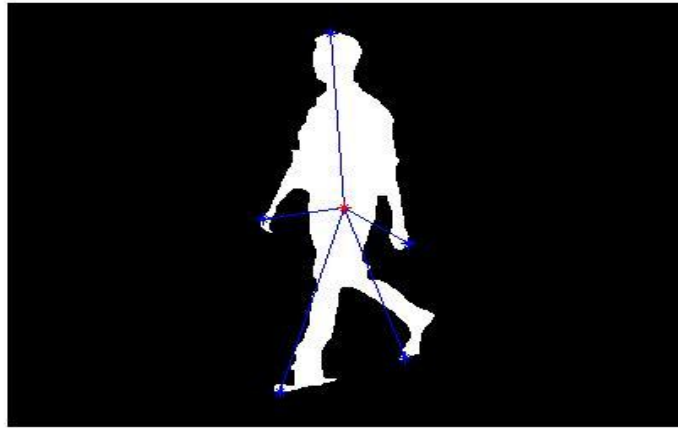


Fig. 2.5. Star Structure with Binary Human Object

Skeleton can be analyzed like distance of head and legs from centroid are relatively same, angle of legs with the line joining centroid to head is greater than 150 degree in maximum cases.

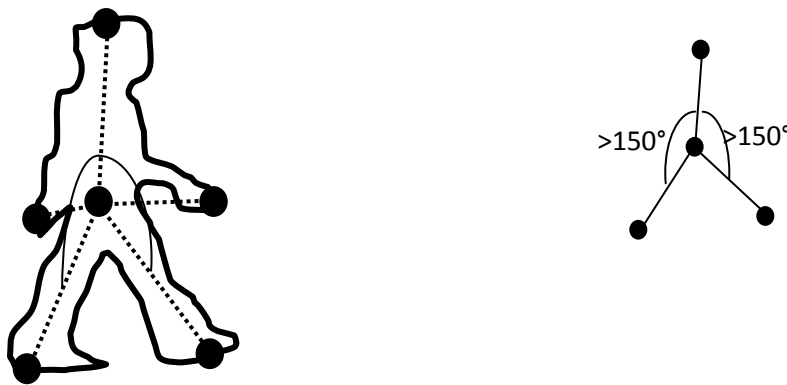


Fig. 2.6. Star Skeleton Analysis

### 2.3.3. Block Matching Technique

Sometimes the object may have fast motion for which mean shift tracker may not get an overlapping area to continue. At this moment similarity function goes below certain threshold and so tracking will be stopped. This is the time when block matching procedure will be initialized. Since block matching is a simple and relatively fast process, it affects the execution time negligibly and it can be called whenever tracking fails to ensure whether it is due to fast motion or not. In block matching technique, motion direction will be calculated



first which can be done from the centroid of previous two frames. After getting the direction, a shifted center will be chosen as in Fig.2.7 and by referencing that point, three blocks will be chosen in the direction of motion. The shifted center position is  $w/2$  pixels away from the original centroid position, where  $w$  is the width of the patch containing the object.

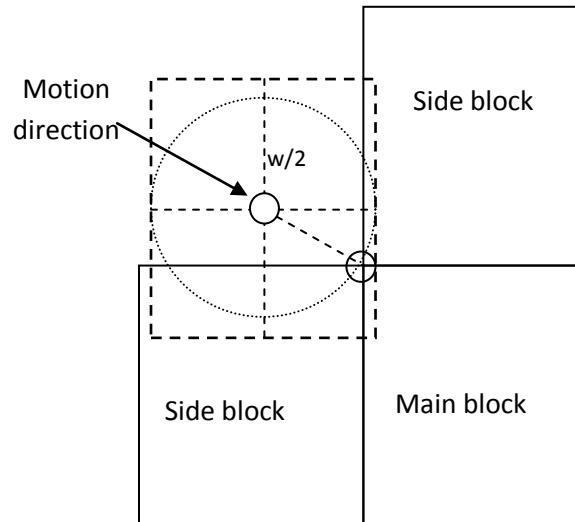


Fig. 2.7. Forward blocks for block matching in the motion direction

Now similarity functions will be calculated between object block and three new blocks. There are three blocks: one main block and two side blocks. Generally main block is the block, which may contain the shifted object in most cases. If the object has changed direction of motion suddenly, then one of the side blocks will contain the object. The block which will have maximum similarity with the object model will be upgraded as candidate model to continue the mean shift algorithm.

### 2.3.4. Occlusion Handling

**Full occlusion handling:** If there is full occlusion for a long time, it is difficult to get the track after occlusion by traditional deterministic and Probabilistic methods. It may happen that the object will halt its motion during occlusion for which probabilistic trackers fail to estimate the position of the object. Sometimes it may happen that the object before to the occlusion and after the occlusion may be different i.e. a new object of similar type will come out. The proposed method so applied which tracks the object after the prolonged occlusion irrespective of motional characteristics variation during prolonged occlusion.

In this approach when the object is being occluded, tracking will stop till the object is not

coming out of that occlusion. Tracking is not needed when the object is going for occlusion because during that period we can't get information about the object. The important thing is that after occlusion the tracking for that object should start automatically. It is done by putting the object into detection procedure to conclude whether that is a human object or not. After detection the object model will be compared with the object model before occlusion by similarity parameter to know whether the same object is coming out of the occlusion or not and if same object is detected, then tracking continues through mean shift procedure.

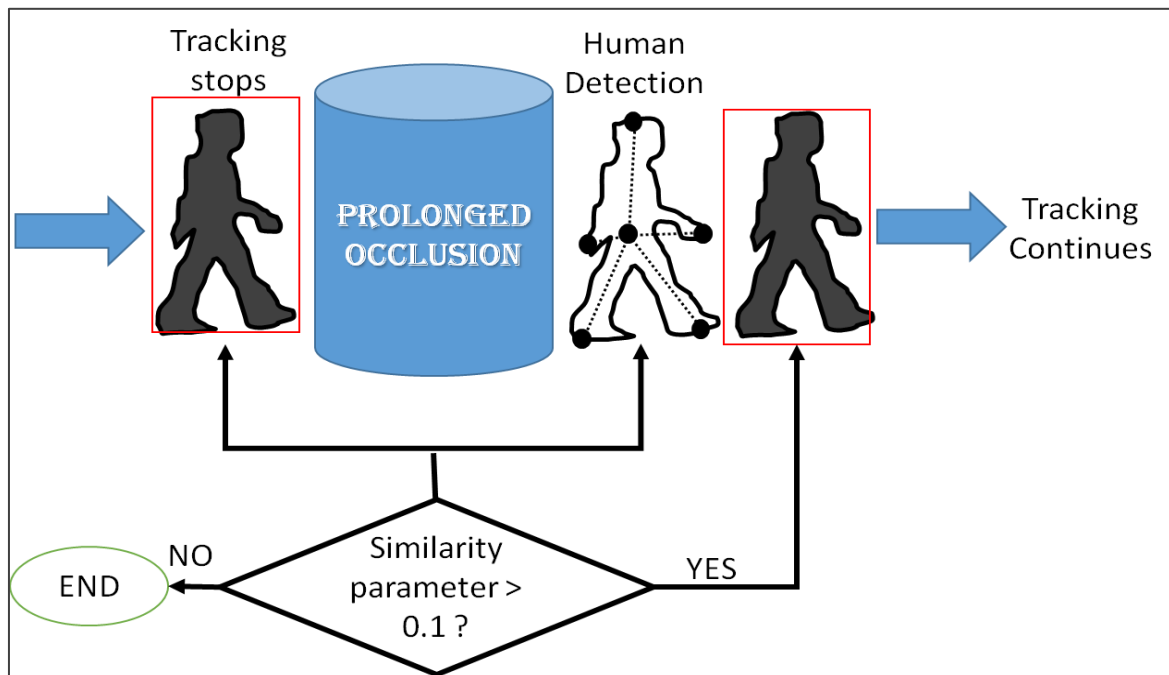


Fig. 2.8. Prolonged Occlusion Handling

**Partial occlusion handling:** If only partial occlusions are there, then another method is more useful. Similar method is used in [13] and which works on the concept of fragmentation based tracking. The advantage of this method is that it can handle sudden illumination change and partial occlusion up to large extent. During this process the patch containing the object will be subdivide into 9 divisions as shown in Fig. 2.9(a). The performance of the process depends on number of divisions. But, if the sub divisions will be more than certain value, the method fails because the object losses its identity and the blocks can be easily mixed with backgrounds to create ambiguity.

As shown in Fig. 2.9(a), the corner sub-blocks will be tracked separately. By taking the similarity parameter comparison, the best block to be tracked will be chosen. According to that block, total patch in the current frame will be redrawn. If due to partial occlusion some

portions of the patch will be hidden, other portions will be tracked correctly. The subjective comparison has been given in Fig. 2.9(b) and 2.9(c). The fragmentation method gives more accurate result in comparison to traditional method.

1	2	3
4	5	6
7	8	9

(a)



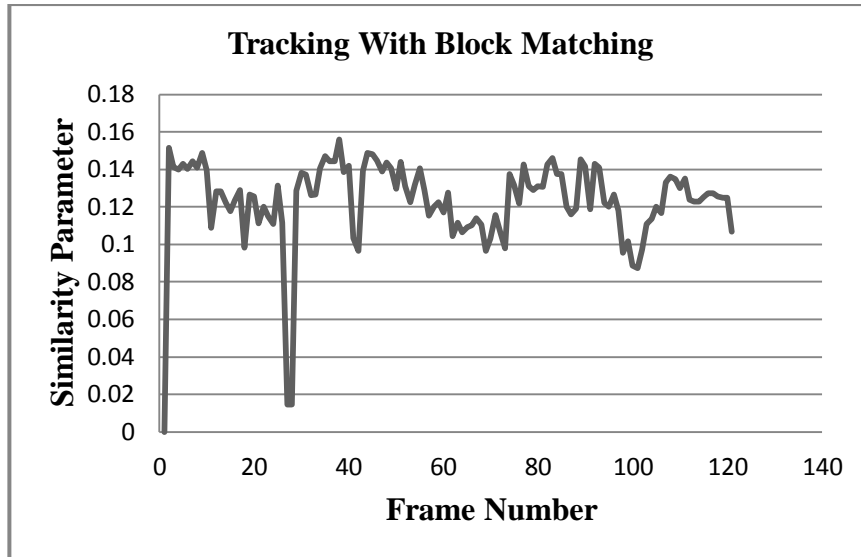
Fig. 2.9. (a) Patch fragmentation (b) Result of traditional tracking (c) Result of patch fragmentation based tracking.

## 2.4. Results and Discussion

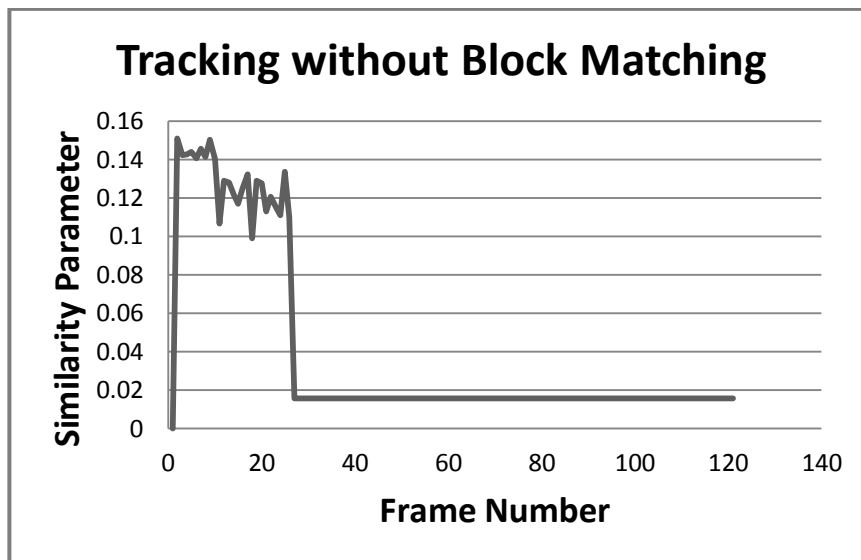
To conduct the experiment, real time videos have been collected by a panasonic lumix dmc lz-30 camera with the dimensions  $1280 \times 720$ , 30fps. Two types of various videos have been taken: one, where objects fails to have overlapping area in next frame and another which is having prolonged occlusion. Test1, Test2, Test3, Test4 are four test videos having different length of prolonged occlusion.

### 2.4.1. Block Matching

Tracking fails for fast moving objects where there is no overlapping patch for mean shift tracking. In Fig. 2.9(a), it is clearly shown that the similarity parameter goes below 0.02 and remains there for further frames which concludes the failure of traditional technique. Block matching technique had supplied an overlapping area to continue the track whose success has been reflected in Fig. 2.9(b) where similarity parameter has become considerably high.



(a)

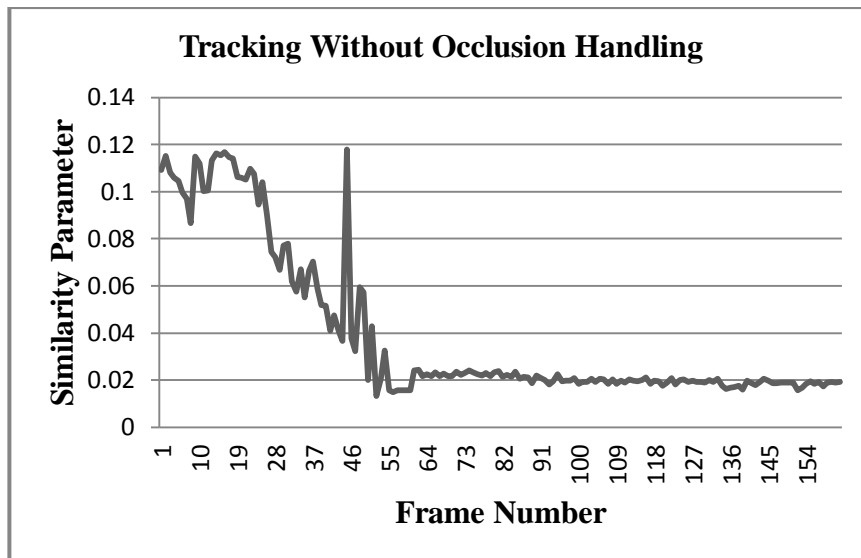


(b)

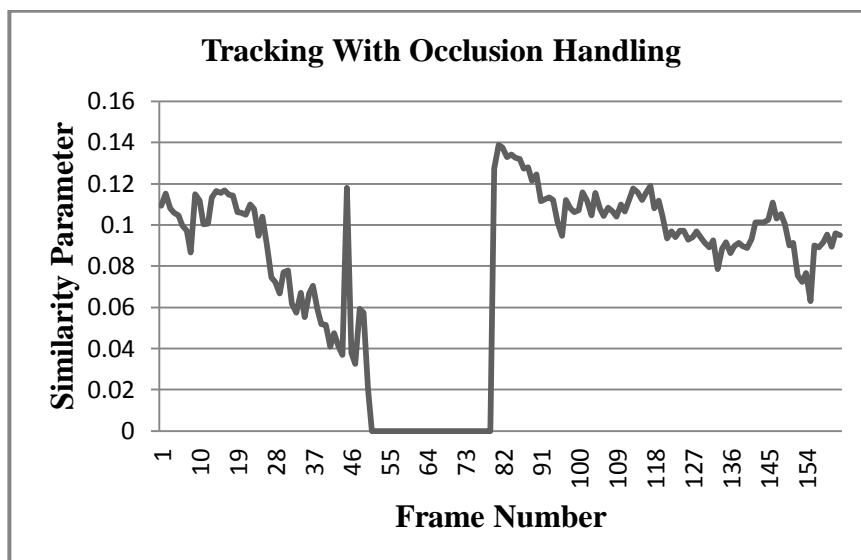
Fig. 2.10. Similarity parameter vs. frame number curves for (a) normal mean shift tracking (b) mean shift with block matching

### 2.4.2. Occlusion Handling

The Fig. 2.10 is for test video Test1 in which occlusion occurs at 158th frame and so similarity parameter has gone below 0.02 and remains at that level for further frames as shown in Fig. 2.10(a). Proposed algorithm stops the tracking when similarity parameter goes below 0.05 as shown in Fig. 2.10(b) where tracking has been stopped for occlusion from 160-189<sup>th</sup> frame. After that, again tracking starts by background subtraction and human detection and it continues till the end of the database as in Fig. 2.10(b).



(a)



(b)

Fig. 2.11. Similarity parameter vs. frame number curves for (a) tracking without Occlusion Handling (b) tracking with Occlusion Handling. (curves have been shown excluding first 110 frames)

The real time frames of test video Test1 has been shown in Fig. 2.11, in which the red colour rectangle shows the success of tracking the human in occlusion condition. The human detection has been indicated in Fig. 2.11(a) and the starting of tracking, Fig. 2.11(b) is the last frame before tracking fails due to prolonged occlusion, Fig. 2.11(c) shows the continuation of tracking by proposed method after occlusion, and the Fig. 2.11(d) is an arbitrary frame which shows still the tracking is continuing after occlusion.

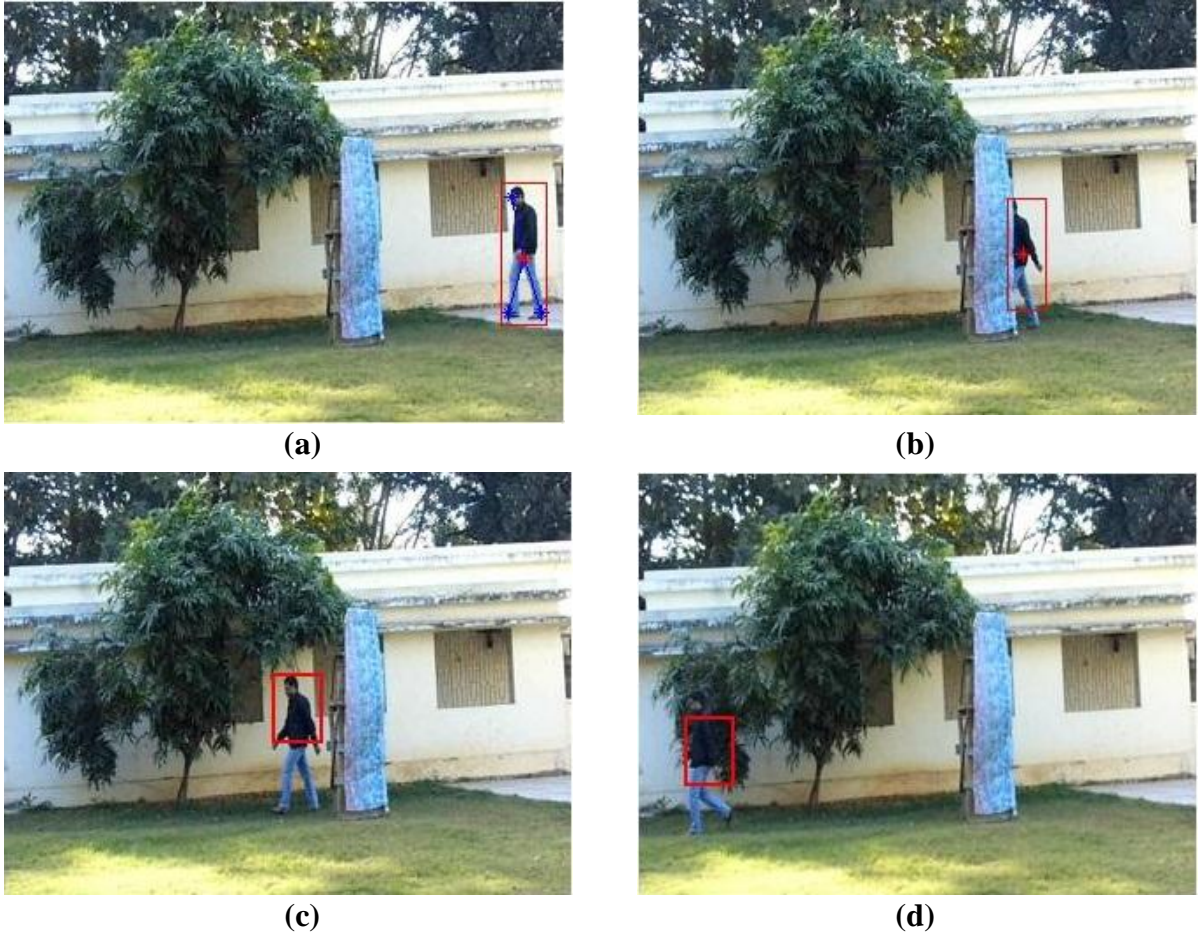


Fig. 2.12. Human tracked frames of the video. (a) Human detected and start of tracking, (b) Tracked frame before occlusion, (c) Tracked frame after occlusion, (d) Tracked frame toward end of frame.

Similarly, the Table 1 shows the success of tracking for four different videos having occlusion problem in which traditional method fails to track. From the table it is clearly understood that tracking stops after 2-5 frames of starting of occlusion and again restarts after 5-10 frames after occlusion. It is due to the degradation of similarity parameter value till threshold and proper human detection after occlusion respectively.

*Table 1: Performance of Proposed Method*

Test videos	Test1	Test2	Test3	Test4
Number of frames	300	495	300	345
Occluded frames	158-184	134-180	123-175	163-210
Failed at frame	160	137	127	167
Track continued by proposed method at frame	189	185	185	220
Tracked till frame	272	465	295	290

## **2.5. Summary**

The chapter is mainly focused on tracking by mean shift algorithm. Some addition to traditional mean shift tracker has been made it more efficient for real time applications. Partial occlusion has been handled by patch fragmentation and prolonged occlusion by proposed method. The automatic starting of tracking has been done introducing human detection method into the procedure. If fast motion is a problem for track failure then proposed block matching algorithm is a solution to large extent.

# **CHAPTER 3**

## **ACTIVITY RECOGNITION**



### 3.1. Introduction

Activity recognition is the study of human activities which will help in surveillance. It can be divided into two levels of recognition: low level & high level. Hierarchy of recognition helps in reducing the complexity and time consumption of the real time system. Low level is mainly based on trajectory of the object. Trajectory is the list of tracks of a single object from beginning to end. Trajectory gives the total positional and motional characteristics of an object. High level recognition is mainly based on shape of the object. Human posture recognition helps the system to detect the abnormality or seriousness of the happening at high level.

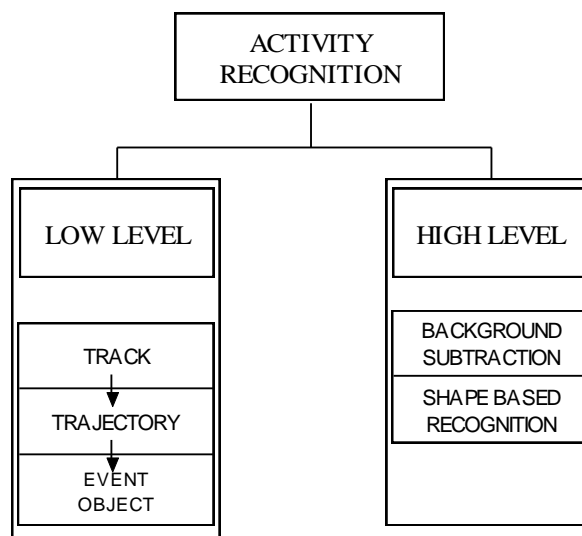


Fig. 3.1. Basics of Activity Recognition

Event object is the collection of information for a trajectory to decide whether it is normal, abnormal, or undecided event. Fig. 3.2 shows which information an event object holds.

**Event type:** It hold the information about the type of happening; whether it is abnormal or unknown.

**Start & End:** It gives the information about the beginning and end of the event.

**Location:** It gives the positional information about the event.

**Trajectory List:** List of associated trajectories from starting to ending of the event

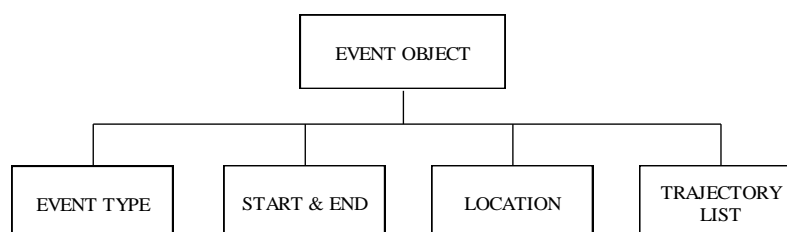


Fig. 3.2. Event Object

### 3.2. Low level Activities

Low level activities which are mainly depends on trajectory, is a spatial and temporal study of tracks from frame to frame. Temporal analysis helps in providing the occurring time for abnormality. If it is not above the threshold then the abnormality will be decided as non-harmful and so no alarm. Spatial analysis helps in study of positions of tracks or density of tracks at a certain region. Density helps to decide for abnormal loitering at certain places, position of tracks helps in deciding whether the object is entering to restricted area or not, and also it will help to detect human collision and fell down abnormality.

#### 3.2.1. Person fell down

From trajectory information it is difficult to say the surety about person collision and fell down. Simply it can be checked that whether two tracks are meeting or not, if met whether new tracks are starting from meeting place or not. Sometimes the disappearing of tracks may be due to occlusion and it may be treated as abnormal condition. To avoid occlusion based false alarm and non-harmful fall downs, a threshold known as  $\eta_{elapse}$  has been introduced. If after meeting place tracks are not generating after  $\eta_{elapse}$  number of frames, then the situation is asking urgency for observation or simply an abnormal condition.

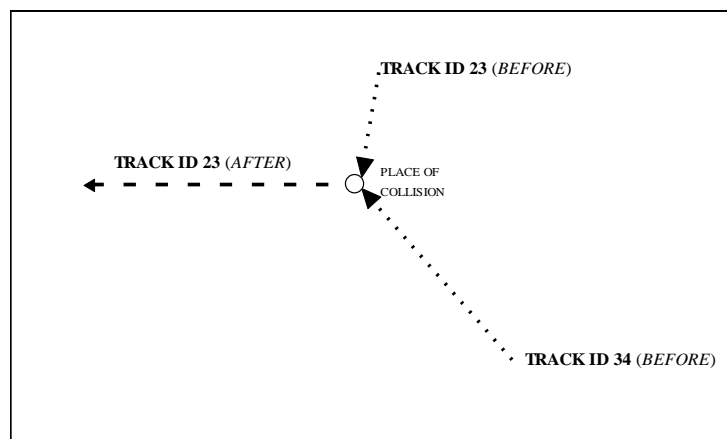


Fig. 3.3. Person Fell down

As shown in Fig. 3.3. two tracks are meeting and one track is coming out the collision place. That indicates another track is missing and it may be due to person fell down [16]. But, if track is missing due to partial or full occlusion, then also it will lead to ‘person fell down’ criteria which again leads to a false alarm. Similarly if the person after falling down, get standing up and walk, then it is not an abnormal situation. So to handle these two problems, another temporal threshold named as  $\eta_{elapse}$  has been introduced. It will check how much time the new track is taking to restart. If it is more than threshold then it will be declared as

abnormal event.

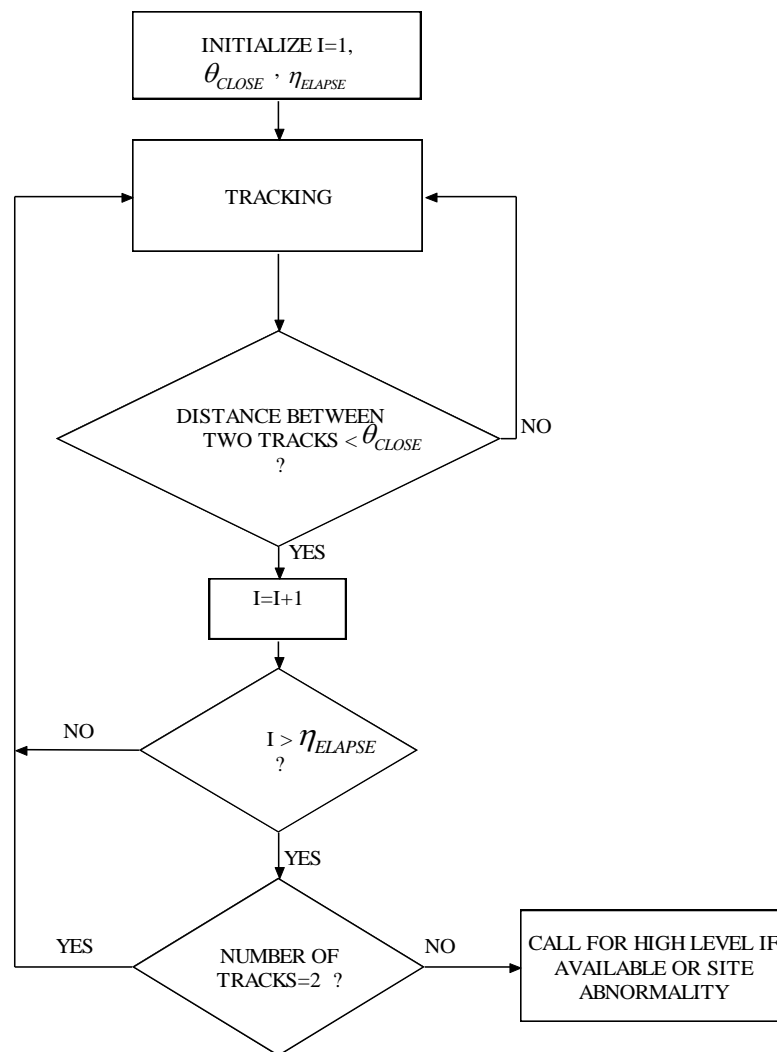


Fig. 3.4. Flowchart for Human Fell down detection

### 3.2.2. Loitering for large time and entering into Restricted area

(a) Loitering [16][17] for a long time in a crowded area may lead to a careful observation as it may be an abnormal case. Maybe someone is trying to do some mischievous work and that is why he is loitering there to get the proper time to do so. So in low level recognition it should be recognized as an abnormal case and should be put into high level recognition to ensure the abnormality.

(b) First the user will select the restricted area [16][17] in the frame by selecting the some region. Maybe it is inside a mall or office where entering to that zone creates large problem in security purpose. So position of tracks of a human should be regularly checked with the restricted position and if the positions are same then it will be an abnormal case it may be happen that the entry is not that much dangerous i.e. it may be due to some flying object or

child. So low level abnormality will be put into high level recognition to ensure the case and so to reduce false alarm rate.

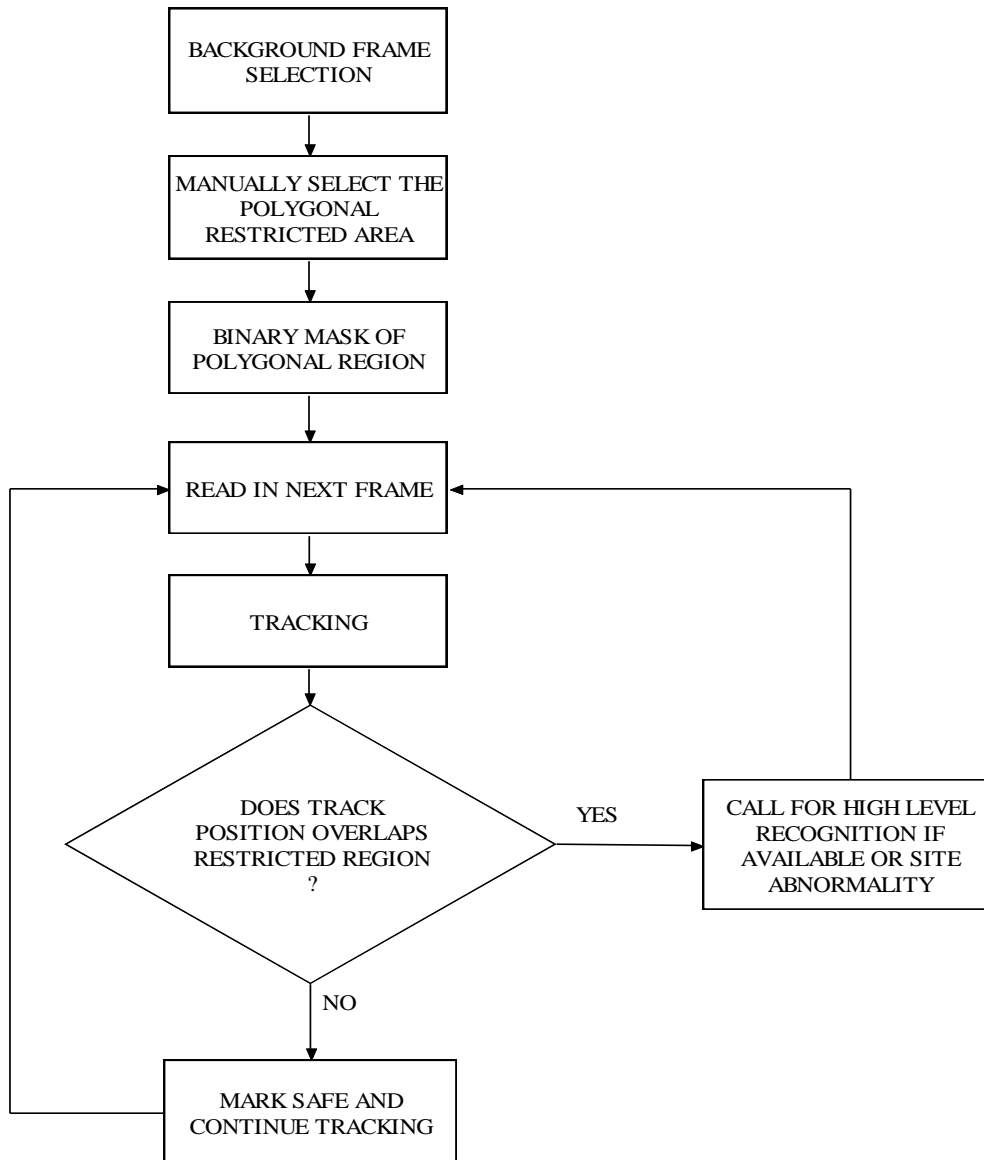


Fig. 3.5. Flowchart for detection of illegal entry

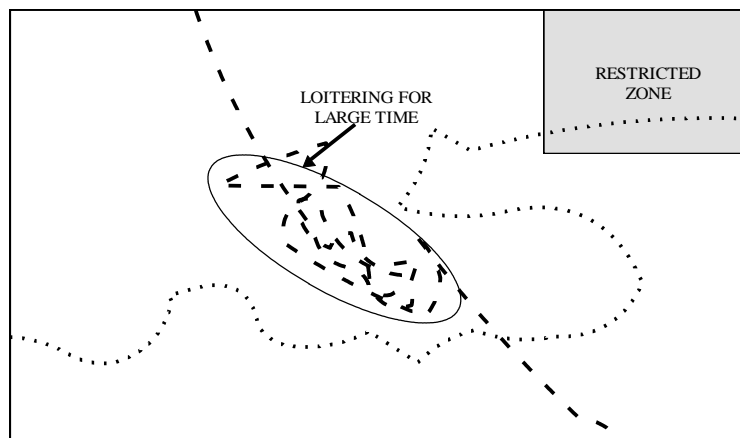


Fig. 3.6. Loitering and Entering into restricted zone

### 3.2.3. Line formation

It is an abnormal event created which creates problem in trafficking at crowded areas. To recognize whether line is getting formed or not, is not an abnormal event, but whether that line is there for long time or not is an abnormal event. Accidentally some lines may be formed which causes false alarm. So it forces us to consider temporal relations along with spatial relations.

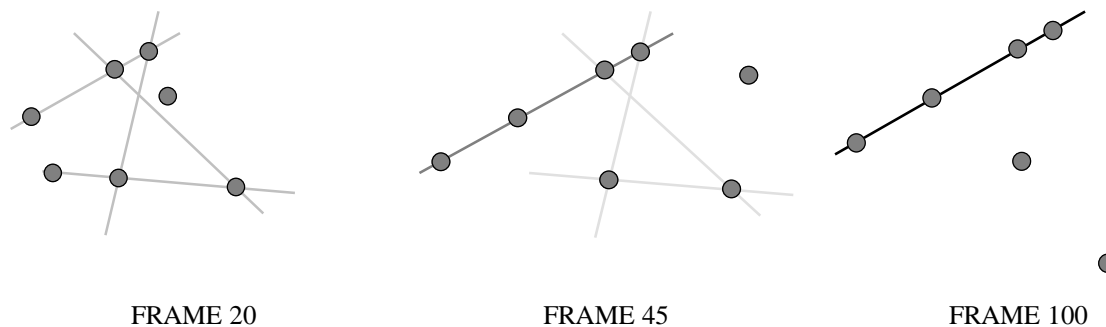


Fig. 3.7. Line Formation

The line event [16] will start with event initiation procedure. Three tracks having distance less than a pre set threshold will be initiated as a line event. The threshold is known as  $\theta_{close}$ . Line is not always a straight line. So another threshold known as distance error  $\theta_{dist}$  will be added which will add tracks onto initial line event. These two thresholds are based on spatial relations as shown in Fig. 3.5(a), (b). To add more evidence to the event object one temporal relation has been taken into consideration.  $\theta_{vote}$  is a temporal threshold which is used to give more evidence to the event and to reduce false alarms. Time to time tracks which is meeting the spatial relations with the event, will be added as votes towards that event. When the number of votes exceeds the threshold, it will be decided as an abnormal event.

### 3.3. High level Activities

High level recognition [16] is mainly based on shape based human pose detection. When abnormality found at low level which is trajectory based, it should be again put into high level recognition module for further investigation to ensure the abnormality. This helps in reducing false alarm rate. The first step in this process is to collect training database. As shown in Fig. 3.8(a), frames containing human being will put into background subtraction to extract and crop them. Now those cropped patches with different poses will be stored as training data set as shown in Fig. 3.8(b). After training data set collection, testing data will be collected and compared by various existing methods such as SVM, HMM, ANN etc. The

example of various human pose detection such as standing walking, lying down have been shown in Fig. 3.8(c).

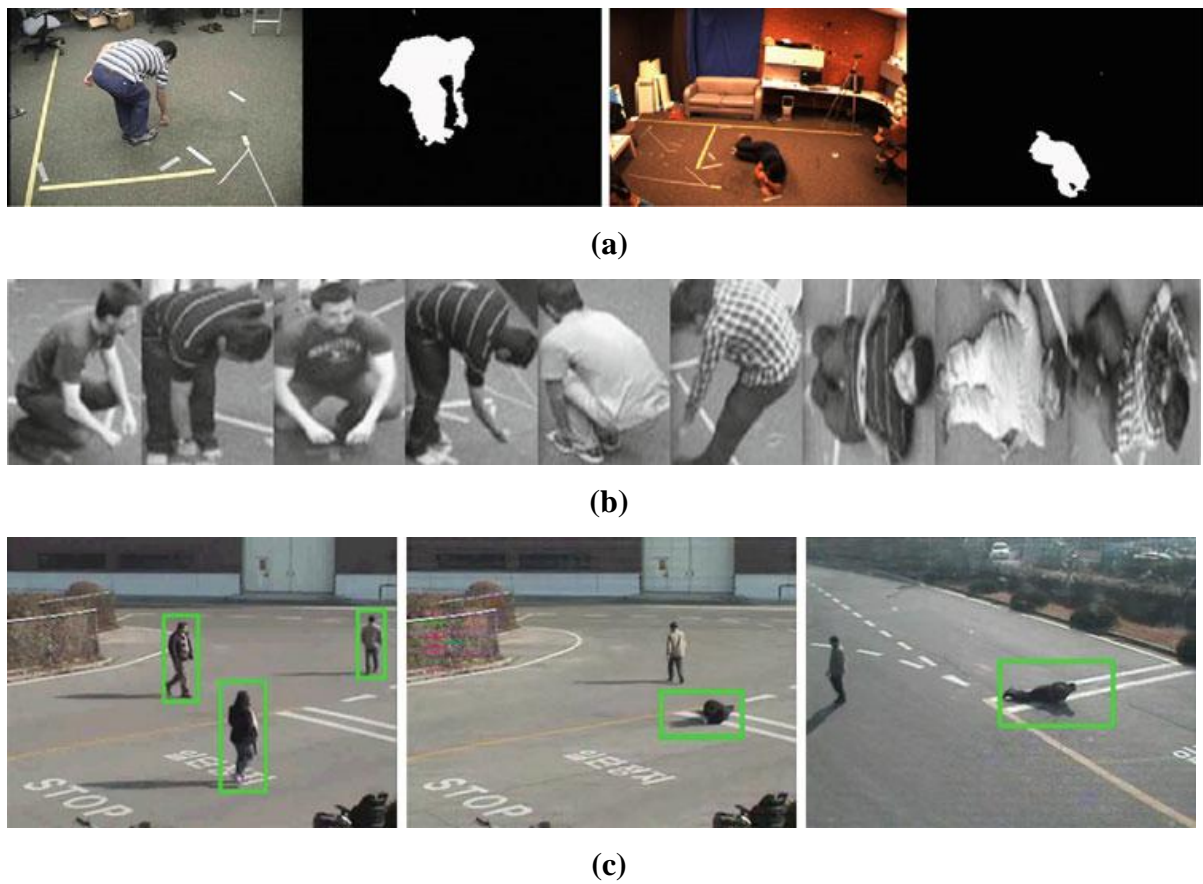


Fig. 3.8. (a) Original images and background subtraction results, (b) Collected training data, (c) Examples of human detection in different poses (left: standing and walking, middle and right: lying down)

*Images From: S. C. Lee, R. Nevatia, "Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system", Machine Vision and Applications (2014) 25:133–143.*

### 3.4. Results and Discussion

To conduct the experiment, real time videos have been collected by a panasonic lumix dmc lz-30 camera with the dimensions  $1280 \times 720$ , 30fps.

#### 3.4.1. Person fell down

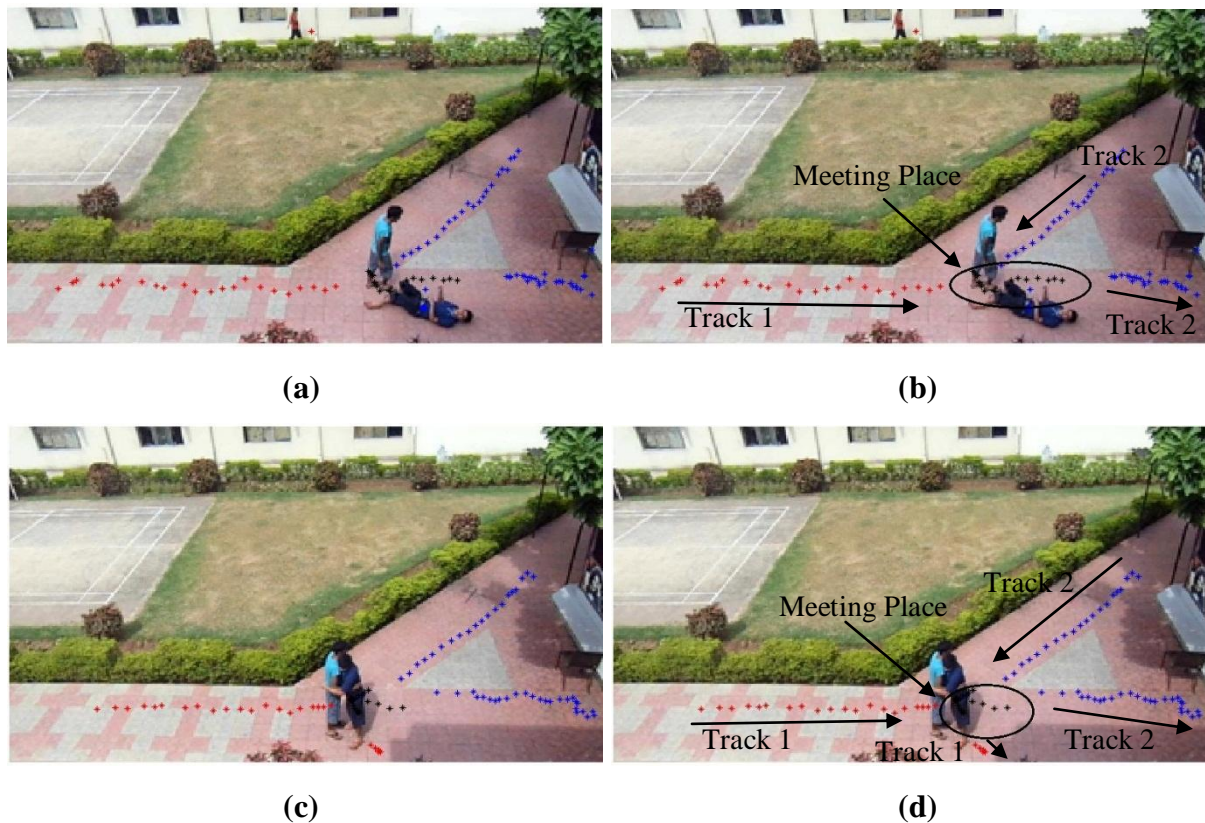


Fig. 3.9. Person Fell Down (a) Tracking of fell down event, (b) Explanation of Fell Down Event, (c) False Alarm due to Occlusion, (d) False Alarm Explanation

Fig. 3.9(a) shows the subjective view of human collision and fell down with their trajectories. Fig. 3.9(b) explains the event. Track 1 and track 2 are meeting at meeting place and two tracks are not being generated after the threshold time. So it concludes the event as an abnormal event. In the similar way Fig. 3.9(c) (d) explains a false alarm case. Two persons are meeting here but it is not a case of collision and fell down. For computer vision, it is proved by checking the number of tracks after the threshold time which concluded the event as 'normal'.

### 3.4.2. Entering into Restricted Area

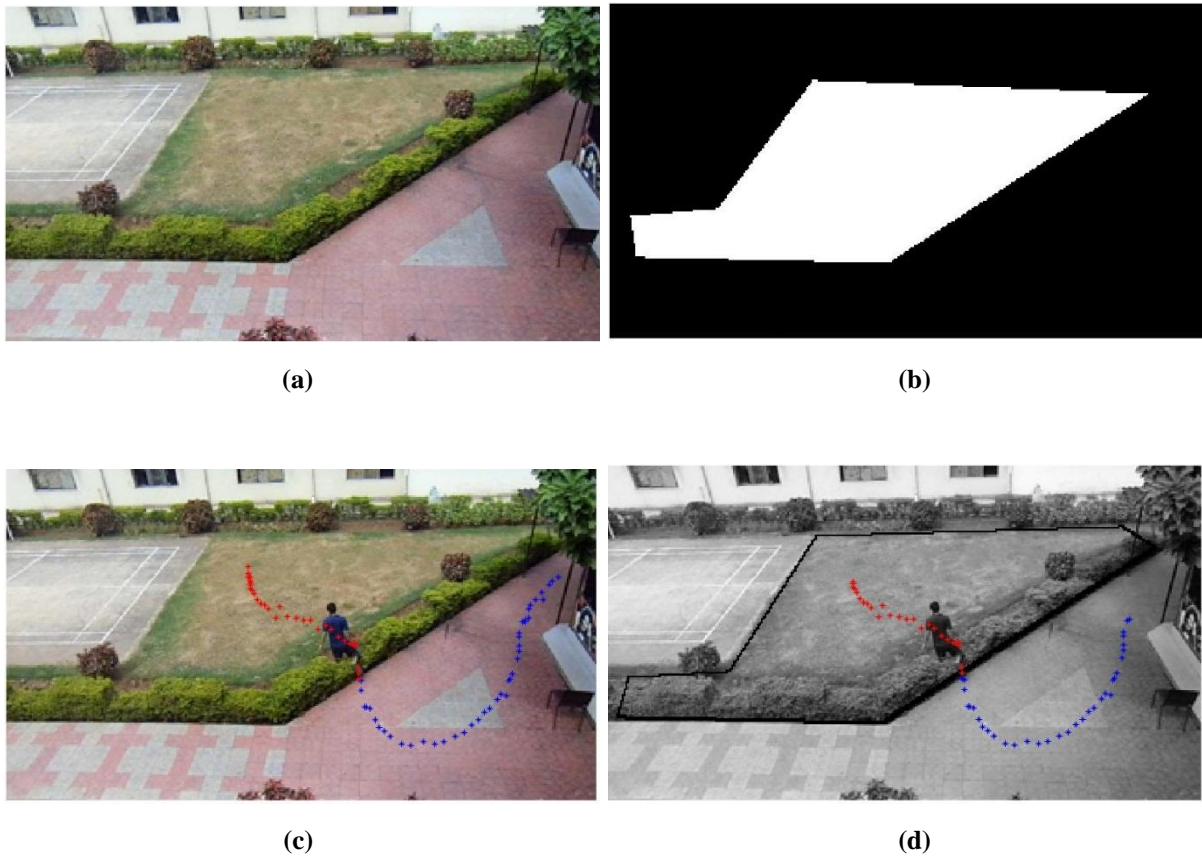


Fig. 3.10. Entering to Restricted Area (a) Background frame (b) Polygonal Restricted area mask (c) Person Enters the Restricted Area (d) Final output with restricted area marker

Fig. 3.10(a) shows the reference background frame on which the restricted area has been selected. In this frame the restricted area is the green garden, walking on which is strictly prohibited. So the garden area has been selected manually and the selected polygonal area has been shown in Fig. 3.10(b). This binary mask helped to determine whether the object's position is in restricted area or not. In Fig. 3.10(c) the track positions have been marked, where the red ones are representing the illegal entry. Finally Fig. 3.10(d) is the output frame of the system.



### 3.4.3. Loitering



Fig. 3.11. Loitering (a) loitering at a place (b) loitering mark up

Table 2 : Objective result of Activity Recognition

Test Video Type		No. Of times executed	Successful detection	False detection
Fell down	Occlusion	10	7	3
	True fell down	10	6	4
Loitering		10	10	-
Illegal entry		10	8	2

### 3.5. Summary

Human activity recognition has been explained in chapter 3. The trajectory based activity recognition is the base of the explanation in this chapter. To detect abnormality trajectory of the human being has been used and the abnormalities like human collision and fell down, illegal entry, abnormal loitering have been discussed briefly in this section. Databases have been taken and processed to prove the detection concept.

**CHAPTER 4**  
**CONCLUSIONS AND FUTURE**  
**WORK**

## **4.1. Conclusion**

This thesis work presents an efficient approach to track single human by advanced mean shift tracker in chapter 2. The human position and patch to start tracking is selected by human detection through star skeletonization. Separate databases are used to study the fast motion and occlusion problems. Block search algorithm has provided overlapping patch to continue the track in case of fast motion. Occlusion handling is successfully applied and executed to prolonged period of occlusion. Results show that the proposed method is efficient for real time tracking where the tracking starts automatically and handles the obstacles like fast motion and prolonged occlusion. The method can be applied to any real time environment and crowded areas for surveillance.

Low level activities have been detected and checked for abnormality in chapter 3. Person fell down has been detected and false alarm due to occlusion has been discarded by temporal processing of data with the help of the elapse time threshold. Illegal entry and abnormal loitering have been marked by the help of the trajectory of single human.

## **4.2. Future Work**

- The human detection and tracking can be extended to mass tracking and activity recognition in crowded areas.
- Multiple cameras can be used for data acquisition and every camera's data can be reflected onto ground plane to get more accurate information about the position of the object.
- Mean shift tracking should handle scale variation and sudden illumination change with more accuracy.
- Human modelling can be extended to distinguish low level human activities as well as high level recognition such that less requirement of classifiers will be required.

## PUBLICATION

Suraj Prakash Sahoo and Samit Ari, “Automated Human Tracking using Advanced Mean shift Algorithm”, in Proc. 4<sup>th</sup> IEEE Int. Conf. on Comm. and Signal processing, pp. 792-796, 2-4 April, 2015, Adhiparasakthi College of Engineering, Melmaruvathur,India.

## REFERENCES

- [1]. C.R. Wren and A.P. Pentland, "Dynamic Models of Human Motion," Proc. Third IEEE Intl. Conf. Automatic Face and Gesture Recognition, April 1998.
- [2]. D. Beymer and K. Konolige, "Real-Time Tracking of Multiple People Using Continuous Detection," Artificial Intelligence Center, SRI International Report, 1998.
- [3]. N. Oliver, B. Rosario, and A. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," MIT Media Laboratory Report, 1998.
- [4]. R. Rosales and S. Sclaroff, "3D Trajectory Recovery for Tracking Multiple Objects and Trajectory Guided Recognition of Actions," Proc. of IEEE Conf. On Computer Vision and Pattern Recognition, June 1999.
- [5]. A. Azarbayejani and A. Pentland, "Real-Time Self-Calibrating Stereo Person Tracking Using 3D Shape Estimation from Blob Features," Proc. of the ICPR, IEEE, 1996.
- [6]. C. Bregler, "Learning and Recognizing Human Dynamics in Video Sequences," IEEE Press, pp 568-574, 1997.
- [7]. C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, No.7, July 1997.
- [8]. J. Ben-Arie, Z. Wang, P. Pandit, and S. Rajaram, "Human Activity Recognition Using Multidimensional Indexing," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, No. 8, August 2002.
- [9]. R. Cutler and M. Turk, "View-Based Interpretation of Real-Time Optical Flow for Gesture Recognition," University of Maryland, College Park Report, 1997.
- [10]. A. Cretual, F. Chaumette and P. Bouthemy, "Complex Object Tracking by Visual Servoing Based on 2D Image Motion," Proc. of the IAPR Intl. Conf. on Pattern Recognition, vol 2, pp 1251-1254, Australia, August 1998.
- [11]. C. Eveland, K. Konolige, and R. Bolles, "Background Modeling for Segmentation of Video-Rate Stereo Sequences," Report, Computer Science Department, University of Rochester, 1997.
- [12]. Z. Wen and Z. Cai, "A Robust Object Tracking Approach using Mean Shift", *3rd International Conference on Natural Computation*, Vol. 2, pp. 170-174, 2007

- [13]. J. Jeyakar, R. V. Babu and R. K. Ramakrishnan, “Robust Object Tracking using Kernels and Background Information,” *IEEE International Conference on Image Processing*, Vol. 5, pp. 49-52, 2007
- [14]. F. Tung, J. S. Zelek, and D. Clausi, “Efficient target recovery using stage for mean-shift tracking”, In *Canadian Conf. Computer and Robot Vision*, pp. 16-22, 2009
- [15]. M Zhao, L Wang, J Han, “An Adaptive Tracking window Based on Mean-Shift Target Tracking Algorithm”, in *Chinese Automation Congress*, pp. 348 – 352, 2013
- [16]. S. C. Lee, R. Nevatia, “Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system”, *Machine Vision and Applications* (2014) 25:133–143
- [17]. R. Bodor, B. Jackson, N. Papanikolopoulos, “Vision-Based Human Tracking and Activity Recognition” Proc. of the 11th Mediterranean Conf. on Control and Automation, June 2003
- [18]. B. Tamersoy, "Background Subtraction – Lecture Notes", University of Texas at Austin, September 29, 2009
- [19]. K. Kim, T.H. Chalidabhongse, D. Harwood, and L. Davis, “Background Modelling and Subtraction by Codebook Construction,” Proc. IEEE Int’l Conf. Image Processing, vol. 5, pp. 3061-3064, 2004
- [20]. M. Zhariy, “Introduction to Optical Flow”, Uttendorf, 2005
- [21]. C. Suliman, C. Cruceru, Florin Moldoveanu, “Kalman Filter Based Tracking in an Video Surveillance System,” 10th International Conference on Development and Application Systems, Suceava, Romania, May 27-29, 2010
- [22]. Wen J, Gao X, Yuan Y, Tao D, JieLi, “Incremental tensor biased discriminant analysis: a new color-based visual tracking method”, *Journal of Neurocomputing*, pp. 827–839, 2010
- [23]. H Zhi-qiang, L. Xiang, Y. Wang-sheng, L. Wu, Huang An-qi, “Mean shift tracking Algorithm with improved Background weighted Histogram”, in *5th International Conference on Intelligent Systems Design and Engineering Applications*, pp. 597 – 602, 2014.
- [24]. Comaniciu D, Ramesh V, Meer P. , “Real-time tracking of non-rigid objects using mean shift”, *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 142–149, 2000.
- [25]. Chen J J, An G C, Zhang S F, et al., “A mean shift algorithm based on modified Parzen window for small target tracking”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1166–1169, 2010

- [26]. S. Birchfield, Elliptical head tracking using intensity gradients and color histograms, in: Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 232–237.
- [27]. R.T. Collins, Mean-shift blob tracking through scale space, in: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. 234–240.
- [28]. D. Comaniciu, V. Ramesh, P. Meer, Kernel-based object tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (5) (2003) 564–577.
- [29]. P. Pérez, C. Hue, J. Vermaak, M. Gangnet, Color-based probabilistic tracking, in: Proceedings of the 7th European Conference on Computer Vision, vol. 1, 2002, pp. 661–675.
- [30]. N.S. Peng, J. Yang, Z. Liu, Mean shift blob tracking with kernel histogram filtering and hypothesis testing, *Pattern Recognition Letters* 26 (2005) 605–614.
- [31]. A. H. Mazinan, A. Amir-Latifi, “Applying mean shift, motion information and Kalman filtering approaches to object tracking,” *ISA Transactions*, Vol. 51, Issue 3, pp. 485-497, 2012
- [32]. A. Ali and K. Terada, “A framework for human tracking using kalman filter and fast mean shift algorithms,” *IEEE Intl. Conf. Computer Vision Workshops*, pp. 1028-1033, 2009.
- [33]. Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, “Review and evaluation of commonly-implemented background subtraction algorithms”, *In IEEE 19th International Conference on Pattern Recognition*, pp. 1-4, 2008
- [34]. Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing* (3rd Edition). Prentice Hall, 3 edition, August 2007.
- [35]. N. Zarka, Z. Alhalah and R. Deeb. "Real-Time Human Motion Detection and Tracking," *3rd International Conference on ICTTA*, pp. 1-6, 2008