Neurocomputing 000 (2017) 1–12

Contents lists available at ScienceDirect

Neurocomputing

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

Compressive perceptual hashing tracking

Long Chen[1,*], Zheng Li[1], Jian-Fei Yang

School of Data and Computer Science, Sun Yat-Sen University, Guangzhou 510006, China

ARTICLE INFO

Article history:
Received 24 December 2015
Revised 27 November 2016
Accepted 1 February 2017
Available online xxx

Communicated by Yi Zhe Song

Keywords:
Tracking
Perceptual hashing
Compressive sensing

ABSTRACT

This paper proposes a novel compressive sensing based perceptual hashing algorithm for visual tracking. A tracking object is represented by dimensionality reduced feature projected from perceptual hashing feature through a sparse measurement matrix. Besides, an updating weight map is assigned for each object and the weight map is updated according to the accumulation of foreground block and the distance between the foreground block and the center of the weight map. Based on above object representation and its weight map, our tracker searches the local region with the maximum similarity in coarse-to-fine way. In addition, we introduce a visual attention knowledge that the object, namely foreground, should be always located in the center of the weight map, to handle the model drift problem. Extensive experiments demonstrate that the proposed tracking method outperforms state-of-the-art methods in challenging scenarios and our tracker is especially insensitive to the location of the initial box.

1. Introduction

Visual object tracking is one of the key, components for numerous vision-based applications, such as assistant/autonomous driving system surveillance, object identification, and security. Although a number of trackers have been proposed, building a robust and fast visual tracker that can effectively handle all changes in appearance caused by illumination change, occlusion, viewpoints variation, scale change, and motion blur, remains a challenging problem.

Generative trackers and discriminative trackers are two main types of appearance-based trackers. The generative ones [1–5] use a particular feature vector or subspace model to present a target object and search for regions with the least reconstruction error from the target object. The discriminative trackers [6–11], namely tracking-by-detection, treat tracking problems as a local search detection problem based on a binary classifier. Generally, a discriminative with a prior data set can perform better but it would involve external training cost.

Wu et al. [12] presented a comprehensive evaluation of online trackers up until 2013. Since then, several effective trackers have been proposed. Locality Sensitive Histograms Tracker (LSHT) [13] is a simple and real-time tracking framework based on the locality sensitive histogram method, which is robust to illumination change. Given the assumption that the noise of reconstruction error is Gaussian–Laplacian distributed, an efficient iteration algorithm based on maximum joint likelihood of parameters is proposed in [14], to solve the Least Soft-threshold Squares Regression problem. In [15], a discriminative tracker based on weak SVM classifiers via randomized weighting vectors is introduced. To address the model drift problem, a multi-expert framework is chosen by [16], in which, an entropy-regularized restoration scheme is utilized to correct the undesirable effects of bad model updates for the base tracker. Extended Lucas–Kanade tracking (ELK) [17] combines template matching with pixel object and background segregation, which is resistive to drift as it disregards template background pixels strategically. Nevertheless, object tracking is still a challenging problem owing to appearance changes caused by illumination change, pose variation, occlusion, and so on.

Object representation, search mechanism, and model update are three crucial parts of a tracking problem. Numerous features and models have been chosen for object representation, some of which have a certain resistance to appearance change caused by illumination, viewpoints, or motion blur. For instance, Haar-like features [6,18,19], global integral histogram [20], locality sensitive histograms [13], sparse representation [1,21,22], adaptive color attributes [23] and weighted histogram representation [24]. Unlike a strategy that uses complex appearance models for attaining robustness in object representation, we choose to represent objects using a simple binary code constructed with the hashing technique, which is efficient and robust, because the Hamming distance is used for similarity estimation with the help of our binary feature and we did not need to train a complex classifier. Recently, hashing techniques have been widely applied to solve

* Corresponding author.
E-mail addresses: chenl46@mail.sysu.edu.cn, lchen.whu@gmail.com (L. Chen).
[1] Long Chen and Zheng Li have the equal contribution.

http://dx.doi.org/10.1016/j.neucom.2017.02.004
0925-2312/

the similarity calculation problem in many vision applications, such as image retrieval [25,26] and image search [27–29], owing to its efficiency and accuracy in data organization.

Perceptual hashing works well in various applications. However, until 2015, limited research had been conducted on hashing-based tracking [30,31]. In [31], Fei et al. present an object tracking approach using the perceptual hashing algorithm. After locating the tracking objects, they scan the next frame using the sliding-window method and present each window by perceptual hashing. Then, they measure the confidence between it and the tracking object using hamming distance. The perceptual hashing feature is unstable for violently shaking objects. To overcome the bottleneck of training hash functions with a large number of samples, which is time-consuming, they proposed a two-dimensional hashing method. During tracking, hash functions are updated using an incremental learning model. The 2D hashing used in [30] is a uniform-fragments based hashing feature, which means that there is no crossover between any two fragments. Besides, for one target, there are numerous fragments with different regions and different sizes. It is obviously a high-complexity and low-efficiency operation to extract features from all the fragments in a one-by-one way. We use a sparse measurement matrix that asymptotically satisfies the restricted isometry property (RIP) in compressive sensing theory, thereby facilitating efficient projection from the image feature space to a low-dimensional compressed subspace.

Compressive sensing theory [32] tells us that sufficiently high-dimensional signals can be represented as their low-dimensional random projections, which contain enough information to reconstruct the original high-dimensional signals. As a dimensionality reduction matter, compressive sensing has been introduced recently into visual tracking to achieve real-time performance [11,33]. In [33], a sparse measurement matrix is constructed to extract the efficient features from a multiscale image feature space. The tracking track is formulated as a binary classification by a naive Bayes classifier. Compressive tracker exhibits good performance for some tracking situations in real-time. However, the performance is unstable due to the shallow Haar feature, which cannot completely represent the object well and adaptively when encountering challenging scenarios such as illumination variation, deformation and etc, and the random mapping process scheme. Considering the complementary attributes of the perceptual hashing based tracker and compressive sensing based tracker, we present a novel tracking mechanism using the compressive tracking framework with perceptual hashing appearance model. For most current advanced tracking method, the object is restricted in a rectangle to represent its position. However, we know the rectangle must contain most proportion of the tracking object as well as a little information of the background context, sometimes causing drifting problem. To tackle the common conundrum, an idea is proposed that the tracking object is most likely to be fixed in the center of the rectangular tracking region. Consequently, we propose a foreground learning method to address the drift problem, which is not considered in their method.

The main components of the proposed compressive hashing tracking algorithm are shown in Fig. 1. For tracking, we extract our simple and robust perceptual features from multi-scale patches which are extracted randomly from the positive and negative samples. However, all of the scales and the positions should be considered, which will lead to high complexity. For computational efficiency, we introduce a sparse measurement matrix under the RIP condition in compressive sensing theory [34] for efficient projection from the image feature space to a low-dimensional compressed subspace. Finally, we generate our discriminative binary code representation by locality-sensitive hashing. Above all, we construct our object representation by Compressive Perceptual Hashing (CPH, Perceptual feature+Compressive sensing+LSH). In the next frame, we sample a series of candidates surrounding the last object position via coarse-to-fine search method and extract CPH in the same way. Then, we generate the confidence map by calculating the similarity between candidate CPH and positive/negative CPH for each candidate, and select the location of the maximum confidence candidate as the new location in the next frame. Besides, it is important to focus on foreground information of actual object instead of background in the tracking box, thus we use improved Gaussian Mixture Model (GMM) [35] to extract foreground of tracking box and take patch foreground weight into consideration to make confidence evaluation more robust, which is shown in Fig. 2. The proposed compressive tracking algorithm runs at real-time and performs favorably against state-of-the-art trackers on challenging sequences in terms of efficiency, accuracy, and robustness.

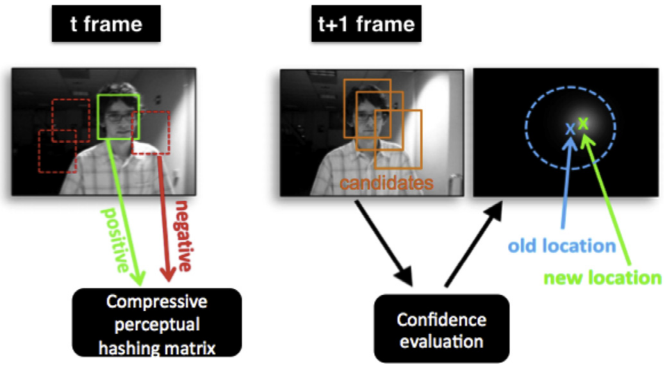

Fig. 1. Main components of the proposed compressive hashing tracking algorithm.

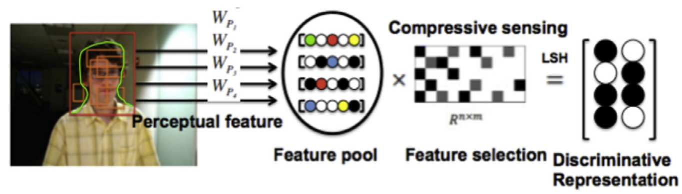

Fig. 2. Compressive perceptual hashing (CPH) representation.

This paper extends our previous work [36] by including a comprehensive description of the introduction, methodology and more thorough experiments. The contributions of the proposed tracking framework are as follows.

- We design a novel discriminative binary object representation method for visual tracking by combining perceptual hashing method with compressive sensing. The proposed perceptual hashing method avoids model training and achieves state-of-the-art performance with comparable computation speed among several discriminative methods.
- We assign each CPH template with an online learning foreground weight map to handle the model draft problem under challenging scenarios, which enhances the robustness of confidence evaluation.
- A visual attention knowledge, that the object should always lie in the center of the tracking box, is first imported into the tracking framework to further solve the model drift problem.

The rest of this paper is organized as follows. Section 2 describes three crucial components, discriminative compressive object representation, visual tracking framework, and foreground learning. Experimental results and comparisons are shown in Section 3. Section 4 concludes the paper.

Please cite this article as: L. Chen et al., Compressive perceptual hashing tracking, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.02.004