Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 58 (2015) 478 – 485

Procedia Computer Science

Second International Symposium on Computer Vision and the Internet (VisionNet'15)

An Adaptive Approach for Validation in Visual Object Tracking

Jasper Princy Shinora W [a]*, Agilandeeswari L [b], Muralibabu K [c]

[a,b]School of InformationTechnology and Engineering (SITE), VIT University, Vellore – 632014, India

[c]Global Institute of Engineering and Technology, Vellore – 632001, India

Abstract

One of the major goals in the field of computer vision is to enable computers to replicate the basic functions of human vision such as motion perception and scene understanding. To achieve the goal of intelligent motion perception, much effort has been spent on visual object tracking. Research interest in visual object tracking comes from the fact that it has a wide range of real-world applications. The uncertainty of validating unpredictable features in object tracking is a challenging task in visual object tracking with occlusion and large appearance variation. To address this uncertainty, we propose an adaptive approach which uses updating model based on the occlusion and distortion parameters. In case of occlusion or large appearance variation, the proposed method uses backward model validation where it updates the invalid appearance and then validates the target feature model. If the target feature did not undergo any kind of clutter or distortions, it simply validates and then updates the appearance model using forward feature validation. The experimental results obtained from this adaptive approach demonstrate effectiveness in terms of OR (Overlap Rate) and Center Location Error, compared with other relevant existing algorithms.

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Keywords: Visual Object Tracking; Adaptive model; Object Occlusion; Backward Model Validation; Forward Model Validation

* Corresponding author. Tel.: 9600490949
E-mail address: shinora.jasper@gmail.com

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)
doi:10.1016/j.procs.2015.08.010

1. Introduction

Despite the importance of object tracking, in the field of computer vision, developing a robust tracking algorithm is still a challenging problem. In the past decades significant progress has been made and produced a variety of interesting results. However due to numerous factors such as partial occlusion, illumination variation, pose change, fast movement and background clutter, developing a robust tracking algorithm is still a challenging problem. Visual object tracking finds many practical applications in motion analysis, video surveillance, human-computer interaction, vehicle navigation and so on. The results produced from tracking is essential for some high level tasks such as suspicious target identification and object behaviour understanding in smart systems. Therefore it is indispensible to develop a robust tracking system that does validation and updating efficiently. Image analysis in tracking can be done by extracting some of the functional details from the captured images. Therefore, if there is a requirement for identifying an object, and notable robust characteristics of an object (color, pattern, edges, intensity, and structure). In this work, a generative and hybrid tracking method based on a novel robust and linear regression algorithm is proposed. In contrast to existing methods, the Least Soft-threshold squares algorithm [1] models the error term with the Gaussian distribution, can be solved efficiently. We derive Least Soft-threshold Squares (LSS), based on maximum joint likelihood of parameters to measure the difference between an observation sample and the dictionary. Compared with the distance derived from ordinary least squares methods, the proposed metric is more effective in dealing with outliers. Several tracking algorithms iVT [2], Struck [3], TLD [4] were developed in the past to efficiently handle the controversies. The target feature identification can be done using particle filter framework where the particles are assigned with weights and re-sampled contiguously from the model pool. When large appearance change (caused by appearance variation or occlusion) is detected, the appearance data is not immediately labelled invalid, but the appearance model is duplicated into the model pool where the duplicated model stays steady and the original one keeps updating with the incoming data.

However in order to handle error accumulation, Mathews et.al. [5] proposes an algorithm for updating, based on training the appearance model. The appearance model is applied to distinguish the appropriate target features. This method is named as forward feature validation in updating the appearance model, as the feature data are updated and then validated. In case of unpredictable target appearance variation forward feature based validation algorithm loses track of target object. Therefore in each step, the target appearance information in the incoming frame backward-check all the appearance models in the model pool. The chosen model which adaptively uses forward as well as backward validation approach is detected as most valid and accurate to estimate the target states and is updated with the estimated target appearance. The experimental results obtained shows that the proposed algorithm which combines both BVT [10] and FVT, in case of occlusion, appearance variation, fast movement, can obtain accurate and consistent tracking results, compared with existing relevant algorithms of visual object tracking.

The rest of the paper is organized as follows: Section II discusses about the related works, Section III explains the proposed work, Section IV represents the Experimental Analysis and Section V describes the Conclusion.

Nomenclature

A	BVT - Backward Validation Tracking
B	IVT - Incremental Visual Tracking
C	IPCA - Incremental Principal Component Analysis

2. Related Works

Incremental learning for robust visual tracking [2]: In this paper, initialized a dictionary using local low-rank features to represent the appearance subspace for the object. In this way, each candidate can be modeled by the sparse linear representation of the learnt dictionary. Then by incrementally updating the local dictionary and learning sparse representation for the candidate, we build a robust online object tracking system. Furthermore, in contrast to the traditional holistic dictionary, the local low-rank features based dictionary contains abundant partial information and spatial information. Experimental results on challenging image sequences show that our method consistently outperforms several state-of-the-art methods.

Struck [3]: This track presents a framework for adaptive visual object tracking based on structured output prediction. It explicitly allows the output space to express the needs of the tracker, by avoiding the need for an intermediate step of classification. A kernel structured output support vector machine (SVM), which is learned online provides the adaptive tracking of the target.

Multiple Instance Learning Track [8]: MIL Track that uses a novel Online Multiple Instance Learning algorithm, by taking into account of weak classifier. MIL algorithm for object tracking is based on the instance probability optimization, achieves superior results with real-time performance. The MIL framework allows update of the appearance model with a set of image patches, even though it is not known which image patch precisely captures the object of interest. MIL track used greedy feature selection method. It requires updating of the selected classifier 'M' times and the objectives for tracking and classification are not explicitly coupled.

Visual Tracking Decomposition [9]: This algorithm is proposed that can work robustly in a challenging scenario such that several kinds of appearance and motion changes of an object occur at the same time. The algorithm is based on a VTD scheme for the set of sample features where the observation model is broken down into multiple basic observation models that are again fabricated by sparse principal component analysis (SPCA). Each basic observation model covers a particular appearance of the target sample. This track, however drifts away and is not effective in separating two objects that are similar and near-by.

From the above literature survey, the shortcomings of the trackers are handled efficiently by combining the pros of both forward feature validation tracking and backward model validation tracking.

3. Proposed System

The proposed visual object tracking is based on particle filtering. In our work, we assume the target object to be tracked is marked/ detected manually. The target candidate states are predicted with a motion model, for each incoming frame. To estimate the actual target state of the object in successive frames, the appearance features of the target are extracted by IPCA and least soft-squares method are computed, and used along with the appearance model pool.

3.1. Particle Filter Framework

This framework helps in estimating the posterior probability sequentially by several random particles and weights related to them considering equation (1), regardless of the distribution of prediction and observation functions [10]. In each frame, several particle/sample states (states can be defined by the position, scale, skew, etc.) are predicted based on the tracking results from the previous frame.

$$P(a_x|b_{1:t})\ \alpha\ P(b_x|a_x)\int P(a_x\ |a_{x-1})P(a_{x-1}|b_{1:x-1})d\,a_{x-1} \tag{1}$$

From the above equation, let a_x be the state variable of the target object at time t, and it is represented by the affine transformation parameters, a_x =(a_x, b_x, θ_x, α_x, ϕ_x), where (a_x, b_x) denotes the center location of the bounding box around the target; θ_x is the rotation angle; s_x is the scale; α_x is the aspect ratio; ϕ_x is the skew direction at time x.

3.2. Motion Model

Since the target object is in motion, the state variable a_x of rotation angle x, scale s_x, aspect ratio α_x and skew direction ϕ_x would change over time. Different from the previous works, which predict a_x by simply adding a zero-mean Gaussian distribution independent to a_{x-1} in the previous frame [10], we demonstrate this change from a_{x-1} to a_x using equation (2), and by considering the velocity of the state as given in equation (3).

$$a_x = \quad \mathrm{a}_{x-1} + \hat{\mathrm{v}}_{x-1} + \varepsilon(\mathrm{x}, \mathrm{y}, \theta, \mathrm{s}, \alpha, \phi) \tag{2}$$

$$\hat{\mathrm{v}}_x = \begin{cases} \hat{\mathrm{a}}_x - \hat{\mathrm{a}}_{x-1} & x > 0 \\ 0 & x = 0, \end{cases} \tag{3}$$

where, $\hat{a}_x$ is the target vector estimated and $\varepsilon(x, y, \theta, s, \alpha, \phi)$ is the Gaussian noise vector.

3.3. Appearance Model

The appearance of the object under tracking may change over time due to changes of shape, view or illumination, and the appearance variation would lead the tracker to lose tracking. In this work, we extract the feature for visual targets based on a feature from IPCA. The IPCA [10] based feature is able to adjust illumination variation and out-of-plane rotation with an eigen space-based representation, since the global appearance of objects under different illumination lies approximately in a low-dimensional subspace.

IPCA [10]: The IPCA algorithm is used to build multi-view based target appearance model. IPCA model generates the probability of a gray-scale image template I_t and is governed by a Gaussian distribution.

3.4. The Appearance Model Selection Tracking Algorithm

Input: The initial state of target x_0. The maximum model number of the model pool M and the other parameters.
Step1: Initialization of the model a_0 with the labelled/detected target appearance predicted by x_0;
Step2: Estimating the Motion Model, for $a_x - 1$.
Step3: Validating the appearance model, a_0
Step4: Initialization of the model with IPCA (Incremental Principal Component Analysis)
Let I_t be the gray scale image template obtained from x_t. The probability obtained from IPCA provides the similarity measure between I_t and target appearance x_t.
Step5: Model Updating: Update a_x with the estimated target appearance at x_t ;
End
Output: Estimated Target State $\hat{x}_t$ at time t.

3.5. The Adaptive Validation Model

The proposed approach focuses either on updating the invalid appearance in case of occlusion or large appearance variation and then validate the target feature model or, it simply validates and then updates the appearance model, if the target feature do not undergo any kind of clutter or distortions. It adaptively chooses to use forward feature or backward model validation process. This can be achieved by assigning thresholds Tr1, Tr2 to detect the occlusion and large appearance variation that are notable, affecting the appearance model updating in object tracking.

Firstly, the appearance model pool $\{A_x\}$, contains M appearance model represented by a feature set of IPCA, at time t, where at least one of the candidates can represent the target accurately. Therefore the true target appearance information is contained in the bag of candidates in the model pool. If the target is visible in the frame, the model with invalid updates will get a low reliability and the model with accurate updates will get a high reliability. After occlusion some of the candidates may dominate by representing the occlusion, in the computation of reliability function. But as the target is visible in the frame most of the time, it wouldn't return a higher value as the updating during occlusion gives in only a smaller part of the model.

In the starting stage of the object track, the target appearance model in the first frame is only initialized and all the remaining appearance models are empty. Generally the appearance model is updated incrementally. Whenever a larger appearance variation is detected, the updated target appearance models are stored in the empty models. The large appearance variations can be detected based on the distance $\hat{d}_1, \hat{d}_2$ of the target estimated $\hat{a}_x$ to the feature set of IPCA and the thresholds related. The models M, in the model pool are initialized, and after a while (M+1)th model is created and stored in place of model M in the model pool which produces the lowest likelihood at that instant. Thus the size of the model pool would not be more than M. However the storing of large values of M tends to result in high computational complexity. To avoid frequent initializations of appearance model, the threshold values should be chosen in way that the smaller variation of the distance $\hat{d}_1, \hat{d}_2$ is tolerated. When the tracking begins, the updating of the appearance model is done incrementally if $\hat{d}_1 < Tr1$ AND $\hat{d}_2 < Tr2$. When the values of sudden increasing distance $\hat{d}_1 \geq Tr1\ OR\ \hat{d}_2 \geq Tr2$, a new appearance model is initialized and then updated using the target appearance. Therefore the target feature model is updated and validated.

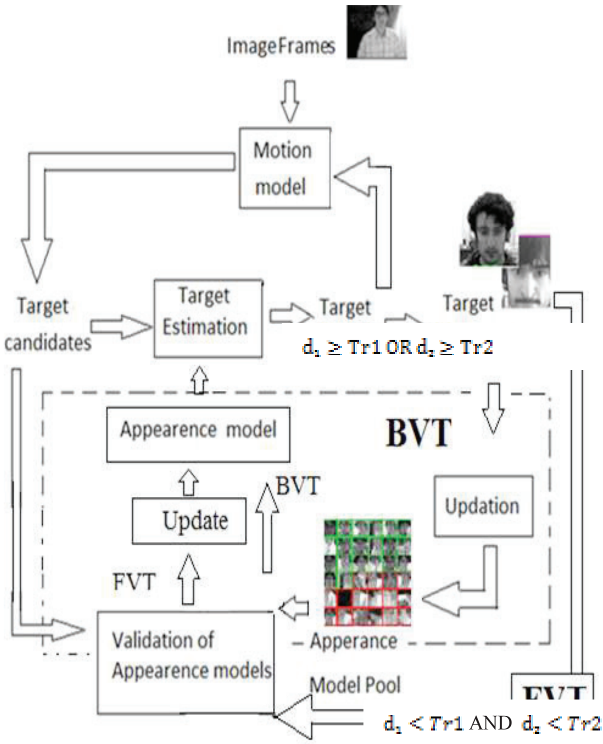

Fig. 1. The Proposed System

Table 1. Feature Comparison

Challenges	IPCA	Sparse Representation
Illumination Variation	Robust	Robust
Shape Deformation	Sensitive	Sensitive
Scale Variation	Robust	Robust
Image Description	Multi-views of Texture	Global Texture
Challenges	IPCA	Sparse Representation

4. Experimental Analysis

For the evaluation of the proposed algorithm, we use two sequences of CAVIAR datasets, one face tracking sequence captured in backward validation tracking model and sequences of object tracking from videos David Indoor, Occlusion1. These include both indoor and outdoor scenes with challenging appearance and illumination variation, fast movement, out of plane rotations and occlusion. We compare the proposed algorithm with the available visual object trackers such as iVT [2], MIL [8] and BVT [10].
The proposed visual object tracking is implemented using MATLAB tool. On an average PC, this algorithm runs at 3 frames per second and has the requirement of Intel 733 MHz Pentium III Processor and 16GB memory. We include patch size value in

IPCA model; model pool size M=4 and the threshold values Tr1 and Tr2, chosen empirically.

Table 2. Characteristics of the Video Sequences

Video sequence	Type	Frames per second
Caviar1	Slow motion	3
Caviar2	Medium	4
David Indoor	Slow motion	3

4.1. Qualitative Comparison

Here, we show the comparison between the proposed method and the state-of-art algorithms. The performances of the trackers on different sequences are analysed.

Severe Occlusion:

The video sequences with heavy occlusion and frequent occlusion (e.g., Caviar1) are used in the qualitative analysis of the trackers such as BVT, MIL and iVT with the proposed tracker. Our tracker performs efficiently as it stores the target appearance information, the target is recaptured. But most of the other trackers fail to track the target object effectively.

Fig. 2. Caviar1 with occlusion

Background clutter and Illumination Variation:

David Indoor is a typical sequence taken with dull lighting and illumination variation. The proposed system uses color histogram which works efficiently in illumination variations. In the fig4 that the proposed model trackers the target well compared with the other trackers such as BVT, MIL and iVT.

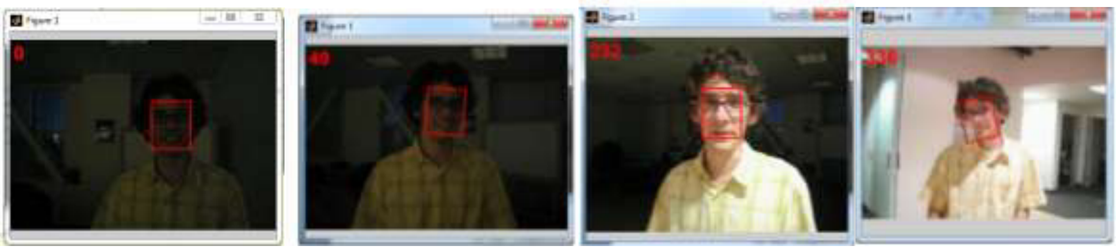

Fig. 3. David Indoor with background clutter and occlusion

Quick Appearance Variation:

To compare the quick appearance variation, the sequences with the deformation of the target or rotations (e.g., Occlusion1) is taken and tested. The proposed model uses IPCA model to get robust target tracking by every batch of frame. The results are shown in figure. 4 prove that our tracker performs with stability compared with the other related algorithms [2], [3], [4].

Fig. 4. Occlusion1 with appearance variation and occlusion

Fig. 5. Caviar2 with occlusion and fast movement

4.2. Quantitative Comparison

We give a comparison between the proposed method and the BVT, MIL [8] and iVT [2] trackers quantitatively to verify the tracking performance. The performance is evaluated in pixels based on the average overlap rate. The center of the box, bounding the target is manually labelled. The distance could be expressed in terms of the overlap between object and the dictionary and the threshold could be set to zero overlap. The Euclidian distance, in 2D or in real 3D coordinates, between target centers and hypotheses is used for object centroid trackers.

Table 3. Average Center Location Error (In Pixel). The Best Result Is Shown In **Bold** Font.

Sequences	**iVT [2]**	**MIL[8]**	**Our Approach**
Caviar1	45.2	48.5	**1.419**
Caviar2	8.6	70.3	**2.35**
David Indoor	**3.1**	34.3	**3.154**
Occlusion1	**10.2**	**14.1**	**3.1214**

Table 4. Average Overlap Rate. The Best Result Is Shown In **Bold** Font.

Sequences	**iVT [2]**	**Our Approach**
Caviar1	0.28	**0.89**
Caviar2	0.45	**0.80**
David Indoor	0.69	**0.76**
Occlusion1	**0.85**	**0.86**

It can be seen from the table that the performance improvement of the proposed tracking model is distinct compared with the state-of-art tracking algorithms.

5. Conclusion

In this paper, we have proposed algorithm for visual object tracking in which the validation and updating are performed based on the thresholds of the input video sequences. The tracking using backward model validation works by updating of the appearance model first and then using the information from the incoming frame to check whether the updating of the appearance model if the model is valid or not. This method avoids the uncertainty in validation of model updating process in existing tracking algorithms, which use the current appearance model to validate the estimated target appearance features. The purpose of model updating is to include new changes in appearance variations to the updated model, while the existing methods do not have the provision to include such changes in model updating by "validating" changes with the existing model. If there is no large appearance variation in the incoming video sequences, then the validation is done first and then it is updated. This is done by using the forward feature validation approach. As demonstrated, the proposed method remedies this drawback, and provides a new way by making use of the advantages of both forward and backward model validation methods in solving the error accumulation problem in appearance model updating of visual object tracking. In addition, we have integrated the colour based feature that is by color histogram with texture based feature adaptively, based on IPCA. Both quantitative and qualitative analysis on challenging image sequences show that the proposed tracker performs favourably compared with several state-of-the-art algorithms. Experimental results on challenging video sequences have demonstrated the proposed algorithm can obtain much better performance than other five existing trackers, especially in scenarios where the occlusion and appearance variation occurs together.

References

1. D. Wang,Huchuan Lu, M-H Yang, "Least Soft-thresold Squares Tracking", In CVPR, 2013.

2. A.Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," International Journal of Computer Vision, vol.77, no. 1-3, pp. 125-141, 2008.K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in Computer Vision–ECCV 2012, pp. 864877, Springer , 2012.

3. S. Hare, A. Saffari, and P. H. S. Torr. Struck: Structured output tracking with kernels. In ICCV, pages 263–270, 2011.

4. Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 7, pp. 1409-1422, 2012.

5. L. Matthews, T. Ishikawa, and S. Baker, "The template update problem," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, no. 6, pp. 810-815, 2004.

6. Kaihua Zhang, Lei Zhang and Ming-Hsuan Yang, "Real-Time Object Tracking via Online Discriminative Feature Selection", IEEE, 2013. Santner, C. Leistner, A. Saffari, T. Pock, and H. Bischof. PROST: Parallel robust online simple tracking. In CVPR, pages 723–730, 2010.

7.B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in Computer Vision and Pattern Recognition, 2009 . CVPR 2009. IEEE Conference on, pp. 983-990, IEEE, 2009.

8. J. Kwon and K. Lee, "Visual tracking decomposition," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp.1269-1276, IEEE, 2010.

9. Y. Yuan, S. Emmanuel, Y. Fang and Weisi Lin. Visual Object Tracking based on Backward Model Validation. IEEE Transactions on Circuits and Systems for Video Technology, 2014.

10. H. Grabner, C. Leistner, and H. Bischof, "Semi-Supervised On-line Boosting for Robust Tracking," in Computer Vision–ECCV 2008, pp. 234–247, Springer, 2008.