applied sciences

MDPI

Article

Adaptive Context-Aware and Structural Correlation Filter for Visual Tracking

Bin Zhou [1,*] and Tuo Wang [1,2]

[1] Systems Engineering Institute, Xi'an Jiaotong University, Xi'an 710049, China; twang@sei.xjtu.edu.cn
[2] Xi'an Jiaotong University Suzhou Academy, Suzhou 215123, China
* Correspondence: captain_zhou@stu.xjtu.edu.cn; Tel.: +86-186-2903-8020

Received: 24 January 2019; Accepted: 26 March 2019; Published: 29 March 2019

Featured Application: There are many application scenarios for tracking. For example, tracking can be used to assist driving to improve driving safety. Moreover, tracking can assist the police to search for suspicious vehicles in massive videos.

Abstract: Accurate visual tracking is a challenging issue in computer vision. Correlation filter (CF) based methods are sought in visual tracking based on their efficiency and high performance. Nonetheless, traditional CF-based trackers have insufficient context information, and easily drift in scenes of fast motion or background clutter. Moreover, CF-based trackers are sensitive to partial occlusion, which may reduce their overall performance and even lead to failure in tracking challenge. In this paper, we presented an adaptive context-aware (CA) and structural correlation filter for tracking. Firstly, we propose a novel context selecting strategy to obtain negative samples. Secondly, to gain robustness against partial occlusion, we construct a structural correlation filter by learning both the holistic and local models. Finally, we introduce an adaptive updating scheme by using a fluctuation parameter. Extensive comprehensive experiments on object tracking benchmark (OTB)-100 datasets demonstrate that our proposed tracker performs favorably against several state-of-the-art trackers.

Keywords: visual tracking; correlation filter; context-aware; partial occlusion; model updating

1. Introduction

Visual object tracking is a crucial research issue in computer vision and has many applications including video security, traffic monitoring, robotics and human computer interface. In the past decade, great improvements have been made by some visual tracking algorithms [1–6], meanwhile large visual tracking datasets and benchmarks, such as the Object Tracking Benchmark (OTB)-50 [7], OTB-100 [8], Visual Object Tracking (VOT) [9,10], Need for Speed (NfS) [11] and Multi-Object Tracking (MOT) [12] have sparked the interest of numerous scholars and given impetus to the research area appreciably. In spite of having made considerable headway in recent decades, the tracking issue is still considered a big challenge in some scenarios such as illumination variation, scale variation, occlusion, fast motion, deformation, background clutters, etc.

Lately, correlation filter (CF) based methods [13–15] have been sought in visual tracking based on their excellent performance. CF-based trackers learn a correlation filter with training samples obtained by cyclically shifting base samples. For each new frame, the trained filter locates to the new target position where has the maximum response. At last, the filter is updated by the new location. By utilizing the circulant construction of training samples, CF-based trackers enable efficient learning and detecting in the Fourier domain.

Despite their appealing performance both in accuracy and high frames per second (FPS) rate, there are some drawbacks that reduce the performance of CF-based trackers. A major drawback is the boundary effect caused by circularly shifting. To inhibit the boundary effect, a cosine window is usually employed. But the cosine window limits search areas and thereby the quantity of samples, specifically the quantity of negative samples, is sharply reduce. Therefore, CF-based trackers usually have insufficient context and thus readily drift in scenes of fast motion or background clutter. Moreover, CF-based trackers only learn a holistic model so that they cannot deal with cases of partial occlusion well.

To further improve the performance of CF-based trackers, several effective methods have been proposed. By handling large scale changes in complex image sequences, the Scale Adaptive With Multiple Features Tracker (SAMF) [16], and the Discriminative Scale Space Tracker (DSST) [17] have achieved state of the art performance. Sum of Template and Pixel-wise LEarners (Staple) trackers [18] have learned a filter to enhance the robustness of color variations and distortions by combining two notations easily affected by complementary factors. However, these trackers have not more effective approaches to mitigate the boundary effect. The Spatially Regularized Discriminative Correlation Filters (SRDCF) tracker [19] induces boundary effects by adopting spatially regularized elements in training to punish correlation modulus according to their position. However, solving this optimization problem of this strategy is time-consuming. Lately, context-aware (CA) [20] and background-aware [21] CF-based trackers have demonstrated notable enhancement on visual tracking performance. These methods incorporate global context information into the filter training to add number of negative samples, the quantity of negative samples will obviously affect tracking results. However, there are two major problems: (1) context patches are simply selected on fixed positions around the target, which causes the problem that these selected negative samples are perhaps non-representative; and (2) only a holistic model is learned, these trackers lack robustness against partial occlusion.

To tackle the above inherent problems, we propose an adaptive context-aware and structural correlation filter for visual tracking from the following three perspectives. Firstly, in order to add informative negative samples to inhibit the boundary effect caused by circularly shifting, we introduce a novel context selecting strategy to obtain negative samples. We evenly divide the target neighborhood into four blocks, then the maximum response patch is selected from each block as the context patches for collecting negative samples. Negative samples selected by this strategy are more informative. Secondly, to gain robustness against partial occlusion, we construct a structural correlation filter by learning both the holistic and local model. The final response of the correlation filter can be obtained by integrating every component with adaptive weight based on the Peak to Sidelobe Ratio (PSR) that is used to measure the reliability of each component. The new location of target is predicted according to the final weighted response. Finally, we propose an adaptive updating scheme by using a fluctuation parameter. The parameter is calculated by the deviation of PSR between two consecutive frames, which can describe the degree that appearance changing is alleviating or deteriorating. The updating scheme can adaptively adjust the updating rate for different frames, thereby updating the filter more accurately. We evaluate our proposed method on OTB-100 benchmark. The benchmark annotated 100 video sequences with 11 sub-categories (including occlusion, background clutter, illumination variation, scale variation, deformation, motion blur, fast motion, in-plane rotation, out-of-plane rotation, out-of-view and low resolution) that describe the different challenges in the tracking problem. The precision plot and success plot are used to carry out quantitative analysis in the benchmark. The procedure of the adaptive context-aware and structural correlation filter shows in Figure 1. Extensive comprehensive experiments on OTB-100 datasets demonstrate that our approach performs favorably against state-of-the-art methods.

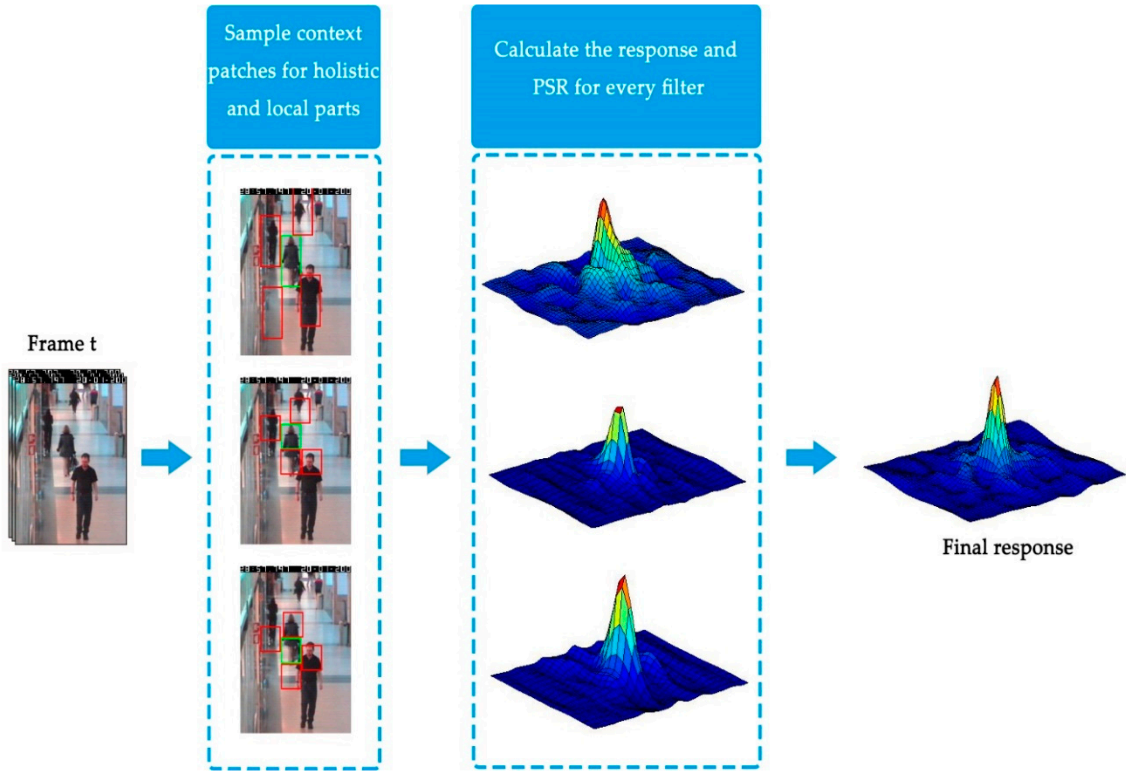

Sample context patches for holistic and local parts
Calculate the response and PSR for every filter
Frame t
Final response

Figure 1. Procedure of the adaptive context-aware and structural correlation filter. For a new frame, sampling optimal context patches for holistic and local parts. Then computing the response and Peak to Sidelobe Ratio for every filter. Finally, combining confidence maps of each parts with adaptive weight to obtain the final confidence map.

The rest of the paper arranged as follows. In Section 2, we discuss recent visual tracking algorithms. Section 3 elaborates the proposed adaptive context-aware and structural correlation filter exhaustively. Section 4 evaluates the performance of our proposed tracker by Visual Tracking benchmark [8] datasets, and compares it with several state-of-art trackers, and Section 5 concludes this paper.

2. Related Works

Visual tracking is a challenging research issue [22] in computer vision. The existing trackers are generally partitioned into two categories [23]. The first category is to achieve the purpose of tracking by modeling the target and combining the target detection method, which is called generative [24–28]. Kwon et al. [2] have proposed a visual tracking decomposition algorithm that employs a diverse observation-motion model to explain a comparatively apparent shape change caused by lighting variation and fast motion. The algorithm has been further developed to search for proper trackers by Markov Chain Monte Carlo sampling [3]. Another category of tracking methods, which is called discriminative, is to treat the issue as a classification task of foreground (target) and background, by training classifiers in combination with detection and tracking to track target. Discriminative methods are usually similar to tracking-by-detection [22]. Tracking-by-detection trackers [1,5,29,30] are more popular in current research, because of their the high performance and efficiency. Hare et al. [1] propose a Structured Output Support Vector Machine (SVM) and Gaussian kernels that locate a target directly. Kalal et al. [5] make use of structural constraints to direct the sampling of a boosting method. Zhang et al. [29] propose compressive tracking that learns a Naive Bayesian Model using a compressive sensing theory for dimensionality reduction of Haar-like features to accelerate calculation. Khattak et al. [31] merge multiple features into tracking to improve robustness. Additionally, the Sparsity-based Collaborative Model (SCM) [4] integrates both discriminative and generative classifier to generate a more robust tracker.

2.1. Correlation Filter Based Trackers

Over the past decade, CF-based trackers [32] have attracted more interest in the research of tracking, as well as attained very effective improvements in diverse challenges [9] and benchmarks [8]. Traditionally, correlation peaks are typically generated for each patch of interest in a frame while producing a low response to the background, which is often used as a detector of the desired model. Although tracking issue can be successfully resolved using these filters, they are not suitable for online tracking as the required training needs. This situation has been changed after the Minimum Output Sum of Squared Error (MOSSE) [13] filter has been proposed. By an adaptive learning method, the MOSSE tracker has robust and efficient results in tracking. According to the fundamental theory of MOSSE, multiple improved algorithms followed. For instance, Henriques et al. [14] improve the MOSSE filter by adopting the Circulant Structure of with Kernels (CSK). Danelljan et al. [33] employ Color Name features that can convey color attributes to improve the tracking performance of the CSK tracker. Since high dimension of color features, the updating scheme [33] is adapted to reduce the dimension of feature by principal Component Analysis (PCA). An improved Kernelized Correlation Filter (KCF) [15] employed multi-channel features to replace raw pixels, which is the most prevalent filter widely adopted since it has good overall performance and high FPS rate. Zhang et al. [34] formulate a distribution scheme in a Bayesian optimization framework to alleviate drifting.

By further coping to scale variation, SAMF [16] and DSST tracker [17] based on correlation filter have appealing performance, and have defeated the rest of the participating methods from aspects of accurateness and robustness in recent competitions. Li et al. [35] introduce a variant of SAMF based on water flow driven minimum barrier distance (MBD) algorithm. Staple [18] has learned a model to improve the robustness of both color variations and distortions by combining two notations easily affected by complementary factors. To induce the boundary effects, SRDCF tracker [19] exploits spatially regularized elements in training to punish correlation modulus according to their position. Recently, convolution features have also been used for visual tracking tasks to improve tracking accuracy and robustness. Deep SRDCF [36] employs deep features of single convolution layer in Convolutional Neural Networks (CNN) for tracking, and Continuous Convolution Operator Tracker (C-COT) [37] extends it to multi-layers of convolution.

2.2. Context-Aware Based Trackers

Furthermore, context-aware trackers have demonstrated notable improvement in tracking performance. Dinh et al. [30] propose a method that automatically explores distractors and supporters around the tracking object adopting a consecutive random forest classifier. Distractors are patches which have similar appearance with the target and consistently have high response value. Xiao et al. [38] utilize contextual information by using multi-level clustering. In more recent research, Mueller et al. [20] incorporate global context information into the filter training to add a number of negative samples, and remodel the original regression function, while offering a closed-form solution for both single and multi-channel features. However, the CA tracker [20] simply samples context patches on fixed locations around the target, which causes the problem that selected negative samples are perhaps non-representative. Qie et al. [39] sample patches of top m response value directly on the whole neighboring region of target and then employ K-means cluster algorithm to select hard negative samples. Our proposed context selecting strategy is similar to [39]. However, the difference between these two methods is evident. The method proposed by [39] directly samples patches of top m response around the target. These patches may be concentrated in a small area, thus ignoring distractors from other areas. Our method divides the target neighborhood evenly into four blocks and then selects the maximum response patch from each block. The distractors sampled by our scheme are more representative. In addition, the clustering process employed by [39] will consumes extra time.

2.3. Part-Based Correlation Filter Trackers

Jeong et al. [40] propose an adaptive partial block scheme to attenuate the influence by partial occlusion. Sun et al. [41] propose a framework that employs shape-preserved scheme to overcome object variations for each individual sub-part. Liu et al. [42] retain the spatial distribution structure of each sub-part and take advantage of the inherent connection of sub-parts to improve tracking performance. However, all these trackers [41,42] just focus on local parts, ignoring the relationship between holistic target and local models. Akin et al. [43] cope with the partial occlusion tasks by applying coupled interactions between a holistic tracker with some local trackers. Fan et al. [44] combine holistic and local models to catch the internal structure of target to make to filter more robust for part-occlusion. Structural models of [43,44] are based on the response value which is perhaps misleading as the response between the tracking window and the current target that is continuously adapted by learning from the previous match. Jeong et al. [45] pick the maximum response part from both global and sub-block as the appointing tracking object, but employing an independent model isn't reliable in occlusion tasks. Different from the above methods, we construct a structural correlation filter by learning both the holistic and local model. The final response of the correlation filter can be obtained by integrating every component with adaptive weight based on the Peak to Sidelobe Ratio (PSR) that is used to measure the reliability of each component. And the final location of target is predicted according to the final weighted response.

2.4. Adaptive Update Schemes

Moreover, in order to ensure long-term tracking, the correlation filter should be updated robustly. In the process of losing for the object, the updating rate must be reduced to avoid updating much harmful confusion into appearance model. Several researches have focused on long-term components for failure tracking strategies. Ma et al. [46] propose an online random fern classifier as re-detection component for long-term tracking (LCT). Jeong et al. [45] introduce an adaptive learning rate by calculating the ratio between a value of current frame and a desire result based on the Peak to Sidelobe Ratio. However, the above schemes neglect the interframe deviation, which can measure the degree that appearance changing is alleviating or deteriorating.

In this paper, we propose an adaptive context-aware and structural correlation filter for object tracking. Firstly, we propose a novel context selecting strategy to obtain negative samples, instead of sampling on fixed locations. Secondly, to gain robustness against partial occlusion, we construct a structural correlation filter by learning both the holistic and local models. Finally, we introduce an adaptive update scheme by using a fluctuation parameter.

3. The Proposed Tracker

In this section, we elaborate our proposed adaptive context-aware and structural correlation filter. Firstly, we review the classical KFC tracker [15], and then a novel context selecting strategy to obtain negative samples will be introduced. Moreover, we construct structural correlation filter to improve robustness against partial occlusion. Finally, a fluctuation parameter is proposed to update the model.

3.1. The KCF Tracker

Like other CF-based methods, our proposed method is also based on the KCF tracker. Therefore, before the exhaustive discussion of our proposed method, we first review the KCF tracker [15]. The KCF tracker achieves excellent results and high-speed performance on the visual tracker benchmark [8], despite the idea and implementation of the KCF tracker being very simple. The KCF tracker collects positive and negative samples around the target using the structure of the circulant matrix, to improve the discriminative capability of the track-by-detector tracker. The circulant matrix can be diagonalized with the Discrete Fourier Transform (DFT), enabling a fast dot-product instead of an expensive Matrix algebra.

The goal of the KCF tracker is to find a function that minimizes the squared error over data matrix **X** and their regression target **y**,

$$\underset{\mathbf{w}}{min} \parallel \mathbf{Xw} - \mathbf{y} \parallel^2 + \lambda \parallel \mathbf{w} \parallel^2 \tag{1}$$

where the square matrix **X** contains all circulant shifts of the base sample **x**, the regression target **y** is Gaussian-shaped, and the λ is a regularization parameter to ensure the generalization performance of the classifier, Equation (1) has the closed-form solution.

$$\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{2}$$

The circulant matrix **X** has some intriguing properties [47,48], and the most useful one is that the circulant matrix can be diagonalized by the Discrete Fourier Transform (DFT) as below:

$$\mathbf{X} = \mathbf{F}^{\mathrm{H}} diag(\hat{\mathbf{x}})\mathbf{F} \tag{3}$$

where **F** is the DFT matrix, and $\mathbf{F}^{\mathrm{H}}$ is the Hermitian transpose. $\hat{\mathbf{x}}$ denotes the DFT of **x**, $\hat{\mathbf{x}} = F(x) = \sqrt{\mathrm{n}}\mathrm{F}\mathbf{x}$.

Applying Equation (3) into the solution of linear regression (Equation (2)), we have the solution as below:

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}} \odot \hat{\mathbf{y}}}{\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}} + \lambda} \tag{4}$$

where $\hat{\mathbf{x}}^*$ is the a complex-conjugate of $\hat{\mathbf{x}}$. The symbol $\odot$ and the fraction denote element-wise product and division respectively.

For detecting the new location of target in the next frame, we can compute the response $f(z)$ for all candidate patches **z**, and diagonalize $f(z)$ to obtain as below:

$$\hat{f}(z) = \hat{\mathbf{w}} \odot \hat{\mathbf{z}} \tag{5}$$

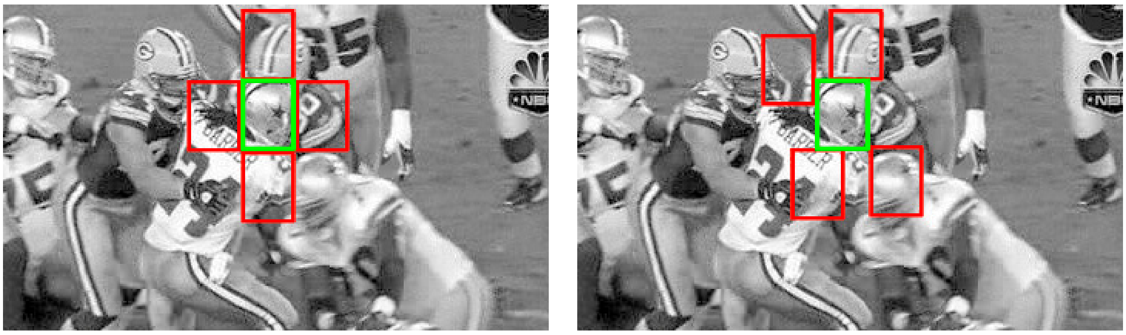
The candidate patch with the maximum response is considered as the new location of target.

3.2. Context Selection Strategy for Tracking

Context information of the tracking target has a significant effect on overall performance. For instance, in the scenery with lots of background clutter, selecting optimal context patches is very important for successful tracking. Mueller et al. [20] introduced a method based on a context-aware framework to add contextual information to the filter and got competitive results in some tracking challenges. However, the conventional CA tracker [20] simply samples context patches on fixed locations around the target, which causes the problem that these selected negative samples are perhaps non-representative (see Figure 2). The ideal strategy of selecting context patches is that context patches are sampled at locations where the filter response is high and spatially far from the maximum.

In this paper, we propose a novel context selecting strategy to obtain negative samples. Firstly, we divide the surroundings of the tracked object into four areas, each area is twice the target. Then, patches that have the maximum response value in their respective regions are selected as the context patches for collecting negative samples (as shown in Figure 3).

The method has two advantages. Firstly the strategy of conventional CA tracker may yield meaningless negative samples. By contrast, sampled negative patches by our proposed strategy are more similar with the target. In other words, these context patches will more probably become background distractors in next frames. Secondly, our method avoids redundancy which led by sampling patches of top m response value directly on the whole neighboring region of target.



(**a**) strategy of conventional CA [20].

(**b**) proposed strategy.

Figure 2. Two strategies of selecting context patches. The green rectangle denotes the target, and red rectangles denote selected context patches.

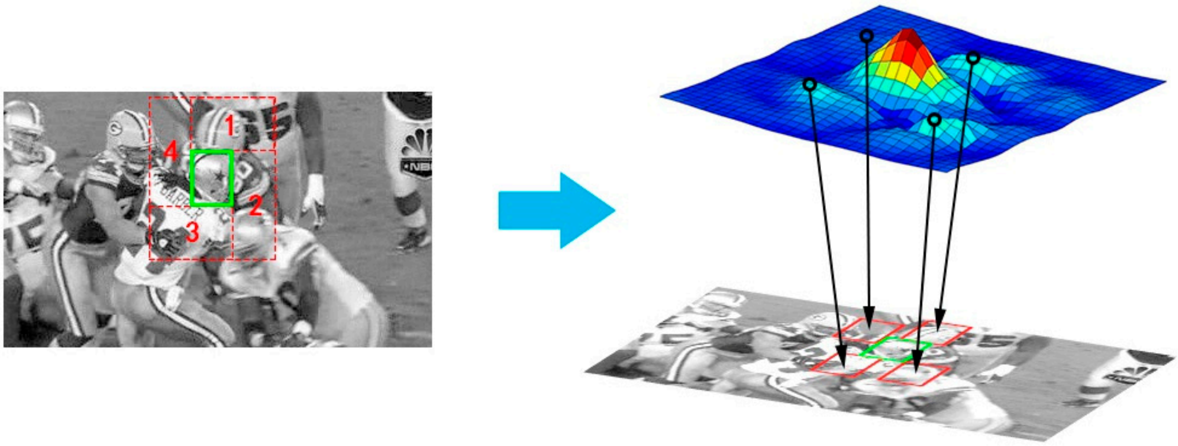


Figure 3. Sampling procedure of proposed strategy.

Next, we learn the correlation filter using the target samples and the negative samples based on CA [20]. Our goal is to train a filter with a high response on the target patch and closes to zero response on context patches. The regression function that adds context patches can be written as Equation (6):

$$\underset{\mathbf{w}}{min} \parallel \mathbf{X_0 w} - \mathbf{y} \parallel^2 + \lambda_1 \parallel \mathbf{w} \parallel^2 + \lambda_2 \sum_{\mathrm{i}} \parallel \mathbf{X_i w} \parallel^2 \tag{6}$$

where $\mathbf{X_0}$ and $\mathbf{X_i}$ are the corresponding circulant matrices of target and context patches, respectively. The parameter λ_2 can control context patches to regress to zeros.

We can stack context image patches under the target patch to establish a new data matrix $\mathbf{B}$, and concatenate $\mathbf{y}$ with zeros to construct a new regression function $\bar{\mathbf{y}}$. The Equation (6) can be rewritten as below:

$$\parallel \mathbf{Bw} - \bar{\mathbf{y}} \parallel^2 + \lambda_1 \parallel \mathbf{w} \parallel^2 \tag{7}$$

where $\mathbf{B} = \begin{bmatrix} \mathbf{X_0} \\ \sqrt{\lambda_2}\mathbf{X_1} \\ \vdots \\ \sqrt{\lambda_2}\mathbf{X_k} \end{bmatrix}$ and $\bar{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$.

Because the Equation (7) is convex, it can be minimized by setting the gradient to zero. The closed-form solution of Equation (7) can be obtained.

$$\mathbf{w} = \left(\mathbf{B^T B} + \lambda_1 \mathbf{I}\right)^{-1} \mathbf{B^T} \bar{\mathbf{y}} \tag{8}$$

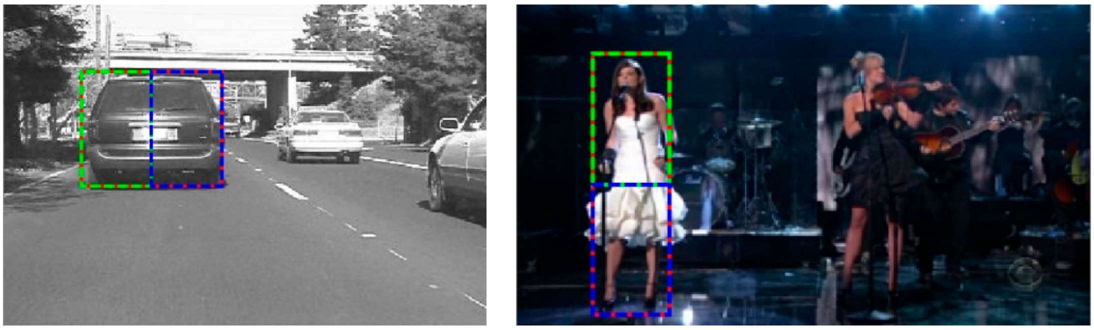
The closed-form solution in the Fourier domain can be calculated by applying Equation (3).

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}_0 \odot \hat{\mathbf{y}}}{\hat{\mathbf{x}}_0^* \odot \hat{\mathbf{x}}_0 + \lambda_1 + \lambda_2 \sum_{\mathrm{i}=1}^{\mathrm{k}} \hat{\mathbf{x}}_0^* \odot \hat{\mathbf{x}}_{\mathrm{i}}} \tag{9}$$

The detection formula is exactly the same as in the standard formulation in Equation (5).

3.3. Structural Correlation Filter

In visual tracking tasks, partial occlusion is one of the major challenges limiting performance of tracker. The correlation filter based on a single holistic appearance model is sensitive to partial occlusion, which may reduce the overall performance and even lead to failure in the tracking challenge. To address this defect, we combine the correlation filters of holistic target and local parts to preserve the inner spatial structure of the target. We employ effective spatial distributions to divide the target into two local parts, one for the horizontally and one for the vertically aligned object based on the ratio of the height and width of the target. The scale and positions of the local parts are strictly confined. As illustrated in Figure 4.



(**a**) (**b**)

Figure 4. Two spatial distributions based on the ratio of the height and width of the target. The red rectangle represents holistic part, the green and blue rectangles represent two local parts. (**a**) horizontally aligned object from car4; (**b**) vertically aligned object from single1.

In most situations, the relative movement between the holistic target and local parts is limited, so we can combine all confidence maps to obtain a robust correlation filter. It is critical to note how to combine confidence maps of each part. If we simply sum confidence maps with the same weight, the false tracked parts maybe unexpectedly highlighted. Intuitively speaking, the response of reliable parts should be given larger weight, while the failure detection part will have less weight. Fortunately, the Peak to Sidelobe Ratio (PSR) is a suitable indicator to quantify the reliability of filter. PSR is used to measure the sharpness of the correlation peak, the higher PSR score means more confident detection. Thence, we adopt PSR to define the weight of every part.

The PSR is defined as:

$$\rho_i = \frac{max(\boldsymbol{f_i}(z)) - \mu_i}{\sigma_i} \tag{10}$$

where μ_i and σ_i are the mean and the standard deviation of *i-th* confidence map, respectively.

The final confidence map can be defined as:

$$\boldsymbol{f}(z) = \sum_{i=1}^{3} c_i \cdot \rho_i \cdot \boldsymbol{f_i}(z) \tag{11}$$

where $c_1 = 1$ denotes the holistic part, $c_{2,3} = 0.5$ denote local parts.

3.4. Updating Strategy

During tracking, the appearance of the object will change due to many factors, such as deformation and rotation. Furthermore, the target object may be occluded by other objects. A conventional CF tracker, such as KCF, used a fixed learning rate to updating the model. If the tracker is occluded or drifts, the correlation filter will be contaminated. To tackle this problem, we propose an adaptive updating scheme by using a fluctuation parameter.

In the process of deforming or occluding for the object, we want to gradually reduce the learning rate to avoid updating much confusion into model. When the target is restoring, the learning rate should gradually increase to adapt the change of model. The deviation of PSR between two consecutive frames can describe the degree which appearance changing is alleviating or deteriorating. Based on the above motivation, we propose a fluctuation parameter m, the fluctuation parameter can adaptively adjust the updating rate for different frames. Moreover, when PSR score drops to a threshold, it is an indication that the object is heavy-occluded or tracking has failed. Therefore, the correlation filter shouldn't be updated when PSR score is less than the threshold.

For both holistic and local components, the fluctuation parameter of *t-th* frame can be computed as below:

$$m_i^t = exp\left(\boldsymbol{g}(\rho_i^t) - \boldsymbol{g}\left(\rho_i^{t-1}\right)\right),\ t > 1 \tag{12}$$

$$\boldsymbol{g}(\boldsymbol{z}) = \frac{1}{1 + e^{-(\rho_t - THR)}} \tag{13}$$

where the THR denotes the PSR threshold.

Finally, if $\rho_i^t > THR$, this component will be updated individually by its own learning rate $\gamma_i^t = m_i^t \beta$, else its model will be prohibited updating. Where the β is an initial learning rate.

4. Experiments

In this section, we introduce the detail and parameters of implementation, and the experimental methodology. Moreover, to evaluate the performance of the proposed adaptive context-aware and structural correlation filter, we implemented our method on the OTB-100 benchmark [8] and VOT2016 with comparisons to several recent state-of-the-art trackers.

4.1. Detail and Parameters

We implement our proposed method on Staple and SAMF trackers, and we name them Staple_SCA and SAMF_SCA, respectively. Staple_SCA tracker integrates HoG [49] and color histogram, and the scale scheme is based on DSST like Staple, the more information refer to [18]. Similar to SAMF tracker [16], SAMF_SCA copes to the scale variation with a scaling pool **S**, and $\mathbf{S} = \{0.985, 0.99, 0.995, 1.0, 1.005, 1.01, 1.015\}$. SAMF_SCA employs HoG and Color-naming [33] features, using a cell size of 4×4 and the number of bin is 9. We also increase the padding windows from 2.5 times of target object to 3, due to the increased robustness from context patches. The number of context patches k is set to 4. The regularization factor λ_1 is set to e^{-4}. We set the other regularization factor λ_2 to {0.5, 0.4} and the initial learning rate β are {0.015, 0.005} for Staple_SCA and SAMF_SCA, respectively.

For our proposed method, the PSR score of filter generally ranges between 16 and 34 under normal tracking conditions. When PSR score drops to around 7, it signifies that the object is heavy-occluded or tracking has failed. Thus, the threshold THR of PSR is set to 7.

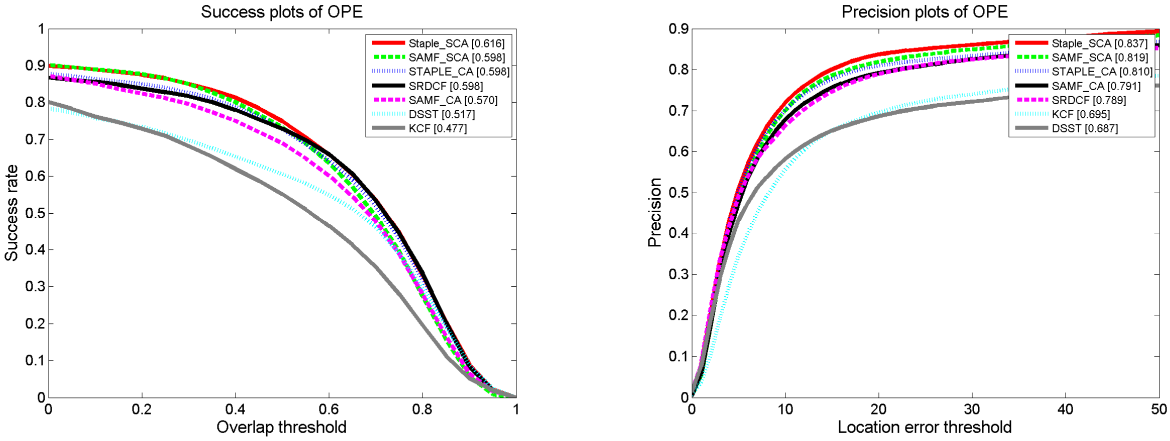
The proposed method is implemented in MATLAB R2014a version. All the experiments are conducted on an Intel Xeon(R) E3-1226 V3 CPU (3.30 GHz) PC with 32GB RAM. Staple_SCA and SAMF_SCA trackers run at 24.9 fps and 10.6 fps, which are still within real time range.

4.2. Experimental Methodology

We select two quantitative evaluation components from the object tracking benchmark [7,8], including precision and success. Precision estimates the center location error between the bounding box of the tracked targets and the bounding box of ground truths. In the precision plot, the center location error in pixel distance varies along the x-axis, and the y-axis represents the percentage of accurately located bounding window per threshold. The threshold is set at 20 pixels [8] for ranking trackers. Success is calculated as the intersection-over-union (IOU) of the tracking bounding window and the labeled ground truths. In the success plot, the IOU overlap rate varies along the x-axis, and the y-axis represents the percentage of accurately located bounding window per threshold. The final ranking is determined by the area under the curve (AUC), computed from the mean of the success rates corresponding to the sampled overlap thresholds from 0 to 1.

4.3. Overall Performances

To indicate the performance improvements of our approach, we compare Staple_SCA and SAMF_SCA trackers with five recent state-of-art trackers that include STAPLE _CA [20], SAMF_CA [20], SRDCF [19], DSST [17], KCF [15] on the OTB-100 datasets. Figure 5 shows the precision and success plots of proposed tracker and other methods.

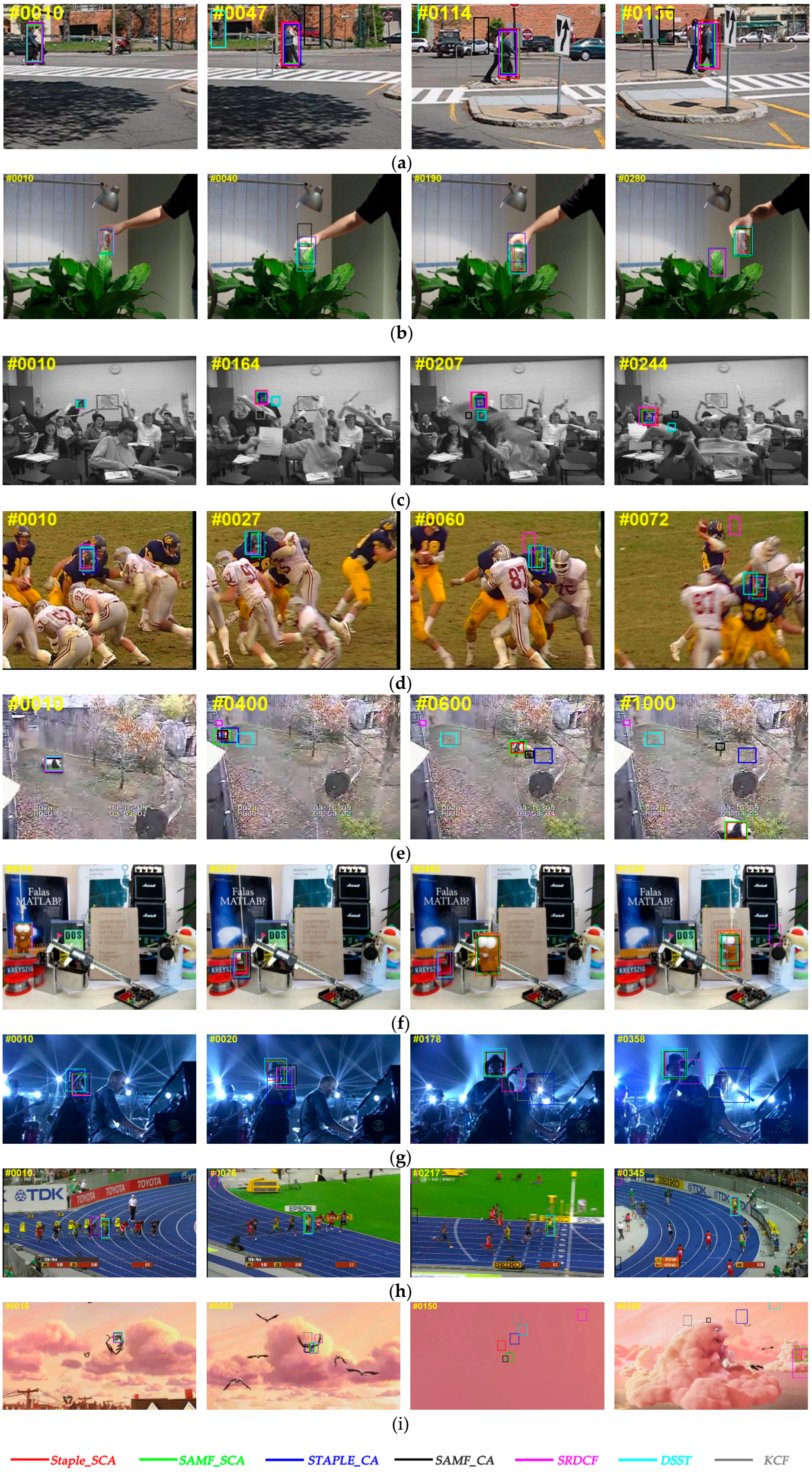


(**a**) (**b**)

Figure 5. Success plots (**a**) and precision plots (**b**) of proposed tracker against five state-of-art trackers [15,17,19,20] on the 100 video sequences in Visual Tracking Benchmark [8].

It is apparent from the success and precision plots of Figure 5 that our proposed trackers have better comprehensive performance than other state of the art trackers. As shown in Figure 5, the success rate and precision rate of the Staple_SCA tracker are 61.6% and 83.7% respectively, which demonstrates the outstanding performance of the proposed method in the visual tracking challenge. Comparing to KCF tracker [15], the proposed tracker gets a 29.1% and 20.4% improvement for the success rate and precision rate, respectively. Furthermore, compared with the conventional Context-aware-based tracker, STAPLE _CA tracker [20], the proposed tracker outperforms it by 3% and 3.3% on success rate and precision rate, respectively.

For qualitative evaluation, we compared our proposed trackers with five state of the art trackers in several challenging situations, such as occlusion, background clutters, fast motion, deformation, motion blur, etc. Figure 6 shows that proposed tracker achieves favorable performance under different challenging scenarios. Specifically, in Panda sequences (show in Figure 6e), our trackers are the only two that have completed the tracking task successfully. However, we also observe that our trackers are unsuccessful in achieving Bird1 sequences, same as other trackers. This failure is caused by the long-term occlusion, trackers are hard to retrieve long missing target in a limited searching window.



(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

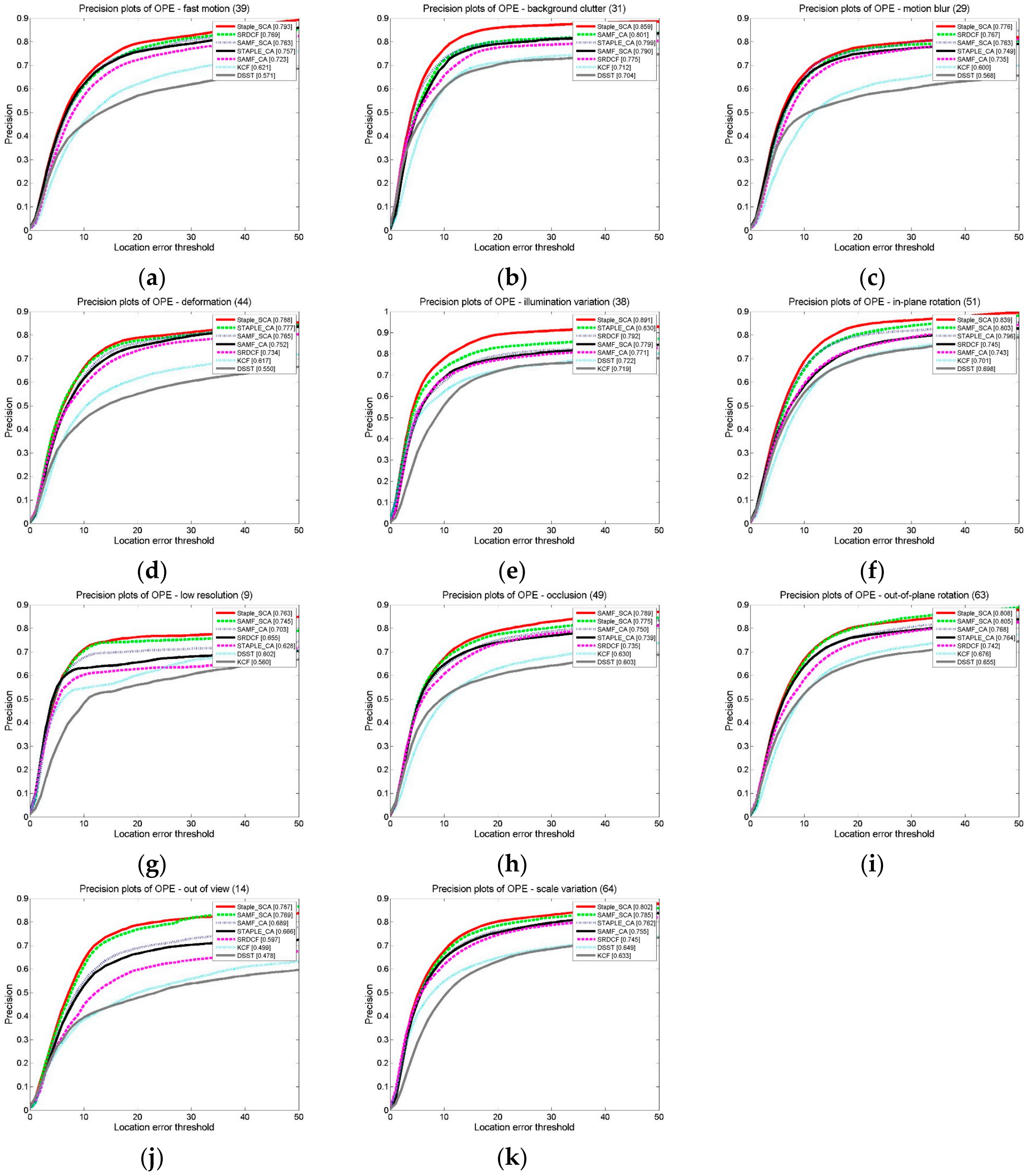
(**g**)

(**h**)

(i)

Figure 6. Comparison of our trackers with other State-of-art trackers [15,17,19,20] in different challenging situations. (**a**) couple; (**b**) coke; (**c**) freeman4; (**d**) football1; (**e**) Panda; (**f**) lemming; (**g**) shaking; (**h**) bolt; (**i**) Bird1.

For attribute-based analysis, we evaluate the overall performance of seven competing trackers under 11 sub-categories tagged in the OTB-100. Figure 7 shows the precision plots of different challenging sub-categories on the OTB-100 datasets for seven competing trackers. Impressively, our proposed Staple_SCA tracker obtains ten the best and one the second places (the first place in this sub-category is our SAMF_SCA tracker.) in 11 sub-categories tasks. In the background clutter attribute (Figure 7b), trackers using context-aware framework outperform other trackers. It demonstrates that incorporating context information into the filter can mitigate the drifting problem led by background clutter, and our proposed context selection strategy is more effective than conventional method. Furthermore, the proposed tracker exceeds other trackers obviously in occlusion task. The result suggests that the proposed structural correlation filter produces the desired effect in the occlusion challenge.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

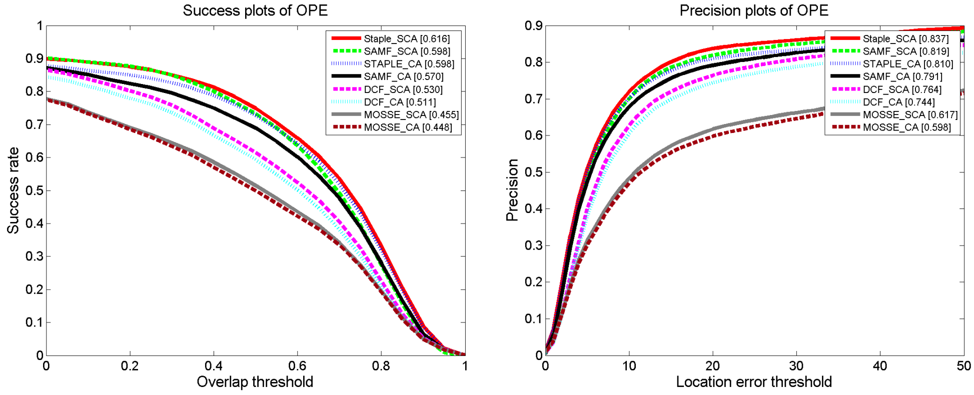
(**g**) (**h**) (**i**)

(**j**) (**k**)

Figure 7. Precision plots over the default 100 video sequence in Visual Tracking Benchmark [8] for seven competing trackers under 11 challenging attributes. (**a**–**k**) indicate the precision plots of different sub-categories, respectively.

Furthermore, in order to compare the effectiveness with the conventional CA method comprehensively, we apply our method to Staple, SAMF, DCF, and MOSSE trackers like [20]. Figure 8 shows the comparing result. Compared to Staple_CA, SAMF_CA, DCF_CA, and MOSSE_CA, our method has improved performance by 3%, 4.9%, 3.7%, and 1.6% on the success plots, respectively. And the precision plots have been improved by 3.3%, 3.5%, 2.7%, and 3.2%, respectively.

Compared to the convention CA method, we divide the target neighborhood evenly into four blocks, then select the maximum response patch from each block, instead of sampling on fixed locations. The more representative negative samples enhanced the discrimination of filter. Moreover, the structure which integrates the holistic and local model, and the updating scheme by using interframe PSR deviation further improved performance of trackers based on our method.



(**a**) (**b**)

Figure 8. Success plots (**a**) and precision plots (**b**) of proposed method against the convention CA method on the 100 video sequences in Visual Tracking Benchmark [8].

4.4. Evaluation on VOT2016

To show the generality of our method, we evaluate the performance of our method on the VOT2016 [10] datasets. VOT2016 includes 60 challenging video sequences, in which background clutters are serious and occlusion is common. The trackers comprehensive performance is evaluated by the expected average overlap (EAO). The EAO combines the raw values of per-frame accuracies and failures in a principled manner and has a clear practical interpretation. We compare our Staple_SCA tracker with five baseline trackers (staple, SRDCF, KCF, SAMF, DSST) on VOT2016. Figure 9 shows the result. Our Staple_SCA tracker with an EAO of 0.298 is superior to other trackers

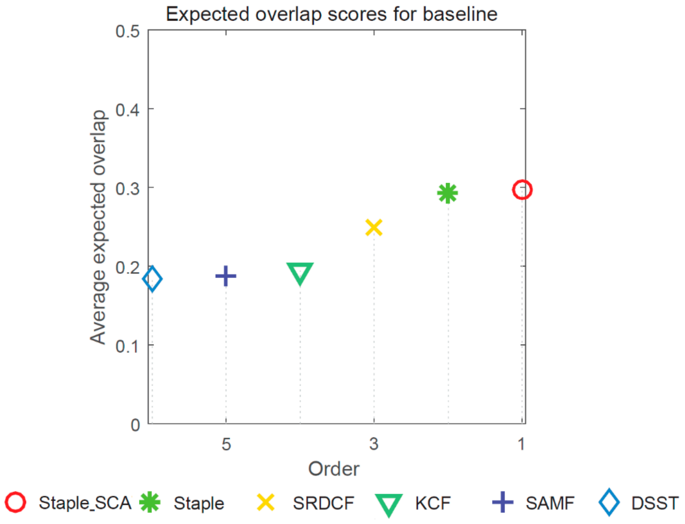


Figure 9. The expected average overlap (EAO) graph with our tracker and five baseline trackers.

5. Conclusions

This paper presents an adaptive context-aware and structural correlation filter for visual tracking from the following three perspectives. Firstly, to collect more informative negative samples to enhance the discrimination of the filter, we proposed a context selecting strategy by selecting the maximum response value context patch in four neighboring regions around the target. Secondly, we construct a structural correlation filter that not only divides the target into two local models to overcome partial occlusion, but also integrates the holistic and local models with adaptive weight to preserve the internal relationship of the target based on the PSR value, and produces the desired effect in the occlusion challenge. Finally, we focus on the degree to which appearance changing is alleviating or deteriorating and exploit the interframe deviation of PSR to update the model, which makes our trackers more robust. Our proposed trackers Staple_SCA and SAMF_SCA run at 24.9 fps and 10.6 fps, which are still within real time range. Extensive experiments have been implemented to demonstrate the validity of our proposed tracker.

Author Contributions: Z.B. designed the main algorithm and the experiments under the supervision of W.T. Z.B. wrote the paper and analyzed the experimental results. W.T. edited the final document. All authors participated in discussions on the paper.

Funding: This research was supported by National Natural Science Foundation of China (No.61071217) and Natural Science Foundation of Jiangsu Province (No.BK20141216).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hare, S.; Golodetz, S.; Saffari, A.; Vineet, V.; Cheng, M.M.; Hicks, S.L.; Torr, P.H.S. Struck: Structured output tracking with kernels. *IEEE Trans. Pattern Anal.* **2016**, *38*, 2096–2109. [CrossRef] [PubMed]
2. Kwon, J.; Lee, K.M. Visual tracking decomposition. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1269–1276.
3. Kwon, J.; Lee, K.M. Tracking by sampling trackers. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1195–1202.
4. Zhong, W.; Lu, H.C.; Yang, M.H. Robust object tracking via sparsity-based collaborative model. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1838–1845.
5. Kalal, Z.; Matas, J.; Mikolajczyk, K. P-n learning: Bootstrapping binary classifiers by structural constraints. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 49–56.
6. Jia, X.; Lu, H.C.; Yang, M.H. Visual tracking via adaptive structural local sparse appearance model. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1822–1829.
7. Wu, Y.; Lim, J.; Yang, M. Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013.
8. Wu, Y.; Lim, J.; Yang, M.H. Object tracking benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1834–1848. [CrossRef] [PubMed]
9. LIRIS, F. The visual object tracking vot2014 challenge results. In Proceedings of the ECCV: European Conference on Computer Vision, Computer Vision—ECCV 2014 Workshops, Zurich, Switzerland, 6–12 September 2014.
10. Kristan, M.; Leonardis, A.; Matas, J.; Felsberg, M.; Pflugfelder, R.; Cehovin, L.; Vojir, T.; Hager, G.; Lukezic, A.; Fernandez, G.; et al. The visual object tracking vot2016 challenge results. In Proceedings of the Computer Vision—ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–10 and 15–16 October 2016; Volume 9914, pp. 777–823.

11. Galoogahi, H.K.; Fagg, A.; Huang, C.; Ramanan, D.; Lucey, S. Need for speed: A benchmark for higher frame rate object tracking. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1134–1143.
12. Milan, A.; Leal-Taixé, L.; Reid, I.; Roth, S.; Schindler, K. MOT16: A benchmark for multi-object tracking. *arXiv*, 2016; arXiv:1603.00831.
13. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.
14. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. *Lect. Notes Comput. Sci.* **2012**, *7575*, 702–715.
15. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596. [CrossRef] [PubMed]
16. Li, Y.; Zhu, J.K. A scale adaptive kernel correlation filter tracker with feature integration. In Proceedings of the European Conference on Computer Vision, Computer Vision—ECCV 2014 Workshops, Zurich, Switzerland, 6–12 September 2014; Volume 8926, pp. 254–265.
17. Danelljan, M.; Häger, G.; Khan, F.; Felsberg, M. Accurate Scale Estimation for Robust Visual Tracking. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014; BMVA Press: London, UK, 2014.
18. Bertinetto, L.; Valmadre, J.; Golodetz, S.; Miksik, O.; Torr, P.H.S. Staple: Complementary learners for real-time tracking. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1401–1409.
19. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Learning spatially regularized correlation filters for visual tracking. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Washington, DC, USA, 7–13 Decmeber 2015; pp. 4310–4318.
20. Mueller, M.; Smith, N.; Ghanem, B. Context-aware correlation filter tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 1387–1395.
21. Galoogahi, H.K.; Fagg, A.; Lucey, S. Learning background-aware correlation filters for visual tracking. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1144–1152.
22. Smeulders, A.W.M.; Chu, D.M.; Cucchiara, R.; Calderara, S.; Dehghan, A.; Shah, M. Visual tracking: An experimental survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1442–1468. [PubMed]
23. Yang, H.X.; Shao, L.; Zheng, F.; Wang, L.; Song, Z. Recent advances and trends in visual tracking: A review. *Neurocomputing* **2011**, *74*, 3823–3831. [CrossRef]
24. Zhou, S.H.K.; Chellappa, R.; Moghaddam, B. Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Trans. Image Process.* **2004**, *13*, 1491–1506. [CrossRef] [PubMed]
25. Lee, K.C.; Kriegman, D. Online learning of probabilistic appearance manifolds for video-based recognition and tracking. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 852–859.
26. Ross, D.A.; Lim, J.; Lin, R.S.; Yang, M.H. Incremental learning for robust visual tracking. *Int. J. Comput. Vis.* **2008**, *77*, 125–141. [CrossRef]
27. Li, X.; Hu, W.M.; Zhang, Z.F.; Zhang, X.Q.; Luo, G. Robust visual tracking based on incremental tensor subspace learning. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio De Janeiro, Brazil, 14–21 October 2007; Volume 1–6.
28. Wen, J.; Gao, X.B.; Li, X.L.; Tao, D.C. Incremental learning of weighted tensor subspace for visual tracking. In Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 11–14 October 2009.
29. Zhang, K.H.; Zhang, L.; Yang, M.H. Real-time compressive tracking. In Proceedings of the Computer Vision—ECCV 2012, 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Volume 7574, pp. 864–877.
30. Dinh, T.B.; Vo, N.; Medioni, G. Context tracker: Exploring supporters and distracters in unconstrained environments. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 1177–1184.

31. Khattak, A.S.; Raja, G.; Anjum, N. Adaptive framework for multi-feature hybrid object tracking. *Appl. Sci.* **2018**, *8*, 2294. [CrossRef]
32. Chen, Z.; Hong, Z.; Tao, D. An experimental survey on correlation filter-based tracking. *arXiv* **2015**, arXiv:1509.05520.
33. Danelljan, M.; Khan, F.S.; Felsberg, M.; van de Weijer, J. Adaptive color attributes for real-time visual tracking. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24–27 June 2014; pp. 1090–1097.
34. Zhang, B.C.; Li, Z.G.; Cao, X.B.; Ye, Q.X.; Chen, C.; Shen, L.L.; Perina, A.; Ji, R.R. Output constraint transfer for kernelized correlation filter in tracking. *IEEE Trans. Syst Man Cybern. Syst.* **2017**, *47*, 693–703. [CrossRef]
35. Li, C.B.; Yang, B. Robust scale adaptive visual tracking with correlation filters. *Appl. Sci.* **2018**, *8*, 2037. [CrossRef]
36. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Convolutional features for correlation filter based visual tracking. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Washington, DC, USA, 7–13 Decmeber 2015; pp. 621–629.
37. Danelljan, M.; Robinson, A.; Khan, F.S.; Felsberg, M. Beyond correlation filters: Learning continuous convolution operators for visual tracking. *Lect. Notes Comput. Sci.* **2016**, *9909*, 472–488.
38. Xiao, J.J.; Qiao, L.B.; Stolkin, R.; Leonardis, A. Distractor-supported single target tracking in extremely cluttered scenes. In Proceedings of the Computer Vision—ECCV 2016, 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9908, pp. 121–136.
39. Qie, C.G.; Guo, G.J.; Yan, Y.; Zhang, L.M.; Wang, H.Z. Improved correlation filter tracking with hard negative mining. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 1646–1651.
40. Jeong, S.; Paik, J. Partial block scheme and adaptive update model for kernelized correlation filters-based object tracking. *Appl. Sci.* **2018**, *8*, 1349. [CrossRef]
41. Sun, X.; Cheung, N.M.; Yao, H.X.; Guo, Y.L. Non-rigid object tracking via deformable patches using shape-preserved kcf and level sets. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5496–5504.
42. Liu, S.; Zhang, T.Z.; Cao, X.C.; Xu, C.S. Structural correlation filter for robust visual tracking. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 4312–4320.
43. Akin, O.; Erdem, E.; Erdem, A.; Mikolajczyk, K. Deformable part-based tracking by coupled global and local correlation filters. *J. Vis. Commun. Image Represent.* **2016**, *38*, 763–774. [CrossRef]
44. Fan, H.; Xiang, J. Robust visual tracking via local-global correlation filter. In Proceedings of the 2017 AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4025–4031.
45. Jeong, S.W.; Kim, G.S.; Lee, S.K. Effective visual tracking using multi-block and scale space based on kernelized correlation filters. *Sensors* **2017**, *17*, 433. [CrossRef] [PubMed]
46. Ma, C.; Yang, X.K.; Zhang, C.Y.Y.; Yang, M.H. Long-term correlation tracking. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5388–5396.
47. Gray, R.M. Toeplitz and circulant matrices: A review. *Found. Trends®Commun. Inf. Theory* **2006**, *2*, 155–239. [CrossRef]
48. Davis, P.J. *Circulant Matrices*; American Mathematical Society: Providence, RI, USA, 2012.
49. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.