Pattern Recognition 47 (2014) 3828–3840

Contents lists available at ScienceDirect

Pattern Recognition

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

Discriminative multi-task objects tracking with active feature selection and drift correction

Baojie Fan [a,*], Yang Cong [b], Yingkui Du [b]

[a] College of Automation, Nanjing University of Posts and Telecommunications, No.9 Wenyuan Road, Nanjing 210046, China
[b] State Key Laboratory of Robotics, Shenyang Institute Automation, Chinese Academy of Sciences, Shenyang 110016, China

ARTICLE INFO

Article history:
Received 20 October 2013
Received in revised form
16 June 2014
Accepted 19 June 2014
Available online 8 July 2014

Keywords:
Discriminative multi-task objects tracking
Active feature selection
Drift correction
Jointly sparsity

ABSTRACT

In this paper, we propose a discriminative multi-task objects tracking method with active feature selection and drift correction. The developed method formulates object tracking in a particle filter framework as multi-Task discriminative tracking. As opposed to generative methods that handle particles separately, the proposed method learns the representation of all the particles jointly and the corresponding coefficients are similar. The tracking algorithm starts from the active feature selection scheme, which adaptively chooses suitable number of discriminative features from the tracked target and background in the dynamic environment. Based on the selected feature space, the discriminative dictionary is constructed and updated dynamically. Only a few of them are used to represent all the particles at each frame. In other words, all the particles share the same dictionary templates and their representations are obtained jointly by discriminative multi-task learning. The particle that has the highest similarity with the dictionary templates is selected as the next tracked target state. This jointly sparsity and discriminative learning can exploit the relationship between particles and improve tracking performance. To alleviate the visual drift problem encountered in object tracking, a two-stage particle filtering algorithm is proposed to complete drift correction and exploit both the ground truth information of the first frame and observations obtained online from the current frame. Experimental evaluations on challenging sequences demonstrate the effectiveness, accuracy and robustness of the proposed tracker in comparison with state-of-the-art algorithms.

1. Introduction

Given the initialized position and size of a target in the first frame (or former frames) of a video, the goal of visual tracking is to estimate the states of the moving target in the subsequent frames. This active topic has been extensively studied in computer vision due to its important pole in many applications such as automated surveillance, robot navigation, video indexing, traffic monitoring, human–computer interaction and so on. Despite that much progress has been made in recent years [1–20], developing a robust tracking algorithm is still a challenging problem due to the following numerous factors: large and dynamic appearance changes caused by illumination, rotation, and scaling, abrupt motion, background clutters, partial or full occlusions, pose variation and shape deformation.

Inspired by the success of sparse representation-based face recognition [18], Mei and Ling [27] propose a novel L1 tracker that uses a series of target templates and trivial ones to model the tracked target, where the target templates are used to describe the tracked object and trivial templates are used to deal with outliers (e.g., partial occlusion) with the sparse constraints. The tracker represents each target candidate as a sparse linear combination of dictionary templates that can be dynamically updated, and its corresponding likelihood is determined by minimizing the reconstruction error. This representation has been shown to be robust against partial occlusions, which improves the tracking performance. Recently, based on the milestone work, there are several methods have been proposed to improve the L1 tracker in terms of both speed and accuracy [27–36], such as using accelerated proximal gradient algorithm [29], replacing raw pixel templates with orthogonal basis vectors [32,33], modeling the similarity between different candidates [37], to name a few. Despite of demonstrated success, the above mentioned L1 trackers have the following shortcomings.

Firstly, sparse coding based trackers perform computationally expensive L1 minimization at each frame. Although recent efforts have been made to speed up this tracking paradigm [27,34], these methods assume that sparse representations of particles are independent and ignore their relationships, which can help and improve the tracking performance.

Second, the trivial templates lack the discriminative ability, and they are used to model any kind of image regions whether they are

* Corresponding author.
E-mail address: jobfbj@gmail.com (B. Fan).

http://dx.doi.org/10.1016/j.patcog.2014.06.015

from the target objects or the background. Thus, the reconstruction errors of images from the occluded target and the background may be both small. As a result of generative formulation where the sample with minimal reconstruction error is regarded as the tracking result, ambiguities are likely to accumulate and cause tracking failure. Overall, the trivial templates decrease the efficiency and effectiveness of the L1 tracking algorithms.

Third, some appearance models (or dictionary templates) are only designed to represent the object. The background pixels in the target templates do not lie on the linear template subspace. The scale of the reconstruction error from background pixels is often larger than that from the target pixels, which might affect the accuracy of the sparse representation. If appearance models consider both the object and its local background, the trackers may perform better than the former ones. In this paper, we focus on the discriminative appearance model since it is the important component of the tracking algorithm.

Furthermore, during the process of the tracking, owing to the appearance variations of the target object and the background, online update schemes is required. Numerous successful approaches have been developed [6,10,3,19,21]. However, they introduce potential drifting problems due to the accumulation of errors during the self-updating.

We observed that target can be reliably represented by the templates of target and background, and only a few part of templates can discriminate the target and background, because they treat tracking as a binary classification problem, which separates target from its local background via a discriminative classifier. Motivated by [30,38], considering above existing problems and our observations. We propose a discriminative multi-task objects tracking method with active feature selection and drift correction. The developed method object tracking in a particle filter framework is viewed as multi-task discriminative tracking. The tracking algorithm starts from the active feature selection scheme, which adaptively chooses suitable number of discriminative features from the tracked target and background in the dynamic environment. Based on the selected feature space, we construct the discriminative dictionary templates that are updated dynamically. Only a few of dictionary templates are used to represent all the particles at each frame. In other words, all the particles share the same dictionary templates. While learning the reliable representation of each particle is viewed as an individual task. The particle that has the highest similarity with the dictionary templates is selected as the next tracked target state. This jointly sparsity and discriminative learning can exploit the relationship between particles and improve tracking performance. To alleviate the visual drift problem encountered in object tracking, a two-stage particle filtering algorithm is proposed to complete drift correction and exploit both the ground truth information of the first frame and observations obtained online from the current frame.

The main contributions of this paper are as follows:

(1) Active feature selection scheme is used to adaptively choose suitable number of the discriminative features from the tracked target and its local background.
(2) In this paper, object tracking in a particle filter framework is viewed as a discriminative multi-task sparse learning problem. As opposed to sparse coding based trackers [27–30] that handle particles independently, we mine the relationships among different particles and learn their representations jointly with the same discriminative dictionary atoms, which is constructed by the selected discriminative features and updated dynamically.
(3) The initial information of the first frame is incorporated into the tracking framework to correct the tracking drift and improve the tracking performance.

The paper is organized as follows: in Section 2, we summarize the works most related to ours. The detailed description of the proposed tracking approach is presented in Section 3. It contains the principle of our method and its advantages over state-of-the-art methods in detail. Section 4 gives the detailed experiment setup and results. Finally, Section 5 concludes the paper.

2. Related work

Much work has been done in object tracking. In this section, we only briefly review nominal tracking methods and those that are the most related to our tracker. We focus specifically on tracking methods that use particle filters, sparse representation and general multi-task learning methods. For a more thorough survey of tracking methods, we refer the readers to [1–4].

Existing tracking algorithms can be roughly categorized as either generative or discriminative.

2.1. The generative trackers

The generative methods represent the target as an appearance model. The tracking problem is formulated as searching for the regions which are the most similar to the tracked targets. These methods are based on either templates [5,6,8,9,12] or subspace models [7,10,11]. Popular generative trackers include eigentracker [5], mean shift tracker [6], fragment-based tracker [7], incremental tracker (IVT) [8], and visual tracking decomposition (VTD) tracker [9]. Black and Jepson [5] learn a subspace model offline to represent target at predefined views and build on the optical flow framework for tracking. The mean shift tracker [6] is a popular mode-finding method, which successfully copes with camera motion, partial occlusions, clutter, and target scale variations. The Fragment tracker [7] aims to solve partial occlusion with a representation based on histograms of local patches. The tracking task is carried out by accumulating votes from matching local patches using a template. But, this template is static, and it cannot to handle changes in object appearance. Ross et al. [8] learn an adaptive linear subspace online for modeling target appearance and implement tracking with a particle filter. However, IVT is less effective in handling heavy occlusion or non-rigid distortion. Kwon et al. [9] extend the classic particle filter framework with multiple dynamic observation models to account for appearance and motion variation. Nevertheless, due to the adopted generative representation scheme, this tracker is not equipped to distinguish between the target and its local background.

2.2. Discriminative trackers

Discriminative methods cast the tracking as a classification problem that distinguishes the tracked targets from their surrounding backgrounds. The trained classifier is online updated during the tracking procedure. Discriminative tracking algorithms use the information from both the target and the background. Examples of discriminative methods are ensemble tracking [13], on-line boosting (OAB) [16], semi-online boosting [17], online multiple instance learning tracking [18], adaptive metric differential tracking [23], P-N learning tracker (PN) [24], Compressive Tracking (CT)[25].

In ensemble tracking [13], a set of weak classifiers are trained and combined for distinguishing the object and the background. The features used in [13] may contain redundant and irrelevant information which affects the classification performance. To improve the classification performance, feature selection is needed. Collins et al. [14] have demonstrated that online selecting discriminative features can greatly improve the tracking performance. Inspired by the advances in face detection [15], many boosting feature selection