Object detection and tracking in video image

Rajkamal kishor Gupta

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela – 769 008, India

Object detection and tracking in video image

Dissertation submitted in

May 2014

to the department of

Computer Science and Engineering

of

National Institute of Technology Rourkela

in partial fulfillment of the requirements

for the degree of

Master of Technology

by

Rajkamal Kishor Gupta

(Roll 212CS1091)

under the supervision of

Prof. Ramesh Kumar Mohapatra

Department of Computer Science and Engineering

National Institute of Technology Rourkela

Rourkela – 769 008, India

dedicated to my Parents, Brother and Sister...

Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela-769 008, India. www.nitrkl.ac.in

Ramesh Kumar Mohapatra
Asst. Professor

Date:

Certificate

This is to certify that the work in the thesis entitled *Object detection and tracking in video image* by *Rajkamal Kishor Gupta*, bearing roll number 212CS1091, is a record of research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of *Master of Technology* in *Computer Science and Engineering*. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Ramesh Kumar Mohapatra

DECLARATION

I, Rajkamal Kishor Gupta (Roll No. 212CS1091) understand that plagiarism is defined as any one or the combination of the following

1. Uncredited verbatim copying of individual sentences, paragraphs or illustrations (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Uncredited improper paraphrasing of pages or paragraphs (changing a few words or phrases, or rearranging the original sentence order).

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did or wrote what.

I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the thesis may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Date:

Rajkamal Kishor Gupta
Roll: 212CS1091
M. Tech in Department of Computer Sc. & Engg.
NIT Rourkela

Acknowledgement

This dissertation, though an individual work, has benefited in various ways from several people. Whilst it would be simple to name them all, it would not be easy to thank them enough.

The enthusiastic guidance and support of *Prof. Ramesh Kumar Mohapatra* inspired me to stretch beyond my limits. His profound insight has guided my thinking to improve the final product. My solemnest gratefulness to him.

I am also grateful to *Prof. Banshidhar Majhi* for his ceaseless support throughout my research work. My sincere thanks to *Prof. S. K. Jena* for his continuous encouragement and invaluable advice.

It is indeed a privilege to be associated with people like *Prof. S. K. Rath*, *Prof. Rameswar Baliarsigh*, *Prof. D. P. Mohapatra*, *Prof. A. K. Turuk*, *Prof. S. Chinara*, *Prof. Pankaj Sa* and *Prof. B. D. Sahoo*. They have made available their support in a number of ways.

Many thanks to my comrades and fellow research colleagues. It gives me a sense of happiness to be with you all. Special thanks to *Rajesh, Anshuman, Karthikeyan, Anoop, Priyesh and Manish* whose support gave a new breath to my research.

Finally, my heartfelt thanks for her unconditional love and support. Words fail me to express my gratitude to my beloved parents, who sacrificed their comfort for my betterment.

Rajkamal Kishor Gupta

Abstract

In recent days, capturing images with high quality and good size is so easy because of rapid improvement in quality of capturing device with less costly but superior technology. Videos are a collection of sequential images with a constant time interval. So video can provide more information about our object when scenarios are changing with respect to time. Therefore, manually handling videos are quite impossible. So we need an automated devise to process these videos. In this thesis one such attempt has been made to track objects in videos. Many algorithms and technology have been developed to automate monitoring the object in a video file.

Object detection and tracking is a one of the challenging task in computer vision. Mainly there are three basic steps in video analysis: Detection of objects of interest from moving objects, Tracking of that interested objects in consecutive frames, and Analysis of object tracks to understand their behavior.

Simple object detection compares a static background frame at the pixel level with the current frame of video. The existing method in this domain first tries to detect the interest object in video frames. One of the main difficulties in object tracking among many others is to choose suitable features and models for recognizing and tracking the interested object from a video. Some common choice to choose suitable feature to categories, visual objects are intensity, shape, color and feature points. In this thesis, we studied about mean shift tracking based on the color pdf, optical flow tracking based on the intensity and motion; SIFT tracking based on scale invariant local feature points. Preliminary results from experiments have shown that the adopted method is able to track targets with translation, rotation, partial occlusion and deformation.

***Keywords*:** Object detection, Frame difference, Vision and scene understanding, Background subtraction, Scale invariant feature transform (SIFT).

Contents

List of Figures

Chapter 1

Introduction

Object detection and tracking is an important challenging task within the area in Computer Vision that try to detect, recognize and track objects over a sequence of images called video. It helps to understand, describe object behavior instead of monitoring computer by human operators. It aims to locating moving objects in a video file or surveillance camera. Object tracking is the process of locating an object or multiple objects using a single camera, multiple cameras or given video file. Invention of high quality of the imaging sensor, quality of the image and resolution of the image are improved, and the exponential increment in computation power is required to be created of new good algorithm and its application using object tracking.

In Object Detection and Tracking we have to detect the target object and track that object in consecutive frames of a video file.

1.1 Challenges of Object Detection and Tracking

Object tracking fundamentally entails estimating the location of a particular region in successive frames in a video sequence. Properly detecting objects can be a particularly challenging task, especially since objects can have rather complicated structures and may change in shape, size, location and orientation over subsequent video frames. Various algorithms and schemes have been introduced in the few

decades, that can track objects in a particular video sequence, and each algorithm has their own advantages and drawbacks. Any object tracking algorithm will contain errors which will eventually cause a drift from the object of interest. The better algorithms should be able to minimize this drift such that the tracker is accurate over the time frame of the application.

In object tracking the important challenge that has to consider while the operating a video tracker are when the background is appear which is similar to interested object or another object which are present in the scene. This phenomena is known as *clutter*.

The other challenges except from cluttering may difficulty to detect interested object by the appearance of the that object itself in the frame plane due to factors which are described as follows:

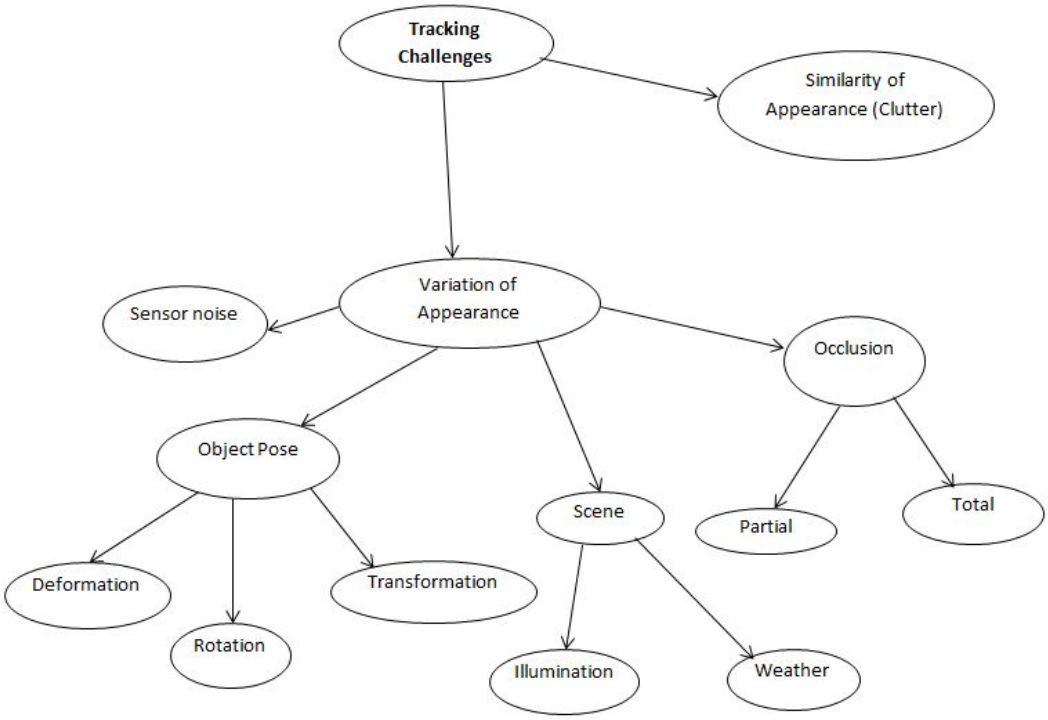

Figure 1.1: Main challenges in video Tracking

- *Object poses in the video frame :* In a video file, since the object is moving so the appearance of an interested object may vary its projection on a video

frame plane.

- *Ambient illumination :* In a video, it is possible to change in intensity, direction and color of ambient light in appearance of interested objects in a video frame plane.
- *Noise :* In the acquisitions process of video, it may possible to introduce a certain amount of noise in the image or video signal. The amount of noise depends upon sensor qualities which are used in acquitting the video.
- *Occlusions :* In a video file, moving object may fall behind some other object which are present in the current scene. In that case tracker may not observe the interested object. This is known as occlusion.

1.2 Object Detection and Tracking Pipeline

To overcome the different challenges issue as discussed in previous section there are following main component of object detection and tracking

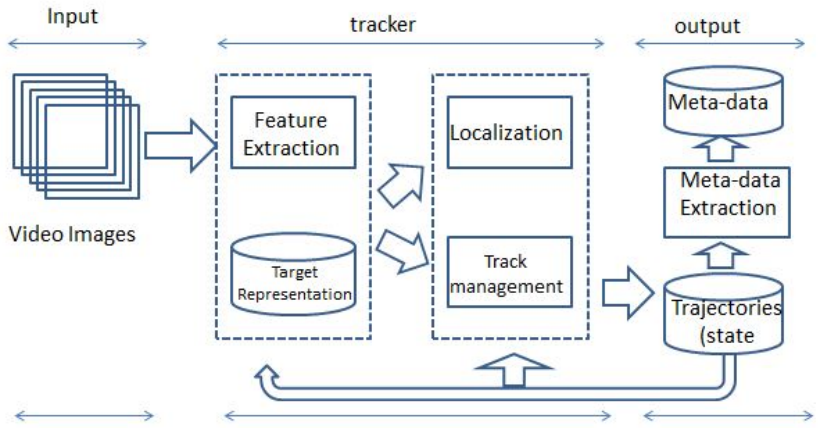

Figure 1.2: Basic component of object tracking algorithm

1.2.1 Feature extraction

Any object tracking algorithm can be analyzed by the quality of information that can be extracted from video frames or an image. To get more exploit information from image, we use image formation technique to extract feature which are more important, significant to identify interested object uniquely without any disambiguation.

- From the image background in the scene and
- From many another objects which are present in the scene

For any tracking algorithm extracting feature is the important step which is allowing us to highlight the information of the interested object from the video frames or target image plane. Extracted feature can be of three types:

- *Low level extraction*, e.g., motion, color, gradient
- *Mid level extraction*, e.g., edge, corner, interest point, region
- *High level extraction*, e.g., centroid, area, orientation, whole object

1.2.2 Target representation

The model that can be used by any tracking algorithm to represent the interested object is known as target representation. That model includes the information of interested object about the shape, size and appearance in an image. The model depends on the interested object and tracking algorithm that are used. There are different ways to model an interested object:-

- It may define priori of interested object
- It may snapshot of interested object
- It may be decided by training sample

There are two ways of target representation

- *Shape representation*, e.g., centroid, rectangle, ellipse, rigid model, contours or point distribution model
- *Appearance representation*, e.g., template, histogram

1.2.3 Localization

In localization, we describe how to localize an interested object over time, depending on the initial position. After initialization the localization step of a video tracker recursively estimates the stake X_k given the feature extracted from the video frames and the previous state estimates $X_1 : X_k$. We can classify methods into two major classes

- Single Hypothesis Localization (SHL)
- Multiple Hypothesis Localization (MHL)

1.2.4 Track management

The tracking algorithms presented generally rely on the estimate of the interested object position in the video frame plane. In a specific application, it is an operational condition which is acceptable while a developing a tracker where it can rely on initialization by user. In the real time varying application of a number of interested objects, the tracker needs to use automated initialization and automated termination capability.

1.2.5 Trajectory

A trajectory (state) is the path that a moving object follows through space as a function of time. A trajectory can be described mathematically either by the geometry of the path or as the position of the object over time. It stores the actual path of the object of interest, i.e. information about the target in consecutive frames. We will get the all information about a target object that in which direction it moves and what is the speed of target.

1.3 Motivation

The rapid improvement in technology makes video acquisition sensor or devices better in compatible cost. This is the cause of increasing the applications in different areas that can more effectively utilize that digital video. Digital videos are a collection of sequential images with a constant time interval. So there is more information is present in the video about the object and background are changing with respect to time. After studying the literature, it is seen that detecting and tracking of objects in a particular video sequence or any surveillance camera is a really challenging task in computer vision application. Video processing is really time consuming due to a huge number of data is present in video sequence.

The area of video tracking is currently immense interest due to its implication in video surveillance, security, medical equipments, robotic systems. Video tracking offers a context for extraction of significant information such as scene motion, background subtraction, object classification, interaction of object with background and other objects from a scene, human identification, behavior of human with object and background, etc. Therefore it is seen that there is a wide range of research possibilities are open in relation to video tracking.

1.4 Problem Statement

In this thesis our aim is to improve the performance of object detection and tracking by contributing originally to two components (a) motion segmentation (b) object tracking.

Automated tracking of objects can be used by many interesting applications. An accurate and efficient tracking capability at the heart of such a system is essential for building higher level vision-based intelligence. Tracking is not a trivial task given the non-deterministic nature of the subjects, their motion, and the image capture process itself. The objective of video tracking is to associate target objects in consecutive video frames. We have to detect and track the object moving independently to the background. In this there are four situations to be

considered in the account:

- Single camera and single object,
- Single camera and multiple objects,
- Multiple cameras and single object,
- Multiple cameras and multiple objects.

From the previous section it is found that there are many challenges in detecting of an object and tracking of objects and also recognition for fixed camera network. 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, and 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 from a stationary camera. Therefore the main objectives are:

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

1.5 Thesis Layout

The thesis is organized in chapter as follows:

Chapter 2: Object Detection and Tacking In this chapter, we introduce the background of object detection and tracking, with object detection, object representation and object tracking. In this chapter we also discuss the literature surveys that have been done during the research work which provides a detailed survey of the literature related to motion detection and object tracking.

Chapter 3: Feature Extraction Method In this chapter, we discuss about some features of objects that can be extracted by feature extraction methods like

Scale Invariant Feature Transform (SIFT), Kanade Lucas Tomasi (KLT) feature tracker and Mean SIFT.

Chapter 4: Result In this chapter, we discuss about experimental result on Object Detection and Tracking in video image using SIFT algorithm with static camera.

Chapter 5: Conclusion and Future Work In this chapter, we conclude the work we have done and proposing the work that can be done in future.

Chapter 2

Object detection and tracking

Detection of object and tracking of that object is an important task in the area of computer vision application. In object detection we locate or detect interested object in consecutive frames of a video file. Tracking is a process to locate moving interested object or multiple objects in a video file or camera with respect to time. Technically, in object tracking we estimate or define the trajectory or path of an interested object in the frame plane as it moving around the image plane. Because of technology increasing in computational power, availability of good quality and low cost video camera and the need of automated video system people are sowing the more interest in object tracking algorithm. In a video analysis, there are three basic or main steps are there:

- Detection of interested object from moving objects,
- Tracking of that interested objects in consecutive frames, and
- Analysis of trajectory of object to understand the behavior of interested object.

2.1 Object Representation

Object tracking is a video processing application with a wide number of applications. Applications may include tracking particular people in a video for security reasons

for tracking planetary objects from satellite data for astronomical studies. An object of interest is defined on the basis of particular application which is present at hand. An object of interest may depend on the type of application. For example, in traffic surveillance application interested object may be human or car, whereas for satellite application interested object may be a planet or for gaming application it may be face of particular person.

Figure 2.1: Example of object of interest for video tracking: (left) face, (right) people

In an object tracking algorithm, an object of interest is defined on the basis of application which are present but it can be used for further analysis. From previous example, it is clear that we have to take such object as objects of interest which help to object tracking. For example, animal in the zoo, missile in war, people in the mall, building on satellite, etc. are the example of set of interested object which may be the most important to detect or track in a particular application. An interested object may be modeled by their appearance and shape. In object representation interested object can be modeled different way that can help to video tracker:

- Point: Interested object can be represented by using a point. In this we can represent our interested object by either single point (e.g., centroid) or

multiple points. We can use point representation of interested object in those tracking object application where target is present in smaller region or target is itself small.

- Primitive geometric shapes: Interested object can be represented by geometric shape. For example, by using circle, ellipse, rectangle etc.. Primitive geometric shape representation can be used for representing a rigid object; it can also useful for tracking simple non rigid object. These types of representation are modeled by projective transformation, affine or translation for object motion.

- Object silhouette and contour: Interested object can be represented by contour. The boundary of the object is defined as contour representation. Region surrounded by the contour is known as the silhouette of interested object. To represent the complex non-rigid shapes we generally use contour or silhouette representation.

- Articulated shape models: Interested object can be represented by articulated shape model. Composition of body part which are held with joint can be represented by this model. For example, the body of a person is an articulated object by hand, torso, feet, legs and head connected by joints. By using kinematic motion model we can establish relations between the parts of the body. Joint angle is an example of kinematic motion model. It can be represented by no of geometric shape (e.g., cylinder, ellipse) can be used.

- Skeletal models: Interested object can be represented by skeletal model. Object skeleton might be created by applying average hub change to the object outline or object silhouette. This model is generally utilized as a shape representation for perceiving objects. The skeleton representation could be utilized to model both enunciated and inflexible objects.

- Probability densities of object appearance: The provability density evaluations of the object appearance can either be parametric, for example, Gaussian and a mixture of Gaussians, or nonparametric, for example, Parzen windows and histograms. The likelihood densities of item appearance characteristics (color, surface) might be figured from the picture districts specified by the shape models (inside district of a circle or a form).

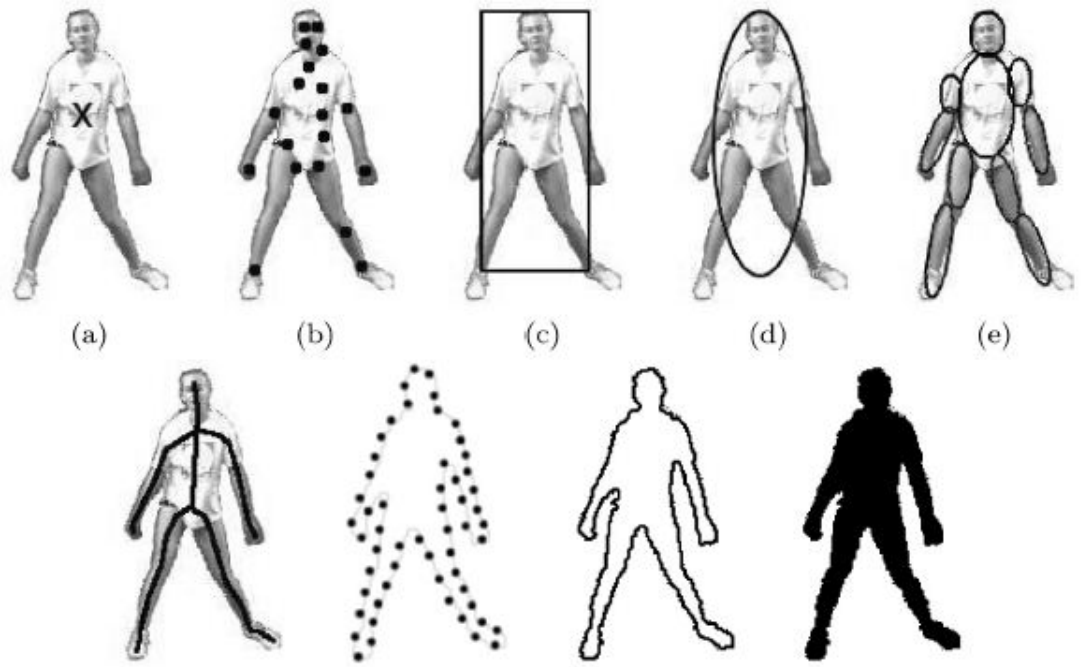

Figure 2.2: Representations of interested object (a) single point, (b) multiple points, (c) rectangular shape, (d) elliptical shape, (e) part-based multiple geometric shape, (f) skeleton of object, (g) contour of object, (h) control points on object contour, (i) silhouette of object.

- Templates: Interested object can be represented by templates model. Template model can be created by using the basic shape of geometry or object silhouettes. This is a better representation of an object because it can contain both appearance and special information about interested object. Since the template only encode the appearance of interested objects from an only one view.

Accordingly, it is suitable for only for tracking those interested object which are not changing the pose during the tracking.

- Active appearance models: Interested object can be also represented by active appearance models. These types of models are produced by all the while demonstrating the object appearance and shape. When all is said in done, the object shape is characterized by a situated of milestones. Like as contour based representation, the historic point can live on the boundary of object or, alternatively, they can dwell inside the object contour. For every historic point, an appearance vector is stored in the form of gradient magnitude, texture or color. Appearance models oblige a preparation stage where both its associated appearance and the shape are gained from a situation of example utilized. For example, Principal component analysis (PCA).

- Multiview appearance models: Interested object can be represented by multi-view appearance models. These models encode diverse perspectives of an object. One methodology to speak to the diverse item perspectives is to create a subspace from the given perspectives. Subspace approaches, for instance, Principal Component Analysis (PCA) and Independent Component Analysis (ICA), have been utilized for both shape and appearance representation.

However, there is a solid relationship between the interested object representations and the tracking algorithm. Object representations are generally picked as indicated by the requisition space. For tracking interested object, which seem little in a frame, point representation is normally fitting. For the object whose shapes could be approximated by rectangles or circles, primitive geometric shape representations are more suitable. For a tracking interested object with complex shapes, for instance, people, a contour or a silhouette based representation is fitting.

2.2 Object Detection

Object detection technique is an important task in any tracking algorithm to detect the interested object in either each frame of video or from that frame where the object first show up on video. Then again some object detection system makes utilization of worldly information resister from the frame sequence to decrease the amount of false detection. For object detection, there are few regular object detection techniques depicted.

- *Point detectors* - One of the object detection technique is point detector. This detector are generally used for discover fascinating point from the video frame which have an expressive surface in their particular area. An alluring nature of an interesting point is its invariance to changes in enlightenment and camera perspective. In literature, regularly utilized interesting point detectors incorporate Harris detector, Moravec's detector, SIFT detector, KLT detector.

- *Background Subtraction* - Another object detection technique is background subtraction. Object detection could be achieved by building a representation of the scene called the establishment display and after that running across deviations from the model for every one approaching frame. Any basic change in a picture district from the establishment model means a moving object. The pixels constituting the territories encountering change are checked for further process of frame. This methodology is insinuated as the Background subtraction. There are diverse schedules for establishment subtraction as discussed in the outline are Hidden Markov models (HMM), Frame differencing Region-based (or) spatial information and Eigen space decomposition.

- *Segmentation* - Another object detection technique is segmentation. The goal of frame segmentation is to divide the picture into perceptually comparable areas. In every segmentation algorithm, it addresses two issues, the criteria for a great allotment and the strategy for attaining productive dividing. In the review of literature, it has been discussed different methods of segmentation

that are applicable to object detection. They are, Active contours, image segmentation using Graph-Cuts (Normalized cuts) and mean shift clustering.

Object identification could be performed by taking in distinctive object sees naturally from a set of cases by method for regulated taking in system.

2.3 Object Tracking

Object tracking decides the movement of the projection of one or more object in video frame plane. This movement is incited by the relative movement between the camera and the watched scene. It is truly characterized as, "Placing a moving object or different protests over a time of time utilizing camera" and actually as, "issue of assessing the trajectory or way of an object in the video frame plane as it moves around a scene". Object tracking could be connected in numerous regions like robotized observation, movement checking, human workstation connection and so forth. Challenges in the same region incorporate commotion in casings, complex item movement and shape, impediment, change in enlightenment and so on.

Systems for object tracking might be arranged into emulating four classifications as per the tool utilized throughout tracking.

- *Region-based methods :* These strategies give a productive approach to decipher and investigate movement in a frame sequence of video. A frame district might be characterized as a set of pixels having homogeneous attributes. It could be determined by image segmentation, which might be focused around different object characteristics like color, edges and so forth. Basically, a district would be the image range secured by the projection of the object of investment onto the frame plane. On the other hand, a locale could be the bouncing box of the anticipated object under examination.

- *Contour-based methods :* An option method for concocting an object tracking algorithm is by representation of object utilizing contour shape information

and tracking it time to time, hence recovering both its position and shape. Such a demonstrating technique is more entangled than displaying whole locales. Then again, contour based tracking are typically more hearty than region based object tracking algorithm, on the grounds that it could be adjusted to adapt to halfway impediments. Additionally the outline information is unfeeling to light varieties.

- *Feature point-based methods :* Feature point-based object tracking could be characterized as, the endeavor to recuperate the motion parameters of a characteristic point in a feature succession. All the more formally, let f = f0, f1, , fn means the N frames of a video file sequence and pi (xi, yi) , i = 0, 1,...,N indicate the positions of the same characteristic point in those frame. The current task is to focus a motion vector di (dx,i, dy,i) that best decides the position of the feature points in the following frame, mi+1 (xi+1, yi+1), that is: mi+1 = mi + di. The interested object to be tracked is generally characterized by the bouncing box or the curved structure of the tracked feature point.

- *Template-based methods :* Template-matching procedures are utilized by numerous scientists to perform object tracking. Template based tracking is nearly identified with region based tracking on the grounds that a template is basically a model of the picture area to be tracked. These routines include two steps for tracking; introduction step took after by matching step. In the first step template might be instated by different on-line and off-line strategies. Throughout matching, it includes the procedure of seeking the interested object to focus the image district that looks like the layout, taking into account a likeness or separation measure.

In present commitment article following is attained utilizing characteristic point-based technique.

2.4 Literature Survey

The research conducted so far for object detection and tracking objects in video surveillance system. Tracking is the process to locating the interested object within a sequence of frames, from its first appearance to its last. The type of object and its description within the system depends on the application. During the time that it is present in the scene it may be occluded by other objects of interest or fixed obstacles within the scene. A tracking system should be able to predict the position of any occluded objects.

In [1], the author suggests an algorithm to isolate the moving objects in video sequences and then presented a rule-based tracking algorithm. The preliminary experimental results demonstrate the effectiveness of the algorithm even in some complicated situations, such as new track, ceased track, track collision, etc. A tracking method without background extraction is discussed in [2]. Because while extracting background from video frame if there are small moving things in that frame they form a blob in thresholding which create confusion in case of tracking that blob as they arent of any use that can be reduced here. The author introduces a video tracking in computer vision, including design requirements and a review of techniques from simple window tracking to tracking complex, deformable objects by learning models of shape and dynamics in [3].

In [22], there are various studies identified with the MS-based CMS (or Camshift) trackers. In the investigation of Stern and Efros, they created a strategy that adaptively switches shade space demonstrates all around the transforming of a feature. Additionally, they proposed another execution measure for assessing following calculation. Their proposed technique is utilized to discover the ideal color space and shade appropriation models fusion in the configuration of versatile shade following frameworks. Their color exchanging strategy was performed

Inside the skeleton of the CAMShift tracking algorithm, they consolidated various methodology to develop an improved face tracking methodology. At every cycle of the CAMShift algorithm, given image is changed over into a likelihood image utilizing the model of shade dispersion of the skin color being tracked.

In the study of Li et al. [23], they proposed a novel methodology for global target following focused around MS strategy. The proposed technique speaks to the model and the applicant regarding background and shade weighted histogram, separately, which can get exact object estimate adaptively with low computational unpredictability. Likewise, they actualized the MS technique by means of a coarse-to-fine path for global greatest looking for. This system was termed as versatile pyramid MS, in light of the fact that it utilizes the pyramid examination procedure and can focus the pyramid level adaptively to diminishing the amount of iteration needed to attain merging. The trial consequences of the study of Li et al. [23] indicate that the proposed system can effectively adapt to distinctive circumstances, for example, camera movement, camera vibration, camera zoom and center, high velocity moving item following, halfway impediments, target scale varieties, and so forth. Yuan et al. [24] proposed another moving object tracking algorithm, which joins together enhanced nearby binary pattern texture surface and tone information to portray moving object and embraces the thought of CAMShift algorithm. With a specific end goal to diminish matching unpredictability on the reason of fulfilling the correctness, numerous sorts of neighborhood twofold example and tint are chopped down. As per Yuan et al. [24], the experiment demonstrate that the proposed algorithm can track adequately moving interested object, can fulfill continuous and has preferred execution over others. In the study of Mazinan and Amir-Latifi [25], an enhanced curved part capacity was proposed to defeat the fractional impediment. Hence, to enhance the MS calculation against the low immersion and additionally sudden light, changes are created out of movement data of the fancied succession. By utilizing both the color feature and the motion information all the while, the competence of the MS calculation was correspondingly expanded. In their study [25], by accepting a steady speed for the article, a hearty estimator, i.e., the Kalman channel was acknowledged to tackle the full impediment issue. As indicated by Mazinan and Amir-Latifi [25], the trial results checked that the proposed technique has an ideal execution progressively protest following, while the aftereffect of the first MS calculation may be unsatisfied.

In [26], the OF strategy is oftentimes utilized as a part of picture movement investigation. The reckoning OF from a picture succession gives exceptionally imperative data to movement investigation. This issue includes moving object detection and tracking, moving object segmentation, and movement distinguishment. In the study of Lai, another movement estimation calculation was displayed that it gives faultless OF processing under non uniform brilliance varieties.

Lipton et al. [27] proposed edge contrast that utilization of the pixel-wise contrasts between two casing pictures to concentrate the moving locales. In an alternate work, Stauffer & Grimson et al. [28] proposed a Gaussian mixture model focused around foundation model to distinguish the item. Liu et al. [29], proposed foundation subtraction to recognize moving districts in a picture by taking the contrast between present and reference foundation picture in a pixel-by-pixel. Collins et al. [30], created a half breed system that joins three-edge differencing with a versatile foundation subtraction model for their VSAM (Video Surveillance and Monitoring) undertaking. Desa & Salih et al [31], proposed a mixture of foundation subtraction and casing contrast that enhanced the past consequences of foundation subtraction and edge distinction. Cheng & Chen, 2006 proposed a color and a spatial characteristic of the item to recognize the track object. The spatial characteristic is concentrated from the jumping box of the article. In the interim, the color characteristics concentrated is mean and standard worth of each one article. Czyz et al., 2007 proposed the color conveyance of the object as perception model. The comparability of the object estimated by Bhattacharya distance. The low Bhattacharya distance relates to the high similitude.

Chapter 3

Feature Extraction Method

In multimedia technology visual object detection and tracking is the important task, especially in provisions, for example, remotely coordinating, reconnaissance and human workstation interface The trouble in visual object detection and tracking is to discover and channel a few Feature that are less delicate to image interpretation, scaling, pivot, brightening progressions, bending and in part impediment. The objective of interested object tracking is to focus the position of the object in frame persistently and dependably against element scenes. To attain this focus on, various exquisite routines have been secure

3.1 Feature Extraction

The execution of a feature tracker relies on upon the nature of the data we can extricate from the pictures. To see how to better adventure picture data, feature extraction is the one of the critical step in the object detection and tracking algorithm. It permits us to show information from image. There are fundamentally three approaches to concentrate characteristic from client

3.1.1 Low level extraction

Low level feature (e.g., color, gradient, motion) introduces a diagram of color representation, slope and movement computational techniques. The objective is

to see how to endeavor low level characteristic in the diverse phase of target representation and limitation. Portraying shade through characteristic that continue as before paying little respect to shifting circumstances included by picture handling is an imperative necessity for some feature following requisition. Changes in imaging condition are identified with the review course, the target surface introduction and brightening condition. This change present antiques, for example, shading shadows and highlights. Nearby force change convey vital data about the appearance of object of investment. These change happen inside the item itself and the limits between article and foundation.

3.1.2 Mid level extraction

Mapping the picture onto low level characteristic may not be satisfactory to attain an objective synopsis of picture substance, consequently diminishing the adequacy of a feature tracker. So we break down the feature utilizing subset of pixel that speak to important structure (e.g., edge, corner, interest point) or uniform locale where Whole pixel impart some regular properties. Most investment point identifier has a tendency to select exceedingly unique neighborhood example, for example, corner, conspicuous edge and region with segregate surface.

3.1.3 High level extraction

So as to characterize a interested object, one could be either aggregate mid level characteristic, for example, premium point and region or can recognize straightforwardly the item overall focused around its appearance. High level characteristic might be centroid, entire range or introduction of interested object. A verity of methodologies for detecting moving objects are focused around background subtraction is known as *background modeling.*

3.2 Scale Invariant Feature Transform

Scale Invariant Feature Transform (SIFT) is a methodology for identifying and concentrating local feature descriptors that are sensibly invariant to changes in enlightenment, scaling, pivot, image noise and little changes in perspective. This calculation is initially proposed by David Lowe in 1999, and afterward further created and moved forward.

SIFT characteristics have numerous preferences, for examples are follows:

- SIFT Features are natural feature of pictures. They are positively invariant to picture interpretation, scaling, revolution, brightening, perspective, commotion and so on.
- Great strength, rich in data, suitable for quick and precise matching in a mass of feature database.
- Richness. Heaps of SIFT feature will be investigated regardless of the possibility that there are just a couple of object.
- Moderately quick speed. The pace of SIFT even can fulfill ongoing process after the SIFT algorithm is advanced.
- Better expansibility. SIFT is extremely helpful to consolidate with other eigenvector, and create much valuable information.

3.2.1 Scale-space extrema detection

The primary phase of calculation inquiries over all scales and image areas. It is actualized productively by method for a difference of- Gaussian capacity to recognize potential investment point that are invariant to scale and orientation. Interest point for SIFT characteristics relate to neighborhood extrema of difference of- Gaussian channels at diverse scales. Given a Gaussian-blurred image described as the formula

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3.1}$$

where

$$G(x, y, \sigma) = \frac{1}{2\Pi\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}} \tag{3.2}$$

is a variable scale Gaussian, whose result of convolving an image with a difference-of-Gaussian filter is given by

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{3.3}$$

Which is just be different from the Gaussian-blurred images at scales σ and kσ

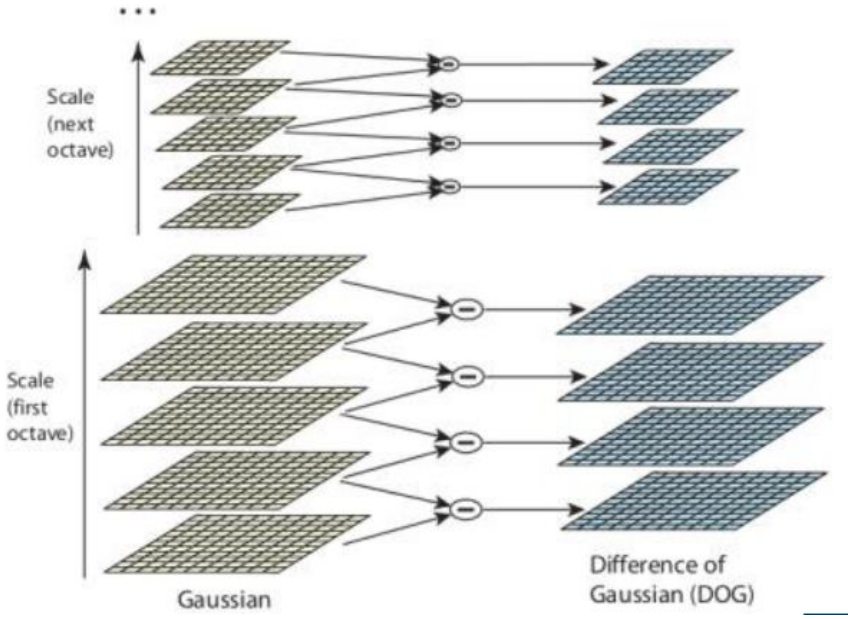

Figure 3.1: Frame at different scales, and the calculation of the difference-of-Gaussian images

Once this Dog are discovered, image are hunt down neighborhood extrema over scale and space. For, e.g., one pixel in a picture is contrasted and its 8 neighbors and also 9 pixels in next scale and 9 pixels in past scales. In the event that it is a nearby extrema, it is a potential keypoint. It essentially implies that keypoint is best spoken to in that scale Interest points (called keypoints in the SIFT framework) are identified as local maxima or minima of the DoG images across scales. Each pixel in the DoG images is compared to its 8 neighbors at the same scale, plus the 9 corresponding neighbors at neighboring scales. If the pixel is a local maximum or minimum, it is selected as a candidate keypoint.

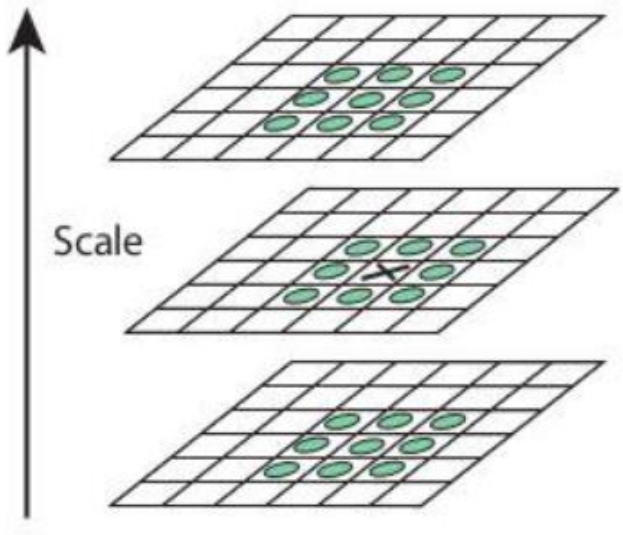

Figure 3.2: Local extrema detection, the pixel stamped x is looked at against its 26 neighbors in a $3 \times 3 \times 3$ area that compasses contiguous DoG Image

3.2.2 Locating keypoint

The key step, additionally is the first venture in object recognition utilizing SIFT technique is to produce the stable feature point. At every competitor area, a definite model is fit to focus scale and area. Keypoints are chosen on premise of measures of their strength.

3.2.3 Orientation assignment

One or more introductions are allocated to every keypoint area on premise of nearby picture inclination bearings. All future operations are performed on picture information that has been converted with respect to the doled out scale, introduction, and area for each one characteristic, consequently giving invariance to these changes. Course parameters to the keypoints are dead set to quantize the depiction. Lowe[10] detailed the determination with the standard and the plot in Euclidean 10 space, with the course of key focuses utilized as standardized the slope bearing of the key point administrator in the accompanying step. After a picture revolvement, the indistinguishable headings requested could be worked out.

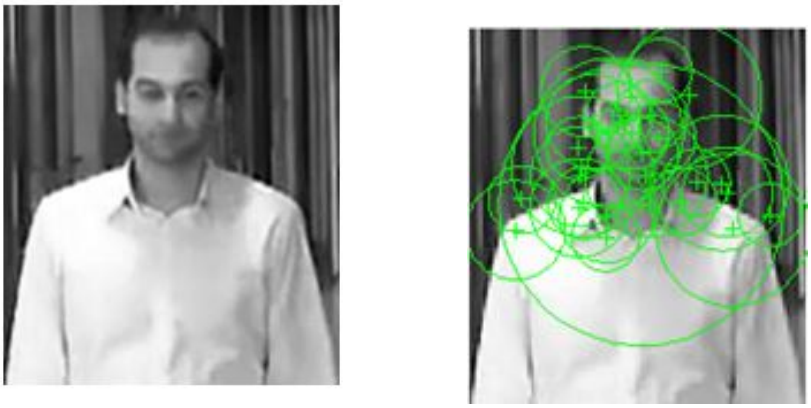

Figure 3.3: Given image and feature point of that image

3.2.4 Generation of keypoint descriptors

The nearby picture inclinations are measured at the coarse scale in the area around every keypoint. These angles are changed into a representation which concedes noteworthy levels of neighborhood change fit as a fiddle mutilation.

3.2.5 Keypoint matching

The next step is to apply these SIFT techniques to video frame successions for object tracking. Filter characteristics are concentrated through the data feature cased successions and put away by their keypoint descriptors. Each one key point allocates 4 parameters, which are 2D area (x direction and y coordinate), introduction and scale. Each one item is followed in another feature edge grouping by independently thinking about each one characteristic point found from the new feature casing arrangements to those on the interested object. The Euclidean separation is presented as a comparability estimation of characteristic characters. The applicants could be safeguarded when the two characteristic's Euclidean separation is bigger than the edge specified past. So the best matches could be selected by the parameters esteem, in the other path, consistency of their area, introduction and scale.

Figure 3.4: Target frame and feature points of that frame

3.3 Kanade - Lucas - Tomasi feature tracker

Kanade-Lucas-Tomasi(KLT) feature tracker is a methodology of characteristic extraction in the area of computer vision. It is proposed basically with the end goal of managing the issue that customary image enrollment strategies are by and large immoderate. KLT makes utilization of spatial intensity information to steer the quest for the position that yields the best match. It is speedier than customary strategies for analyzing far fewer potential matches between the images.

KLT is an implementation, in the C programming language, of a feature tracker for the computer vision community. The source code is in general society space, accessible for both business and non-business utilization. The tracker is focused around the early work of Lucas and Kanade [11], was produced completely by Tomasi and Kanade [12], and was clarified obviously in the paper by Shi and Tomasi [13]. Later, Tomasi proposed a slight change which makes the processing symmetric concerning the two pictures - the ensuing comparison is determined in the unpublished note without anyone else's input [14]. Briefly, great characteristics are spotted by inspecting the base eigenvalue of every 2 by 2 slope framework, and characteristics are followed utilizing a Newton-Raphson system for minimizing

the contrast between the two windows. Multi determination following considers moderately substantial removals between pictures. The relative processing that assesses the consistency of characteristics between non-sequential casings [13] was actualized by Thorsten Thormaehlen a few years after the first code and documentation were composed.

To track the object over time, we also uses the Kanade-Lucas-Tomasi (KLT) algorithm. While it is possible to use the cascade object detector on every frame, it is computationally expensive. It may also fail to detect the object, when the object turns or tilts. This limitation comes from the type of trained classification model used for detection. The example detects the object only once, and then the KLT algorithm tracks the object across the video frames.

The KLT algorithm tracks a set of feature points across the video frames. Once the detection locates the object, the next step is identify feature points that can be reliably tracked. It uses the standard, "good features to track" proposed by Shi and Tomasi.

3.4 Mean Shift

Accurate visual item following under the state of low computational multifaceted nature displays a test. Ongoing undertaking, for example, reconnaissance and observing [15], perceptual client interfaces [16], keen rooms [17, 18], and feature layering [19] all require the capability to track moving object. As a rule, following of visual items is possible either by forward - tracking or by backward - tracking. The forward - tracking methodology appraises the positions of the areas in the current casing utilizing the division result got for the past picture. The backtracking based methodology sections frontal area areas in the current images and after that makes the correspondence of regions between the past image. For securing correspondence, a few object layouts are used. A conceivable forward-following method is mean-shift dissection. Mean movement method was initially presented in 1975, yet just following 20 years after the fact in 1995, this system has been re-presented by D. Fuiorea [20]. In his article, a portion capacity is characterized

to figure the separation between specimen focuses and its mean movement, likewise a weight coefficient is reverse with the separation. The closer the separation is, the bigger the weight coefficient is. The mean movement calculation is a non-parametric method[21]. It gives faultless restriction and proficient matching without exorbitant exhaustive hunt. It is an iterative process, that is to say, first register the mean movement esteem for the current point position, then move the point to its mean movement esteem as the new position, then figure the mean movement until it satisfy certain condition. The guideline of mean movement strategy might be picked up from Figure 3.5.

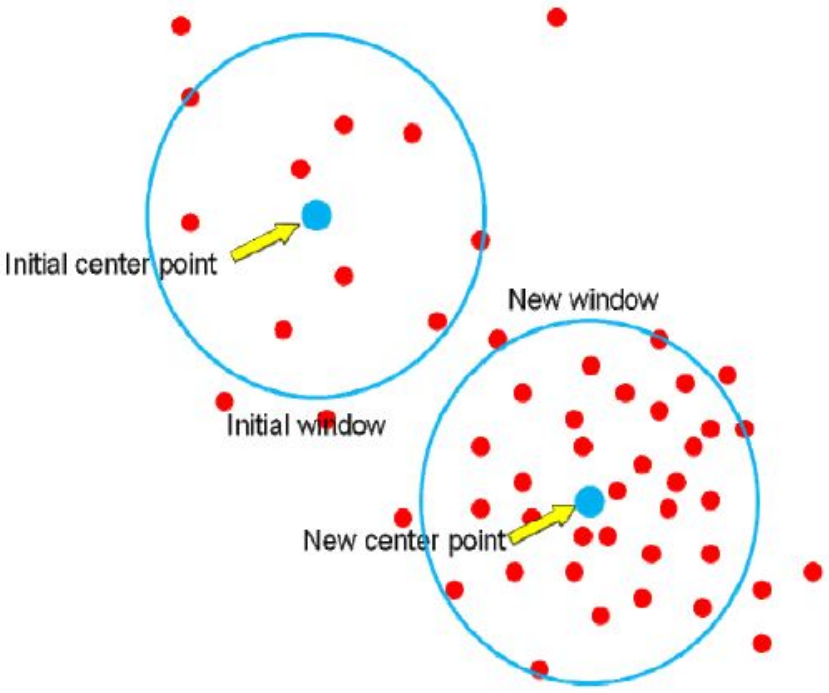

Figure 3.5: The principle of mean shift procedure

Chapter 4

Experimental Result

Several experiments had been done to evaluate the tracking algorithms. These sequences used in experiments consist of indoors and outdoors testing environments so that the proposed scheme can be fully evaluated.

First, target object of interest is defined from the first some frames. Then SIFT features are obtained from the target object. The features are stored by their keypoints descriptors. Each keypoint specifies 4 parameters.

Figure 4.1: Target image and feature point of target image

Then, SIFT features are obtained from the consecutive frames to match the

feature from interested object. The features of frames are also stored by other keypoints descriptors.

Figure 4.2: Corresponding frame and feature point of that frame

The target object is tracked in the next frame by individually comparing each feature point found from the next frame to those on the target object. The Euclidean distance is worked out. The candidate can be preserved when the two features Euclidean distance is larger than a threshold. So the good matches are

picked out by the consistency of their location, orientation and scale.

Figure 4.3: Matched Feature of target object in frame

Tracking results on the example video sequence are illustrated in Fig 4.4. They represent the outcomes of the SIFT and tracker.

Figure 4.4: Detected target object in frame

A trajectory is the path that a moving object follows through space as a function of time. A trajectory can be described mathematical either by the geometry of the path or as the position of the object over time. It will store the actual path of object of interest i.e. information of target in consecutive frames. We will get the all information about target object that in which direction it moves and what is the speed of target. Trajectory of our given object of interest is give below:

Figure 4.5: Trajectory of target object in video

Chapter 5

Conclusion and future Work

5.1 Conclusion

Object detection and tracking is an important task in computer vision field. In object detection and tracking it consist of two major processes, object detection and object tracking. Object detection in video image obtained from single camera with static background that means fixing camera is achieved by background subtraction approach. In this thesis, we tried different videos with fixed camera with a single object and multiple objects to see it is able to detect objects. Motion based systems for detecting and tracking given moving object of interest can be created. Using SIFT feature extraction first feature of the object and the frame has detected to match the interested object. Since for feature extraction, SIFT algorithm has been used so tracker is invariant to representation of interested object.

5.2 Future Work

In the future, we can extend the work to detect the moving object with non-static background, having multiple cameras which can be used in real time surveillance applications.

Bibliography

[1] Yiwei Wang, John F. Doherty and Robert E. Van Dyck, "Moving Object Tracking in Video", in proceedings of 29th applied imagery pattern recognition workshop, ISBN 0-7695-0978-9, page 95,2000.

[2] Bhavana C. Bendale, Prof. Anil R. Karwankar, "Moving Object Tracking in Video Using MATLAB", International Journal of Electronics, Communication and Soft Computing Science and Engineering ISSN: 2277-9477, Volume 2, Issue 1.

[3] Marcus A. Brubaker, Leonid Sigal and David J. Fleet, "Video-Based People Tracking", hand book of ambient intelligence under smart environments 2010, pp 57-87.

[4] Emilio Maggio and Andrea Cavallaro, "Video Tracking: Theory and Practice", first edition 2011, John Wiley and Sons, Ltd.

[5] Y.Alper, J.Omar, and S.Mubarak. "Object Tracking: A Survey" ACM Computing Surveys, vol. 38, no. 4, Article 13, December 2006.

[6] B. Triggs, P.F. McLauchlan, R.I. Hartley and A.W. "Fitzgibbon. Bundle adjustment - a modern synthesis". In Proceedings of the International Conference on Computer Vision, London, UK, 1999, 298?372.

[7] G.C. Holst and T.S. Lomheim. "CMOS/CCD Sensors and Camera Systems". Bellingham, WA, SPIE Society of Photo-Optical Instrumentation Engineering, 2007.

[8] E. Maggio, M. Taj and A. Cavallaro. "Efficient multi-target visual tracking using random finite sets". IEEE Transactions on Circuits Systems and Video Technology, 18(8), 1016?1027, 2008.

[9] G. David Lowe. Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision. 2. pp. 1150?1157,1997.

[10] G. David Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), pp, 91-110, 2004.

[11] Bruce D. Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. International Joint Conference on Artificial Intelligence, pages 674-679, 1981.

[12] Carlo Tomasi and Takeo Kanade. Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

[13] Jianbo Shi and Carlo Tomasi. Good Features to Track. IEEE Conference on Computer Vision and Pattern Recognition, pages 593-600, 1994.

[14] Stan Birchfield. Derivation of Kanade-Lucas-Tomasi Tracking Equation. Unpublished, January 1997.

[15] Y. Cui, S. Samarasekera, Q. Huang. Indoor Monitoring Via the Collaboration Betweena Peripheral Senson and a Foveal Sensor, IEEE Work-shop on Visual Surveillance, Bomba y, India, 2-9, 1998.

[16] G. R. Bradski, Computer Vision Face T racking as a Component of a Perceptual User Interface, IEEE Work. on Applic. Comp. Vis., Princeton, 214-219, 1998.

[17] S.S. Intille, J.W. Davis, A.F. Bobick, Real-Time Closed-World T racking. IEEE Conf. on Comp. Vis. and Pat. Rec., Puerto Rico, 697-703, 1997.

[18] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, Pfinder: Real-Time Tracking of the Human Body, IEEE Trans. Pattern Analysis Machine Intell, 19:780-785, 1997.

[19] A. Eleftheriadis, A. Jacquin. Automatic Face Location Detection and Tracking for Model-Assisted Coding of Video Teleconference Sequences at Low Bit Rates, Signal Processing- Image Communication, 7(3): 231-248, 1995.

[20] D. Fuiorea, V. Gui, D. Pescaru, and C. Toma. Comparative study on RANSAC and Mean shift algorithm, International Symposium on Electronics and Telecommunications Edition 8. vol. 53(67) Sept. 2008, pp. 80-85.

[21] Y.Cheng. Mean Shift, Mode Seeking, and Clustering, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 17, No 8, 790-799,1995

[22] Stern H, Efros B (2005) Adaptive color space switching for tracking under varying illumination. Image Vis Comput 23(3):353364. doi : 10.1016 /j. imavis 2004.09.005

[23] Li S-X, Chang H-X, Zhu C-F (2010) Adaptive pyramid mean shift for global real-time visual tracking. Image Vis Comput 28(3):424437. doi : 10.1016 / j. imavis 2009.06.012

[24] Yuan G-W, Gao Y, Xu D (2011) A moving objects tracking method based on a combination of local binary pattern texture and Hue. Procedia Eng 15:39643968. doi:10.1016/j. proeng 2011.08.742

[25] Mazinan AH, Amir-Latifi A (2012) Applying mean shift, motion information and Kalman filtering approaches to object tracking. ISA Trans 51(3):485497. doi: 10.1016/j. isatra 2012.02.002

[26] Lai S-H (2004) Computation of optical flow under non-uniform brightness variations. Pattern Recognit Lett 25(8):885892. doi : 10.1016 / j. patrec 2004.02.001

[27] Alan J Lipton, Hironobu Fujiyoshi, and Raju S Patil. Moving target classification and tracking from real-time video. In Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on, pages 814. IEEE, 1998.

[28] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., volume 2. IEEE, 1999.

[29] Ya Liu, Haizhou Ai, and Guang-you Xu. Moving object detection and tracking based on background subtraction. In Multispectral Image Processing and Pattern Recognition, pages 6266. International Society for Optics and Photonics, 2001.

[30] Changick Kim and Jenq-Neng Hwang. Fast and automatic video object segmentation and tracking for content-based applications. Circuits and Systems for Video Technology, IEEE Transactions on, 12(2):122129, 2002.

[31] Shahbe Mat Desa and Qussay A Salih. Image subtraction for real time moving object extraction. In Computer Graphics, Imaging and Visualization, 2004. CGIV 2004. Proceedings. International Conference on, pages 4145. IEEE, 2004.