Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

ELSEVIER

Contents lists available at ScienceDirect

Signal Processing

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

SIGNAL PROCESSING

Single object tracking via robust combination of particle filter and sparse representation

Q1 Shuangyan Yi [a], Zhenyu He [a,*], Xinge You [b], Yiu-Ming Cheung [c,d]

[a] School of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, China
[b] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, China
[c] Department of Computer Science, Hong Kong Baptist University, Hong Kong
[d] United International College, Beijing Normal University, Hong Kong Baptist University, Zhuhai, China

A R T I C L E I N F O

Article history:
Received 12 May 2014
Received in revised form
13 September 2014
Accepted 17 September 2014

Keywords:
Visual object tracking
Sparse representation
Occlusion prediction
Template update
Particle filter

A B S T R A C T

The drifting problem is a core problem in single object tracking and attracts many researchers' attention. Unfortunately, traditional methods cannot well solve the drifting problem. In this paper, we propose a tracking method based on the robust combination of particle filter and reverse sparse representation (RC-PFRSR) to reduce the drifting. First, we find the ill-organized coefficients. Second, we propose a diagonal matrix $\boldsymbol{\alpha}$, whose diagonal line includes each patch contribution factor, to function each patch coefficient value of one candidate obtained by sparse representation. Third, we adaptively discriminate the power of each patch within the current candidate region by an occlusion prediction scheme. Our experimental results on nine challenging video sequences show that our RC-PFRSR method is effective and outperforms six state-of-the-art methods for single object tracking. Q3

1. Introduction

Single object tracking, which plays an indispensable role in motion analysis, activity recognition, video surveillance and traffic monitoring, is a basic work in vision community and has achieved many progresses recently. While, it is still a challenging task to design a robust visual tracking method because of many negative factors existing in video sequences, such as occlusions, background variations and pose variations. Among these factors, the drifting is one core problem and has not been solved thoroughly.

Before the popularity of sparse model, the particle filter (PF), which is a classical framework in the field of single object tracking, is paid many attentions. Usually, approaches for single object tracking can be divided into two categories: discriminative models and generative models. The discriminative models [1–4] aim to discriminate the target from the background by training a classifier according to the information from both the target and the background. While, the generative models [5–8] aim to search for regions, which are extremely similar with the true object targets, based on templates or subspace. In recent years, sparse representation, which origins from compression sensor [9,10], has been successfully applied on face recognition (FR) [11] and has highly attracted researchers' interests on object tracking because of its ability on feature representation. And then, the sparse representation, which is used to reconstruct the object target by matching templates with the minimum reconstruction error, starts to play an extremely essential role in single object tracking. Generally, single object tracking approaches, based on the sparse representation, can be divided into three categories: the sparsity-based discriminative classifier (SDC), the sparsity-based generative model

* Corresponding author.
E-mail addresses: zyhe@hitsz.edu.cn (Z. He), youxg@hust.edu.cn (X. You), ymc@comp.hkbu.edu.hk (Y.-M. Cheung).

http://dx.doi.org/10.1016/j.sigpro.2014.09.020
0165-1684/

(SGM) [12–14], and the combination of SGM and SDC [15,16]. For example, Mei and Ling first applied sparse representation for tracking in [12], in which, the object target candidate (it corresponds to one particle and means the probable location of the object target in next frame) is represented as a linear combination of target templates and trivial templates, and every trivial template has only one nonzero element. Moreover, the good object target candidate is assumed to be sparsely represented by the learned template set and finally this sparse optimization problem is solved as a L1 minimization problem with no negative constraints. Mei et al. [13] proposed a bounded particle resampling (BPR)-L1 tracker based on sparse representation by combining PF framework to improve the tracker speed. Liu et al. [15] developed a robust tracker based on sparse representation by combining generative model with discriminative model. However, the above methods have a detrimental effect on the quality of tracking when the video sequences face with the partial occlusion and background clutter. With respect to this problem, Xu et al. [14] proposed a method based on adaptive structural local sparse appearance (ASLSA) and obtained a relatively optimal tracking result. Moreover, Zhong et al. [16] first proposed the SDC and then combined it with SGM to get the combination of SGM and SDC.

The tracking methods based on sparse representation [12–16] can be summarized as

$$\overline{\mathbf{b}} = \min_{\mathbf{b}} \|\mathbf{Y}-\mathbf{D}\mathbf{b}\|_2^2+\lambda\|\mathbf{b}\|_p, \tag{1}$$

where $\|\mathbf{Y}-\mathbf{D}\mathbf{b}\|_2^2$ means the error term, $\mathbf{Y}\in\mathbb{R}^d$, $\mathbf{D}$ is dictionary, $\mathbf{b}$ is the coefficient corresponding to $\mathbf{Y}$ and $\|\mathbf{b}\|_p$ means the sparse term. The tradeoff between sparse term and reconstruction error term is governed by the parameter λ. In detail, when $p=0$, Eq. (1) becomes a NP-hard problem; when $p=2$, the sparsity of the sparse term in Eq. (1) becomes weak; when $p=2,1$, Eq. (1) has the property of row sparsity; when $p=1$, Eq. (1) is actually a LASSO problem [17] which gets its popularity as a convex optimization problem and has been accepted as a most useful tool in different fields [18–24]. Furthermore, it also can be applied into object tracking by solving Eq. (2) [12]:

$$\overline{\mathbf{b}} = \min_{\mathbf{b}} \quad \|\mathbf{Y}-\mathbf{D}b\|_2^2+\lambda\|\mathbf{b}\|_1,$$
$$\text{s.t.} \quad \mathbf{b}\succcurlyeq\mathbf{0} \tag{2}$$

where $\mathbf{Y}\in\mathbb{R}^d$ (d is the dimension of the image vector) means one candidate, $\mathbf{D}\in\mathbb{R}^{d\times n}$ means the dictionary composed of tracking results from n frames, λ, which usually satisfies $\lambda>0$, controls the sparsity of the solution [25] and $\mathbf{b}\succcurlyeq\mathbf{0}$ means all the elements in $\mathbf{b}$ are nonnegative. Second, the useful coefficients, which cannot represent the object target invalidly, are abstracted from the trivial coefficients [26]. Third, target templates are replaced by PCA basis [27] in order to avoid the redundancy of templates set and can represent the object target effectively. Fourth, the patch-dictionary, which is formed by dividing each template into N overlapped local image patches [16,14], is paid more and more attentions by researchers. Moreover, the more description about patch term can refer to Section 2.

Traditional tracking methods usually use Eq. (2) to reconstruct the candidate by matching dictionary composed of n tracking templates. While, we use reverse sparse representation, which is shown in Eq. (3), to reconstruct the template set by matching dictionary composed of candidates.

To the best of our knowledge, the ASLSA method [14], compared with other state-of-the-art methods, has obtained better partial information by assembling the patches at the same positions of the patch-dictionary. Unfortunately, the ASLSA method still suffers from the drifting because it does not consider geometric information between patches and assigns the same contribution factor for every patch within one candidate region. Generally speaking, the drifting starts to appear when the right candidate is not suitably represented instead of the false negative candidate. Therefore, the drifting is very easy to occur under an occlusion case. In addition, the drifting also tends to appear if the way of dividing patches is not suitable. Hence, the core scale of dividing patches is that the patches should be more discriminative. Therefore, we adopt our own patch way for some datasets, which is described in the experimental section, to focus on the key foreground appearance information.

Therefore, in this paper, we propose a robust combination of particle filter and reverse sparse representation (RC-PFRSR) method to reduce the drifting. In summary, our key contribution in this paper is to explore how to integrate the partial information into the integral one to improve the tracking result. That is, our method exploits the spatial information of each local patch with an occlusion prediction scheme. Different with the above proposed methods, our RC-PFRSR method, the framework of which is shown in Fig. 1, first adds a contribution factor $\boldsymbol{\alpha}$ (translation matrix), which is inspired by investigating the origin coefficients with occlusion information, into the object function in [14] to reduce the drifting. In mathematics, the function of $\boldsymbol{\alpha}$ is to make the dictionary have a parallel translation, thereby represent the candidate in the best. In this way, the value of object function approximates zero and the reconstruction error reaches

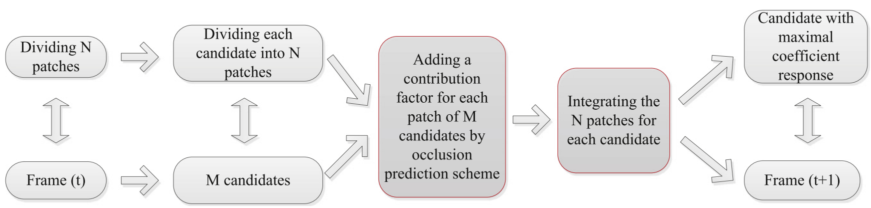

Fig. 1. The framework of our tracking method. The part boxed by a red line is our core idea. We use Frame (t) and Frame ($t+1$) to represent the tracking result in frame t and $t+1$, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Please cite this article as: S. Yi, et al., Single object tracking via robust combination of particle filter and sparse representation, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.09.020