Journal of Visual Communication and Image Representation 51 (2018) 104–111

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Learning adaptively windowed correlation filters for robust tracking☆

Yangliu Kuai*, Gongjian Wen, Dongdong Li

College of Electronic Science, National University of Defense Technology, Changsha, Hunan, China

A R T I C L E I N F O

Keywords:
Correlation filter
Target likelihood
Window adaptation

A B S T R A C T

Visual tracking is a fundamental component for high-level video understanding problems such as motion analysis, event detection and action recognition. Recently, Discriminative Correlation Filters (DCF) have achieved enormous popularity in the tracking community due to high computational efficiency and fair robustness. However, the underlying boundary effect of DCF leads to a very restricted target search region at the detection step. Generally, a larger search area is adopted to overcome this disadvantage. Such an expansion of search area usually includes substantial amount of background information which will contaminate the tracking model in realist tracking scenarios. To alleviate this major drawback, we propose a generic DCF tracking framework which suppresses background information and highlights the foreground object with an object likelihood map computed from the color histograms. This object likelihood map is merged with the cosine window and then integrated into the DCF formulation. Therefore, DCF are less burdened in the training step by focusing more on pixels with higher object likelihood probability. Extensive experiments on the OTB50 and OTB100 benchmarks demonstrate that our adaptively windowed tracking framework can be combined with many DCF trackers and achieves significant performance improvement.

1. Introduction

Visual tracking is a classical computer vision problem with many applications in multimedia such as video surveillance, augmented reality and human-computer interaction. Generic tracking means single-camera, single-object, short-term and model-free tracking [1–3], which estimates the trajectory of a target in the whole video given only its initial state (usually an axis-aligned rectangle) in the first frame. *Short-term* implies re-detection modules are unnecessary while *model-free* means neither pre-learned object models nor class-specific prior knowledge are permitted. Despite significant progress in recent years, robust tracking under complicated scenarios is still challenging due to illumination change, self-deformation, partial occlusion, fast motion and background clutter.

Recently, discriminative correlation filters (DCF) have achieved enormous popularity in the tracking community due to the high computational efficiency and excellent performance. Numerous negative samples are generated from the periodic extension of the positive sample to train the correlation filter. With the circular structure, DCF transform computationally consuming spatial correlation into efficient element-wise operation in the Fourier domain and achieve extremely high tracking speed. However, standard DCF suffer from the boundary effect produced by the periodic assumption. The detection scores are only accurate near the center of the search area, which leads to a very restricted target search area at the detection step. Generally, this restricted target search area hampers the tracking performance of DCF in presence of challenging factors such as fast target motion and partial occlusion. An naive solution to this problem is expanding the target search area used for training and detection in DCF. However, such an expansion results in a larger periodicity and includes substantial amount of background information which contaminates the tracking model and degrades the discriminate power of DCF, thus leading to inferior tracking performance.

Generally, there exist two popular ways to mitigate the contamination of the background information, cosine window and regularization window. In standard DCF, the features extracted from the cropped image patch are weighted by the cosine window to suppress the excessive background and focus on the target. Later, Spatially Regularized Discriminate Correlation Filters (SRDCF) [4] introduce a spatial regularization window into the standard DCF formulation. This regularization window penalizes correlation filter coefficients residing in the background region to mitigate the emphasis on the background information in the learned classifier. The SRDCF formulation allows the correlation filters to be learned on a significantly larger set of negative samples without corrupting the positive samples and achieves significant performance gain against standard DCF formulation. However,

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

* Corresponding author.
E-mail address: kuaiyangliunudt@163.com (Y. Kuai).

https://doi.org/10.1016/j.jvcir.2018.01.008
Received 16 October 2017; Received in revised form 11 January 2018; Accepted 14 January 2018
Available online 04 February 2018

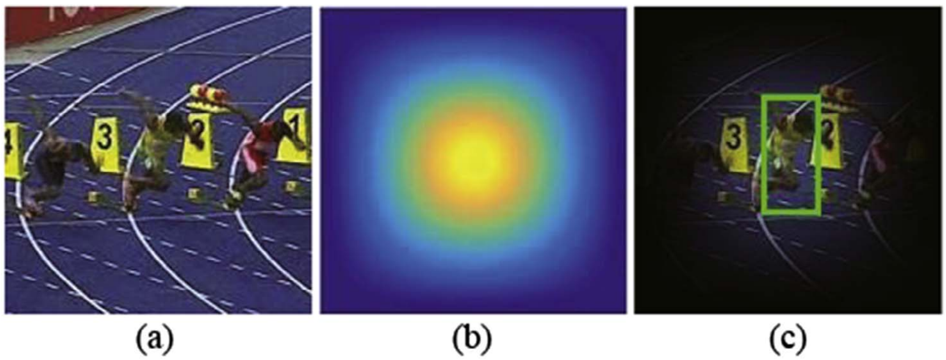

Fig. 1. Image patch (a), cosine window (b) and image patch weighted by the cosine window (c). In (c), the target bounding box contains certain background information and much background information remains unsuppressed around the target.

there are several limitations of the above mentioned two ways which will hamper the tracking performance. (1) The initial target bounding box in the first frame includes the target appearance along with certain amount of background information especially for irregular and deformable targets, as shown in Fig. 1(a). This background information severely corrupts the positive sample. (2) The cosine window is centrosymmetric and fixed during tracking while the target is deformable, noncentrosymmetric and changing its appearance from frame to frame, as shown in Fig. 1(b). (3) The cosine window suppresses the background information near the boundary of the target search area more severely than near the center, as shown in Fig. 1(c), which leaves much background around the target and contaminates the fast-moving target. (4) The introduced regularization window destroys the regularity of DCF formula and needs iterative solution, which is quite time-consuming.

In this work, we tackle the aforementioned limitations by introducing adaptive window into the standard DCF formulation, the flowchart is shown in Fig. 2. The adaptive window is generated by merging the fixed cosine window with an adaptive likelihood map. The likelihood map is built from the global color histogram and gives high values in regions resembling the target appearance, and otherwise low values to regions resembling the background. This global color histogram can be computed and updated efficiently in each frame, thus increasing only little burden to the calculation of correlation filters. Extensive experiments on the popular OTB50 [5] and OTB100 [6] benchmarks show that our adaptively windowed tracking framework can be combined with many DCF trackers to achieve significant overall performance improvements.

2. Related work

There have been many advances in the object tracking literature in the recent years. Due to space limitations, here we focus on those that are most relevant to our work.

2.1. Discriminate correlation filters

In recent years, discriminate correlation filters (DCF) based trackers

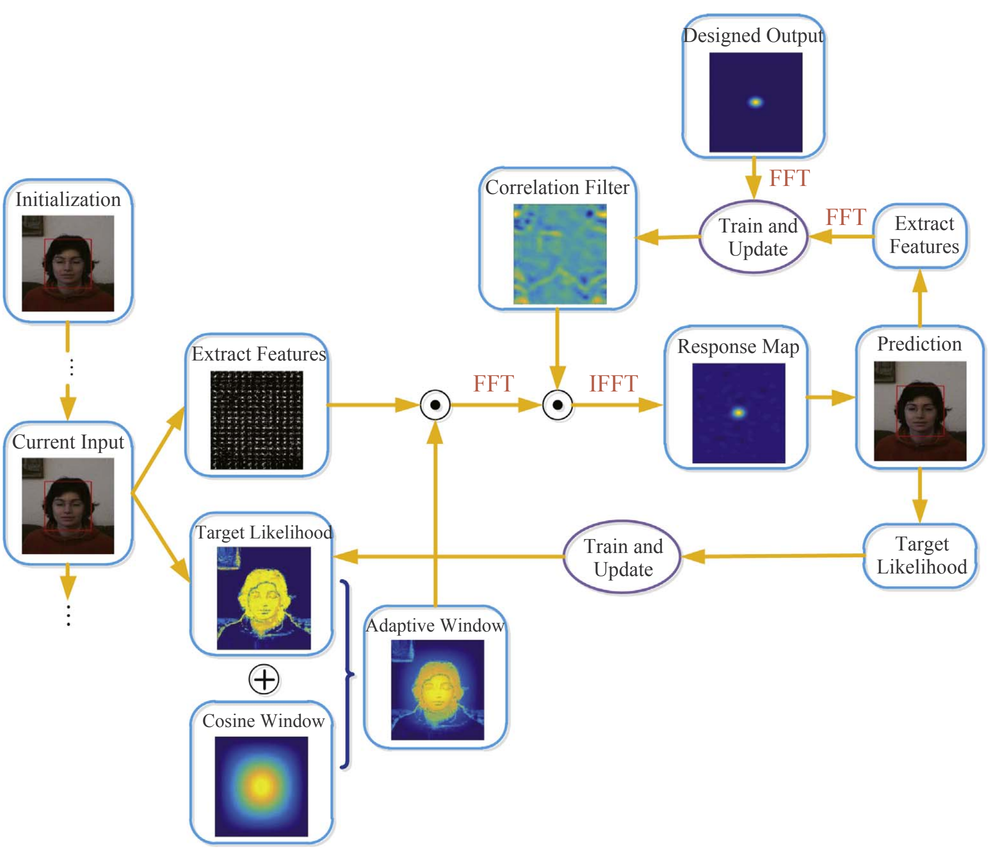

Fig. 2. The flowchart of our proposed tracking framework. We introduce an adaptive window into the existing DCF based trackers to better suppress the background while maintaining the target information. The adaptive window is generated by merging the fixed cosine window with an adaptive likelihood map. The likelihood map is built from the global color histogram and gives high values in regions resembling the target appearance and thus suppressing the background information at the same time. This global color histogram is computed and updated efficiently in each frame to better capture the target.