Image and Vision Computing 38 (2015) 24–32

Contents lists available at ScienceDirect

Image and Vision Computing

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

Visual tracking based on online sparse feature learning ☆

Zelun Wang [a], Jinjun Wang [b,*], Shun Zhang [b], Yihong Gong [b]

[a] Texas A&M University, College Station, TX 77843, USA
[b] Xi'an Jiaotong University, 28 Xianning West Road, Xi'an, Shaanxi, 710049, China

A R T I C L E I N F O

Article history:
Received 17 July 2014
Received in revised form 19 December 2014
Accepted 2 April 2015
Available online 24 April 2015

Keywords:
Visual tracking
Sparse coding
Sparse feature
Bayesian classifier
Haar-like features

A B S T R A C T

Various visual tracking approaches have been proposed for robust target tracking, among which using sparse representation of the tracking target yields promising performance. Some earlier works in this line used a fixed subset of features to compress the target's appearance, which has limited modeling capacity between the target and the background, and could not accommodate their appearance change over long period of time. In this paper, we propose a visual tracking method by modeling targets with online-learned sparse features. We first extract high dimensional Haar-like features as an over-completed basis set, and then solve the feature selection problem in an efficient L_1-regularized sparse-coding process. The selected low-dimensional representation best discriminates the target from its neighboring background. Next we use a naive Bayesian classifier to select the most-likely target candidate by a binary classification process. The online feature selection process happens when there are significant appearance changes identified by a thresholding strategy. In this way, our proposed method could work for long tracking tasks. At the same time, our comprehensive experimental evaluation has shown that the proposed methods achieve excellent running speed and higher accuracy over many state-of-the-art approaches.

1. Introduction

Visual tracking is currently one of the most important research topics in the field of computer vision, especially for the application of video surveillance, vehicle navigation, and human computer interaction. In practical problems, analyzing video sequences by human labor force can be impractical due to the explosive growth of video volume. Although many tracking algorithms have been proposed, it remains a challenging problem due to factors such as occlusions, illumination changes, pose changes, view point variations, etc. One of the key issues to separate the foreground targets from the background is to propose suitable appearance models. A model with high dimensional features is effective because it can preserve adequate information of the target, but these features are often redundant and often limit the speed for processing. Several methods have been proposed to find the compressive features out of the high dimensional features as sparse representation. These compressive features are low-dimensional and can preserve most information of the targets. Several tracking methods based on sparse representation have been proposed. Zhang et al. [1] introduced in their compressive tracking method a non-adaptive random matrix to project high dimensional features to a low-dimensional space. The data-independent projection matrix can achieve high processing speed and low computational cost on one hand, but on the other hand, its performance can be unstable due to the random characteristic of the matrix. Mei et al. [2] proposed a method by casting tracking as a sparse approximation problem in a particle filter framework, in which the target is represented in the space spanned by target templates and trivial templates, and the sparsity is achieved by solving an L_1-regularized least squares problem. Jia et al. [3] introduced a structural local sparse appearance model which used sparse codes of local image patches with spatial layout in an object, and employed a template update strategy which combines incremental subspace learning and sparse representation. However, these methods require to discover basis functions from the unlabeled data and can be computationally expensive.

In this paper, we model the targets with sparse Haar-like features. At the beginning, high dimensional Haar-like features are extracted in order to preserve sufficient information of the target. Since these features might be redundant and may hinder the speed for tracking, we next introduce sparse coding into the tracking process for dimensionality reduction. Every dimension of the feature can be viewed as a basis function, thus we would only need to solve an L_1-regularized least squares problem to get the sparse coefficients [4]. The process is a ranking mechanism that evaluates the large set of Haar-like features, and all of the coefficients corresponding to the basis functions should vanish except for a few. With the sparse features, we construct a naive Bayesian classifier to evaluate the target candidates [5] selected from near the current target. Positive and negative features extracted from the neighborhood of the target are used to update the classifier online. This approach can be viewed as a combination of a generative tracker

☆ This paper has been recommended for acceptance by Ming-Hsuan Yang.

* Corresponding author. Tel./fax: +86 29 83395146.
E-mail address: jinjun@mail.xjtu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.imavis.2015.04.005
0262-8856/

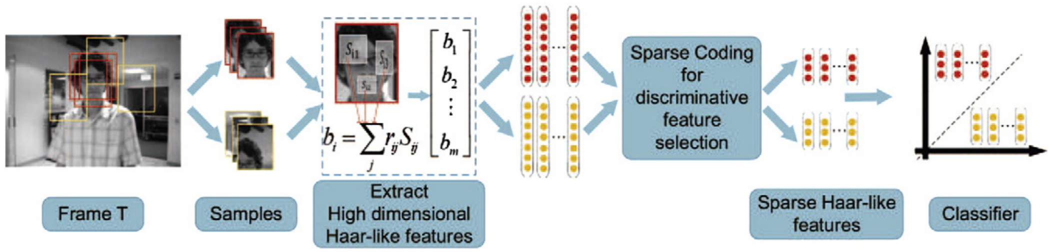

Fig. 1. Sparse feature selection.

and a discriminative tracker. Furthermore, since the appearance of the target changes through the video sequences, we also introduce an adaptive feature update scheme which compares the latest observation with previous target template, i.e., sparse coding is carried out again when target appearance changes significantly. During the tracking process, this method guarantees that the selected features are the most discriminative one. Experiments on several public datasets demonstrate that the proposed tracking method performs favorably against several state-of-the-art methods, and at the same time achieves high tracking speed.

The main contributions of this paper include:

- An online sparse feature selection method for modeling tracking target from its neighboring background,
- An automatically feature updating strategy to accommodate significant appearance changes of the target,
- More stable and accurate tracking results compared to several state-of-the-art methods, as well as real-time processing speed

The rest of the paper is organized as follows. First we review some most relevant works on target tracking in Section 2. Then we introduce the sparse feature selection process in Section 3. We elaborate the construction and updating of the naive Bayesian classifier in Section 4 and next we introduce the tracking process and the online feature selection strategy in Section 5. In Section 6, we list the evaluation results of our algorithm on 7 public dataset, and finally in Section 7, we conclude our work.

2. Related work

According to the type of the adopted appearance model, visual tracking algorithms can be categorized into generative, discriminative, or hybrid approaches. Generative trackers locate the targets using a maximum-likelihood or maximum-a-posterior formulation relying only on the target appearance model. These appearance models represent object appearance without considering its discriminative power with respect to the appearance of the background or other targets. Jepson et al. [6] introduced an appearance model that involves a mixture of stable image structure, learned over long time courses, along with 2-frame motion information and an outlier process. In [7], Matthews et al. introduced a template update method that can reduce the drifting problem by aligning with the first template to reduce drifts. Kwon et al. [8] proposed a method that decomposed the observation model and motion model into multiple basic observation models and basic motion models that are constructed by sparse principle component analysis (SPCA) of a set of templates. In [9], Ross et al. presented a tracking method that incrementally learns a low-dimensional subspace representation and adapt online to the changes in the appearance of the target.

Discriminative trackers aim to distinguish the targets from the background using a classifier that learns a decision boundary between the appearance of the target and that of the background or other targets. Avidan proposed [10] an ensemble tracking method that constantly updates a collection of weak classifiers to separate the foreground object from the background. Tang et al. [11] introduced a semi-supervised learning approach that built an online support vector machine (SVM) for each independent feature and fuses the classifiers by combining the confidence map from each classifier. Babenko et al. [12] introduced a discriminative learning paradigm called multiple instance learning (MIL) that puts all ambiguous positive and negative samples into bags to learn a discriminative model for tracking. Grabner and Bischof proposed [13] an online boosting based feature selection framework.

Hybrid trackers use a combination of the previous two approaches, in which a generative model and a discriminative classifier are combined to capture appearance changes and allow reacquisition of an object after total occlusion. Yu et al. [14] proposed a generative model using a number of low dimension linear subspaces to describe the target appearance, as well as a discriminative classifier using an online support vector machine which is trained to focus on recent appearance variations. In [15], Zhang et al. proposed a hybrid compressive tracking algorithm. The targets are represented by a multiscale convolution with rectangle filters. Then they employed non-adaptive random projections over filtered images using a very sparse measurement matrix, and then used the projected features to formulate the tracking task as a binary classification via a naive Bayesian classifier. They also introduced a coarse-to-fine target search algorithm, which reduces the computational complexity. In [16], Zhong et al. developed a sparsity-based discriminative classifier (SDC) and a sparsity-based generative model (SGM) that exploited both holistic templates and local representations. Notice that Zhong's objective function for SDC is very similar to ours. However, the entire workflows are significantly different. In [16], the SDC learns a sparse classification model while in our work, Eq. (3) is only used for feature selection while a more robust Bayesian classifier

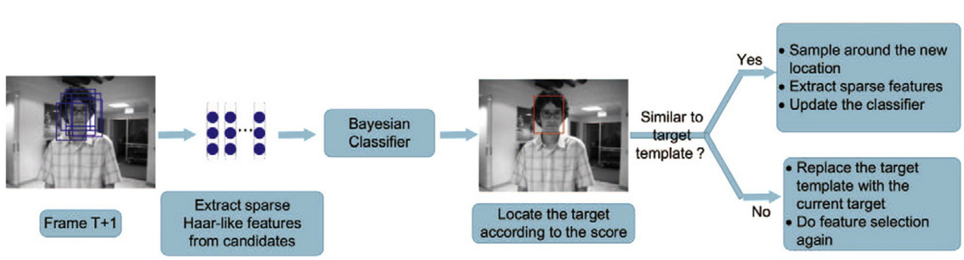

Fig. 2. Target search and feature updating.