J. Vis. Commun. Image R. 35 (2016) 146–154

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Discriminative descriptors for object tracking ☆

Yang Ruan, Zhenzhong Wei *

Beihang University, Key Laboratory of Precision Opto-Mechatronics Technology of Ministry of Education, XueYuan Road No. 37, HaiDian District, Beijing 100083, China

ARTICLE INFO

Article history:
Received 2 September 2015
Accepted 14 December 2015
Available online 19 December 2015

Keyword:
Object tracking
Correlation filter
Color descriptors
Computer vision
Scale variation
Target representation
Model-free tracking
Real-time processing

ABSTRACT

Object tracking is one of the most challenging problems in computer vision. Only Fast trackers can satisfy the real-time requirements and can be used in many artificial intelligence applications. Due to the impressive high-speed, correlation filters have received much attention within the field of object tracking. Recently, trackers using luminance information or color names for image description have been performed in a correlation filters framework. In this paper, we propose the usage of discriminative color descriptors to improve the tracking performance of the traditional correlation filters tracker. Discriminative color descriptors are compact and efficient. Moreover, our tracker incorporates scale estimation into the traditional correlation filters, which results in increased tracking performance. Extensive experiments demonstrate that the proposed tracker can obtain superior results compared to existing trackers using correlation filters and it is also able to outperform state-of-the-art trackers on the CVPR2013 object tracking benchmark.

1. Introduction

Despite that object tracking has been studied for several decades [1–3], it remains a challenging task to handle pose variation, illumination variation, scale variation, and partial occlusion, etc. The robust tracking in complex environments is significant for a lot of applications including motion analysis, vision systems, surveillance or human–computer interaction. In this field, researchers [4–9] tracked an arbitrary object with no prior knowledge. The approach is referred to as model-free tracking.

Single object tracking is the most common task within the field of computer vision. Several approaches have been proposed for single object tracking. Adam et al. [10] presented a novel algorithm called FragTrack (Frag) which can overcome partial occlusions and pose change by combining multiple patch votes. Grabner and Bischof [11] proposed a novel on-line AdaBoost (OAB) feature selection method. In [12], Babenko et al. showed that using Multiple Instance Learning (MIL) instead of traditional supervised learning can lead to a more robust tracker. Kalal et al. [13] proposed a novel tracking framework (Tracking–Learning–Detection, TLD for short) that the long-term tracking task was explicitly decomposed into tracking, learning and detection. In [14], Hare et al. provided an online learning tracker with a kernelized structured output support vector machine (SVM). In [15], they presented Extended Lucas Kanade or ELK that it casts the original LK algorithm [16] as a maximum likelihood optimization and then they extended it by considering pixel object/background likelihoods in the optimization. These algorithms rely on intensity or texture information for image description and include complex appearance models and optimization methods. It is difficult for them to keep up with the 25 frames per second (FPS) demanded for real-time processing.

Recently, a number of excellent algorithms using correlation filters have been proposed for a variety of applications like object detection and recognition, alignment, tracking etc. Due to the impressive high-speed, correlation filters have received much attention. In [18], authors devised Kernel classifiers with the same characteristics as correlation filters. The Kernel classifiers can be trained and evaluated quickly with the Fast Fourier Transform (FFT). The tracker exploiting the Circulant Structure of Tracking-by-detection with Kernels (CSK) can be implemented with only a few lines of code and runs at hundreds of frames-per-second. The article [19] proposed a vector correlation filter (VCF) directly using with vector features likes color names (CN) [21]. The vector correlation filter is an extension of the unconstrained correlation filters. This is different from traditional correlation filters designed using scalar features (most commonly pixel values). Multi-channel correlation filters [20] are similar to vector correlation filter (VCF) in the core of the study. A multi-channel image correlated/convolved with a multi-channel filter gives a single-channel response. In [22], authors extended the CSK tracker with color names (CN). Color names have good balance between photometric invariance and discriminative power and have shown been suitable for object recognition [23].

☆ This paper has been recommended for acceptance by M.T. Sun.
* Corresponding author.

http://dx.doi.org/10.1016/j.jvcir.2015.12.009

In this paper, we propose the usage of discriminative color descriptors to improve the tracking performance of the traditional correlation filters tracker. Discriminative color descriptors (DD) are compact and efficient, which is designed by Khan et al. [24]. The discriminative descriptor (DD) was found to outperform various other pure color descriptors and the color names descriptor [25,26]. Moreover, our tracker incorporates scale estimation into the traditional correlation filters, which results in increased tracking performance. Adding this technique helps the capability to handle target scale changes. Extensive experiments demonstrate that the proposed tracking algorithm is able to outperform state-of-the-art trackers on various challenging videos. Fig. 1 shows screenshots of sampled tracking results. Only top five trackers on success rate are presented.

The remainder of this paper is organized as follows. In Section 2, we review the CSK tracker based on a traditional correlation filter framework. In Section 3, we describe our approach. In Section 4, we show experimental results. Finally, we present our conclusions.

2. The CSK tracker

Correlation filters have shown superior performance on a number of computer vision problems. The CSK tracker [18] is based upon a kernelized single-channel correlation filter and runs at hundreds of frames-per-second. In the CVPR2013 object tracking benchmark, the CSK tracker has shown to obtain competitive performance and the highest speed. In this section, we briefly describe the CSK tracker.

2.1. Training Samples and Labels

Training Samples and Labels are used as inputs of the classifier. A set of image patches are obtained around the target center, which are referred to as training samples. The image patch x of size $m' \times n'$ is training sample and its corresponding confidence score is y. The label in a large majority of trackers is a binary value in general. In this paper, the labels computing by a Gaussian function are continuous values. The confidence score will be 1 near the object center, and is gradually reduced to 0 as far away from the target. The total of locations is $m' \times n'$.

The label of ith location is,

$$y_i = \exp(-0.5/s^2 \times (r_i^2 + c_i^2)) \tag{1}$$

where the spatial bandwidth $s = \sqrt{m'n'}/16$, and the r_i and c_i represent the horizontal and vertical pixel differences between location and target center.

2.2. Training

Training Samples and Labels $(x_1, y_1), \ldots, (x_m, y_m)$ are used as inputs of the classifier. A classifier $f(x)$ is trained by finding the parameter w which minimizes the cost function. The cost function minimization problem is written as:

$$\min_f \sum_{i=1}^{m} L(y_i, f(x_i)) + \lambda ||w||^2 \tag{2}$$

where $L(y, f(x))$ is a loss function, and λ is a parameter for regularization. Regularized Least Squares (RLS) uses the quadratic loss $L(y, f(x)) = (y - f(x))^2$.

The kernel trick can convert inputs to a higher dimensional feature space. The kernel is defined as $k(x, x') = \langle \varphi(x), \varphi(x') \rangle$, where $\varphi(x)$ is a mapping. The solution $w = \sum_i \alpha_i \varphi(x_i)$ is a linear combination of the training samples. The online classifier coefficients are updated by updating the solution w over time. So, the goal is to find the solution w.

2.3. Simple closed form solution

The simple closed form solution of RLS with Kernels (KRLS):

$$\alpha = (K + \lambda I)^{-1} y \tag{3}$$

where K is the kernel matrix with elements $K_{ij} = k(x_i, x_j)$, and I is the identity matrix. The vector y has elements y_i.The λ is a parameter for regularization. Under suitable conditions, the kernel matrix becomes circulant. The properties of circulant matrices make them particularly amenable to manipulation, since their sums, products and inverses are also circulant. This makes online learning more efficient.

The simplest kernel is linear kernel. The form of linear kernel is $k(x, x') = \langle x, x' \rangle$. There are some other kernel functions. The Gaussian Kernel is used in the CSK tracker.

Using circulant matrices, the Fast Fourier Transform can be used in fast training and detection. The cost function minimization problem has the simple closed form solution:

$$\alpha = F^{-1}\left(\frac{F(y)}{F(k) + \lambda}\right) \tag{4}$$

where F and F^{-1} denote the Fourier transform and Fourier inverse transform respectively.

In summary, the goal has changed to find the vector α. The solution w is implicitly represented by the vector α.

2.4. Fast detection

In the new frame, a grayscale patch z of size $m' \times n'$ is obtained in a search region around object location. The Kernel classifier can detect quickly with the Fast Fourier Transform (FFT). A classier response is computed for each single input. All the responses are evaluated simultaneously. The confidence map of the target center is obtained by,

$$\hat{y} = F^{-1}(F(\bar{k}) \odot F(\alpha)) \tag{5}$$

where $\bar{k}$ is a vector with elements $\bar{k}_i = k(z_i, \hat{x})$, F and F^{-1} denote the Fourier transform and Fourier inverse transform respectively. The learned target appearance $\hat{x}$ is updated overtime. The best object location can be estimated by maximizing the confidence map.

3. Our approach

The goal of our approach is to improve the tracking performance of the CSK tracker. We propose the usage of discriminative color descriptors to improve the tracking performance. Moreover, our tracker incorporates scale estimation into the traditional correlation filters, which results in increased tracking performance. The details of our approach are described as follows.

3.1. Discriminative color descriptors for tracking

The right features are exactly those that make the tracker work best. Color descriptors have been successful in many computer vision applications. Here, we investigate them for the visual tracking problem. Discriminative color descriptors (DD) are compact and efficient, which is designed by Khan et al. [24]. In their paper they took an information theoretic approach to color description and clustered color values together based on their discriminative power in a classification problem. The discriminative color descriptors have two settings, namely 11 and 25 clusters. The descriptor with 25 dimensions can obtain better tracking performance in