International Journal of Infinite Innovations in Technology|ISSN:2278-9057
IJIIT|Volume-II|Issue-IV|2013-2014 April|Paper-17
Reg. No.:20140317|DOI:V2I4P17

Tracking System with effective Learning

Prachi Arora
D. J. Sanghvi College of Engineering
Vileparle(w),Mumbai

Prof. Vivek Deodeshmukh
D. J Sanghvi College of Engineering
Vileparle(w),Mumbai

Abstract—There are different methods that are used for object tracking perform adaptive tracking-by-detection, meaning that a detector predicts the position of an object and adapts its parameters to the object's appearance at the same time. While suitable for cases when the object does not disappear from the scene, but these methods fails on occlusions, illumination changes etc. In this work, we build on a novel approach called Tracking-Learning-Detection (TLD) that overcomes this problem. In methods based on TLD, a detector is trained with examples found on the trajectory of a tracker that itself does not depend on the object detector. We achieve high robustness and outperform existing adaptive tracking-by-detection methods. We show that by using simple features for object detection and by employing a feedback approach a considerable reduction of computing time is achieved.

***Index Terms*—Long Term Tracking, Principal component Analysis, TLD Framework, Template matching, PN learning.**

INTRODUCTION

In a video stream a bounding box is defined which is defining the object of interest in a single frame, our goal is to automatically determine the object's bounding box or indicate that the object is not visible in every frame that follows.[1] The video stream is to be processed at frame rate and the process should run indefinitely long .We refer to this task as long-term tracking. To enable long-term tracking, there are a number of problems which need to be addressed. The main problem is the detection of the object when it reappears in the camera's field of view. This problem is aggravated by the fact that the object may change its appearance, thus making the appearance from the initial frame irrelevant. Next, a successful long-term tracker should handle scale and illumination changes, background clutter, partial occlusions, and operate in real time.[2]

The long-term tracking can be approached either from tracking or from detection perspectives. Tracking algorithms estimate the object motion. Trackers only require initialization, are fast, and produce smooth trajectories. On the other hand, they accumulate error during runtime (drift) and typically fail if the object disappears from the camera view. Research in tracking aims at developing robust trackers that track "longer." The long-term tracking can be implemented either from tracking or from detection perspective. Tracking algorithms tracks object motion. Trackers only require initialization, are fast, and produce smooth trajectories. On the other hand, they accumulate error during runtime (drift) and typically fail if the object disappears from the camera view [3]. The post failure behaviour is not directly addressed.

Detection based algorithms estimate the object location in every frame independently. Detectors do not drift and do not fail if the object disappears from the camera view. However they required an offline training stage and therefore cannot be applied to unknown objects [4].

It is found in research that neither tracking nor detection can solve the long term tracking problem. If both opearate together there is potential to benefit from one another. A tracker can provide weakly labeled training data for a detector and thus improve it during runtime. A detector can reinitialize a tracker and thus minimize the tracking failures [6] [5].

RELATED WORK

To perform video tracking an algorithm analyzes sequential video frames and outputs the movement of targets between the frames. There are a variety of algorithms, each having strengths and weaknesses. Considering the intended use is important when choosing which algorithm to use. There are two major components of a visual tracking system: target representation and localization and filtering and data association [5]. Target representation and localization is mostly a bottom-up process. These methods give a variety of tools for identifying the moving object. Locating and tracking the target object successfully is dependent on the algorithm. For example, using blob tracking is useful for identifying human movement because a person's profile changes dynamically [4]. Typically the computational complexity for these algorithms is low. The following are some common target representation and localization algorithms [6].

- **Blob tracking**: segmentation of object interior (for example blob detection, block-based correlation or optical flow)

- **Kernel-based tracking** (mean-shift tracking): an iterative localization procedure based on the maximization of a similarity measure (Bhattacharyya coefficient).
- **Contour tracking**: detection of object boundary (e.g. active contours or Condensation algorithm)
- Camshift algorithm uses color feature for real time object tracking. Camshift fails when video is under rapid motion illumination changes and background distraction.[1].
- Adaptive Local Search and Kalman Filter are proposed to predict position of Moving object.

Kalal has given work TLD framework for tracking and detection and PN learning algorithm to learn about the characteristics of the moving object in video stream [2]. Improved Camshift reduces the effect of illumination interference and judges whether the target is lost. Tracking objects can be complex due to

- Noise in images
- Complex object motion
- Non rigid or articulated nature of objects Partial and full object occlusions, complex object shapes
- Scene illumination changes
- Real-time processing requirements
- Change in illumination
- Shadows of moving object

The implementation of object tracking system is based on following key concepts:
Detection of interested moving objects in a frame, Such objects are tracked from frame to frame.
Analysis of object tracks to recognize their behavior and trajectory of object can be estimated

CURRENT WORK

TLD FRAMEWORK:

Tracker estimates the object motion under the assumption that the object is visible and its motion is limited. A tracker can provide weakly labeled training data for a detector and thus improve it during runtime.[1] Detector performs full scanning of the image to localize all the appearances that have been observer in the past. A detector can reinitialize a tracker and thus minimize the tracking failures. Detection based algorithms estimates the object location in every frame independently. [1][3]

Detectors do not drift and do not fail if the object disappears from the camera. However, they require an offline training stage. The starting point of my work says that neither tracking nor detection can solve long term tracking task independently. But if they operate simultaneously, there is potential to benefit

Wide range of trackers and detectors exist, we are not aware of any learning method that would be suitable to the TLD framework [5]. Such a learning method should:

- Deal with arbitrarily complex video streams where the tracking failures are frequent.
- Never degrade the detector if the video does not contain relevant information, and operate in real time.

To tackle all these challenges, we rely on the various information sources contained in the video. Consider for instance, a single-path denoting the object location in a single frame. This path defines not only the appearances of the object, but also determines the surrounding patches, which define the appearances of the background [4]. When tracking the patch, one can discover different appearances of the same objects as well as more appearances of the background [7]. This is in constant to standard machine learning approaches, where a single example is considered independent from other examples. This opens interesting questions of how to effectively exploit the information in the video during learning [6].

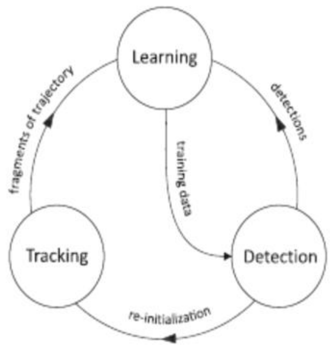

Fig. 1. Block diagram of TLD framework

In every frame, the P-N learning performs the following steps: 1) evaluation of the detector on the current frame. 2) Estimation of the detector errors using the P-N experts, and 3) update of the detector by labeled examples output by the experts.

- P-experts – recognizes missed detections and

- N-experts –recognizes false alarms

The estimated errors augment a training set of the detector, and the detector is retained to avoid these errors in the future. As with any other process, the P-N experts are also making errors themselves [8]. However, if the probability of an experts error is within certain limits (which will be analytically quantified), the errors are mutually compensated, which leads to stable learning [10].

A real-time long-term tracking system based on the TLD framework and the P-N learning is shown. The system tracks, learns, and detects an object in a video stream in real time [5].

TEMPLATE MATCHING:

Template matching is a technique for finding areas of an image that match (are similar) to a template image (patch).

Two primary components are required:

Source image (I): The image in which we expect to find a match to the template image

Template image (T): The patch image which will be compared to the template image. Goal is to detect the highest matching area:[16]

Fig. 2.

To identify the matching area, we have to c*ompare* the image against the source image by sliding it:[16] By sliding, we mean moving the patch one pixel at a time (left to right, up to down). At each location, a metric is calculated so it represents how "good" or "bad" the match at that location is (or how similar the patch is to that particular area of the source image).[16] For each location of **T** over **I**, you *store* the metric in the *result matrix* **(R)**. Each location (x, y)in **R** contains the match metric:[12]

Result R of sliding the patch with a metric TM_CCORR_NORMED. The brightest locations indicate the highest matches. As you can see, the location marked probably the one with the highest value, so that location (the rectangle formed by that point as a corner and width and height equal to the patch image) is considered the match. In practice, we use the function minMaxLoc to locate the highest value (or lower, depending of the type of matching method) in the R matrix. Different Template matching techniques are mentioned below.[16]

1. Method=CV_TM_SQDIFF

$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2$$

2. Method=CV_TM_SQDIFF_NORMED

$$R(x,y) = \frac{\sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$

3.Method=CV_TM_CCORR

$$R(x,y) = \sum_{x',y'} (T(x',y') \cdot I(x+x',y+y'))$$

4. Method=CV_TM_CCORR_NORMED

$$R(x,y) = \frac{\sum_{x',y'} (T(x',y') \cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$

5. Method=CV_TM_CCOEFF

$$R(x,y) = \sum_{x',y'} (T'(x',y') \cdot I(x+x',y+y'))$$

Where

$$T'(x',y') = T(x',y') - 1/(w \cdot h) \cdot \sum_{x'',y''} T(x'',y'')$$
$$I'(x+x',y+y') = I(x+x',y+y') - 1/(w \cdot h) \cdot \sum_{x'',y''} I(x+x'',y+y'')$$

6.Method=CV_TM_CCOEFF_NORMED

$$R(x,y) = \frac{\sum_{x',y'} (T'(x',y') \cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T'(x',y')^2 \cdot \sum_{x',y'} I'(x+x',y+y')^2}}$$

LIMITATION OF EXISTING SYSTEM

When object changes its appearance or object is moving out of camera frame and comes back, it does not recognize. Long term tracking fails due to rotation, illumination changes, background clutters and operate in real time. TLD does not perform well in case of full out-of-plane rotation.

PROJECT FRAMEWORK

A method is proposed for predicting the object motion and detecting the abnormal activities from surveillance videos, which is based on the learning of statistical motion patterns [12]. In these applications, the movements of objects are constrained by structured environments. Therefore, the relationship between objects and environments can be exploited as additional information for improving the performance of tracking [11].

The proposed system is implemented using Open CV tools. Proposed method is evaluated in a quantitative manner. Tracking-Learning-Detection, which is tightly coupled with an adaptive background generation to overcome the limit of block matching [5]. The proposed algorithm is robust to the object's sudden movement or the change of features [1]. It can be extended to complex and dynamic changing environment [2]. Trackers only require initialization, are fast, and produce smooth trajectories.[8]

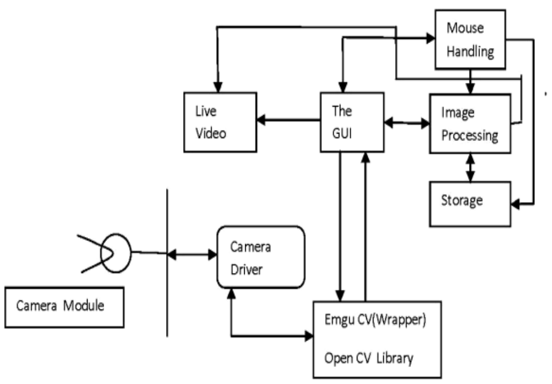

Fig. 4. System Architecture

ALGORITHM

1 Initialize all the system variables
2 Initialize Camera
3 Fetch the first frame from the camera
4 Select object to be tracked .i.e. 1st template
5 Store the template in Database
6 Create ROI at 20 pixel distance. ROI can be increased with the increment of 20 pixels if object is not found in the current ROI.
7 Fetch the next Frame from video stream.
8 Apply the template matching in imgROI for ith index of learned template to get the Object location. Template matching algorithms used to get highest intensity location and mark the object location.[13] If ROI fails then background subtraction technique can be used. [4][14]
9 Check for the % of matching of the ith template and store in the array
10 If matching is greater than matching limit update the tracker for the matched location and regenerate ROI
11 Repeat from step 7
12 If match not found at all find the best match in the array.
13 if match is greater than learning limit apply PCA algorithm [15][9][17] with the best matched template and the new template from ROI
14 If matched percentage is greater than learning limit apply PN learning algorithm for identification of detectors error and learning from it by pair of two experts P and N experts. P experts detect the true image while N expert detects background image.

PN learning Algorithm

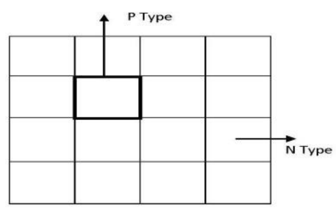

Fig. 5 P N Algorithm

Object/template from the image is selected by the user. Generate grid on the frame i.e divide the frame into no. of square or rectangular box, object to be tracked is called as P-type of object and remaining is background which is divided into the numbers of N-type objects as shown in figure. This P-type Object becomes reference template. Apply template matching on the ith negative image with the new template. Percentage of P-Type and N-Type matching is calculated.[17] If percentage matching is greater than limit it is a background and if it is less it is a positive response. If response is negative increase the region of Interest and go back to step 7 similarly response is positive store it in database and repeat from step7. As Object moves, position of P-Type and N-Type of Objects may change accordingly. Percentage of P-Type of Object is calculated and behavior of object is learned. Maximum percentage of P-type of object gives the object of interest

APPLICATIONS

1. TLD system is designed to track the objects continuously in Long-Term videos
2. In existing system to track any object we must watch and observe the each and every frame of the footage to track the object movements

3. It is very much time consuming process and it can be overcome by TLD We can track the object as well as learn the object
4. Video indexing: automatic annotation and retrieval of in multimedia databases
5. Automated surveillance or wildlife surveillance: monitoring a scene to detect activities or unlikely events.
6. Traffic monitoring: real-time gathering of traffic statistics to direct traffic flow 6. Vehicle navigation: video-based path planning and obstacle avoidance capabilities [1]

CONCLUSION

The project undertaken effectively detects the object of interest, discarding the smaller unwanted objects causing false alarms in the form of p and n experts. Implementation also presents a way to track an interested object, from a group of moving and detected object, in the subsequent frames of the continuous video stream. TLD framework is simple in implementation and needs very less computational requirements; this can be suitably and efficiently applicable to darker objects over light background. For the reverse condition, i.e., for lighter object over dark background, a slight shift in the position of the detected object is being observed..
Various algorithms and techniques are studied to enhance the performance of long term tracking by reducing the complexity caused by complex object shapes motion, illumination changes, scaling, rotation and partial and full object occlusions compared to existing systems.

REFERENCES

[1] Zdenek Kalal, jiri Matas, ―Tracking- learning –Detection‖ IEEE Transactions on pattern analysis and machine intelligence‖, vol. 34, no. 7, july 2012 1409, 0162-8828/12/$31.00 2012 IEEE. [2] Rangachar Kasturi, Fellow, IEEE, Dmitry Goldgof,

[2] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," Proc. Seventh Int'l Joint Conf. Artificial Intelligence, vol. 81, pp. 674-679, 1981.

[3] J. Shi and C. Tomasi, "Good Features to Track," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, 1994.

[4]Marwa abdel el Azeem Marzouk, ―"Modified background subtraction algorithm for motion detection in surveillance systems" vol 1,Number 2,(2010), pp -112-123.

[5] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," Pattern Recognition, vol. 36, no. 3, pp. 585–601, 2003.

[6] Prof. Y. Vijaya Lata1, Chandra Kiran Bharadwaj Tungathurthi, "Facial Recognition using Eigen faces by PCA", Transaction Paper International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009

[7] Prof. Y. Vijaya Lata1, Chandra Kiran Bharadwaj Tungathurthi2, H. Ram Mohan Rao3, Dr. A. Govardhan4, Dr. L. P. Reddy, ―Facial Recognition using Eigenfaces by PCA‖, International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009

[8] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," Conference on Computer Vision and Pattern Recognition,1998.

[9] Tim K. Lee and Mark S. Drew, ―"3D Object Recognition by Eigen-Scale-Space of Contours", Cancer Control Research Program, BC Cancer Reserach Centre, 675

[10] C. Bibby and I. Reid, "Robust real-time visual tracking using pixel-wiseposteriors," European Conference on Computer Vision, 2008.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints,"International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110,2004.

[12] B. Babenko, M.-H. Yang, and S. Belongie, "Visual Tracking with Online Multiple Instance Learning," Conference on Computer Visionand Pattern Recognition, 2009.

[13] The OpenCV Tutorials, Release 2.4.2 July, page 1-355

[14] Priti Kuralkar*, Prof. V.T.Gaikwad ―"Background Subtraction and Shadow Detection Techniques A Review Paper", International Journal of Computer, Electronics & Electrical Engineering (ISSN: 2249 –9997) Volume 2–Issue1

[15] Chi-Farn Chen and Yun-Te Su ―"The Use of PCA for Moving Objects Tracking on the, Center for Space and Remote Sensing" Research National Central University Jhongli, TAIWAN Image Sequence

[16] open CV documentation
http://docs.opencv.org/doc/tutorials/tutorials.html

[17] "Adaptive Automatic Tracking, Learning and Detection of Real-time Objects in the Video Stream*" International* Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA International Conference & workshop on Advanced Computing 2013 (ICWAC 2013) – www.ijais.org 3.