International Journal of Engineering and Applied Computer Science (IJEACS)
Volume: 02, Issue: 07, ISBN: 978-0-9957075-8-0, July 2017

Detection and Tracking of Objects: A Detailed Study

Kuldeep N. Shukla
Department of Electrical & Electronics Engineering
NITTTR, Bhopal, India

Anjali Potnis
Department of Electrical & Electronics Engineering
NITTTR, Bhopal, India

Prashant Dwivedy
Department of Electrical & Electronics Engineering
NITTTR, Bhopal, India

Shahbaz Soofi
Department of Electrical & Electronics Engineering
NITTTR, Bhopal, India

***Abstract*—Detecting and tracking objects are the most widespread and challenging tasks that a surveillance system must achieve to determine expressive events and activities, and automatically interpret and recover video content. An object can be a queue of people, a human, a head or a face. The goal of this article is to state the Detecting and tracking methods, classify them into different categories, and identify new trends, we introduce main trends and provide method to give a perception to fundamental ideas as well as to show their limitations in the object detection and tracking for more effective video analytics.**

Keywords-Detection, tracking, representations, descriptors, features.

I. INTRODUCTION

A visual surveillance environment attempts to detect, track, and identify objects from various videos, and usually to understand object behaviors and activities. For example, its purposes to automatically compute the flux of things at public areas such as stores and travel sites, and then accomplish congestion and analysis to support in track organization and targeted advertisement. Such systems would substitute the old-style surveillance setups where the number of cameras beats the capacity of costly human operators to monitor them.

Proceeding with a feature to high-level incident understanding method, there are three main steps of visual analytics: detection of objects [1], tracking of such objects and pointers from frame to frame, and estimating tracking results to describe and conclude semantic events and hidden phenomena. This analogical be extended to other applications with motion-based recognition, access control, video indexing, human and computer communication, and track monitoring and navigation. This paper reviews important characteristics of the detection and tracking steps to support a deeper appreciation of many applications. Suppose you are waiting for your turn in a shopping line at a busy store. You can simply sense humans and classify deferent things of their interactions. As with other tasks that our brain does simply, visual analytics has turned long out to be entwined for machines. Not amazingly, this is also a problem for visual insight. The main challenge is the problem of changeability. A visual detection and tracking system requirements to simplify across vast variations in object presence such due for a case to lookout, posture, facial expressions, lighting conditions, imaging quality or occlusions while preserving specificity to not claim everything it sees are objects of attention. In addition, these tasks should preferably be performed in real-time on conservative computing stages. In detection, motion changes and appearance signs can be used to differentiate objects, which classically reduces it quite easily, and tracking techniques are often activated by detection results. Grouping of statistical analysis of visual features and time-based motion information typically lead to more robust styles. For those systems which face noisy environments, however, tracking is recommended to be tracked by detection to gather sufficient statistic as sufficient track-before-detect algorithms propose. Also, tracking direct to choose detection areas, source and sink areas. In any case, it has been common in the past few years, to accept that deferent approaches are required for these deferent tasks. Here we take the hypothetical view that detection and tracking, rather than being two distinct tasks, represent two points in a spectrum of generalization levels [2].

II. OBJECT DETECTION

Object detection includes detecting instances of objects from a specific class in any image. The aim of object detection is to detect all instances of objects from a known class, such as cars, people or faces in any image. Typically, only a small number of instances of the object exist in the image, but there are many numbers of possible scenes and scales at which they can occur and that need to somehow be explored. Each detection is described with some form of posture information. This could be as simple as the location of the object, or the extent of the object defined in terms of a bounding box. In other conditions, the posture information is more detailed and covers the parameters of a linear or non-linear transformation. For example, a face detector may compute the locations of the eyes, nose and mouth, in addition to the bounding box of the face. The posture could also be defined by a three-dimensional transformation postulating the location of the object comparative to the camera.

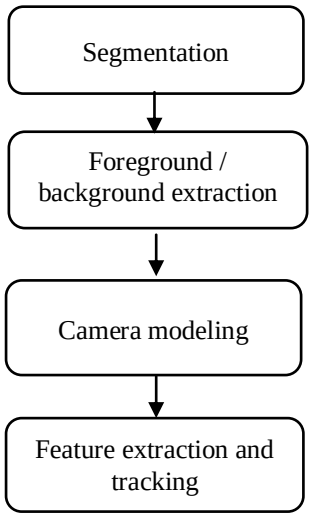
III. OBJECT TRACKING

Object tracking is defined as the procedure of segmenting an object from a video section and track its movement, direction, obstruction etc. to extract useful information [3].

Object tracking in video surveillance follows the separation step and is equivalent to the 'recognition' step in the image processing. Detection of moving objects in video streams is the first appropriate step of information abstraction in many computer vision applications, including traffic monitoring, remote video surveillance, and people tracking. There are fundamentally three approaches in object tracking. Feature based methods aim at extracting features such as points, line sectors from image sequences, tracking phase is then confirmed by a matching process at every time instant. Differential methods are based on the optical flow calculation, i.e. on the apparent gesture in image sequences, under some regularization expectations. The third class uses the relationship to measure temporary time of life shifts. Selection of an approach mainly depends on the domain of the problem.

Steps in object tracking

The object tracking process is summarized in the block diagram given below:



Basic steps in object tracking can be listed as:

- Segmentation

Segmentation is the procedure of finding components of the image. Segmentation includes procedures such as boundary detection, connected element labeling, thresholding etc

- Foreground / background extraction

This is the procedure of separating the foreground and background of the image. It is expected that foreground holds the objects of interest. In this method, we use subtraction of images to find objects that are moving and those that are not.

- Camera modeling

Camera model is a significant aspect of any object-tracking algorithm. The all present objects tracking systems use a camera model. In words camera model is directly derived from the domain knowledge.

- Feature extraction and tracking

The next step is to extract useful features from the sequence of frames. Depending on the algorithm, definition of 'feature' may vary.

IV. OBJECT DETECTION AND TRACKING APPROACHES OBJECT TRACKING

A. *Feature-based object detection*

Feature-based object detection contains correction of image structures and registering of mention points. The images may require to be changed to additional space for control modifications in clarity, size and arrangement. One or more types are removed and the objects of importance are modeled in terms of these types. Object detection and appreciation can be then changed into a chart similar problematic.

- Shape-based approaches

Shape-based object detection [4] is the solidest problems due to the trouble of segmenting objects of attention in the images. To sense and control the edge of an object, an image may must to be preprocessed. The preprocessing procedure or filter depends on the application. Changed object types such as publics, cars, and aircraft can need changed processes. For extra multipart units, sound deduction and exchanges invariant to scales and spin may be wanted. When the object is sensed and placed, its border can be found by edge finding and boundary-following processes.

- Color-based approaches

Different many other image types (e.g. shape) shade is comparatively constant below lookout changes and it is simply to be developed. Even if color is not at all times fit as the only means of sensing and tracking objects, but the small computational rate of the processes planned makes color a necessary feature to achievement when proper.

B. *Template-based object detection*

If a template relating an exact object is available, object detection becomes a procedure of similar types between the template and the image order in study. Object detection with an exact equal is normally computationally expensive and the quality of similar depends on the specifics and the degree of accuracy providing by the object template. There are dual types of object template same, stable and deformable template matching [6][5].

- Fixed template matching

Fixed templates are beneficial when object profiles do not variation with respect to the watching direction of the camera. Dual main techniques have been used in fix template matching.

Here technique, the template location is resolute from minimizing the space function between the template and many locations in the image.

Correspondence

Similar by correspondence uses the location of the normalized cross-correlation top between a template and an image to find the best match. This technique is normally immune to sound and lighting possessions in the images.

- Deformable template matching

Deformable template matching methods are other proper for cases where objects due to rigid and non-rigid bend. These dissimilarities can be produced by each the deformation of the object or just by different object position relation to the camera. Because of the deformable behavior of objects in maximum video, deformable models are other attractive in tracing tasks.

In this method [7], a template is denoted as a bitmap describing the specific outline/edges of an object figure. A probabilistic alteration on the prototype outline is applied to deform the template to set salient edges in the input image. An objective function with transformation limitations which correct the shape of the template is formulated imitating the cost of such transformations. The best main application of deformable template matching techniques is motion detection of objects in video edges.

C. Motion detection

Detecting moving objects, or motion detection, visibly has very important meaning in video object detection and tracking. A great quantity of research struggles of object detection and tracking absorbed on this problematic in last period. Equaled with object detection without motion, on single hand, motion detection complicates the object detection difficult by addition objects temporal modification requests, on the other hand, it also offers additional info source for detection and tracking.

A great variation of motion detection algorithms has been proposed. They can be categorized into the following sets almost.

- Thresholding technique over the interface difference

These methods depend on the detection of temporal variations either at pixel or block level. The difference map is usually binaries using a predefined beginning value to obtain the motion/no-motion organization.

- Statistical tests constrained to pixel wise independent decisions

These tests accept basically that the detection of temporal changes is equal to the motion detection. But, this assumption is legal when either large displacement appears or the object estimates are suitably textured, but be unsuccessful in the case of moving objects that preserve uniform regions. To duck this limitation, temporal change detection masks and filters have also been measured. The usage of these masks increases the efficiency of the change detection processes, especially in the case where various a priori information about the size of the moving objects is available, since it can be used to define the type and the size of the masks.

- Global energy frameworks

The motion detection problematic is formulated to minimalize a global objective function and is generally done using stochastic (Mean-field, Simulated Annealing) or deterministic reduction processes (Iterated Restricted Modes, Highest Confidence First). In that way, the spatial Markov Random Arenas have been widely used and motion detection has been measured as an arithmetical estimate problem. While this estimate is a very powerful, usually it is very time consuming.

V. THE TYPICAL KALMAN FILTER

The Kalman filter has widely used in engineering application. The Kalman filter has two characteristics. One is its mathematical model; it is described in terms of state-space concepts. The other is that its solution is computed recursively. Usually, the Kalman filter is described by system state model and measurement model.

The state-space model is described as

System state model: $s(t) = \ddot{o}(t-1)s(t-1) + \omega(t)$ (1)

and

Measurement model:

$z(t) = H(t)s(t) + v(t)$ (2)

where $\ddot{o}(t-1)$ and H(t) are the state transition matrix and measurement matrix respectively. The $\omega(t)$ and $v(t)$ are white Gaussian noise with zero mean and

$E\{w(k)w^T\} = Q\delta_{kl}$,

$E\{v(k)v^T\} = R\delta_{kl}$,

where δ_{kl} denotes the Kronecker delta function [8]; Q and R are covariance matrices of $w(t)$ and $v(t)$, respectively.

The state vector $s(t)$ of the current time t is predicted from the previous estimate and the new measurement $z(t)$.

The tasks of the Kalman filter have two phases: prediction step and correction step. The prediction step is responsible for projecting forward the current state, obtaining a priori estimate of the state $s^-(t)$. The task of correction step is for the feedback. It incorporates an actual measurement into the a priori estimate to obtain an improved a posteriori estimate $s^+(t)$. The $s^+(t)$ is written as

$$s^+(t) = s^-(t) + K(t)\big(z(t) - H(t)\, s^-(t)\big), \quad (4)$$

where $K(t)$ is the weighting and is described as

$$K(t) = P(t)^{-}H(t)^{T}\left(H(t)P^{-}(t)H(t)^{T} + R(t)\right)^{-1} = \frac{P(t)^{-}H(t)^{T}}{\left(H(t)P^{-}(t)H(t)^{T} + R(t)\right)} \quad (5)$$

In Eq. (10), the $P(t)^{-}$ is the priori estimate error covariance. It is defined as

$$P^{-}(t) = E[e^{-}(t)e^{-}(t)^{T}]$$

Where $e^{-}(t) = s(t) - s^{-}(t)$ t) is the priori estimate error. In addition, the posteriori estimate error covariance $P^{+}(t)$ is defined as

$$P^{+}(t) = E[e^{+}(t)e^{+}(t)^{T}]$$

where$e^{+}(t) \equiv s(t) - s^{+}(t)$ is the posteriori estimate error.

VI. Kernal Based Mean Algorithm

Mean shift is a non-parametric statistical method which was introduced for object tracking applications. To characterize the target, first a feature space is chosen. Then reference target model is represented by its probability density function (PDF) in the feature space. Similarly, a candidate model is represented with PDF function. A similarity density is calculated between the target model and candidate model to match the maximum likeness with the help of Bhattacharya coefficient ρ [p(x), q]. For example, the reference model can be chosen to be the color PDF of the target. [10] Without loss of generality, the target model can be considered as cantered at the spatial location 0. In the subsequent frame, a target candidate is defined at location y, and is characterized by the PDF p(y). Both PDFs are to be estimated from the data.

VII. Mean Shift Algorithm

Mean shift algorithm for moving object tracking was initially proposed in the estimation of Probability density function. Mean shift algorithm iteratively shifts a data point to the average of data point in its neighborhood. If we have distribution points. Then according to the mean shift algorithm modes or Peaks in density function is determined. This method is called non-parametric method this method of tracking tracks the object for long time and more robust compare to other tracking algorithm. To find the new location of the object that we are going to track, we need to find a vector which can suggest the direction of the moving object. This vector [11] is called mean shift vector First we need to draw the ROI around the object and get the data points, approximate location of the mean of this data. Then estimate the exact location of the mean of the data by determining the mean shift vector from the initial mean.

- In the first frame, tracking object is selected and object model has probability distribution of colour

Histogram [9]. If $y0$ is the centre of an object, then the position of pixels are $\{x *_i\}\ i = 1 \ldots N$, where N is the number of pixels in the image. statistical histogram distribution model of target area given by

$$q_h = C \sum_{I=1}^{n} k\left(||x_i||2\right)\delta[b(x_i^{*} - h)] \quad (6)$$

- At the current frame, the statistical histogram distribution given by

$$\widehat{p_h}(y_0) = C_h \sum_{i=1}^{nh} k(||\frac{y_0 - x_i}{w})\delta[b(x_i) - h] \quad (7)$$

(1)

- Computing the measurement between the object and candidate template by Bhattacharyya coefficient.

$$P(\widehat{p_h}(y_0), \widehat{q_h}) = \sum_{h=0}^{H-1} \sqrt{p_h(\widehat{y_0})\widehat{q_h}}$$

- Weight of the window of pixels in tracking window

$$w_i = \sum_{h=0}^{H-1} \delta[b(x_i) - h] \sqrt{\frac{\widehat{q_h}}{\widehat{q_h(y_0)}}}$$

- New object position search by mean shift value given by

$$y_i = \frac{\sum_{i=0}^{nh} x_i\, w_i g(|\left|\frac{y_0 - x_1}{W}\right| 2|\ )}{\sum_{i=0}^{nh} w_i\, g(|\left|\frac{y_0 - x_1}{W}\right| |2\ )}$$

Then computing the Bhattacharya coefficient given by

$$P(P_h(\widehat{y_1}), \hat{q}) = \sum_{h=0}^{H-1} \sqrt{\widehat{q_h}\,(y_1)\widehat{q_h}}$$

- Comparing coefficients and update the candidate window.
- If $||y1 - y0\ || < \varepsilon$ iteration stops, and going to step 2.

VIII. SIFT Algorithm

Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. This algorithm was published by David

Lowe. The SIFT algorithm can identify two objects as similar even the object is partly concealed in either one of the images has changed orientation, or the object is viewed at different angles.

The SIFT algorithm has split into four main phases such as,

A. Extrema Detection

The first phase inspects the image under various scales and octaves to separate points of the picture that are different from their backgrounds. These points are called extrema which is the potential candidates for image features.

B. Key point Localization

The Key Point Detection, starts with the extrema and selects some points to be key points, that are a whittled down a set of feature candidates. This refinement rejects extrema, which are caused by edges of the picture and by low contrast points.

C. Orientation Assignment

Each key point and its neighborhood are converted into a set of vectors by computing a magnitude and a direction for them. It also identifies other key points that may have been missed in the first two phases; this is done based on a point having a significant magnitude. The algorithm now has identified a final set of key points.

D. Key point Descriptor Generation

Key point Descriptor Generation, takes a collection of vectors in the neighborhood of each key point and consolidates this information into a set of eight vectors called the descriptor. Each descriptor is transformed into a feature by computing a normalized sum of these vectors.

CONCLUSION

In this paper, we researched various filters for image tracking and found that the use of Kalman filter shows better results. The effectiveness and robustness have been proved.

REFERENCES

[1] G. L. Foresti, Object Recognition And Tracking For Remote Video Surveillance, IEEE Trans. Circuits Syst. Video Technol., 9(7):1045-1062, October 1999.

[2] A. J. Lipton, H. Fujiyoshi, R. S. Patil, Moving Target Classification And Tracking From Real-time Video, Applications of Computer Vision,1998. WACV '98. Proceedings., Fourth IEEE Workshop on,pp. 8-14, 1998.

[3] Y. Li, A. Goshtasby, and O. Garcia, Detecting and tracking human faces in videos, Proc. ICPR'00 vol. 1, pg. 807-810 (2000).

[4] Gabriel, P.; Hayet, J.-B.; Piater, J.; Verly, J., "Object Tracking using Color Interest Points," IEEE Conference on Advanced Video and Signal Based Surveillance, 2005. International Journal of Information Technology, Modeling and Computing (IJITMC) Vol.1, No.2, May 2013

[5] C.Harris; M.Stephens., "A combined corner and edge detector," 4th Alvey Conference, pages 147– 151, 1988.

[6] J.J. Koenderink and A.J. Van Doorn,"Representation of local geometry in the visual system," Biological Cybernectics,55(6), 1987.

[7] Haritaoglu, I., Harwood, D., and Davis, L., "real-time surveillance of people and their activities," IEEE Trans. Patt. Analy. Mach. Intell. 22, 8, 2000.

[8] Sato, K. and Aggarwal, J., "Temporal spatio-velocity transform and its application to tracking and interaction," Comput. Vision Image Understand. 96, 2, 100–128,2004.

[9] Chun-Te Chu, Jenq-Neng Hwang, Shen-Zheng Wang, Yi-Yuan Chen ,"Human tracking by adaptive Kalman filtering and multiple kernels tracking with projected gradients," IEEE International Conference on Distributed smart Cameras, 2011.

[10] Zhi Liu; Liquan Shen; Zhongmin Han; Zhaoyang Zhang, "A Novel Video Object Tracking Approach Based on Kernel Density Estimation and Markov Random Field ," IEEE International Conference on image processing, 2007.

[11] Khatoonabadi, S.H.; Bajic, I.V., "Video Object Tracking in the Compressed Domain Using Spatio- Temporal Markov Random Fields," IEEE Transaction on Image Processing, 2013.