Object Tracking using Spatio-Temporal Information of Video Data Cuboid

Aradhana Kushwaha
Department of Computer Science and Engineering, PSIT
Kanpur, Uttar Pradesh, India

Awanish Mishra
Department of Computer Science and Engineering, PSIT
Kanpur, Uttar Pradesh, India

ABSTRACT

Object tracking in a video sequence is a challenging Problem. The difficulties in the object tracking are arises due to its motion, shape, size and speed. In this paper, a new method has been introduced for object tracking. Generally, the tracker needs an initialization, often done manually or by object detection, but in present approach there is no need to initialize the tracker.

A video can be represented as 3D data cuboid having spatial (X - Y axis) and temporal information (T axis). This cuboid is can be represented as continuous Y number of XT frame. For video having stationary background, the horizontal lines in XT frames are due to static scene. The inclined lines in XT frames are due to linear moving object with constant speed. Thus the information containing object motion in XY frames correspond to inclined line in XT frames. Depending upon the size of the object these inclined lines can be thin or thick. For small and large object it will be thin and thick respectively. Hough transform based line detection algorithm is used to extract these inclined lines. Binary edge map of XT frame is obtained by Canny edge detection algorithm which is used by Hough transform. Two nearby and parallel line is appeared due to the large object. The morphological operations are used to combine these lines into a single thick line. These lines correspond to motion due to a single object.

If the object is not moving with constant speed, there will be curve in XT frame due to object motion. A curve is set of consecutive line. Using Hough transform based line detection algorithm, curve trajectory can be trajectory, which correspond to the object motion in XY frame.

General Terms

Object detection, Tracking

Keywords

Canny edge detection, Hough transform, morphological operation, erosion and dilation

1. INTRODUCTION

Tracking can be defined as the problem of approximating the trajectory of an object. A tracker allocates consistent labels to the tracked objects in the frames of a video. It can provide object-centric information depending on the tracking domain i.e location, area, or contour of an object [1]. The rise of high-powered computers, the accessibility of high quality and low-cost video cameras, and the growing need for programmed video analysis has produced a great deal of interest in object tracking algorithms. It is an essential part of computer vision [9] [2].
Application of object tracking is relevant in the area of,

—Video indexing, in multimedia databases, automatic explanation and recovery of the videos.

—Automated surveillance, that is, to detect mistrustful actions observing a scene or doubtful events.

—Traffic monitoring, that is, for direct traffic flow real-time assembly of traffic figures.

—Vehicle navigation that is, hurdle prevention abilities video-based path scheduling.

—Motion-based recognition, that is, automatic object detection, human identification based on way of walking.

—Human-computer interaction, that is, motion recognition, eye observation, tracking for data input to processors.

The problem in object tracking can rise due to unexpected object motion, varying appearance arrangements of the object and the scene, non ridged object structures, object-to-object and object-to-scene occlusions, processing requirements in real time, camera motion and data loss by projection of 3D world to 2D image.
The tracking method is categorized on the basis of object and motion representation. Based on shape and appearance of track, object can be represented as centroid, rectangular patch, elliptical patch, multiple points, part based multiple patches, complete object contour, object skeleton, control point on object contour and object silhouette depending on tracking method used.
Feature selection is another very important step of object tracking. To identify the unique feature, a criterion like color, edges, motion and texture is used. After identify the unique feature from the images the important step, object detection mechanism are applied. For the object detection the following methods area used like point detectors, background subtraction, segmentation and supervised learning [4]. After detecting the objects tracking mechanism are applied to track the object .These methods are applied to track the object like point tracking, kernel tracking [6] and, silhouette tracking [1].

2. METHODLOGY

2.1 Spatio-temporal Information of Video Data Cuboid

Basically, video is a collection of sequence of XY frame. The sequence number of XY frame represents T axes. Thus a video V_{XYT} can be represented with three axes X,Y and T. X and Y represents column and row and T represent number of such XY frame. By taking any two axes we can form a spatial frame and other axes can be temporal axis [10].

$$V_{XYT} = \{I(t)_{XY}, \forall t\epsilon\ \{1,2,3.......p\}\} \tag{1}$$

Where $I(t)_{XY}$ is a XY frame of video sequence at any particular time t on T axis. Total number of frame is p.

The data cuboid V_{XYT} can represented as XT slices $I(y)_{XT}$.

$$V_{XYT} = \{I(y)_{XT}, \forall y\epsilon\{1,2,3...........m\}\} \tag{2}$$

Where $I(y)_{XT}$ is a XT frame of video sequence at any row y and m is the total number of rows.

The data cuboid V_{XYT} is represented as YT slices $I(x)_{YT}$.

$$V_{XYT} = \{I(x)_{YT}, \forall x\epsilon\{1,2,3...........n\}\} \tag{3}$$

Where $I(x)_{YT}$ is YT frame of video sequence at any column x and n is the total number of column. This slicing operation is shown in fig 1 [7].

2.2 Spatio-temporal Information of XT slice

(a) Frame in a XY plane (traditional frame)

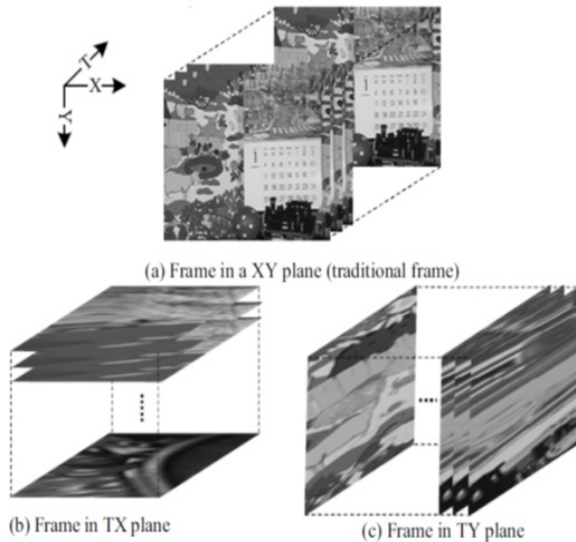(b) Frame in TX plane

(c) Frame in TY plane

Fig. 1. Slicing Operation

In the XT frames, there are several horizontal lines. These horizontal lines appeared due to static background of the scene in the XY frame. The curve appeared in XT frame is due to trajectory of moving object. If object speed is constant and linear, then there will be inclined line instead of curve in XT frames. Depending on object pixels contain in X direction, XT slice will have thick inclined line. The most important result is that the trajectory of object motion

in XY frame is contained in few XT slices. This trajectory can be approximated by combining a number of inclined line segments. To extract trajectory of motion we have to extract all inclined line segments in few XT slice. To extract the trajectory, take those XT slices which are having inclined line. Hough transform based line detection technique can be used to detect these line.

2.3 Canny edge detection

Basically edges are formed due to change in intensity of pixels, discontinuities in depth, variation in scene brightness. The main objective of edge detection to reduce the huge amount of data while preserving the structural properties of an image. The Canny edge detection algorithm is become very old but it is a very efficient and effective method of edge detection. Thus, the Canny edge detection algorithm is used in current research work. There are some points should always remember if we perform edge detection algorithm:

—Detection: There should be highest probability to detect the real edge and lowest probability to detect the falsely edge.

—Localization: Less error between detected edge and the real edge.

—Number of responses: here are only one real edge than the edge detection algorithm should detect only one edge not more than that. [3]

Canny's algorithms have these following steps.

(1) **Smoothing** To remove the noise, blurring operation is performed with the help of Gaussian filter. The 2D Gaussian function is as follows:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2}e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{4}$$

(2) **Finding gradients** Gradient is basically first derivative. Calculate the Gradient in both X and Y direction of an image. Gradient is a scalar vector thus magnitude and direction should be calculated and marked the edge where the magnitude is largest. First order derivatives of Gaussian function in x direction

$$\frac{\partial G(x,y,\sigma)}{\partial x} = -\frac{x}{2\pi\sigma^4}e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{5}$$

First order derivatives of gaussian function in y direction

$$\frac{\partial G(x,y,\sigma)}{\partial y} = -\frac{y}{2\pi\sigma^4}e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{6}$$

(3) **Non-maximum suppression** We calculate the local maxima and marked these edge.

(4) **Hysteresis Thresholding** By the suppression of edges which are not connected with strong edge ,the final edge is determined.

Using the edge map found by canny detector, these lines can be extracted by Hough tranform.

2.4 Hough transform based line detection

(1) Parameter Space (c, m)

—The equation of line is :

$$y = mx + c \tag{7}$$

,where m is slope and c is interception on y axis.

—After rearrangement it is written as,

$$c = (-x)m + y \tag{8}$$

—For any particular edge point i the equation becomes: c = (-xi)m + yi

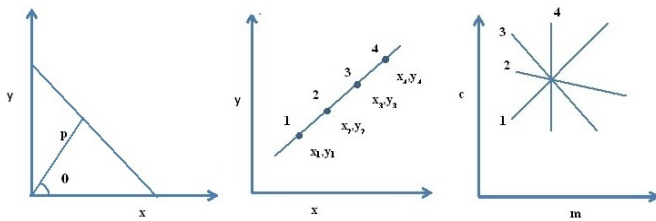—Now the equation of line is in parameter space space as shown in Fig 2.

Fig. 2. Mapping of collinear point of xy plane in mc plane

There is problem for vertical lines m and c grows to infinity so the polar form of equation of line is used. The polar form of line equation is:

$$p = x \cos \theta + y \sinh \theta \qquad (9)$$

as shown in fig 2. Now the Hough transform is calculated for polar form of the line equation and the steps are as follows:

—Quantize the parameter space $P[\theta_{min} \dots \theta_{max} , p_{min} \dots p_{max}]$.

—For every edge point (x,y) calculate the $p = x \cos \theta + y \sinh \theta$.

—$P[\theta, p] = [\theta, p] + 1$.

—In the parametric space find the local maxima.

Hough transform extracts two parallel and nearby lines are appeared due to motion of large object. These lines represent motion of single large object [11] [5].

2.5 Thick line detection

After the extraction of trajectory two lines are appeared. The two lines are detected due to the large size of object. For extracting the track the lines has to merge into a single line. This can be done using morphological operation. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. With the help of dilation operation the gap between the two lines are filled. But this will add some pixel on the outside boundary of the parallel lines. This can be removed by erosion operation. This process is called morphological closing. Then the two parallel lines are converted into one thick line [8].

3. RESULT

Fig 3 shows the result of slicing and extraction of track in XT frame. Fig 5 shows labeled track in red color in XY frame. These tracking result demonstrates tracking of multiple and moving objects.

In Fig 3(a), the XT frames are generated with the slicing operation of video data cuboid. The horizontal lines in XT frames are appeared due to the stationary object in XY frame and the inclined lines are appeared due to liner moving object. Due to change in
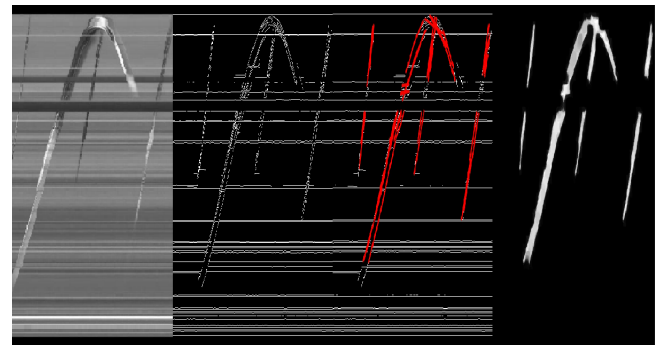

Fig. 3. (a) Frame of XT slice (b) Edge map obtained by Canny edge detector (c) Hough transform of binary image (d) Thick line obtained by morphological operation

velocity of object, the curve is appeared. Edge map of XT frame using Canny edge detector is shown in Fig 3(b) .Fig 3(c) shows detected lines using hough transform. The morphological operation(dilation, erosion) are performed on extracted line segment to detect thick line. Fig 3(d) shows thick line detection operation.

Fig. 4. Tracking result of video sequence 1 (a) Frame 172 (b) Frame 204 (c) Frame 390 (d) Frame 427

Fig. 5. Tracking result of video sequence 2 (a) Frame 290 (b) Frame 273 (c) Frame 390 (d) Frame 654

4. CONCLUSION

In this paper a new tracking technique is identified. A video can be represented as data cuboid. In this cuboid object's motion cues in XY plane is mapped to XT frame. If motion is linear, there will be inclined line in XT frame. To extract the trajectory in XY frame, inclined line extraction in XT frame is done. Hough transform with morphological operation is used to extract thick inclined line. Data cuboid form by extracted line segment will have object location in XY frame. There are some areas of interest for future research. In this approach, all XT frames are used for extracting the track information in XY frame. One can use inclined slicing operation on these video cuboid which can have complete trajectory of track. Inclined slicing operation can reduce complexity than XT slicing.

5. REFERENCES

[1] Omar Javed Alper Yilmaz and Mubarak Shah. Object tracking. 2006.

[2] Rita Cucchiara Simone Calderara Afshin Dehghan Mubarak Shah Arnold W. M. Smeulders, Dung M. Chu. Visual tracking: An experimental survey. *IEEE Transactions on pattern analysis and machine intelligence*, 2014.

[3] Shai Avidan. Support vector tracking. 2004.

[4] Shai Avidan. *Support Vector Tracking*. 2014.

[5] Christopher M. Brown. Inherent bias and noise in the hough transform. 1983.

[6] Visvanathan Ramesh Dorin Comaniciu and Peter Meer. Kernel-based object tracking. date.

[7] Josef Sivic Ivan Laptev Jose Lezama, Karteek Alahari. Track to the future: Spatio-temporal video segmentation with long-range motion cues.

[8] Pedro Henrquez Miguel Aleman-Flores, Luis Alvarez and Luis Mazorra. Morphological thick line center detection. 2010.

[9] Omar Javed Mubarak Shah and Khurram Shafique. Automated visual surveillance in realistic scenarios. 2007.

[10] Manoranjan Paul and Weisi Lin. Efficient video coding considering a video as a 3d data cube. *2011 International Conference on Digital Image Computing: Techniques and Applications*, 2011.

[11] Jun Zhou and Arturo Sanchez-Azofeifa Walter F. Bischof. Extracting lines in noisy image using directional information. 2006.