Leveraging Attention for Improved Discriminative Correlation Filters in Single Object Tracking

Ahmed O. Elsaid[1†], Mohamed M. Fouad[2†], Tarek S. Ghoniemy[3*]

[1,3]Department of Computer Engineering & Artificial Intelligence, Military Technical College, Cairo, Egypt.
[2]Faculty of Information Systems & Computer Science, October 6*th* University, Central axis, Giza, Egypt.

*Corresponding author(s). E-mail(s): ghoniemy_t@mtc.edu.eg;
Contributing authors: ahmed.osaz@gmail.com; mmafoad.cs@o6u.edu.eg;
†These authors contributed equally to this work.

orrelation Filter-based Trackers have shown impressive results in the object tracking research, outperforming classical trackers in several benchmarks. However, accurately tracking objects with deformation, fast motion, or occlusion remains a main challenge in the process of tracking. The cyclic suggestion of training samples used in correlation filter tracking usually lead to undesirable boundary effects, that significantly reduce the tracking efficiency. To address this issues, a hybrid attention-based correlation filter method (Att-DCF) is proposed for robust object tracking. By developing a discrimination filter and reducing the limitations of irrelevant features, the proposed approach provides more accurate and consistent estimated target locations. The proposed Att-DCF tracker is evaluated using the recent and standard datasets, OTB-100, Temple-Colour128, and UAV123. Compared to the baseline DCF tracker, the proposed tracker achieves an improvement of 2.71% and 1.38% in terms of area under the curve and precision measures using the OTB-100 dataset, 3.99% and 3.43% using the Temple-Colour128, and 4.72% and 1.76% using the UAV123, respectively.
keywords: Visual tracking; attention; correlation filter.

1 Introduction

Computer vision and intelligent systems are rapidly developing in many areas such as target detection, recognition and tracking [1]. Object detection involves tracking consecutive frames in a video sequence while preserving the size and position of the objects. Object tracking is widely used in human computer interaction, medical image analysis, and military operations. This field has been extensively researched and computer vision and intelligent systems have many specific applications such as: (1) Tracking and capturing objects using unmanned vehicles (UAVs) [2], such as equipment for military operations. Object tracking is also used to track and identify an object to avoid collision [1]. (2) Smart changes have occurred in the industry and recently product tracking has been adopted in production [3]. (3) Intelligent Transportation System (ITS) uses a tracking system to monitor and track vehicles in real time using tracking devices [4].

There are two main types of tracking techniques: Generative ones involve training the tracking model to represent the object appearance and then finding the object that best matches the object appearance [5]. This is usually done using particle [6], Kalman filter [7], or mean-shift [8]. Discrimination techniques, on the other hand, attempts to separate the target from the surrounding environment to achieve higher accuracy using DCF-based trackers [9]. The correlation filter (CF) is designed to use a circular sliding window over the test sample, assuming that the sample will expand over time. Using Fast Fourier Transform (FFT), the time hypothesis leads to rapid training and detection [10]. First, the required search is narrow and the cosine window generally suppresses boundary effects resulting from migration [11]. Therefore, it may cause tracking drift, especially in the case of occlusion. Second, even with partial occlusion, the target area still utilize the response map to determine the location of the moving object. The limitations of the traditional DCF model mentioned above will affect the tracking accuracy as follows:

- DCF-based trackers [9] will encounter problems in high-speed situations due to the limited search space.
- Lack of inadequate training patches affects the accuracy of the presence of large changes [12].
- The above limitations in training reduce the tracker's ability to re-identify objects under occlusion [13].
- Using a larger time frame is associated with an increase in negative images used for training. That's why good training models include a lot of background information. This information spread reduces the model discrimination ability, resulting in poor tracking accuracy

. A hybrid approach for visual tracking is proposed, that integrates the attention technique into the LADCF tracker [14]. In particular, spatial attention techniques are introduced that focus on distinctive features of the target, that in turn discard most of the irrelevant features.

The rest of this manuscript is organized as follows. Sec. 2 overviews the conventional correlation filters techniques in visual tracking. In Section 3, the proposed tracker is presented that utilizes the visual attention and the criteria for selecting relevant information. Section 4 provides objective as well as subjective evaluations of the proposed tracker with a discussion of its effectiveness. Finally, conclusions are drawn in Section 5.

2 Related Work

In traditional CF tracking techniques, an image of a specific target object extracted from the very initial frame is employed for training, then the pattern is located in the incoming frames. The DFT usually perform correlation by replacing element-wise arithmetic operations with overloaded convolutions. After the correlation process, an inverse FFT is used to generate a spatial confidence (response) map. The new object location is estimated as the location with the highest score in response map. To train and update the CF, the appearance at predicted position is employed.

CF technique was first appeared in MOSSE filter [15] which uses gray-scale images to acquire data. The basic idea behind correlation is that two signals that have a higher correlation are more similar. The optimization process can be simplified as correlation diagram (g), filter (h) and input (c); the result is g $= c{*}h$, where $*$ denotes the circulate correlation. FFT is used to speed up the calculation. In this area the corresponding map is represented by G = C.H; where C and H are the Fourier transforms of c and h, respectively. However, this method cannot guarantee accurate tracking when the target template appearance changes greatly..

Henriques et al. presented the circular Tracking with Kernels (CSK) which uses a dense sample instead of the sparse one used by the MOSSE method [16]. However, this causes the calculation to increase and hence affect the computational efficiency. To solve this problem, CSK method adopted FFT to speed up the process. Danwelljan et al. get. introduced the Kernel Correlation Filter (KCF), which improves accuracy by using multiple-level features [17]. However, the object rotation, get out of view and moving fast are still challenging.

To solve the different-scale variation problem, Danelljan et al. proposed a discriminative Scale Space tracking system (DSST) using pyramid features [18]. The fDSST method has been proposed as a faster DSST that uses smoothing and interpolation techniques to speed up the process [19]. The fDSST method increases the the search area, improves tracking accuracy, and uses PCA features to reduce the size.

The SRDCF algorithm [20] solves the boundary problem that arises in CF by incorporating a regularization term. SRDCF solves the conversion problem by dividing the conversion scale into multiple scales. However, it is sensitive ti the training process. To solve this problem, STRCF [21] uses the alternating direction multiplier method (ADMM) algorithm to provide a closed solution for each problem and remove bounds by adding 5 acceleration when using manual features. The SRDCFDecon [22] algorithm improves the model computational cost and learning of SRDCF technology and saves early data to avoid model drift due to CF defective CF samples.

STAPLE algorithm proposed by L. Bertinetto et al. [23] combines HOG and color features at 80 fps, it solves the problem of insufficient deformation robustness, outperforming other tracking accuracy methods. The LMCF method [24] combines CF methods with SVM to improve discrimination while using multi-modal forward recognition to avoid interference from similar objects. The DCF model mentioned above has limitations that affect tracking accuracy in he form of limited search area that makes it difficult for the tracker to keep up with fast moving targets, over-fitting due to reduced negative sample training, and inability to recover from occlusion.

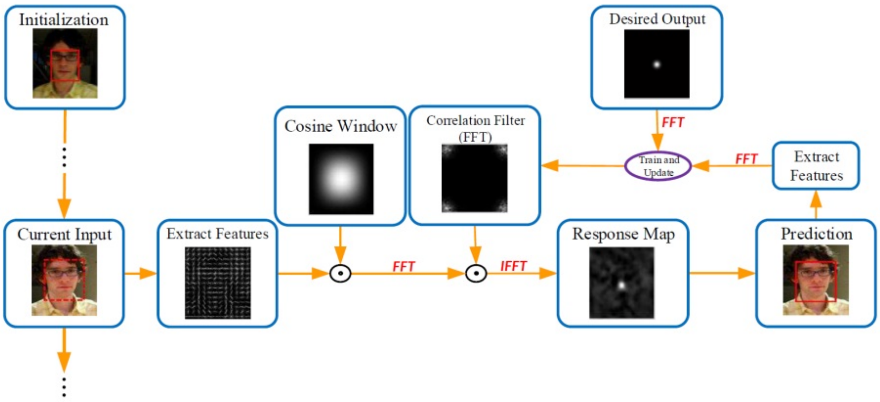


Fig. 1 Steps involved in typical correlation filter-based object tracking technique [25].

3 Methodology

CF tracking (CFT) is a technique used in computer vision to track objects in video sequences [25]. The process usually begins by feeding the CF with a target patch from the first frame as Fig. 1 illustrates. The image pattern of the estimated position in the previous frame is then used to initialize next frame. Visual features are then extracted from the input image and boundary effects are smoothed using cosine windows. Convolution and FFT are used to calculate the correlation between the current input and the learned filter. The results were converted back to spatial reference using inverse FFT and a confidence map was created. The maximum of this map is used to determine the approximate target location, and this information is used to update the CF.

The simple idea is to first use the relationship between candidate objects and object models (filters) in the spectral domain to find the most compatible model in order to be faster. While continuous coupling is used to generate input data, the basic DCF method requires a good relationship between image blocks, resulting in boundary interference and variation variation, thus limiting the sample to the small search space in Figure Fig. 2.

The CF phase usually includes sample as well as filter training. The cyclic matrix $X = F\,\mathrm{diag}(\hat{x})F^H$ is similar to the cyclic transformation model with the positive model and the design x. The F denotes the discrete Fourier matrix; $\hat{x} = Fx$ stands for the simple Fourier conjugate transpose of x and F^H is the transform of F. The filter training compromises a ridge regression problem created using the equation $f(z) = w^\top z$ that in turn reduces the sum of squared differences between predicted values and actual values as shown in (1). Similarly, the objective regression y_i. The edges of the object are specified in the first frame of phase classification and follow the regression target on the 2D Gaussian data. External features of the target are obtained

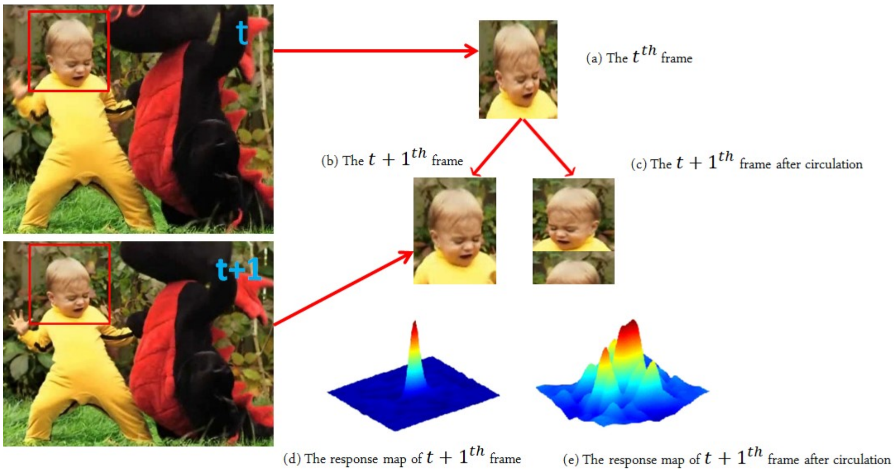


Fig. 2 Circular shifts don't actually represent real translations [25].

by knowing the target boundary and extracting the HOG features.

$$\min_{w} \sum_{i} (A_0 w - y_i)^2 + \lambda \|w\|^2, \tag{1}$$

where A_o denotes the object area's extracted features, and λ is a filter hyper parameter, and w represents the filter. **SRDCF** incorporates an additional regular coefficient, as

$$\min_{w} \|A_0 w - y\|_2^2 + \lambda \|s \odot w\|_2^2, \tag{2}$$

where s represents the spatial attention element, and the penalty increased as we moves away from the target area, $\odot$ represents vector-element multiplication. Then, in **CFCA**, the global background info is considered and incorporated in the learning process, as

$$\min_{w} \|A_0 w - y\|_2^2 + \lambda_1 \|w\|_2^2 + \lambda_2 \sum_{i=1}^{k} \|A_i w\|_2^2, \tag{3}$$

where A_i represents the context feature information of the cropped patch, λ_1, λ_2 are hyper parameters, and k indicates the number of context patches. The framework for target response y_o generation is given as

$$\min_{w} \|A_0 w - y\|_2^2 + \lambda_1 \|w\|_2^2 + \lambda_2 \|y - y_0\|_2^2, \tag{4}$$

The tracking accuracy is affected by the boundary effect, but CF-based tracker is largely rely on the spatial characteristics for training filters. Accordingly, the spatial attention is presented to achieve as much precise position information and lessen the dismissal of features.

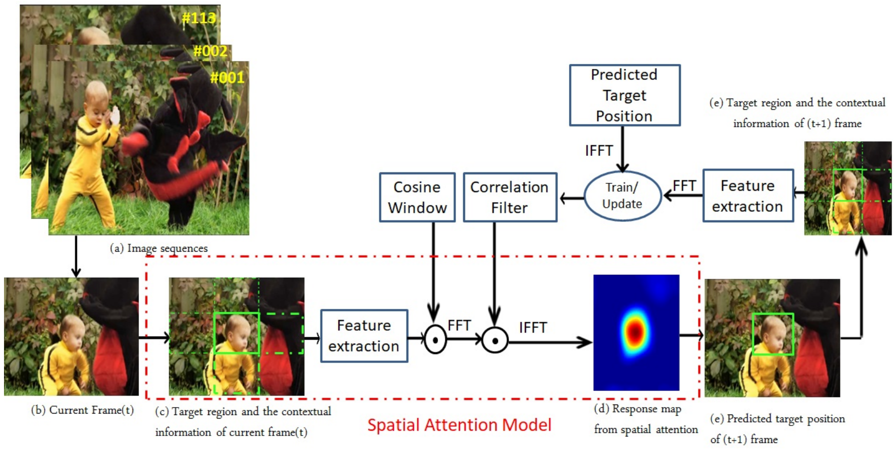


Fig. 3 The framework of the proposed tracker Att-DCF.

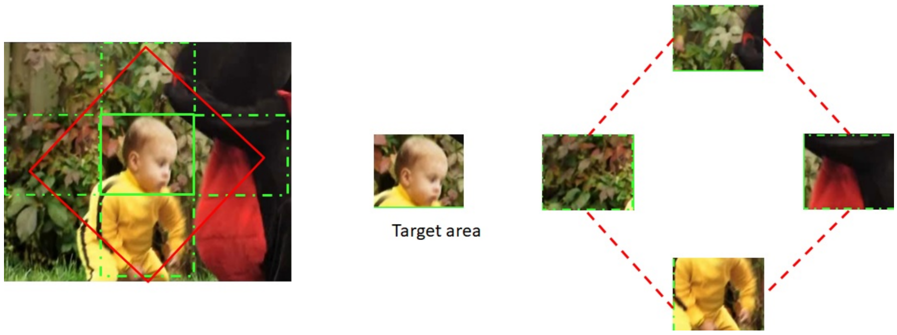


Fig. 4 Context data gathering: 4 areas near selected object region.

4 The Proposed Approach.

In this paper, we propose a method that uses spatial attention to suppress an important component of an object and focus on its important components, as shown in Figure Fig. 3. Irrelevant content can be reduced by using visualization. Data points around the target contain important information. However, this information may not be relevant. To obtain specific points, we select four point patches near the target, as shown in Figure Fig. 4. By combining these spatial tracking methods, our scheme can focus on specific features of the target, thus improving tracking accuracy. The target features plays a crucial role in target discrimination. The bounding area selection during extraction of features affect the target detection and tracking. The computational complexity increases with the target area and as well, the impact of the boundary effect grows. The spatial attention part of the equations $s : \Omega \to R$ is adopted to penalize the filter quotient area. The target area is important to decide which spatial features is more attentive. Features located close to the target's boundary are often less reliable than those closer to the centroid of target. The spatial attention is adopted To address this issue [26] that suppress the background information while assigns more weights to the central regions, which significantly lessens the learned classifier's reliance on background information while focuses on weights of the target region, the formula can be given as,

$$\min_{w} \|A_0 w - y\|_2^2 + \lambda_1 \|s \odot w\|_2^2 + \lambda_2 \|y - y_0\|_2^2 + \lambda_3 \sum_{i=1}^{k} \|A_i w\|_2^2 . \tag{5}$$

Equation (5) combines the target features A_o, and the context patch A_i, to produce a filter w, for object tracking using CF. The equation incorporates a spatial attention component s, which represents the importance of distinctive parts of the context patch as compared to the target area. The λ_1,λ_2, and λ_3 parameters are adopted to control the target, context patch's, and the spatial attention component contributions, respectively. The term μ represents constant and η represents scaling factor, and can be formulated as,

$$s(m,n) = \mu + \eta(m/P)^2 + \eta(n/Q)^2. \tag{6}$$

The component s, is calculated using the quadratic equation, in which (m, n) pair is the coordinates of a pixel within the context patch, and (P, Q) pair is the context patch dimension. The outcome of the regularizer is set to η.=3 and the values of s is set to μ=0.1. The filter w, and a are given as,

$$\hat{w} = \frac{\lambda_2 \left(\hat{a}_0 \odot \hat{y}_0\right)}{\lambda_2 \left(\hat{a}_0 \odot \hat{a}_0^*\right) + (1+\lambda_2)\left(\lambda_1 \hat{s} \odot \hat{s}^* + \lambda_3 \sum_{i=1}^{k} a_i \odot \hat{a}_0^*\right)} \tag{7}$$

$$\hat{a} = \frac{\lambda_2 \hat{y}_0}{\lambda_2 \left(\hat{a}_0 \odot \hat{a}_0^*\right) + (1+\lambda_2)\left(\lambda_1 \hat{s} \odot \hat{s}^* + \lambda_3 \sum_{i=1}^{k} \hat{a}_i \odot \hat{a}_0^*\right)} \tag{8}$$

The video sequence frame [t] and the initial sampling position $[p_s]$ are the inputs, the frame number is t^h's and the estimated output is p_t[t].

5 Experiments and Results

5.1 Implementation Setup

An Intel Core i7, 2.20 GHz Processor is used to implement the proposed Att-DCF in MATLAB Version 2018A. All features have been defined as mixture of both the color feature as well as the spatial gradient histogram feature (HOG) [27] (Color Name, CN).

5.2 Dataset Description.

Our comprehensive benchmarks are used to measure our tracking effectiveness. Temple Color [28], OTB100 [29] and UAV123 [30]. First, the OTB100 dataset contains 100 annotated movie sequences and 11 sequence features. Second, the UAV123 dataset contains 123 complex sequences, and the temple color dataset contains 128 color video sequences.

5.3 Evaluation Metrics

In our experiments, first, the one pass evaluation (OPE) [31] metric is employed to evaluate the tracking accuracy. It incorporates two main characteristics, precision and success plots to measure the accuracy of the trackers. These plots are generated by comparing the center location error and bounding box (BB) overlap between the predicted location of the object and the ground truth location. As well, the area under curve (AUC) and the overlap precision (OP) metrics have been used to evaluate the percentage of overlap ratios that exceeds 0.5. The AUC metric is calculated as the ratio of the intersection of the estimated BB and the ground truth BB to the union of the both, as

$$AUC = \frac{A_t^G \cap A_t^T}{A_t^G \cup A_t^T}, \tag{9}$$

where A_t^G is the ground truth BB, A_t^T represents the estimated BB. The higher the ratio, the better the accuracy of the tracker. Fourth, the distance precision (DP) metric is used to measure the percentage of location errors in the range of 20 pixels. The Precision (the higher the better) is defined as the ratio between the TP and the number of all detections as,

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

where the true positive (TP) and false positive (FP) represent the number of correct and incorrect detections made by the model respectively. The proposed Att-DCF method offers a reliable and accurate solution for visual object tracking, with significant improvements in accuracy compared to existing methods. Our method's 17.8% fps makes it a suitable for real-time applications, which is essential for various real-world scenarios.

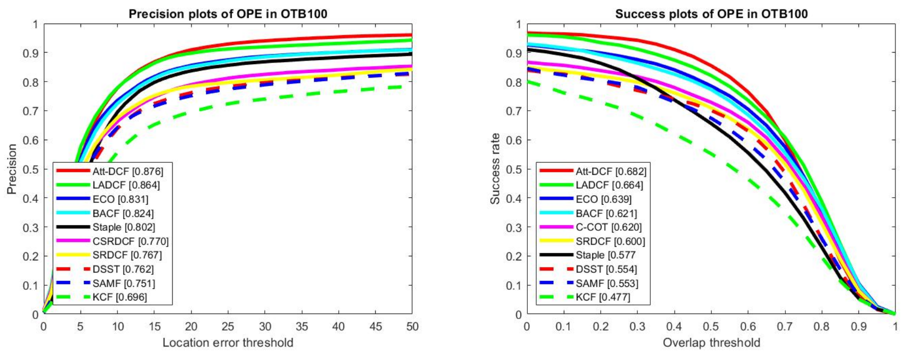
5.4 Results and Discussion

5.4.1 Experimental Results of OTB100 Dataset

In a comprehensive comparison, the proposed Att-DCF approach is evaluated against 9 recent top trackers: LADCF [14], Eeco [32], BACF [33], C-cot [34], SRDCF [20], STAPLE [23], DSST [18], SAMF [35], and KCF [17]. The evaluation is adopted using the One Pass Evaluation (OPE) [36], which uses distance precision and overlaps metrics to measure the trackers' accuracy. Using both distance precision and overlap metrics for OPE, the results show that the proposed tracker outperforms competing trackers in terms of both metrics as shown in Table 1, and Fig. 7.

Table 1 AUC and Precision results of recent trackers on OTB-100.

Tracker	AUC ↑	Precision ↑	Tracker	AUC ↑	Precision ↑
proposed (Att-DCF)	0.682	0.876	LADCF [14]	0.664	0.864
ECO [32]	0.639	0.831	BACF [33]	0.621	0.824
C-cot [34]	0.620	0.770	SRDCF [20]	0.600	0.767
STAPLE [23]	0.577	0.802	DSST [18]	0.554	0.762
SAMF [35]	0.553	0.751	KCF [17]	0.477	0.696

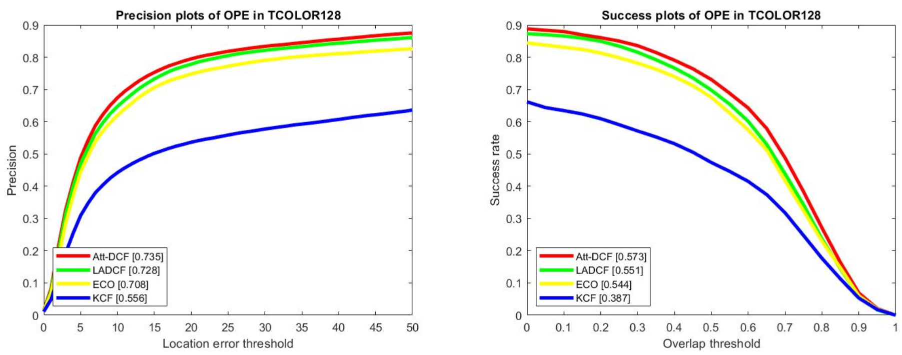


(a) (b)

Fig. 5 Precision and Success plots using OTB100 dataset

5.4.2 Experimental Results of Temple-Colour-128 Dataset

In a qualitative comparison, the application (Att-DCF) was compared with status monitors: LADCF[14], ECO[32], and KCF[17]. A one-pass evaluation using distance accuracy and overlapping performance tests shows that the tracker outperforms competing trackers on both metrics, as shown in Table Table 2 and Figure Fig. 6.



(a) (b)

Fig. 6 Precision and Success plots using Temple-Colour 128 dataset.

Table 2 AUC and Precision metrics using Temple-Colour 128 dataset.

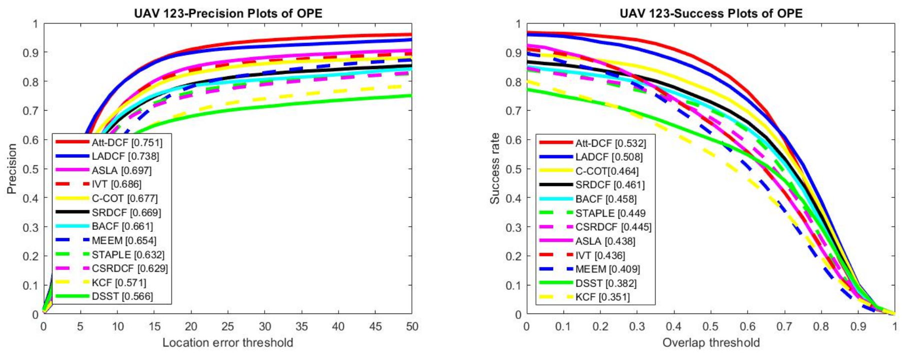
Tracker	Precision ↑	AUC ↑	Tracker	Precision ↑	AUC ↑
proposed (Att-DCF)	0.735	0.573	LADCF [14]	0.728	0.551
ECO [32]	0.708	0.544	KCF [17]	0.556	0.387

5.4.3 Experimental Results of UAV-123 Dataset

In comparison, we validate our relation discrimination (Att-DCF) against 11 recent top trackers: LADCF [14], ASLA [37], IVT [38], C-cot [34], SRDCF [20], BACF [33], MEEM [39], STAPLE [23], CSRDCF [40], DSST [18], and KCF [17]. A one-pass evaluation using distance accuracy and overlap performance testing shows that the tracker outperforms competing trackers in both tests, as shown in Table Table 3 and Figure Fig. 7. Additional testing was also performed to evaluate the proposed Att-DCF in terms of size and flexibility. In particular, the UAV123 dataset [30], which includes large-scale and interlaced videos, is used. The results show that Att-DCF outperforms the benchmark, with AUC and accuracy improving by 5.2% and 3.8% in AUC and precision, respectively.

Table 3 AUC and Precision plots using UAV123 dataset.

Tracker	Precision ↑	AUC ↑	Tracker	Precision ↑	AUC ↑
proposed (Att-DCF)	0.751	0.532	LADCF [14]	0.738	0.508
ASLA [37]	0.697	0.438	IVT [38]	0.686	0.436
C-cot [34]	0.677	0.464	SRDCF [20]	0.669	0.461
BACF [33]	0.661	0.458	MEEM [39]	0.654	0.409
STAPLE [23]	0.632	0.449	CSRDCF [40]	0.629	0.445
DSST [18]	0.566	0.382	KCF [17]	0.571	0.351



(a) (b)

Fig. 7 Precision and Success plots using UAV123 dataset.

5.4.4 Qualitative Analysis

Figure Fig. 8 compares proposed tracker's accuracy using a correlation filter and a local object with several state-of-the-art trackers. The results are shown in two different videos: Dragonbaby and Bird1, chosen because they offer unique challenges regarding acceleration, closing, and transitions. The purpose of the comparison is to evaluate the ability of the process to overcome these problems and track products in real life.

The results show that the method can track objects in difficult situations with rapid movements, occlusions, and measurement changes, which are common problems in the field of object tracking. The use of these objective measurements provides a rigorous scientific method for evaluating the accuracy of a method. Figure Fig. 8 shows not only this comparison but also the results of the various monitoring strategies compared. It shows the accuracy of LADCF [14], ECO [32], BACF [33], C-cot [34], SRDCF [20], STAPLE [23], DSST [18], SAMF [35], and KCF [17] as well as the proposed Att-DCF approach on various challenging sequences, including those with rapid changes in the target's appearance change, object-deformations, out-of-view, partial-occlusions, and in-plane and out-of-plane rotations.

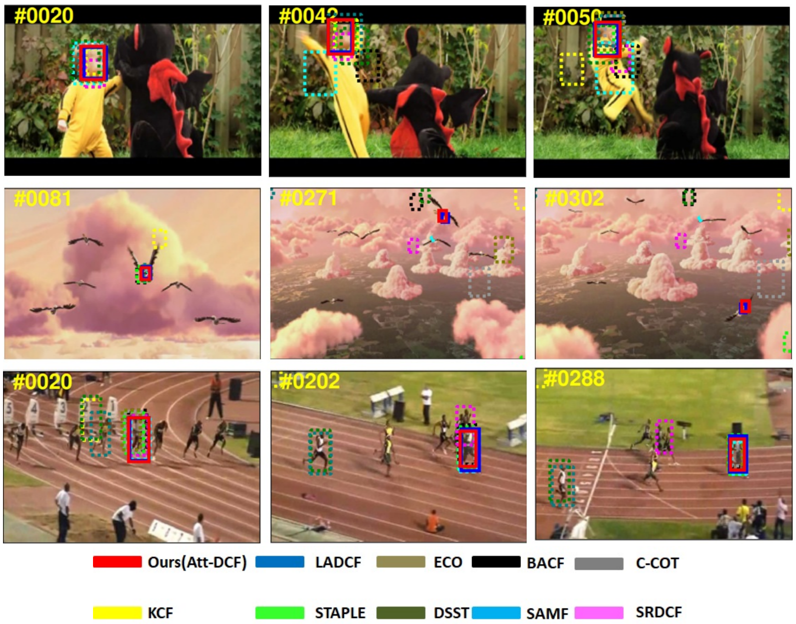


Fig. 8 Subjective analysis on challenges in video sequences.

6 Conclusions

In summary, an object tracking method (Att-DCF) that is robust to deformations and changes in object shape is proposed. By combining tracking methods that consider event data and use spatial tracking, both target discrimination and tracking reliability are improved. Att-DCF is evaluated on a wide range of measurement data and achieves good state-of-the-art performance compared to other existing trackers. This method extracts four dot patches near the target to obtain information about the dot and uses spatial color to improve the accuracy of location estimation in terms of semantics, texture, and context. The proposed approach outperforms other DCF-based trackers, yielding improved AUC and precision by an average increase of 2.71% and 1.38% using the OTB-100 dataset, by 3.99% and 3.43% using the Temple-Colour128 dataset, and by 4.72% and 1.76% using the UAV123 dataset, respectively.

References

[1] Wu, Q. & Zhou, Y. The evolution of video surveillance: an overview. *Machine Vision and Applications* **19**, 574–579 (2019).

[2] Huang, Z., Fu, C., Li, Y., Lin, F. & Lu, P. Learning aberrance repressed correlation filters for real-time uav tracking. *IEEE/CVF International Conference on Computer Vision* 2891–2900 (2019).

[3] Choudhry, N., Abawajy, J., Huda, S. & Rao, I. A comprehensive survey of machine learning methods for surveillance videos anomaly detection. *IEEE Access* **11**, 114680–114713 (2023).

[4] Yang, Z. & Cheng, L. Vehicle detection in intelligent transportation systems and its applications under varying environments: A review. *Image and Vision Computing* **69**, 143–154 (2018).

[5] Luo, H., Guo, Y. & Pan, T.-S. Target detection and recognition for wide area border defense intelligent surveillance. *Journal of Information Hiding and Multimedia Signal Processing* **11**, 199–208 (2020).

[6] Luo, Y., Xu, J., Wang, P. & Jiang, D. Improved particle filtering based motion target detection method. *International Conference on Intelligent Computing and Signal Processing* 1137–1140 (2023).

[7] Li, Y. *et al.* Adaptive kernel learning kalman filtering with application to model-free maneuvering target tracking. *IEEE Access* **10**, 78088–78101 (2022).

[8] Li, H., Cui, G., Kong, L., Guo, S. & Wang, M. Scale-adaptive human target tracking for through-wall imaging radar. *IEEE Geoscience and Remote Sensing Letters* **17**, 1348–1352 (2020).

[9] Mandal, M. & Vipparthi, S. An empirical review of deep learning frameworks for change detection: Model design, experimental frameworks, challenges and research needs. *IEEE Trans. on Intell. Transp. Sys.* **23**, 6101–6122 (2022).

[10] TANG, M., ZHENG, L., YU, B. & WANG, J. Fast kernelized correlation filter without boundary effect. *IEEE Conf. on App. of Comp. Vis.* 2998–3007 (2021).

[11] Zhou, W. *et al.* Siamcan: Real-time visual tracking based on siamese center-aware network. *IEEE Transactions on Image Processing* **30**, 3597–3609 (2021).

[12] Dai, K., Wang, D., Lu, H., Sun, C. & Li, J. Visual tracking via adaptive spatially-regularized correlation filters. *IEEE/CVF Conference on Computer Vision and*

Pattern Recognition 4670–4679 (2019).

[13] Chen, L., Zhu, X. & Tang, B. Visual long-term target tracking method based on uncertain estimation principle. *International Conference on Frontiers of Electronics, Information and Computation Technologies* 595–598 (2022).

[14] 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. on Image Proc.* 5596–5609 (2019).

[15] Zhang, J., He, Y. & Wang, S. Learning adaptive sparse spatially-regularized correlation filters for visual tracking. *IEEE Signal Proc. Letters* **30**, 11–15 (2023).

[16] Li, Y. & Zhu, J. A scale adaptive kernel correlation filter tracker with feature integration. *Computer Vision - ECCV 2014 Workshops* **8926**, 254–265 (2015).

[17] Peng, F., Xu, Q., Li, Y., Zheng, M. & Su, H. Improved kernel correlation filter based moving target tracking for robot grasping. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–12 (2022).

[18] Danelljan, M., Häger, G., Khan, F. & Felsberg, M. Discriminative scale space tracking. *IEEE Trans. on Pat. Ana. and Machine Intell.* **39**, 1561–1575 (2017).

[19] Walid, W., Awais, M., Ahmed, A., Masera, G. & Martina, M. "real-time implementation of fast discriminative scale space tracking algorithm". *Journal of Real-Time Image Processing* **18**, 2347–2360 (2021).

[20] Han, R., Feng, W. & Wang, S. Fast learning of spatially regularized and content aware correlation filter for visual tracking. *IEEE Transactions on Image Processing* **29**, 7128–7140 (2020).

[21] Li, F., Tian, C., Zuo, W., Zhang, L. & Yang, M.-H. Learning spatial-temporal regularized correlation filters for visual tracking. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* 4904–4913 (2018).

[22] Tang, Z.-Q. & Arakawa, K. Visual tracking via spatial-temporal regularized correlation filters with advanced state estimation. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* 1145–1149 (2020).

[23] Bertinetto, L., Valmadre, J., Golodetz, S., Miksik, O. & Torr, P. H. S. Staple: Complementary learners for real-time tracking. *IEEE Conference on Computer Vision and Pattern Recognition* 1401–1409 (2016).

[24] Wang, M., Liu, Y. & Huang, Z. Large margin object tracking with circulant feature maps. *IEEE Conf. on Comp. Vis. & Pat. Recog.* 4800–4808 (2017).

[25] Zhang, Y. & Zheng, Y. Object tracking in uav videos by multi-feature correlation filters with an adaptive update strategy. *International Conference on High*

Performance Big Data and Intelligent Systems 233–238 (2022).

[26] Elsaid, A. O., Fouad, M. M. & Ghoniemy, T. E. Att-siammask: Attention-based network for enhanced visual object tracking. *Journal of Information Hiding and Multimedia Signal Processing* **13**, 22–38 (2022).

[27] Zeng, X., Zhang, L., Luo, Z. & Xiong, X. An improved visual tracking approach based on hierarchical convolutional features. *IEEE Asia Conference on Information Engineering* 42–46 (2021).

[28] Zhang, Y., Yu, Y.-F., Chen, L. & Ding, W. Robust correlation filter learning with continuously weighted dynamic response for uav visual tracking. *IEEE Transactions on Geoscience and Remote Sensing* **61**, 1–14 (2023).

[29] Lai, P. *et al.* Ncsiam: Reliable matching via neighborhood consensus for siamese-based object tracking. *IEEE Trans. on Image Proc.* **32**, 6168–6182 (2023).

[30] Keawboontan, T. & Thammawichai, M. Toward real-time uav multi-target tracking using joint detection and tracking. *IEEE Access* **11**, 65238–65254 (2023).

[31] Zhang, J. & Tao, T. Long-term single-target pedestrian tracking combined with reliability discrimination and re-detection. *International Conference on Intelligent Computing and Signal Processing* 1604–1608 (2023).

[32] Danelljan, M., Bhat, G., Khan, F. S. & Felsberg, M. ECO: Efficient convolution operators for tracking. *IEEE International Conference on Computer Vision and Pattern Recognition* 6931–6939 (2017).

[33] Galoogahi, H. K., Fagg, A. & Lucey, S. Learning background-aware correlation filters for visual tracking. *IEEE Intern. Conf. on Comp. Vis.* 1144–1152 (2017).

[34] Zhou, Y. & Wu, Q. An anti-occlusion tracking system for uav imagery based on discriminative scale space tracker and optical flow. *IEEE International Conference on Advanced Robotics and Mechatronics* 616–621 (2019).

[35] Jiang, S., Li, S. & Zhu, C. Kernelized correlation filter tracking with scale adaptive filter and feature integration. *IEEE Intern. Conf. on Computer and Communications* 201–206 (2018).

[36] González, J., Castillo, H., Sánchez, J. & Murillo, M. Assessment of computer vision methods for motion tracking of planar mechanisms. *Journal of Mechanical Engineering Science* **236**, 4093–4104 (2022).

[37] Javed, S., Mahmood, A., Dias, J. & Werghi, N. Structural low-rank tracking. *IEEE Intern. Conf. on Advanced Video and Signal Based Surveillance* 1–8 (2019).

[38] Zhang, T., Xu, C. & Yang, M.-H. Robust structural sparse tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**, 473–486 (2019).

[39] Fang, Z., Cao, Z., Xiao, Y., Gong, K. & Yuan, J. Mat: Multianchor visual tracking with selective search region. *IEEE Trans. on Cybernetics* **52**, 7136–7150 (2022).

[40] Lukezic, A., Vojir, T., Cehovin Zajc, L., Matas, J. & Kristan, M. Discriminative correlation filter with channel and spatial reliability. *IEEE Conference on Computer Vision and Pattern Recognition* **126**, 671–688 (2018).