J. Vis. Commun. Image R. 56 (2018) 92–105

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Visual tracking via context-aware local sparse appearance model ☆

Guiji Li [a], Manman Peng [a,*], Ke Nai [a], Zhiyong Li [a], Keqin Li [a,b]

[a] College of Information Science and Engineering, Hunan University, Changsha 410082, Hunan, China
[b] Department of Computer Science, State University of New York, New Paltz, NY 12561, USA

ARTICLE INFO

Article history:
Received 11 April 2018
Revised 11 July 2018
Accepted 3 September 2018
Available online 5 September 2018

Keywords:
Visual tracking
Local sparse representation
Spatial-temporal context
Dictionary update

ABSTRACT

Most existing local sparse trackers are prone to drifting away as they do not make use of discriminative information of local patches. In this paper, we propose an effective context-aware local sparse appearance model to alleviate the drift problem caused by background clutter and occlusions. First, considering that different local patches should have different impacts on the likelihood computation, we present a novel Impact Allocation Strategy (IAS) with integration of the spatial-temporal context. Varying positive impact factors are adaptively assigned to different local patches based on their ability distinguishing the spatial context, which provides discriminative information to prevent the tracker from drifting. Furthermore, we exploit temporal context to introduce some historical information for more accurate locating. Second, we present a new patch-based dictionary update method being able to update each patch independently with the validation of effectiveness. On the one hand, we introduce sparsity concentration index to check whether the local patch to be updated is a valid local patch from the target object. On the other hand, spatial context is further employed to eliminate the effect of the background. Experimental results show the superiority and competitiveness of the proposed method on the benchmark data set compared to other state-of-the-art algorithms.

1. Introduction

As one of the most fundamental and important topics in pattern recognition and computer vision, visual tracking has received widespread concerns [1–9]. After obtaining the unique initial states of an uncertain object (maybe a person or a car), visual tracking aims at estimating the size and location of this specific object continuously in an image sequence. Although there are many breakthroughs in recent years, visual tracking still remains a quite challenging task due to the factors such as occlusion, varying illumination, background clutter, etc.

Recently, a large number of sparse representation based trackers have sprung up in visual tracking with demonstrated success [10]. The basic idea of sparse representation in tracking is to represent each target candidate by the linear combination of dictionary atoms with sparse constraint. According to the conventional representation scheme, sparse representation based appearance model can be classified into global or local pattern.

Global sparse trackers [11–16] adopt the holistic template of a target as the appearance representation. Although good performance are reported, their methods may fail with high possibility when the local appearance changes occur. Partial variations will cause an imprecise similarity measurement between candidates and the object since they are treated as single entities. Compared to global sparse appearance models, local sparse appearance models [17–22] are more attractive due to their effectiveness in handling partial occlusions. However, there still exist several drawbacks in local sparse appearance models as follows. (1) Different impacts of different local patches are not significantly considered on the likelihood computation. Due to the appearance variation difference of local patches during tracking, it is necessary to make a distinction among them rather than treating them equivalently. (2) Local sparse appearance models are less effective in dealing with background clutter. The underlying reason for this weakness lies in that discriminative information is rarely used in local sparse appearance models. (3) Most local sparse trackers update the dictionary based on the holistic pattern. That is to say, once the tracking result is updated, all local patches within it are updated naturally. When the tracking result contains some occluded patches, error will be accumulated if updated and thus result in a dirty dictionary. Otherwise, some effective appearance changes may fail to be captured, leading to degraded representation ability of the dictionary.

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: guiji.li@hnu.edu.cn (G. Li), pengmanman@hnu.edu.cn (M. Peng), naike_hnu@hnu.edu.cn (K. Nai), zhiyong.li@hnu.edu.cn (Z. Li), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.jvcir.2018.09.004
1047-3203/

To address above issues, we propose a context-aware local sparse appearance model for robust visual tracking. First, considering that different local patches should have varying degrees of impact on the likelihood computation of target candidates, we develop a novel Impact Allocation Strategy to adaptively allocate positive impact factor for each local patch using spatial-temporal context. Specifically, a local patch which is more distinguishing from the spatial context will obtain higher positive impact factor on the final likelihood function. To this end, we separately construct local object dictionary and local context dictionary to represent each patch inside the tracking result, which introduces some discriminative information to alleviate the drift problem. Furthermore, we exploit temporal context for more robust tracking. Historical information provided by the previous frame is utilized to help locate the target more accurately. To the best of our knowledge, there are few tracking methods which integrate the spatial-temporal context information into local sparse appearance model to consider appearance difference among different local patches. And extensive experiments have demonstrated the effectiveness of this integration.

Second, to ensure that all effective appearance changes can be captured from the tracking result, a new patch-based dictionary update method is proposed in this paper. All local patches are updated independently with the validation of effectiveness using sparsity concentration index [23] and spatial context. On the one hand, we exploit sparsity concentration index to check whether the local patch to be updated is a valid local patch from the target object. On the other hand, spatial context is used to provide some discriminative information to eliminate the effect of the background. In this way, even though most of local patches inside the target object undergo heavy occlusions, the tracker is still capable of capturing effective appearance changes from un-occluded local patches without missing.

The main contributions are summarized as follows.

1. An effective context-aware local sparse appearance model is proposed for robust tracking. The reliability of each local patch is measured through the whole tracking process with spatial-temporal context.
2. A novel impact allocation strategy (IAS) is developed to consider appearance difference among local patches. Different local patches are adaptively allocated varying positive impact factors on the likelihood computation of target candidates.
3. A patch-based dictionary update scheme is presented to ensure that all effective appearance changes can be updated into the dictionary without missing, even if the tracking result is mostly under occlusion.

2. Related works

There is extensive literature about various tracking methods, we advise readers to refer to [24–26] for thorough acquaintance. Here, we only talk about the most related works to ours and the tracking technology used in this study.

2.1. Part-based trackers

The part-based tracking methods become more popular depending on their robustness against partial appearance variations. Zhang et al. [27] propose a part matching tracker (PMT) for robust visual tracking. PMT matches parts from multiple adjacent frames via a locality-constrained low-rank sparse learning method. In [28], Li et al. develop a structured patch-based tracking method by simultaneously modeling the appearance of individual patches and the spatial information among them. Li et al. [29] attempt to identify and exploit the reliable patches to perform robust tracking. The final target location and scale are determined via a Hough voting-like scheme based on all reliable patches. Liu et al. [30] develop a real-time part-based tracker that learns multiple part correlation filters and combines their response maps with consideration of the reliability and temporal smoothness property of each part.

2.2. Sparse representation for tracking

Sparse representation is a popular technology that is widely used in computer vision task including face recognition [23,31] and visual tracking. In recent years, a variety of sparse representation based trackers have been developed with significant success [2,11,12,14–17,20–22,32–34].

Global sparse trackers: Mei and Ling [15] represent each target candidate as a linear combination of target templates and trivial templates with sparse constraint in a particle filter framework, and then determine the tracking result by finding the candidate with the minimal reconstruction error using target templates. This is the first time sparse representation is applied to visual tracking. Zhang et al. [14] exploit the similarities among target candidates to enforce joint sparsity on their representations for handling tracking drift. Hong et al. [35] propose a multi-task multi-view joint sparse tracker with integrating richer feature representations to deal with various tracking challenges. Wang et al. [12] present a novel appearance model which exploits both incremental subspace learning and sparse representation scheme for robust object tracking. In order to improve the discriminative power of sparse coding coefficients at a low computation cost, Hong-tu et al. [16] decompose the origin sparse coding problem into two sub sparse coding problems and compute the final sparse representation by Cartesian product.

Local sparse trackers: Zhang et al. [21] make full use of the intrinsic relationship among and inside target candidates and the spatial layout structure of the local patches for robust tracking with efficient computation. In [17], a structural local sparse appearance model is developed by performing alignment-pooling operation on the sparse codes of all local patches. Zhao et al. [36] propose a dual-scale structural local sparse appearance model, where larger-scale model is used to capture the structural quasi-holistic feature of the target and the smaller-scale model is used to capture the structural local features of the target. Qi et al. [19] propose a structure-aware local sparse appearance model to obtain more powerful discriminative ability. In their work, both global and local sparsity constraints are applied to the sparse representation of target candidates. Zhong et al. [2] propose a sparse collaborative framework, where sparse generative model pays much attention to occluded local patches in terms of similarity measurement and template update. Ma et al. [20] design a robust tracking framework using a strong classifier and structural local sparse descriptors to deal with partial occlusion and deformation. To further take the difference among local patches into account, He et al. [22] turn to treat patches differently and emphasize the contributions of key patches which are determined by their location and the states of occlusion. Tian et al. [34] utilize the color similarities between the local patches from templates and candidates as weights of them for obtaining a more reliable tracker. Nai et al. [37] propose an effective local sparse appearance model to deeply explore the appearance characteristics of different local patches. They respectively define stable patches, valid patches and invalid patches and consider their different importance on the target locating.

Low-rank trackers: In [38], Zhang et al. propose a novel low-rank sparse tracker, which formulates the tracking problem as a