Neurocomputing 144 (2014) 581–595

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Structured partial least squares based appearance model for visual tracking

Jia Yan [a], Xi Chen [a,b,*], Dexiang Deng [a], Qiuping Zhu [a]

[a] Department of Electrical Engineering, School of Electronic Information, Wuhan University, Wuhan 430072, Hubei, PR China
[b] International School of Software, Wuhan University 430072, Hubei, PR China

ARTICLE INFO

*Article history:*
Received 6 June 2013
Received in revised form
7 February 2014
Accepted 14 February 2014
Communicated by Jungong Han
Available online 13 April 2014

*Keywords:*
Appearance model
Partial least squares
Visual tracking
Structural information

ABSTRACT

The appearance model is an important issue in the visual tracking. In this paper, we present a structured partial least squares (SPLS) based appearance model for object tracking which not only discriminates the target from its surrounding background but also able to tolerate its appearance variations. The target is represented as a low-dimensional feature vector, which is the structured combination of the subspaces learned by SPLS. In order to capture the variations in appearance during the tracking, we update the appearance model by taking the estimated class labels of the new targets obtained online into account. Furthermore, aiming to alleviate the drift problems, negative samples generated by disordering the structured appearance model are added in the training samples. Both qualitative and quantitative evaluations on challenging benchmark video sequences demonstrate that the proposed tracking algorithm performs favorably against several state-of-the-art methods.



## 1. Introduction

Visual tracking is essential for applications like activity analysis [1], man–machine interaction [2] and visual surveillance [3]. However, for many real-world problems, the ambiguities make it difficult to develop accurate trackers, such as intrinsic (e.g., pose change and shape deformation) and extrinsic factors (e.g., varying illumination, motion blur and occlusions).

In the literature, a variety of tracking algorithms have been proposed and detailed reviews can be found in [4–6]. Usually, a tracking method typically consists of three components: object appearance model, motion model and search strategy. In this paper, we study the problem of designing a robust online appearance model that confronts the aforementioned difficulties. Hence, we only discuss key issues related to appearance models. Appearance modeling is adopted to represent the tracked target using information extracted from the target region. In general, the recently proposed appearance models can be categorized into two classes: discriminative models and generative ones.

Discriminative models address the tracking problem as a classification problem which aims to distinguish the target from the background online. This method is also termed as tracking-by-detection [7–11], in which a binary classifier separates target from background in the continuous frames. To handle appearance changes, the classifier is updated incrementally using the new information over time. However, frequent updates may gradually result in drifts due to accumulated errors. In order to alleviate the drifting problem, Grabner and Leistner [9] propose a semi-supervised boosting method to combine decisions of a given prior and an on-line classifier. Babenko et al. [10] propose a tracking method to treat ambiguous positive and negative samples into bags to overcome the drifting based on the online multiple instance learning method.

Generative models focus on learning an appearance model and formulate the tracking problem as finding the target observation most similar to the learned appearance or with minimal reconstruction error. These methods are based on either templates [12–14] or subspace models [15–17]. Most of these methods use the holistic model to represent the target and hence cannot handle partial occlusion or distracters. Adam et al. [14] utilize multiple fragments to design an appearance model which is robust to partial occlusions. Recently, sparse representation has been introduced to the tracking task [18–22], demonstrating good performance to partial occlusions, illumination changes and pose variations. However, these generative models do not take the background information into consideration, therefore throwing away some very useful information that can help to discriminate object from background.

In addition, several algorithms that exploit the advantages of both generative and discriminative models have been proposed

* Corresponding author at: International School of Software, Wuhan University 430072, Hubei, PR China.
Tel.: +86 27 68778696.
*E-mail address:* robertchenxi2011@gmail.com (X. Chen).

http://dx.doi.org/10.1016/j.neucom.2014.02.034
0925-2312/

[23–26]. Recently, Wang et al. [26] propose to use partial least squares (PLS) analysis to learn low-dimensional discriminative feature subspace by modeling the correlation of object appearance and class labels from positive and negative samples. The learned subspace is then utilized to construct an appearance model. During tracking, the candidate sample which has the smallest distance with the learned appearance model is chosen as the new target. This appearance model is discriminative since the PLS takes both target and surrounding background into consideration. It is also generative because the target can be well represented by the feature subspaces. The PLS tracker has demonstrated good performance to handle appearance changes. However, the learned appearance model in the PLS tracker does not take into account any information about the corruption or occlusion.

This paper proposes a novel structured PLS (SPLS) based appearance model, which is robust against occlusion and pose variations. The particular structural information is incorporated in the entries of the original feature space, and the appearance model generated by the SPLS is a structured combination of the learned subspaces. We also propose an effective method to update the appearance model online with an occlusion handing scheme rather than simply update the model as a whole. To further alleviate the drift problem, we enrich the training samples by including the generated negative samples. Finally, the appearance model is embedded into the particle filter framework to form a tracking algorithm. Empirical results on challenging video sequences demonstrate the superior performance of our method in robustness and stability to state-of-the-art methods.

The remainder of the paper is organized as follows: We give a brief summary of the related methods aiming at handling target appearance changes in Section 2. Section 3 details our proposed appearance model based on SPLS. Following this, in Section 4, we present our experimental results on several challenging sequences. Finally, we conclude this paper in Section 5.

## 2. Related work

### 2.1. Adaptive appearance models

Handling appearance variations is a very challenging problem for visual tracking. Since it is difficult to develop a static representation invariant to all appearance changes, adaptive appearance models are needed for robust tracking performance. In this subsection, we review the related tracking algorithms which also employ the adaptive appearance models.

In [23], holistic templates are incorporated to construct a sparsity-based discriminative classifier and local representations are adopted to form a robust histogram that considers the spatial information among local patches. Zhang et al. [25] propose to simultaneously achieve discriminating the target from its background and being robust to its appearance variations separately based on sparse representation in a particle filter framework. The target's appearance variants are modeled as the error in the linear representation. The error is spanned by a set of basis functions which are learned online when new observations are available over time. In [27], discriminative stable regions are extracted from target by using the criterion of maximal entropy and these regions present high appearance stability and strong discriminative power to tolerate the appearance variations.

### 2.2. Subspace representation

Subspace representation aims at adapting the appearance variations with a low-dimensional subspace based on the core desire for dimensionality reduction. In the tracking literature, numerous methods have been proposed to learn feature subspace to represent the appearance models, from the static subspace representation to adaptive learning subspace ones. In this subsection, we review some of these methods.

The eigenspace representations are used for evaluation the similarity between the observed images and the model in [28]. However, this method relies on training offline that may lead to failure if the target variations are outside of the set of training samples. The rich sample sets are often not available in the tracking tasks, therefore, the adaptive model is needed to learn the target variations online. Lin et al. [16] propose the Fisher linear discriminant (FLD) based learning subspace model, which renders the largest separation between the class means of samples and the smallest variance within each class. In [15], object tracking is by incrementally learning a linear subspace model with both mean update and eigenbasis update. Consequently, the focus has been made on developing the image-as-matrix learning algorithm for effective subspace analysis. Li et al. [29] employ a three-dimensional temporal tensor subspace learning for visual tracking. In [30], Li et al. present an online subspace learning algorithm based on the Log-Euclidean Riemannian metric. Wu et al. [31] present a tracking approach that incrementally learns a low-dimensional covariance tensor representation. These methods, benefiting from the adaptive learning or updating scheme, usually exhibit superiority compared to the static models. However, the above algorithms, static or adaptive, usually do not have mechanisms to handle occlusions and they could suffer from failure caused by occlusion in a long duration. Wang et al. [20] introduce l1 regularization into the PCA reconstruction, and develop a novel algorithm to represent an object by sparse prototypes that account explicitly for data and noise. In this paper, we also explicitly take the occlusion into account in the subspace learning and model updating.

### 2.3. Structural information based representation

Many methods prefer to integrate structural information into the appearance representation of the target to better the tracking results. In [32], a local sparse representation scheme is employed to model the target appearance and then represent the basis distribution of the target with the sparse coding histogram. Due to the representation of local patches, their method performs well especially in handling the partial occlusion. In [33], the sparse coding of overlapped local image patches with a spatial layout helps locate the target more accurately and handle partial occlusion. In addition, the dictionary for local sparse coding is generated from the dynamic templates, which are updated online based on both incremental subspace learning and sparse representation. In [19], the appearance of an object is modeled as a sparse linear combination of structured union of subspaces in a basis library, which consists of a learned Eigen template set and a partitioned occlusion template set. Further, Block Orthogonal Matching Pursuit is adopted to solve the structured sparse representation problem to reduce the computational cost.

## 3. Structured partial least squares based appearance model

In this section, we give the details of the proposed appearance model based on SPLS. We briefly review the PLS analysis first, and then formulate visual tracking as the representation problem with our proposed SPLS. The appearance model update scheme is also introduced together with the generation of the training samples.