Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-11, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

Real Time Object Tracking Using Tracking Learning Detection

Mr. Atul D Mairale[1] & Prof. Ravindra Kumar Gupta[2]
[1] Dept. of Computer Engg., RKDFIST, Bhopal, India
[2] Dept. of Computer Engg., RKDFIST, Bhopal, India.

Abstract: In this paper we investigate long-term tracking of an objects in a video clips. The object is defined by its location from where it is captured and extent in a single frame. In each and every frame that follows, the task is to find the location of object and extent or show that the object is not present. We propose a tracking framework (TLD) that explicitly focuses on long-term tracking task into tracking, learning, and detection. The tracker follows the object frame by frame. The detector stores all appearances that have been observed so far and improves the correctness of tracker if necessary. Learning monitors the performance of both tracker and detector, calculates detector's errors, and generates training examples to avoid these errors in the future. A novel learning method (P-N learning) which estimates the errors by a pair of "experts": (i) P-expert estimates missed detections, and (ii) N-expert estimates false alarms. The learning process is modelled as a discrete dynamical system and the conditions under which the learning guarantees improvement are found.

1. Introduction

One of the great open challenges in computer vision is to build a system that can kinematically track people. A reliable solution opens up tremendous possibilities, from human computer interfaces to video data mining to automated surveillance. This task is difficult because people can move fast and unpredictably, can appear in a variety of poses and clothes, and are often surrounded by clutter. Because of the technical challenge and the attractive rewards, there exists a rich body of relevant literature. However, in practice, the resulting systems typically require some limiting assumption such as multiple cameras, manual initialization, or controlled/ simplistic backgrounds.

Object tracking is an important task within the field of computer vision. The proliferation of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. In the simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object. Tracking object can be complex due to loss of information to projection of 3D world on 2D image, noise, complex object shape, nature of object, partial and full object occlusions etc.

Consider a video stream taken by a hand-held camera depicting various objects moving in and out of the camera's field of view. Given a bounding box 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. The video stream is to be processed at frame rate and the process should run indefinitely long.

2. Background Overview

A. The Tracking Approaches

- Tracking People by Learning Their Appearance
- Object tracking using SIFT features & mean shift
- Fast Occluded Object Tracking by a Robust Appearance Filter
- Fast Multiple Object Tracking via a Hierarchical Particle Filter
- A Linear Programming Approach for Multiple Object Tracking
- Tracking the Invisible: Learning Where the Object Might be

B. Drawbacks of Existing System

Existing system has the problems when object changes its appearances or object is moving out of camera view and again coming in front of camera. Tracking fails due to scaling, rotation, illumination changes and does not perform in case of full out of plane rotation. The proposed system overcomes the above mentioned drawback and system can Track

and Learn the Real-time object automatically using Principle component analysis, P-N learning, template matching and Eigen object detection.

C. Proposed System

Proposed system presents an automatic long term tracking and learning and detection of real time objects in the live video stream. In this system, Object to be tracked also called as cropped image is defined by its location and the extent in the single frame by selecting the object of interest in the live video. Many existing systems for tracking objects fails due to loss of information caused by complex shapes, rapid motion, illumination changes, scaling and projection of 3D world on 2D image. Proposed modified PN learning algorithm which uses background subtraction technique to increase speed of the frame processing for object detection. Proposed Modified PN learning algorithm considers the object to be tracked as P-Type Object and background is divided into the numbers of N-Type objects.

Initially input image is matched with the N-Type of objects for rejection and then with P-type for acceptance. Proposed system uses the Template Matching algorithm to match cropped image with region of interest in the current frame to mark the Object Location. If match is found then Principle Component Analysis algorithm is used for detection of the fast moving object which is the advantage over the existing systems. If match does not found then Proposed Modified PN learning processing is applied to detect the image in rapid motion video. Proposed system uses background subtraction to increase the performance for detection of any moving object as the background remains still and we get approximate location of the moving object. Proposed System is expected to minimize delay for frame processing and reduce average localization errors to improve in matching percentage irrespective of scaling of the input image. Thus proposed system is expected to overcome the drawbacks of existing system for efficient tracking of any real time object.

3. The Proposed System

A. System Overview

Video surveillance plays a crucial role due to security issues involved in various areas like crowded public places, departmental stores, traffic monitoring, banks and boarders between two countries. The System is expected to track and learn the real time objects. Video stream will be processed at frame rate and process should run indefinitely long. The task is called as long term tracking. In the simplest form, Tracking is defined as is process of continuously finding an object of interest in the video. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object. Tracker estimates the object motion under the assumption that the object is visible and its motion is limited. A tracker can provide weakly labelled training data for a detector and thus improve it during runtime.

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.

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 one from another.

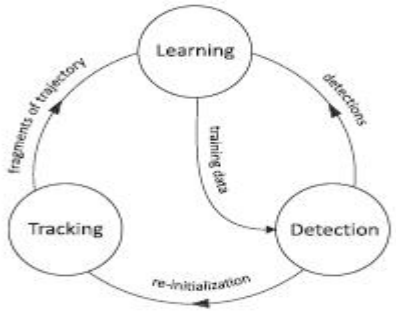

Figure 1 The block diagram of TLD framework

Learning observes the performance of both tracker and detector, estimates detector's errors, and generates training examples to avoid these errors in the future. The learning component assumes that both the tracker and the detector can fail. By virtue of the learning, the detector generalizes to more object appearances and discriminates against background.

B. Machine learning

Detectors are traditionally trained assuming that all training examples are labelled. Such an assumption is too strong in our case since we wish to train a detector from a single labelled example and a video stream. This problem can be formulated as a semi-supervised learning that exploits both labelled and unlabelled data. These methods typically assume independent and identically distributed data with

certain properties, such as that the unlabelled examples form "natural" clusters in the feature space. A number of algorithms relying on similar assumptions have been proposed in the past including EM, Self-learning and Co-training. Expectation-Maximization (EM) is a generic method for finding estimates of model parameters given unlabelled data. EM is an iterative process, which in case of binary classification alternates over estimation of soft-labels of unlabelled data and training a classifier. EM was successfully applied to document classification and learning of object categories .In the semi-supervised learning terminology, EM algorithm relies on the "low density separation" assumption , which means that the classes are well separated. EM is sometimes interpreted as a "soft" version of self-learning. Self-learning starts by training an initial classifier from a labelled training set, the classifier is then evaluated on the unlabelled data. The examples with the most confident classifier responses are added to the training set and the classifier is retrained. This is an iterative process. The self-learning has been applied to human eye detection in. However, it was observed that the detector improved more if the unlabelled data was selected by an independent measure rather than the classifier confidence. It was suggested that the low density separation assumption is not satisfied for object detection and other approaches may work better. Co-training is a learning method build on the idea that independent classifiers can mutually train one another. To create such independent classifiers, co-training assumes that two independent feature-spaces are available. The learning is initialized by training of two separate classifiers using the labelled examples. Both classifiers are then evaluated on unlabelled data. The confidently labelled samples from the first classifier are used to augment the training set of the second classifier and vice versa in an iterative process. Co-training works best for problems with independent modalities, e.g. text classification (text and hyper-links) or biometric recognition systems (appearance and voice). In visual object detection, co-training has been applied to car detection in surveillance and moving object recognition.

C. P-N Learning

This part investigates the learning component of the TLD framework. The goal of the component is to improve the performance of an object detector by online processing of a video stream. In every frame of the stream we wish to evaluate the current detector, identify its errors, and update it to avoid these errors in the future. The key idea of P-N learning is that the detector errors can be identified by two types of "experts." P-expert identifies only false negatives, N-expert identifies only false positives. Both of the experts make errors themselves; however, their independence enables mutual compensation of their errors.

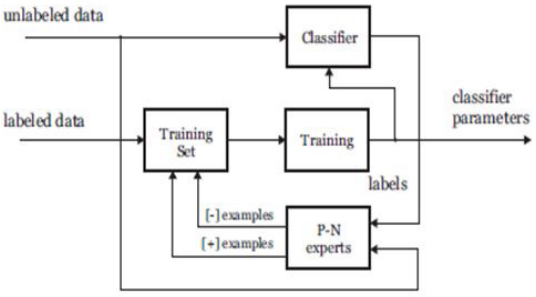

Figure 2. The block diagram of the P-N learning

The P-N learning consists of four blocks. A classifier to be learned.

1. Training set- a collection of labelled training examples.
2. Supervised training-a method that trains a classifier from training set.
3. P-N experts-functions that generate positive and negative training examples during learning.

D. PCA Algorithm

The purpose of PCA is to produce PC (Principal) images that contain both static and dynamic information on the interested clip. Firstly, since the video images normally include three bands (R, G, and B), every band in the interested clip will be assembled to form a single-band image set. Then, PCA is performed on every image set and produce a series of Principal Components for every individual band. Finally, all Principal Components are collected together to become a PC image sequence.

Three steps are used to reach the goal;

1. Interested clip extraction, the purpose to reduce the computational time of PCA.
2. PCA, the process to produce PC (Principal Component) images that contain both static and dynamic information on the image sequence.
3. Track extraction, the step to extract the tracks of moving objects from each PC.

E. Template Matching

Template matching is a technique for finding areas of an image that match (are similar) to a template image (patch). We need two primary components: Source image: The image in which we expect to find a match to the template image & Template image: The patch image which will be compared to the template image our goal is to detect the highest matching area. To identify the matching

area, we have to c*ompare* the image against the source image by sliding it.

E. Architecture of TLD

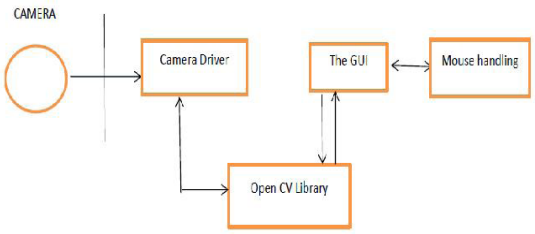

Figure 3. Architecture of TLD

This architecture consist of four main parts. The camera, OpenCV, GUI, Mouse handling.

1. Camera: It is used to capture the video.
2. OpenCV: It is a development tool. Camera interfacing can be done by adding OpenCV library function into visual studio.
3. GUI: It is the interfacing between the user and the software.
4. Mouse Handling: It is used to select the object that the user wants to track.

F. Implementation of TLD

Step 1: Camera interfacing
Step 2: Take the video streams into the system
Step 3: Object selection (Cropped image) bythe user.
Step 4: Creation of image array for learning.
Step 5: Store the Cropped object at array index 0.
Step 6: Create ROI. ROI can be increased with the increment of 20 pixels if object is not found in the current ROI.
Step 7: Fetch the next Frame from video stream.
Step 8: Apply the template matching for cropped image for cropped image at the index 0 in the ROI frame to get the Object location. Template matching algorithm is to get highest intensity location and mark the object Location. If ROI fails then background subtraction technique can be used.
Step 9: If match is found then apply the PCA for tracking the moving object on the video image then Eigen Object Detector to detect the object. Check for percentage matching. If matching percentage is less then store the new object in the learned array at the index 0 and shift other images in the array.
Step 10: If match is not found then Identification of detector errors and learning from it can be done from by P-N learning. P-N learning estimates the errors by pair of experts. P-experts detects missed detection whereas N-experts detects alarms.

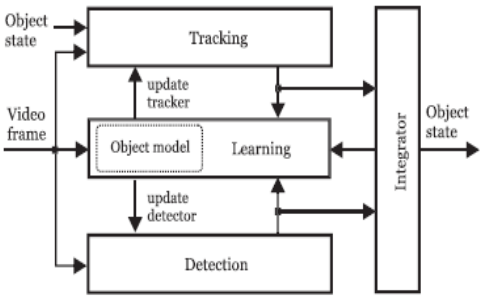

Figure 4. Detailed block diagram of the TLD framework.

4. Conclusion

In this paper, we studied the problem of tracking of an unknown object in a video stream, where the object changes appearance frequently moves in and out of the camera view. We designed a new framework that decomposes the tasks into three components: tracking, learning, and detection. The learning component was analysed in detail. The learning component analysis shows that an object detector can be trained from a single example and an unlabelled video stream using the following strategy:

1. Evaluate the detector,
2. Estimate its errors by a pair of experts, and
3. Update the classifier.

Each expert is focused on identification of particular type of the classifier error and is allowed to make errors itself. The stability of the learning is achieved by designing experts that mutually compensate their errors. The theoretical contribution is the formalization of this process as a discrete dynamical system, which allows specifying conditions, under which the learning process guarantees improvement of the classifier. The experts can exploit spatio-temporal relationships in the video. TLD framework is a real-time approach.

5. References

[1] Chavan, Rupali S., and Mr SM Patil. "Object Tracking Based On Tracking-Learning-Detection."

[2] Nemade, Bhushan, and R. R. Sedamkar. "Adaptive Automatic Tracking, Learning and Detection of Real-time Objects in the Video Stream."

[3] Kalal, Zdenek, Krystian Mikolajczyk, and Jiri Matas. "Tracking-learning-detection." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*34.7 (2012): 1409-1422.

[4] Ramanan, Deva, David A. Forsyth, and Andrew Zisserman. "Tracking people by learning their appearance." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.1 (2007): 65-81.