ARTICLE IN PRESS

ISPRS Journal of Photogrammetry and Remote Sensing xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Visual object tracking by correlation filters and online learning

Xin Zhang [a,b], Gui-Song Xia [b], Qikai Lu [a,*], Weiming Shen [b], Liangpei Zhang [b]

[a] Electronic Information School, Wuhan University, Wuhan 430079, PR China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, PR China

ARTICLE INFO

Article history:
Received 28 February 2017
Received in revised form 20 July 2017
Accepted 20 July 2017
Available online xxxx

Keywords:
Object tracking
Correlation filters
Feature integration
Online learning

ABSTRACT

Due to the complexity of background scenarios and the variation of target appearance, it is difficult to achieve high accuracy and fast speed for object tracking. Currently, correlation filters based trackers (CFTs) show promising performance in object tracking. The CFTs estimate the target's position by correlation filters with different kinds of features. However, most of CFTs can hardly re-detect the target in the case of long-term tracking drifts. In this paper, a feature integration object tracker named correlation filters and online learning (CFOL) is proposed. CFOL estimates the target's position and its corresponding correlation score using the same discriminative correlation filter with multi-features. To reduce tracking drifts, a new sampling and updating strategy for online learning is proposed. Experiments conducted on 51 image sequences demonstrate that the proposed algorithm is superior to the state-of-the-art approaches.

1. Introduction

Visual object tracking is one of the hottest areas in computer vision, which can be regarded as the process of estimating the state (e.g., position and scale) of a moving object in videos or image sequences. It has been applied in motion-based recognition, unmanned aerial vehicle (UAV) surveillance, human-computer interaction, autonomous driving, traffic monitoring, and other photogrammetry and computer vision fields (Yilmaz et al., 2006; Wu et al., 2013; Klinger et al., 2017; Jokinen, 2013; Tian et al., 2016; Zhou et al., 2016). In the tracking scenario, the target which is located by a bounding box in the first frame should be tracked in subsequent image frames (Zhong et al., 2014; Ma et al., 2015b; Ma et al., 2016). Numerous algorithms have been proposed to address the tracking task over the past years. These algorithms can be divided into generative and discriminative methods. The generative methods, such as templates matching (Zhong et al., 2000; Jurie and Dhome, 2002; Kaneko et al., 2003), sparse representations (Xie et al., 2014; Bai and Li, 2012; Hu et al., 2015; Han et al., 2011) and subspace representations (Ross et al., 2008; Ma et al., 2015c; Li et al., 2008), aim to search for the patch that is most similar to the target. Different to the generative methods, discriminative methods (Smeulders et al., 2014; Godec et al., 2013; Babenko et al., 2011) regard the tracking task as a classification problem. And lots of machine learning methods have been used, such as supervised learning, unsupervised learning. Some researchers develop discriminative methods by modeling the target appearance with online semi-supervised boosting (Grabner et al., 2008), multiple instance learning (Babenko et al., 2011) and online discriminative learning (Jang et al., 2011). However, visual object tracking is still a challenging problem due to the target appearance changes caused by occlusion, deformation, scale or illumination variation, etc.

The prior knowledge for object tracking is the initial bounding box of the target in the first frame. A target model is trained with this prior knowledge and updated in the subsequent image frames. Due to the changes of target appearance, how to efficiently update the target model is a challenging problem. In the traditional discriminative methods, such as Boosting (Grabner et al., 2008; Grabner et al., 2006), Support Vector Machines (SVM) (Malisiewicz et al., 2011; Avidan, 2004), Multiple Instance Learning (MIL) (Babenko et al., 2011; Babenko et al., 2009), positive and negative patches are sampled from the image frames to train a discriminative classifier for object tracking. Obviously, sampling of positive and negative patches plays an important role in training the classifier.

Recently, correlation filters, which were widely used in recognition (Savvides et al., 2002; Savvides and Kumar, 2003; Savvides et al., 2004) and detection (Casasent et al., 1994; Mahalanobis et al., 2004; Bolme et al., 2009), have attracted great attention in object tracking owing to their high speed and promising precision.

* Corresponding author.
E-mail address: qikai_lu@whu.edu.cn (Q. Lu).

http://dx.doi.org/10.1016/j.isprsjprs.2017.07.009
0924-2716/

Please cite this article in press as: Zhang, X., et al. Visual object tracking by correlation filters and online learning. ISPRS J. Photogram. Remote Sensing (2017), http://dx.doi.org/10.1016/j.isprsjprs.2017.07.009

In the Fourier domain, correlation score is computed by the element-wise multiplication between image features and the complex conjugate of correlation filter (Bolme et al., 2010). Inverse fast Fourier transform (IFFT) is utilized to transform correlation back into the spatial domain. The sharp peak of the correlation score indicates the center of target (Henriques et al., 2012; Danelljan et al., 2014b). In recent researches, trackers based on correlation filters alone (Liu et al., 2016; Danelljan et al., 2014a; Henriques et al., 2015; Ma et al., 2015a) or in combination with other methods (Ma et al., 2015b; Bertinetto et al., 2016; Liu et al., 2015; Hong et al., 2015) show impressive performance in object tracking, comparing with the traditional tracking methods (Grabner et al., 2008; Babenko et al., 2011; Malisiewicz et al., 2011; Avidan, 2004; Babenko et al., 2009). Bolme et al. (2010) propose a tracker based on correlation filters by minimizing the output sum of squared error, and this tracker shows a distinct advantage with high speed. Following this idea, many researches have been developed, such as the circulant structure with kernels (CSK) tracker (Henriques et al., 2012), color names (CN) tracker (Danelljan et al., 2014b), discriminative scale space tracker (DSST) (Danelljan et al., 2014a), etc. The CSK tracker trains a kernelized correlation filter by kernel regularized least squares with dense sampling. Nevertheless, this tracker is not robust in dealing with many challenging scenarios (e.g.,occlusion, deformation, etc.) since the correlation filter is trained by intensity only. In a comprehensive study, a tracker based on kernelized correlation filters (KCF) (Henriques et al., 2015) trained by HOG features shows higher precision than CSK tracker, which indicates that robust features are more suitable to track targets. Subsequently, lots of CFTs using robust features have been proposed to address the tracking tasks (Danelljan et al., 2014b; Henriques et al., 2015; Ma et al., 2015a). Danelljan et al. (2014b) propose the CN tracker, where a kernelized correlation filter is trained by intensity and color attributes (De Weijer et al., 2009). Danelljan et al. (2014a) estimate the target's position by training correlation filters named translation filters with intensity and HOG features extracted from the window centered around the target.

However, due to the limited prior knowledge and the variation of target appearance, many CFTs fail to track targets preciously when facing challenging scenarios (Henriques et al., 2012; Danelljan et al., 2014b; Bolme et al., 2010; Danelljan et al., 2014a; Henriques et al., 2015). Therefore, some recent trackers estimate the state of a target by combining correlation filters and other methods (Ma et al., 2015b; Bertinetto et al., 2016; Liu et al., 2015; Hong et al., 2015). Ma et al. (2015b) use an online random fern classifier to re-detect the targets when the maximum value of the correlation score (response map) is lower than a threshold. Kiani Galoogahi et al. (2013) propose the MUlti-Store tracker (MUSTer) by incorporating a short-term processing as well as long-term processing into the tracking method. The short-term processing uses kernelized correlation filters for position estimation and a one-dimensional correlation filter for scale estimation. The long-term processing is used to model target appearance with local scale-invariant features and correct the target's position when a mistake position is estimated by correlation filters. In this context, we develop a method to reduce target drifts and improve the tracking performance.

In this paper, an object tracking method named correlation filters and online learning (CFOL) is proposed by combining correlation filters and online learning. The proposed CFOL can be decomposed into position estimation and scale estimation. For position estimation, we give a discriminative correlation filter trained by multi-features including intensity, HOG and color attributes. Specifically, the combination of effective features benefits the description of the target appearance (Zhou et al., 2012; Xia et al., 2010). Meanwhile, in order to reduce the tracking drifts, we utilize an online learning method which is based on logistic regression classifier. In CFOL, the initial positive and negative patches are collected from the first frame. Furthermore, HOG and intensity of these patches are extracted to train a logistic regression classifier. The classifier will be activated when the maximum value of the correlation score is less than a threshold. Due to the changes of target appearance (e.g., deformation, scale variation, etc.) and the complexity of background scenarios (e.g., illumination variation, background clutters, etc.), the classifier that was trained before may cannot accurately detect the candidate target in the current frame. Before estimating the position of targets, this classifier should be updated with a new training set, where the positive samples come from the target patches tracked in the previous frames. Then, the logistic regression classifier is utilized to detect the target from a set of candidate patches. For the scale estimation, we follow Danelljan et al. (2014a) and estimate the scale of the target by scale filters. The main contribution of this paper can be summarized as follows:

- We propose a tracking method based on correlation filters and online learning to estimate the position and scale of target. To estimate the target's position, a two-dimensional discriminative correlation filter is trained by using intensity, HOG and color attributes. We develop an online learning method to re-detect the target in the case of tracking drifts. To estimate the target's scale, a one-dimensional discriminative correlation filter is constructed by multi-scale patches.
- We present a new sampling and updating strategy for online learning. In the process of online learning, we train a logistic regression classifier by positive and negative samples, where these samples are selected by a dense sampling strategy. To ensure the reliability of tracking performance, we update the classifier before re-detecting the candidate position of target. Then, if the probability of candidate position is less than a threshold, the classifier will be updated again to refine the accuracy of classifier.
- We conduct a comparison between our proposed algorithm and the state-of-the-art trackers on a publicly available benchmark dataset (Wu et al., 2013). Furthermore, in the analysis and discussion, we carry out a comparison to analyze the factors that impact the performance of trackers. The experimental results illustrate that robust features and online learning play important role in reducing tracking drifts.

The remainder of this paper is organized as follows. In Section 2, we describe the proposed CFOL for in detail. Experimental results and discussion are reported in Section 3. Finally, we draw conclusions for this paper in Section 4.

2. Methodology

In this section, a new efficient method for estimating the position and scale of a single object is shown in Fig. 1. This section introduces the methodology of position estimation (Section 2.1) and scale estimation (Section 2.2). Furthermore, correlation filters based tracking and online learning are described in Sections 2.1.1 and 2.1.2, respectively.

2.1. Position estimation

The methods based on correlation filters achieve promising performances in object tracking (Henriques et al., 2012; Danelljan et al., 2014a; Henriques et al., 2015). In this context, we develop a correlation filter based method using multi-features, which can estimate the target's position more accurately.

Please cite this article in press as: Zhang, X., et al. Visual object tracking by correlation filters and online learning. ISPRS J. Photogram. Remote Sensing (2017), http://dx.doi.org/10.1016/j.isprsjprs.2017.07.009