Pattern Recognition 48 (2015) 2459–2473

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust visual tracking via efficient manifold ranking with low-dimensional compressive features

Tao Zhou [a,b], Xiangjian He [c], Kai Xie [a,b], Keren Fu [a,b], Junhao Zhang [a,b], Jie Yang [a,b,*]

[a] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240 China
[b] Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China
[c] Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 20 April 2014
Received in revised form
26 December 2014
Accepted 9 March 2015
Available online 19 March 2015

Keywords:
Visual tracking
Appearance model
Manifold ranking
Random projections
Low-dimensional compressive features
Spatial context

ABSTRACT

In this paper, a novel and robust tracking method based on efficient manifold ranking is proposed. For tracking, tracked results are taken as labeled nodes while candidate samples are taken as unlabeled nodes. The goal of tracking is to search the unlabeled sample that is the most relevant to the existing labeled nodes. Therefore, visual tracking is regarded as a ranking problem in which the relevance between an object appearance model and candidate samples is predicted by the manifold ranking algorithm. Due to the outstanding ability of the manifold ranking algorithm in discovering the underlying geometrical structure of a given image database, our tracker is more robust to overcome tracking drift. Meanwhile, we adopt non-adaptive random projections to preserve the structure of original image space, and a very sparse measurement matrix is used to efficiently extract low-dimensional compressive features for object representation. Furthermore, spatial context is used to improve the robustness to appearance variations. Experimental results on some challenging video sequences show that the proposed algorithm outperforms seven state-of-the-art methods in terms of accuracy and robustness.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Visual tracking is a long standing research topic due to its wide range of applications such as behavior analysis, activity recognition, video surveillance, and human–computer interaction [1,2]. Although it has had a significant progress in the past decades, developing an efficient and robust tracking algorithm is still a challenging problem due to numerous factors such as partial occlusion, illumination variation, pose change, abrupt motion, and background clutter. These factors can lead to wrong association, and result in drift and even failure in tracking.

The main tracking algorithms can be categorized two classes: generative [3–6] and discriminative methods [7–13].

Generative methods focus on searching for the regions which are the most similar to the tracked targets with minimal reconstruction errors of tracking. Adaptive models including the WSL tracker [3] and IVT method [14] have been proposed to handle appearance variation. Recently, sparse representation methods have been used to represent an object by a set of trivial target templates and trivial templates [6,15] to deal with partial occlusion, pose variation and so on. Therefore, it is critical to construct an effective appearance model in order to handle various challenging factors. Furthermore, generative methods discard useful information surrounding target regions that can be exploited to better separate objects from backgrounds.

Discriminative methods treat tracking as a classification problem that distinguishes the tracked targets from the surrounding backgrounds. A tracking technique called tracking by detection has been shown to have promising results in real-time. This approach trains a discriminative classifier online to separate an object from its background. Collins et al. [7] selected discriminative features online to improve the tracking performance. Boosting method has been used for object tracking through combining weak classifiers to establish a strong classifier to select discriminative features, and some online boosting feature selection methods have been proposed for object tracking [8,16]. Babenko et al. [9] proposed a novel online MIL algorithm that achieved superior results with real-time performance for object tracking. An efficient tracking algorithm based on compressive sensing theories was proposed by Zhang et al. [10]. It uses low dimensional features randomly extracted from high dimensional multi-scale image features in the foreground and background, and it achieves better tracking performance than other methods in terms of robustness and speed. Moreover, although some efficient feature extraction techniques have been proposed for visual tracking [8,10,12], there

* Corresponding author at: Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 800 Dongchuan road, Shanghai, 200240 China. Tel./fax: +86 21 3420 4033.
E-mail address: jieyang@sjtu.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.patcog.2015.03.008
0031-3203/© 2015 Elsevier Ltd. All rights reserved.

often exist a large number of samples from which features need to be extracted for classification, thereby entailing computationally expensive operations [9].

The above tracking methods have shown promising performance. However, they have some shortcomings. Firstly, although the goal of a generative method is to learn an object appearance model, an effective searching algorithm and an effective measuring method to match candidate samples to an object model are difficult to obtain. Secondly, background varies broadly during a tracking process, so it is difficult to achieve the aim of a discriminative method to distinguish a target region from a complicated background when the target looks similar to its background. Therefore, it is very difficult to construct a discriminative object representation. Thirdly, feature selection is of crucial importance for generating an effective appearance model. However, approaches using a large amount of features make the computational load very heavy. Therefore, the computational complexity of tracking methods is rather high, and this limits the real-time applications of these methods.

Graph-based ranking algorithms have been widely applied to information retrieval and have proved to have excellent performance and feasibility on a variety of data types [17–19]. The manifold ranking algorithm first constructs a weighted graph by considering each data node as a vertex. The ranking score of the query is iteratively propagated to nearby node via a weighted graph. Finally, nodes will be ranked according to the ranking scores, in which a larger score indicates higher relevance. In this paper, we develop a novel and robust tracking method based on manifold ranking, which regards tracking as a ranking problem. As shown in Fig. 1, we mark the tracked results as labeled nodes, while candidate samples are regarded as unlabeled nodes. The tracking objective is to estimate the corresponding likelihood that is determined by the relevance between the queries and all candidate samples. We use a manifold structure to measure the relevance between a model and samples, and in our method low-dimensional compressive features can efficiently compress features of foreground objects and their background. Experimental results on some challenging video sequences are demonstrated to show the effectiveness and robustness of the proposed model and algorithm in comparison with seven state-of the-art tracking methods.

The main contributions of this paper are as follows:

1. A novel visual tracking method based on graph-manifold ranking is proposed.
2. An efficient manifold ranking algorithm is adopted. It can reconstruct graph efficiently in each tracking round and reduce the computation complexity.
3. Low-dimensional compressive features of an image are extracted by a very sparse measurement matrix for object representation. This matrix preserves the structure of the image and discriminates objects from their cluttered background effectively.
4. Our method exploits both temporal and spatial context information, and it is robust to appearance variations caused by abrupt motion, occlusion and background clutters.
5. Experimental results show that the proposed algorithm outperforms seven state-of-the-art methods in terms of accuracy and robustness.

This is an extension of our paper showing preliminary results in [20]. The rest of this paper is organized as follows. The graph-manifold ranking algorithm, the efficient manifold ranking algorithm and low-dimensional compressive features are described in Section 2. Details of our proposed method based on an efficient manifold ranking with low-dimensional compressive features are demonstrated in Section 3. Experimental results are shown and analyzed in Section 4. The conclusion is presented in Section 5.

2. Preliminaries

2.1. Graph-based manifold ranking

Manifold ranking (MR), a graph-based ranking algorithm, has been widely applied in information retrieval and shown to have excellent performance and feasibility on a variety of data types [17,18]. The manifold ranking method is described as follows: given a query node, the remaining unlabeled nodes are ranked based on their relevance to the given query. The goal is to learn a ranking function to define the relevance between unlabeled nodes and this query [18,19].

In [19,21], a ranking method that exploits the intrinsic manifold structure of data for graph labelling is proposed. Given a data set $X = \{x_1, x_2, \ldots, x_n\} \in \Re^{m \times n}$, where m is the dimension of feature vector and n is the number of data in the data set, some data points are labelled queries and the rest need to be ranked according to their relevance to the queries. Let $f: X \rightarrow \Re^n$ denote a ranking function which assigns a ranking value r_i to each point x_i, and r be a column vector defined by $r = [r_1, r_2, \ldots, r_n]^T$. Let $y = [y_1, y_2, \ldots, y_n]^T$ denote an indication vector, in which $y_i = 1$ if x_i

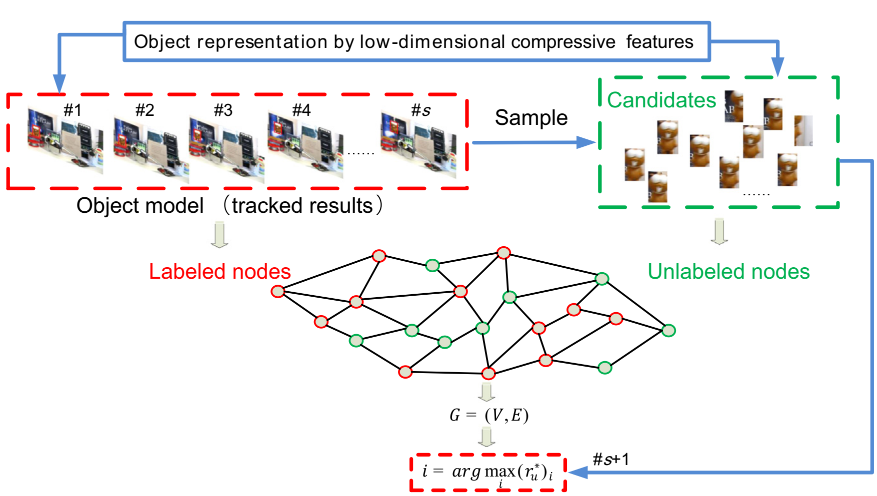

Fig. 1. Basic flow of our tracking algorithm. A graph is established combining labeled nodes (tracked results) and unlabeled nodes (candidate samples), and ranking scores represent the relevance between the object model and candidate samples.

Download English Version:

https://daneshyari.com/en/article/10361277

Download Persian Version:

https://daneshyari.com/article/10361277

Daneshyari.com