Pattern Recognition Letters 000 (2017) 1–8

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Online structural learning with dense samples and a weighting kernel

Xianguo Yu*, Qifeng Yu

College of Aerospace Science and Engineering, National University of Defense Technology. Changsha, 410073, China

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Visual tracking
Structural learning
Dense sampling
Weighting kernel

ABSTRACT

A great deal of visual tracking algorithms, especially the tracking-by-detection methods, have been reported in recent years. Among them the structural learning based have shown great performance. One major problem with online learning a classifier is the limited number of training samples. Meanwhile, the success of correlation filters reveals the importance of allowing dense sampling in the training process. In this paper we propose to boost the robustness and the efficiency of an online learned structural support vector machine (SVM). Specifically, we find training with dense samples could be very efficient by applying the Fourier techniques and careful implementations. Furthermore, we propose to use a weighting kernel to improve tracking performance and the performance gain does not come with a sacrifice in the efficiency. Actually, the weighting kernel is as efficient as the linear kernel. Finally, we show favorable results on the latest VOT challenge sequences. An extended experiment incorporating a 34 dimensional HOG feature representation into our method results in top3 performance on the VOT-TIR2016 dataset.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Visual object tracking, especially the online single object tracking problem, has been attracting numerous researchers because of its many practical applications such as autonomous driving, visual surveillance and human computer interaction. Visual tracking is also an important preliminary step of higher-level vision tasks such as behavior analysis and scene parsing.

Visual tracking methods can be broadly categorized as either generative or discriminative. Generative methods model the tracking problem as searching for the region with maximum likelihood with the target [9], while discriminative approaches aim at distinguishing the target from the background by a classifier [25] or a regressor [13]. Various tracking algorithms have been developed thus far. However, some major difficulties are still challenging this field, e.g. illumination variations, viewpoint changes, background clutters, target deformations and occlusions. As discriminative methods are more resistant to the drifting problem, they are gaining more popularity than the generative ones.

Traditional discriminative models train a classifier with online generated samples. Samples are labeled with a $+1$ if they overlap the target by more than a certain threshold, otherwise they are labeled with a -1. Then a binary classifier is trained and updated to reflect the difference between the target and the background over time. Early researches focused on trying different binary classifiers and equipping them with an online updating strategy [1]. Although a better classifier could improve tracking performance, the label noise problem which arises from designating samples to a wrong category is generally overlooked.

To relieve the labeling problem, some new learning approaches are introduced to the tracking field, e.g. the semi-supervised learning [12] and the multiple instance learning [4]. While these algorithms partly alleviate the situation, the problem is still unsolved until the Struck algorithm [26] and the correlation filters [8,13]. Struck leverages a structured output prediction classifier that couples the tracking objective, namely estimating the target position, with the training objective function. Thus it naturally bypasses the labeling process. A correlation filter is actually a regressor trained and applied in the Fourier domain. As the regression target are continuous real values, samples are no longer simply assigned as either positive or negative.

Another problem with conventional discriminative models is how to generate samples for training. Most existing methods feed the classifiers by samples randomly extracted from local image regions [14]. However, to meet the time requirement of online tracking, the number of training samples are often limited to a very low level, which deteriorates the classifier performance. Correlation filter based trackers overcome this shortage by learning from dense samples derived from a base sample with different cyclic shifts. As illustrated by Danelljan et al. [18], the underlying periodic assumptions of correlation filters not only brings about the boundary effect problem, but also leads to the samples to fail to capture the true image content.

* Corresponding author.
E-mail addresses: yuxianguo_chn@163.com, yuxianguo2011@gmail.com (X. Yu).

http://dx.doi.org/10.1016/j.patrec.2017.05.017
0167-8655/© 2017 Elsevier B.V. All rights reserved.

Please cite this article as: X. Yu, Q. Yu, Online structural learning with dense samples and a weighting kernel, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.05.017

This work extends the Struck algorithm with the ability to efficiently learn from dense samples. As with Struck, we resolve the online structural learning problem by LaRank [3] and SMO [15]. SMO breaks a large quadratic programming (QP) problem into a series of small QP problems, each of which optimizes a pair of dual coefficients analytically. The computation time of SMO is dominated by the classifier evaluation, thus it is fast for training linear support vector machines. LaRank adopts SMO as the basic component and combines partial gradient information with a perceptron like stochastic training methodology to achieve both fast convergence rate and low computation time. The major problem with directly adopt dense sampling in Struck is the unaffordable computational burden of the training process even with a linear kernel. We propose to overcome this by using the Fast Fourier Transform (FFT) and an efficient caching mechanism.

We also propose to use a weighting kernel that gives more importance to the center of the sample images. Researches on different vision tasks [17] have exposed the necessity to give different weights to different image locations, e.g. the building of the SIFT descriptor [7] involves assigning each sample point a Gaussian weight relative to its distance to the descriptor center. In this paper we will show that applying a weighting kernel dramatically improves the robustness of the classifier while can be as efficient as using the linear kernel.

We build simple trackers with the proposed learning algorithm to perform tracking by detection. Our experimental evaluation demonstrates that the proposed method achieves competitive results on the VOT2016 dataset while being very efficient. Contrast experiments with different parameter settings also reveal a significant performance gain by taking advantage of the weighting kernel. Finally, the proposed tracker equipped with a 34 dimensional HOG feature representation wins third place on the VOT-TIR2016 dataset according to the VOT evaluation kit.

2. Related work

In this section we briefly review tracking by detection methods, especially those based on the SVMs. Advanced learning methods such as online structural learning, correlation filters and convolutional neural networks based tracking algorithms are also introduced. We then present our motivations of this paper.

Tracking by detection methods have long been prevalent in the visual tracking field because they can take advantage of the huge progress in object detection researches. Among them, large margin classifiers based algorithms are most popular because they generalize well when there are only a few training samples. Saffari et al. [2] propose the online multi-class LPBoost algorithm which tries to maximize the multi-class soft margin of the samples online. They report state-of-the-art results on both detection and tracking tasks. The support vector machines and the variants are very popular in tracking. In [27] the ranking SVM is used to robustly locate the object. Ranking SVMs learn relative relations between different samples. The target is supposed to be ranked higher than other samples around it. Supancic and Ramanan [14] apply the formalism of self-paced curriculum learning to select better examples to train a better linear SVM classifier online.

Structural learning is introduced to object tracking as it predicts structured output rather than a simple class label, thus being more appropriate for tracking. Li et al. [16] propose an online metric-weighted linear representation model by online structural metric learning. They learn a Mahalanobis metric matrix and incorporate it into a linear representation of appearance. A significant improvement in the discriminative power of the tracker is reported. Thanks to the efficiency and the fast convergence rate of the LaRank optimization method [3], the structural SVM [26] has been adapted for online tracking. As structural learning for object tracking no longer depends on a labeler, the potential labeling noise problem often encountered in binary classifiers is avoided. The deformable part based models [22] have achieved great success in object detection. It learns a structural SVM classifier with latent variables. Yao et al. [25] propose to solve this learning problem online to tackle with the tracking problem and obtain state-of-the-art tracking performance.

Correlation filters: Speed is always an important factor for online tracking. Traditional tracking by detection methods struggle with increased computations brought by complex learners, more training samples and larger search region. The correlation filters [8] are famous for their capability to efficiently learn with dense samples. Henriques et al. [13] enable the correlation filters to incorporate rich features thus obtaining top-performing tracking results. The underlying circulant property makes ridge regression with dense samples extremely efficient by using the discrete Fourier transform. However, the periodic assumption of correlation filters introduces further problems such as the boundary effects. To enlarge the search range and refrain the boundary effects, Danelljan et al. [18] propose to use a spatial regularization matrix to penalize correlation filter coefficients near the boundaries. While they have boosted the tracking performance to a new state-of-the-art, the simplicity and efficiency of correlations filters are weakened.

Deep learning: One major problem with online tracking is the lack of enough training samples. Therefore, some researchers have averted their eyes to learn a general tracking model offline. Nam and Han [19] pretrains a convolutional neural network (CNN) using a large set of videos with tracking ground-truths to get a general object representation. During tracking, the output of the shared layers in the pretrained network is combined with a binary classifier. Ma et al. [6] claims that different convolutional layers in a deep network encode different levels of target appearance information. They learn separate correlation filters on each layer and locate targets by integrating all correlation response.

Our first motivation: Structural learning has proved to be highly effective for object tracking. However, it is still difficult to train with increased number of samples efficiently. With the development of latest researches, the problem related to the number of training samples has been shown very important. Simpler learning algorithms such as a ridge regressor can also make up extremely successful trackers when dense sampling is possible [13]. However, the structural learning based Struck tracker is relatively ranked behind according to recent VOT competitions. This work will attempt to fill this gap, that is training with dense samples, for tracking by online structural SVMs. Analogous to correlation filters, we leverage the Fourier trick for fast training.

Our second motivation: Despite the number of training samples, designing a robust similarity function is also of importance for visual tracking. The SVM based trackers represent the similarity measure by a decision function which decomposes into a set of kernel inner productions. Kernel tricks have long been used in the SVM learning problem and the correlation filters. By applying a kernel function, samples data are implicitly mapped to a high dimensional space where a separating plane can be better learned. Interestingly, though complex kernel functions are thought to perform better, most tracking methods prefer the linear kernel [14,25]. An explanation is that the linear kernel is computationally more efficient. This paper proposes another kernel function, the weighting kernel, which can greatly boosts the tracking performance without raising the dimensions of samples data, thus being as efficient as the linear kernel.

The rest of the paper is organized as below: Section 3 details the online structural learning problem and the proposed method to learn with dense samples as well as a weighting ker-

Please cite this article as: X. Yu, Q. Yu, Online structural learning with dense samples and a weighting kernel, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.05.017

Download English Version:

https://daneshyari.com/en/article/6940507

Download Persian Version:

https://daneshyari.com/article/6940507

Daneshyari.com