LONG TERM TRACKING-LEARNING-DETECTION OF MOVING OBJECT

[1]Prof. Pravin Adivarekar, [2]Rushikesh Jadhav, [3]Rajaram Rane, [4]Prathamesh Salvi

[1]Asst. Professor, [2,3,4]BE Student, [1,2,3,4]Comp. Engg. Dept, SSJCET, Asangaon, India.

[1]engineerpravin2008@gmail.com, [2]rushi1156@gmail.com, [3]rajcol11@yahoo.in, [4]prathamsalvi27@gmail.com

Abstract : This paper examines long-term tracking of unknown objects in a video stream. The object is defined by its location and extent in a single frame. In every frame that follows, the task is to determine the object's location and extent or indicate that the object is not present. We propose a framework (TLD) that explicitly decomposes the long-term tracking task into tracking, learning and detection. The tracker follows the object from frame to frame. Proposed framework selects the best result from several independent components and estimates the error at the same time. We develop a 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 modeled as a discrete dynamical system and the conditions under which the learning guarantees improvement are found. We describe our real-time implementation of the TLD framework and the P-N learning.

*Keywords — **learning from video, Long-term tracking, real-time, P-N learning, semi-supervised learning, TLD framework.***

I. INTRODUCTION

The video stream is processed at frame rate and the process runs continuously which is called long term tracking. In this video stream, various objects in video moving in and out of the camera range. Then the selection is done of the object to be detected by giving a bounding box of single frame.

To execute long term tracking number of problems are taken into consideration. The main problem is the detection of the object when it reappears in the camera's field of view. This problem is irritated by the fact that the object may change its appearance thus making the appearance from the initial frame irrelevant. So successful long term tracker should handle this change such as background clutter, partial occlusions and operate in real-time.

A tracker can provide weakly labeled training data for a detector and thus improve it during run-time. A detector can re-initialize a tracker and thus minimize the tracking failures. The design of a TLD framework that decomposes the long-term tracking task into three sub-tasks: tracking, learning and detection. The tracker follows the object from frame to frame. The detector localizes all appearances that have been observed so far and corrects the tracker if necessary. The learning estimates detector's errors and updates it to avoid these errors in the future.

While a wide range of trackers and detectors exist, there is no awareness of any learning method that would be suitable for the TLD framework. Such a learning method should: (i) deal with arbitrarily complex video streams where the tracking failures are frequent, (ii) never degrade the detector if the video does not contain relevant information and (iii) operate in real-time.

Tracking method is used to locate identical object in the sequent frames as much as possible. Multi-object tracking in real world is not easy because of occlusions, changing background, noise and so on. Recently tracking-by-detection methods become more and more popular. When the tracking is performed in a cluttered environment where multiple targets can be present , problems related to the validation and association of the measurements arise. Gating techniques are used to validate only measurements whose predicted probability of appearance is high.

 © 2016, IJREAM All Rights Reserved.

The detector is evaluated in every frame of the video. Its responses are analyzed by two types of "experts":(i) P-expert – recognizes missed detections, and (ii) N-expert – recognizes false alarms. The estimated errors augment a training set of the detector, and the detector is retrained to avoid these errors in the future. As any other process, also the P-N experts are making errors them self.

II. LITERATURE SURVEY

Long-term tracking is a complex problem that is closely related to tracking, detection and machine learning and in many cases it is studied from one point of view only. These terms are understood as follows. Tracking estimates the object motion between consecutive frames relying on temporal coherence in the video. Detection considers the video frames as independent and localizes all objects that correspond to an object model. Machine learning is often employed in both of these approaches. Trackers use machine learning to adapt to changes of the object appearance. Detectors use machine learning to build better models that cover various appearances of the object.

A. Tracking

Tracking is a task of estimating object motion. Various definitions are considered in the literature. In this section consideration of tracking as the task of estimating the object motion between consecutive frames is done. The implicit assumption of such algorithms is that the location of the object in the previous frame is known. This is in contrast to long-term tracking where this location might not be defined. In the following, the term tracking will be sometimes substituted with more accurate frame-to- frame tracking to emphasize the meaning.

B. Classification

One of the most distinctive properties of a tracking algorithm is the object state, which determines the variables that are estimated during tracking. Here use of the object state to classify tracking algorithms into five categories.

"Points" are often used to represent the smallest objects that do not change their scale dramatically. Algorithms that represent the object by a point will be called point trackers. Point trackers estimate only translation of the object. The estimation can be performed using frame-to-frame tracking , key-point matching , key-point classification , or linear prediction. Recent work is directed towards optimizing performance of these methods.

"Geometric shapes" such as bounding boxes or ellipses, are often used to represent motion of objects which undergo significant changes in scale. These methods typically estimate object location, scale and in-plane rotation, all other variations are typically modeled as the changes of the object appearance.

"Contours" are used to represent non-rigid objects. Parametric representation of contours has been used for tracking of human heads or arbitrarily complex shapes. Non-parametric representations have been applied for the tracking of people in sport footage , or various non-rigid objects including animals and human hands.

"Articulated" models are used to represent the motion of non-rigid objects consisting of several rigid parts. These models typically consist of several geometric shapes, for which relative motion is restricted by a model of their geometric relations. Articulated models have been used for tracking of humans or human arms.

"Motion field" is a non-parametric representation of the object motion which gives the displacement of every pixel of the object between two frames. Recent developments aim at producing long, continuous trajectories of image points. In this thesis the representation of this object state by a bounding box. This representation balances the tradeoff between the expressive power of the representation and the difficulty to reliably estimate the object motion. The related methods will be now analyzed in detail.

C. Detection

Object detection is the task of localizing objects in an input image. In long-term tracking, detection capability is essential as the object freely moves in and out of the camera field of view. Object detectors do not make any assumptions about the number of objects nor their location in the image. The objects are described by a model that is built in a training phase. At run-time, the model remains typically fixed. This section reviews the detection approaches starting from the simplest up to the most complex.

D. Machine learning

Machine learning reviews strategies for learning of sliding window-based object detectors. At the core of these detectors is a binary classifier, which classifies patches in an input image. During training, the image patches are interpreted as points in the feature space (training

 © 2016, IJREAM All Rights Reserved.

examples), and the goal is to find a decision boundary that separates the positive examples from the negative examples. Detectors are traditionally trained using supervised learning. While this setting is not directly relevant for long-term tracking of unknown objects, it becomes valuable when the class of the object is known in advance. For instance, if it is known that the object of interest will be a face, it is possible to train a face detector in advance.

E. Semi-supervised learning

A number of algorithms relying on similar assumptions have been proposed in the past including Expectation-Maximization, self-learning and co-training.

Expectation-Maximization (EM) is a generic method for finding estimates of model parameters given unlabeled data. EM is an iterative process, which in case of binary classification alternates over: (i) estimation of soft-labels of unlabeled data, and (ii) training a classifier exploiting the soft-labels. 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 in the feature space. EM is sometimes interpreted as a “soft” version of Self-learning .

Self-learning starts by training an initial classifier from a labeled training set, the classifier is then evaluated on the unlabeled 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. Self-learning has been applied to training of a human eye detector. However, it was observed that the detector improved more if the unlabeled data was selected by an independent measure rather than the classifier confidence. Rosenberg et al. 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 built 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 training is initialized by the training of two separate classifiers using the labeled examples. Both classifiers are then evaluated on unlabeled data. The confidently labeled 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 or moving object recognition.

III. AIM AND OBJECTIVE

A. Aim

Long-term tracking of unknown objects in a video stream. The object is defined by its location and extent in a single frame. In every frame that follows , the task is to determine the object’s location and extent or indicate that the object is not present. A wide range of trackers and detectors exist, there is no awareness of any learning method that would be suitable for the TLD framework.

B. Objective

Consider a video stream depicting various objects moving in and out of the camera field of view. Given a bounding box defining the object of interest in a single frame, this 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 full frame-rate and the process should run indefinitely. Thus this task is referred as long term tracking.

A number of algorithms related to long-term tracking have been proposed in the past. However, these typically make strong assumptions about the task. In particular, tracking based algorithms assume that the object moves on a smooth trajectory and typically fail if the object moves out of the image.

Detection-based algorithms assume that an object is known in advance and require a training stage. In contrast, this is to track an arbitrary object that moves in and out of the camera view immediately after initialization. The difficulty of the considered data and the achieved results.

The main objective is to implement the TLD framework and P-N learning. A range of trackers and detectors exist, there is no awareness of any learning method that would be suitable for the TLD framework. The detector is evaluated in every frame of the video with the use of the P-N learning.

IV. COMPARATIVE STUDY OF EXISTING SYSTEM

 © 2016, IJREAM All Rights Reserved.

Existing System	DataSet used	Parameter used for evaluation	Result
1. Face-TLD: Tracking-Learning-Detection Applied To Faces	Face should be tracked	Generic detector Validator	Detected face from complex environment
2. Multi-Object Tracking Based on Tracking-Learning-Detection Framework	Multiple Object should be tracked	P-expert and N-expert	Multiple object are get tracked in from single frame
3. Robust detection and tracking an object particle filter kalman filter	video from traffic	false negative and false positive	detection and tracking of lane marking using visual inputs from a camera
4. Tracking Learning and Detection of Multiple Objects using Static Camera	Multiple Object should be tracked using static camera	HSV colour color model Camshift	Multiple object are get detected and tracked by HSV
5. Visual Tracking with Online Multiple Instance Learning	Learning an adaptive appearance model for object tracking	Multiple Instance Learning	From frame the multiple objects are get tracked at same instance

V. ALGORITHM IMPLEMENTATION

A. Tracker

Step 1:Input: b_1, i,j;
Step 2:P1…….Pn //generate points(b_1)
Step 3:For all p_i do
Step 4:$P'_i \leftarrow LK(p_i)$
Step 5:$P''_i \leftarrow LK(p'_i)$
Step 6:$E_i \leftarrow (pi-p''_i)$
Step 7:$n_i \leftarrow NCC(w(p_i),w(p'_i))$
Step 8:end for
Step 9:$med_{NCC} \leftarrow median(n_i \ldots n_n)$
Step 10:$med_{FB} \leftarrow median(E_1 \ldots E_n)$
Step 11:if$med_{FB} > O_{FB}$ then
Step 12:$B_1=0$
Step 13:Else
Step 14:$C \leftarrow \{(p_i,p'_i)|p'_i=0, E<=med_{FB}, n_i>=med_{NCC}\}$
Step 15:$B_i \leftarrow transform(B_i,C)$
Step 16:End if

B. Camshift

Step 1: calculate the back projection image.
Step 2: Use the kalman filter to predict new location and ℓ_{ext}
Step 3: $W_{tmp} = A_{cs}(W_{t-1})$
Step 4: calculate similarity $P=P_{cs}(W_{t-1})$
Step 5: if p>T then
Step 6: $W_t=W_{tmp}$
Step 7: else
Step 8: $W_t=ALS\ (W_{t-1}, \ell_{ext})$
Step 9: end if
Step 10: update the kalman filter
Step 11: return W_t

C. Multiple instance learning:

Step 1. Crop out a set of image patches,$X^s = \{x|s> \|l(x)-l^*_{t-1}\|\}$, and compute feature vectors.
Step 2. Use MIL classifier to estimate p(y = 1|x) for x = X^s
Step 3. Update tracker location $l^*_t = l$
Step 4. Crop out two sets of image patches, $X^r=\{x|r> \|l(x)-l^*_t\|\}$ and $X^{r,b} = \{r|b> \|l(x)-l^*_t\|>r\}$
Step 5. Update MIL appearance model with one positive bag X^r and $X^{r,b}$ negative bags,
each containing a single image patch from the set $X^{r,b}$

VI. RESULT AND DISCUSSION

Performance analysis of P-N learning. The Initial Detector is trained on the first frame. The Final Detector is trained using the proposed P-N learning.

VII. CONCLUSION

The problem of tracking of an unknown object in a videostream, where the object changes appearance frequently moves in and out of the camera view. A new framework exists that decomposes the tasks into three components: tracking, learning and detection. The learning component analysis shows that an object detector can be

 © 2016, IJREAM All Rights Reserved.

trained from a single example and an unlabeled video stream using the following strategy:

I) evaluate the detector,

II) estimate its errors by a pair of experts, and

III) 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.

REFERENCES

[1] A. Blum and T. Mitchell, "*Combining labeled and unlabeled data with co-training*," Conference on Computational Learning Theory, p. 100, 1998.

[2] B. D. Lucas and T. Kanade, "*An iterative image registration technique with an"*, 1981.

[3] J. Shi and C. Tomasi, "*Good features to application to stereo vision*," International Joint Conference on Artificial Intelligence, vol. 81, pp. 674–679*track*," Conference on Computer Vision and Pattern Recognition, 1994.

[5] P. Sand and S. Teller, "*Particle video: Long-range motion estimation using point trajectories*," International Journal of Computer Vision,*- vol. 80, no. 1, pp. 72–91, 2008.

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

[7] D. Ramanan, D. A. Forsyth, and A. Zisserman, *"Tracking people by learning their appearance,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 65–81, 2007.

[8] P. Buehler, M. Everingham, D. P. Huttenlocher, and A. Zisserman, "Long term arm and hand tracking for continuous sign language TV broadcasts*," British Machine Vision Conference,*" 2008.

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

[10] M. Isard and A. Blake, *"CONDENSATION - Conditional Density Propagation for Visual Tracking,"* International Journal of Computer Vision, vol. 29, no. 1, pp. 5–28, 1998.

 © 2016, IJREAM All Rights Reserved.