Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.2832

2024, 43(4) 14-28

An adaptive learning based aberrance repressed multi-feature integrated correlation filter for Visual Object Tracking (VOT)

Mubashar Masood [a, *], Gulistan Raja [a]

[a] *Department of Electrical Engineering, University of Engineering and Technology Taxila, Pakistan*

[*] Corresponding author: Mubashar Masood, Email: mubashar.masod@students.uettaxila.edu.pk

Received: 15 May 2023, Accepted: 26 September 2024, Published: 01 October 2024

KEYWORDS

Correlation filter

Multi-feature fusion

Object tracking benchmark

Peak-to-sidelobe Ratio (PSR)

Adaptive learning

ABSTRACT

Target tracking via Correlation Filter (CF) is a hot research area of computer vision domain, and offers various credible benefits. Existing CF algorithms face challenges when there are target appearance variations due to background noise, scale and illumination changes, occlusion, and fast motion, which severely degrades the overall tracker performance. To get maximum benefits, an object tracker should perform well with the less computational burden in the presence of real time challenging situations. To address this issue, a novel visual object tracker is proposed based on multi feature fusion and adaptive learning technique with aberrance suppression. At first, multiple features i.e., Histogram of gradient (HOG), Color Naming (CN), saliency, and gray level intensities are combined using feature fusion technique. Further, based on the evaluation of final fused response map using Peak-to-Sidelobe Ratio (PSR), an adaptive learning strategy is integrated to improve the learning phase of tracker. Tracking results show that the proposed strategy beats the other modern CF trackers with Distance Precision (DP) scores of 88.2%, 85.9%, and 74.1% and 64.7% over OTB2013, OTB2015, and TempleColor128 and UAV123 datasets respectively.

1. Introduction

Object tracking is an essential problem in computer vision applications including security and supervision [1], human-machine interaction [2], Intelligent Transportation Systems (ITS) [3], and various UAV applications [4, 5]. The prime goal of object tracking is to estimate the tracking path of a single or multiple objects in the video sequence. With effective object tracking, more information about the recognition and activities of the object can be extracted. Object tracking becomes a challenging task as the targeted object experiences many real time challenging factors i.e., partial/full occlusion, target deformation, light and scale variations, background clutter, and fast motion.

Generative and discriminative are the two broad categories of object tracking techniques. The generative approach focuses on the general characteristics of the object model and searches for the target location where the candidate region has the highest similarity [6]. Optical flow is a popular generative technique used for object tracking [7]. It tracks the object by calculating the pixel movement between two consecutive frames. Although, optical flow is a simple tracking technique but it grapples with challenges in various situations i.e., fast movement, high resolution, scale and light variations, occlusion. The discriminative approach distinguishes the target from the surrounding in an image by considering it a classification problem. The discriminative method includes Deep Learning (DL) [8], Discriminative Correlation Filter (DCF) [9], or combination of both. YOLO (You Only Look Once) is a highly efficient and fast DL technique used for object detection that extends to object tracking by processing frames sequentially [10]. Although DL-based tracking

© Mehran University of Engineering and Technology 2024

methods have shown greater accuracy and outstanding performance on challenging datasets i.e., Siamese Network has achieved 160 Frames Per Second (FPS) [11] and Continuous Convolution Operator Tracker (C-COT) [8] ranked as top in the VOT2016 challenge [12], they have high computational complexity and thus require high Graphical Processing Units (GPU's) for training on the datasets. This bottleneck makes the DL-based tracker not feasible for real-time applications. Alternatively, Correlation Filter (CF) based trackers provide a favourable solution for real-time applications due to their low computational burden, and the good trade-off between accuracy and tracking speed.

This research proposes an effective solution to cope with the aforesaid issues for single object tracking. Inspired by the work presented in [13], we further improved the tracker's performance especially for aerial videos by incorporating aberrance repression feature. To assess the tracker performance, numerous experiments have been performed on challenging benchmarks i.e., OTB-2013, OTB-2015, TempleColor128, and UAV123.

The rest of the paper is organized as follows: Section 2 provides the detailed critical analysis of related work on object tracking. Section 3 discusses the main framework of proposed tracking approach. Section 4 describes the research methodology employed for this work and presents analysis of results. Finally, conclusion and future research directions is provided in section 5.

2. Related Work

Since the last decade, correlation filters (CF) for visual object trackers have attracted the object-tracking community due to their superior performance. A typical outline of a CF tracker requires the object information from the first frame only. Afterward, the tracking is performed over the entire image sequence. The effectiveness of localizing the targeted object is determined by how efficiently the filter is learned in the prior frame. CF filter utilized the frequency domain to performs the correlation operation and provides a sharp response over the target in the search window. Performing operations in the frequency domain, enable the CF trackers to achieve a higher frame rate.

Bolme et al., first introduced CF-based target tracking as the Minimum output sum of the Squared error (MOSSE) which becomes a baseline tracker for the later development of Visual Object Tracking (VOT) [14]. MOSSE tracker takes the grayscale image patches as input and computes the correlation in the Fourier domain for rapid calculations which leads to achieving the unprecedented tracking speed (669 FPS). Although MOSSE trackers have achieved very high running speed, the tracking accuracy was not satisfactory due to the utilization of the single channel Histogram of Gradient (HoG) feature. Based on the MOSSE framework, different trackers have been proposed to achieve high performance such as scale estimation [15, 16], integration of circulant structure [17, 18], multi-feature fusion [19, 20], and various filter learning and updation strategies [21, 22]. For high-speed tracking, these techniques use the Fast Fourier Transform (FFT) which produces unwanted boundary effects, causing degradation in the accuracy and precision of the tracker.

Different techniques have been proposed [23-28,32] to overcome the boundary effects. Danelljan et al. proposed a Spatially Regularized DCF (SRDCF) tracker in which the regularization parameter is injected to penalize the filter coefficients for learning more weights over the central part of the target [24]. The authors utilized the densely extracted training samples from the foreground as well as the surrounding background of the target. However, the expansion of the search region makes the tracker more sensitive to background clutter and aberrance.

Later, the SRDCF framework became the baseline of different trackers that shows impressive performance e.g., Continuous Convolution Operator Tracker (CCOT) [8], and Learning Spatial-Temporal Regularized Correlation Filter (STRCF) [21]. Although the SRDCF tracker has effectively mitigated the influence of boundary effects, the tracking speed was too low (5.3 FPS). The STRCF tracker is an amended version of the SRDCF tracker, in which spatial-temporal information is introduced to enhance the tracking speed and reduce the complexity. However, the STRCF tracker failed to tackle object appearance variations, particularly in aerial videos due to the fixed regularization component. To handle this issue, an adaptive spatial-temporal regularization parameter is introduced [42]. Wang et al. utilized the adaptive cosine window instead of a fixed cosine window to enlarge the search region which leads to the elimination of boundary effects [19]. Gao et al. introduced the dynamic saliency information into the DCF tracker to increase the positive samples for training to eliminate the boundary effects [13]. Yuan et al. proposed the Adaptive Spatial-Temporal Context Aware (ASTCA) tracker for aerial tracking scenarios. Adding spatial context information in the object model effectively alleviates the boundary effect and improves accuracy by distinguishing the target from its background [27].

© Mehran University of Engineering and Technology 2024

Recently, different trackers have been proposed with multiple features i.e., Deep features, HoG, Colour Naming (CN), Saliency, and greyscale features [9,16,19,20,29]. Wang et al., have combined the HoG and CN features to develop a reliable objective model [19]. Xu et al. incorporated the HoG, CN, and deep features for feature representations and achieved outstanding accuracy on OTB2013/2015 and VOT2017/2018 datasets. However, it achieves a tracking speed of 7.8 FPS only [29].

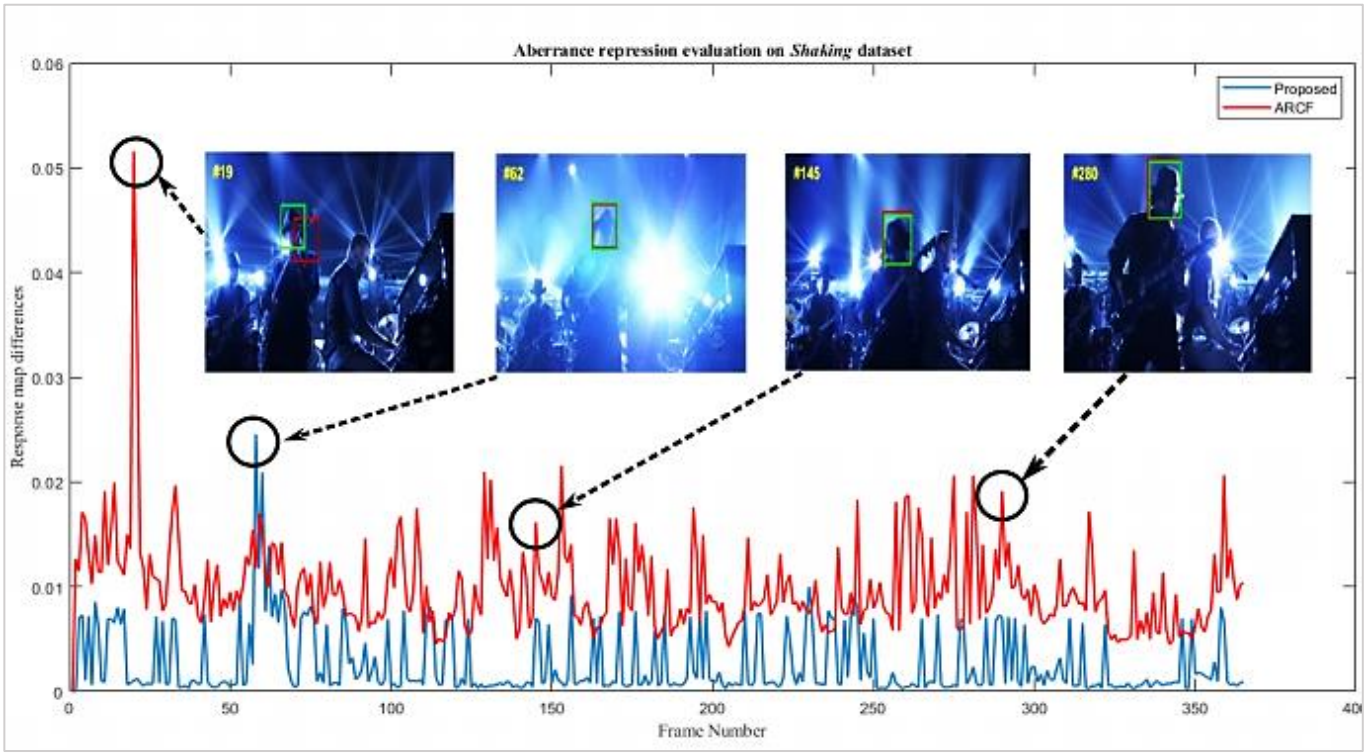

Fig. 1. Aberrance Repression Evaluation Of Proposed Tracker With Baseline ARCF Tracker On Shaking Dataset. Sudden Change Of Target Object From Frame #19 To #21 Are Well Addressed By The Proposed Method. Red And Green Bounding Boxes Represents Baseline And Proposed Trackers Respectively

Higher dimensional features produce enormous training samples, increasing computational complexity and reducing tracking speed. For example, the CCOT tracker updates 800,000 parameters during the learning and updation phase of the filter [8]. Such large training parameters limit the efficiency of the tracker in terms of computational cost and tracking speed. Moreover, these trackers are not feasible for real-time tracking, as embedded systems have limited memory size. Based on the incorporation of handcrafted (HoG + CN) and deep object features, Danelljan et al. presented an Efficient Convolution Operator (ECO) tracker that utilized the convolutional operator to suppress the training parameters up to 80% (with deep features) [30]. Another similar work is presented where the authors used the same features as our method but they did not cope with aberrance [31].

In [32], Ji et al., addresses the boundary effects and filter corruption for UAV tracking by introducing adaptive contextual learning and key filter selection in regularized correlation filter. Ma et al., proposed Spatial disturbance suppression and Object saliency-aware Correlation Filter (SOCF) tracker to handle similar objects disturbance in aerial tracking [33].

Based on the above-highlighted issues, we proposed an aberrance-repressed multi-feature integrated correlation filter with adaptive learning for VOT to enhance the tracker's performance in visually challenging scenarios. The strength of response map is assessed using Peak-to-Sidelobe Ratio (PSR) and an adaptive learning strategy is introduced based on these PSR values. In addition, multiple features are integrated by utilizing the feature fusion technique to effectively develop the object appearance model. Furthermore, an Alternating Direction Method of Multipliers (ADMM) is used to resolve the optimization problem with low computational cost.

3. Proposed Tracking Approach

This section discusses the main framework of the proposed aberrance repressed multi-feature integrated correlation filter with an adaptive learning technique.

3.1 Baseline Tracker

In this study, an Aberrance Repressed Correlation Filter (ARCF) tracker is taken as the baseline tracker that utilized the HoG features only for the representation of the object appearance model [17].

Let x and w present a vectorized input sample and a filter respectively, then response $R(x)$ is given as,

© Mehran University of Engineering and Technology 2024

$$R(x) = \sum_{c=1}^{C} x^c \star w^c \quad (1)$$

where $\circledast$ shows the circular convolution operation. The C is a dimensional feature map that is obtained using input sample x and the ideal response y. The tracker estimates a similar object in the next frame using the filter w. The optimal parameters of filter w are determined during the filter training by minimizing the loss function ε(w) given as:

$$\varepsilon(w) = \frac{1}{2}\left\| y - \sum_{c=1}^{C} Bx^c \star w^c \right\|_2^2 + \frac{\lambda}{2}\sum_{c=1}^{C}\|w^c\|_2^2 \quad (2)$$

where B is a binary cropping matrix, used to select central M elements of the vectorized input sample x^c, y is the desired correlation output in Gaussian form with a peak centred over the target of interest, and λ regularization parameter. The size of the cropped matrix B is $M \times N$ where M and N represent rows, and columns respectively, and usually, $N \gg M$.

3.2 Proposed Approach

Fig. 2 demonstrates the block diagram of the proposed tracking approach. The image patch which contains the area of interest including the target object is extracted from the current frame and then multiple features i.e., fHOG, CN, Gray level intensities, and Saliency, are obtained to develop a resilient object model. To combine the response maps of all extracted features effectively, a fusion strategy based on the PSR is utilized. Besides multiple feature extraction, an aberrance suppression scheme is employed to suppress the background interference. Finally, the combined response map is obtained based on the fused and aberrance repression response maps.

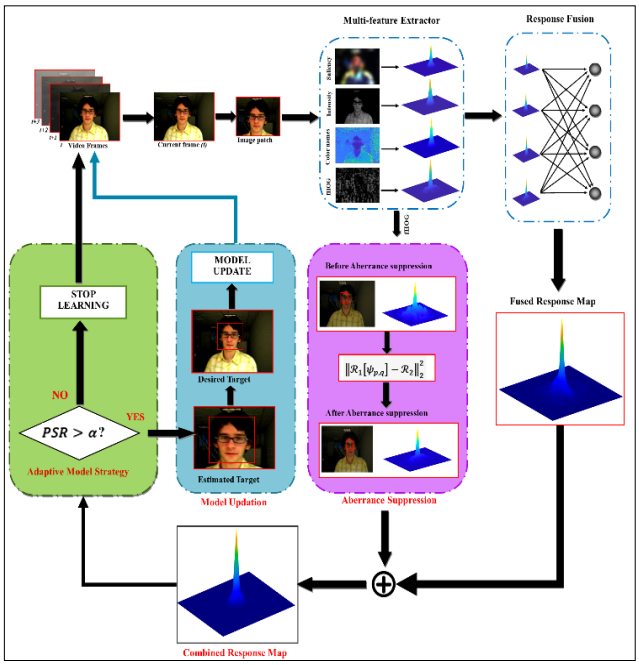

Fig.2. Overview Of Proposed Tracking Method

3.2.1 Multi-feature extractor

I. HOG feature extraction: The Histogram of Gradient (HOG) is a single feature descriptor that utilized the gradient information and is widely used in object detection to provide texture information about the targeted object. HOG is better than other similar descriptors i.e., Scale Invariant and Feature Transform (SIFT), canny edge detector, etc., as it utilized both the angle as well as the magnitude to calculate the gradient information of the target. Thus, it provides a better object appearance model against geometric variations. Felzenszwalb's HOG (fHOG) is a faster implementation of the HOG feature descriptor in which the feature dimensions are reduced for rapid gradient calculations [34]. Taking this advantage, fHOG is used as a feature descriptor and 31 channels of HOG features are adapted in this work. The fHOG map of the cropped image patch is shown in Fig. 3.

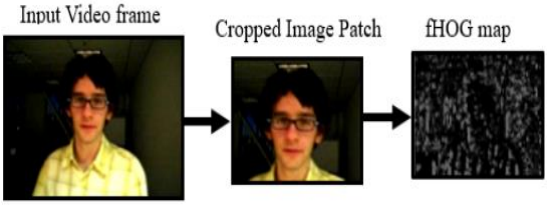

Fig. 3. fHOG Feature Map

II. Colour Naming (CN) feature extraction: For developing a robust appearance model, textual information's are not enough to handle different visual challenges i.e., object deformation, illumination, and scale variations. The colour Naming (CN) technique performs well in such scenarios. In CN, the input RGB image is transformed into an 11-dimensional annotated colour space using 11 linguistic colour labels. CN map of the input image patch is shown in Fig. 4.

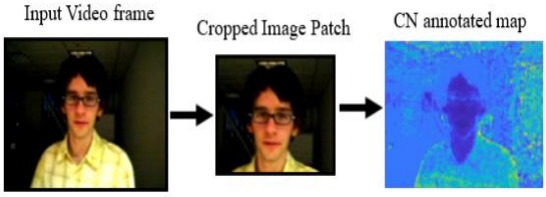

Fig. 4. CN Feature Map

III. Grey feature extraction: In some visually challenging scenarios, CN-based feature descriptors do not model the object effectively. For example, in aerial tracking, when the context around the object is similar

© Mehran University of Engineering and Technology 2024

to the foreground, the CN features fail and are labelled with the same colour name which increases the spurious detection rate. To handle this issue, Gray-level intensities that are robust to motion blur are also integrated to build a robust appearance model. The gray feature map G(t) is shown in Fig. 5, and obtained by converting the RGB image patch into the gray-scale image using the following Eq:

$$\mathcal{G}^{(t)}(i,j) = \beta_R x_R^{(t)}(i,j) + \beta_G x_G^{(t)}(i,j) + \beta_B x_B^{(t)}(i,j) \quad (3)$$

where *x(t)* is the input image patch, the values of β_R, β_G and β_B are taken as 0.2989, 0.5870, and 0.1140 respectively.

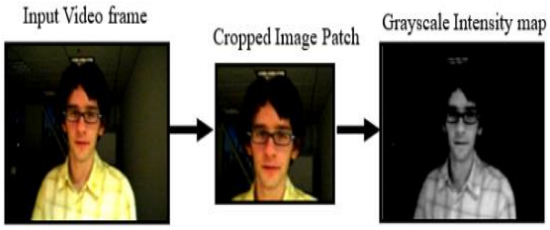

Fig. 5. Grayscale Feature Map

IV. Saliency feature extraction: Saliency refers to the exclusive features which make an object more prominent than its background. In this work, this feature is adopted to suppress background interference and enhance the object foreground. The procedure described in [35] is utilized for constructing the saliency map as depicted in Fig. 6.

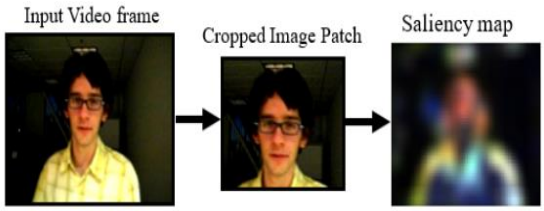

Fig. 6. Saliency Feature Map

3.2.2 Response fusion

Single feature descriptors rely on a single feature such as HOG, CN, or deep features and failed to perform well in challenging scenarios when object appearance changes rapidly. Multiple features, on the other hand, achieve a more robust object appearance model and can effectively tackle visually challenging factors. Feature fusion is a technique in which different features are combined in a way to develop a more robust target appearance model.

In the proposed tracking framework, four features are utilized including fHOG, CN, Gray, and Saliency. PSR is utilized with the response of each feature to filter the noise for creating a more stable response as depicted in Fig. 7. Final fused response map $\mathcal{F}(t)$ is obtained using the following Eq:

$$\mathcal{F}^{(t)} = \frac{1}{6}\Big[\Big(\beta^{\mathcal{H}^{(t)}\mathcal{C}^{(t)}}\mathcal{R}^{\mathcal{H}^{(t)}\mathcal{C}^{(t)}} \oplus \beta^{\mathcal{H}^{(t)}\mathcal{G}^{(t)}}\mathcal{R}^{\mathcal{H}^{(t)}\mathcal{G}^{(t)}} \oplus \beta^{\mathcal{H}^{(t)}\mathcal{S}^{(t)}}\mathcal{R}^{\mathcal{H}^{(t)}\mathcal{S}^{(t)}} \oplus \beta^{\mathcal{C}^{(t)}\mathcal{G}^{(t)}}\mathcal{R}^{\mathcal{C}^{(t)}\mathcal{G}^{(t)}} \oplus \beta^{\mathcal{C}^{(t)}\mathcal{S}^{(t)}}\mathcal{R}^{\mathcal{C}^{(t)}\mathcal{S}^{(t)}} \oplus \beta^{\mathcal{G}^{(t)}\mathcal{S}^{(t)}}\mathcal{R}^{\mathcal{G}^{(t)}\mathcal{S}^{(t)}}\Big) + \mathcal{R}_{AR}\Big] \quad (4)$$

where $\mathcal{H}$, $\mathcal{C}$, $\mathcal{G}$ and $\mathcal{S}$ represents response maps of HOG, CN, Grayscale, and Saliency respectively. β is the weighted PSR that helps to eliminate the interference from the response map. In the Eq. 4, $\oplus$ denotes element-wise addition and R_{AR} is the aberrance suppressed response map. The final response has only one highest peak which is the estimated location of the target as demonstrated in Fig. 7.

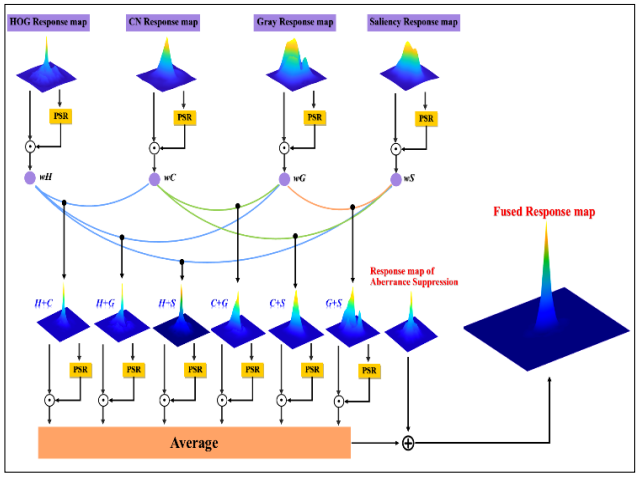

Fig. 7. Multi-Features Fusion Technique

3.2.3 Aberrance suppression

Aberrance takes place in different real-time scenarios, especially in aerial videos. i.e., partial/full occlusion, in/out plan rotation, and background clutter. In the presence of aberrance, abrupt changes occurred in the response map which leads to poor performance and tracking failure. To suppress the aberrance, L^2 norm or Euclidean norm is used to find the difference in response maps of the previous and current frame $\mathcal{R}_1$ and $\mathcal{R}_2$ as follows:

$$\left\|\mathcal{R}_1[\psi_{p,q}] - \mathcal{R}_2\right\|_2^2 \quad (5)$$

where *p,q* denote the location differences between both responses in 2D space, and $\psi_{p,q}$ represents shifting operation, performed to coincide both peaks. In an aberrance scenario, produces a high value due to the dissimilarity between the peaks of both response maps. With aberrance repression, the objective function in Eq. 2 becomes:

© Mehran University of Engineering and Technology 2024

$$\varepsilon(w) = \frac{1}{2}\left\|y - \sum_{c=1}^{C} Bx^c \star w^c\right\|_2^2 + \frac{\lambda}{2}\sum_{c=1}^{C}\|w^c\|_2^2 + \frac{\gamma}{2}\left\|\sum_{c=1}^{C}(Bx_{t-1}^c \star w_{t-1}^c)\left[\psi_{p,q}\right] - \sum_{c=1}^{C}(Bx_t^c \star w_t^c)\right\|_2^2 \quad (6)$$

where t and $t-1$ represents current and previous frames respectively. Parameter γ is an aberrance penalizing factor.

3.2.4 Adaptive model strategy

Occlusion handling is a crucial step in object detection and tracking. In various signal processing applications, the peak signal strength is measured by the PSR. Based on this idea, PSR is used to efficiently tackle occluded situations. To calculate the PSR, the correlation response map is divided into the correlation peak and side-lobes. The peak of the correlation response is the predicted target object location, while the side-lobe is the rest of the pixels in the search window. The PSR is calculated as:

$$PSR = \frac{\max\left(\mathcal{R}(z)\right) - \mu_{sl}\left(\mathcal{R}(z)\right)}{\sigma_{sl}\left(\mathcal{R}(z)\right)} \quad (7)$$

where R(z) represents correlation response map, max R(z) is the response peak, μsl and σsl are the mean and standard deviation of side-lobe respectively. The target patch is denoted by z. PSR can be considered an effective indicating factor for occlusion detection. In target occluded conditions (79th frame of the Jogging dataset), the response map has a low value of PSR because of the presence of side-lobe which increases the standard deviation σsl as shown in Fig. 8(a). In this case, there are many minor peaks (sidelobes) appearing in the response map therefore, the PSR value given by Eq. 7 is reduced to 9.2. In occlusion-free situations (52nd frame), only one sharp peak centred at the targeted object will appear while the rest of the area is flat. For this case, the value of Eq. 7 will be high (15.7) due to the absence of minor lobes as demonstrated in Fig. 8(b).

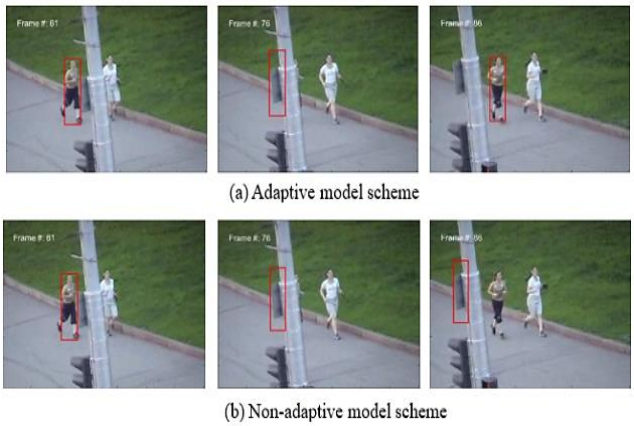

Fig. 8. Peak-To-Side-Lobe Ratio (Psr) Values In Occluded And Non-Occluded Scenarios In Two Different Frames (Dataset: Jogging)

In the structure of most CF trackers, the model is modified in each frame to adjust the model variations in the target. The target location in the succeeding frame is estimated based on the modified learned model. However, as the target experiences occlusion or goes through out-of-view situations, the model drifting may occur due to the uncertain training samples for filter learning. To get superior tracking performance, it is imperative to strict the tracker in the learning phase to not update the model in challenging conditions.

To cope with these situations, an adaptive model learning strategy is introduced based on the values of PSR. Mathematically, it can be written as:

$$\eta = \begin{cases} \eta_o & if\ PSR < \alpha \\ 0 & otherwise \end{cases} \quad (8)$$

The values of α and η_o are shown in Table 1. In occluded or out-of-view scenarios, the learning rate η will be set to zero, no model updation will be performed in such situations according to Eq. 8. A comparison of both schemes (adaptive and non-adaptive) is demonstrated in, where the target is occluded from 61st to 79th frame of Jogging dataset. Our proposed adaptive model strategy successfully re-detects the target when it reappears from the occlusion (i.e., 86th frame) as shown in Fig. 9(a). Conversely, if the adaptive model scheme is not employed, the tracking failure occurs due to the corruption of the object model during occlusion as shown in Fig. 9(b).

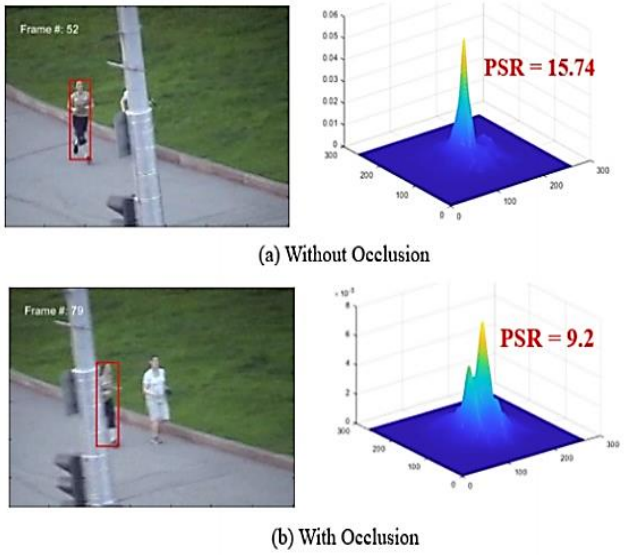

Fig. 9. Tracking Results Of Adaptive And Non-Adaptive Schemes. Dataset: Jogging

3.2.5 Model updation

The model x is updated in each frame to adjust the object appearance changes. The model $\hat{x}^{model}$ is updated using the following Eq:

© Mehran University of Engineering and Technology 2024

$$\hat{x}_t^{model} = (1-\eta)\hat{x}_{t-1}^{model} + \eta\hat{x}_t \quad (9)$$

where η shows the learning rate, *t*, and $t-1$ present the current and previous frame respectively.

4. Simulation Environment and Results

4.1 Simulation Setup

The simulation of the proposed tracker is performed using MATLAB® 2018 on Intel® Core™ i7-11700 CPU at 2.40 GHz with 16GB of RAM. In this research, HoG, CN, saliency, and greyscale features are adopted to build a more robust object model. The proposed tracker is evaluated with the different parameter settings as listed in Table 1.

4.2 Evaluation Datasets

Extensive experiments are carried out on the well-known challenging benchmarks for the assessment of our tracker including OTB-2013 [36], OTB-2015 [37], TempleColor128 [38], and UAV123 [39] which consists of 50, 100, 128, and 123 video sequences respectively. It can be classified into 11 attributes according to the challenges including illumination variations (IV), scale variations (SV), occlusion (OCC), deformation (DEF), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out-of-view (OV), background clutter (BC), and low resolution (LR).

Table 1
Major parameters used in the proposed method

Parameter	Value	Parameter	Value
Feature cell size	4×4	ADMM iterations	2
Number of scales (S)	5	Learning rate η	0.019
Scale step	1.01	Bandwidth of Gaussian function	$\frac{\sqrt{w \times h}}{16}$
Regularization parameter $\boldsymbol{\lambda}$	0.01	Aberrance penalty factor γ	0.71
$\boldsymbol{\eta_o}$	0.019	α	10

4.3 Evaluation Methodology

One Pass Evaluation (OPE) is a common method for robustness evaluation which requires the ground truth of the target from the first frame only, and then mean accuracy and success rate are obtained from all the subsequent frames. All the DCF trackers are evaluated based on three metrics: Distance Precision (DP), Overlap Success (OS), and Centre Location Error (CLE). A short description of all three metrics is given below:

Centre Location Error (CLE) represents the difference (in pixels) between the tracked and ground truth bounding boxes. The least value of CLE is an indication of the higher tracker accuracy. CLE is calculated as:

$$CLE = \sqrt{\left(x_{gt} - x_t\right)^2 + \left(y_{gt} - y_t\right)^2} \quad (10)$$

where x_{gt}, y_{gt} and x_t, y_t represents the central location of ground truth and estimated bounding box respectively.

Distance Precision (DP) rate counts the frames in which the CLE of the predicted bounding box is less than the pre-set threshold from the ground truth. The precision plot is obtained by calculating the precision over various thresholds.

Overlap Success (OS) rate shows the number of video frames whose intersection over union (IoU) is greater than a pre-set threshold t (normally 0.5). The larger value of the success rate is an indication of more accuracy. Let b_t *and* b_g represents the bounding box of our tracker and ground truth respectively, success rate S is calculated as follows:

$$S = \frac{|b_t \cap b_g|}{|b_t \cup b_g|} > t \quad (11)$$

4.4 Performance Evaluation

4.4.1 Evaluation of OTB2015 dataset

The precision and success plot of the proposed tracker along with other modern trackers on the OTB2015 dataset with the percentage score of precision and success is shown in Fig. 10. From the precision plot, it is obvious that our method beats the baseline ARCF tracker [40] and obtains a DP score of 85.9% which is 6.1% greater than the ARCF score (79.8%). Our tracker has achieved a running speed of 13.4 FPS. Table 2 shows the mean CLE, FPS, DP, and OS score percentages of various trackers, the Centre Location Error (CLE) of our tracker is 17.7 only which is 3.6% and 5% less than the BACF (second best), and STRCF (third best) trackers respectively. From Table 2, it is obvious that the proposed scheme has obtained the lowest mean CLE (17.7) in comparison with other trackers which shows the greater accuracy of proposed method.

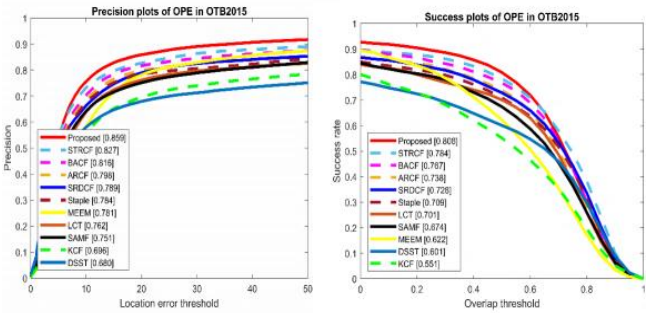

Fig. 10. Precision And Success Plot Over OTB2015 Dataset

4.4.2 Evaluation of OTB2013 dataset

Fig. 11 shows the precision and success plot over OTB2013 dataset. Tracking results show that our tracker has outperformed all other trackers on both precision and success plots by obtaining a DP score of 88.2% and an OS score of 66.7%. BACF trackers appear as the second-best tracker for this dataset and achieve a DP score of 84.9% and an OS score of 64.4%. The baseline ARCF tracker has achieved a DP score of 83.2% and an OS score of 62.1%. Compared to the ARCF tracker, our method gains 3.3% and 2.3% improvements in precision and success rate respectively. Table 3 shows the DP and OS score (%) of all other trackers. Fig. 14(b) shows the graphical representation of precision and AUC score achieved by the different trackers over the OTB2013 dataset.

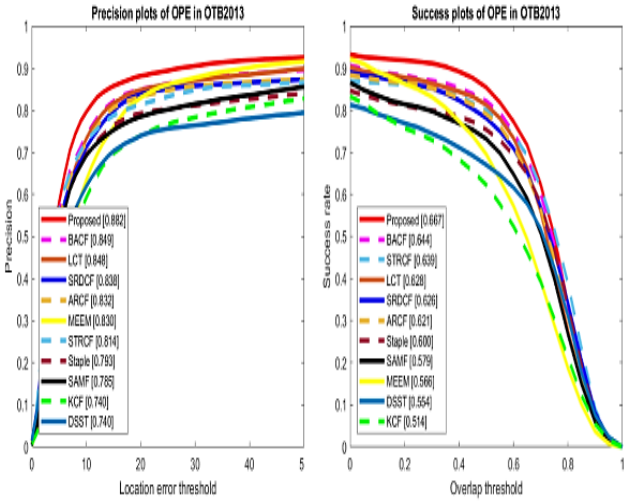

Fig. 11. Precision and Success Plot Over OTB2013 Dataset

4.4.3 Evaluation of TempleColor128 dataset

The performance of our tracker is also investigated on the TempleColor128 dataset [38] which consists of 128 challenging videos. Tracking results in comparison with other existing trackers CSK [18], STRCF [21], BACF [22], SRDCF [24], DSST [29], ARCF [40], SAMF [41], MUSTER [42], TLD [43] are shown in Fig. 12. Based on the tracking results, it is clear that the proposed method has performed well against the abovementioned trackers with a DP score of 74.1% and OS score of 68.9%. Moreover, our method gains improvement by 6%, 7.4%, and 7.8% precision over STRCF, ARCF, and SRDCF trackers. Fig. 14(c) shows the graphical representation of precision and AUC score of the aforementioned trackers over TempleColor128 dataset.

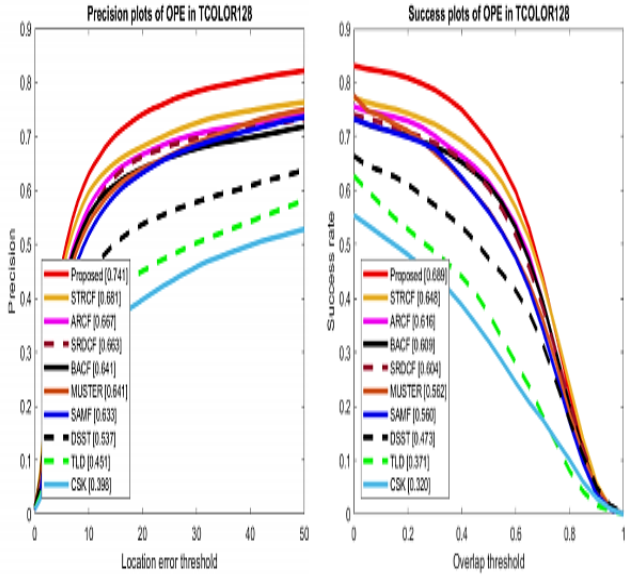

Fig. 12. Precision And Success Plot Over Templecolor128 Dataset

4.4.4 Evaluation of UAV123 dataset

Fig. 13 demonstrates the comparison of proposed and other trackers including CSK [18], STRCF [21], BACF [22], SRDCF [24], DSST [29], ARCF [40], SAMF [41], MUSTER [42], TLD [43] trackers over UAV123 dataset. The tracking results show that our tracker performs well among other trackers, and achieves 64.7% precision and 46.1% success rate, while the baseline ARCF tracker obtains 61.2% and 43.3% precision and success scores respectively. Based on the tracking results, it is clear that our tracker outperforms ARCF, MEEM, and SRDCF trackers by 3.5%, 6.3%, and 7.2% DP scores respectively.

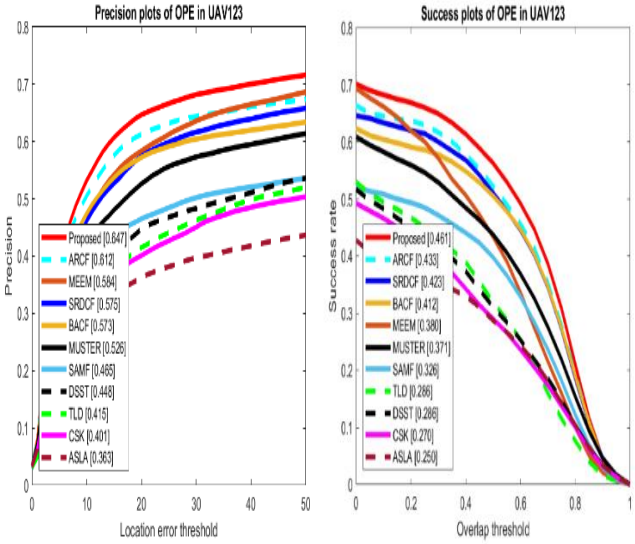

Fig. 13. Over All Precision And Success Plots Over UAV123 Dataset

© Mehran University of Engineering and Technology 2024

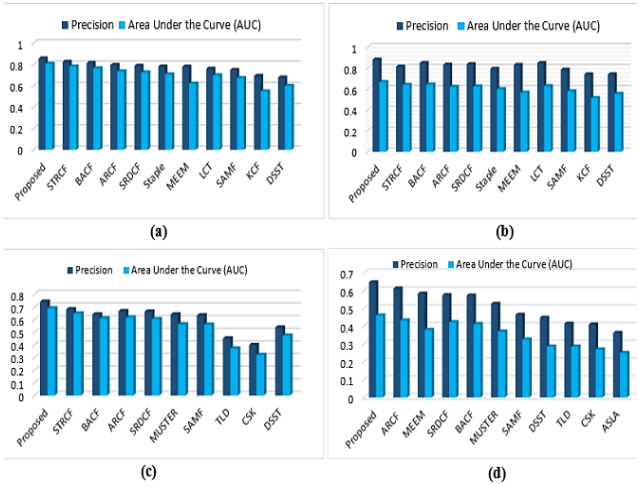

Fig. 14. Precision And Area Under The Curve (AUC) Of Different Trackers Over: (A) OTB2015 (B) OTB2013 (C) Templecolor128 (D) UAV123

4.5 Attribute-Based Performance Analysis

In OTB benchmarks [36,37], authors categorized the video sequences into 11 attributes as described in Section 4.2. The performance of our tracker is further investigated for each attribute of used datasets. Fig. 16 demonstrates the attribute-based performance comparison with other trackers over the OTB2015 benchmark.

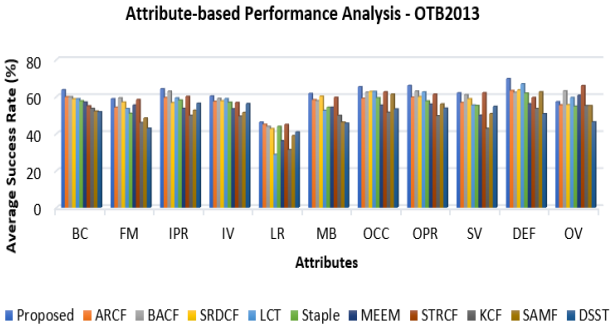

Fig. 15. Attribute-Based Performance Analysis Of Our Method In Comparison With Other Trackers Over The OTB2013 Dataset

Among other trackers, STRCF appears as a strong candidate for deformation (58.5%), in-plane-rotation (58.3%), illumination variation (63.2%), and out-of-plane rotation (59.8%) attributes, while BACF performs well for Background Clutter (62.8%). Table 4 presents the attribute-based effectiveness of all the trackers on 9 more challenging sequences. From the results, it is clear with a score of 5.491. Fig. 15 illustrates the average success score (%) of our method with other trackers over 11 dominant challenging attributes of the OTB2013 benchmark. that our tracker has achieved a 5.516 overall score, and STRCF appeared as the second-best tracker.

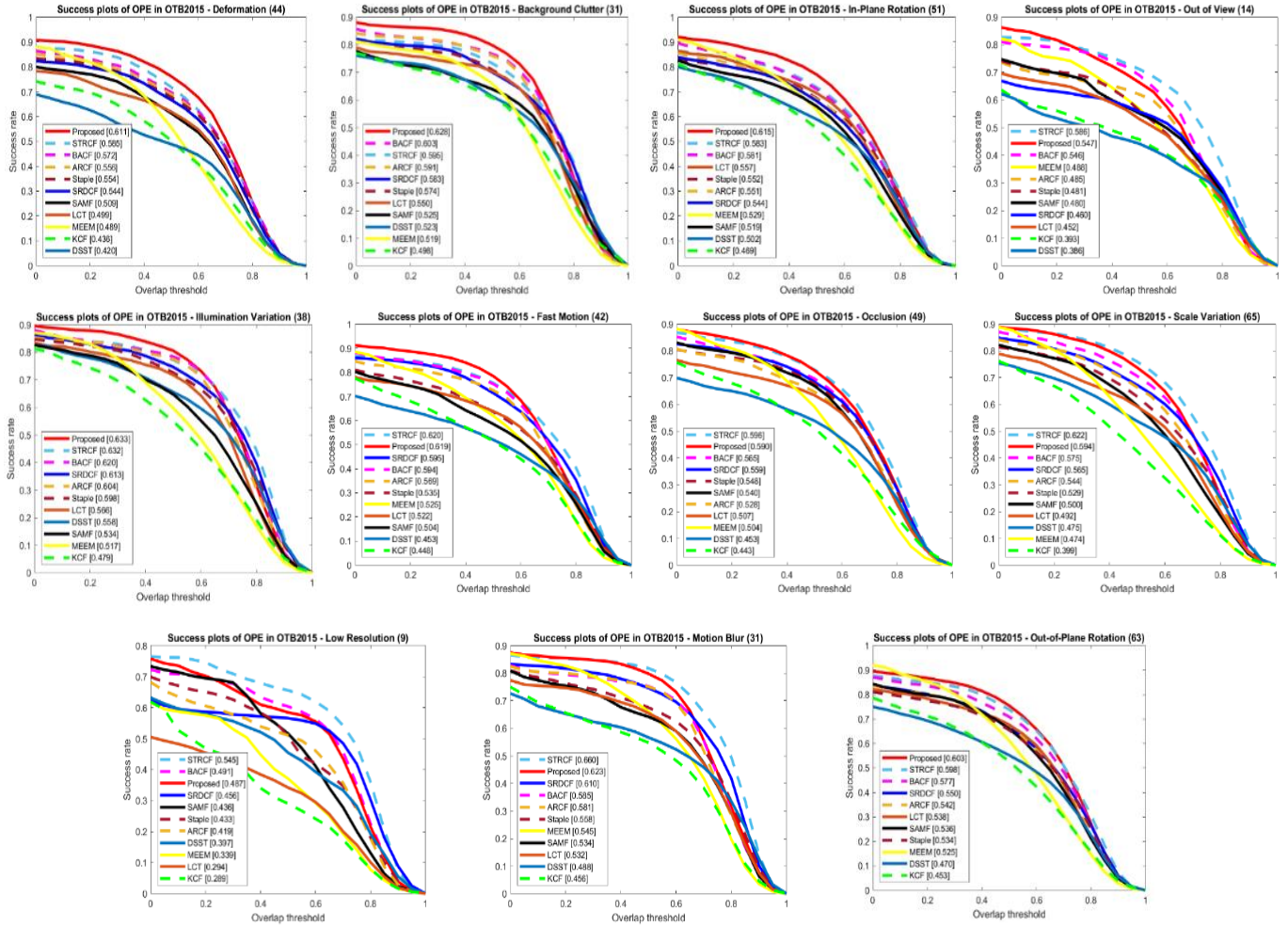

Fig. 16. Success Plot Of 11 Challenging Attribute Over OTB2015 Dataset

© Mehran University of Engineering and Technology 2024

Table 2
Performance comparison over OTB-2015 benchmark

Method	STRCF [21]	BACF [22]	ARCF [40]	SRDCF [24]	Staple [46]	MEEM [45]	LCT [44]	SAMF [41]	KCF [47]	DSST [29]	Proposed
CLE	22.7	21.3	27.2	25.2	36.8	47.1	48.7	31.0	47.8	47.0	17.7
DP (%)	82.7	81.6	79.8	78.9	78.4	78.1	76.2	75.1	69.6	68.0	85.9
OS (%)	78.4	76.7	73.8	72.8	70.9	62.2	70.1	67.4	55.0	60.1	80.8
FPS	28.9	26.7	23.1	5.8	76.6	22.4	21.7	23.2	216	40.9	13.4

Table 3
Performance comparison over OTB-2013 Benchmark

	STRCF [21]	BACF [22]	ARCF [40]	SRDCF [24]	Staple [46]	MEEM [45]	SAMF [41]	KCF [47]	DSST [29]	Proposed
DP (%)	81.4	84.9	83.2	83.8	79.3	83.0	78.5	74.0	74.0	88.2
OS (%)	63.9	64.4	62.1	62.6	60.0	56.6	57.9	51.4	55.4	66.7

Table 4
Success score over 9 more challenging attributes of OTB2015 dataset. red, green, and blue bold values show the best, second best, and third best respectively

	Proposed	STRCF	BACF	ARCF	SRDCF	Staple	MEEM	LCT	SAMF	KCF	DSST
SV	0.594	0.622	0.575	0.544	0.565	0.529	0.474	0.492	0.500	0.399	0.475
OPR	0.603	0.598	0.577	0.542	0.550	0.534	0.525	0.538	0.536	0.453	0.470
OCC	0.590	0.596	0.565	0.528	0.559	0.548	0.504	0.507	0.540	0.443	0.453
MB	0.623	0.660	0.585	0.581	0.610	0.558	0.545	0.532	0.534	0.456	0.488
IV	0.633	0.632	0.620	0.604	0.613	0.598	0.517	0.566	0.534	0.479	0.558
IPR	0.615	0.583	0.581	0.551	0.544	0.552	0.529	0.557	0.519	0.469	0.502
FM	0.619	0.620	0.594	0.569	0.595	0.535	0.525	0.522	0.504	0.448	0.453
DEF	0.611	0.585	0.572	0.556	0.544	0.554	0.489	0.499	0.509	0.436	0.420
BC	0.628	0.595	0.603	0.591	0.583	0.574	0.519	0.550	0.525	0.498	0.523
Overall	5.516	5.491	5.272	5.066	5.163	4.982	4.627	4.763	4.701	4.081	4.342

Table 5
Success score over 11 challenging attributes of the OTB2013 dataset. red, green, and blue bold values show best, second best, and third best respectively

	Proposed	STRCF	BACF	ARCF	SRDCF	Staple	MEEM	LCT	SAMF	KCF	DSST
SV	0.619	0.620	0.609	0.567	0.587	0.551	0.498	0.553	0.507	0.427	0.546
OV	0.571	0.658	0.630	0.554	0.555	0.547	0.606	0.594	0.550	0.550	0.462
OPR	0.659	0.612	0.629	0.596	0.599	0.575	0.558	0.624	0.559	0.495	0.536
OCC	0.652	0.625	0.623	0.590	0.627	0.593	0.552	0.627	0.612	0.514	0.532
MB	0.616	0.595	0.577	0.582	0.601	0.541	0.541	0.524	0.461	0.497	0.455
LR	0.461	0.448	0.438	0.449	0.426	0.438	0.36	0.286	0.388	0.312	0.408
IV	0.602	0.567	0.589	0.573	0.576	0.568	0.533	0.588	0.513	0.493	0.561
IPR	0.641	0.60	0.628	0.594	0.566	0.580	0.535	0.592	0.525	0.497	0.563
FM	0.587	0.583	0.592	0.541	0.569	0.508	0.553	0.534	0.483	0.459	0.428
DEF	0.696	0.594	0.622	0.632	0.635	0.618	0.56	0.668	0.625	0.534	0.506
BC	0.636	0.549	0.598	0.598	0.587	0.576	0.569	0.587	0.520	0.535	0.517
Overall	6.74	6.451	6.535	6.276	6.238	6.095	5.865	6.177	5.743	5.313	5.514

© Mehran University of Engineering and Technology 2024

Table 6
Success score over 11 challenging attributes of Templecolor128 dataset. red, green, and blue bold values show best, second best, and third best respectively

	Proposed	STRCF	BACF	ARCF	SRDCF	MUSTER	TLD	SAMF	CSK	DSST
SV	0.539	0.516	0.479	0.480	0.481	0.434	0.327	0.450	0.276	0.404
OV	0.462	0.479	0.400	0.343	0.395	0.337	0.350	0.339	0.204	0.279
OPR	0.547	0.477	0.463	0.471	0.456	0.442	0.285	0.469	0.286	0.399
OCC	0.520	0.467	0.445	0.441	0.476	0.455	0.309	0.446	0.266	0.368
MB	0.483	0.446	0.391	0.405	0.424	0.357	0.271	0.410	0.240	0.339
LR	0.449	0.478	0.382	0.420	0.354	0.351	0.399	0.242	0.205	0.269
IV	0.529	0.511	0.522	0.517	0.512	0.530	0.300	0.484	0.302	0.438
IPR	0.529	0.449	0.466	0.459	0.453	0.430	0.273	0.409	0.279	0.394
FM	0.514	0.472	0.441	0.435	0.447	0.410	0.280	0.451	0.264	0.371
DEF	0.584	0.540	0.529	0.560	0.528	0.509	0.261	0.537	0.247	0.372
BC	0.524	0.494	0.507	0.517	0.499	0.531	0.286	0.433	0.294	0.387
Overall	5.68	5.329	5.025	5.048	5.025	4.786	3.341	4.67	2.863	4.02

Table 7
Success score over 11 challenging attributes of the UAV123 dataset. red, green, and blue bold values show best, second best, and third best respectively

	Proposed	ARCF	BACF	MEEM	MUSTER	SRDCF	TLD	SAMF	CSK	ASLA	DSST
SV	0.270	0.267	0.246	0.241	0.235	0.237	0.242	0.166	0.149	0.186	0.171
OV	0.373	0.325	0.308	0.322	0.281	0.333	0.160	0.238	0.189	0.180	0.210
OPR	0.392	0.354	0.333	0.327	0.299	0.346	0.239	0.274	0.205	0.194	0.235
OCC	0.345	0.256	0.275	0.231	0.203	0.311	0.148	0.220	0.151	0.099	0.157
MB	0.385	0.360	0.326	0.337	0.302	0.355	0.219	0.282	0.220	0.202	0.246
LR	0.444	0.404	0.396	0.361	0.333	0.399	0.252	0.271	0.214	0.159	0.225
IV	0.424	0.398	0.373	0.340	0.343	0.390	0.268	0.306	0.247	0.238	0.262
IPR	0.336	0.274	0.274	0.311	0.225	0.263	0.163	0.160	0.134	0.148	0.137
FM	0.396	0.330	0.320	0.325	0.295	0.368	0.202	0.289	0.227	0.162	0.245
DEF	0.231	0.199	0.172	0.211	0.229	0.229	0.135	0.197	0.150	0.134	0.156
BC	0.403	0.358	0.352	0.348	0.316	0.356	0.251	0.264	0.203	0.199	0.231
Overall	3.999	3.525	3.375	3.354	3.061	3.587	2.279	2.667	2.089	1.901	2.275

Table. 5 provides attribute-based quantitative data of the proposed and other trackers on the OTB2013 dataset. Our method has achieved an overall score of 6.74, while BACF and STRCF appeared as the second and third-best performance trackers. Table 6 provides the attribute-based analysis of the TempleColor128 dataset. Based on the results, it is noticeable that our method has achieved an overall score of 5.68, while STRCF and ARCF appeared as second and third-best trackers for these 11 challenging attributes.

Table 7 shows the attribute-based performance over UAV-123 dataset. The overall score of all trackers is less compared to the other benchmarks because this dataset contains aerial videos which are more challenging. Our method shows an overall score of 3.99, while SRDCF and ARCF trackers achieve 3.587 and 3.525 scores, appearing as second and third-best trackers respectively.

4.6 Qualitative Comparison

Fig. 18, demonstrates the qualitative comparison of proposed and other trackers over OTB benchmarks. In the *basketball* video sequence, the target experiences background clutter (BC) because all players wear the same dress. Moreover, the target object (#9 player) faces partial occlusion at frame #655. All aforementioned trackers keep tracking the target object except BACF. For this challenging situation, the bounding box drawn by our method is nearer to the bounding box of the ground truth.

Similarly, the target has a fast motion (FM) attribute at frame #72 in the Bird1 challenging sequence. For this case, all other trackers failed except the proposed tracker which keeps tracking the target despite fast motion. Our tracker tackles the Scale Variation (SV) challenge effectively as shown in frames #83 and #90 of Dragonbaby and shaking

© Mehran University of Engineering and Technology 2024

sequences respectively. For the Dragonbaby sequence, only the proposed tracker performs accurate tracking, other trackers lose the object. For the shaking sequence, DSST, and SRDCF miss the target as shown in #90, #220, and #344. The Soccer sequence suffers from occlusion, motion blur, scale variation, and fast motion attributes. For all these challenges, our tracker successfully tracks the object, while STRCF and ARCF perform poor tracking. Tiger2 sequence experiences occlusion, deformation, background clutter, and in-plane rotation. Our tracker keeps tracking the target accurately, while DSST and KCF trackers failed to perform well in this case.

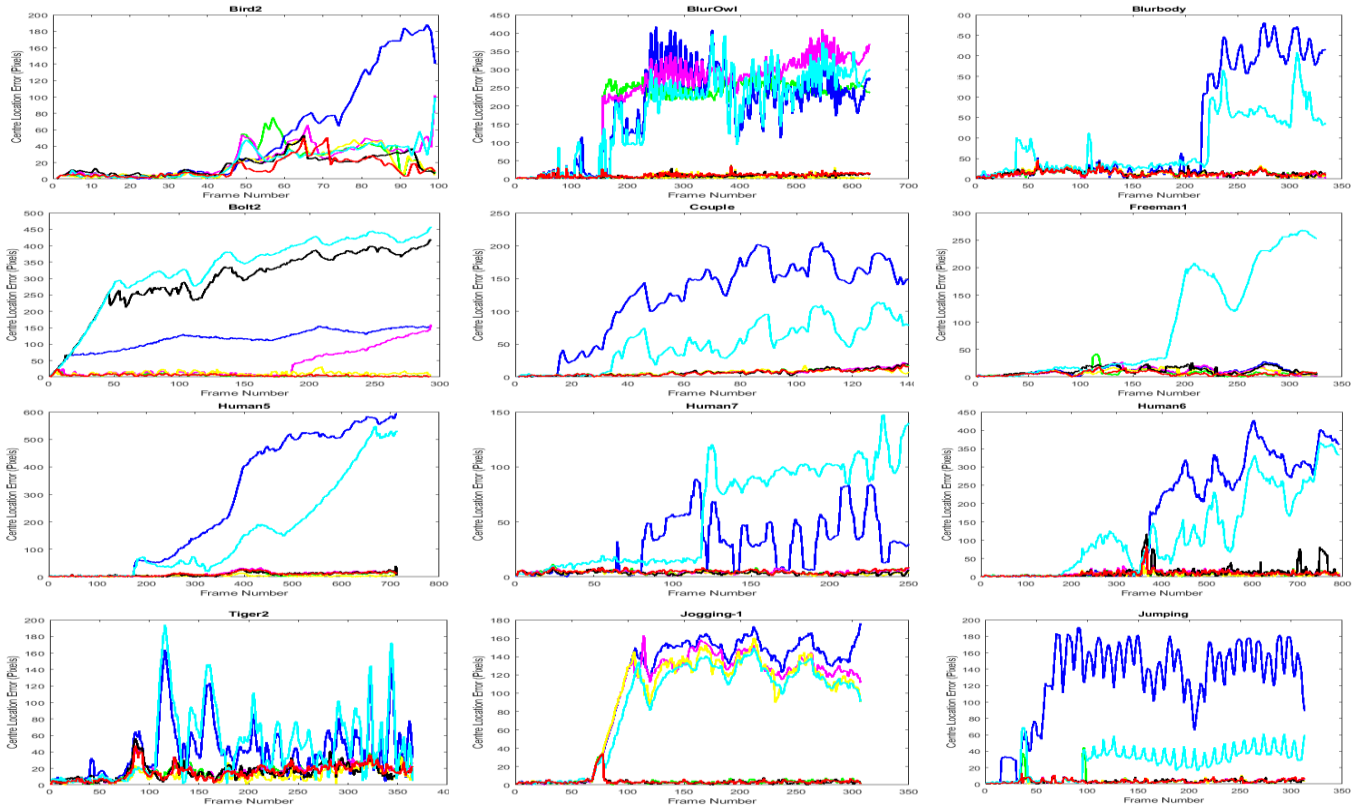

Fig. 17. Comparison Of Centre Location Error (Pixels) On 12 Challenging Videos Of The OTB2015 Dataset

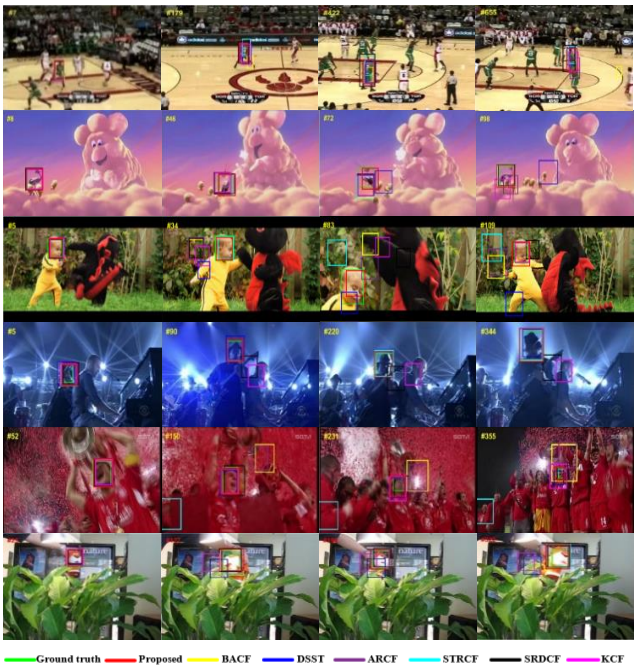

Fig. 18. Tracking Performance Of Our Method With Other Modern Trackers: BACF, DSST, ARCF, STRCF, SRDCF, KCF

A frame-to-frame comparison of different trackers on 12 challenging videos of the OTB2015 dataset is shown in Fig. 17. The CLE of our method is represented by the red line as shown in legends. From all the tracking results, it is clear that our tracker has the least CLE as compared to other trackers. The mean CLE of our tracker over the whole OTB2015 dataset is only 17.74 as shown in Table 2.

5. Conclusion and Future Work

This paper presented an effective object tracker based on the Discriminative Correlation Filter (DCF) framework. Our method is an enhanced version of the ARCF tracker which utilize HoG features only. We proposed an online multifeatured learning-based visual object tracker with an aberrance repression scheme is proposed. Firstly, we developed a robust object appearance model by utilizing the fHOG, CN, Grayscale intensities, and Saliency features. All features are combined effectively with adaptive weights which are calculated using the PSR of each response map individually, to suppress the noise and interference. Secondly, we introduced an adaptive learning rate updation scheme based on the PSR of the final fused response map to alleviate the spurious

© Mehran University of Engineering and Technology 2024

learning in case of occlusion or target loss. ADMM method is utilized to efficiently solve DCF calculations. With the incorporation of adaptive model updation, our proposed tracker efficiently redetected the target after long-term occlusion. Numerous experiments are performed over OTB2013/15, TempleColor128, and UAV123 datasets. Among other trackers, our tracker achieves the least CLE (17.7) on the OTB2015 dataset, which shows the greater accuracy of the proposed method. Both qualitative and quantitative tracking results over other datasets show that the proposed method outperforms the other modern handcrafted visual object trackers.

Although our tracker has shown outstanding results against challenging scenarios still, there are research gaps for improvements like motion estimation, low resolution, etc. To handle these issues more effectively, we will incorporate the deep neural network in future research work.

6. Reference

[1] G. Liu, S. Liu, K. Muhammad, A. K. Sangaiah, and F. Doctor, “Object tracking in vary lighting conditions for fog based intelligent surveillance of public spaces”, IEEE Access, vol. 6, pp. 29283–29296, 2018, doi: https://doi.org/10.1109/access.2018.2834916.

[2] M. Fiaz, A. Mahmood, and S. Jung, “Tracking noisy targets: A review of recent object tracking approaches”, arXiv, vol. 1802, no. 03098, Feb. 2018.

[3] X. Zhang, H. Gao, C. Xue, J. Zhao, and Y. Liu, “Real-time vehicle detection and tracking using improved histogram of gradient features and Kalman filters”, International Journal of Advanced Robotic Systems, vol. 15, no. 1, p. 172988141774994, Jan. 2018, doi: https://doi.org/10.1177/1729881417749949.

[4] C. Martinez, C. Sampedro, A. Chauhan and P. Campoy, "Towards autonomous detection and tracking of electric towers for aerial power line inspection", 2014 International Conference on Unmanned Aircraft Systems (ICUAS), Orlando, FL, USA, 2014, pp. 284-295, doi: 10.1109/ICUAS.2014.6842267

[5] M. Mueller, G. Sharma, N. Smith and B. Ghanem, "Persistent aerial tracking system for UAVs", 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea (South), 2016, pp. 1562-1569, doi: 10.1109/IROS.2016.7759253.

[6] J. Liu and X. Zhong, “An object tracking method based on Mean Shift algorithm with HSV color space and texture features”, Cluster Computing, vol. 22, no. S3, pp. 6079–6090, Feb. 2018, doi: https://doi.org/10.1007/s10586-018-1818-7.

[7] M. Alipour Sormoli, M. Dianati, S. Mozaffari and R. Woodman, "Optical flow based detection and tracking of moving objects for autonomous vehicles", IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 9, pp. 12578-12590, Sept. 2024, doi: 10.1109/TITS.2024.3382495.

[8] M. Danelljan, A. Robinson, F. Khan, and M. Felsberg, “Beyond correlation filters: learning continuous convolution operators for visual tracking”, Springer, pp. 472–488, Aug. 2016, doi: https://doi.org/10.48550/arXiv.1608.03773.

[9] C. Fu, F. Lin, Y. Li, and G. Chen, “Correlation filter-based visual tracking for UAV with online multi-feature learning”, Remote Sensing, vol. 11, no. 5, pp. 549–549, Mar. 2019, doi: https://doi.org/10.3390/rs11050549.

[10] Moksyakov, Y. Wu, Stephen Andrew Gadsden, J. Yawney, and M. AlShabi, “Object detection and tracking with YOLO and the sliding innovation filter”, Sensors, vol. 24, no. 7, pp. 2107–2107, Mar. 2024, doi: https://doi.org/10.3390/s24072107.

[11] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu, “Distractor-aware siamese networks for visual object tracking”, arXiv (Cornell University), Aug. 2018, doi: https://doi.org/10.48550/arxiv.1808.06048.

[12] B. Lee, E. Erdenee, S. Jin, M. Y. Nam, Y. G. Jung, and P. K. Rhee, “Multi-class multi-object tracking using changing point detection”, Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14, Springer, 2016, pp. 68–83.

[13] M. Masood and G. Raja, "Multi-feature integration with adaptive learning based correlation filter for visual object tracking", 2023 International Multi-disciplinary Conference in Emerging Research Trends (IMCERT), Karachi, Pakistan, 2023, pp. 1-4, doi: 10.1109/IMCERT57083.2023.10075225.

[14] D. S. Bolme, J. R. Beveridge, B. A. Draper and Y. M. Lui, "Visual object tracking using adaptive correlation filters", 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 2544-2550, doi: 10.1109/CVPR.2010.5539960.

© Mehran University of Engineering and Technology 2024

[15] D. Yuan, X. Shu, and Z. He, “TRBACF: Learning temporal regularized correlation filters for high performance online visual object tracking”, Journal of Visual Communication and Image Representation, vol. 72, no. 102882, p. 102882, Oct. 2020, doi: https://doi.org/10.1016/j.jvcir.2020.102882.

[16] M. Danelljan, G. Häger, F. S. Khan and M. Felsberg, "Discriminative scale space tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 8, pp. 1561-1575, 1 Aug. 2017, doi: 10.1109/TPAMI.2016.2609928.

[17] S. Zhang, W. Lu, W. Xing, and L. Zhang, “Learning scale-adaptive tight correlation filter for object tracking”, IEEE Transactions on Cybernetics, vol. 50, no. 1, pp. 270–283, Oct. 2018, doi: https://doi.org/10.1109/tcyb.2018.2868782.

[18] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “Exploiting the circulant structure of tracking-by-detection with kernels”, Computer Vision – ECCV 2012, vol. 7575, pp. 702–715, 2012, doi: https://doi.org/10.1007/978-3-642-33765-9_50.

[19] H. Wang, W. Ma, S. Zhang, G. Chen, H. Ge and Y. Du, "Robust visual object tracking with multiple features and reliable re-detection scheme", IEEE Access, vol. 8, pp. 98810-98826, 2020, doi: 10.1109/ACCESS.2020.2997072.

[20] D. Yuan, X. Zhang, J. Liu, and D. Li, “A multiple feature fused model for visual object tracking via correlation filters”, Multimedia Tools and Applications, vol. 78, no. 19, pp. 27271–27290, Jun. 2019, doi: https://doi.org/10.1007/s11042-019-07828-2.

[21] F. Li, C. Tian, W. Zuo, L. Zhang and M. -H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4904-4913, doi: 10.1109/CVPR.2018.00515.

[22] C. Fu, Z. Huang, Y. Li, R. Duan, and P. Lu, “Boundary effect-aware visual tracking for uav with online enhanced background learning and multi-frame consensus verification”, arXiv (Cornell University), Jan. 2019, doi: https://doi.org/10.48550/arxiv.1908.03701.

[23] L. Gao, B. Liu, P. Fu, M. Xu, and J. Li, “Visual tracking via dynamic saliency discriminative correlation filter”, Applied Intelligence, vol. 52, no. 6, pp. 5897–5911, Aug. 2021, doi: https://doi.org/10.1007/s10489-021-02260-2.

[24] M. Danelljan, G. Häger, F. S. Khan and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking", 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 4310-4318, doi: 10.1109/ICCV.2015.490.

[25] H. K. Galoogahi, A. Fagg and S. Lucey, "Learning background-aware correlation filters for visual tracking", 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 1144-1152, doi: 10.1109/ICCV.2017.129.

[26] F. Li, C. Tian, W. Zuo, L. Zhang and M. -H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4904-4913, doi: 10.1109/CVPR.2018.00515.

[27] Y. Li, C. Fu, F. Ding, Z. Huang, and G. Lu, “AutoTrack: towards high-performance visual tracking for UAV with automatic spatio-temporal regularization”, arXiv (Cornell University), Jun. 2020, doi: https://doi.org/10.1109/cvpr42600.2020.01194.

[28] Y. Yuan, J. Chu, L. Leng, J. Miao, and B.-G. Kim, “A scale-adaptive object-tracking algorithm with occlusion detection”, EURASIP Journal on Image and Video Processing, vol. 2020, no. 1, Feb. 2020, doi: https://doi.org/10.1186/s13640-020-0496-6.

[29] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, “Joint group feature selection and discriminative filter learning for robust visual object tracking”, arXiv (Cornell University), pp. 7949–7959, Jan. 2019, doi: https://doi.org/10.48550/arxiv.1907.13242.

[30] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, “ECO: efficient convolution operators for tracking,” arXiv.org, Apr. 10, 2017. https://arxiv.org/abs/1611.09224 (accessed Jul. 30, 2023).

[31] M. Masood and G. Raja, "Multi-feature integration with adaptive learning based correlation filter for visual object tracking", 2023 International Multi-disciplinary Conference in Emerging Research Trends (IMCERT), Karachi, Pakistan, 2023, pp. 1-4, doi: 10.1109/IMCERT57083.2023.10075225.

[32] Z. Ji, K. Feng, Y. Qian, and J. Liang, “Sparse regularized correlation filter for UAV object tracking with adaptive contextual learning and

keyfilter selection", Information Sciences, vol. 658, pp. 120013–120013, Dec. 2023, doi: https://doi.org/10.1016/j.ins.2023.120013.

[33] S. Ma, B. Zhao, Z. Hou, W. Yu, L. Pu, and X. Yang, "SOCF: A correlation filter for real-time UAV tracking based on spatial disturbance suppression and object saliency-aware", Expert Systems with Applications, vol. 238, pp. 122131–122131, Oct. 2023, doi: https://doi.org/10.1016/j.eswa.2023.122131.

[34] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627–1645, Sep. 2010, doi: https://doi.org/10.1109/tpami.2009.167.

[35] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach", 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 2007, pp. 1-8, doi: 10.1109/CVPR.2007.383267.

[36] Y. Wu, J. Lim and M. -H. Yang, "Online object tracking: A benchmark", 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 2411-2418, doi: 10.1109/CVPR.2013.312.

[37] Y. Wu, J. Lim and M. -H. Yang, "Object tracking benchmark", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1834-1848, 1 Sept. 2015, doi: 10.1109/TPAMI.2014.2388226

[38] P. Liang, E. Blasch and H. Ling, "Encoding color information for visual tracking: Algorithms and benchmark", IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5630-5644, Dec. 2015, doi: 10.1109/TIP.2015.2482905.

[39] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking", Computer Vision – ECCV 2016, vol. 24, no. 12, pp. 445–461, 2016, doi: https://doi.org/10.1007/978-3-319-46448-0_27.

[40] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning aberrance repressed correlation filters for real-time UAV tracking", arXiv (Cornell University), Jan. 2019, doi: https://doi.org/10.48550/arxiv.1908.02231.

[41] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration", Springer, vol. 8926, pp. 254–265, Jan. 2015, doi: https://doi.org/10.1007/978-3-319-16181-5_18.

[42] Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov and D. Tao, "Multi-store tracker (MUSTer): A cognitive psychology inspired approach to object tracking", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 749-758, doi: 10.1109/CVPR.2015.7298675

[43] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 7, pp. 1409–1422, Jul. 2012, doi: https://doi.org/10.1109/tpami.2011.239

[44] C. Ma, X. Yang, Chongyang Zhang and M. -H. Yang, "Long-term correlation tracking", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 5388-5396, doi: 10.1109/CVPR.2015.7299177.

[45] J. Zhang, S. Ma, and S. Sclaroff, "MEEM: Robust tracking via multiple experts using entropy minimization", Lecture Notes in Computer Science, pp. 188–203, 2014, doi: https://doi.org/10.1007/978-3-319-10599-4_13.

[46] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik and P. H. S. Torr, "Staple: Complementary learners for real-time tracking", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 1401-1409, doi: 10.1109/CVPR.2016.156.

[47] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed tracking with kernelized correlation filters", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 583-596, 1 March 2015, doi: 10.1109/TPAMI.2014.2345390.