CHAPTER-1

INTRODUCTION

1.1 OVERVIEW

Object Tracking is one of the most important processes in the area of computer vision. Its objective is to locate a moving object or several objects in video sequences, using a single camera or multiple cameras. A tracking algorithm is used, whose purpose is to analyze the video frames in isolation and predict the location of moving targets in the video frames. So, we can define tracking as the task of localizing an object of interest from a video scene by means of segmentation and afterwards keeping track of its motion, orientation, occlusion etc. in the successive frames. This is done to extract useful information from the video sequences which can be used for various purposes, for example, traffic surveillance, crowd monitoring, robot navigation and many such applications.

The sheer growth in the number of high-powered computers in today's world, and the increasing need for analyzing automated video have led to generation of a great deal of interest in visual object tracking algorithms. The use of visual object tracking is of most relevance in the tasks of automated surveillance, traffic monitoring, vehicle navigation, human-computer interaction etc. Automated video surveillance observes people or vehicles on real time in busy crowded environments, whereby their activities can be tracked, monitored and analyzed as long as they are in the field of view.

There are three key steps in object tracking:

1. Detection of interested moving objects
2. Tracking of such objects from frame to frame, and
3. Analysis of object being tracked to recognize their behaviour.

1.2 FEATURE SELECTION FOR TRACKING

One of the most significant tasks in object tracking is selection of proper feature for tracking and it is directly related to the representation of the object. For example, we use color as a feature for histogram based methods, whereas edges are the features commonly used for contour based representation. Now-a-days, most of the tracking algorithms combine more than one feature and then use it for tracking.

The visual features are as follows:

- **Color**: Color of an object is affected by two factors:
  1. Spectral power distribution of the surface reflectance and
  2. Illuminant properties of the object

  Different color models commonly used are RGB, L*a*b and L*u*v that we use to represent color.

- **Edges**: We use edge detection to identify heavy changes in intensities of the image generated by boundary of the object. Edges are more sensitive to color features as compared to illumination changes. Canny Edge detector is one of the most popular edge detection approach.

- **Optical Flow**: Optical Flow is a dense field of displacement vector which defines how each pixel translates in a region. It is measurable, and is calculated using the brightness constraint (which assumes brightness of corresponding pixels in consecutive frames would be constant). It is generally used as a feature in tracking application and motion based segmentation.

- **Texture**: It refers to the measure of a surface's intensity variation. For this measurement, properties such as smoothness and regularity are quantified. A processing step is required to generate the descriptors. There are various texture descriptors: Grey Level Co-occurrence Matrices, loss texture measures, wavelets, and steerable pyramids.

1.2 ISSUES IN OBJECT TRACKING

Some of the constraints that are generally imposed during object tracking are:

- Smooth motion of object having no sudden changes in the motion
- No abrupt changes of the background
- Gradual changes in the appearance of object
- Fixed camera
- Number and size of objects
- Limited amount of occlusion

Some of the fundamental open problems in object tracking are:

- Abrupt object motion
- Noise in the image sequences
- Changes in scene illumination
- Changing appearance patterns of the object and the scene
- Object-to-object and object-to-scene occlusions
- Non-rigid object structures
- Camera motion
- Real time processing requirements

There are a number of issues involved in the development of a robust object tracking system, which needs to be understood.

Object Modeling is an important issue in visual object tracking. One of the major tasks of object modeling is to find an appropriate visual description that makes the object distinguished from other objects and background.

Changes in appearance and shape are issues that should also be considered during visual object tracking. The appearance of an object can vary as camera angle changes. Deformable objects such as human can change their shape and appearance during different video frame sequences. The appearance and shape can also change due to perspective effect i.e. objects farther from the camera appears smaller than those near to the camera.

Handling illumination changes is also one of the challenging issues for visual object tracking. The appearance of an object can be largely affected by illumination changes. An object may look different in indoor environment (artificial light) than outdoor environment (sun light). Even the time of day (morning, afternoon, evening) and weather conditions i.e. cloudy, sunny etc. can be the causes of illumination changes.

Shadows and reflections are also difficult to handle during object tracking. Motion, shape and background are some of the features such that are very sensitive to shadows cast on the ground because it looks just like the object that casts it. Same kind of problem can be caused by reflections of moving objects on smooth surfaces.

Occlusion is also an important issue for visual object tracking. Occlusion can be classified into different categories on the basis of what occludes the object being tracked. These are self occlusion; inter object occlusion and object-to-background occlusion. During occlusion, an ambiguity occurs in the objects and their features. The tracking methods must be capable to resolve the individuality of the objects involved in the occlusion, before and after the occlusion takes place.

The issues mentioned above are significant to both single-object tracking and multiple object tracking. However, tracking of multiple objects also requires resolution of some other issues such as modeling of multiple objects. Tracking method should be able to distinguish different objects in order to keep them consistently labeled. Although during the last few years, there has been a substantial progress towards moving object detection and tracking. But tracking an object in an unconstrained, noisy and dynamic environment still makes this problem a central focus of research interest.

1.4 PROBLEM STATEMENT

This thesis aims to improve the performance of object detection and tracking by contributing originally to two components (a) motion segmentation (b) object tracking.

The goal of the work in this thesis is twofold:

- To set up a system for automatic segmentation and tracking of moving objects in stationary camera video scenes, which may serve as a foundation for higher level reasoning tasks and applications
- To make significant improvements in commonly used algorithms. Finally, the aim is to show how to perform detection and motion-based tracking of moving objects in a video sequences taken from a stationary camera.

Therefore the main objectives are:

- To analyze Graph-cut segmentation algorithm to detect the objects.
- To analyze Kalman tracking method for tracking the single objects and multiple objects.

1.5 THESIS LAYOUT

The report is organized in six chapters. Brief outline of the chapter contents is given below:

Chapter-1 gives an overview of object tracking, issues in object tracking, the problem statement and objectives of the thesis.

Chapter-2 discusses the literature review of object tracking relevant to this thesis.

Chapter-3 provides a brief introduction to techniques used in the thesis, which include Graph cut, segmentation, Kalman tracking and background subtraction.

Chapter-4 discusses the presented approach along with the flow chart diagrams and algorithms.

Chapter-5 presents the results of the presented technique.

Chapter-6 highlights the conclusions drawn about the method from the results shown in Chapter-6 and the scope of future research on this method.

CHAPTER-2

LITERATURE REVIEW

2.1 OVERVIEW

Different authors have tried different methods to track multiple objects like particle filters, background-foreground segmentation approaches and so on. The review includes applicability of such methods to multiple object tracking, their advantages and limitations and the scope of any future work, if possible, in these methods.

2.2 LITERATURE SURVEY

This section provides a literature review of various research papers that have used different methods to handle occlusion in the field of object tracking. It mainly emphasizes on graph cut methods. Graph cut based tracking for multiple object tracking has been presented in this thesis.

A template matching based method was proposed by **Hieu T. Nguyen et al.** in [18] to track rigid objects in video sequences. Template updation is done by smoothing appearance features with the help of Kalman filters that are robust. This stabilizes the tracking results and also, the noise is suppressed. Position, shape, motion and appearance are the features used for tracking. Photometric features that represent the region of target in the video frame are included in appearance attributes. Robust error norm and the Mahalanobis distance are used to down weight the outliers that occur due to the presence of partial occlusion in video sequence. Because partial occlusions are well handled by the method, it is an accurate detection method and can be used where severe partial occlusions are present. Any abrupt lighting condition changes are also handled by this method and it can be used in real time applications because of its low computational complexity. Low computational complexity is due to the use of very less parameters by the method.

Karthik Hariharakrishnan et al. in [19] proposed a method for tracking which does not use segmentation. Segmentation is only used to initialize an object in the first few frames. Because segmentation is omitted, the algorithm is simple and fast. Bloch motion vectors are the features used by them to track objects by predicting their boundaries. After the prediction step, updation of contour is done using a new concept of occlusions and disocclusion. Motion estimation in the video frames is done by an adaptive block matching algorithm. Occlusion and disocclusion are dual of each other. To detect disocclusion, motion compensation of the current frame is done by the use of previous frame to get uncovered regions; and to detect occlusion, it is done by the use of next frame to get covered regions. When disocclusion (occlusion) occurs, the uncovered (covered) regions of the object possess motion features which are like (dissimilar from) those of the object. The disocclusion determining algorithm checks for the similarity of these motion features. This duality principle can be exploited to detect occlusions from an algorithm formulated to detect disocclusion. Contextual information is also used by the authors. The proposed approach can be used for real time applications because parallel processors can be used to implement the approach making the process fast. The new concept of disocclusion and the duality principle allows tracking of objects for a longer period of time without the need of re-initialization or re-segmentation. But the proposed algorithm fails to track for smaller objects.

Zhen Jia et al in [20] proposed to combine Bayesian framework with the already existing template matching methods to track targets in video sequences. The template matching methods used in [20] by the authors are Sum of the Squared Errors (SSE) and Gaussian Mixture Models (GMMs). These are first weighted by their matching accuracies and then fused with the Bayesian theory to achieve robust results in template updation and matching. The SSE template matching is used to locate the template in the image with the minimum Sum of the Absolutely Squares (SAD) between the image and the template’s pixel intensity value, while the GMMs is used to represent the appearance of template. The GMMs also provides the mixture Gaussian distribution functions for estimating the accuracy of matching process. This algorithm can deal with the problems of template drifting, shape deformation and occluded object matching. But the resort to use first template as the benchmark to reduce drift makes the method ineffective where there are major changes in the target being tracked.

Jiyan Pan et al. in [11] proposed an algorithm that combines the spatiotemporal context information, information of reference target and the motion constraints altogether. It is named Content Adaptive Progressive Occlusion Analysis (CAPOA) algorithm. Targets and outliers can be clearly differentiated by this method. An occluded target can be accurately tracked by correcting the location of target using the variant-mask template matching (VMTM). The problem of template drift is severe during template update process. It is handled with the help of a specially designed Kalman filter, known as drift-inhibitive masked Kalman appearance filter (DIMKAF). A local best match authentication (LBMA) algorithm is also proposed to deal with targets that are completely occluded. It is a robust tracking method and one of the very few methods that handle complete occlusion.

A new method to deal with completely occluded targets was proposed by **G.Di Caterina et al.** in [22]. It is a template matching method, which is adaptive in nature and is based on SWAD tracker. Different updation rates are used for different parts of the template and at the same time, the template structure is conserved by not allowing any major changes or erroneous updates in it. The method has the ability to indicate whether the target is occluded or not. It is able to detect the completely occluded target after the target overcomes the complete occlusion and is visible again.

A new graph cut based method to track single and multiple objects in the region of interest was proposed by **James Malcolm et al.** in [12], [13]. Distance based pixel weights are assigned to each object so as to restrict segmentation in that area only.

Location prediction of objects is done by the use of a filter. The pixel weights based upon distance are then put at the predicted location of the objects and are scaling is done. This scaling is adaptive in nature and a prediction confidence measure is used to achieve this. It can track multiple objects that differ in intensity profiles and tracking is accurate in both color and gray scale videos.

The use of min-cut/max flow energy minimization algorithms in tracking and segmentation was proposed by **Aur´elie Bugeau and Patrick P´erez** in [9], [10]. No data association method is used in this to track multiple objects. A combination of low level measures and high level observations is used to form objective functions. Low level measures are pixel-wise color and motion. High level observations are acquired by three different processes: detection which is independent of any other process, motion prediction and contextual regularization which is sensitive to contrast. The regions of video frames (pixel sets) are the observations that can be obtained from any kind of a detector. A cost function is minimized by min-cut/max flow optimization algorithms and it allows "detection-before-track" tracking and segmentation of tracked targets at the same time. A second stage of minimization process is required when the tracked objects lie close to each other and the detection process mixes them. This helps to accurately track and segment closely spaced objects despite the failure of external detection process. The proposed method performs accurate tracking in the presence of partial occlusions and it is robust to observations that are missed and does not require 100% accurate observations to provide good segmentations. Because a second stage of energy minimization is used, the method helps to track and segment objects that are missed in the first stage. But the method fails when there are extreme and abrupt changes in illumination.

A graph cut based tracking method is proposed by **Nicolas Papadakis, and Aure´ lie Bugeau** in [8] within video sequences. The major reason to use graph cuts is its low computational complexity as compared to exiting segmentation techniques. The paper mainly focuses on representation of occluded and visible parts of the objects. The aim is to track both these parts. Since the location of objects does not change significantly in video frames, predictions can be used to track the objects. The predicted location of objects is divided into good and bad labels with respect to segmentation results. Each object is associated with a label and an energy function is minimized using graph cuts minimization for multiple labels. It provides a fast and accurate tracking method for multiple objects. Partial and total occlusions are correctly handled by the method.

The new concept of ‘bag of features’ (BoF) is used by **F. Yang et al.** in [21] to track objects in video sequences. A codebook is needed to implement BoF algorithms. To construct this, first of all, some initial frames are taken on which incremental PCA visual tracking (IVT) is performed. From these frames image patches are randomly taken. In each frame, these are then sampled inside the region of tracked object. A constant number of candidate targets are generated by this sampling process for the purpose of tracking. Tracked object can now finally be represented as a bag. [21] uses two features to construct codebooks, viz, RGB color features and texture features represented by Local Binary Pattern (LBP). Thus, two codebooks have been used in this method so as to get more information about the objects being tracked. The pose and appearance changes of the targets are also taken care of by the use of an updating mechanism. Bags are trained by the use of training video sequences. Thus, candidate’s patch similarity can be computed with the codewords and its bag similarity with trained bags. The target in test videos is located by calculating the maximal combined similarity of patches and bags. This is a rather new area in object tracking but gives good results.

Bing-Fei Wu et al. [23] proposed a relative discriminative histogram of oriented gradients (HOG) (RDHOG)-based particle filter (RDHOGPF) approach to traffic surveillance with occlusion handling. Based on the conventional HOG, an extension known as RDHOG is proposed, which enhances the descriptive ability of the central block and the surrounding blocks RDHOGPF can be used to predict and update the positions of vehicles in continuous video sequences. RDHOG was integrated with the particle filter framework in order to improve the tracking robustness and accuracy. To resolve multiobject tracking problems, a partial occlusion handling approach is addressed, based on the reduction of the particle weights within the occluded region. RDHOGPF is highly effective and it delivers good performance in many different traffic situations. The predicted trajectory is closer to that of the real rigid body. The proposed RDHOGPF can determine the target by using the feature descriptor correctly, and it overcomes the drift problem by updating in low-contrast and very bright situations.

Efforts were made to track multiple objects in the presence of occlusion by **Mohammad Azari et al.** in [24] by the use of a new distance measure in General Nearest Neighbor (GNN) method to calculate data association. Object localization is done by background subtraction method. When there is no occlusion, Kalman filter, whose measurement noise covariance is kept constant, is used to track objects. But when occlusion occurs, the measurement noise covariance of the Kalman filter is made adaptive. It is done by correlation coefficient method based local template matching, whose outputs are used to vary the measurement noise covariance of the Kalman filter in case of occlusions. This method is able to detect occlusions and track occluded objects. Segmentation errors are minimized by this method.

CHAPTER-3

BASIC CONCEPTS

3.1 OVERVIEW

This chapter deals with the basic concepts that have used to develop this thesis. The concepts are explained in relevance to their use in the thesis in the following sections.

3.2 GRAPH CUTS

In graph theory, a graph is shown by **G=(V, E)** where V is the set of vertices or nodes and E is the set of edges. In a graph **G**, when the vertices are partitioned into two subsets such that the subsets are disjoint, the partition is known as a **graph cut** or simply a cut. Any cut determines a cut-set, which is the set of edges having one endpoint in each subset of the partition.

In a flow network, there are two terminal nodes known as the source and the sink. In such networks, a cut is defined as an **s-t cut.** It is one such cut where the source and the sink are in different subsets. The cut-set of an s-t cut only contains edges going from the source's side to the sink's side.

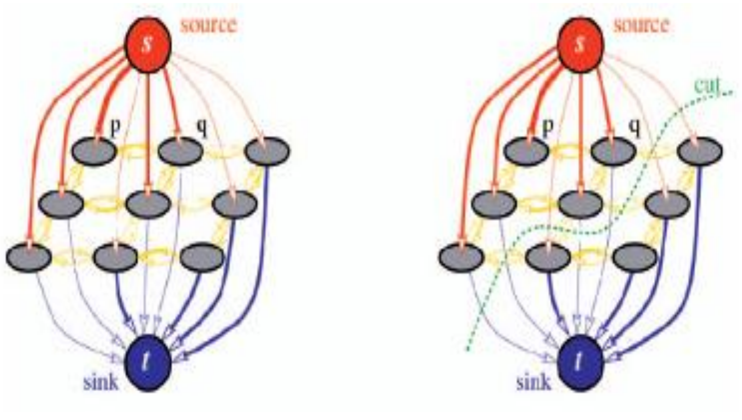

Fig 3.1 Example of a directed capacitated graph. Thickness reflects edge weights.

3.2.1 Definition

A cut **C= (S, T)** is a partition of the set V of a graph G = (V, E) into two disjoint subsets *S* and *T*, where V is a set of vertices or nodes and E is a set of directed edges. The cut-set of a cut $C = (S, T)$ is the set $\{(u, v) \in E | u \in S, v \in T\}$ of edges that have one endpoint in *S* and the other endpoint in *T*. If *s* and *t* are specified vertices of the graph *G*, then an *s*–*t* cut is a cut in which *s* belongs to the set *S* and *t* belongs to the set *T*.

3.2.2 Capacity and Flow

The capacity of a cut **C=(S, T)** is the sum of weights of edges leaving S.

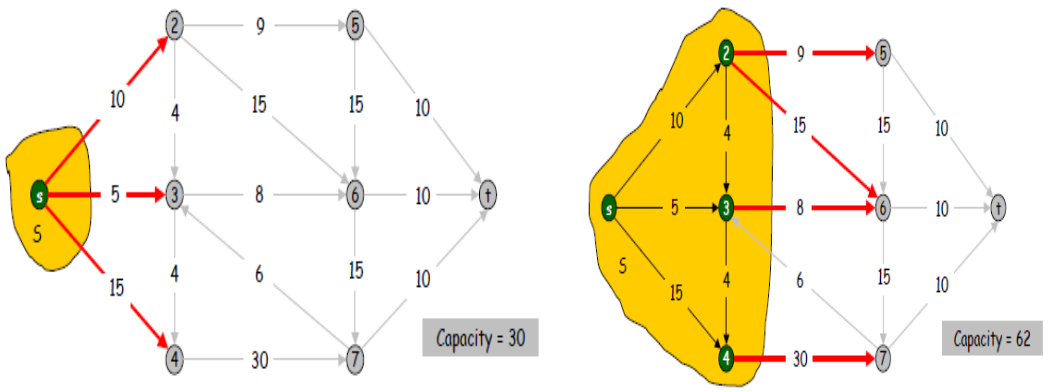

Fig 3.2 Illustration of Capacity of a Cut

A flow **f** is an assignment to weights of edges such that:

Capacity: $0 \leq f(e) \leq u(e)$

The flow conservation theorem states that the total flow leaving *v* is equal to the total flow entering *v*, except at *s* or *t*.

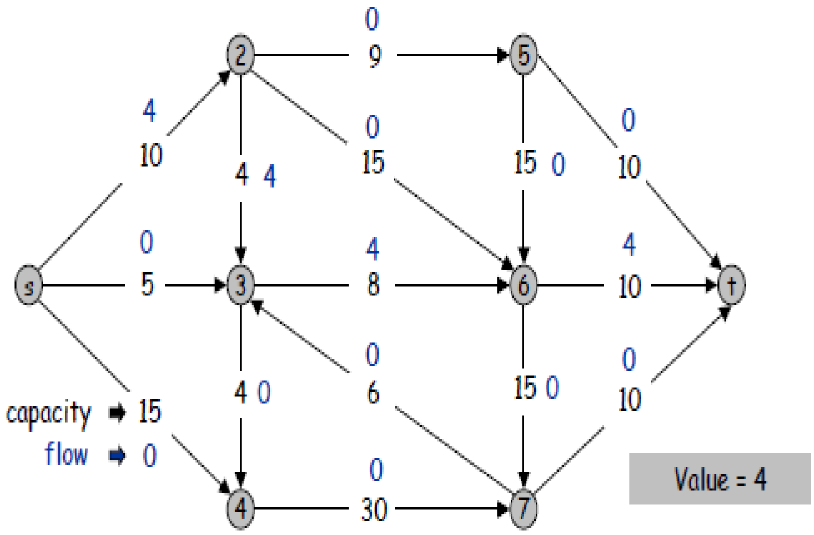

Fig 3.3 Illutration of Flow

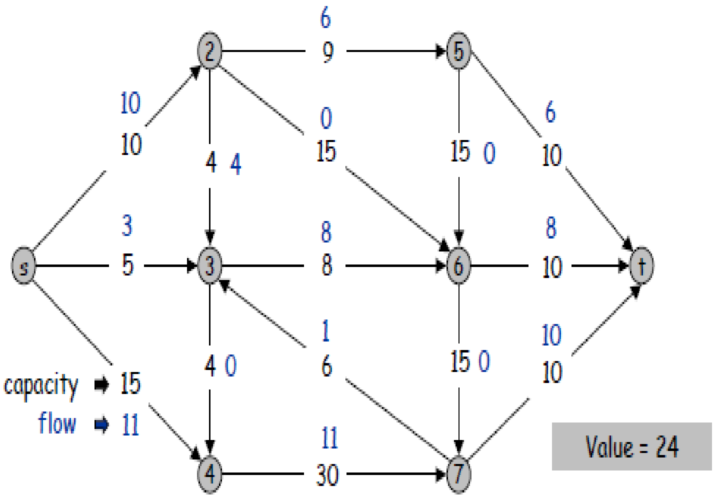

Fig 3.4 Flow Assignment

3.2.3 Minimum Cut

A cut is said to be a minimum cut if the size or weight of the cut is the smallest as compared to any other cut of the same graph. The max-flow min-cut theorem proves that the maximum network flow and the sum capacity of any minimum cut that separates the source and the sink are equal. This is discussed in the following sections.

3.2.4 Maximum Cut

In a graph, a cut whose size is at least the size of any other cut of the graph is known as **maximum cut**. The problem of finding a maximum cut in a graph is known as the **Max-Cut Problem.**

The Max-Cut problem can be defined simply as follows: To extract a subset *S* of the vertex set in such a way that the number of edges between *S* and the complementary subset is as large as possible.

In **Weighted Max-Cut** each edge has an associated real number, that is, its weight, and the objective is to maximize the total weight of the edges between *S* and its complement.

3.2.5 Max-Flow Min-Cut Theorem

The Max-Flow Min-Cut Theorem was first proposed by Ford-Fulkerson in 1956. It states that in any flow network, the maximum amount of flow passing from the source to the sink is equal to the minimum capacity that, when removed in a specific way from the graph allows no flow to pass from the source to the sink, that is, the value of max flow equals the capacity of min cut.

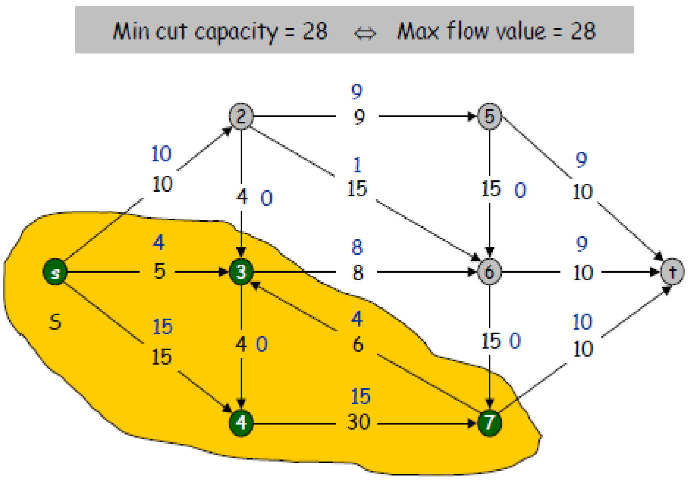

Fig 3.5 Max-Flow Min-Cut Theorem

3.3 GRAPH CUTS IN COMPUTER VISION

The problems which aim for energy minimization in low level computer vision problems can be solved by the method of graph cuts. These include image smoothing, image segmentation, stereo correspondence problem and so on.

Greig et al. [1] were the first to discover that powerful min-cut/max-flow algorithms can be used to solve important energy minimization problems in computer vision. They used graph cut technique to solve the problem of binary image restoration. It was for nearly about 10 years that the graph cut technique proposed by Greig et al. in [1] remained unpopular. The reason for this can be cited because of the confined area of binary image restoration. Early attempts to use graph cuts were restricted to image clustering. Later, the problem of which energy functions can be minimized by graph cuts was solved by Kolmogorov in [3], who provided a simple, necessary and sufficient condition to be applied on such energy functions.

A growing number of publications now address the use of graph cuts in image segmentation and other computer vision problems.

3.4 GRAPH CUT SEGMENTATION

This section describes how graph cuts are applied to the problem of image segmentation. The technique has been used for the segmentation of video frames in the thesis.

3.4.1 Graph Construction

A graph $G = \{ V, E\}$ is a set of nodes or vertices V and a set of edges E connecting neighboring nodes. Figure 3.6 shows a simple example of an undirected graph that can be used for image segmentation.

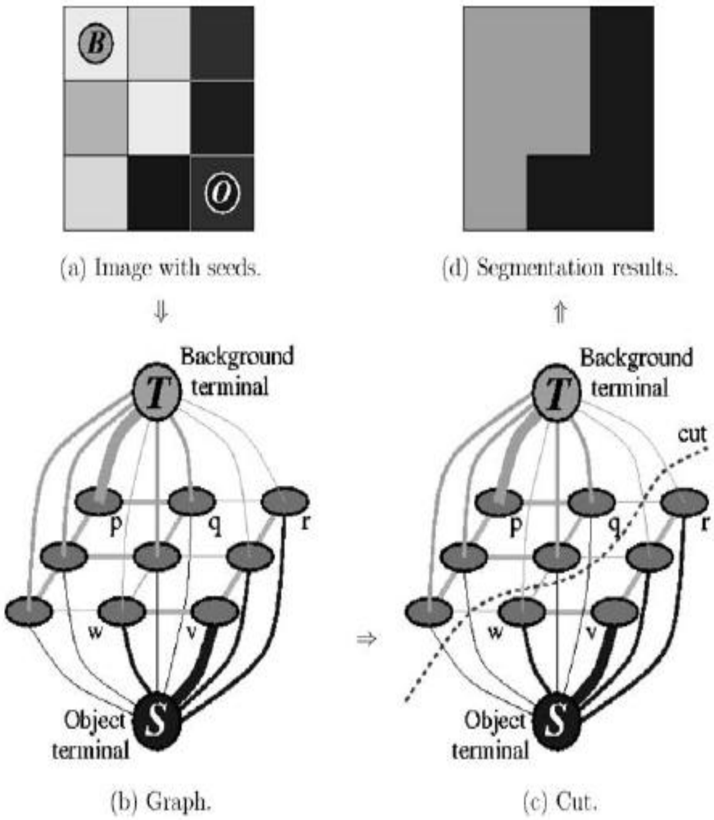

Fig 3.6 A simple 2D segmentation example of 3X3 image

The nodes of graphs in image segmentation problems represent image pixels. There are also two specially designated terminal nodes *S* (*source*) and *T* (*sink*) that represent "object" and "background" labels. Typically, neighboring pixels are interconnected by edges in a regular grid-like fashion. Edges between pixels are called *n-links* where *n* stands for "neighbor". Another type of edges, called *t-links*, are used to connect pixels to terminals. All graph edges $e \in E$ including n-links and t-links are assigned some nonnegative weight (cost) w_e.

3.4.2 Segmentation Energy

Image Segmentation problem can be modeled as Energy minimization of MRF with Graph containing image pixels. In this framework, one seeks the labeling f that minimizes the energy,

$$E(f) = E_{smooth}(f) + E_{data}(f)$$

Here $E_{smooth}(f)$ measures the extent to which f is not piecewise smooth, while $E_{data}(f)$ measures the disagreement between f and the observed data. Many different energy functions have been proposed in the literature. The form of $E_{data}(f)$ is typically,

$$E_{data}(f) = \sum_{p \in P} D_p(f_p)$$

where D_p measures how well label f_p fits pixel p. Mostly used model for E_{smooth} is Potts model given as,

$$E_{smooth}(f) = \sum_{\{p,q\} \in N} u_{\{p,q\}} . T(f_p \neq f_q)$$

where T is indicator function i.e. it will output 1 if the input condition is true.

3.4.3 Segmentation

Each pixel *p* has two t-links {*p*, *S*} and {*p*, *T* } connecting it to each terminal. Each pair of neighboring pixels {*p*, *q*} in *G* is connected by an n-link. Without any ambiguity, an n-link connecting a pair of neighbors *p* and *q* is also denoted by {*p*, *q*}. The graph *G* is now completely defined. The segmentation boundary between the object and the background is drawn by finding the minimum cost cut on the graph *G*.

This is done by the minimization of energy using α- expansion algorithm.

1. Start with an arbitrary labeling f
2. Set success := 0
3. For each label $\alpha \in \mathcal{L}$
    3.1. Find $\hat{f} = \arg\min E(f')$ among f' within one α-expansion of f
    3.2. If $E(\hat{f}) < E(f)$, set f := $\hat{f}$ and success := 1
4. If success = 1 goto 2
5. Return f

3.5 BACKGROUND SUBTRACTION

The background subtraction method is one of the very simple and promising approaches for extracting moving objects from video sequences. In this approach, we have a reference frame (known as the background image) which is compared with the current frame. In case the difference between the frames is significant, it means that moving objects are present in it. However, there is a catch. The reference frame needs to be modeled / updated adequately and regularly, or else this approach becomes vulnerable to environmental condition changes such as structural background changes and illumination. So background modeling is one of the primary and challenging tasks for background subtraction. The background subtraction algorithm should be robust against environmental changes i.e. it should be capable to handle changes in illumination conditions and intelligent enough to be able to ignore the movement of small elements in the background.

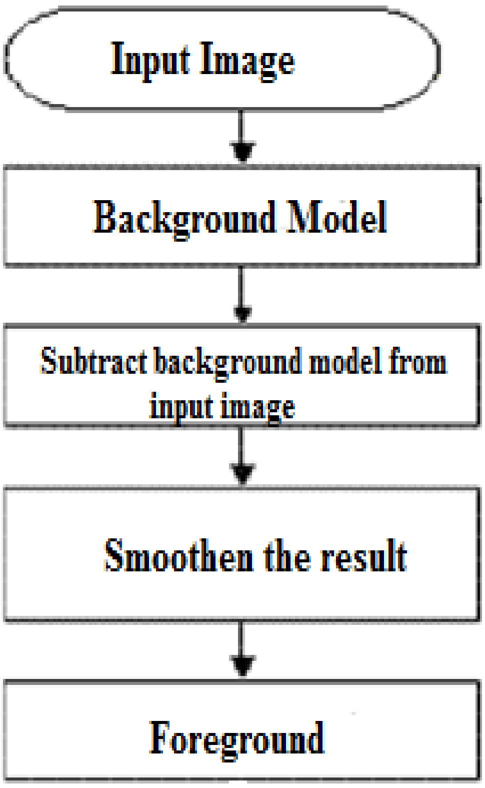

Fig 3.7 Flow Chart of Background Subtraction

In the thesis wok, background subtraction is applied to segmented frames to extract the foreground. Since the background is stationary, foreground consists of moving objects only.

3.6 KALMAN FILTERING

The Kalman filter estimates the position of the object in each frame of the sequence. Because of the movement of the object during the sequence, the input parameters of the Kalman filter, the position of the object in the image at time k, the object’s size and the width and length of the search window of the object vary. These parameters represent Kalman filter’s state vector and the measurement vector. In general, the process of estimation of parameters followed with a Kalman filter requires the following steps:

- The measure, to take the tracking parameters.
- The estimate, which updates the position of the object.
- The prediction, which calculates the position of the object in the next frame.

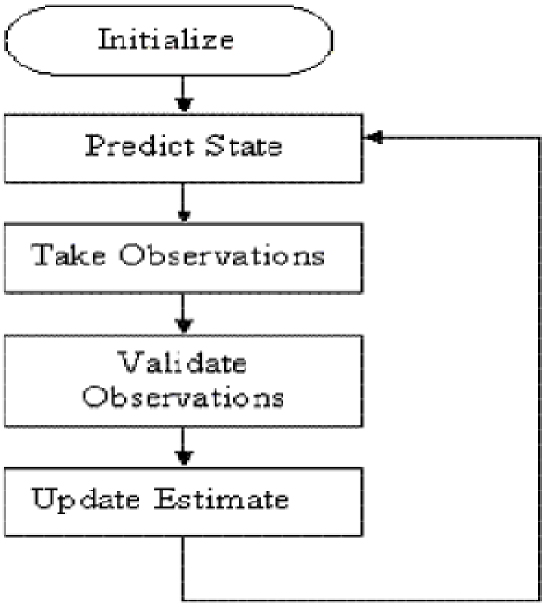

Fig 3.8 Flow chart of Kalman Filter

The variable parameters of the Kalman filter are the state vector and the measurement vector:

The state vector is presented by equation:

$$s_k = (x_k, y_k, W_k, L_k, x_c, y_c)$$

Where W_k is the width of the search window; L_k is the length of the search window; (x_k,y_k) is the center of mass of the object at time t_k; and (x_c,y_c) is the initial position.

The measurement vector of the Kalman filter is given by equation

$$z_k = (x_k, y_k, W_k, L_k)$$

Where W_k is the width of the search window; L_k is the length of the search window; (x_k,y_k) is the initial position of the object at time t_k

3.6.1 Basics of Kalman Filter

Kalman filtering is composed of the following steps:

1. **Process to estimate**: The Kalman filter estimates the states of a discrete process. This state is modelled by the linear equation

   $$s_k = A \times s_{k-1} + w_{k-1}$$

   where A is the transition matrix, w_k is the noise process and dt is the difference between the two moments k and k −1.

   The measurement model is defined by the equation:

   $$z_k = H \times s_k + v_k$$

   with H as the measurement matrix. As the noise processes "w_{k-1}" and "v_k" are assumed independent of state vectors and measurement, these are modeled as:

$$p(w) \sim N(0, Q)$$
$$p(v) \sim N(0, R)$$

2. **The equations for updating**: Finally, the output equations for the two blocks of prediction and correction of Kalman filter are:

- **Equations for predicting**

$$\hat{s}_k = A \times \hat{s}_{k-1}$$
$$P_k^- = A \times P_{k-1} \times A^T + Q$$

- **Correction Equations**

$$K_k = P_k^- \times H^T \times (H \times P_k^- \times H^T + R)^{-1}$$
$$\hat{s}_k = \hat{s}_k^- + K_k \times (z_k - H \times \hat{s}_k^-)$$

$$P_k = P_k^- - K_k \times H \times P_k^-$$

3.6.2 Kalman Filter for Multiple Object Tracking

The Kalman filter algorithm involves two steps:

1. Prediction - It uses the previous states to predict the current state.
2. Correction (update step) - It uses the current measurement, such as object location, to correct the state.

Tracking multiple objects poses some additional challenges for us:

- Multiple detections have to be associated with the correct tracks
- New objects that appear in a scene should be handled with utmost care
- Object identity must be maintained when multiple objects merge into a single one

All these issues are well handled by the Kalman filter approach for tracking multiple objects in the scene. The limitation of using this approach is that it can handle only small occlusions.

Kalman filter is applied to track multiple objects in this thesis. After the segmentation of the moving objects has been done by the use of graph cuts and background subtraction, some process preparations is required for tracking moving objects by Kalman filter approach subsequently. Firstly, each moving object in the scene should be assigned a tracking window. Secondly, tracking window should not be kept too large. This is to prevent noise from interfering with the tracking process. We make the tracking window such that its size is slightly larger than the object image size, so that it not only can decrease the noise interference, but also can decrease time taken for image processing and increase speed of operation.

CHAPTER 4

GRAPH CUT BASED TRACKING METHOD

4.1 OVERVIEW

This chapter explains the Graph Cut Based Tracking of Multiple Objects. It discusses the presented method along with a flow chart diagram and the implementation details of the method. The presented method is relatively a new method to be used for object tracking. The method offers several advantages over existing methods, which are discussed in the following chapters.

4.2 PRESENTED METHOD

In this thesis, a graph cut based method is used to track multiple objects in a video sequence. The presented method lies in the category of "**detect before track**" methods. The method is briefly explained in this section.5

The video sequence, wherein objects are to be tracked, is first converted into frames. Then, graph cuts are used to segment these video frames and from these segmented video frames, moving objects are extracted using background subtraction. Once moving objects are detected in the video sequence, Kalman filter approach is used to predict the location of these moving objects in subsequent frames.

Thus, the presented method uses Graph Cuts for segmentation to detect moving objects in video sequences and Kalman filter approach for tracking these detected objects.

4.3 FLOW DIAGRAM

The flow diagram for the presented approach is shown in figure 4.1.

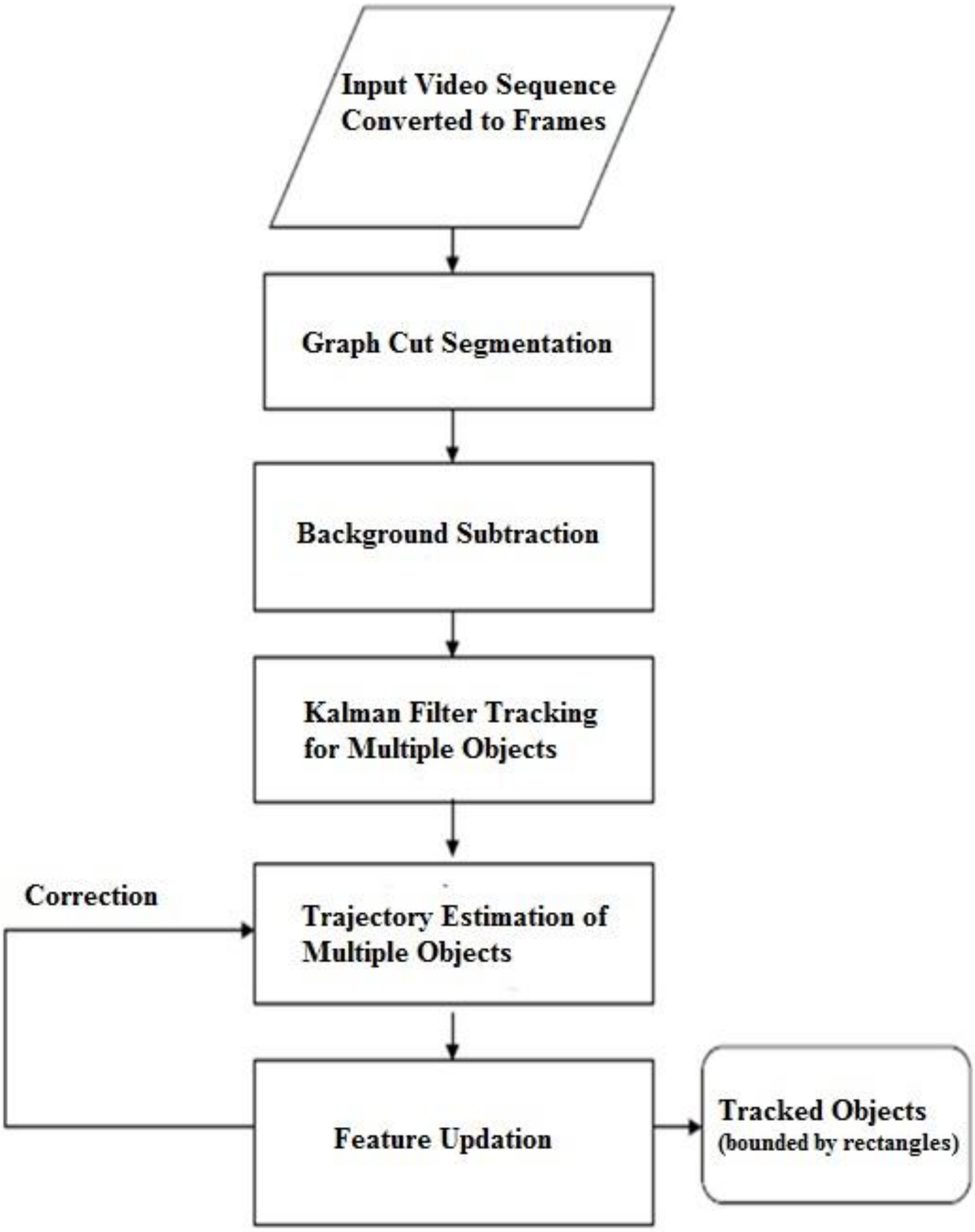

Fig 4.1 Flow chart of Presented Method

4.4 IMPLEMENTATION STEPS

The presented approach to track multiple objects can be outlined as below:

Step 1: Read the video sequence in which targets need to be tracked.

Step 2: Convert the input video into frames.

Step 3: Reserve some initial frames for background modeling.

Step 4: Apply Graph Cut Segmentation to get segmented video frames.

Step 5: Once the segmented frames are available, apply background subtraction to extract moving objects from those frames.

Step 6: Using Kalman filter approach, predict the location of moving objects in subsequent frames. This is the tracking step, which includes prediction, updation and correction.

Step 7: Tracked objects in the video sequence are shown by a bounded rectangular box.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 RESULTS

The presented method is tested on different videos, wherein tracking is done for both single object and multiple objects. The method has been implemented on MATLAB 2012B software.

5.1.1 Graph Cut Segmentation Results

The segmentation results obtained by applying graph cuts on various videos are shown in this section.

Segmented Frames of Test Video 1

The original dataset can be referred from Appendix I.

Fig 5.1 Segmented Video Frames of Test Video 1 (2 Frames)

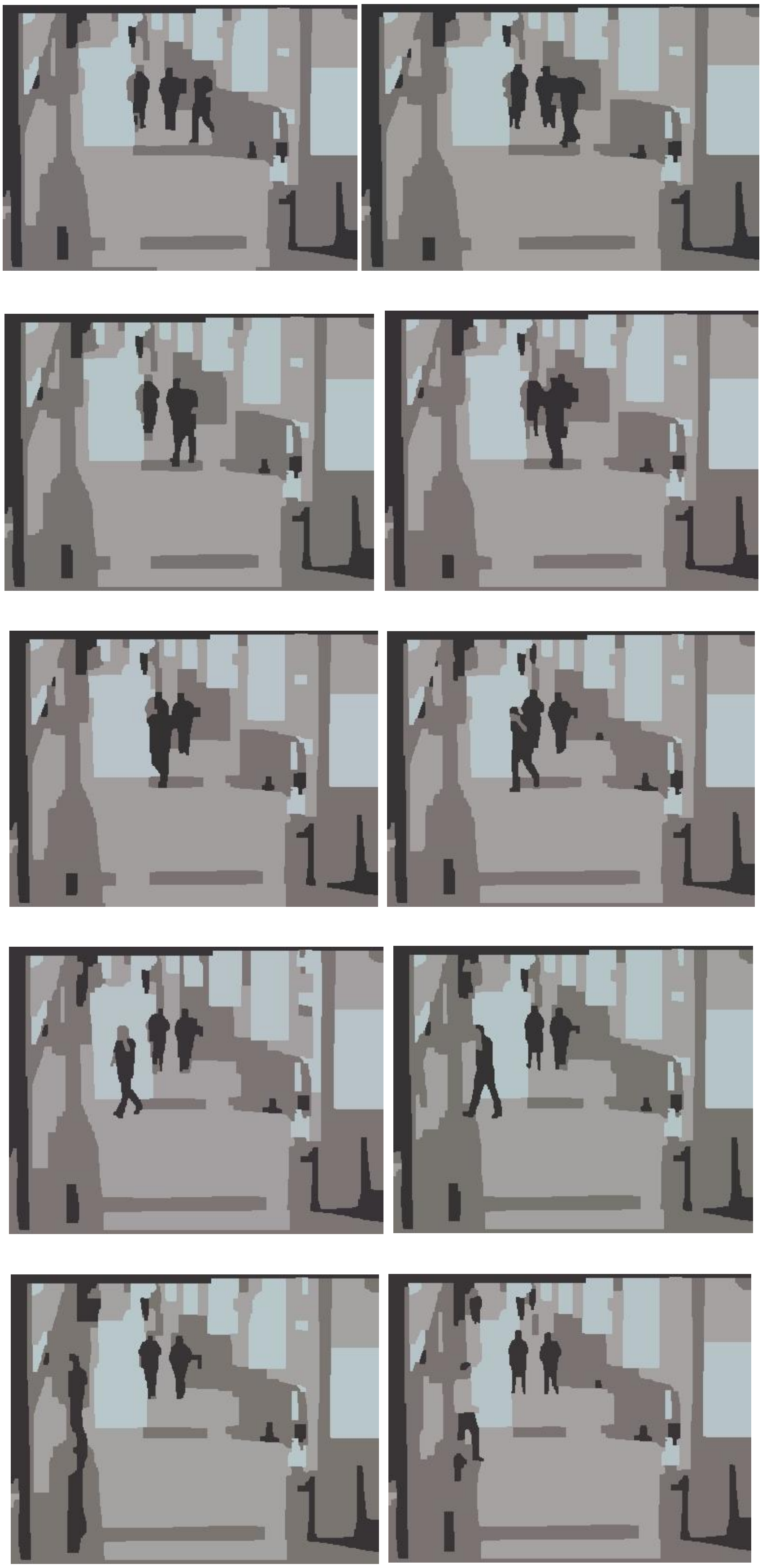

Fig 5.2 Segmented Video Frames of Test Video 1 (Next 10 Frames)

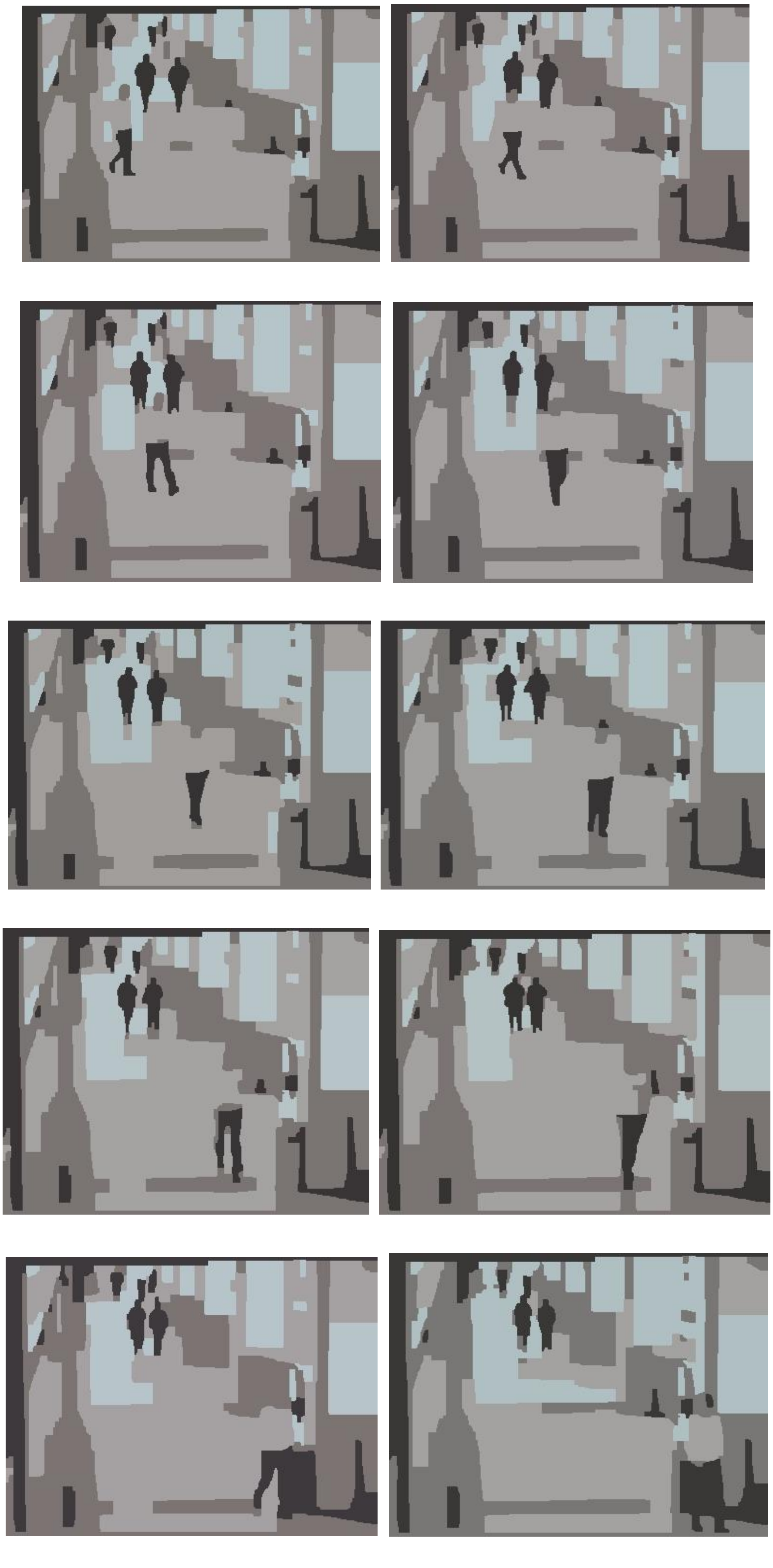

Fig 5.3 Segmented Video Frames of Test Video 1 (Next 10 Frames)

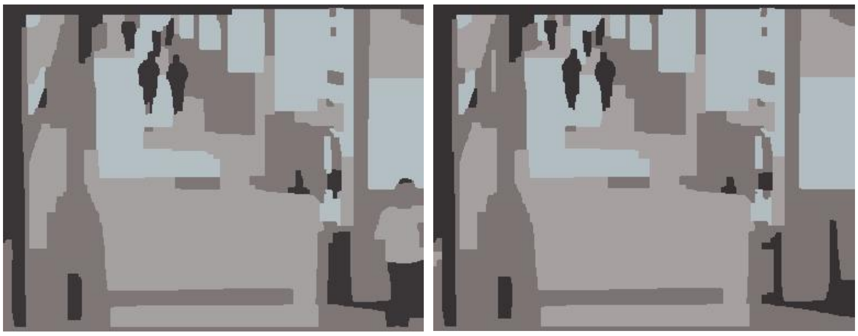

Fig 5.4 Segmented Video Frames of Test Video 2 (Last 2 Frames)

<u>Segmented Frames of Video 2</u>

The original dataset can be referred from Appendix I.

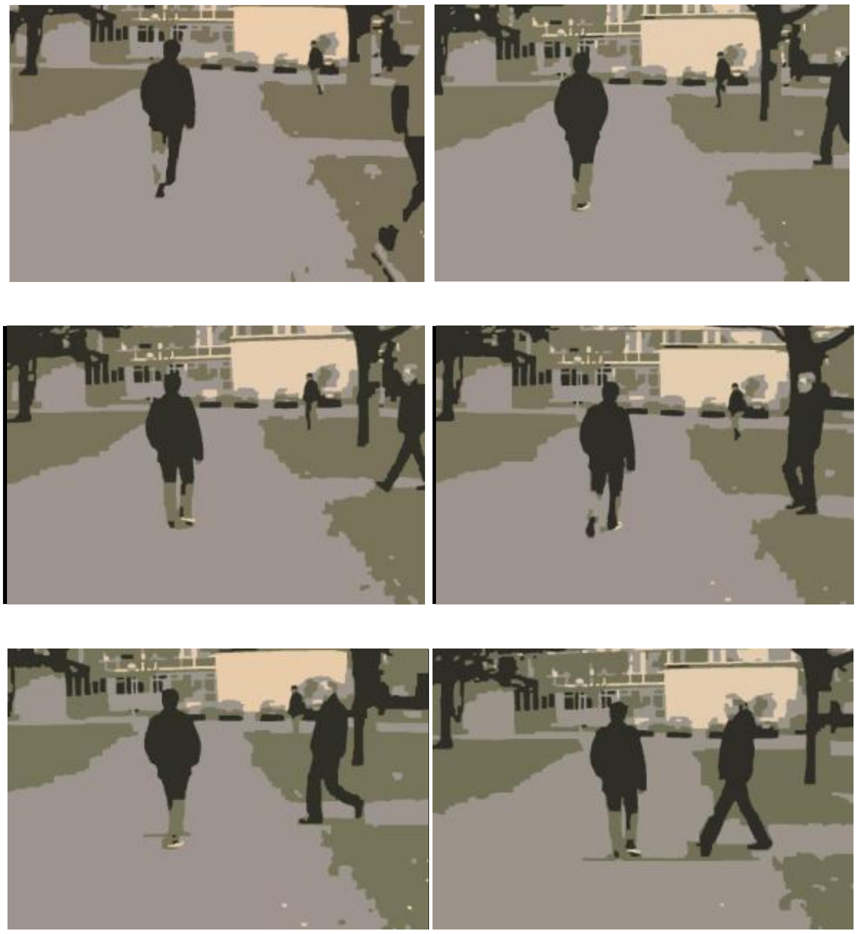

Fig 5.5 Segmented Video Frames of Test Video 2 (6 Frames)

Fig 5.6 Segmented Video Frames of Test Video 2 (Last 6 Frames)

<u>Segmented Frames of Video 3</u>

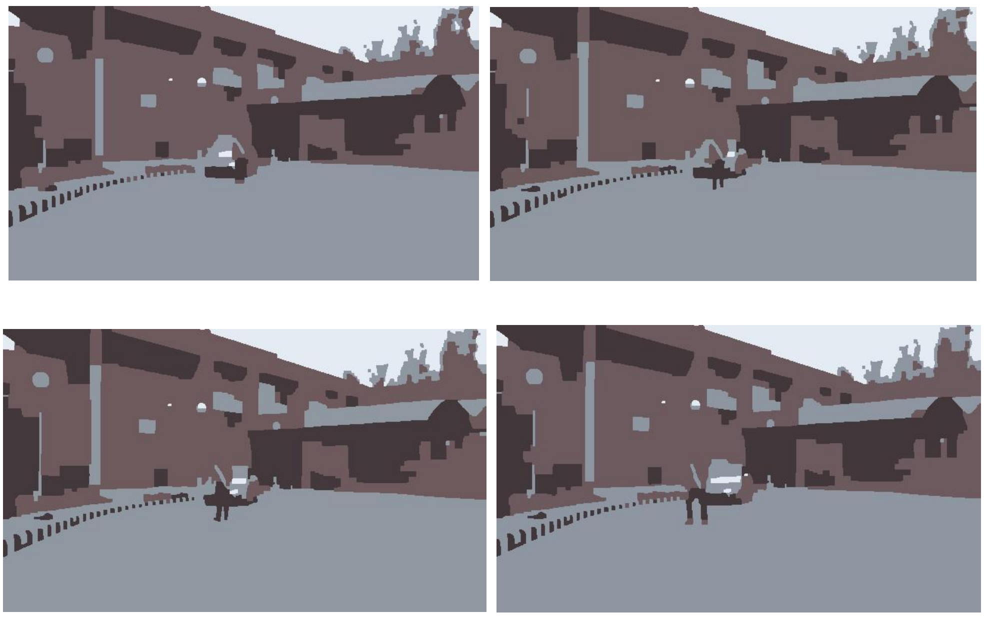

Fig 5.7 Segmented Video Frames of Test Video 3 (4 Frames)

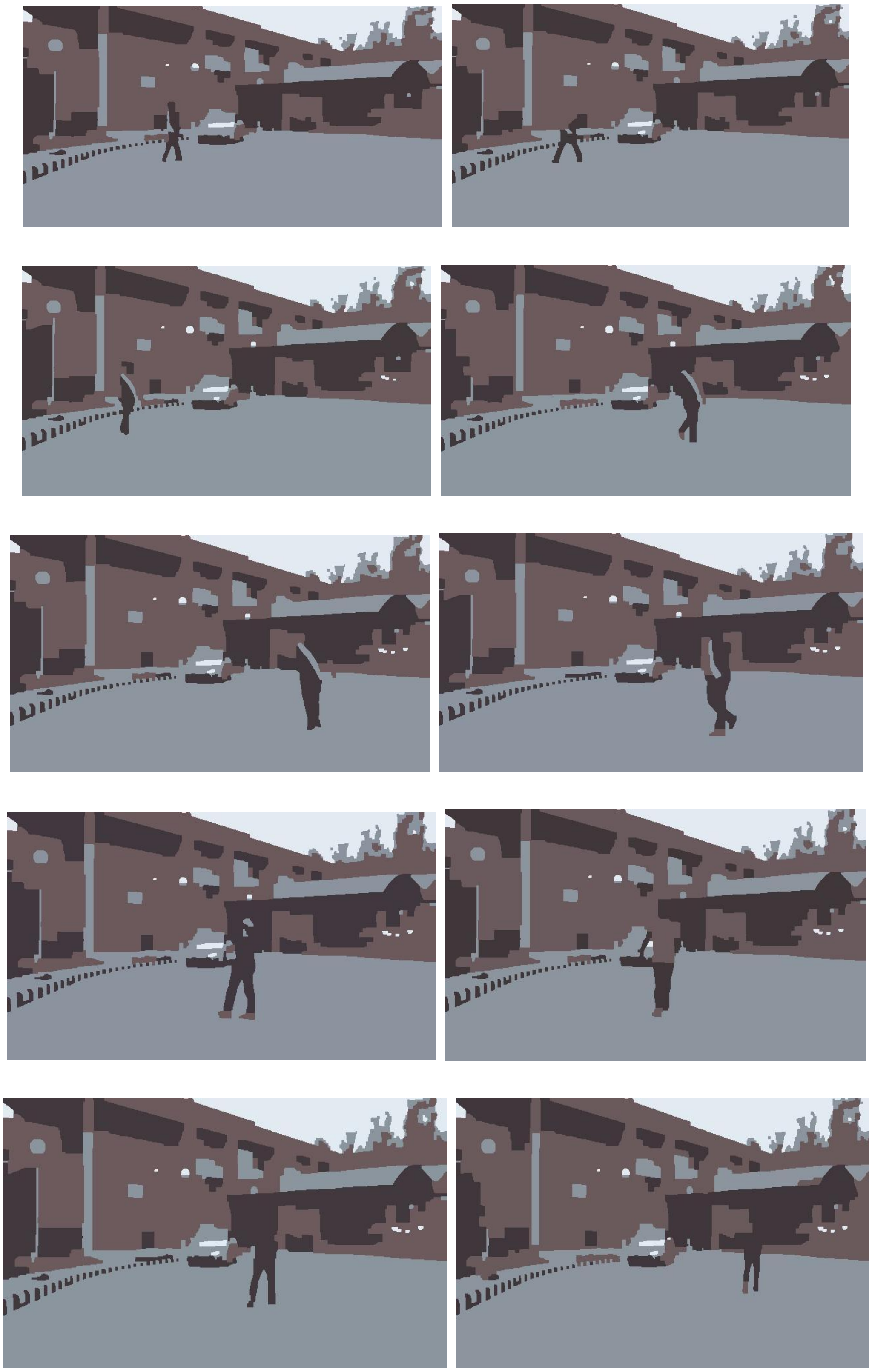

Fig 5.8 Segmented Video Frames of Test Video 3 (Next 10 Frames)

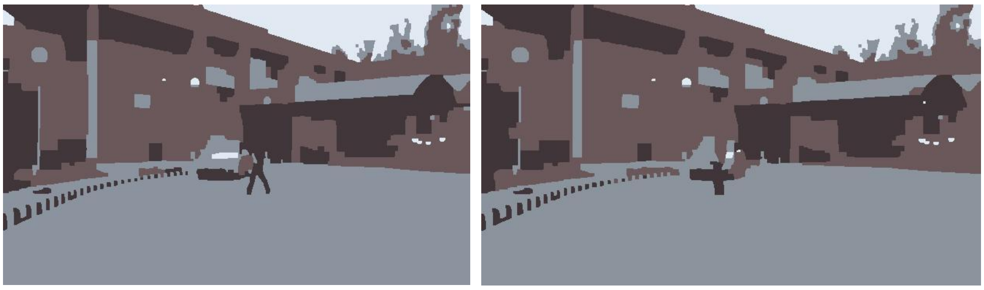

Fig 5.9 Segmented Video Frames of Test Video 3 (Last 2 Frames)

5.1.2 Tracking Results

The tracking results obtained by applying Kalman filter on the segmented video frames are shown in this section.

<u>Test Video 1</u>

Fig 5.10 Tracking Results on Test Video 1 (4 Frames)

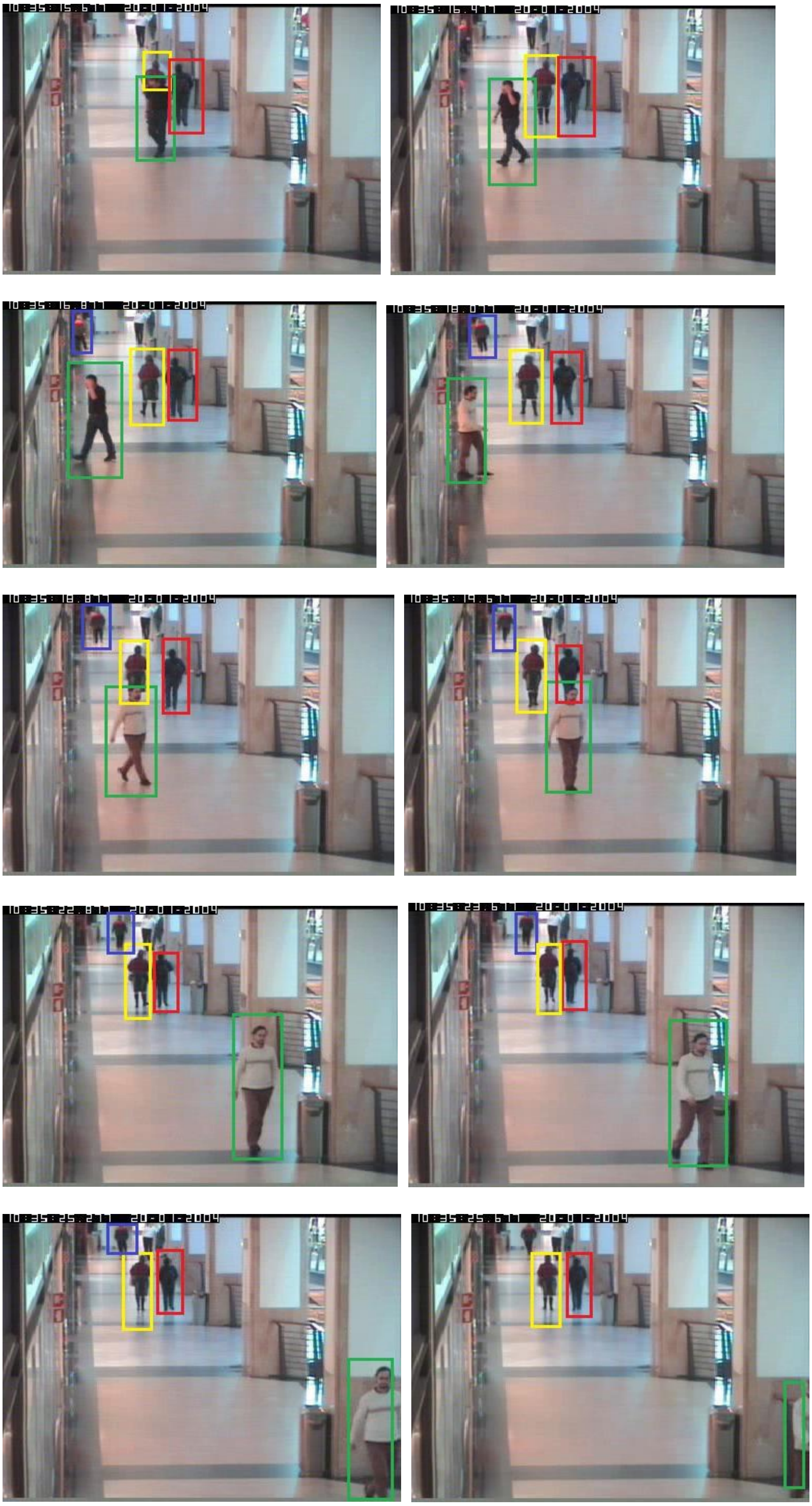

Fig 5.11 Tracking Results on Test Video 1 (Next 10 Frames)

Test Video 2

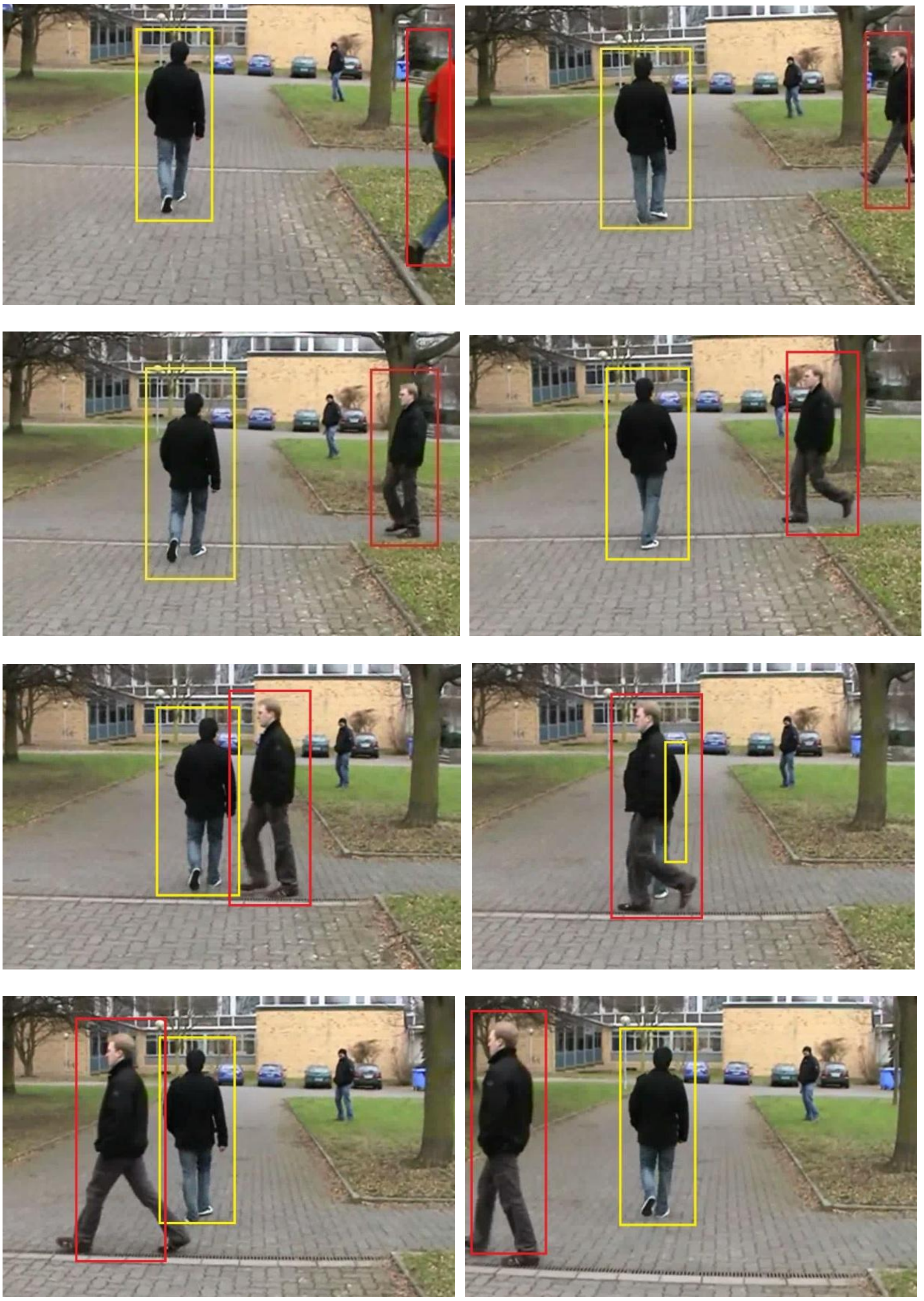

Fig 5.12 Tracking Results on Test Video 2

Test Video 2

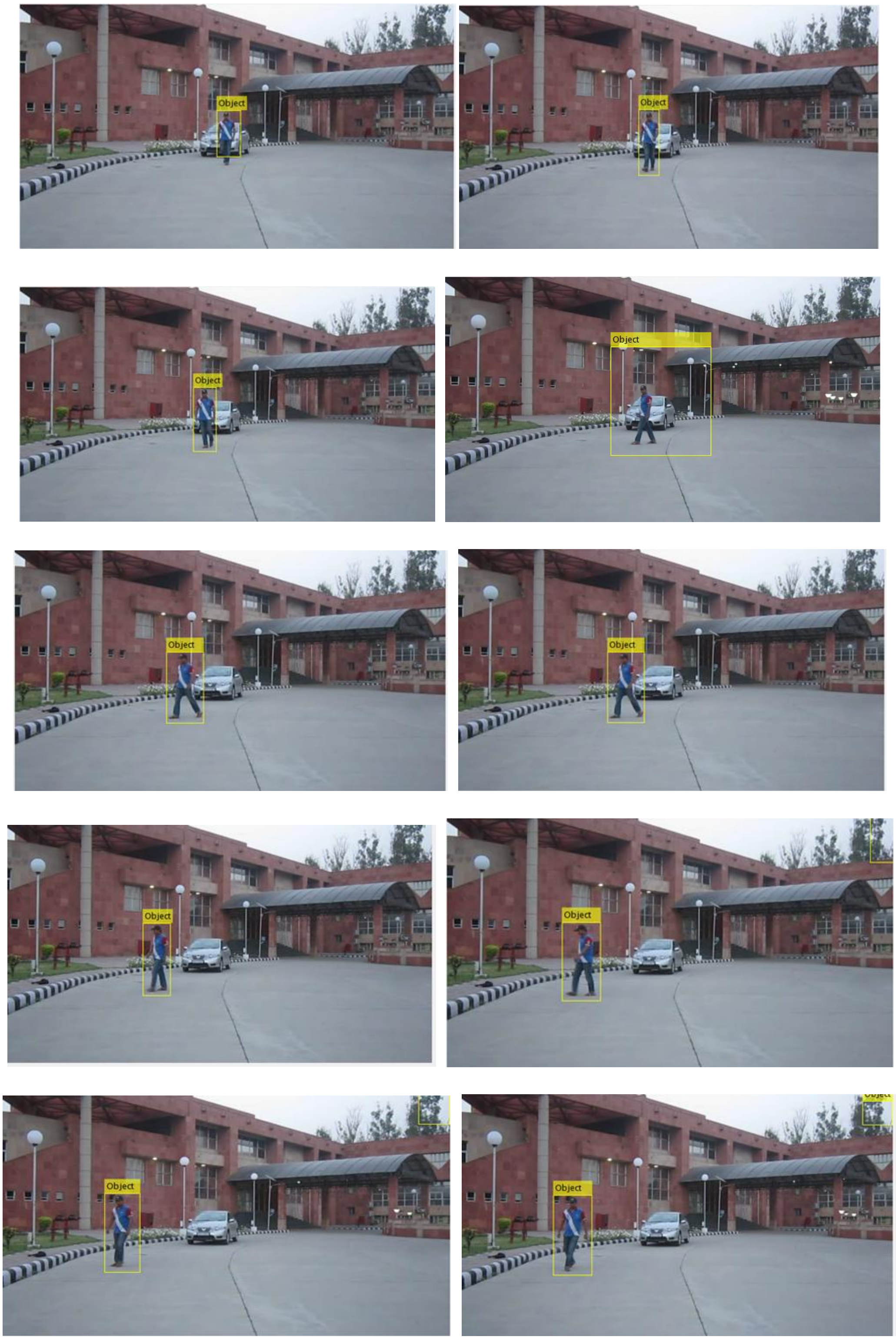

Fig 5.13 Tracking Results on Test Video 3 (10 Frames)

Fig 5.14 Tracking Results on Test Video 3 (Last 10 Frames)

5.2 DISCUSSION

From the results shown on various videos in the previous section, it is evident that the presented method works well for both single object tracking and multiple object tracking. The complexity of the presented method is improved due to the use of graph cuts, which accelerates the process of segmentation. The method is versatile, robust and flexible in both cases, that is, single object tracking and multiple object tracking.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

The graph cut based tracking method presented in this thesis worked well on the video sequences it was tested on. From the results shown in the previous section, following conclusions about the approach developed in this thesis can be drawn.

1. The presented method is shown to work well for tracking of both single object and multiple objects.
2. The efficiency of the method is verified by testing it on the standard database and our own database.
3. The method detects the object before tracking. It is a "**detect before track**" method.
4. It is an automated object detection and tracking method.

5.2 SCOPE OF FUTURE WORK

The presented method can be used in various applications. These may include the following.

1. **Human–computer Interaction**, comprises the planning, study, and design of the interaction between users (people) and computers.
2. **Anomaly detection** or outlier detection is a process to identify patterns in an already known data set that follow a model that is different from a normal behaviour set by the user. These identified patterns are known as outliers or anomalies.

3. A real time application of the method is **traffic surveillance system**. This method can be used for tracking of vehicles passing on the road in a robust and fast manner.
4. To count the number of objects in a video, multiple object tracking can be used.
5. The steering system of a **robot navigation system** has to detect various obstacles in the path of the robot so as to prevent it from colliding. A real time object tracking system is required when the obstacles are not stationary.

Future research in this field can be devoted to the use of Particle Filter for tracking rather than Kalman Filter. Tracking objects in the presence of shadows is also one area where future work can be carried out using the presented method.

REFERENCES

[1] D. Greig, B. Porteous, and A. Seheult, “Exact Maximum A Posteriori Estimation for Binary Images,” *J. Royal Statistical Soc.*, Series B, vol. 51, no. 2, pp. 271-279, 1989.

[2] Y. Boykov and V. Kolmogorov, “An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124-1137, Sept. 2004.

[3] V. Kolmogorov and R. Zabih, “What Energy Functions Can Be Minimized via Graph Cuts?” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147-159, Feb. 2004.

[4] Y. Boykov and M.P. Jolly, “Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in n-d Images,” *Proc. IEEE Int’l Conf. Computer Vision*, vol. 1, pp. 105-112, 2001.

[5] Y. Boykov and G. Funka Lea, “Graph Cuts and N-D Image Segmentation,” *International Journal of Computer Vision* 70(2), 109–131, 2006.

[6] A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov, “Bilayer Segmentation of Live Video,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 53-60, 2006.

[7] H. Ishikawa and D. Geiger, “Segmentation by Grouping Junctions,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 125-131, 1998.

[8] Nicolas Papadakis and Aure´ lie Bugeau, ”Tracking with Occlusions via Graph Cuts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 144-157, Jan. 2011,

[9] A. Bugeau and P. Pe´rez, “Track and Cut: Simultaneous Tracking and Segmentation of Multiple Objects with Graph Cuts,” *Proc. Int’l Conf. Computer Vision Theory and Applications*, vol. 2, pp. 447- 454, 2008.

[10] A. Bugeau and P. Pe´rez, “Track and Cut: Simultaneous Tracking and Segmentation of Multiple Objects with Graph Cuts,” *EURASIP J. Image and Video Processing*, vol. 2008, pp. 1-14, 2008.

[11] Jiyan Pan, BoHu, and Jian Qiu Zhang,”Robust and Accurate Object Tracking Under Various Types of Occlusions,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, no.2, pp. 223-236, Feb. 2008.

[12] J. Malcolm, Y. Rathi, and A. Tannenbaum, “Multi-Object Tracking through Clutter Using Graph Cuts,” *Proc. IEEE Int’l Conf. Computer Vision*, 2007.

[13] J. Malcolm, Y. Rathi, and A. Tannenbaum, “Tracking through Clutter Using Graph Cuts,” Proc. *British Machine Vision Conf.*, 2007.

[14] Jianbo Shi and Jitendra Malik, “Normalized Cuts and Image Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-905, Aug. 2000.

[15] J. Shi and J. Malik, [a]Normalized Cuts and Image Segmentation,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 731-737, 1997.

[16] J. Shi and J. Malik, [a]Motion Segmentation and Tracking Using Normalized Cuts,” *Proc. Int'l Conf. Computer Vision*, pp. 1,154- 1,160, 1998.

[17] H. T. Nguyen, M.Worring, and R. v. d. Boomgaard, “Occlusion robust adaptive template tracking,” in *Proc. Int. Conf. Comput. Vis.*, vol. 1, pp. 678–683, July 2001.

[18] H.T. Nguyen and A.W.M. Smeulders, “Fast Occluded Object Tracking by a Robust Appearance Filter,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1099-1104, Aug. 2004.

[19] K. Hariharakrishnan and D. Schonfeld, “Fast object tracking using adaptive block matching,” *IEEE Trans. Multimedia*, vol. 7, pp. 853–859, Oct. 2005.

[20] Z. Jia, A. Balasuriya, and S. Challa, “Target tracking with Bayesian fusion based template matching,” in *Proc. Int. Conf. Image Process.*, vol. 2, pp. II–826, Sep. 2005.

[21] F. Yang, H. Lu, W. Zhang and G. Yang, “Visual Tracking via Bag of Features,” *Image Processing IET*, vol. 6, no. 2, pp. 115-128, March 2012

[22] G.Di Caterina and J.J. Soraghan, “Robust complete occlusion handling in adaptive template matching target tracking,” *Electronic Letters*, vol. 48, no. 14, pp. 831-832, July 2012.

[23] Bing-Fei Wu, Chih-Chung Kao, Cheng-Lung Jen, Yen-Feng Li, Ying-Han Chen, and Jhy-Hong Juang, “A Relative-Discriminative-Histogram-of-Oriented-Gradients-Based Particle Filter Approach to Vehicle Occlusion Handling and Tracking,” *IEEE Transactions on Industrial Electronics*, vol.. 61, no. 8, pp. 4228-4237, Aug. 2014.

[24] Mohammad Azari , Ahmad Seyfi and Amir Hossein Rezaie, “Real Time Multiple Object Tracking and Occlusion Reasoning Using Adaptive Kalman Filters,” *Proc, IEEE Machine Vision and Image Processing*, Nov. 2011.

[25] A.D. Jepson, D.J. Fleet, and T.F. El-Maraghi, “Robust Online Appearance Models for Visual Tracking,” *Proc. Computer Vision and Pattern Recognition*, 2001.

[26] A. Senior *et al.*, “Appearance models for occlusion handling,” *J. Image Vis. Comput.*, vol. 24, no. 11, pp. 1233–1243, 2006.

[27] A. Yilmaz, O. Javed, and M. Shah, “Object Tracking: A Survey,” *ACM Computing Surveys*, vol. 38, no. 4, 2006.

APPENDIX A

Original Video Frames of Video 1

Fig A.1 Original Video Frames of Test Video 1

Original Video Frames of Video 2

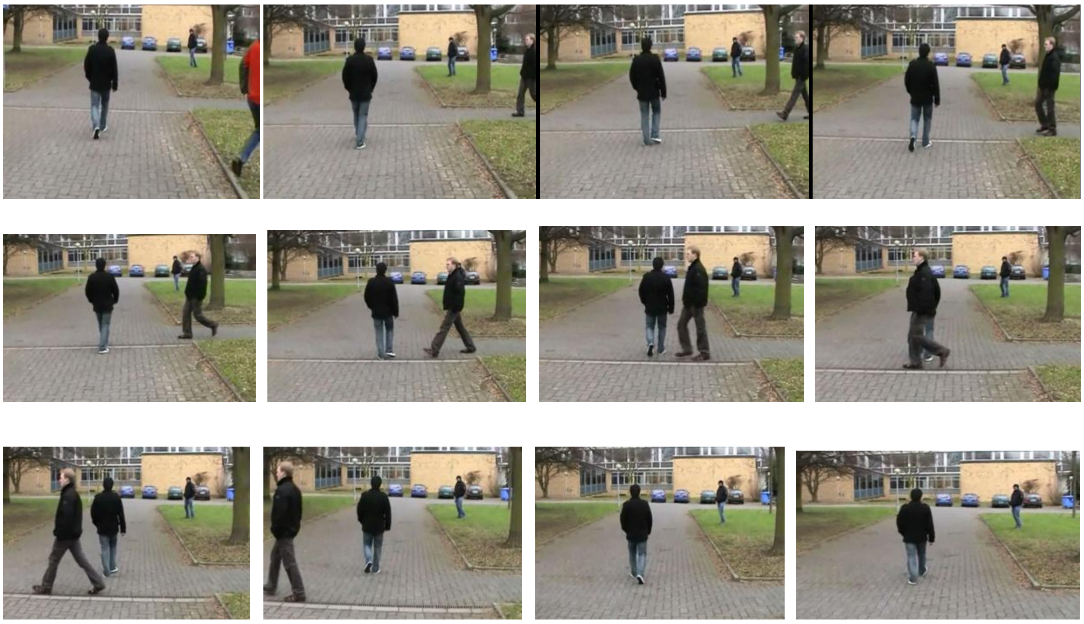

Fig A.2 Original Video Frames of Test Video 2

Original Video Frames of Video 3

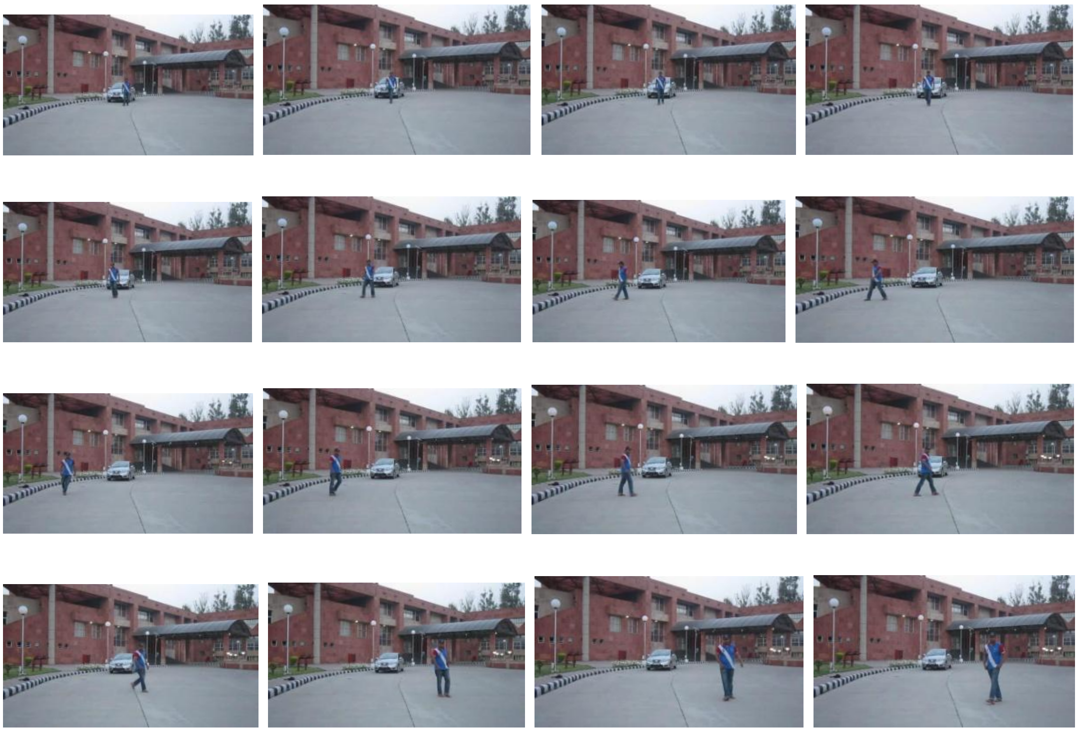

Fig A.3 Original Video Frames of Test Video 3 (16 Frames)

Fig A.4 Original Video Frames of Test Video 3 (Last 18 Frames)