MOVING OBJECT DETECTION AND TRACKING IN VIDEO

MODI NIRAV D.
Department of Electronics, Bharati Vidyapeeth Deemed University College of Engineering, Pune

Abstract—Object tracking means to track an object over a sequence of images. Video tracking is to associate target objects in consecutive video frames. Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition object tracking is to monitor and objects spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc.

I. INTRODUCTION

As we know that videos are made up of subsequent images (frames) which are move fast enough. So that, human eyes realize them continuous. Now, for any processing on video we need to the frame. The video analysis is done by following three steps:
1. Detection of object
2. Tracking of that object
3. Analysis of that object

Visual content can be modeled as follow:
1. At first level, it gives the information of color or brightness with raw pixel.
2. At second level, it gives information of edges, corners, lines, curves, and color regions.
3. At higher level, these levels are combined and made objects.

II. STEPS INVOLVED IN OBJECT TRACKING

The main task of object tracking is to foreground object detection, object classification and moving object tracking.

A. FOREGROUND OBJECT DETECTION:
The common first step in the process of object tracking is to identify objects of interest in the video sequence and to cluster pixels of these objects. Since moving objects are typically the primary source of information, most methods focus on the detection of such objects.

Object detection can be done by following techniques:
1) Frame differencing
2) Optical flow
3) Background subtraction

1. Frame differencing:
The presence of moving objects determined by calculating the difference between two consecutive images. Its calculation is simple and easy to implement. For a variety of dynamic environments, it has a strong adaptability, but it is generally difficult to obtain complete outline of moving object, responsible to appear the empty phenomenon, as a result the detection of moving object is not accurate.

2. Optical flow:
Optical flow method is to calculate the image optical flow field, and do clustering processing according to the optical flow distribution characteristics of image. This method can get the complete movement information and detect the moving object from the background better, however, a large quantity of calculation, sensitivity to noise, poor anti-noise performance, make it not suitable for real-time demanding occasions.

3. Background subtraction:
First step for background subtraction is background modelling. Background Modeling to yield reference model. This reference model is Used in background subtraction in which each video sequence is compared against the reference model to determine possible Variation. The variations between current video frames to that Of the reference frame in terms of pixels signify existence of Moving objects The background subtraction method is to use the difference method of the current image and background image to detect moving objects, with simple algorithm, but very sensitive to the changes in the external environment and has poor anti-
interference ability. However, it can provide the most complete object information in the case background is known. Background subtraction have mainly two approaches:

1) Recursive algorithm
2) Non-recursive algorithm

![Background Subtraction](image)

**i. Recursive Algorithm:**

1) Approximate median method

The running estimate of the median is incremented by one if the input pixel is larger than the estimate, and decreased by one if it is smaller. The disadvantage of this method is that it does not provide smoother results in all conditions.

2) Adaptive background method

This method is based on gain values. Typical value of gain values $\alpha = 0.1$ to $0.9$ only. The main advantage of this algorithm is that the increased efficiency of the segmented results by the gain values. Also auto threshold is included in this method.

3) Mixture of gaussians

The background model is parametric. Each pixel location is represented by a number (or mixture) of gaussian functions that sum together to form a probability distribution function. Even though the mixture of gaussian method is complex and tedious, it provides better results by our modified parameters.

**ii. Non-recursive algorithms**

1. Median filtering:

Median filtering is the commonly used technique to find foreground images. The pixel parameters minimum, maximum, and largest inter-frame absolute difference are used. The main advantage of this method is that, the computation speed is high.

**C. Object Tracking**

Tracking can be defined as the problem of approximating the path of an object in the image plane as it moves around a scene. The purpose of an object tracking is to generate the route for an object above time by finding its position in every single frame of the video.

Object is tracked for object extraction, object recognition and tracking, and decisions about activities. Object tracking can be classified as point tracking, kernel based tracking and silhouette based tracking. For illustration, the point trackers involve detection in every frame; while geometric area or kernel based tracking or contours-based tracking require detection only when the object first appears in the scene.

Classification is based on:

- Which Features of image should be used? Features can be color, edges, textures etc.
- How should the motion, appearance, and shape of the object be showed?

Tracking methods can be divided into following categories:
**Point Tracking Approach**

In an image structure, moving objects are represented by their feature points during tracking. Point tracking is a complex problem particularly in the incidence of occlusions, false detections of object. Recognition can be done relatively simple, by thresholding, at of identification of these points.

The cost of correspondences is mainly by Kinematic limits.

1. Proximity assumes the position of the object will not change especially from one frame to other.
2. Extreme velocity limits the possible correspondences to the neighborhood around the object in circular manner.
3. Minor velocity change (even motion) assumes the direction and speed of the object does not change hugely.
4. Mutual motion: velocity of objects in a small neighborhood to be similar with regular interval of duration as shown in figure 4. This constraint is appropriate for objects represented by multiple points.

![Figure 4 Point base Tracking](image)

Point Tracking is capable of dealing with:
- Tracking very small objects.

**Kalman Filter:**

They are based on Optimal Recursive Data Processing Algorithm. The Kalman Filter performs the restrictive probability density propagation. The Kalman filter is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process in several aspects: it supports estimations of past, present, and even future states, and it can do the same even when the precise nature of the modeled system is unknown.

The Kalman filter estimates a process by using a form of feedback control. The filter estimates the process state at some time and then obtains feedback in the form of noisy measurements. The equations for Kalman filters fall in two groups: time update equations and measurement update equations. The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the priori estimate for the next time step. The measurement update equations are responsible for the feedback.

It is necessary to update the background image frequently in order to guarantee reliable object detection. The basic idea in background adaptation is to integrate the new incoming information into the current background image using a Kalman filter:

$$ B_{n+1} = B_n + (u_{1} \times (1 - \alpha_2) + \alpha_2 \times M_2) \times D_2 $$

Where, $B_n$ represents the background model at time $n$ is the difference between the present frame and the background model, and $M_2$ is the binary moving objects hypothesis mask. The gain $u_{1}$ and $u_{2}$ are based on an estimate of the rate of change of the background. The larger it is, the faster new changes in the scene are updated to the background frame.

**Kalman Tracking** is capable of dealing with:

1. Kalman filters always give optimal solutions.
2. Another potential approach is to handling noise $S$.
3. Tracking is applicable only for single and multiple objects.

**Particle Filter:**

One restriction of the Kalman filter is the assumption of state variables are normally distributed (Gaussian). Thus, the Kalman filter is poor approximations of state variables which do not Gaussian distribution. This restriction can be overwhelmed by using particle filtering.

This algorithm usually uses contours, color features, or texture mapping. The particle filter is a Bayesian sequential importance Sample technique, which recursively approaches the later distribution using a finite set of weighted trials.

It also consists of fundamentally two phases: prediction and update as same as Kalman Filtering. It was developing area in the field of computer vision communal and applied to tracking problematic and is also known as the Condensation algorithm. In the method of staring with population of particle, each will assign value to no variables and weight of 1.

At each step, the procedure will be:

1. Select a variable that has not sampled. For each particle, it samples the variable according to some proposal distribution. The weight of the particle is reorganized as in importance sampling.
2. Select a piece of evidence to absorb which is not absorbed previously. Weight of the particle is increased.
3. Population should be resampled. Resampling creates a new population of particles, each with the similar weight. Some particles may be elapsing or repeated.

**Multiple HYPOTHESES Tracking (MHT):**

In MHT algorithm, several frames have been observed for better tracking outcomes MHT is an
iterative algorithm. Iteration begins with a set of existing track hypotheses. Each hypothesis is a crew of disconnect tracks. For each hypothesis, a prediction of object’s position in the succeeding frame is made. The predictions are then compared by calculating a distance measure.

MHT is capable of dealing with:
1. Tracking multiple object
2. Ability to tracks for objects entering and exit of Field Of View (FOV).
3. It also handles occlusions.

**Kernel Based Tracking Approach:**

Kernel tracking is usually performed by computing the moving object, which is represented by an embryonic object region, from one frame to the next. The object motion is usually in the form of parametric motion such as translation, conformal, affine, etc.

These algorithms diverge in terms of the presence representation used, the number of objects tracked, and the method used for approximation the object motion. In real-time, illustration of object using geometric shape is common. But one of the restrictions is that parts of the objects may be left outside of the defined shape while portions of the background may exist inside. This can be detected in rigid and non-rigid objects. They are large tracking techniques based on representation of object, object features, appearance and shape of the object.

**Simple Template Matching:**

Template matching is a brute force method of examining the Region of Interest in the video. In template matching, a reference image is verified with the frame that is separated from the video. Tracking can be done for single object in the video and overlapping of object is done partially. Template Matching is a technique for processing digital images to find small parts of an image that matches, or equivalent model with an image (template) in each frame. The matching procedure contains the image template for all possible positions in the source image and calculates a numerical index that specifies how well the model fits the picture that position.

It capable of dealing with:
1. Tracking single image.
2. Partial occlusion of object.
3. Necessity of a physical initialization.

**Mean Shift Method:**

Mean-shift tracking tries to find the area of a video frame that is locally most similar to a previously initialized model. The image region to be tracked is represented by a histogram. A gradient ascent procedure is used to move the tracker to the location that maximizes a similarity score between the model and the current image region. In object tracking algorithms target representation is mainly rectangular or elliptical region. It contain target model and target candidate. To characterize the target color histogram is chosen. Target model is generally represented by its probability density function (PDF). Target model is regularized by spatial masking with an asymmetric kernel.

**Support Vector Machine (SVM):**

SVM is a broad classification method which gives a set of positive and negative training values. For SVM, the positive samples contain tracked image object, and the negative samples consist of all remaining things that are not tracked.

It capable of dealing with:
1. Tracking single image.
2. Partial occlusion of object.
3. Necessity of a physical initialization.
4. Object motion by translation.

**Layering based tracking:**

This is another method of kernel based tracking where multiple objects are tracked. Each layer consists of shape representation (ellipse), motion such as translation and rotation, and layer appearance, based on intensity. Layering is achieved by first compensating the background motion such that the object’s motion can be estimated from the rewarded image by means of 2D parametric motion. Every pixel’s probability of calculated based on the object’s foregoing motion and shape features.

It capable of dealing with:
1. Tracking multiple images.
2. Fully occlusion of object.
3. Object motion by translation, scaling and rotation.

**Silhouette Based Tracking Approach:**

Some object will have complex shape such as hand, fingers, shoulders that cannot be well defined by simple geometric shapes. Silhouette based methods afford an accurate shape description for the objects. The aim of a silhouette-based object tracking is to find the object region in every frame by means of an object model generated by the previous frames. It capable of dealing with:
1. Handling of large variety of object shapes easily.
2. Handling Occlusion
3. Dealing with object split and merge

Contour Tracking:
Contour tracking methods, iteratively progress a primary contour in the previous frame to its new position in the current frame. This contour progress requires that certain amount of the object in the current frame overlay with the object region in the previous frame. Contour Tracking can be performed using two different approaches. The first approach uses state space models to model the contour shape and motion. The second approach directly evolves the contour by minimizing the contour energy using direct minimization techniques such as gradient descent. The most significant advantage of silhouettes tracking is their flexibility to handle a large variety of object shapes.

Shape Matching:
These approaches examine for the object model in the existing frame. Shape matching performance is similar to the template based tracking in kernel approach.

Another approach to Shape matching is to find matching silhouettes detected in two successive frames. Silhouette matching, can be considered similar to point matching.

Detection based on Silhouette is carried out by background subtraction. Models object are in the form of density functions, silhouette boundary, object edges. Capable of dealing with:
1. Edge based template, Silhouette tracking feature of shape matching are able to track only single object.
2. Occlusion handling will be performed in with Hough transform techniques.

CONCLUSION
In this paper, we discussed the various methods to track the objects in brief. We divide tracking method in three categories namely, point tracking, Kernel tracking and Silhouette tracking.

REFERENCES