Neurocomputing 204 (2016) 77–86

Contents lists available at ScienceDirect

Neurocomputing

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

Robust and fast visual tracking via spatial kernel phase correlation filter

Lichao Zhang, Duyan Bi, Yufei Zha*, Shan Gao, Hongxun Wang, Tao Ku

Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an, Shaanxi 710038, China

ARTICLE INFO

Article history:
Received 20 February 2015
Received in revised form
25 October 2015
Accepted 26 October 2015
Available online 13 April 2016

Keywords:
Phase correlation
Spatial attributes
Circulate structure
Multiple features
Visual tracking

ABSTRACT

In this paper, we present a novel robust and fast object tracker called spatial kernel phase correlation based Tracker (SPC). Compared with classical correlation tracking which occupies all spectrums (including both phase spectrum and magnitude spectrum) in frequency domain, our SPC tracker only adopts the phase spectrum by implementing using phase correlation filter to estimate the object's translation. Thanks to circulant structure and kernel trick, we can implement dense sampling in order to train a high-quality phase correlation filter. Meanwhile, SPC learns the object's spatial context model by using new spatial response distribution, achieving superior performance. Given all these elaborate configurations, SPC is more robust to noise and cluster, and achieves more competitive performance in visual tracking. The framework of SPC can be briefly summarized as: firstly, phase correlation filter is well trained with all subwindows and is convoluted with a new image patch; then, the object's translation is calculated by maximizing spatial response; finally, to adapt to changing object, phase correlation filter is updated by reliable image patches. Tracking performance is evaluated by Peak-to-Sidelobe Ratio (PSR), aiming to resolve drifting problem by adaptive model updating. Owing to Fast Fourier Transform (FFT), the proposed tracker can track the object at about 50 frames/s. Numerical experiments demonstrate the proposed algorithm performs favorably against several state-of-the-art trackers in speed, accuracy and robustness.

1. Introduction

Computer vision aims to enable a computer to implement the basic functions of human vision such as, motion perception and scene understanding. To achieve the goal of intelligent motion perception, many efforts have been spent on visual object detection, recognition and tracking in videos. For example, in order to recognize human activities, active learning approach is introduced to video indexing to reduce human labeling cost [1]. A new approach for multi-camera group activities detection, which characterizes the intra-camera and inter-camera contexts by the structure of hidden variable, is proposed in [2].

Visual object tracking [3] is definitely also one of the most challenging problems in computer vision. It plays a crucial role in many applications, especially in human-computer interaction, surveillance and robotics. The challenge [4] is how to adapt to the object appearance variation, such as occlusion, scale change, deformation pose etc. In addition, illumination change and moving camera are also significant causes leading to tracking failure. We observed that extending the application domain of phase correlation to dense sampling can alleviate these issues.

Recently, a tracker based on the Minimum Output Sum of Squared Error (MOSSE) filter [5] is prevalently used for its fast speed. The circulant matrix is exploited by circulant structure tracker (CSK) [6] to obtain the large overlapping samples. In the latest work, the kernelized correlation filters [7] is proposed to process multi-channel features. The spatio-temporal context [8] by Gaussian weights is proposed for fast object tracking. Adaptive Color Attributes (ACA) [9] tracker extends the CSK tracker to multi-dimensional color attributes, by which the object model is represented more saturatedly. What is more, the scale estimation approaches in VOT [10,11] are also based on the CSK and MOSSE framework and the main reasons for their excellent performance in "VOT2014" are the usage of multi-feature and scale space (or multi-scale).

In fact, all above methods belong to spatial template matching method, which is accelerated by FFT. The aim of tracking is to find an object's translation in spatial domain which is precisely encoded by phase difference in frequency domain. It is well known that phase correlation is to estimate an object's translation only by using phase information and achieves high accuracy and robust performance. Meanwhile, phase correlation has been proved to be

* Correspondence to: Building No. 1, Ba Ling road Ba Qiao District, Xi'an, Shaanxi 710038, China. Tel.: +86 18049634540; fax: +86 02984787724.
E-mail address: zhayufei@126.com (Y. Zha).

http://dx.doi.org/10.1016/j.neucom.2015.10.131
0925-2312/

an excellent method in image registration [12,13]. The purpose of this paper is to research the essential law that rules the spatial translation of an object in visual tracking. Fortunately, phase correlation filter provides this powerful tool to resolve this core issue!

By using the approach of circulant date structure and kernel trick, the circulant kernel phase correlation is proposed to locate the object's new position by implementing dense sampling. We demonstrate that it is also suitable to process multi-channel circulant features. Given that conventional dirac delta function is sensitive to noise and cluster, thus always leading to failure, we adopt a new spatial response distribution to learn the object's spatial context model to resolve this problem.

To conclude, a novel spatial kernel phase correlation filter based tracker (SPC) is proposed to estimate object's translation in our work. In contrast with the classic correlation filter based tracker, SPC is intrinsic and simplex to reflect the translation of the object. We conduct extensive experiments to compare the previous studies of the correlation filter based trackers [5–10] with our SPC tracker, and the latter achieves a very appealing performance both in robustness and speed. The pipeline of our SPC tracker is vividly shown as Fig. 1 and the main contributions of this paper can be summarized as follows:

Firstly, phase correlation filter is firstly proposed to be trained by using all subwindows of an image, namely using dense sampling. We demonstrate that circulant structure is likewise suitable to phase correlation. While in contrast with all of the correlation tracking algorithms which exploit full spectrum information including magnitude spectrum and phase spectrum simultaneously in frequency domain, we intrinsically only uses the phase information to estimate the translation of an object.

Secondly, kernel phase correlation based tracker is firstly proposed in this paper. Thanks to the circulant attributes of kernel matrix, we can process all patches simultaneously, achieving surprising performance with the usage of kernelized features. We adopt a multi-feature integration scheme which employs the raw pixel, histogram of gradient and color attributes [14] to further enhance the proposed tracker to deal with more challenging scenarios.

Thirdly, our work focuses on the spatial attributes of phase correlation tracker. As the spatial shift of an object is fully dependent on phase spectrum, our approach re-formulates the spatial context model in phase correlation. The spatial response distribution models the statistical phase correlation between the low-level features (i.e., image intensity and position) of the object and those of surrounding regions.

2. Phase correspondence matching

In this section, we briefly introduce a *Phase Correlation function* [12,13]. Denote x and x' as the image signals, the response of phase correlation is

$$\mathbf{R} = \frac{\hat{\mathbf{x}} \odot \hat{\mathbf{x}}'^{*}}{\|\hat{\mathbf{x}} \odot \hat{\mathbf{x}}'^{*}\|} = e^{j(\theta_{\hat{\mathbf{x}}} - \theta_{\hat{\mathbf{x}}'})}, \tag{1}$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ are the 2D DFTs of $\mathbf{x}$ and $\mathbf{x}'$ respectively, $\hat{\mathbf{x}}'^{*}$ denotes the complex conjugate of $\hat{\mathbf{x}}'$, θ denotes the phase spectrum and $\odot$ is the dot production. The 2D phase correlation function $\mathbf{r}$ between $\mathbf{x}$ and $\mathbf{x}'$ is given as the 2D inverse DFT (2D IDFT) of $\mathbf{R}$. When two images are similar, their phase correlation function gives a distinct sharp peak. When two images are not similar, the peak drops significantly. The height of the peak gives a good similarity measure for image matching, and the location of the peak shows the translational displacement between two images.

The important techniques for improving the accuracy of 2D image correspondence matching (i) function fitting for high-accuracy estimation of peak position, (ii) window filtering for reducing boundary effects, (iii) spectral weighting for reducing aliasing and noise effects and (iv) averaging 2D phase correlation functions to improve peak-to-noise ratio.

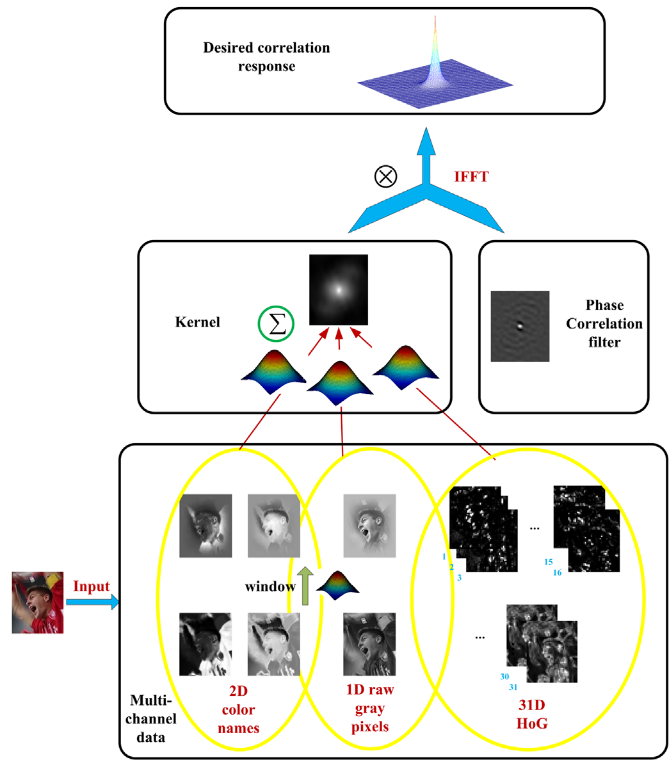

Fig. 1. Tracking framework of SPC. Firstly, low-level multiple channels features are integrated into high-level single channel features by linear regression and kernel trick; then, kernelized features are convoluted with phase correlation filter; finally, spatial response is achieved by IFFT.

3. Spatial kernel phase correlation filter

3.1. Phase correlation filter

The data matrix $\mathbf{X}$, which is called circulant matrix [6,7], is purely generated by the cyclic shifts of $\mathbf{x}$. It has an intriguing property [16] that all the circulant matrices can be expressed as below

$$X = F^{H}\mathrm{diag}(\hat{x})\mathrm{F}, \tag{2}$$

where $\mathbf{F}$ is a constant matrix that does not depend on $\mathbf{x}$, and $\hat{\mathbf{x}}$ denotes the DFT of the generating vector, $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{x})$. From now on, we will always use a hat ˆ as shorthand for the DFT of a vector. The key of KCF tracker is that the augmentation of negative samples are employed to enhance the discriminative ability of the track-by-detector scheme, while exploring the structure of the circulant matrix for its high efficiency.

The goal of training is to find a function $f(\mathbf{z}) = \mathbf{h}^{T}\mathbf{z}$ that minimizes the squared error over samples $\mathbf{x}_i$ and their regression targets y_i [5]

$$\min_{\mathrm{h}} \sum_{i}^{n} (f(\mathrm{x}_i) - y_i)^2 + \lambda \|\mathbf{h}\|. \tag{3}$$

The ridge regression has the close-form solution, $\mathbf{h} = (\mathbf{X}^{T}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{T}\mathbf{y}$. Substituted by Eq. (2), the correlation filter in Fourier domain can be calculated $\hat{\mathbf{h}}^{*} = \frac{\hat{\mathbf{x}}^{*} \odot \hat{\mathbf{y}}}{\hat{\mathbf{x}}^{*}} \odot \hat{\mathbf{x}} + \lambda$, where $\hat{\mathbf{x}}^{*}$ denotes the complex-conjugate of $\hat{\mathbf{x}}$. In Comparison with other prevalent solutions, this method saves the computational cost