Object tracking and event detection

mohamed.outghratine And Soufiane.Ouali

February 2021

Contents

1 Introduction

The fields of image and video analysis have gained a lot of attention over the last years, in part due to the success of deep learning models. Most state-of-the-art methods use convolutional neural networks (CNNs) – deep networks which have shown to produce good results without the need for manual feature extraction. The shift of attention towards CNNs for image analysis happened in 2012, when Krizhevsky et al. convincingly won the ILSVRC (Image Net Large-Scale Visual Recognition Challenge) , a competition used for benchmarking image analysis models. The challenge was in image classification, where the goal is to label each image with one of multiple classes – e.g. to tell if a picture is of a cat or a dog. More accurate models have since been created, and in 2015 computers could outperform humans in this task. The use of CNNs has not been limited to image recognition, but has been successfully adopted in tasks such as object detection (localising objects of given classes), semantic segmentation (labelling each pixel in the image), and image captioning (producing a descriptive text of the content). Video analysis is closely related to the field of image analysis, a video being multiple images stacked in time. Similar challenges as in image analysis have been tackled, including video classification and object detection in video, but also tasks exclusive to video, such as object tracking (identifying objects across multiple frames), trajectory prediction (estimating paths of objects) and action recognition (classifying actions in a video sequence). As in image analysis, CNNs have produced great results for these tasks . Having accurate image and video analysis models is fundamental to autonomous vehicles and robots. Object detection models can enable self-driving cars to avoid collisions, and some level of scene understanding is required for robots to react to their surroundings. With the increased use of unmanned aerial vehicles (UAVs) – commonly known as drones – for tasks such as surveillance, delivery, and search and rescue , it is important to develop tools suitable for these applications and adapted to UAVs. Often, it is required to identify and track humans and other moving objects. For example, in surveillance tasks, it can be essential to track humans in order to detect unusual behaviour. Unfortunately, not all video analysis models are usable on UAVs; the most powerful models require high-end computational hardware, but for UAVs that should be light and inexpensive, the hardware is necessarily restricted.

2 Object detection

Object Detection is a process to identify objects of interest in the video sequence and to cluster pixels of these objects. Object detection can be done by various techniques such as temporal differencing , frame differencing , Optical flow and Background subtraction.
Every tracking method requires an object detection mechanism either in every-frame or when the object first appears in the video.This 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.

2.1 Object detection methods

2.1.1 Temporal Differencing

Temporal differencing method uses the pixel-wise difference between two or three consecutive frames in a video imagery to extract moving regions from the background [16]. It has high adaptability with dynamic scene changes although it cannot always extract all relevant pixels of a foreground object mostly when the object moves slowly or has uniform texture [18, 19]. When a foreground object stops moving, temporal differencing method cannot detect a change between consecutive frames and results in loss of the object.

2.1.2 Frame Differencing

Some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections. This temporal information usually in the frame differencing, highlights changing regions in consecutive frames. Given the object regions in the image, it is then the tracker's task to perform object correspondence from one frame to the next to generate the tracks. In this method, presence of moving objects is 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, as a result the detection of moving object is not accurate [5].

2.1.3 Optical Flow

Optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and the scene. Optical flow method is to calculate the motion between two image frames which are taken at times t and t+ t at every position [4]. These methods are called differential since they are based on local Taylor Series approximation of the image signal; that is, they use partial derivatives with respect to the spatial and temporal coordinates. 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 antinoise performance, make it not suitable for real-time demanding occasions [4].

2.1.4 Background Subtraction

Background subtraction is a technique for segmenting a foreground object from its background. The main step in background subtraction is background modelling. It is the core of background subtraction algorithm. Background Modelling must be sensitive enough to recognize moving objects [1]. Background Modelling is 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. Currently, mean filter and median filter are widely used to realize background modelling [6]. 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 has mainly two approaches [4]:

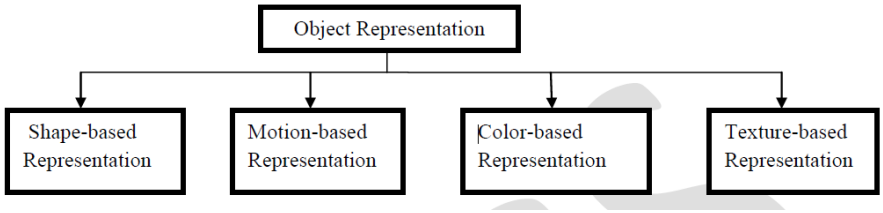
a-Recursive Techniques : Recursive techniques do not maintain a buffer for background estimation. Instead, they recursively update a single background model based on each input frame. As a result, input frames from distant past could have an effect on the current background model. Compared with non-recursive techniques, recursive techniques require less storage, but any error in the background model can linger for a much longer period of time. This technique includes various methods such as approximate median, adaptive background, Gaussian mixture [6, 11].

b-Non-Recursive Techniques : A non-recursive technique uses a sliding-window approach for background estimation. It stores a buffer of the previous L video frames, and estimates the background image based on the temporal variation of each pixel within the buffer. Non-recursive techniques are highly adaptive as they do not depend on the history beyond those frames stored in the buffer. On the other hand, the storage requirement can be significant if a large buffer is needed to cope with slow-moving traffic [11, 6].

The problem with background subtraction [20, 21] is to automatically update the background from the incoming video frame and it should be able to overcome the following problems: Motion in the background, Illumination changes, Memory, Shadows,Camouflage and Bootstrapping.

3 Object representations methods

In a tracking scenario, an object can be defined as anything that is of interest for further analysis. Objects can be represented by their shapes and appearances [17]. The extracted moving object may be different objects such as humans, vehicles, birds, floating clouds, swaying tree and other moving objects [5]. Hence shape features are usually used to represent motion regions. As per literature survey, approaches to represent the objects are as follows :



3.1 Shape-based Representation :

Different shape information of motion regions such as representations of points, box and blob are available for representing moving objects. Input features to the network is a combination of image-based and scene-based object parameters such as image blob area, apparent aspect ratio of blob bounding box and camera zoom [9]. Representation is performed on each image blob at every frame and results are stored in histogram.

3.2 Motion-based Representation :

Non-rigid articulated object motion shows a periodic property. This method has been used as a reliable approach for moving object representation. Some optical flow methods such as residual flow can be used to analyze rigidity and periodicity of moving entities. Rigid objects typically present little residual flow where as a non rigid moving object has higher average residual flow and displays a periodic component [9].

3.3 Color-based Representation :

Unlike many other image features color is relatively constant under viewpoint changes and easy to be acquired. Although color is not always appropriate as the only means of detecting and tracking objects, but the algorithms that have low computational cost makes color as an important feature to use when appropriate. To detect and track vehicles or pedestrians in real-time, among other techniques, color histogram based technique [23] is used. A Gaussian Mixture Model is used to describe the color distribution within the sequence of images. Object occlusion is handled using an occlusion buffer [6].

3.4 Texture-based Representation :

Texture based technique [14] counts the occurrences of gradient orientation in localized portions of an image, then computes the data on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for better accuracy.

4 Object tracking

Object tracking is a discipline within computer vision, which aims to track objects as they move across a series of video frames. Objects are often people, but may also be animals, vehicles or other objects of interest, such as the ball in a game of soccer.

4.1 Multiple object tracking

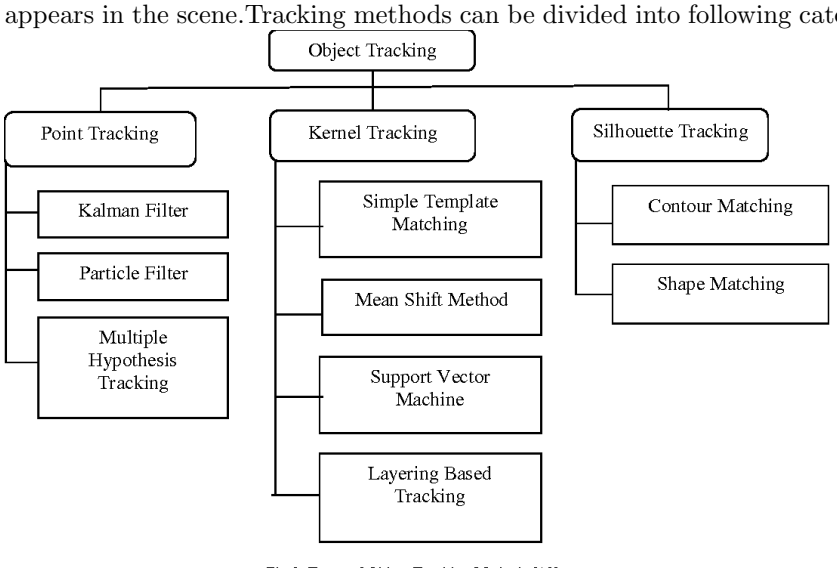
As the name suggests, multiple object tracking consists of keeping track of objects in a video as they move around. A formal description would be: for each frame in a video,localise and identify all objects of interest, so that the identities are consistent throughout the video. In other words, a good model has to accurately detect objects in each frame, and provide a consistent labelling of them. Challenges arise when objects are partially or completely occluded, or temporarily leave the field of view, since ideally the objects should keep their former IDs when reappearing.
Furthermore, objects whose paths intersect might confuse the model and cause it to erroneously switch their IDs. Different scenarios exist that allow for different types of models. An important distinction is that of online versus offline models . An online model receives video input on a frame-by-frame basis, and has to give an output for each frame. This means that, in addition to the current frame, only information from past frames can be used. Offline models, on the other hand, have access to the entire video, which means that information from both past and future frames can be used. The task can then be viewed as an optimisation problem, where the goal is to find a set of paths that minimise some global loss function. This has been solved using linear programming and k-shortest path optimisation methods. Since offline trackers have access to more information, one can expect better performance from these models. It should be noted, however, that real-time usage requires online models, as future frames are obviously unavailable. For a discussion on real-time models. Models can be further categorised into single class and multiple class models .

In the former, all tracked objects are of the same class, person, while in the latter, multiple classes are present, e.g. pedestrian, bicycle and car. On one hand, having multiple. classes introduces the problem of classifying each object. On the other hand, this could act as a distinguishing feature when the paths of objects of separate classes intersect. An extensive survey from 2006 of

object tracking systems showed that research until then had focused on finding good features, object representations and motion models to improve the tracking quality. It was noted that most methods assumed some restriction on the setting that greatly facilitated the tracking. For example, by assuming a static camera, it was possible to find all moving objects by applying background subtraction. Other common assumptions were that objects had a high contrast with respect to the background, moved smoothly and with little to no occlusion. Yilmaz et al. identified these assumptions to be greatly limiting when it comes to real-world application, stating that while said assumptions increased the performance, they made the models too specific.

4.2 Object Tracking Methods

Tracking can be defined as the problem of approximating the path of an object in the image plane as it moves in a scene. The purpose of an object tracking is to generate the route for an object by finding its position in every single frame of the video . Object is tracked for object extraction, object recognition and decisions about activities. Object tracking can be classified as point based 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.Tracking methods can be divided into following categories:



4.2.1 Point based Tracking

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 object detections. Recognition can be done relatively simple, by thresholding for the identification of these points. Point based tracking approaches are described below:

1-Kalman Filter :They are based on Optimal Recursive Data Processing Algorithm. The Kalman Filter performs the restrictive probability density propagation. 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 modelled 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. Kalman filters always give optimal solutions.

2-Particle Filter :The particle filter generates all the models for one variable before moving to the next variable. Algorithm has an advantage when variables are generated dynamically and there can be unboundedly numerous variables. It also allows for new operation of resampling. 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 are 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 is applied in developing area such as computer vision communal and applied to tracking problematic.

3-Multiple Hypothesis 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 tracking multiple object, handles occlusions and Calculating of Optimal solutions.

4.2.2 Kernel Based Tracking

Kernel tracking is usually performed by the moving object, which is represented by a 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 representation used, the number of objects tracked, and the method used for approximating 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 number of tracking techniques based on representation of object,object features ,appearance and shape of the object. Kernel based approaches are described below:

1-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 at that position. This method is capable of dealing with tracking single image and partially occluded object.

2-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. To characterize the target color histogram is chosen. Target model is generally represented by its probability density function. Target model is regularized by spatial masking with an asymmetric kernel.

3-Support Vector Machine (SVM) : SVM is a broad classification method which is termed by a set of positive and negative sample values. For SVM, the positive samples contain tracked image object, and the negative samples consist of all remaining things that are not tracked. It can handle single image, partial occlusion of object but necessity of physical initialization and training is must.

4-Layering based tracking : This is another method of kernel based tracking where multiple objects can be 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 b y means of 2D parametric motion. Every pixel's probability is calculated based on the object's foregoing motion and shape features. It can be capable of tracking multiple images and full occlusion of object.

4.2.3 Silhouette Based Tracking

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. This method is capable of dealing with variety of object shapes, Occlusion and object split and merges. Silhouette based tracking approaches are described below :

1-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.

2-Shape Matching : These approaches examine the object model in the existing frame. Shape matching performance is similar to th 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 in the form of density functions, silhouette boundary, object edges capable of dealing with single object and Occlusion handling will be performed in with Hough transform techniques.

5 Conclusion

Significant progress has been made in object tracking during the last few years. Several robust trackers have been developed which can track objects in real time in simple scenarios.We saw all the major aspects of object detection, object representation and object tracking have been addressed. Various methods in these aspects have been explained in brief and a number of merits and demerits were highlighted in each and every technique. Different object detection methods are temporal differencing, frame differencing, optical flow and background subtraction. It can be summarized as background subtraction is a simplest method providing complete information about object compared to other methods. Among the different methods of object representation, most of the researchers prefer texture based and color based object representation. Object tracking can be performed using various methods based on point, kernel, and silhouette. Advance study may be carried out to find efficient algorithm to reduce computational cost and to decrease the time required for tracking the object for variety of videos containing diversified characteristics.

References

[1] 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, 242-247.

[2] Saravanakumar, S.; Vadivel, A.; Saneem Ahmed, C.G., *"Multiple human object tracking using background subtraction and shadow removal techniques," Signal and Image Processing (ICSIP), 2010 International Conference on , vol., no., pp.79,84, 15-17 Dec. 2010*

[3] Abhishek Kumar Chauhan, Prashant Krishan, —*Moving Object Tracking Using Gaussian Mixture Model And Optical Flow‖, International Journal of Advanced Research in Computer Science and Software Engineering, April 2013*

[4] Sen-Ching S. Cheung and Chandrika Kamath, —*Robust techniques for background subtraction in urban traffic video.*

[5] Rupali S.Rakibe, Bharati D.Patil, —*Background Subtraction Algorithm Based Human Motion Detection,International Journal of Scientific and Research Publications, May 2013*

[6] M.Sankari, C. Meena, —Estimation of Dynamic Background and Object Detection in Noisy Visual Surveillance, International Journal of Advanced Computer Science and Applications, 2011, 77-83

[7] K.Srinivasan, K.Porkumaran, G.Sainarayanan,—Improved Background Subtraction Techniques For Security In Video Applications

[8] Rahul Mishra, Mahesh K. Chouhan, Dr. Dhiraj Nitnawwre,—Multiple Object Tracking by Kernel Based Centroid Method for Improve Localization‖, International Journal of Advanced Research in Computer Science and Software Engineering, July-2012, pp 137-140.

[9] Hitesh A Patel, Darshak G Thakore,—Moving Object Tracking Using Kalman Filter‖, International Journal of Computer Science and Mobile Computing, April 2013, pg.326 – 332.

[10] Greg Welch, Gary Bishop," An introduction to the Kalman Filter‖, In University of North Carolina at Chapel Hill,Department of Computer Science. Tech. Rep. 95-041, July-2006.

[11] Ruolin Zhang, Jian Ding, —Object Tracking and Detecting Based on Adaptive Background Subtraction‖, International Workshop on formation and Electronics Engineering, 2012, 1351-1355.

[12] Mr. Joshan Athanesious J; Mr. Suresh P, Implementation and Comparison of Kernel and Silhouette Based Object Tracking‖, International Journal of Advanced Research in Computer Engineering Technology, March 2013, pp 1298- 1303.

[13] J. Shotton, "Contour and texture for visual recognition of object categories," Doctoral of Philosophy, Queen's College, University of Cambridge, Cambridge, 2007.

[14] Jae-Yeong Lee; Wonpil Yu, "Visual tracking by partition-based histogram backprojection and maximum support criteria,"Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on, vol., no., pp.2860,2865, 7-11 Dec. 2011

[15] Rupesh Kumar Rout ,—A Survey on Object Detection and Tracking Algorithms‖ Department of Computer Science and Engineering National Institute of Technology Rourkela – 769 008, India.

[16] Kinjal A Joshi, Darshak G. Thakore —A Survey on Moving Object Detection and Tracking in Video Surveillance System —International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012

[17] Alper Yilmaz, Omar Javed, and Mubarak Shah. —*Object tracking: A survey. Acm Computing Surveys‖ (CSUR), 38(4):13, 2006.*

[18] N. Paragios, and R. Deriche.. —Geodesic active contours and level sets for the detection and tracking of moving objects.‖ IEEE Trans. Pattern Analysis Machine Intelligence. 22, 3, 266–280, 2000.

[19] S. Zhu and A. Yuille. —Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation.‖ IEEE Trans. Pattern Analysis Machine Intelligence 18, 9, 884–900, 1996.

[20] Changick Kim and Jenq-Neng Hwang. —Fast and automatic video object segmentation and tracking for content-based applications.‖ Circuits and Systems for Video Technology, IEEE Transactions on, 12(2):122–129, 2002.

[21] Zhan Chaohui, Duan Xiaohui, Xu Shuoyu, Song Zheng, and Luo Min. —An improved moving object detection algorithm based on frame difference and edge detection.‖ In Image and Graphics, 2007. ICIG 2007. Fourth International Conference on, pages 519–523. IEEE, 2007.

[22] Shai Avidan. —Support Vector Tracking.‖ IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 8,August 2004.

[23] Dorin Comaniciu, Visvanathan, Ramesh, Peter Meer.Kernel-Based Object Tracking.‖ IEEE Conference on Computer Vision and Pattern Recognition, 2000.