Low-Rank Multi-Channel Features for Robust Visual Object Tracking

Fawad [1,*], **Muhammad Jamil Khan** [1], **MuhibUr Rahman** [2,*], **Yasar Amin** [1] **and Hannu Tenhunen** [3,4]

[1] ACTSENA Research Group, Telecommunication Engineering Department, University of Engineering and Technology Taxila, Punjab 47050, Pakistan
[2] Department of Electrical Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada
[3] Department of Electronic Systems, Royal Institute of Technology (KTH), Isafjordsgatan 26, SE 16440 Stockholm, Sweden
[4] Department of Information Technology, TUCS, University of Turku, 20520 Turku, Finland
* Correspondence: engr.fawad@students.uettaxila.edu.pk (F.); muhibur.rahman@polymtl.ca (M.R.)

check for updates

Abstract: Kernel correlation filters (KCF) demonstrate significant potential in visual object tracking by employing robust descriptors. Proper selection of color and texture features can provide robustness against appearance variations. However, the use of multiple descriptors would lead to a considerable feature dimension. In this paper, we propose a novel low-rank descriptor, that provides better precision and success rate in comparison to state-of-the-art trackers. We accomplished this by concatenating the magnitude component of the Overlapped Multi-oriented Tri-scale Local Binary Pattern (OMTLBP), Robustness-Driven Hybrid Descriptor (RDHD), Histogram of Oriented Gradients (HoG), and Color Naming (CN) features. We reduced the rank of our proposed multi-channel feature to diminish the computational complexity. We formulated the Support Vector Machine (SVM) model by utilizing the circulant matrix of our proposed feature vector in the kernel correlation filter. The use of discrete Fourier transform in the iterative learning of SVM reduced the computational complexity of our proposed visual tracking algorithm. Extensive experimental results on Visual Tracker Benchmark dataset show better accuracy in comparison to other state-of-the-art trackers.

Keywords: circulant matrix; texture; tracking; convolution; filter

1. Introduction

Visual tracking is the process of Spatio-temporal localization of a moving object in the camera scene. Object localization has potential applications, including human activity recognition [1], vehicle navigation [2], surveillance and security [3], and human-machine interaction [4]. The researchers are developing robust trackers to reduce the computational cost of the visual object tracking algorithm. Several reviews on robust tracking techniques have been published [5–7]. The results discussed in this paper show that the tracking is affected by the geometric and photometric variations in the object appearance. Visual tracking is desired to be robust against intrinsic variations (e.g., pose, shape deformation, and scale) and extrinsic variations (e.g., background clutter, occlusion, and illumination) [8,9]. Significant efforts have been made to extract invariant features through handcrafted and deep learning methods. The deep learning approaches have achieved a higher accuracy; however, a massive training data is required, which is not often available in many surveillance applications. The handcrafted method, on the other hand, only requires a careful selection of discriminative features in the object appearance.

Tracking techniques reported in the literature can generally be classified into three groups: generative, discriminative, and filtered based tracker. The generative models [10] identify the target

among many sampled candidate region through similarity function. The test instance is decided as the target location when it has the highest similarity to the appearance model among the sampled candidate regions. The generative tracker, therefore, undergoes through a high computational overhead. The discriminative trackers [11] use the target samples to learn a classifier that can differentiate the target from its background. A discriminative tracker largely depends on the positive and negative samples to update the classifier during tracking. The large sample set can make the classifier more robust; however, it is not available due to time-sensitivity. During the last decade, considerable research on correlation filter-based (CFB) trackers [12] is performed. CFB frameworks brought various improvements in the visual tracking process. The correlation filters include the circulant matrix and fast Fourier transform (FFT), which can employ an extensive sample set to train the classifier. The weighted mask of the kernelized correlation filter (KCF) [13] makes the tracker more robust to the variations in the visual scene. Despite the robustness, KCF continuously requires to update the learned kernel along with the changes in the target appearance. However, such a model update mechanism is sensitive to occlusions. The correlation filter-based tracker performance depends on the quality of the features. Moreover, large dimensionality of the feature vector is also a barrier for the tracker to participate in the real-time scenarios.

In this work, we identified a feature set, that improves the tracking performance even in the presence of intrinsic and extrinsic variations. We have fused multiple handcrafted feature channels to get a response map of the target object's position in each new frame of the video. The dimensionality of the fused feature vector is reduced by selecting ten high entropy variables from the Principal Component Analysis (PCA) output. Instead of the euclidean distance of the feature variables, Pearson's coefficients of skewness is used as PCA input, as shown in Figure 1. Moreover, we have used the circulant matrix along with fast Fourier transform in the kernel correlation filter to reduce the computation complexity of our tracker. Extensive experiments performed over benchmark dataset reveal that the proposed descriptor provides considerable improvement in precision and success rate.
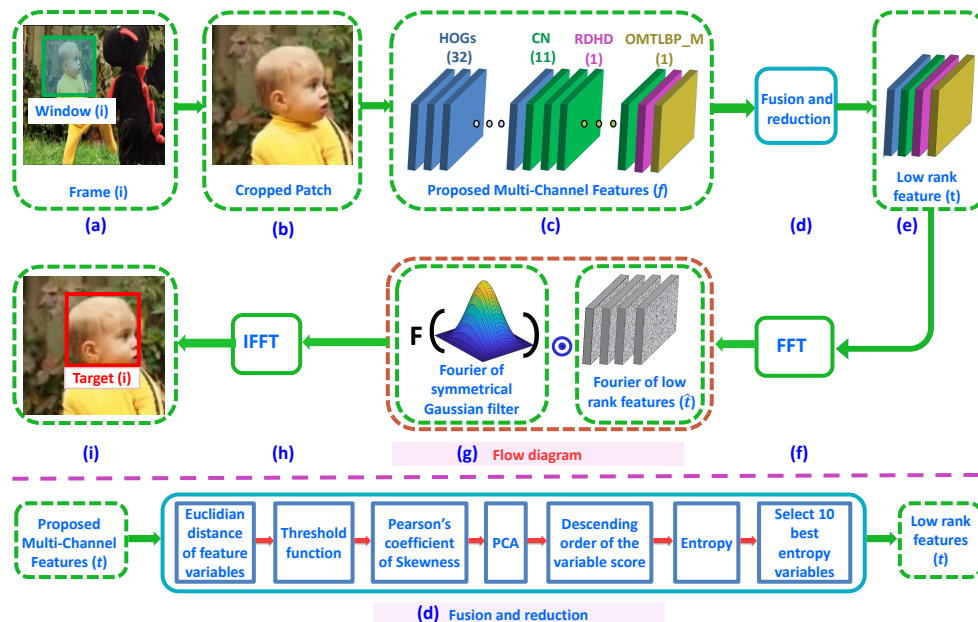

Figure 1. Framework of the proposed tracker.

The benchmark dataset [8] is a useful tool to evaluate the performance of visual trackers. The dataset provides the ground truth position of the target in each frame. The dataset set contains 100 sequences labeled with 11 different attributes. Each frame sequence is manually specified with multiple challenges. The overlap ratio of estimated and the ground truth bounding box describes the accuracy of the trackers.

The remaining discussion of the paper is organized as follows. Section 2 discusses the related work from the literature. Section 3 contains the detail about the proposed tracker methodology. Section 4 presents the results in the shape of graphs related to precision and success rate, bar plots, and comparison table. The conclusion of the paper is drawn in Section 5.

2. Related Work

Visual object tracking has been extensively studied and discussed in [14,15]. The visual trackers can be grouped into single vs. multi-object trackers, context-aware vs. unaware, and generative vs. discriminative trackers. Single object trackers [16] can only track a single target at a time, while multi-object trackers [17] can monitor more than one object at the same instant. The discriminative models in [18,19], employ the handcrafted features to train the ensemble classifier. In [20,21], tracking by detection through deep learning is studied. To deal with appearance variability, the discriminative methods, update their classifier at each location of the candidate, which results in a massive computational cost. In [22,23] sparse representation and metric learning [10] are used to build the generative model-based trackers. The generative model is updated at each location of the candidate to avoid the tracking drift. Similar to the discriminative model, the generative model also suffers from a huge computational load.

Sparsity-based Collaborative Model (SCM) [24] and Adaptive Structural Local-sparse Appearance model (ASLA) [25] are proposed to deal with extreme variations in the target appearance. SCM and ASLA both suffer from a significant scale drift, whenever the target has rotating motions and fast scale changes. The MUlti-Store Tracker (MUSTer) [26], and Multiple Experts Using Entropy Minimization (MEEM) [27] employ ensemble-based methods to solve the drift problem in online tracking. However, MEEM fails whenever identical objects appear in the visual scene.

The correlation filter-based tracker [28] attempts to minimize the sum of squared error between the actual and desired correlation response. In [29,30] the use of Fast Fourier Transform has reduced the correlation cost. The work in [28] has introduced, multi-channel feature map by employing color and texture descriptor combination. Kernelized correlation filters were developed in [31–33]. The synthetic training samples generated by the circular shift operation creates a boundary effect. The training of the kernelized correlation filters is severely affected by such synthetic data. In [34,35], spatial regularization is applied over the correlation filter to remove the boundary effect. The algorithm developed in [36] overcomes the boundary effect by simultaneous learning of the correlation filter and its desired response map. Discriminative Correlation Filter (DCF) [37], trackers employ Fourier transforms to efficiently learn a classifier on all patches of the target neighborhood. However, the discriminative correlation filter based tracker also suffers from the spatial boundary effect. The Spatially Regularized Discriminative Correlation Filters (SRDCF) [35] utilize the spatial regularization to vanish the boundary effect of the DCF. The SRDCF fails when the target object is hollow at the center; the filter then considers the background pixels as a target, that leads to the drift problem. The Channel and Spatial Reliability Discriminative Correlation Filters (CSRDCF) [38] includes a color histogram-based segmentation mask, which avoids the background pixels. The Discriminative Scale Space Tracking (DSST) [39] enhances the tracking speed, but the performance is inferior in comparison to CSRDCF. In SRDCFdecon [40] adaptive decontamination is used which adaptively learns the reliability of each training sample and eliminates the influence of contaminated ones. The minimum barrier distance (MBD) [41] is developed to mitigate the impact of background on the tracker accuracy. The MBD consider the dissimilarity value to weight the extracted feature at each target position. The MBD based tracker can precisely locate the target an all attributes of OTB database, but it fails at low resolution, and clutter background.

Correlation filters utilize multiple different combinations of color and texture features, extracted from the patches in the search window. The multi-channel HoG feature [42] integrated with color naming features [43] provide the basis for kernelized correlation filters based tracking methods cite19. The handcrafted features have produced excellent results on the existing benchmark datasets; However, they provide poor performance when there is a rapid variation in the object appearance.

The Color Histograms (CHs) based handcrafted features are robust to the fast motion of the object, but they result in a poor performance in the presence of illumination variations, and background clutters [44]. A robust descriptor can significantly improve the performance of visual tracking. In [45], the concentration of the feature map has shown favorable performance even in the presence of target object state and color variations. Recently, SVM-based support correlation filters (SCFs) [43] have increased efficiency by utilizing the circulant matrix. Moreover, multi-channel SCF (MSCF), kernelized SCF (KSCF), and scale-adaptive KSCF (SKSCF) further improves the accuracy and efficiency of the trackers.

3. Proposed Method

The proposed tracker incorporates the discriminant color and texture features to achieve a better precision and success rate. A novel fusion and reduction approach is employed to reduce the computational cost. The details of the proposed multi-channel feature and dimensionality reduction are discussed as follows.

Multi-Channel Feature

The image target patch in Figure 1b is described using multi-channel features of the proposed visual tracking approach. A combination of a total of 45 channels consisting of HoG, Color Naming, RDHD [46,47], and the magnitude component of Overlapped Multi-oriented Tri-scale Local Binary Pattern (OMTLBP) [48] is used to describe the object with better accuracy. The Felzenszwalb's HoG (FHoG) feature [42] vector is extracted from the patch of each input frame shown in Figure 1a. Each patch is divided into a grid, with each grid cell of size 4×4. The 32-dimensional feature vector represents each cell in the grid. Total nine orientations described through 27 variables are used in combination with four texture and one truncation variable to describe the grid cell. Let $f_1 = \{w/w \in \mathbb{R}^{32}\}$ is the FHoG feature vector for each grid cell. The color names (CN) [43] feature with vocabulary size 11 is extracted from each grid cell of the input patch. Suppose $f_2 = \{x/x \in \mathbb{R}^{11}\}$ is the feature vector representing the colour naming shown in the Figure 1c.

The RDHD [47] enhances the discriminative capability of the proposed multi-channel feature. The extrema responses of both first and second-order symmetrical Gaussian derivative filters are quantized to obtain the robust features. The symmetrical Gaussian derivative filters applied on the target patch S, which is presented through the following mathematical equations.

$$H_\theta^{2max} = \frac{1}{2}\left(S_{xx} + S_{yy} + \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}\right). \tag{1}$$

$$H_\theta^{2min} = \frac{1}{2}\left(S_{xx} + S_{yy} - \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}\right). \tag{2}$$

$$D = H_\theta^{2max} - H_\theta^{2min} = \sqrt{(S_{xx} + S_{yy})^2 + 4S_{xy}^2}. \tag{3}$$

$$G(x, y) = \frac{1}{2\pi\sigma} e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)} \tag{4}$$

where $S_x = G_x * S$, $S_y = G_y * S$, $S_{xx} = G_{xx} * S$, $S_{yy} = G_{yy} * S$ and $S_{xy} = G_{xy} * S$. while G_{xx}, G_{yy} and G_{xy} represent the second order partial derivatives of the symmetrical Gaussian function with respect to x, y, and (x, y). The symmetrical Gaussian function is shown in Equation (4). The channel D of RDHD obtained through Equation (3) is used as a multi-channel feature in the proposed descriptor. The $f_3 = \{w/w \in \mathbb{R}^1\}$ represent the noise robust RDHD feature.

The rotation invariant magnitude component of Overlapped Multi-oriented Tri-scale Local Binary Pattern (OMTLBP_$M_{P,R}^{riu2}$) [48] is also included in the multi-channel feature to increase the robustness of the visual object descriptor. The total number of sample points $P = 8$ and radius $R = \{1, 2, 3\}$ is used to extract the OMTLBP_$M_{P,R}^{riu2}$ as shown in Equation (5). The $\chi(u, v)$ tends to be 1, when u is greater

than v, otherwise 0. In Equations (5) and (6) the μ_m is the mean value of the kth magnitude component $m_{r,c}^k$ extracted from the segment (r, c) of the input image. We assume $f_4 = \{z/z \in \mathbb{R}^1\}$ represent the OMTLBP_M$_{P,R}^{riu2}$ feature vectors. The Multi-channel feature 'R' described in Equation (7) is obtained by the concatenation of FHoG, CN, RDHD [47], and OMTLBP_M$_{P,R}^{riu2}$.

$$OMTLBP\_M_{P,R}^{riu2} = \begin{cases} \chi(m_{r,c}^k, \mu_m), & \text{if} \quad M\_U_{P,R} \leq 2 \\ P+1, & \text{otherwise} \end{cases} \tag{5}$$

$$M\_U_{P,R} = |\chi(m_{r,c}^{P-1}, \mu_m) - \chi(m_{r,c}^0, \mu_m)| + \sum_{k=1}^{P-1} |\chi(m_{r,c}^k, \mu_m) - \chi(m_{r,c}^{k-1}, \mu_m)| \tag{6}$$

$$R = HoG||CN||RDHD\_D||OMTLBP\_M_{P,R}^{riu2} \tag{7}$$

where the symbol $||$ represent the concatenation operation. R denote the final multi-channel feature. The dimensionality reduction operation is performed over the multi-channel feature to reduce the complexity of the proposed tracker.

Fusion and Reduction

Recently, fusion, and reduction [49–51] approach shown in the Figure 1d has been developed, to reduce the dimensionality of the feature vector. The robust variables are selected, among the fused features set to increase the recognition and decrease the computation time. The feature vector R in Equation (7) is the fused vector. After fusion, robust variables are selected based on the entropy value of their coefficient of skewness. Let r_i be the variables in the fused feature vector R.

$$\psi = \left\{ \sqrt{\sum_{i=1}^{45} (r_i - r_{i+1})^2} \right\}_{j=1}^{M} \tag{8}$$

$$\tau = h(\psi) \tag{9}$$

where M represents the total number of grid cells in the frame. The ψ denote the Euclidean distance between the variables of the fused feature. The h is a thresholding function, that is defined as:

$$h = \begin{cases} R_m, & \text{if} \quad \psi \leq 0.4 \\ 0, & \text{if} \quad \psi > 0.4 \end{cases} \tag{10}$$

$$\gamma = \frac{R_m - \tilde{x}}{\sqrt{\sigma^2}} \tag{11}$$

where h is the threshold function, which is derived for selection of minimum distance features and R_m denotes the minimum distance feature vector. R_m is used to compute skewness of values. We have selected the threshold parameter as 0.4 in Equation (10). We tested the performance of our tracker with various values of threshold parameter used in Equation (10) of the manuscript. Through experimentation, we found that the threshold parameter 0.4 in Equation (10), provides the highest joint maxima for both AUC and precision when validated on OTB-50 and OTB-2013 dataset. The symbol σ^2 and $\tilde{x}$ denote the variance and median values of R_m. Then the technique of principal component analysis is applied on γ to calculate the score of each feature. Finally, the entropy values associated with each variable of the feature are sorted into descending order. Ten highest score variable is selected as a feature vector. The output low-rank representation is denoted by t, as shown in Figure 1e.

$$y_{m,n} = exp(-\frac{m^2 + n^2}{2\sigma^2}) \tag{12}$$

The $t \in \mathbb{R}^{M \times N \times D}$ denote extracted multi-channel feature map of the input template training patch. The size of each channel is $M \times N$ variables, with total number of $D = 45$ channels in the feature map. The $y \in \mathbb{R}^{M \times N}$ is the symmetrical Gaussian label of dimension equal to the feature map, and $\sigma = 0.02\sqrt{M \times N}$. The proposed method include a symmetrical Gaussian filter label to locate the position of the target in the subsequent frame of the video. The Gaussian filter label used is symmetrical shape similar to maxican hat [52,53] shown in the Figure 1f-g. Let $\hat{t} = \mathcal{F}(t)$ and $\hat{y} = \mathcal{F}(y)$, where $\mathcal{F}(.)$ represent the Discrete Fourier Transformation of the multi-channel features.

The training process identifies the function $f(t) = w^T t$, that can reduce the lease square error of the Ridge Regression, shown in Equation (8). The aim is to learn the support correlation filter w and the regularization parameter λ that reduces the least square error in the Ridge Regression. Equation (13) gives the close from of the minimization.

$$\min_{w} = \sum_{m,n}(f(m,n) - y(m,n))^2 + \lambda\|w\| \tag{13}$$

$$\hat{w} = \frac{\hat{y}}{T^H.T + \lambda} \odot T^H \tag{14}$$

where T is the circulant matrix of the multi-channel feature vector t. The T^H represent the hermitian transpose, i.e., $T^H = (T^*)'$, and T^* is the complex-conjugate of T. For any *ith* input frame of the video, the position of the target object is the location of the maximum response value in R_i, described in Equation (15).

$$(R_i) = \mathcal{F}^{-1}(\hat{z}_i \odot w_i) \tag{15}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform shown in the Figure 1h. The $\hat{z}_i = \mathcal{F}(z)$ is the Fourier domain of the test multi-channel feature z. The symbol $\odot$ represent the element-wise product of the matrices.

$$(x_i, y_i) = \underset{m,n}{argmax} R_i(x_m, y_n) \tag{16}$$

where x_i and y_i represent the position of the target in the *ith* frame of the video. The target position in the *ith* frame is found at the largest response R. The bounding box of the target size around x_i and y_i is used to label the target in the *ith* frame of the video as shown in Figure 1i.

4. Experimental Setup

4.1. Dataset and Evaluated Trackers

The performance of the proposed multi-channel feature is evaluated on Tracker Benchmark (TB) dataset [8]. The TB contains a total of 100 images of annotated sequences with eleven different attributes. The dataset includes variations in illumination, scale, and resolution. The objects in the dataset suffer from deformation, occlusion, and background clutter. The video sequence is captured in the in-plane and out-of-plane rotation. The blurriness due to fast motion and out of view object is captured in the sequence. The TB is classified into three different test suits, namely OTB2013, OTB50, and OTB2015. The information about the target object true bounding box is included with the first frame of each sequence. The ground truth positions of the subject in each frame are provided with the dataset to calculate the accuracy.

For comprehensive comparisons, we employed the proposed descriptor in KSCF and SKSCF methods. The SKSCF methods with linear, polynomial, and Gaussian kernel is studied in this work. The KSCF and SKSCF of Gaussian kernel are evaluated on multi-channel HoG and Color naming features as shown in Table 1. The results in Table 1 show that the tracker SKSCF along with Gaussian kernel provide 87.04% DP and 62.30% AUC, which is better in comparison to linear and polynomial kernel type. The results of the proposed tracker with Gaussian kernal is also compared with other correlation filters based trackers (e.g., HDT [54], MEEM [27], Struck [55], KCF [32], BIT [56],

CSRDCF [38], CT [57], LCT [58], MIL [18], ASLA [25], L1APG [59], TLD [60], CXT [61], SCM [24], Staple [62], and SRDCF [35]).

Table 1. Performance comparison of various correlation filter and the kernel combination for the multi-channel HoG and Color naming features.

|	**Kernels**	**Linear**	**Polynomial**	**Gaussian**
	Mean DP (%)	83.89	85.91	**87.04**
SKSCF	Mean AUC (%)	59.33	60.82	**62.30**
	Mean FPS (%)	14	11	8
	Mean DP (%)	82.00	84.20	**85.00**
KSCF	Mean AUC (%)	56.20	57.10	**57.50**
	Mean FPS (%)	**94**	55	35

4.2. Evaluation Procedure

The one-pass evaluation (OPE) plots [8] are used to evaluate and compare the performance of the proposed descriptor with other methods. The OPE consist of precision and success plots used for the evaluation of our proposed tracker.

The precision vs. location error threshold plot evaluates the tracking precision. The precision illustrates the number of estimated target positions which lie within the defined threshold distance from the ground truth. The location error is the average Euclidean distance between the object's estimated position and the annotated ground truth position. The location error describes the gap in term of pixels only and does not care for the size and scale of the target object. The threshold error ranges between 0 and 50 pixels is used to calculate the average precision of each sequence. The distance precision (DP) is at location error of 20 pixels.

The success rate vs. overlap threshold is also used in parallel to evaluate our tracker. The success rate is the percentage of frames with an overlap ratio of γ greater than a given threshold of τ. Where γ is the overlap ratio between estimated B_{est} and ground truth B_{gth} bounding box, i.e., $\frac{B_{est} \cap B_{gth}}{B_{est} \cup B_{gth}} > \tau$. The range of threshold τ is between 0 and 1. The trapezoidal integration is used to calculate the Area Under the Curve (AUC) of the success plot. The precision and success plot shown in Figure 2 show a comparison of the proposed tracker with other state-of-the-art trackers. The DP and AUC in Figure 3 describe the results more conveniently.

Figure 2. *Cont.*
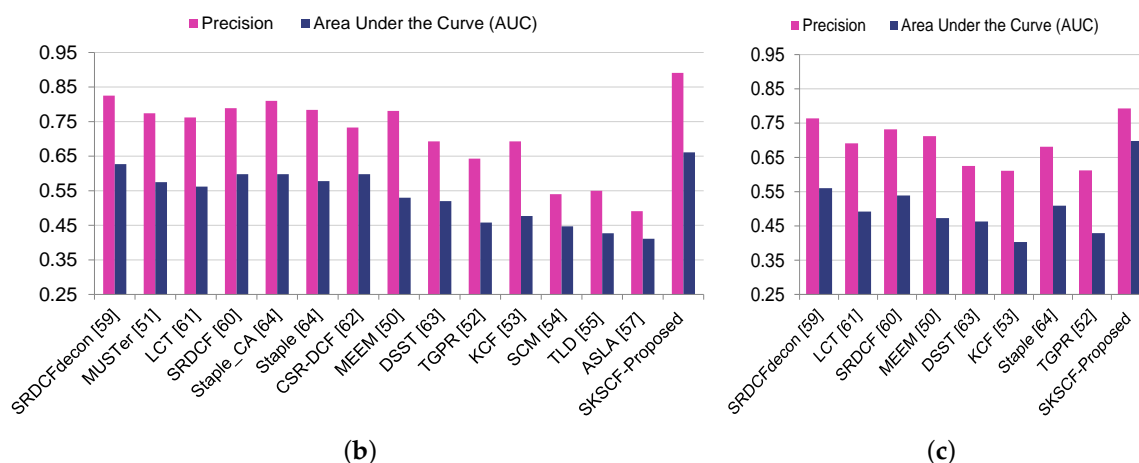
(**c**)

(**d**)

(**e**)

(**f**)
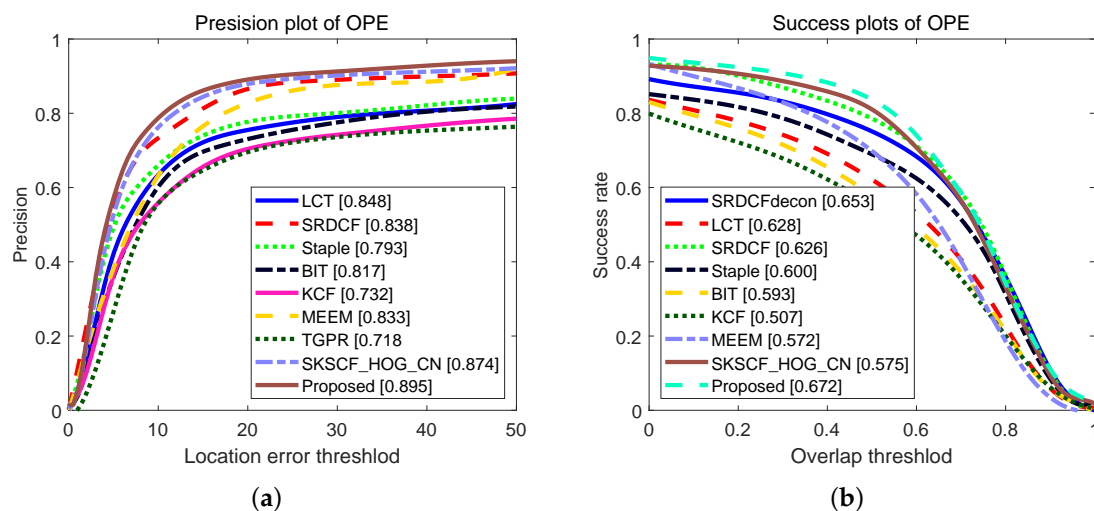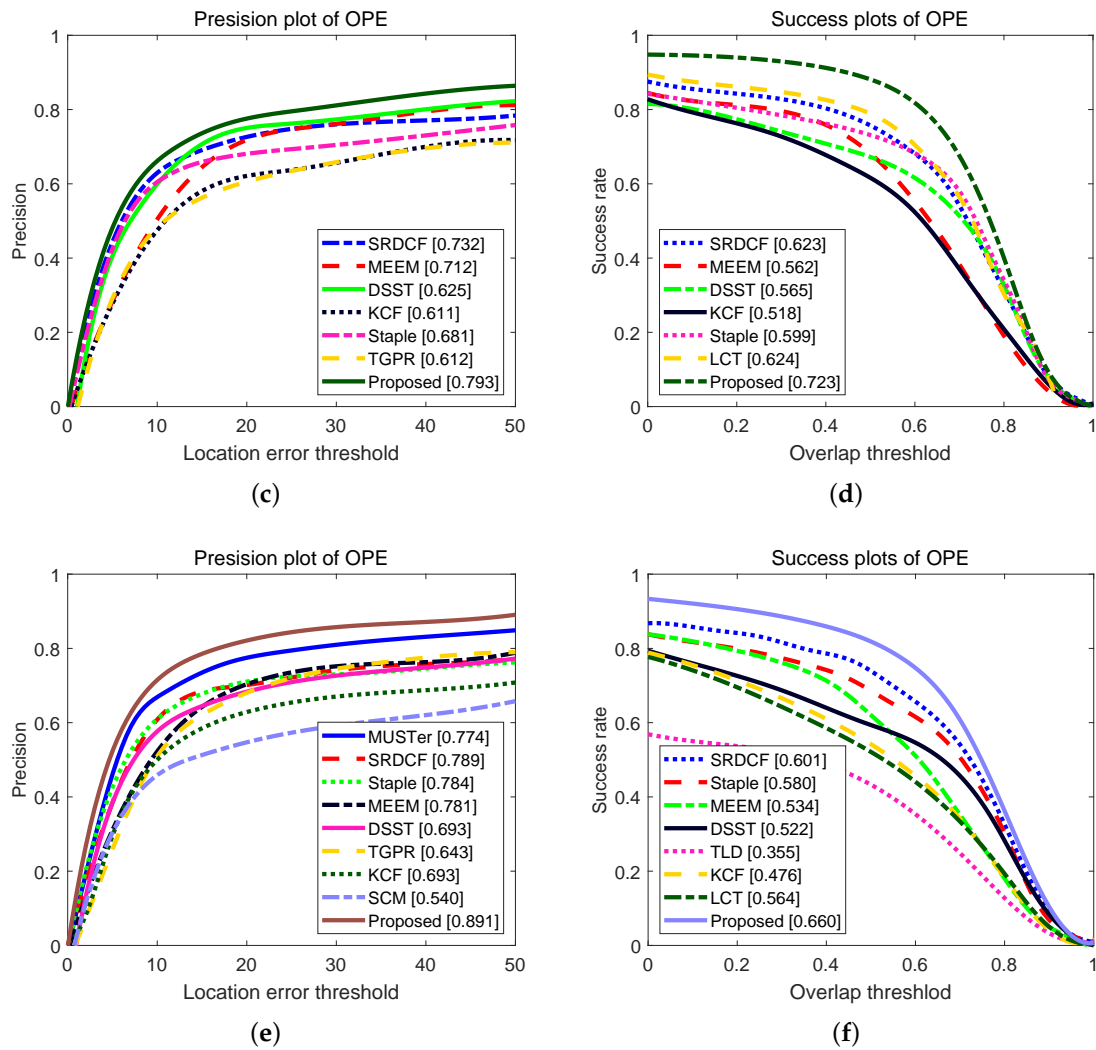
Figure 2. Comparison of the various tracker performance on OTB2013, OTB50, and OTB2015. (**a**) precision-OTB2013; (**b**) AUC-CVPR2013; (**c**) precision-OTB50; (**d**) AUC-OTB50; (**e**) precision-OTB2015; (**f**) AUC-OTB2015.

(**a**)
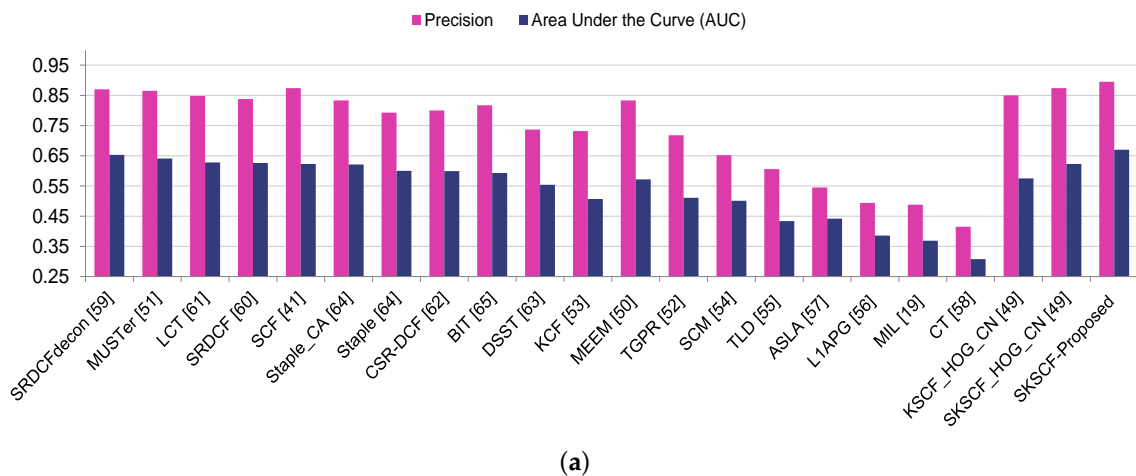
Figure 3. *Cont.*

Figure 3. The precision and area under the curve for various trackers on OTB test suits. (**a**) OTB2013; (**b**) OTB2015; (**c**) OTB50.

4.3. Parameter Setting

The experimental results are obtained on the desktop computer with an Intel 2 core 2.2 GHz CPU and 8 GB RAM and Nvidia GTX 750 Ti GPU. The parameter scales ψ and scales factor γ of SKSCF are 21 and 1.04, respectively. The standard deviation σ of the kernel is 0.5 for SKSCF. The optimal setting for the upper and lower threshold (θ_l θ_u) are (0.3, 0.6). The cell size and orientation of the HoG feature are set to 4 and 9, respectively. The OMTLBP is defined for 8 sample point with scale values ranges from 1 to 3.

4.4. Overall Performance

Table 1, summarizes the results of SKSCF method for the different type of kernels on 50 challenging image sequences of TB-100 dataset [8]. Table 1 shows that SKSCF with Gaussian kernel outperforms the polynomial and linear in mean DP and AUC. However, the linear kernel can process more frames per second (FPS) in comparison to Gaussian and polynomial. In this work, we have selected the Gaussian kernel with different variants of correlation filters. Eleven visual attributes classify the annotated sequences of the benchmark dataset [8]. The 11 characteristics of the sequences includes background clutter (BC), illumination variation (IV), scale variation (SV), deformation (DEF), fast motion (FM), in-plane rotation (IPR), low resolution (LR), motion blur (MB), occlusion (OCC), out-of-view (OV), and out-of-plane rotation (OPR).

The proposed tracker evaluated on three benchmark test suits of OTB dataset. The OTB-2015 test suit is used to evaluate the proposed tracker for dimensionality reduction using the Gaussian kernel. In the absence of dimensionality reduction, the SKSCF-proposed gives 89.1%, and 66.1% mean DP and AUC respectively at speed seven frames per second speed. The same SKSCF-proposed when tested in the presence of the dimensionality reduction, it provided 88.87% average DP and 65.6% average AUC at rate eight frames per second speed. By employing the dimensionality reduction, the mean DP by 0.04% while the AUC reduced by 0.05% at the cost of one frame per second increase in the speed.

The precision and success rate mentioned in Tables 2 and 3 demonstrate that the proposed tracker leads other methods both in precision and success rate. Table 2 compare the performance of our proposed tracker with other recently developed methods on OTB-2013 test suits. Table 2 shows that our proposed tracker has a better precision on all attributes except in the case of BC, SV, and IPR. The precision of the proposed tracker is 1.9%, 3.3%, 0.6%, 1.2%, 0.2%, 1.8%, 0.8%, and 1.5% higher than the other trackers for FM, MB, DEF, IV, LR, OCC, OPR, and OV, respectively. The success rate is lower only for the case of MB, IPR, and SV. The success rate is 0.5%, 0.1%, 0.7%, 1.6%, 3.9%, 0.6%, 0.5%, and 0.5% higher that of other trackers for FM, BC, DEF, IV, LR, OCC, OPR, and OV, respectively.

Table 3 demonstrates a performance comparison on OTB-2015 test suit. Table 3 shows that our proposed tracker have 4.5%, 2%, 1.1%, 2.8%, 2.6%, 1.5%, and 1.2% higher precision that other tracker on FM, MB, DEF, IV, LR, OCC, and OPR attributes. The precision of the proposed tracker is 0.6%, 3.3%, 1.3%, and 1.5% lower than the best trackers for BC, IPR, SV, and OV respectively. The success rate of the proposed tracker is 0.6%, 0.4%, 0.1%, 0.3%, 0.1, 2.2%, and 5.3% higher for BC, MB, DEF, IV, LR, OCC, and OV respectively. The success rate is lower for the case of FM, IPR, OPR, and SV.

Table 2. SKSCF-Proposed descriptor comparison with other trackers on OTB-2013 [8] test suit.

	Attributes	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
Precision	SKSCF-Proposed	**0.798**	0.829	**0.835**	**0.899**	**0.853**	0.811	**0.598**	**0.890**	**0.865**	**0.815**	0.800
	SKSCF_HoG_CN [13]	0.779	**0.859**	0.802	0.893	0.841	0.810	0.596	0.872	0.857	0.800	**0.809**
	KSCF_HoG_CN [13]	0.680	0.825	0.761	0.854	0.805	**0.816**	0.555	0.852	0.836	0.697	0.768
	MEEM [27]	0.745	0.802	0.721	0.856	0.771	0.796	0.529	0.801	0.840	0.726	0.795
	TGPR [63]	0.579	0.763	0.570	0.760	0.695	0.683	0.567	0.668	0.693	0.535	0.637
	KCF [32]	0.564	0.752	0.599	0.747	0.687	0.692	0.379	0.735	0.718	0.589	0.680
	SCM [24]	0.346	0.578	0.358	0.589	0.613	0.613	0.305	0.646	0.621	0.429	0.672
	TLD [60]	0.557	0.428	0.523	0.495	0.540	0.588	0.349	0.556	0.593	0.576	0.606
	ASLA [25]	0.255	0.496	0.283	0.473	0.529	0.521	0.156	0.479	0.535	0.333	0.552
	L1APG [59]	0.367	0.425	0.379	0.398	0.341	0.524	0.460	0.475	0.490	0.329	0.472
	MIL [18]	0.415	0.456	0.381	0.493	0.359	0.465	0.171	0.448	0.484	0.393	0.471
	CT [57]	0.330	0.339	0.314	0.463	0.365	0.361	0.152	0.429	0.405	0.336	0.448
	SRDCFdecon [40]	0.686	0.597	0.704	0.543	0.593	0.563	0.605	0.585	0.606	0.643	0.633
	SRDCF [35]	0.728	0.641	0.734	0.697	0.665	0.674	0.696	0.694	0.724	0.734	0.694
	LCT [58]	0.541	0.504	0.559	0.537	0.572	0.558	0.508	0.525	0.566	0.595	0.555
	CSRDCF [38]	0.728	0.859	0.838	0.839	0.792	0.782	0.879	0.790	0.806	0.827	0.808
	STAPLE [62]	0.550	0.654	0.553	0.573	0.617	0.596	0.530	0.581	0.624	0.621	0.593
	BIT [56]	0.492	0.562	0.531	0.498	0.539	0.508	0.493	0.546	0.565	0.560	0.491
Success	SKSCF-Proposed	**0.734**	**0.796**	0.755	**0.870**	**0.759**	0.710	**0.581**	**0.794**	**0.762**	**0.813**	0.679
	SKSCF_HoG_CN [13]	0.729	0.795	**0.757**	0.863	0.743	**0.720**	0.542	0.788	0.757	0.808	**0.682**
	KSCF_HoG_CN [13]	0.629	0.741	0.689	0.779	0.649	0.690	0.389	0.696	0.697	0.705	0.540
	MEEM [27]	0.706	0.747	0.692	0.711	0.653	0.648	0.470	0.694	0.694	0.742	0.594
	TGPR [63]	0.542	0.713	0.570	0.711	0.632	0.601	0.501	0.592	0.603	0.546	0.505
	KCF [32]	0.516	0.669	0.539	0.668	0.534	0.575	0.358	0.593	0.587	0.589	0.477
	SCM [24]	0.348	0.550	0.358	0.566	0.586	0.574	0.308	0.602	0.576	0.449	0.635
	TLD [60]	0.475	0.388	0.485	0.434	0.461	0.477	0.327	0.455	0.489	0.516	0.494
	ASLA [25]	0.261	0.468	0.284	0.485	0.514	0.496	0.163	0.469	0.509	0.359	0.544
	L1APG [59]	0.359	0.404	0.363	0.398	0.298	0.445	0.458	0.437	0.423	0.341	0.407
	MIL [18]	0.353	0.414	0.261	0.440	0.300	0.339	0.157	0.378	0.369	0.416	0.335
	CT [57]	0.327	0.323	0.262	0.420	0.308	0.290	0.143	0.360	0.325	0.405	0.342
	SRDCFdecon [40]	0.533	0.457	0.558	0.402	0.458	0.424	0.447	0.430	0.452	0.481	0.469
	SRDCF [35]	0.564	0.461	0.591	0.503	0.503	0.484	0.463	0.511	0.519	0.537	0.502
	LCT [58]	0.413	0.369	0.437	0.374	0.398	0.391	0.294	0.371	0.389	0.418	0.364
	CSRDCF [38]	0.592	0.616	0.627	0.574	0.585	0.556	0.567	0.572	0.573	0.589	0.558
	STAPLE [62]	0.437	0.485	0.446	0.433	0.490	0.440	0.335	0.453	0.459	0.467	0.431
	BIT [56]	0.386	0.392	0.409	0.378	0.377	0.368	0.285	0.394	0.390	0.398	0.333

The bar plots in Figure 3 show a comparison of DP and AUC on all three test suits of TB-100 dataset. The Figure 3a–c show that the DP and AUC of our proposed tracker are higher than the other recently developed trackers when tested on OTB-2013, OTB-2015, and OTB-50, respectively.

The OPE plots in Figure 2 show a comparison of precision and success rate with the recently reported trackers. Figure 2a presents a higher precision at each location error threshold in contrast to other trackers when tested on OTB-2013. The proposed tracker's success rate on OTB-2013 higher except on the overlapped threshold values between 0.7 to 1 shown in Figure 2b. Figure 2c shows the precision comparisons for various location error threshold when the tracker is evaluated on OTB-50. The Figure 3d present the success rate vs. overlap threshold evaluated on OTB-50 test suit. Both Figure 2c,d show the proposed tracker outperforms the state-of-the-art recently reported work. The OPE plots in Figure 2e,f show a comparison of precision and success rate over OTB-2015 test suit. In Figure 2e,f present a higher precision and success rate on the overall range of their respective thresholds.

Table 3. SKSCF-Proposed descriptor comparison with other trackers on OTB-2015.

	Attributes	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
	SKSCF-Proposed	**0.878**	0.837	**0.809**	**0.846**	**0.847**	0.811	**0.872**	**0.798**	**0.828**	0.698	0.804
	MBD [41]	0.780	0.783	0.757	0.813	0.765	0.801	0.808	0.783	0.816	**0.713**	0.799
	HDT [54]	0.833	**0.844**	0.789	0.835	0.819	**0.844**	0.846	0.783	0.815	0.633	**0.817**
	MEEM [27]	0.781	0.746	0.731	0.786	0.746	0.794	0.631	0.769	0.813	0.685	0.757
	Struck [55]	0.633	0.550	0.577	0.549	0.558	0.628	0.671	0.546	0.604	0.468	0.611
Precision	KCF [32]	0.630	0.712	0.600	0.627	0.713	0.693	0.560	0.631	0.678	0.498	0.641
	TLD [60]	0.582	0.515	0.530	0.515	0.571	0.606	0.531	0.558	0.588	0.463	0.584
	CXT [61]	0.568	0.444	0.545	0.415	0.511	0.599	0.509	0.454	0.538	0.388	0.538
	SCM [24]	0.331	0.578	0.269	0.572	0.608	0.544	0.602	0.574	0.582	0.425	0.577
	Staple [62]	0.659	0.560	0.597	0.569	0.588	0.648	0.634	0.566	0.674	0.494	0.675
	SRDCF [35]	0.677	0.580	0.600	0.579	0.598	0.665	0.645	0.579	0.683	0.510	0.683
	SKSCF-Proposed	0.595	**0.589**	**0.596**	**0.559**	**0.616**	0.538	**0.515**	**0.621**	0.545	**0.534**	0.560
	MBD [41]	0.595	0.577	0.592	0.558	0.553	**0.562**	0.457	0.557	**0.573**	0.483	0.526
	HDT [54]	0.568	0.578	0.574	0.543	0.409	0.555	0.401	0.528	0.510	0.472	0.486
	MEEM [27]	0.542	0.530	0.522	0.489	0.516	0.513	0.367	0.504	0.465	.0475	0.502
	Struck [55]	0.408	0.416	0.447	0.452	0.373	0.382	0.372	0.328	0.389	0.333	0.374
Success	KCF [32]	0.459	0.497	0.468	0.436	0.401	0.468	0.388	0.443	0.457	0.396	0.394
	TLD [60]	0.392	0.438	0.399	0.420	0.369	0.355	0.399	0.306	0.368	0.300	0.349
	SCM [24]	0.355	0.409	0.442	0.397	0.421	0.386	0.407	0.345	0.396	0.432	0.406
	Staple [62]	0.537	0.574	0.546	0.554	0.598	0.552	0.422	0.548	0.534	0.481	0.525
	SRDCF [35]	**0.597**	0.583	0.594	0.544	0.613	0.544	0.514	0.599	0.550	0.463	**0.561**

5. Conclusions

We propose robust low-rank descriptor for kernel support correlation filter. The proposed multi-channel feature-based tracker provides favorable results on several attributes of OTB-2013, OTB-50, and OTB-2015 test suit of the database. The low rank of the feature obtained by employing the novel fusion and reduction approach. The accuracy and speed increased by employing the circulant data matrix in the training procedure to the support vector machine. The results show the proposed tracker outperforms the recently developed trackers in the case of deformation, illumination variation, low resolution, occlusion, and out-of-view. The distance precision (DP) of the proposed tracker at location error of 20 pixels is 2.1%, 2.9%, and 6.6% higher than the best performing trackers on OTB-2013, OTB-50, and OTB-2015 respectively. The AUC is 4.2%, 3%, and 3.4% on OTB-2013, OTB-50, and OTB-2015 respectively.

Author Contributions: F. has developed the algorithm of the project, evaluated the framework, and wrote the paper. M.R. helped in formating and the correction of the grammer mistakes. M.J.K., Y.A., and H.T. supervised the research and improved the paper.

References

1. Aggarwal, J.K.; Xia, L. Human activity recognition from 3d data: A review. *Pattern Recognit. Lett.* **2014**, *48*, 70–80 [CrossRef]
2. Zhang, X.; Xu, H.; Fang, J. Multiple vehicle tracking in aerial video sequence using driver behavior analysis and improved deterministic data association. *J. Appl. Remote. Sens.* **2018**, *12*, 016014. [CrossRef]
3. Sivanantham, S.; Paul, N.N.; Iyer, R.S. Object tracking algorithm implementation for security applications. *Far East J. Electron. Commun.* **2016**, *16*, 1. [CrossRef]
4. Yun, X.; Sun, Y.; Yang, X.; Lu, N. Discriminative Fusion Correlation Learning for Visible and Infrared Tracking. *Math. Probl. Eng.* **2019**. [CrossRef]
5. Li, P.; Wang, D.; Wang, L.; Lu, H. Deep visual tracking: Review and experimental comparison. *Pattern Recognit.* **2018**, *76*, 323–338. [CrossRef]

6. Yazdi, M.; Bouwmans, T. New trends on moving object detection in video images captured by a moving camera: A survey. *Comput. Sci. Rev.* **2018,** *28*, 157–177. [CrossRef]

7. Pan, Z.; Liu, S.; Fu, W. A review of visual moving target tracking. *Multimed. Tools Appl.* **2017**, *76*, 16989–17018. [CrossRef]

8. Wu, Y.; Lim, J.; Yang, M.H. Online object tracking: A benchmark. In Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, OR, USA, 23–28 June 2013; pp. 2411–2418.

9. Liu, F.; Gong, C.; Huang, X.; Zhou, T.; Yang, J.; Tao, D. Robust visual tracking revisited: From correlation filter to template matching. *IEEE Trans. Image Process.* **2018**, *27*, 2777–2790. [CrossRef] [PubMed]

10. 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]

11. Hare, S.; Golodetz, S.; Saffari, A.; Vineet, V.; Cheng, M.M.; Hicks, S.L.; Torr, P.H. Struck: Structured output tracking with kernels. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 2096–2109. [CrossRef] [PubMed]

12. Zhang, K.; Zhang, L.; Liu, Q.; Zhang, D.; Yang, M.H. Fast visual tracking via dense spatio-temporal context learning. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.

13. Zuo, W.; Wu, X.; Lin, L.; Zhang, L.; Yang, M.H. Learning support correlation filters for visual tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1158–1172. [CrossRef] [PubMed]

14. Kristan, M.; Matas, J.; Leonardis, A.; Vojir, T.; Pflugfelder, R.; Fernandez, G.; Cehovin, L. A novel performance evaluation methodology for single-target trackers. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 2137–2155. [CrossRef] [PubMed]

15. Li, A.; Lin, M.; Wu, Y.; Yang, M.H.; Yan, S. Nus-pro: A new visual tracking challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 335–349. [CrossRef] [PubMed]

16. Fan, H.; Lin, L.; Yang, F.; Chu, P.; Deng, G.; Yu, S.; Ling, H. Lasot: A high-quality benchmark for large-scale single object tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5374–5383.

17. Kim, D.Y.; Vo, B.N.; Vo, B.T.; Jeon, M. A labeled random finite set online multi-object tracker for video data. *Pattern Recognit.* **2019**, *90*, 377–389. [CrossRef]

18. Babenko, B.; Yang, M.H.; Belongie, S. Visual tracking with online multiple instance learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 983–990.

19. Grabner, H.; Grabner, M.; Bischof, H. Real-time tracking via on-line boosting. *Bmvc* **2006**, *1*, 6.

20. Nam, H.; Han, B. Learning multi-domain convolutional neural networks for visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 4293–4302.

21. Li, H.; Li, Y.; Porikli, F. Deeptrack: Learning discriminative feature representations online for robust visual tracking. *IEEE Trans. Image Process.* **2016**, *25*, 1834–1848. [CrossRef] [PubMed]

22. Zhong, W.; Lu, H.; Yang, M.H. Robust object tracking via sparse collaborative appearance model. *IEEE Trans. Image Process.* **2014**, *23*, 2356–2368. [CrossRef]

23. Lan, X.; Zhang, S.; Yuen, P.C.; Chellappa, R. Learning common and feature-specific patterns: A novel multiple-sparse-representation-based tracker. *IEEE Trans. Image Process.* **2018**, *27*, 2022–2037. [CrossRef]

24. Zhong, W.; Lu, H.; Yang, M.-H. Robust object tracking via sparsity based collaborative model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1838–1845.

25. Jia, X.; Lu, H.; Yang, M.H. Visual tracking via adaptive structural local sparse appearance model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 16–21 June 2012; pp. 1822–1829.

26. Hong, Z.; Chen, Z.; Wang, C.; Mei, X.; Prokhorov, D.; Tao, D. Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 749–758.

27. Zhang, J.; Ma, S.; Sclaroff, S. MEEM: Robust tracking via multiple experts using entropy minimization. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 188–203.

28. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.

29. Wang, L.; Lu, H.; Yang, M.H. Constrained superpixel tracking. *IEEE Trans. Cybern.* **2018**, *48*, 1030–1041. [CrossRef]

30. Lukezic, A.; Zajc, L.C.; Kristan, M. Deformable parts correlation filters for robust visual tracking. *IEEE Trans. Cybern.* **2018**, *48*, 1849–1861. [CrossRef]

31. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 702–715.

32. 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]

33. Montero, A.S.; Lang, J.; Laganiere, R. Scalable kernel correlation filter with sparse feature integration. In Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 7–13 December 2015; pp. 587–594.

34. Galoogahi, H.K.; Sim, T.; Lucey, S. Correlation filters with limited boundaries. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4630–4638.

35. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Learning spatially regularized correlation filters for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4310–4318.

36. Bibi, A.; Mueller, M.; Ghanem, B. Target response adaptation for correlation filter tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 419–433.

37. Xu, T.; Feng, Z.H.; Wu, X.J.; Kittler, J. Learning Adaptive Discriminative Correlation Filters via Temporal Consistency preserving Spatial Feature Selection for Robust Visual Object Tracking. *IEEE Trans. Image Process.* **2019**, *28*, 5596–5609. [CrossRef] [PubMed]

38. Lukei, A.; Voj, T.; Zajc, L.E.; Matas, J.; Kristan, M. Discriminative correlation filter with channel and spatial reliability. *CVPR* **2017**, *126*, 6309–6318, .

39. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Discriminative scale space tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1561–1575. [CrossRef] [PubMed]

40. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1430–1438.

41. Tu, Z.; Guo, L.; Li, C.; Xiong, Z.; Wang, X. Minimum Barrier Distance-Based Object Descriptor for Visual Tracking. *Appl. Sci.* **2018**, *8*, 2233. [CrossRef]

42. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645. [CrossRef] [PubMed]

43. Danelljan, M.; Khan, F.S.; Felsberg, M.; Weijer, J.V.D. Adaptive color attributes for real-time visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.

44. Possegger, H.; Mauthner, T.; Bischof, H. In defense of color-based model-free tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2113–2120.

45. Lukezic, A.; Vojir, T.; Zajc, L.C.; Matas, J.; Kristan, M. Discriminative correlation filter with channel and spatial reliability. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6309–6318.

46. Song, T.; Li, H.; Meng, F.; Wu, Q.; Cai, J. Letrist: Locally encoded transform feature histogram for rotation-invariant texture classification. *IEEE Trans. Circuits Syst. Video Techol.* **2018**, *28*, 1565–1579. [CrossRef]

47. Saeed, A.; Fawad; Khan, M.J.; Riaz, M.A.; Shahid, H.; Khan, M.S.; Amin, Y.; Loo, J.; Tenhunen, H. Robustness-Driven Hybrid Descriptor for Noise-Deterrent Texture Classification. *IEEE Access* **2019**, *7*, 110116–110127. [CrossRef]

48. Khan, M.J.; Riaz, M.A.; Shahid, H.; Khan, M.S.; Amin, Y.; Loo, J.; Tenhunen, H. Texture Representation through Overlapped Multi-oriented Tri-scale Local Binary Pattern. *IEEE Access* **2019**, *7*, 66668–66679.

49. Khan, M.A.; Sharif, M.; Javed, M.Y.; Akram, T.; Yasmin, M.; Saba, T. License number plate recognition system using entropy-based features selection approach with SVM. *IET Image Process.* **2017**, *12*, 200–209. [CrossRef]

50. Xiong, W.; Zhang, L.; Du, B.; Tao, D. Combining local and global: Rich and robust feature pooling for visual recognition. *Pattern Recognit.* **2017**, *62*, 225–235. [CrossRef]

51. Zhang, L.; Zhang, Q.; Zhang, L.; Tao, D.; Huang, X.; Du, B. Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. *Pattern Recognit.* **2015**, *48*, 3102–3112. [CrossRef]

52. Arsalan, M.; Hong, H.; Naqvi, R.; Lee, M.; Kim, M.D.; Park, K. Deep learning-based iris segmentation for iris recognition in visible light environment. *Symmetry* **2017**, *9*, 263. [CrossRef]

53. Masood, H.; Rehman, S.; Khan, A.; Riaz, F.; Hassan, A.; Abbas, M. Approximate Proximal Gradient-Based Correlation Filter for Target Tracking in Videos: A Unified Approach. *Arab. J. Sci. Eng.* **2019**, 1–18. [CrossRef]

54. Qi, Y.; Zhang, S.; Qin, L.; Yao, H.; Huang, Q.; Lim, J.; Yang, M.H. Hedged deep tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 4303–4311.

55. Hare, S.; Saffari, A.; Struck, P.H.T. Structured output tracking with kernels. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 263–270.

56. Cai, B.; Xu, X.; Xing, X.; Jia, K.; Miao, J.; Tao, D. Bit: Biologically inspired tracker. *IEEE Trans. Image Process.* **2016**, *25*, 1327–1339. [CrossRef] [PubMed]

57. Zhang, K.; Zhang, L.; Yang, M.H. Real-time compressive tracking. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 864–877.

58. Ma, C.; Yang, X.; Zhang, C.; Yang, M.H. Long-term correlation tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5388–5396.

59. Bao, C.; Wu, Y.; Ling, H.; Ji, H. Real time robust l1 tracker using accelerated proximal gradient approach. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1830–1837.

60. Kalal, Z.; Mikolajczyk, K.; Matas, J. Tracking-learning-detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1409–1422. [CrossRef] [PubMed]

61. Dinh, T.B.; Vo, N.; Medioni, G. Context tracker: Exploring supporters and distracters in unconstrained environments. *CVPR* **2011**, 1177–1184. [CrossRef]

62. Bertinetto, L.; Valmadre, J.; Golodetz, S.; Miksik, O.; Torr, P.H. Staple: Complementary learners for real-time tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1401–1409.

63. Gao, J.; Ling, H.; Hu, W.; Xing, J. Transfer learning based visual tracking with gaussian processes regression. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 188–203.