A survey of Detection and Tracking of Moving Objects in Video Frames

Bedekar K.A.

P.G. Student, Dr. J. J. Magdum College of Engg, Jaysingpur, Maharashtra

Dr. Patil V.V.

Professor, Electronics Dept, Dr. J. J. Magdum College of Engg, Jaysingpur, Maharashtra

Abstract - In many computer vision applications object detection and tracking are the most important and challenging tasks. Some of the applications are video surveillance, vehicle navigation, and autonomous robot navigation etc. Designing of the efficient video surveillance system in complex environments is the most significant task. The detection of moving objects from a video is crucial for object detection, target tracking, and behavior understanding. Background subtraction, foreground segmentation are primary and popular steps in detection of moving objects in video streams. In this paper, we have considered various object detection and tracking methods for study and comparison. Here, the problem of illumination variation, background clutter and shadows should also be considered while examining various object detection and tracking techniques.

Keywords: Object detection, tracking, background subtraction, foreground segmentation

I.   INTRODUCTION

Object detection is a method of following an object through successive image frames to determine its relative movement with respect to other frames.Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition[1].

Object tracking is a process of segmenting a region of interest from a video and keep tracking of the motion, position and also the occlusion. The tracking is performed by monitoring objects' temporal and spatial changes during the video sequence, also its presence, position, size, shape, etc. [2].

Actually videos are the sequences of images, and here each image is called a frame, displayed in fast enough frequency so that human eyes can percept continuity of the content. It is obvious that all image processing techniques can be applied to individual frames.The contents of the two consecutive frames are usually closely related [3].

II.   CHALLENGES IN OBJECT DETECTION AND TRACKING:

The detection and tracking of objects in video can be complex due to:
  a. Loss of information caused by projection of the 3D world on a 2D image,
  b.   Noise in images,
  c. The complex object motion,
  d. Non-rigid or the articulated nature of the objects,
  e. Partial and full object occlusions,
  f. Scene illumination changes
  g. Complex object shapes,
  h. The real time processingrequirements [4].

The basic steps involved in this task are as follows:

```
┌─────────────────────────┐
│      Input Video         │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│ Extract frames from input│
│          video           │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│     Object Detection     │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│   Object Representation   │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│      Object Tracking     │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│      Moving Object       │
└─────────────────────────┘
```
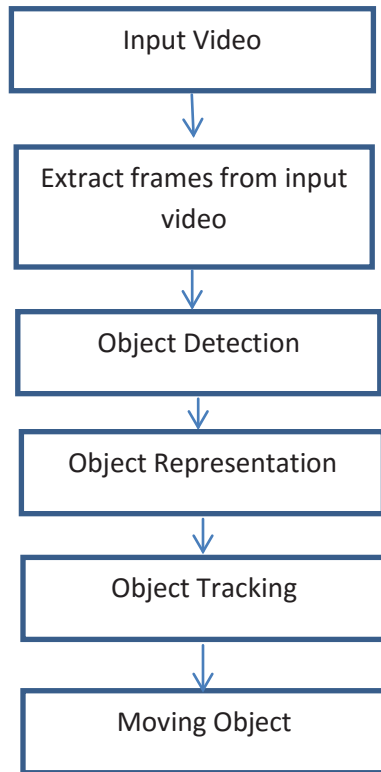
Fig.1 Steps in Object Detection and Tracking

III.   FEATURES OF OBJECT DETECTION AND TRACKING

The feature selection is done first which plays a critical part in detection and tracking. It is strictly associated to the object representation. One of the examples is: color is used as a feature for histogram based representations, and for contour based representations, the object edges are usually used as features.

3.1. *Edges:* The strong changes in image intensity values are generally observed at object boundaries. These changes can be recognized by edge detection. The edges are not less sensitive to illumination changes than that of the color features. Most of the object detection and tracking algorithms use edge detection as main procedure.

3.2. *Color***:** Different color models or color spaces are the vital part of various video frame formats. The data from the different frames can be stored in dissimilar color spaces ranging from grayscale, RGB, HSV, YIQ and YCbCr color spaces. In terms of brightness information in each spectral band, the data is stored in each frame. Generally, the color images are represented as Red i.e. R, Green i.e. G and Blue i.e. B layers or RGB color space. The other characteristic color spaces used in Object detection and tracking are HSV, YCbCr [3].

3.3. *Texture:* The texture is the degree of intensity dissimilarity of a surface that considers the properties like smoothness and regularity. This feature requires a processing step as compared to color space model.

The texture features are not more sensitive to the illumination changes as compared to the edge features.

IV.   OBJECT REPRESENTATION

An object is nothing but an entity of interest. The objects can be represented by their shapes and appearances. Some of the examples are, planes in the air, boats on the sea, vehicles on a road, fish in an aquarium, people walking on a road.

There are various representations of object shape, which is commonly used for tracking and then addresses the joint shape and appearance representations in describing as follows.

4.1 Points: The object can be represented by a point, which is centroid or set of points (Fig 2 (a), (b)). For the objects occupied small regions in an image, the point representation is suitable.

4.2 Primitive geometric shape: The object shape is represented by geometric shapes such as rectangle, ellipse etc. (Fig 2 (c), (d)). The primitive geometric shapes are more suitable for representing simple rigid objects as well as non-rigid objects.

4.3 Object silhouette and contour: The boundary of an object is known as contour. (Fig 2(g), (h)). The region inside the contour is known as the silhouette of the object (Fig 2 (i)). These representations are suitable for tracking of the complex non rigid shapes.

4.4 Articulated shape models: The articulated objects are composed of the body parts that are held together by joints (Figure 2 (e)). The example is the human body. It is an articulated object with legs, hands, torso, head, and feet connected by joints.
   The relationship between the parts is governed by kinematic motion models,for example, joint angle, etc.

4.5 Skeletal model: The object skeleton can be obtained by applying the medial axis transform to the object silhouette. This model is used as a shape representation for recognizing objects. The skeleton representation can be used to model both articulated and rigid objects (see fig 2 (f)).
   In the same way, there are various ways to represent the appearance features of objects.
   Shape representation can be combined with the appearance representations for tracking. Some of the common appearance representations in the case of object tracking are given as follows.

4.6 Probability densities of object appearance: The probability densities of object appearance features (color, texture) can be calculated from the image regions specified by the shape models (the interior region of ellipse or contour).

4.7 Templates: Templates are formed using the simple geometric shapes or silhouettes.It carries both spatial and appearance information. But the templates encode only the object appearance generated from a single view. In this way, they are suitable only for tracking the objects whose poses don't vary much during the course of tracking.

4.8 Active appearance models: These models are generated by modeling the object shape and appearance. The object shape can be defined by a set of landmarks. Each landmark i.e. an appearance vector is stored in the form of texture, color, or the gradient magnitude. These models require a training phase where both shapes & its associated appearance can be learned from a set of samples.
4.9 Multi-view appearance models: These models encode the different views of anobject. One method to represent different object views is to generate a subspace from the given view. The example of subspace approaches is PCA - Principal Component Analysis, ICA - Independent Component Analysis. The limitation of multi-view appearance models is that the appearances in all views have required a lot of time.

   From the above discussion, we can say that there is a strong relationship between the object representation and the tracking algorithms. Object representations are chosen in accordance with the application area.
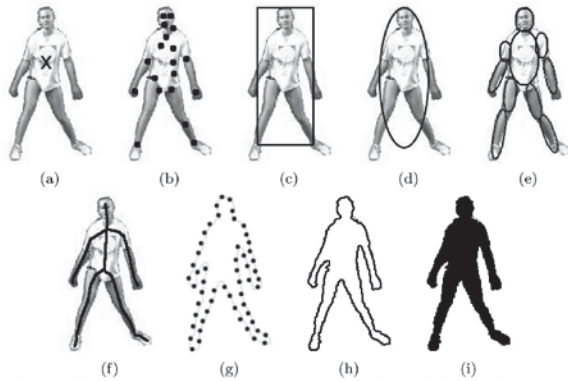
Fig. 2Object representations (a) Centroid, (b) multiple points, (c) rectangularpatch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g)complete object contour, (h) control points on object contour, (i) object silhouette.

V.   OBJECT TRACKING METHODS

The video surveillance system include task such as motion detection, tracking, and activity recognition. Detection of moving object is the first important step and successful segmentation of moving foreground object from the background ensures object classification, personal identification, tracking, and activity analysis, making these later steps more efficient. We can categorize the motion detection into major classes such as frame differencing, background subtraction.

Some of the methods used for object tracking are as given below:

5.1 Background Subtraction: This method is used for segmenting foreground object from the background. The necessary step for background subtraction is background modeling. The Background Modeling must be sensitive enough to recognize the moving objects [5]. To yield reference model, Background Modeling is done. This reference model is used in background subtraction. Here, each video sequence is compared with the reference model to determine the possible variation. The variations between current video frames and the reference frame in terms of pixels signify the existence of moving objects.

*5.2 Meanshift Algorithm:*MeanShift algorithm is a robust non-parametric technique for getting density gradients to find the mode or peak of the probability distribution [6]. This means that if given a set of data-points the algorithm associates each point with nearby maxima of the underlying probability density function and iterates a predefined search window over the data. After each iteration, search window moves to a denser region until it is centered over the maxima.The mean-shift algorithm can be used for finding the location of the object in static images.

5.3 CAMShift algorithm: It was originally developed for a head and face-tracking applications [6] based on the color histogram of the target object. The CAMShift algorithm is improvement over the traditional mean shift algorithm that finds the peak of a probability distribution generated from the color histogram. The CAMShift algorithm is used as an important part of other applications such as head and face tracking applications, human computer interaction, and video surveillance. The location and size of search window is initialized to include area containing the target object in a captured image. The hue value for each pixel in the search window is sampled to generate probability density function. This density function is saved as the histogram model of the target object. Then in the next step, estimation of the probability distribution map is done. To determination of the probability of that pixel belonging to the target object, the histogram model scans through each pixel of the captured scene.

5.4 Kalman Filter: The Kalman Filter is Optimal Recursive Data Processing Algorithm based technique. It consists of mathematical equations used to provide an efficient computational (recursive) for estimation of the state of a process in several aspects: it can support estimations of past state, present state and future states. It can estimate even when there is the unknown nature of the modeled system. This filter also estimates a process by using a form of the feedback control. The equations for Kalman filters [5] can be categorized in two groups: time update equations and measurement update equations. The time update equations for projecting forward i.e. in time the current state and error covariance estimates to obtain the priori estimation of the next time step. Kalman filters always give optimal solutions.

5.5 *Particle Filter:* The particle filter [5] generates all the models for one variable before moving to the next variable. The particle filter algorithm has an advantage when variables are generated dynamically and there can be unboundedly numerous variables. It also includes the new operation: resampling. One restriction of the Kalman

filter is the assumption of state variables are normally distributed i.e. Gaussian. Thus, the Kalman filter is poor approximations of state variables which are not Gaussian distribution. This restriction of Kalman filter can be overcome by using the particle filter. This algorithm usually uses color features, contours, or texture mapping. The particle filter [5] is a Bayesian sequential importance sample technique and it recursively approaches the later distribution using a finite set of weighted trials. It also consists of basically two phases: prediction and update which are same as the Kalman Filtering. It is applied in areas such as computer vision and applied to tracking problematic scenes.

VI.   CONCLUSION

In this paper, we present a survey of object detection and tracking. Various phases of object tracking, five popular object detection methods are addressed here. The Background Subtraction method is used for extracting foreground object from the background. It is the basic algorithm used for object extraction. The MeanShift algorithm is used for obtaining the density gradients which are further used to find the mode or peak of the probability distribution. It works fine on static probability distributions but not on dynamic ones as in a video. CAMShift is based on principles of the MeanShift and it is used for the dynamically changing distributions in videos. It is able to handle the dynamic distributions in frames by readjusting the size of search window for the next frame based on zeroth moment of current frames distribution. This allows the algorithm to anticipate object movement to quickly track the object in the next scene. Also, during quick movements of an object, CAMSHIFT is able to track the object correctly. Kalman Filters are linear quadratic estimators i.e. they are best for calculating linear systems with Gaussian noise. In non-Gaussian environments, we have to use Particle Filters. Depending on our object tracking application we can use one or more methods from the above discussed methods.

REFERENCES

[1]   ManishaChate,S. Amudha, VinayaGohokar, *"Object Detection and Tracking in Video Sequences"*,   ACEEE Int. J. on Signal & Image Processing, Vol. 03, No. 01, Jan 2012, pp: 36-41
[2]   GandhamSinduja, Dr. Renuka Devi S. M., "*A Survey on Detection and Tracking of Objects in Video Sequence",* International Journal of Engineering Research and General Science Volume 3, Issue 2, Part 2, March-April, 2015ISSN 2091-2730 pp:418-426
[3]   Abhishek Kumar Chauhan, PrashantKrishan, *"Moving Object Tracking Using Gaussian Mixture Model And Optical Flow"*, International Journal of Advanced Research in Computer Science and Software Engineering*,* April 2013 pp:243-246
[4]   Yilmaz A, Javed O, Shah M. *"Object Tracking: A Survey"*, ACM Comput. Surv., 2006; 38(4):13.pp: 1-46
[5]   J.Joshan Athanesious, P.Suresh, ―*"Systematic Survey on Object Tracking Methods in Video"*, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) October 2012,pp 242-247.
[6]   Yang Li, Hui Ding, Shudong Zhang, Yilei Wang*, "The Research and Implementation of Camshift Algorithm Based on Automatically Target Extraction",* Journal of Convergence and Information Technology, Vol. 8(6), pp 935-941
[7]   P. Imrankhan, C. Chetan*"A Survey on Moving Object Detection and Tracking Methods"*, International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, pp 5212-5215