Journal of Visual Communication and Image Representation 50 (2018) 74–82

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Improved kernelized correlation filter tracking by using spatial regularization[☆]

Long Gao[a,b], Yunsong Li[a,b,⁎], Jifeng Ning[a,c]

[a] State Key Laboratory of Integrated Service Networks, School of Telecommunications Engineering, Xidian University, No. 2, South Taibai Street, Hi-Tech Development Zone, Xi'an 710071, China

[b] Joint Laboratory of High Speed Multi-source Image Coding and Processing, School of Telecommunications Engineering, Xidian University, No. 2, South Taibai Street, Hi-Tech Development Zone, Xi'an 710071, China

[c] College of Information Engineering, Northwest A&F University, Angling, 712100, China

ARTICLE INFO

Keywords:
Visual tracking
Correlation filter
Spatial regularization
Kernel method

ABSTRACT

The correlation filter based trackers have drawn much attention due to their encouraging performance on precision, robustness and speed. In this paper, we introduce the spatial regularization component into the ridge regression model used by classical kernelized correlation filter (KCF) to improve its performance. It overcomes the fact that the traditional KCF does not consider the prior spatial constraint of the feature distribution of the target. We found that, after adding the spatial regularized function, we can solve the ridge regression formula efficiently with the property of circulant matrices. In this way, we can simultaneously keep the realtime and improve the tracking performance. Finally, we evaluate the proposed SRKCF tracker on the OTB-2013 and OTB-2015 comparing with 36 trackers and our tracker achieves state-of-art. Comparing with the SRDCF which applies the spatial regularized function, our algorithm achieves comparable performance with the obvious advantages in speed.

1. Introduction

Object tracking is an important problem in computer vision as it has many applications, for instance: human computer interaction, traffic monitoring and driverless vehicle [1,2]. The primary task of object tracking is the analysis of video sequences for the purpose of establishing the location of the target over a sequence of frames starting from the bounding box given in the first frame [1]. Although great progress has been made during the past decades, designing a robust tracking algorithm is still a challenging task due to facts such as occlusion, deformation, fast motion, illumination changes and so on.

Recently, the correlation filter based trackers achieve encouraging performance on precision, robustness and speed [3–6]. Especially, the KCF runs at hundreds of frames-per-second. These algorithms assume the training samples and candidate detection samples are periodical. They train a correlation filter from a set of samples which are cyclic shifts of a base sample, and use this trained correlation filter to find the location of the target in the detection patch. The candidate detection samples are also cyclic shifts of a base sample. This periodical assumption enables efficient training and detection by utilizing the Fast Fourier Transform [4]. However, periodical assumption brings several disadvantages. Using the cyclic shifts of a base sample as training samples lead to an unwanted problem: the training samples cannot represent the image content accurately, especially when the target shifted to the boundary area of the searching patch (see Fig. 1). These training samples reduce the discriminative power of the learned model. In the detection part, the response values are accurate only near the center of the searching patch because of using the cyclic shifts of the base sample as candidate samples. This leads to a very restricted target search region at the detection part.

To address these challenges, Danelljan et al. [4] introduce spatial regularized function into discriminatively correlation filters (DCF) and significantly improve the performance of the DCF [4]. However, the introduction of spatial regularized function leads to substantial increase in the amount of computation. SRDCF uses a series of optimization methods to improve the efficient of the algorithm. Still, SRDCF is complex and slow. As inspired by SRDCF, we introduce a spatial regularized matrix to the ridge regression formula used by the Kernelized Correlation Filter. We find that the new regression formula can be efficiently solved by applying the property of circulant matrix. This

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

⁎ Corresponding author at: State Key Laboratory of Integrated Service Networks, School of Telecommunications Engineering, Xidian University, No. 2, South Taibai Street, Hi-Tech Development Zone, Xi'an 710071, China.

E-mail address: ysli@mail.xidian.edu.cn (Y. Li).

https://doi.org/10.1016/j.jvcir.2017.11.008
Received 11 July 2017; Received in revised form 25 October 2017; Accepted 16 November 2017
Available online 20 November 2017

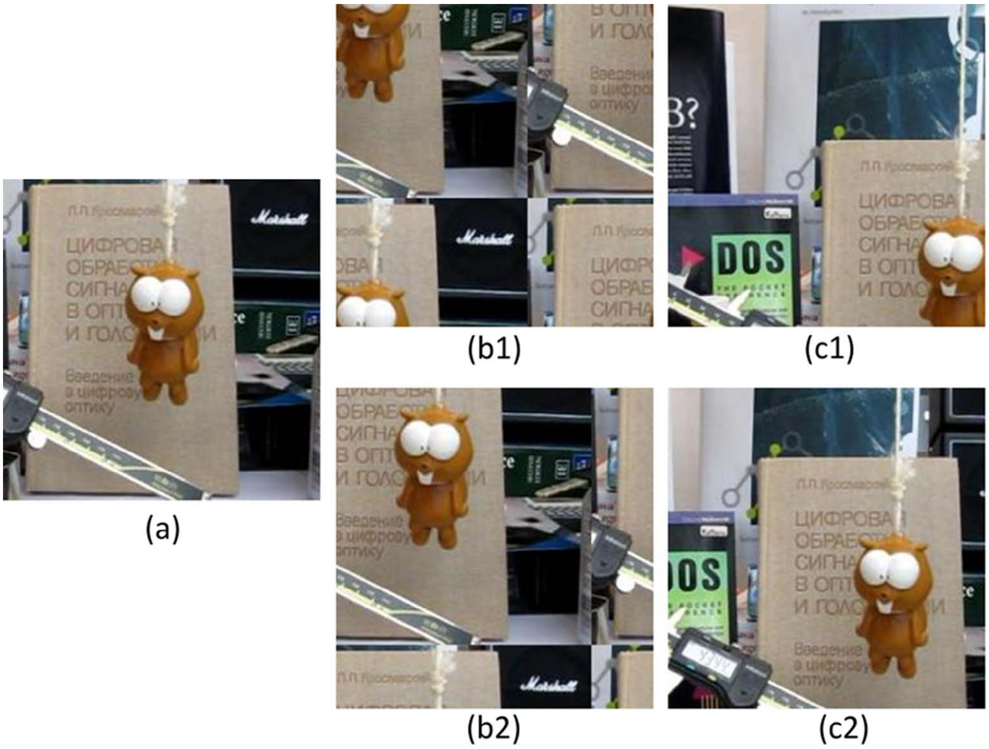

Fig. 1. Training samples obtained by cyclic shifts of one base sample cannot represent the image content properly. (a) is the base sample. We circle shift (a) by [−100 −100] and [−50 −50] separately. Then we get (b1) and (b2). (c1) and (c2) are shifts of (a) by s−100 −100] and [−50 −50]. Compare (a), (b1) and (c1), we can see that target in (b1) is separated into two parts and the training sample supposes to be (c1). Compare (a), (b2) and (c2), the background of training sample should be the background of (c2).

solution of the regression formula lead to fast training of the correlation filter. Meanwhile, the training samples and candidate detection samples can be acquired in larger area. More training samples lead to more discriminative models. This can overcome the challenges caused by target deformation. And more candidate detection samples enable to get larger area of accurate response values near the center of the searching patch. This can overcome the challenges caused by occlusion and fast target motion. Naive expand of the size of the training and detection patch will bring more background information into the feature map and reduce the performance of the tracker.

The main contribution of this paper can be summarized as follow. Firstly, we introduce the spatial regularization component into the ridge regression formula used by KCF to address the challenges caused by the cyclic shift samples. Secondly, we use the property of circulant matrices to solve the ridge regression formula and obtain a closed form of correlation filter. Thirdly, base on the correlation filter, we propose a real-time tracking algorithm whose performance achieves state-of-art. The new algorithm achieves comparable performance with SRDCF and three times the speed of SRDCF.

2. Related work

Tracking-by-detection trackers have been extensively studied [7–11]. As these trackers usually distinguish the tracking target from the background with classifiers, they are categorized as discriminative trackers. Classification algorithms such as support vector machines [8,10–13], boosting [7,14], multiple instance learning [15], random forests [16], self-similarity learning [17], and convolutional networks [18,19] have been applied in tracking-by-detection methods. Struck [8] is a classical SVM based tracker and shows appealing result in [20]. MIL [15] uses multiple instance learning and shows robustness and accuracies of the tracker. Ma et al. [18] adaptively learn correlation filters on each convolutional layer and hierarchically infer the maximum response of each layer to locate the target. For long term tracking, TLD employs positive-negative learning to choose samples and use boosting classifier [9]. The redetection part of TLD prevents target drift in the challenge scenes.

Recently, the correlation filter based tacking-by-detection trackers draw much attention because of its computational efficiency [4,5,21–23]. These algorithms use the fact that the convolution of two matrices in spatial domain equals to the element-wise product in the Fourier domain to realize fast training and detecting. Based on this theory, [3] proposes MOSSE which obtain the discriminative classifier through minimize the output sum of square error. Danelljan et al. [24] extends the MOSSE tracker with scale estimation. Galoogahi et al. [21] improves the MOSSE to handle multi-channel signals. Henriques et al. [22] uses the property of circulant matrices and kernel correlation filter to achieve hundreds of frames-per-second. And they further improve the algorithm with HOG features [5]. Facing the challenging facts such as occlusion deformation, fast motion, illumination changes and so on, many new methods are proposed to improve the performance of the correlation filter based trackers [25–28]. To overcome the drift caused by the scale variation, Li and Zhu [23] proposes effective scale adaptive scheme and multiple features which include HOG and color naming to boost the tracking performance. Part-based improvement methods are applied to address partially occlusion problems [26,28]. Zhang et al. [27] introduce an output constraint transfer (OCT) method to model the distribution of correlation response to mitigate the drifting problem of the KCF. To address the problems caused by the periodical assumption, Danelljan et al. [4] introduces a spatial regularization function in the processing of learning the discriminative classifier. This allows the size of the training and detection samples to be increased, hence better performance can be achieved.

3. The proposed tracking method

We introduced a spatial regularization matrix into the ridge regression formula for learning a classifier. And we derive the solution of this ridge regression formula in spatial domain. With the help of the property of circulant matrices and Gaussian kernel, we calculate this solution in frequency domain for better computational efficiency.