Neurocomputing 184 (2016) 145–167

Contents lists available at ScienceDirect

Neurocomputing

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

Robust object tracking based on local region sparse appearance model

Guang Han *, Xingyue Wang, Jixin Liu, Ning Sun, Cailing Wang

Engineering Research Center of Wideband Wireless Communication Technique, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

ARTICLE INFO

Article history:
Received 22 January 2015
Received in revised form
25 July 2015
Accepted 26 July 2015
Available online 26 November 2015

Keywords:
Object tracking
Local region descriptors
Local sparse representation

ABSTRACT

We propose a robust object tracking algorithm based on local region sparse appearance model in this paper. In this algorithm, the object is divided into several sub-regions, and the sparse dictionaries are obtained by clustering in each sub-region. Therefore spatial structure information of the object can be captured well, and the change of object appearance can be also resisted effectively. First, the object is divided into many small patches. Then the object is divided into several sub-regions according to patch distribution again. The establishment of object dictionary base is based on combination of the dictionaries from all the sub-regions, and then space alignment between different parts of the object can be achieved. Meanwhile, noise removal and other operations in the existing sparse reconstruction error maps are performed to retain valuable information. In the updating framework, a novel flexible template set update mechanism is introduced in this paper. In this update mechanism, valuable object samples will be put into the template set. If samples are not valuable, they should not be put into the template set, even when the template set is not full. Then we use patch sparse coefficient histogram of updated templates to extract time domain information of the object in the form of weighted sum. Therefore, it can provide a reliable template basis for obtaining good candidate object. In addition, when tracking result deviates from the actual position of the object, we use a dynamic sub-region resampling method based on cosine angle to correct the position deviation timely. Therefore this method can effectively prevent the object from being completely lost. Both qualitative and quantitative evaluations on challenging video sequences demonstrate that the proposed tracking algorithm performs favorably against several state-of-the-art methods.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Video-based object tracking has become one of the hotspots of computer vision [1,2], mainly because it has a very important application value in the civilian and military [3,4]. Currently it remains a challenging issue for achieving stable and reliable tracking results in a complex situation, for example, heavy occlusion, short-time complete occlusion, illumination variation, object posture, shape and scale change, background clutter, appearance change caused by camera moving, etc. These may result in mismatch or lose object during tracking and reduce the accuracy of tracker. To solve the above problems, scholars have proposed a lot of algorithms which can be divided into two categories, one is discriminative method [5–25]; the other is generative method [26–37]. In this paper, we will not discuss these two methods which have been introduced in many literatures, more details can be found in [38,39]. Another important classification method is that it is divided into whole-based method and patch-based method according to the modeling form of the object appearance.

Whole-based method considers tracking object as a holistic template and extracts a holistic feature. This method is more suitable when the object is comparatively complete, but the performance will be not robust when the object is occluded partially. Ross et al. [34] propose an Incremental Visual Tracking (IVT) method which is a typical whole-based method by using a low dimensional Principle Component Analysis (PCA) and particle filter framework. The incremental PCA is used to train previous tracking results in IVT method, and then feature subspace is obtained. The candidate with the smallest distance is selected as the tracking result in current frame. Because incremental PCA and a forgetting factor are used in the model updating, IVT algorithm can effectively adapt the deformation of the object appearance and illumination change. However, it also has demonstrated that IVT with using the whole-based representation scheme is sensitive to partial occlusion. In [15,40], the proposed Multiple Instance Learning (MIL) tracker is also based on the holistic template. The MIL algorithm is improved based on the semi-supervised learning algorithm, and it formulates tracking problem as a classification

* Corresponding author. Tel.: +86 25 8349 2050.
E-mail address: hanguang8848@163.com (G. Han).

http://dx.doi.org/10.1016/j.neucom.2015.07.122
0925-2312/© 2015 Elsevier B.V. All rights reserved.

and detection problem. Tracking drifts will be relieved by a pre-trained discriminative online classifier. But when inaccurate positive or negative samples are used to update a classifier, it will result in the deviation and degradation of the classifier. In addition, in [41–43], the tracking algorithms based on sparse representation with the holistic template are proposed, and particle filter framework is used in these algorithms. To find the tracking object in current frame, each object candidate is linearly represented in the space spanned by object templates and trivial templates. After L1-regularized Least Mean Square problem is solved, they can obtain the sparse coefficient and reconstruction error of each candidate. Then the candidate with the smallest reconstruction error is the tracking object. This algorithm is robust to trivial noise and illumination variation, but it cannot deal with partial occlusion effectively.

Patch-based method divides the object into a number of patches, and then features are extracted in these patches. Due to the ability of capturing the local appearance information [37,39], patch-based method can effectively resist the appearance change caused by the illumination variation, deformation and partial occlusion. In [12], a large number of experiments show that the patch-based tracker is superior to the whole-based tracker. Adam et al. [26] propose a tracker based on the patch histogram feature. In this method, the object is divided into a number of patches, and then the object is located in the next frame according to a voting map, which is formed by comparing histograms of candidate patches and the corresponding templates. It is successful to handle with partial occlusions. However, the template set is not updated properly. Thus tracking will fail if the object has large deformation. In [44] object template and the candidate are divided into overlapping patches with the same size. After the sparse coefficients for each patch are solved, the sparse coefficient histograms of candidate patches and the corresponding templates will be compared. Then the tracking object is located by the candidate with the largest confidence value. Although this method is robust with partial occlusion, it may cause the misidentification of patches easily. Because its dictionaries are obtained by clustering all the patches, tracking accuracy is not robust. In addition, Jia et al. [39] propose an alignment method to extract the object local sparse feature and space information. It is also robust to partial occlusion. But it lacks particle re-sampling mechanism. Once tracking result deviates from the actual object location, the tracker may lose the object completely.

In view of the advantages and problems of patch-based sparse representation in the object tracking, this paper gives some innovative designs and improvements on the basis of the existing technology. A robust object tracking algorithm based on local region sparse appearance model is proposed. The main contributions of this paper can be summarized as follows: first, a novel local region appearance model based on sparse representation and space alignment method is presented. The object is divided into many small patches in this method. Then the object is divided into several sub-regions according to patch distribution again. The establishment of the object dictionary base is based on combination of the dictionaries from all the sub-regions. The dictionaries of each sub-region are obtained by clustering all the patches in each sub-region. Thus space alignment between different parts of the object can be achieved. False recognition of patch is prevented. This method can extract sparse features and spatial structure information of the object effectively. Second, initial sparse reconstruction error map will be corrected again. The small isolated regions with the bigger or smaller sparse reconstruction error will be deleted by using the region growing method. Then the noise caused by illumination, pose and scale change, etc. can be resist effectively. Third, a new flexible template set update mechanism is introduced in this paper, i.e. only valuable object samples will be put into the template set. If samples are not valuable, they should not be put into the template set, even when the template set is not full. Then patch sparse coefficient histogram of updated templates is used to extract the time domain information of the object in the form of weighted sum. This new update mechanism can provide a reliable template basis for obtaining good candidate object. Fourth, a novel and effective dynamic sub-region resampling method based on the change of the cosine angle is proposed. The sub-regions are dynamically divided in abnormal position (for example, object drifts or is lost) according to the change of the cosine angle. Then the particle resampling is performed in every sub-region. Thus the tracking object position can be corrected effectively. The situation of the object being completely lost can be also prevented. Finally, we evaluate the proposed algorithm in 25 challenging video sequences. The proposed algorithm shows robust tracking performance.

2. Related work and context

Sparse representation has been applied to the object tracking [34,37,39,41–44] recently and show good experimental results. The object tracking based on sparse representation has become one of the popular areas in the current research. Mei and Ling [43] introduce the sparse representation to the object tracking for the first time. In this method, each candidate object is linear represented in the space spanned by object templates and trivial template. They solve L1-regularized Least Mean Square to obtain the sparse coefficient and reconstruct error for each candidate. The tracking result is decided by largest confidence score from the reconstruct error. Although this method can track the object steadily, inadequate issues still exist: its computation cost is very high. The particle filter requires a lot of candidate particles to ensure accuracy, and L1-regularized Least Mean Square need be solved for each candidate particle. In addition, L1 method only uses holistic template. The trivial template not only is used to model the object region, but also it can model the background. Therefore when partial occlusion occurs, the reconstruction errors of image regions from the object and the background may be very small. The tracking drifting or failure occurs easily. In order to solve the above high computation complexity problem, Bao et al. [41] propose an Accelerated Proximal Gradient (APG) method to solving the L1 minimization problem. The Accelerated Proximal Gradient method is used to solve the sparse problem iteratively until the result converges. Its computation complexity is much less than Lasso method. Bai et al. [45] apply a structured sparse representation model to the visual tracking. They also claim a Block Orthogonal Matching Pursuit (BOMP) algorithm based on orthogonal matching pursuit and mutual relations of the object on the spatial structure. During the iterative procedure, this method uses the predefined patch in matching stage instead of each dimension feature. Once the best patch is matched, the corresponding coefficients are estimated. Then residual error is used in next iterative stage. The algorithm fully combines the structural characteristics of the object to reduce the amount of calculation in the sparse representation. Thus it has good robustness for the occlusion.

Zhang et al. [13] propose a multi-task sparse learning method to improve the tracking algorithm. This multi-task method can make full use of the correlation among particles to solve the sparse problem. Because there is certain relevance among the particles, and reconstruction of most similar particles also use the same template, the joint sparsity is captured based on these advantages. Therefore the overall particle sparse representation is more reasonable. Processing procedure is also accelerated. Zhang et al. [10,29] make use of the properties of compress sensing to

Download English Version:

https://daneshyari.com/en/article/405949

Download Persian Version:

https://daneshyari.com/article/405949

Daneshyari.com