Neurocomputing 160 (2015) 191–205

Contents lists available at ScienceDirect

Neurocomputing

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

Global Coupled Learning and Local Consistencies Ensuring for sparse-based tracking

Yehui Yang [a], Yuan Xie [a], Wensheng Zhang [a,*], Wenrui Hu [a], Yuanhua Tan [b]

[a] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] Karamay Hongyou Software CO., LTD, Karamay 834000, China

ARTICLE INFO

Article history:
Received 27 June 2014
Received in revised form
10 December 2014
Accepted 20 December 2014
Communicated by D. Tao
Available online 10 March 2015

Keywords:
Visual tracking
Sparse representation
Dictionary learning
Coupled learning
Consistency ensuring

ABSTRACT

This paper presents a robust tracking algorithm by sparsely representing the object at both global and local levels. Accordingly, the algorithm is constructed by two complementary parts: Global Coupled Learning (GCL) part and Local Consistencies Ensuring (LCE) part. The global part is a discriminative model which aims to utilize the holistic features of the object via an over-complete global dictionary and classifier, and the dictionary and classifier are coupled learning to construct an adaptive GCL part. While in LCE part, we explore the object's local features by sparsely coding the object patches via a local dictionary, then both temporal and spatial consistencies of the local patches are ensured to refine the tracking results. Moreover, the GCL and LCE parts are integrated into a Bayesian framework for constructing the final tracker. Experiments on fifteen benchmark challenging sequences demonstrate that the proposed algorithm has more effectiveness and robustness than the alternative ten state-of-the-art trackers.

1. Introduction

Visual tracking is an indispensable topic in computer vision, due to its numerous applications in vehicle navigation, surveillance, and human–computer interaction [1,2]. Although efforts have been made by many researchers for constructing more effective trackers in the past years [3–6,40,44], tracking is still a challenging problem, since only few groundtruth in the first frame can be used, and the targets may undergo pose variation, occlusion, illumination changing, background cluttering, etc. All these challenges may contribute to track error and turn to drift.

To design a robust tracker which can handle the aforementioned challenges, various representation schemes are introduced into tracking task, such as pixel-based tracker [7], adopted features based trackers (e.g., texture [8], color [9], sparse-based tracker [9–23]), description models based trackers (e.g., histogram [15], subspace representation [11,40]) and multilevel quantization tracker [16]. Among the schemes listed above, sparse representation is wildly considered to be an effective tool for dealing with the aforementioned challenges.

As to sparse-based trackers, Mei and Ling [10] sparsely represent each object in a space spanned by trivial templates to tackle occlusion, and corruption challenges. Jia et al. [17] propose a tracking method based on a structural local sparse appearance model which exploits both local and spatial information of the target. Moreover, Bai et al. [13] model the object as a sparse linear combination of structured union of subspaces in a basis library. In addition, Hong et al. [18] integrate a multi-task and multi-view sparse learning problem into particle filter framework, which aims to explore underlying relationships between different particles and various types of visual features for tracking. However, all these works only address the foreground samples to construct generative models while ignoring the information from background.

To add the background information, Liu et al. [19] and Wang et al. [20] construct discriminative models based on sparse representation; however, they only encode the local patches of both object and background, while lose the holistic information provided by the object. Additionally, Xie et al. [14] utilize the sparse representation of target and background by combining both generative model and discriminative model, but they only encode the object in a holistic level without taking the local information into consideration. Meanwhile, some other algorithms also try to integrate both generative and discriminative models for tracking [24–27]; however, they are not sparse-based algorithms and they do not exploit the combination of holistic and local features.

To address the above problems, in this paper, we aim to integrate the advantages of both discriminative model and generative model to exploit the holistic and local information from the object. Thus, the

* Corresponding author. Tel.: +86 10 82614489; fax: +86 10 62545229.
E-mail addresses: yehui.yang@ia.ac.cn (Y. Yang), yuan.xie@ia.ac.cn (Y. Xie), wensheng.zhang@ia.ac.cn (W. Zhang), wenrui.hu@ia.ac.cn (W. Hu), tanyh66@petrochina.com.cn (Y. Tan).

http://dx.doi.org/10.1016/j.neucom.2014.12.060
0925-2312/

proposed algorithm is constructed by both global and local parts. In the global part, we encode holistic information of both object and background via a global dictionary, then the sparse codes are used to train a classifier to roughly distinguish the target from the backgrounds. As to the update scheme, we coupled learning the global dictionary and classifier instead of updating the dictionary and classifier as two separate parts as traditional algorithms in [20,21]. In local part, we first partition the candidates into patches, then use a local dictionary to encode the patches into sparse codes. Finally, between two consecutive frames, both temporal and spatial consistencies of the patches are ensured to refine the tracking results. Moreover, the global part and local part are two complementary parts which contain both holistic and local information of the object, and we integrate them into a Bayesian inference framework to construct the final tracker.

The contributions of this paper are as follows:

(1) We sparsely represent the object in both global and local levels via two complementary parts, and these two parts give novel aspects to utilize the object's holistic and local information for tracking.
(2) To construct an adaptive GCL part, we employ an online algorithm to coupled learning of the global dictionary and classifier.
(3) In LCE part, we propose a new method to calculate the candidates' local confidences based on the temporal and spatial consistencies among the object patches.

Similar with our work, Zhong et al. [31] propose a sparse-based collaborative model which exploits both holistic and local information of the object. But we are different from them in both the way of sparse representation and the dictionary updating algorithm. Moreover, we use two stage filtering to combine the global and local parts instead of simply multiplying the confidence values of the holistic template and local patches in [31], and more detailed differences between the two trackers will be discussed in Section 2.

The paper is organized as follows. Firstly, we briefly discuss some related work in Section 2, and the details of our proposed tracker will be presented in Section 3. Then Section 4 presents the quantitative and qualitative comparisons between the proposed algorithm and some state-of-the-art trackers. Finally, conclusions and future work are followed in Section 5.

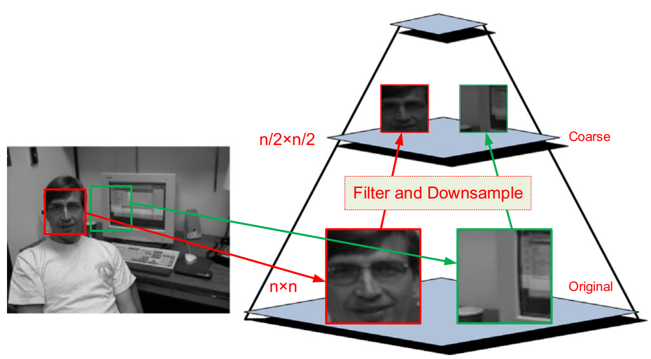

Fig. 1. Illustration of holistic features extraction via Gaussian pyramid. Both positive and negative samples are put into Gaussian pyramid, and the pyramid filters and downsamples the original samples, then returns the coarse version of the samples.

Fig. 3. Illustration of image partitioning and labeling step. In this paper, we partition the positive samples via 4×4 grid, and label the patches from 1 to 16 as illustrated above.

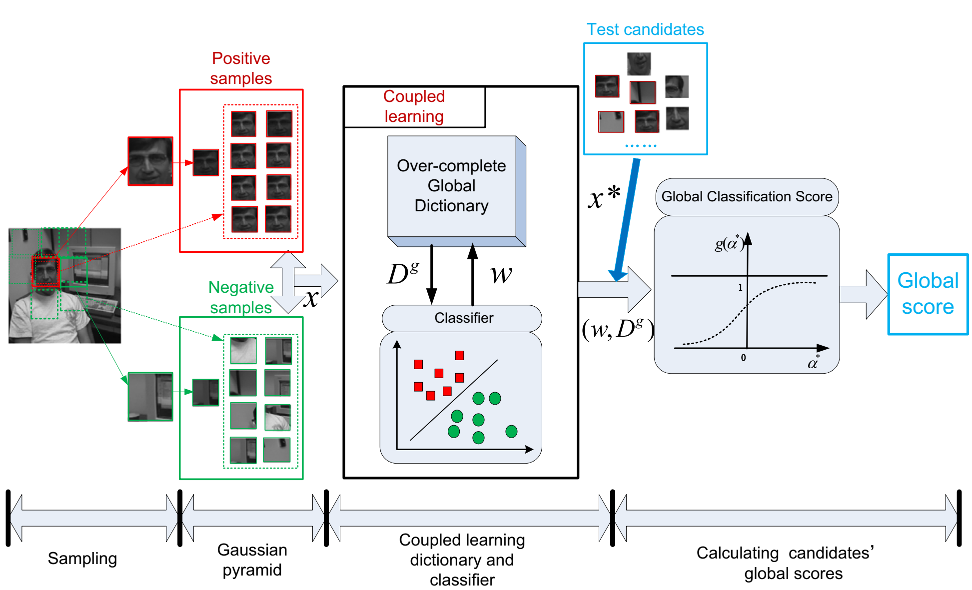

Fig. 2. Workflow of discriminative GCL part, we put both positive and negative samples into Gaussian pyramid, then use the coarse samples to coupled learn D^g and w as Algorithm 1, when global dictionary and the classifier are learnt, we can encode the test candidates, and obtain the global scores of the candidates by Eq. (5).