Overview of Object Detection and Tracking based on Block Matching Techniques

Apurva S. Samdurkar¹, Shailesh D. Kamble², Nileshsingh V. Thakur³, Akshay S. Patharkar⁴

¹PG Scholar, Computer Science & Engineering, Yeshwantrao Chavan College of Engineering, India
²Computer Science & Engineering, Yeshwantrao Chavan College of Engineering, India
³Computer Science & Engineering, Prof Ram Meghe College of Engineering & Management, India
⁴Computer Technology, K.D.K. College of Engineering, India

¹aapurva2009@gmail.com, ²shailesh_2kin@rediffmail.com, ³thakurnisvis@rediffmail.com, ⁴akshay.patharkar7@gmail.com

Abstract - Object tracking is one of the vital fields of computer vision that detects the moving object from a video sequence. Object detection is used to detect the object present in the video and to find the exact location of that object. The object tracking can be applied in various fields that include video surveillance, robot vision, traffic monitoring, automated civil or military surveillance system, traffic monitoring, human-computer interaction, vehicle navigation, biomedical image analysis, medical imaging and much more. The object tracking algorithm requires tracking the object in each frame of the video. A common approach is to use the background subtraction, which eliminates the common static background, resulting into foreground region showing the presence of the desired object. Block matching technique is the most popular technique for computing the motion vectors between the two frames of video sequences and different searching techniques are available to compute motion vectors between frames. Still, there is a scope for improvement in modifying or developing a new shape pattern for block matching motion estimation to find out and track the object in the video. This paper presents the several object detection and tracking methods and how block matching can be used to track object from a video.

Keyword- Object Tracking; Object Detection; Background Subtraction; Block Matching

I. INTRODUCTION

Detection of an object is typically the first step towards tracking process. The tracking methods need an object detection mechanism, either in sequence of frames or when the object first appears in the frame of the video. Object tracking means the process of locating the object of interest from a video sequence. The object is tracked based on monitoring the motion of an object in the video. Videos are the collective and sequential representation of the image frames. Each of the frames can be divided into two set of objects, foreground and background objects. The foreground objects are the moving objects which can be a bird, car, person, etc. and the background can be the static things. The complete process of object tracking can be categorized into the following three steps, object detection, object classification and object tracking as depicted in Figure 1.

Object detection: It is done to find out the region of interest from the frames of the video. Various methodologies are present for object detection from a video sequence. The object detection techniques used are frame differencing, optical flow and background subtraction. Out of three major classes of moving object detection techniques, namely, frame differencing, optical flow, and background subtraction, the last is somewhat robust, as compared to the others [1][2].

Object Classification: The detected object then can be classified as various moving objects. There are many approaches to classifying the moving objects shape-based classification, motion-based classification, color-based classification, and texture-based classification.

Object Tracking: Aim is to generate a trajectory of an object by locating its position in every frame of the video. The approaches to tracking the object are point tracking, kernel tracking, and silhouette tracking.

![Figure 1: Steps in Object Tracking](image-url)
techniques and the cost functions associated with block matching algorithms. Section VI describes a related work on object tracking using block matching. Section VII discusses and roundup the conclusion from the studied existing approaches of object tracking and block matching. Ultimately, the paper ends with the future scope in section VIII followed by references.

II. OVERVIEW OF OBJECT DETECTION

Detecting the object from the video is the first step to tracking an object. Detection of an interesting moving object can be achieved by different existing techniques such as frame differencing, optical flow, background subtraction, segmentation, point detectors [3] as shown in Figure 2.

Displacement position of the objects are the only fundamental source of information. Detailed explanation for various methods is as follows:

Frame differencing: Difference between two consecutive frames is determined to find the moving object. This calculation is easy and simple to implement.

Optical flow: This method [1] is used to calculate the images optical flow field and do clustering process agreeing to the characteristics of optical flow distribution of images. Optical flow presents complete information about the movements and can detect the moving object from the background better.

Background subtraction: It is achieved by building a background model and then finding deviation from each incoming frame [3] [4]. The change with respect to the background model denotes the moving object. The background frame without any object is captured, afterward, when a moving object enters, the second picture is formed [2]. Subtracting the second frame from the first background frame gives the dissimilarity between two frames and the position of moving object can be obtained as shown in Figure 3. Background subtraction technique can be divided into two categories: recursive and non-recursive technique [1] [4]. The recursive technique does not maintain a buffer for background evaluation. On the contrary, they recursively update a single background model based on each input frame[1]. Recursive techniques require less storage. On the other hand, a non-recursive technique makes use of a sliding-window approach for background estimation. A buffer of previous video frames is stored, and tally the background image based on the changes of each pixel within the buffer. Non-recursive techniques are independent of the history beyond those frames stored in the buffer.

Segmentation: The intention of image segmentation is to divide the image into similar regions. The two major problems faced by segmentation are, (i) the standard reference for a good partition and (ii) systematic way for achieving efficient partition. Mean shift clustering, segmentation using graph cuts and active contours are the different segmentation techniques [3].

Point detectors: They are used to detect the points of attention in images which have an effective texture. These interest points are used in the context of motion, stereo, and tracking [3].

III. OVERVIEW OF OBJECT TRACKING

Object tracking is locating the position of the object of interest in each of the sequence of frames. Sequence of frames combines to form a video. Each frame undergoes the process of object tracking. Object extraction, recognition, tracking and decision about tracking can be done by object tracking. Object tracking can be categorized as point tracking, kernel tracking, silhouette tracking [1]. These three methods can be categorized into subtypes as depicted in Figure 4.

Point tracking: The moving objects in the image are portrayed as feature points during the tracking process. It involves detection in every frame. The relation of detected objects is served as points across the frames. The point correspondence seems to be a difficult problem- principally in the existence of occlusions, false detections [3]. Point tracking is divided into deterministic and statistical methods [3].
(1) **Deterministic Method:** This method works on qualitative motion heuristics by forming a connection between each object in the previous frame with the single object in the current frame. This is performed with the assistance of a set of motion constraints. This method for point correspondence associates a cost of each object in frame $f$ to a single object in frame $(f-1)$, using a set of motion constraints [3].

(2) **Statistical Method:** It is also called as Probabilistic Method. This method works by determining the position of an object in the frame with detection mechanism. A probabilistic method considers the object’s measurement and uncertainties so as to establish the relation. The statistical methods resolve the tracking problems like the noise present in the measurements. Object motions can undergo disturbance.

**Kernel tracking:** It is computed by representing the moving object’s region from frame to frame. It is based on the object's motion. Based on the object’s representation, shape and appearance, the number of objects tracked, the kernel tracking method are divided into two subcategories as template based and multi view model.

(1) **Template based models:** This method is based on searching the image, for the object template defined in the previous frame. Templates and density-based appearance models are commonly used due to its simplicity and low computational cost.

(2) **Multi-view appearance models:** It is the new approach used for objects that have different views in different frames of the video. There are some difficulties faced in another method to track object from different views. This model represents the information gathered through the most recent observations of the object. The object appears different from different views and if the view of object changes during tracking, then this model is invalid.

**Silhouette-based tracking:** Complex objects such as hand, fingers are difficult to define by geometric shapes. This method provides shape descriptors for the object. And aims at finding the object region in each frame with the help of an object model generated using the previous frames. The model can be in the form of an object edge or the object contour and color histogram. Shape matching and contour tracking are two main categories of silhouette tracking [3].

(1) **Shape Matching:** It is somewhat similar to the template-based tracking in kernel approach. It can be performed by tracking where an object silhouette and its associated model are present. In this approach, the search is done by means of checking the similarity of the object with the model generated from the two successive frames.

(2) **Contour Tracking:** Contour tracking method repeatedly unfolds a primary contour in the preceding frame to its new position in the present frame [1]. This tracking method requires some overlapping part of the object in the current frame with the object area in the previous frame. Tracking by contour tracking method can be performed using two different approaches. (i) The state space models to model the contour shape and its motion. (ii) Minimizing the contour energy using direct minimization techniques like gradient descent.

**Figure 4:** Object Tracking Methods and Types

**IV. RELATED WORK ON OBJECT TRACKING**

Parekh et al. [1] have suggested various methods for detecting an object and tracking an object from a video sequence. The different approaches to detecting the objects are briefly explained which include frame differencing, subtracting the background and optical flow. Each of these objects detecting methods has advantages and drawbacks. Also, the moving objects can be classified based on its shape, color, texture-based, and motion. The actual object tracking approaches suggested by authors include the point tracker, kernel-based tracking, and silhouette based tracking. From these available and suggested methods, background subtraction is the simplest and easy method that provides the complete information about the object rather than the other methods. A brief review of object tracking methods is presented by Yilmaz [3]. These tracking methods are divided into three categories based on object representation and require detecting the object at some point. They have categorized as point correspondence, primitive geometric model, and contour evaluation. The later focus was moved towards the object detection approaches which include the point detectors, background subtraction, and segmentation and supervised learning. The issues related to object tracking like occlusion, view through multiple cameras are also discussed. Background subtraction is a widely-used approach for detecting moving objects from static cameras [2]. Humans can easily detect the objects present in an image or video. But it is not so easy for the machine to do the same, for this, we require more intelligent machines. One way is to use the concept of Steiner tree [5]. The method that is widely used to detect a moving object is background subtraction.
Some of the most popular cost functions in terms of computational expense are:

**Mean difference or Mean Absolute Difference (MAD):** The mean absolute difference or the MAD is the "average" or "Mean", of the absolute difference of two variables X and Y independently. Mathematically it is given by:

\[
\text{MAD} = \frac{1}{N^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} |C_{ij} - R_{ij}|
\]

**Mean Squared Error (MSE):** The mean squared error (MSE) calculates the average of the squares of the errors and is given by:

\[
\text{MSE} = \frac{1}{N^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (C_{ij} - R_{ij})^2
\]

In the equation, N is the size of the macro-block, and \(C_{ij}\) and \(R_{ij}\) are the pixels being compared in current macroblock and reference macro block, respectively.

**Peak signal-to-noise ratio (PSNR):** The image with motion is created using the motion vectors and macro blocks from the reference frame is characterized by Peak signal-to-noise ratio (PSNR) and given by:

\[
\text{PSNR} = 10 \log_{10} \left( \frac{\text{peak value of original data}^2}{\text{MSE}} \right)
\]

To track the object from the video sequence, background subtraction is performed as shown in Figure 3. Now, if the existence of moving object occurs in both the adjacent frames, the tracking area will be overestimated as shown in Figure 5.

![Figure 5: Background Subtraction without Block Matching Algorithm [18]](image)

Figure 5 represents the generation of the redundancy in the video sequence. To overcome this redundancy, the Block matching algorithm is used in which motion estimation is applied to adjust the tracking area size. The basic concept of BMA applied is that dividing the current frame into small blocks, of equal size. Then for each of the small block we find...
the adjacent block from the search area of the last frame that matches mostly to the current block. Hence the matched block from the previous frame is selected as the motion source of the current block and the resulting position of these two blocks gives the motion vector (MV) that needs to be found. When all the motion vectors of the tracking area have been computed, the most frequently occurred motion vector is selected for correcting the tracking area as shown in Figure 6.

![Figure 6: Object Tracking using Block Matching Algorithm](image)

**VI. RELATED WORK ON OBJECT TRACKING USING BLOCK MATCHING ALGORITHM**

Block matching is a standard technique for encoding motion in video compression algorithms and explores the abilities of the block matching algorithm when applied for object tracking [15]. Gyaourova et al. [15] carried out an experiment to reach goal having two aspects: (i) exploring the performance of the motion estimation algorithm and (ii) improving the motion estimation/detection performance by using different block matching algorithms (BMA) for gaining good object tracking results. A comparative approach for block matching is discussed by Hussain and Haque [17]. Various block matching algorithms are discussed, implemented and compared six different types of block matching algorithms, starting from the basic Exhaustive Search to the recent and fast adaptive algorithms like Adaptive Rood Pattern Search [16, 17]. Various block matching algorithms are explained and implementation of full search motion estimation and three step search motion estimation is carried out and the comparison is done between these two algorithms on basis of computational complexity and PSNR. At last, it is concluded from the result that computational complexity of three step search is 10 times less than the full search algorithm [18]. The comparison among fast block matching algorithms (FBMAs) for motion estimation and object tracking are derived from the experiment by Sherie et al. [20]. A fast BMA is developed that was best from the rest and efficient and compared with other BMA. The resemblances are discovered for both motion estimation and object tracking over the standard test data sets. Another object tracking approach using block matching algorithm is introduced in [21], new object tracking technique is presented. The aim of the presented method is to enhance tracking exactness while keeping the tracking process fast. This technique is based on finding motion vectors. The modified block matching algorithm considers solely the area of motion and thus reduces the computational cost [21]. Based on the various methods and categories available for object detection and tracking, Sugandi et al. [22] has designed an object tracking methodology, which describes the region based tracking. The region-based tracking algorithms track objects based on the modifications of the image regions that corresponds to the moving objects. Regions with motion are detected by deducting the background from the current image [22]. The diamond search algorithm and its modified algorithm for motion estimation consist of a small diamond shape pattern (SDSP) and large diamond shape pattern (LDSP) used in video processing [23, 24, 25]. Diamond search is the basic algorithm which is extended into the cross-diamond search and novel cross diamond search. All these algorithms differ in terms of their searching points. Less the number of searching points, more efficient is the algorithm. Block matching motion estimation is the most popular and efficient techniques used to removes the temporal redundancy present between the two frames, the current frame is divided into blocks and for each block one searches for the best-matched block in an available previous frame. Motion Vector (MV) is the displacement between the candidate block and the best matched in the previous frame. The motion estimation is a technique [30], which tries to minimize the temporal redundancy between the successive frames of the video. Motion estimation is computationally very expensive and consumes about 75% of the computational cost during motion estimation process. The three-step search (TSS) and its improved algorithm for block matching algorithm are defined in terms of the search points required [32]. Block matching algorithm are popular for their simplicity and effectiveness in processing. Block matching algorithms are widely adopted due to the motion analysis of objects, tracking the objects and also in video compression and processing [28] [34]. Block matching motion estimation is used for the video compression and fractal coding [33, 34]. The standards for video compression are utilized for the video coding. Motion estimation and compensation are employed in accomplishing the minimal temporal redundancy between the frames of the video. Motion estimation means finding motion vectors of an object in an image. Data compression is achieved on sequential images with the help of the information of the moving object. Several techniques are available to estimate the motion between two frames. Motion is the important feature in tracking an object from the video. The motion estimation includes a lot of challenges like the computational complexity etc. An approach for reducing the time complexity is possible by creating small block size, parallel computation of the motion vectors [35, 36, 37]. These approaches may help in speeding up the process of finding the motion vectors.
VII. DISCUSSION AND CONCLUSION

Obtaining correct track of the object of interest is crucial in object tracking approaches. This paper summarizes the different approaches to object detection and tracking followed by the block matching algorithm for motion estimation. The different stages/phases of object tracking have been studied. Various methods for object detection are frame difference, optical flow, and background subtraction. Object tracking can be done by using Kalman filter, particle filter, etc. The background subtraction is a simple and efficient technique to provide the complete information regarding the object of interest. Also, motion plays very important role in detecting the object. Among all the motion estimation methodologies, the block matching received very much attention in terms of their simplicity and regularity. The different block matching motion estimation techniques are studied: FS, TSS, NTSS, FSS, DS, LDS, etc and we conclude that the object tracking can be upgraded and improved by applying the block matching technique.

VIII. FUTURE SCOPE

Developing a new way of tracking object from the video scene by combining the object detection approach with the block matching algorithm gives a scope of improvement to existing object tracking techniques. An extended version of the block matching algorithm can be designed for finding out the motion vectors of the object from the video and specific focus on faster tracking of the object. In future, we can develop the block matching algorithm by combining the existing block matching algorithm for increasing its efficiency to track the objects in a video.

REFERENCES


