Chapter 1

Introduction

1.1 Preamble

In the context of modern societal need, the demonstration of human-like intelligence by machine attempts to emerge as Artificial Intelligence (AI). It is an interdisciplinary subject in nature. The Computer Vision (CV) is the broader area in which process of acquiring, analyzing and understanding real world images of high-dimensional data to fabricate numerical or symbolic information.

The AI and CV share allied and stem level topics such as Image Processing (IP), Pattern Recognition (PR) and Machine Learning (ML) techniques. Consequently, CV is sometimes seen either as a part of AI or of computer science. The IP and PR techniques are showing their best potential to exploit and solve abundant attractive and challenging problems associated with machine vision and intelligence area of research in recent past. Therefore, few important and open topics for research under CV are autonomous vehicle, detecting events, modeling objects or environments, computer-human interaction, automatic inspection, scene reconstruction, event detection, indexing, motion estimation, image restoration, object recognition and object tracking.

Out of these research avenues, the object tracking problem has been gathering huge significance in the past decade and accordingly, numerous methodologies were proposed to resolve it to a certain extent. In this respect, to rejuvenate
the much potential strategies, perhaps the urgency to cater and nurture the break through contributions to track the object in given scene has been adopted.

The process of locating the existence of an interested single or multiple objects in a sequence of the frames is object tracking. It is imminence since human cannot focus on the interested moving object (MO) in a scene of video for a long time due to natural constraint. Therefore, with the help of computer intelligence, a visible bounding box is placed over the interested moving object. In other words, viewer is forced to focus on the particular object to understand the behavior, movement, for closer observation, to detect crimes and to monitor suspected movement etc. Object tracking applications are in surveillance, traffic controlling, robot routing, media, medical field, law enforcement and sports etc. It can be broadly classified as point, kernel and silhouette tracking.

In current state-of-art, huge high dimensional video dataset have been continuously heaped through the best available advanced instruments. Thus, these devices are capable of capturing the scene to certain possible extent. Using a single recording instrument, it will not be possible to capture complete information. To solve this problem, it is necessary to add many video recording instruments placed at different positions. Further, multi-dimensional scenes are more capable to provide visual effects profoundly but at the cost of high computational effort.

This research work attempts to focus on the scene or video captured by a single camera. The process of object tracking will be trivial with static background and single object. As the number of objects to be tracked in a scene increases, tracking requires more effort. Thus, increase in the number of objects and dynamic background raise the difficulty levels due to non-capability of tracking strategy to distinguish clearly moving objects. Similarly, scene with dynamic background and single object needs careful attention and radical approach. Further, multiple object oriented scene exploration is much complex in nature. This necessitates, a deeper investigation of the problems related to object tracking in offering various approaches to tackle such intricacies. This research work, proposes techniques which represent spatial, frequency or hybrid domains through IP and PR approaches.

Normally, the moving objects are detected in order to trace its location in subsequent frames. Those detection techniques are background subtraction,
frame differencing and optical flow etc. Surprisingly, some proposed tracking techniques do not require the extraction of moving objects like Normalized Cross Correlation (NCC) and Meanshift criteria etc. Therefore, the work portrayed in this thesis, can be categorized based on set of features erected to represent its location and path of motion as spatial and frequency domain contributions. The prime purpose of research work is to explore the different potentials and the extents of IP and PR techniques to create unified approaches of object tracking.

1.2 Object Tracking Strategies

Object tracking is the part of study of CV and the effort is put forward to reveal through some of the basic methods of object tracking. Figure 1.1 portrays broad classification of the object tracking.

![Object Tracking Classification Diagram](image)

Figure 1.1: Object tracking classification.

1.2.1 Object Tracking: Definition and State-of-Art

Video analysis involves the major parts like detecting, tracking and analyzing the object. Object tracking is a process of establishing the correspondence of the object in subsequent frames. It is essential for recognition, detection and surveillance in sensitive places. Many trackers have been developed to track the object robustly. Though the innumerable methods of object tracking have been
Chapter 1. Introduction

laid down, they differ in accumulating unique features of object and its path of motion. The following are the stages through which object tracking is achieved:

- Extraction of objects.
- Shape representation of the objects using points, primitive geometric shapes, object contours and appearance representations such as histograms etc.
- Extraction of features of object like edges, color and motion etc.
- Object identification.

The objects which are under consideration for analysis such as boats on the sea, fish inside an aquarium, vehicles on a road, planes in the air, people walking on a road or bubbles in the water. Here, interested objects are tracked in a specific domain and usually the objects are represented by their point, shapes and appearances etc. [1].

Firstly, the point representation is suitable for tracking objects that occupy small regions in an image. Otherwise point along with bounding box is considered. Consequently, primitive geometric shapes are used to represent the object such as rectangle, ellipse, circle and square etc. Those are typically used to track simple rigid and non-rigid objects. Modeling of the object motions is typically done with transformations like translation, rotation and scaling which may be affine or projective in nature. The object silhouette and contour are also the modeling representatives. The contour is a boundary of an object and silhouette is the inside portion of the contour. These are used for tracking complex and flexible shapes. Different parts are tied together in a body with joints, which appear as articulated object. The body of human being is best example of an articulated object with two legs, two hands, head, fingers, knee, feet and torso. In the case of the skeletal models, the transformation such as medial axis is used on silhouette of an object to extract the object skeleton. An improved strategy based on appearance representation in object tracking is by using probability density function or histogram.

The selection of unique feature of the object plays a vital role in tracking it. In some cases color has been exploited for histogram representation and in other
cases object boundary may act as contour representation. The combination of these features also has been utilized in most of the cases.

### 1.2.2 Point based Object Tracking: State-of-Art

Here, the object detection and tracking with the help of spatial information of moving object such as centroid is used. Point tracking is complicated in the presence of occlusions, misdetections, entries and exits of objects. Point tracking is divided into two categories namely deterministic and statistical. In deterministic method, qualitative motion based on heuristics observation and statistical methods of object measurement along with uncertainties are taken into account to establish correspondence. Further, deterministic correspondence cost is defined following combination of constraints. Proximity, maximum velocity, small velocity change, common motion, rigidity and proximity uniformity are the constraints in statistical methods. In statistical methods state space approach is used for modeling the object such as position, velocity and acceleration. Object motions can undergo disturbance and this strategy solve the tracking problem by taking measurement and model uncertainties. If object state variables are Gaussian distribution then Kalman filter is used. Kalman filtering has two steps – initial prediction and eventual correction. The prediction state employs a state model to predict new state of variables. Correction step performs observations to update object’s state. If object state variables are not Gaussian distribution, then particle filter is needed. Both criteria estimate the state of a single object. The work of Lowe David proposed an astonishing local descriptor extraction algorithm named Scale Invariant Feature Transform (SIFT)[2] which has the merits of being invariant to scale, rotation, translation and illumination. For each feature, there will be a corresponding vector which describes the feature of the neighboring point making it more distinctive in nature. Hence, matching process between the points in different images proves to be much easier.

The author, P. Gabriel et al [3] have offered an approach in which an object is characterized by a set of Interest Points (IPs) attained through a color Harris detector. Here, set of IPs is described by its local appearance and permits to track an object with partial occlusion provided one or more points remain visible.
Further, the robustness has been increased by exploiting the potential geometric relationships among the IPs.

The authors, Wei Zeng et al [4] have proposed a new algorithm which is based on Kalman filter and point matching assessment. Here, prediction of target location and extraction of multi-scale corner points which are geometrically invariant with different weights to the resulting function has been achieved. The object location is tracked by average vector of image which is divided into blocks.

The work of Yang Yang and Qixin Cao [5] has proposed object tracking system for robot grasp. The object has been detected with the help of SIFT features as points. Image position in tracking stage is computed with homography constraints and interest window is created.

### 1.2.3 Kernel based Object Tracking: State-of-Art

Generally, the kernel is used to represent the shape of object, its appearance and rectangular template or elliptical shape. Tracking is done by computing motion of kernel in consecutive frames. The motion is parametric transformation such as translation, rotation and affine. Object is represented by primitive shapes and motion of object is calculated from frame to frame. There are two sub categories in kernel tacking. First one is templates and density-based appearance and the second one is multiview appearance models.

Tracking templates and density-based appearance models is simple with low cost of computation. In the case of single object tracking template matching has been employed and the brute force method for searching the similar region corresponding to the template in the next frame is done. The color features and intensity of the images are used to create the templates. The template position in the present image is calculated by similarity measure like cross correlation.

Authors Comaniciu et al [53] have suggested a kernel-based tracking approach where the target is characterized through the features of weighted combination of pixels. The central pixels of object represent more reliability than pixels far from the center. The object region, considered as a unit circle, is normalized and a monotonically decreasing kernel which is masked over the circle is used to reveal the importance of each pixel.
Chapter 1. Introduction

The research team Ghaeminia et al [6] have intended an adaptive motion model which signified the improvement in tracking the human with the help of particle filter.

Author, Chia et al [7] have proposed another system of kernel-based tracking. The particle filter sampling process has been applied with meanshift algorithm to decrease the number of particles for reducing computation. In other words, meanshift is an optimizing algorithm to reduce the computational time.

The kernel based object tracking strategy using color histogram technique and covariance algorithm has been suggested to tackle different challenging situations like illuminations and occlusions [8]. Here, the histogram based method proves to be efficient in terms of computation time and the covariance helps in increasing the detection rate.

The work of Mohammad Mahdi Dehshibi et al [9] has proposed the system based on Incremental Bhattacharyya Dissimilarity (IBD) to constantly differentiate the particles of the object region from the others in the background scene. Here, the IBD measure reliably distinguishes the foreground particles from the cluttered background with partial occlusion too. The system of kernel-based object tracking has shown significant improvement in the speed and reliable handling of partial occlusion.

1.2.4 Silhouette based Object Tracking: State-of-Art

Objects that are difficult to represent by geometric shapes like head, hands, shoulders and arms are conveniently represented by silhouette methods. Further, it attempts to reveal the object region in each frame through the object model of the previous frame. Color histogram or object contour are used as object models. Two strategies are noticed in the same category like shape matching and contour evolution. In shape matching object silhouette is found in present frame. In contour tracking, initial contour is extracted to its new position in the current frame either by using state space models or by direct minimization of some energy function.

Shape matching is similar to the template matching. In this, object silhouette and its related model is searched in the current frame. Search is computed on similarity between the model generated and hypothesized object silhouette which
Chapter 1. Introduction

is based on the previous frame. In this method, non-rigid objects are not handled properly. Object model is reinitialized in every frame for appearance, view point and non-rigid object motion changes.

The authors, Chirag Patel and Ripal Patel [10] have suggested a method in which contour of the object in the first frame represented by color histogram. The low frequency colors are removed in order to nullify the effect of noise. The current position of the object contour is represented through its coordinates of its centroid, velocity and acceleration. Consequently, the state of the contour is predicted by Kalman filter to obtain the correct contour in the subsequent frames. The centroid of the contour is acquired in the new frame and then the best fit contour of maximum energy is estimated by searching the neighborhood of the contour.

The work of Junqiu Wang et al [11] has offered a system of gait recognition in order to identify individuals far away from the camera. The tracking module creates initial sequences which consist of bounding boxes and foreground likelihood images. The optimal silhouette-based gait models are searched by gait recognition module based on the tracking results. Further, the segmentation module attempts to segment people using the gait silhouette as shape.

The researcher, Yilmaz A. et al [12] proposed a robust tracking method with complete object regions, adjusting to varying visual features and handles occlusions. The object tracking is accomplished by way of developing the contour frame by frame through minimization of some energy function estimated in the contour neighborhood. This method consists two components as visual features and the object shape. The visual features like color and texture etc. are modeled with the help of semi-parametric models and combined by independent polling of opinion. The shape level sets as shape priors are exercised to recover the missing objects in occlusion.

The author, Huttenlocher D. et al [13] have developed the system based on silhouette shape matching. In shape matching technique of tracking, the silhouette in a frame is searched by way of similarity computation in subsequent frames. Here, the estimation of silhouette matching score has been achieved. The Hausdorff metric measures the magnitude of differences between two sets of points. This is employed on an edge map to represent position of the silhouette matching with the input frame.
1.3 Object Tracking Applications and Challenges

In recent times, the broader area of CV extended the attention towards the object tracking as a major topic of study. The foremost ambition of object tracking is maintaining track and attention of interested particular object in video progression. The movement of object occurs in variety of directions. The scaling problem enhances when the object travels away or towards in the field of view. There may be change in the illumination of the object, in the appearance of object because of deformation and partial occlusion.

The interested object may make an inclination or twists which causes rotation problem. Additional constraint such as sudden entry or exit of the object creating occlusion in the scene is a challenging task [1]. Therefore, the main issues such as translation, scaling, rotation, occlusion, change in illumination, appearance and perspective projection etc. have to be focused. Object tracking process is significant to provide some other applications like surveillance related to security, recognizing suspicious people, providing superior and logical security with the help of video contents, enhancement of the medical facilities to the patients and differently-abled people, traffic flow evaluation to reduce the density by diversion and detection of accidents etc. In the field of entertainment, tracking is employed for video editing for creating innovative effects etc. Challenges or difficulties posed in the process of tracking have been listed below:

- Camera motion.
- Images corrupted with noise.
- Unpredicted motion of the object.
- Flexibility in the object shape.
- Transformation of 3D world information into 2D.
- Occlusions of objects.
- Complexity in object shapes.
- Change in illumination and need of real time processing.
1.4 Problem and Dataset

The goal of research work is to track a single object or multiple objects in the image sequences which are generated by single camera in static or dynamic environment. Different tracking methodologies have been proposed using IP, PR and statistical schemes. These methods have been validated by employing indoor and outdoor benchmark video sequences depicting different environmental conditions.

Popular benchmark dataset such as Ball, Sample and Performance Evaluation of Tracking and Surveillance (PETS) [68], Video Surveillance Online Repository (VISOR), Hamburg-Taxi, dt neuWinter [69] have been used. PETS dataset has PETS2001(1), PETS2001(2), PETS2001(3) and PETS2000 video sequences.

Figure 1.2: Benchmark dataset.
Table 1.1: Details of benchmark dataset.

The Ball dataset shows a falling ball with increasing velocity in a static background and dynamic illumination. The Sample dataset shows a human being walking towards the camera resulting in scaling up of the moving object in static background with dynamic illumination. PETS dataset shows people, vehicles moving in random directions with dynamic background and illumination. Additionally, there is another important traffic dataset of VISOR having dynamic background and illumination. The Hamburg-Taxi dataset has two dark vehicles moving in opposite directions and one white car moving in
Chapter 1. Introduction

Another direction. This dataset is complicated because of dynamics in illumination, similar colored objects and background. The dtneuWinter is another complex dataset representing the snowy area and similarly colored objects in which different sized vehicles move in random directions. Figure 1.2 shows the nine dataset used for experimentation. Table 1.1 shows the characteristics of these dataset along with the challenges involved in tracking.

1.5 Performance Yardsticks

Experiments by applying proposed methods have been conducted on benchmark dataset. Performance analysis and the object tracking worths are demonstrated through qualitative, quantitative and time analysis. Qualitative analysis of tracking performance is observed by the visual analysis. The quantitative analysis is ascertained by the True Positive (TP), False Positive (FP) and False Negative (FN). The moving objects and detected regions are same in case of True Positive (TP) where as wrong matching of objects is treated as False Positive (FP). If the bounding box is enclosed on neither of the moving objects then it is declared as False Negative (FN).

\[
P = \frac{TP}{TP + FP} \times 100 \quad (1.1)
\]

\[
DR \text{ or } R = \frac{TP}{TP + FN} \times 100 \quad (1.2)
\]

\[
FAR = \frac{FP}{TP + FP} \times 100 \quad (1.3)
\]

\[
CR = \frac{DR}{DR + FAR} \times 100 \quad (1.4)
\]

\[
F = \frac{2 \times P \times R}{P + R} \quad (1.5)
\]

\[
G = \sqrt{PR} \quad (1.6)
\]

The tracking performance is estimated by significant metrics like Precision (P), Detection Rate (DR) or Recall (R) and False Alarm Rate (FAR) which use the TP, FP and FN as per Eqs. (1.1)-(1.3). The values of P and R or DR are relative to the FP and FN respectively. For FP and FN being zeros or minimum, P and R will be 100% or high respectively. The Combined Result (CR), F-measure
and G-measure are used to furnish the refined results of tracking by employing P, R and FAR as given in Eq. (1.4), (1.5) and (1.6) respectively.

1.6 Proposed Strategies and Motivation

In order to track the object under difficult situations, some novel algorithms have been proposed like exhaustive image based NCC for tracking the object and updating by NCC itself. In the same category of algorithms, only updating part is replaced by PCA and Histogram Regression Line (HRL) respectively. Similarly, other methods have been proposed for partitioned image with same updating strategies.

Subsequently, the meanshift algorithm has been proposed with joint histogram of color and CCV for similar background contexts. Color Optical flow methods of Horn-Shcunck and Lukas-Kanade have been employed to extract the moving objects. The concepts of HRL and Doyle’s distance are exploited for updating purposes.

In an effort to propose the methodologies in frequency sphere, an optical technique of Horn-Shcunck has been employed to obtain the moving regions. The Fourier Descriptors have been utilized to update the template for various conventional shape signatures along with the proposed Geometric Mean of Segmented Centroid (GMSC) distance function.

The other proposed systems use third level DWT or Lifting based DWT (LDWT) to get moving objects using three consecutive frames. Accordingly, moving objects are created through Double Change Detection (DCD). Further, in order to update the template, Weighted Standard Deviation (WSD) Descriptors for each of the template and moving regions are estimated. The minimum Euclidean distance between the template and the candidate templates will decide the best match to update.

In the earlier works no evidences are found of this kind. Hence, we are motivated to carry out the research to propose these novel methodologies which enhance tracking performance.
1.7 Contributions

The focal point on object tracking research effected in copious fascinating contributions. These contributions have brought in numerous exciting issues and possibilities which could result in a number of upshots. The chief outcomes based on this thesis are listed below:

1. The Routine approaches of respective object tracking methodologies.
   a) Background Subtraction
   b) Frame Differencing
   c) Normalized Cross Correlation (NCC)

2. NCC based tracking for
   a) Exhaustive image Template matching (ET)
      i. Updating by NCC
      ii. Updating by PCA
      iii. Updating by Histogram Regression Line (HRL)
   b) Partitioned image template matching (PT)
      i. Updating by NCC
      ii. Updating by PCA
      iii. Updating by Histogram Regression Line (HRL)

3. The robust tracking systems and tracking efficiency is improved using Meanshift with
   a) Color
   b) CCV
   c) Joint Histogram of Color and CCV
Chapter 1. Introduction

4. The color optical flow has been utilized to offer systems which give better performance.
   a) Optical flow of Horn-Schunck with Histogram Regression Line (HRL) for updating
   b) Optical flow of Lukas-Kanade with Histogram Regression Line (HRL) for updating
   c) Optical flow of Horn-Schunck with Doyle’s distance for updating
   d) Optical flow of Lukas-Kanade with Doyle’s distance for updating

5. Moving objects have been extracted through gray optical flow of Horn-Schunck and updated by
   a) Fourier Descriptors of conventional Centroid distance function
   b) Fourier Descriptors of Geometric Mean of Segmented Centroid (GMSC) distance function

6. Frequency domain technique used to extract the moving regions
   a) Third level DWT with updating by Weighted Standard Deviation (WSD) descriptors
   b) Lifting based DWT (LDWT) with updating by Weighted Standard Deviation (WSD) descriptors

1.8 Organization of the Thesis

The above contributions headed for object tracking are organized as follows to present an incremental stream of thoughts from trivial techniques such as background subtraction, frames difference and NCC matching to advanced techniques based on DWT or LDWT and WSD etc.

Chapter 2 proposes the basic object tracking techniques such as the background subtraction, frame difference and NCC for matching. Here, the spatial information is used to create the bounding box over the object by experimentation with Ball and Sample dataset.
Chapter 3 introduces the NCC based tracking in two fold. Firstly, an exhaustive image based object tracking schemes are proposed with the respective updating schemes as NCC, PCA and HRL. Secondly, partitioned image based tracking with respective updating schemes as NCC, PCA and HRL.

Chapter 4 presents the meanshift criteria with color, CCV and joint histograms respectively and detailed experimentation has been conducted on the different dataset to reveal the outcome. The performance outcome of meanshift criterion with the joint histogram is praiseworthy.

Chapter 5 offers the systems based on color optical flow like Horn-Schunck and Lukas-Kanade to obtain the moving regions. Two schemes like HRL and Doyle’s distance are attached to update the templates. Therefore, there are four proposed methodologies in this chapter such as Horn-Schunck with HRL updating, Lukas-Kanade with HRL, Horn-Schunck with Doyle’s distance and Lukas-Kanade with Doyle’s distance updating respectively.

Chapter 6 elaborates the gray optical flow technique of Horn-Schunck to extract the moving regions. The updating process has been achieved through creating the Fourier Descriptors (FD) for various shape signatures like boundary, edge, area, curvature and centroid. The analysis with these signatures revealed that the outcome is best for conventional Centroid shape function. Further, tracking performance has been increased by providing a novel shape signature of Geometric Mean of Segmented Centroid (GMSC) distance function.

Chapter 7 employs the third level DWT or LDWT to get the moving regions. Here, Weighted Standard Deviation (WSD) descriptor is engaged to update the template. The offered methodologies are third level DWT with WSD and LDWT with WSD for updating respectively. The experimental analysis portrays that the performance is better in case of LDWT based tracking with WSD updating scheme.

Chapter 8 comprehends various proposed object tracking models in this thesis for various dataset through the performance yardsticks.

Chapter 9 summarizes the findings and lists out the future scope and challenges in object tracking for further research.