Volume: 1 Issue: 14 Pages: 38-42

International Journal of Research in Academic World

E-ISSN: 2583-1615
Impact Factor: 4.714

Received: 24/September/2022 Int. J Res. Acad. World. 2022; 1(14):38-42 Accepted: 01/October/2022

Moving Object Tracking Using Machine Learning

*[1]Seethal Prince E, [2]Asha S and [3]Greshma P Sebastian

*[1, 2, 3]Assistant Professor, SCMS College, KTU Karukutty, Thrissur, Kerala, India.

Abstract
Object tracking is the process of locating the object which is in motion. It is an important task in the field of computer vision. Due to the variations in pose, size, illumination, motion of object and the like, object tracking becomes a challenging mission. Widely established tracking method is tracking-by- detection method. Object may change its appearance throughout the image sequences or video. If the tracker is not learning this change in the appearance dynamically, the chance for drift is high and that may leads to reduce the tracker efficiency. To overcome this issue, this study proposing a method to adopt the change in appearance of a moving object with Active Appearance Model. To represent the object in more natural manner SVM based Multiple Instance Learning (MIL) representation is also used. The location of the moving object in the first frame is given as the input to the tracker. The proposed method achieves good result with real time performance. The error plots and the precision plots for different standard object tracking datasets are generated.

Keywords: Object tacking, detection method, multiple instance learning (MIL)

1. Introduction

Object Tracking is a critical task in the area of computer vision. Real-time object tracking is a difficult task. The main reasons for this difficulty is unconstrained environment due to the variations in factors such as pose, size, illumination, partial occlusion and motion blur. Widely established tracking method is tracking-by-detection method. Its application includes in video surveillance, navigation, traffic management, human-computer interaction etc. Real time object tracking is still a very challenging problem since the appearance of the target object can be drastically changed due to the factors such as illumination changes, pose variations, full or partial occlusions, abrupt motion, etc. Object Tracking is to find the object location in each frame of a video at every time instant. It is finding the trajectory of an object over time. Tracking is usually performed with the objects features and its location. There exist a number of approaches for object tracking. These approaches primarily differ from one another based on the way they approach the suitable object representation for tracking, image features, and motion, appearance, and shape model. These depend on the context or the environment in which the object tracking is accomplished as well as the end use for which the tracking information is being searched. Object tracking is one of the most important components of computer vision application. Its application includes in video surveillance, navigation, traffic management, human-computer interaction etc. Due to motion blur, noise, partial occlusion, complex motion object detection and tracking are very challenging tasks in its application. Real time object tracking is still a very challenging problem since the appearance of the target object can be drastically changed due to the factors such as illumination changes, pose variations, full or partial occlusions, abrupt motion, etc.

To address the problem of object tracking, many approaches use a representation of the image objects by a collection of local descriptors of the image content. Some of the commonly used features are color, edges, texture, optical flow, gradient etc. The main three steps in the object tracking are detection of interesting objects, tracking of such objects from frame to frame, and analysis of object tracks. First we need to select the object to track. Then with features of that selected object find the object in following frames. As the final step object tracks that found is analysed. Numerous approaches for object tracking have been proposed. These primarily differ from each other based on the way they approach the suitable object representation for tracking, image features, and motion, appearance, and shape model. These depend on the context/environment in which the tracking is performed and the end use for which the tracking information is being sought. The tracking system mainly consists of three components: an appearance model, which can evaluate the likelihood that the object of interest is at some particular location, a motion model, which relates the locations of the object over time, and a search strategy for finding the most likely location in the current frame. Here an active appearance model evaluates the objects presence in each frame, and then the motion model includes the object location in each frame and the size of that object in that frame. The search strategy for finding the object is the histogram matching approach. It matches the object of interest and the patches from the next frame. An active appearance model, which evolves during the tracking process as the appearance of the object changes, is the key to good performance. Tracking an arbitrary object with no prior knowledge other than its location in the first video frame need a robust way of updating an adaptive appearance model. For this a discriminative learning paradigm called Multiple Instance Learning (MIL) is using.

***Corresponding Author:** Seethal Prince E

2. Motivation

In an input video for object tracking application, the object is in motion and its appearance changes due to its complex motion, illumination change, occlusion etc. An important slice of a tracker to achieve good performance is its ability to update the representation of the object of interest as the conditions changes. Most of the trackers greatly depend on the internal representation of the target object [16]. Appearance model that used in early works kept it fixed throughout the tracking process. It is modelled at the beginning and there will be no change in its representation throughout the process [6], [15], [5]. The motion of the object of interest affects the tracker output. If static appearance model is used for the tracker, it is very difficult to obtain a good result. The inaccurate location of the object in the current frame causes a drift in the coming video frames and degrades the tracker performance. So we use adaptive appearance model. The key gain of MIL lies in the point that by adopting the multiple instance representation of many real object leads to more natural representation of that object. It also captures more information than using single instance representation [5], [23]. Multiple Instance Learning is a special learning framework which deals with uncertainty of instance labels. In this setting training data is available only as pairs of bags of instances with labels for the bags. Instance labels remain unknown and might be inferred during learning. Supervised learning model has one drawback that it is not always possible to provide labelled examples for training. Multiple Instance learning provides a new way of modelling this weakness. Instead of receiving a set of instances which are labelled positive or negative, the learner receives a set of bags that are labelled positive or negative. Each bag contains many instances. A positive bag label indicates that at least one instance of that bag can be assigned a positive label. This instance can therefore be thought of as witness for the label. Instance in negative labelled bags are altogether of the negative class, so there is no uncertainty about their label. By combining these two concepts, Appearance model and Multiple Instance Learning, developed a robust Object Tracking system.

3. Problem Statement

The objective is to develop a robust object tracking system which updates the appearance model with the discriminative learning paradigm called Online Multiple Instance Boosting.

A. Area of Research

Object Tracking Object tracking is estimating the trajectory of an object as it moves around the sight. It is a challenging problem. The challenges in tracking can ascend due to the object motion, change in appearance of object and also the background, occlusion and camera motion. With these challenges, it is one of the most important components of computer vision application. Its application includes in video surveillance, navigation, traffic management, human-computer interaction etc. due to motion blur, noise, partial occlusion, complex motion object detection and tracking are very challenging tasks in its application. To address the problem of object recognition, many approaches use a representation of the image objects by a collection of local descriptors of the image content. Some of the commonly used features are color, edges, texture, optical flow, gradient etc.

- Object representation
- Feature selection
- Object detection
- Object tracking

Anything that is of interest for further analysis is termed as object. Vehicles on a road, people walking on a road are examples for object. Its shape and appearance are the two ways used to represent. Points, primitive geometric shapes, contour and skeletal models are some shape representation for an object. Common appearance representations in object tracking are probability densities of object appearance, templates, active appearance models and multiview appearance models. Features play a critical role in object tracking. If the feature is unique then the object can be easily distinguished from the image. Feature selection is based on the object representation. For example, contour based representation use object edges as the feature. Commonly used features are color, edges, optical flow and texture. Object detection can be performing in every frame or in the first frame only. Object tracking can be done in two ways. One, possible object regions in every frame are obtained by means of an object detection algorithm, and then the tracker corresponds objects across frames. Other, the object region and correspondence is jointly estimated by iteratively updating object location and region information obtained from previous frames.

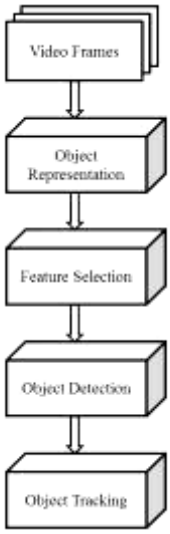

Fig 1: Different steps in object tracking.

B. Appearance Model

Object tracking significantly depend on the representation of the object. To achieve better result the tracker should update the appearance or the representation of the target as it changes to time or frame [16]. The appearance of the object may get change due to its complex motion, illumination change, occlusion etc. Appearance model that used in early works kept it fixed throughout the tracking process. It is modelled at the beginning and there will be no change in its representation throughout the process [6, 15, 5].

C. Multiple Instance Learning

Multiple Instance Learning (MIL) is a variation on supervised learning. In Supervised Learning learner receives a set of instances which are labelled positive or negative but in Multiple Instance Learning learner receives a set of bags that are labelled positive or negative. Each bag contains many instances. If all the instances in it are negative then the assumption is that bag is labelled negative and on the other hand, a bag is labelled positive if there is at least one instance in it which is positive. Multiple Instance Learning is a special learning framework which deals with uncertainty of instance

labels. In this setting training data is available only as pairs of bags of instances with labels for the bags. Instance labels remain unknown and might be inferred during learning. Supervised learning model has one drawback that it is not always possible to provide labelled examples for training. Multiple Instance learning provides a new way of modelling this weakness. Instead of receiving a set of instances which are labelled positive or negative, the learner receives a set of bags that are labelled positive or negative. Each bag contains many instances. A positive bag label indicates that at least one instance of that bag can be assigned a positive label. This instance can therefore be thought of as witness for the label. Instance in negative labelled bags are altogether of the negative class, so there is no uncertainty about their label. Consider the following learning problem. Suppose there is a keyed lock on the door to the supply room in an office. Each staff member has a key chain containing several keys. One key on each key chain can open the supply room door. For some staff members, their supply room key opens only the supply room door; while for other staff members, their supply room key may open one or more other doors (e.g., their office door, the mail room door, the conference room door). Suppose you are a lock smith and you are attempting to infer the most general required shape that a key must have in order to open the supply room door. If you knew this required shape, you could predict, by examining any key, whether that key could unlock the door. What makes your lock smith job difficult is that the staff members are uncooperative. Instead of showing you which key on their key chains opens the supply room door, they just hand you their entire key chain and ask you to figure it out for yourself! Furthermore, you are not given access to the supply room door, so you can't try out the individual keys. Instead, you must examine the shapes of all of the keys on the key rings and infer the answer. This kind of learning problem is Multiple Instance Learning.

4. Literature Review

In [11], A. Jepson *et al* proposed a context for appearance model for motion based tracking of objects. This method involves a combination of stable image assembly along with 2 frame motion and an outlier process. To adapt the appearance model EM-algorithm is adapted. This implementation is based on the filter comebacks from a steerable pyramid. This method provides robustness even if the object faces occlusions in its motion and has the capability to adapt the changes in appearance. It keeps a measure of stability of the object structure throughout the tracking process. The stable model determines the best consistent framework for tracking. The authors proposed an adaptive appearance model that finds the stable feature. The EM-algorithm that used here is an online version that adapts the parameters. The proposed tracking algorithm is incrementally estimates the motion and appearance. Do not lump all figures at the end of the paper! If you have difficulties with the titles on your figures, you can always elect to add in the titles as separate text boxes, rather than importing the titles with the graph. This is sometimes helpful in getting a lengthy vertically-oriented title to display correctly. The authors of [8] proposed a fast, robust method for interpreting face images. They used active appearance model for it. The method finds the matching image by determining the model factors that reduce the difference between the image and blended face. In the training stage, the model learns a linear correlation between parameter shifts and the induced residuals. During searching it finds the portion the residuals and uses this model to precise the existing parameters, leading to a better fit. They adopted a direct optimization approach and it leads to robust and accurate method. The authors deed the fact that the optimization problem can be learn offline. They also discussed the idea of image interpretation by synthesis. In [6] authors proposed a new approach for target representation and localization. It is centred on feature histogram based representation. It is regularized by spatial covering with an isotropic kernel. The masking makes spatially smooth similarity functions and which is suitable for gradient based optimization. The target localization problem is formulated using the local maxima. The Bhattacharya coefficient is used as the similarity measure and to perform optimization tracker used mean shift procedure. The proposed method is positively managed with camera motion, partial occlusions, clutter and target scale variations. The key input of their method is to introduce a new framework for tracking non rigid objects. The new target object representation and localization method can be combined with various motion filters and data association procedures.

K. Ali and K. Saenko proposed a new MIL method for object detection in [1] that is capable of handling the noisier automatically obtained annotations. This method for MIL that is capable of handling strong labeling ambiguities. It is a twostep procedure. First, a committee of randomized MIL Boost learners is used to obtain confidence estimates on the labels of instances belonging to positive bags. Next, we incorporate the estimates within a new Boosting procedure, built by general sizing the MIL Boost loss to incorporate a prior over the latent space and applying Friedmans gradient Boosting. The resulting method is shown to be particularly effective in the case where most positive bags contain little to no positive instances. For any given image or bag, the returned set of bounding boxes consists either exclusively of background patches, or a small number of positives. A batch mode framework, namely Batch mode Adaptive Multiple Instance Learning (BAMIL) is proposed by Wen Li *et al*. in [14] to accelerate the instance level MIL methods. Instead of using all training bags at once, they divided the training bags into several sets of bags called batches. Each time it train the instance level classifier using the instance level MIL methods in a batch-by-batch fashion which is simultaneously adapted from the latest prelearned classifier. Such batch mode framework significantly accelerates the traditional MIL methods for large scale applications and can be also used in dynamic environments such as object tracking. Since training the classifier using one batch of training bags does not take too much time, the training process with our framework is much faster. Moreover, such framework can be applied for dynamic settings such as object tracking.

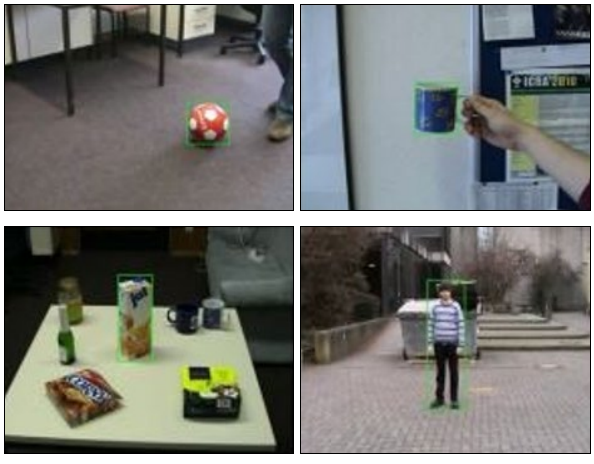

Fig 2: Object tracking

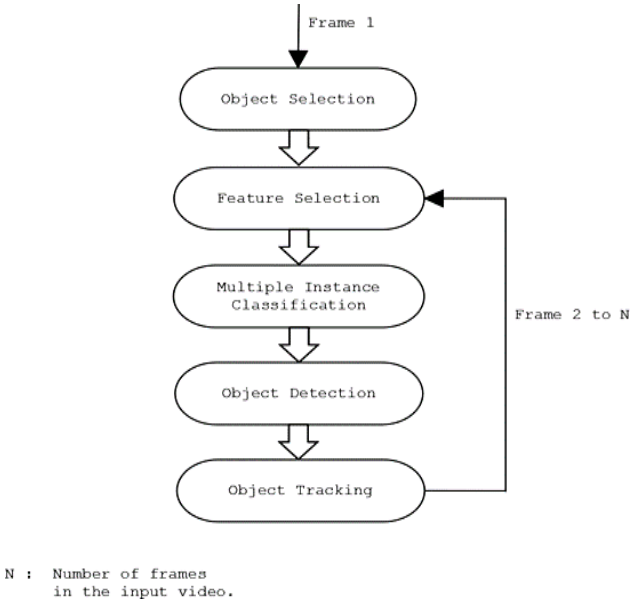

Fig 3: Proposed system working

5. Proposed Object Tracking

1 .Object Selection

Object is selected by drawing a bounding box around the target object. This process is known as land marking. The coordinate values of the bounding box that drawn by the user are saved for the further processing. Figure A shows sample with some datasets.

2. Feature Selection

The Edge and the texture flow are the two features that used here for creating the appearance model. The edge feature is obtained by Canny Edge Detection method. 3. Multiple In stance Classification Multiple Instance Classification classifies the image patches into positive and negative patches. For the learning phase, the classifier uses the image patches around the target object. These patches will be in same size. There will be 72 positive patches and negative patches each for a target. The patches for this are obtained from the neighboring region of the object of interest.

4. Object Detection

Next step is finding the object of interest from the next frame. As mentioned in the methodology, we need to define a search area. Tracker searches the target in this area. Search area for the proposed tracker is 10 pixel radiuses. It is taken though experiments. The number of patches from this search area is 305 (average calculated from different datasets). As the search area increases, number of patches increases and this consume more time. Another reason for taking this as the search radius is that the object will not move a large distance when compared to the previous frame.

Conclusion

A novel method for Object Tracking has been introduced in this study. This method includes appearance model and multiple instance learning concepts. Normally the object tracking is done with static appearance model which learn the appearance of the object only at the early stage of tracking. This may lead to drift in the object path. To overcome this challenge the appearance model based object tracking is proposed. This appearance model is obtained with the combination of shape model and the texture. The shape model is obtained from edge detection method. Canny edge detection method is applied to obtain this. The texture of the object is obtained from its grayscale level. SVM based multiple instance learning is used for the classification purpose. This helps to reduce the time consumed for tracking. The proposed method is evaluated with the ground truth value. The results shows an average 90 per accuracy. When referencing a journal article

.

References

1. Jeena Rita KS. Bini Omman, ”Edge based Object Tracking”, Paper presented in Students Research Symposium (SRS) in Fourth IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), August, 2015.
2. Jeena Rita KS. Bini Omman, A Technical Assessment on License Plate Detection System”, Chapter Proposal accepted in IGI Global for the book, ’Multi Core computer Vision and Image Processing for Intelligent Applications’.
3. Ali K, Saenko K. Confidence rated multiple instance boosting for object detection. In CVPR, pages 2433–2440. IEEE, 2014.
4. Alper Yilmaz OJ, Shah M. Object tracking: A survey. In ACM Computing Surveys, volume 38, 2006.
5. Babenko B, Yang MH, Belongie S. Robust object tracking with online multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell. 2011; 33(8):1619-1632.
6. Balan AO, Black MJ. An adaptive appearance model approach for model based articulated object tracking. In CVPR IEEE Computer Society. 2006; (1):758-765.
7. Collins RT, Lipton A, Kanade T, Fujiyoshi H, Duggins D, Tsin Y, Tolliver, Enomoto N, Hasegawa O, Burt P, *et al*. A system for video surveillance and monitoring, volume 2. Carnegie Mellon University, the Robotics Institute Pittsburg, 2000.
8. Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. 2003; 25(5):564-575,
9. Dietterich T, Lathrop R, Lozano Prez T. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence. 1997; 89(1-2):31-71,
10. Edwards GJ, Taylor CJ, Cootes TF. Interpreting face images using active appearance models. In FG, pages 300–305. IEEE Computer Society, 1998.
11. Freund Y., Schapire RE. A decision-theoretic generalization of on line learning and an application to boosting. In European Conference on Computational Learning Theory, 1995, 23-37.
12. Howse J. Open CV Computer Vision with Python. Packt Publishing Ltd, 2013.
13. Jepson AD, Fleet DJ, El-Maraghi TF. Robust online appearance models for visual tracking. In CVPR Computer Society. 2001; (1):415-422.
14. Klein D, Schulz D, Frintrop S, Cremers AB. *et al*. Adaptive real time video tracking for arbitrary objects. In Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference, 2010, 772–777.
15. Klein DA, Cremers AB. Boosting scalable gradient features for adaptive real time tracking. In ICRA, 2011, 4411–4416.
16. Li W, Duan L, Tsang IWH, Xu D. Batch mode adaptive multiple instance learning for computer vision tasks. In CVPR, Computer Society, 2012, 2368–2375.
17. Prez P, Hue C, Vermaak J, Gangnet M, Color based probabilistic tracking. In Heyden A, Sparr G, Nielsen M,

Johansen P. editors, ECCV (1), volume 2350 of Lecture Notes in Computer Science, pages Springer, 2002, 661-675.
18. Salti S, Cavallaro A, di L. Stefano. Adaptive appearance modeling for video tracking: Survey and evaluation. IEEE Transactions on Image Processing. 2012; 21(10):4334-4348.
19. Viola P, Jones M, Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001; (1):I–511–I–518.
20. Viola PA, Platt JC, Zhang C. Multiple instance boosting for object detection. In NIPS, 2005, 1417-424.