J. Vis. Commun. Image R. 25 (2014) 793–804

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Online semi-supervised compressive coding for robust visual tracking

Si Chen, Shaozi Li *, Songzhi Su, Donglin Cao, Rongrong Ji

School of Information Science and Technology, Xiamen University, Xiamen 361005, China
Fujian Key Laboratory of the Brain-like Intelligent Systems (Xiamen University), Xiamen 361005, China

ARTICLE INFO

Article history:
Received 15 July 2013
Accepted 16 January 2014
Available online 24 January 2014

Keywords:
Visual tracking
Generative tracking
Discriminative tracking
Semi-supervised compressive coding
Online learning
Compressive sensing
Semi-supervised coding
Random projection

ABSTRACT

In this paper we propose an online semi-supervised compressive coding algorithm, termed SCC, for robust visual tracking. The first contribution of this work is a novel adaptive compressive sensing based appearance model, which adopts the weighted random projection to exploit both local and discriminative information of the object. The second contribution is a semi-supervised coding technique for online sample labeling, which iteratively updates the distributions of positive and negative samples during tracking. Under such a circumstance, the pseudo-labels of unlabeled samples from the current frame are predicted according to the local smoothness regularizer and the similarity between the prior and the current model. To effectively track the object, a discriminative classifier is online updated by using the unlabeled samples with pseudo-labels in the weighted compressed domain. Experimental results demonstrate that our proposed algorithm outperforms the state-of-the-art tracking methods on challenging video sequences.

1. Introduction

As a research hotspot in computer vision, visual tracking has a wide spread of applications ranging from human–computer interaction and motion-based recognition to automated surveillance [1,2]. Given a video containing multiple frames, visual tracking essentially assigns consistent labels to the tracked objects with non-stationary appearances. Despite a large number of algorithms in the literature [3–17], tracking an object under complex environments retains challenging due to the drastic appearance changes caused by pose variations, partial or full occlusions, fast motions, illumination changes, etc.

Recent advance in visual tracking prefers an online learning formulation, which is good at dealing with the object appearance changes. Online learning based visual tracking algorithms can be coarsely divided into two categories, i.e., generative [3–7,16,17] or discriminative [8–13]. The former [3–7] incrementally updates an appearance model of the target, based on which the object region with the minimal reconstruction error is located, for example, Eigentracking [3], IVT [4] and L1 tracker [5–7]. However, most generative algorithms [3–7] only consider the holistic representation and do not make full use of the discriminative information between the target and the background, so that these algorithms are difficult to model the target correctly in uncontrolled conditions. On the other hand, the discriminative algorithms [8–13] formulate tracking as a binary classification problem and train a discriminative classifier online to separate the object from the background, such as online boosting [8,9], online SemiBoost [10] and the MIL tracker [11]. However, the existing discriminative algorithms [8–13] are mainly based on the boosting framework, which consumes much time to iteratively update the weak classifier pools for selecting the best features. In the recent work, both generative and discriminative methods are combined, e.g., compressive tracking (CT) [14] and Wang's method [15]. In [14], the CT tracker extracts the features of the object by using a fixed random measurement matrix in the compressed domain, which is, however, not effective enough to deal with dramatic object appearance changes, thus suffering from the drifting problem.

To alleviate the drifting problem towards robust visual tracking, we propose a semi-supervised compressive coding algorithm, termed SCC. The proposed SCC algorithm adopts an adaptive compressive sensing for appearance modeling by using the weighted random projection, where the Fisher discrimination criterion is used to evaluate the discriminability of each random feature (corresponding to a local patch within an image sample). Moreover, a novel semi-supervised coding technique is adopted for online sample labeling. To predict the pseudo-labels of unlabeled samples from the current frame, the SCC algorithm considers both the inconsistency among the unlabeled samples (i.e., local smoothness regularizer) and the similarity between the prior probability

* Corresponding author at: School of Information Science and Technology, Xiamen University, Xiamen 361005, China. Fax: +86 592 2580258.
E-mail address: szlig@xmu.edu.cn (S. Li).

http://dx.doi.org/10.1016/j.jvcir.2014.01.010
1047-3203/

and the current model. In addition, a discriminative classifier is learned on weighted sparse features and is updated by using the unlabeled samples with their predicted pseudo-labels. The effectiveness of the proposed algorithm is born out by experiments on several challenging video sequences.

The main contributions of our work are summarized as follows:

(1) The SCC algorithm develops an adaptive compressive sensing technique to effectively represent an image sample in a low-dimensional feature space, which makes full use of both local and discriminative information of the target object. Such enhances the robustness of tracking to handle the dramatic appearance changes.
(2) A new semi-supervised coding technique is proposed for online sample labeling, which incorporates a local smoothness regularizer into semi-supervised learning, i.e., the inconsistency among unlabeled samples from the current frame. It ensures the effectiveness of our SCC algorithm for incrementally updating the distributions of positive and negative samples during the online tracking process.
(3) Our SCC algorithm learns a linear discriminative classifier based on the adaptive compressive sensing technique to discriminate between the target and the background. Such a strategy naturally combines the generative and discriminative characteristics to improve the tracking performance.

The rest of this paper is organized as follows. Section 2 reviews the related work and the motivation of our work. In Section 3, we present the proposed SCC algorithm in details. Experimental results are given in Section 4 and this paper concludes in Section 5.

2. Related work

There is a rich literature [3–17] in visual tracking, and the current online tracking algorithms can be roughly categorized as either generative or discriminative. The generative algorithms [3–7,16,17] formulate the tracking problem as searching for the region most similar to the target object, based on template matching [3] or subspace learning [4]. In [4], a tracking algorithm was presented to incrementally learn a low-dimensional subspace representation, so as to adapt to appearance changes. Recently, sparse representation has been successfully applied to visual tracking [5–7,16,17]. In these approaches, each candidate sample is represented as a sparse linear combination of dictionary templates, which are online updated in order to maintain an up-to-date object appearance model. Mei et al. [5,6] adopted the holistic representation of the object as the appearance model and then solved the L_1 minimization problem to track the object. Afterwards, Bao et al. [7] developed a new L_1 norm related minimization model by adding the L_2 norm regularization and solved it based on the accelerated proximal gradient technique. In [16], Liu et al. proposed a tracking algorithm that employs histograms of the local sparse representation and uses mean-shift to locate the target object. In [17], Liu et al. proposed a method which selects a sparse and discriminative set of features to improve tracking efficiency. Despite the demonstrated successes, the sparse representation based trackers [5–7,16,17] are computationally expensive to perform the L_1 minimization for each frame, and their target templates lack the discriminative information to effectively separate the object from the surrounding background.

Discriminative algorithms formulate tracking as a binary classification problem so as to distinguish the object from the background. Grabner et al. [8] presented an online boosting algorithm, which passes a labeled sample to boost through N selectors, each containing M ($M > N$) weak classifiers. However, this tracker uses the classification results to update the classifier itself, which is therefore sensitive to label noises. In [10], an online semi-supervised boosting algorithm was developed to update the classifier by using unlabeled data drawn from the current frame. However, the unlabeled samples are predicted only by using the labeled samples from the first frame, and the inherent ambiguity of labeling may degrade the tracking precision. To handle this ambiguity, Babenko et al. [11] proposed an online MIL tracker, which introduces multiple instance learning to improve the flexibility of the classifier. Zhang et al. [12] presented an online weighted MIL tracker (WMIL), which integrates the sample importance into the online learning procedure to distinguish the samples in the positive bag. However, the above MIL based trackers [11,12] use the positive labels for all samples in the positive bag to update the weak classifiers during boosting, thus confusing the discriminability of the classifier.

Recently, both generative and discriminative methods are combined in the modern online trackers [14,15,18]. In [18], a sparsity-based discriminative classifier and a sparsity-based generative model were developed. In [15], an over-complete dictionary is constructed to represent local image patches of a target object, and then a classifier is learned to separate the object from the background. However, the computational cost makes the above methods unsuitable to search for target candidates by using the over-complete dictionary in a high-dimensional feature space. In [14], the appearance model is generative since the object is represented based on features extracted in the compressed domain, where the random projection matrix is computed offline and remains fixed throughout the tracking process. Then the tracking task is formulated as a discriminative classification problem via a naïve Bayes classifier, which is updated with the sets of positive and negative samples drawn from the current frame. However, visual tracking with a static random projection matrix is likely to fail in dynamic scenes due to large appearance changes. Besides, similar to MIL trackers [11,12], the classifier is updated with positive labels for all samples in the positive bag, which might introduce errors to degrade the tracking performance.

In this paper, instead of using the fixed random projection [14] or the computationally expensive sparse representation [5,6], we develop an adaptive semi-supervised compressive coding algorithm (SCC) for robust visual tracking. The proposed algorithm employs adaptive compressive sensing for feature extraction and a semi-supervised coding technique for sample labeling.

3. Semi-supervised compressive coding

In this section we present the proposed SCC algorithm in details. We begin with some preliminaries of compressive sensing in Section 3.1. In Section 3.2, we give the detailed descriptions about our proposed adaptive compressive sensing. Next, the semi-supervised coding technique for sample labeling is presented in Section 3.3. The construction of weak classifiers is introduced in Section 3.4, and the pseudo-codes of the proposed SCC algorithm are further given in the end.

3.1. Preliminaries

Recent years have witnessed a rapid development in the area of compressive sensing (CS) [19–22]. In CS, if the signal, such as a natural image or audio, is compressible, the signal can be reconstructed from a very limited number of measurements. In the field of visual tracking, random projection has been successfully applied to reconstruct the original visual appearance of the tracked target [14,23]. Random projection is a linear technique for dimensionality reduction, which is based on the Johnson–Lindenstrauss (JL) lemma [24]. This lemma states the fact that the distances between the points in a vector space have a high probability to be preserved if they are projected onto a randomly selected subspace