1. Ramdane TAGLOUT, 2. Bilal SAOUD

DOI: 10.15199/48.2025.06.12

Artificial bee colony algorithm combined with CNN, Kalman filter used for visual tracking

Sztuczny algorytm kolonii pszczół połączony z CNN i filtrem Kalmana używany do śledzenia wizualnego

Abstract: *Visual object tracking constitutes a significant focal point in computer vision research, with applications extending to various fields such as transportation and smart industry technologies. While correlation filter-based tracking systems have exhibited strong performances, they encounter challenges such as scale variation, occlusion, boundary effects, and background clutter. To address these issues, we have introduced an approach that combines Artificial Bee Colony optimization (ABC), deep neural networks and the Kalman filter. Our method involves an initial stage where the tracking process reliability is analyzed, followed by the calculation of target confidence information at the current estimated location. Subsequently, an updating stage is employed. In our approach, the adaptation of the target scale is determined through Artificial Bee Colony optimization (ABC). Additionally, we incorporate the Kalman filter to represent the target's motion between frames. Rigorous testing on OTB2013, OTB2015, TempleColor128 and UAV123 datasets attests to the effectiveness, accuracy, and robustness of our method, showcasing superior performance compared to several state-of-the-art techniques*

Streszczenie: *Wizualne śledzenie obiektów stanowi ważny punkt w badaniach nad wizją komputerową, a zastosowania obejmują różne dziedziny, takie jak transport i inteligentne technologie przemysłowe. Chociaż systemy śledzenia oparte na filtrach korelacyjnych wykazały dobrą wydajność, napotykają wyzwania, takie jak zmienność skali, okluzja, efekty brzegowe i bałagan w tle. Aby rozwiązać te problemy, wprowadziliśmy podejście łączące optymalizację sztucznej kolonii pszczół (ABC), głębokie sieci neuronowe i filtr Kalmana. Nasza metoda obejmuje początkowy etap, w którym analizowana jest niezawodność procesu śledzenia, a następnie obliczana jest informacja o pewności celu w aktualnie szacowanej lokalizacji. Następnie przeprowadza się etap aktualizacji. W naszym podejściu dostosowanie skali docelowej określa się poprzez optymalizację sztucznej rodziny pszczół (ABC). Dodatkowo stosujemy filtr Kalmana, który reprezentuje ruch celu pomiędzy klatkami. Rygorystyczne testy na zestawach danych OTB2013, OTB2015, TempleColor128 i UAV123 potwierdzają skuteczność, dokładność i solidność naszej metody, wykazując wyższą wydajność w porównaniu z wieloma najnowocześniejszymi technikami.*

Keywords: Object tracking, confidence information, correlation filter, artificial bee colony, Kalman filter
Słowa kluczowe: Śledzenie obiektów, informacje poufne, filtr korelacji, sztuczna kolonia pszczół, filtr Kalmana

Introduction

Computer vision plays an important role today. Its importance lays in the problems which can solve. It can be seen as a tool that allows the digital world of an image to interact with its physical world. Computer vision has been used to solve many problems like self-driving cars [1], facial recognition [2], augmented and mixed reality [3], health care and health technology [4], Internet of Things (IoT) [5] etc. Furthermore, computer vision is a vast domain of research with many axes. Among these axes of research, we find visual object tracking. The major goal of this axe of research is to estimate the location of target in all frames based on the initial frame target. In other words, after identifying a target in a video frame, it is advantageous to follow its position in the next frames. Every frame where the target is correctly followed produces additional details about the target's identity and motion. Nowadays, visual object tracking is a hot subject of research, where many works have been published [6]. These works aim to improve the existing methods and overcome challenging scenarios such difficult situations including alterations in the target's appearance and tracking targets through intricate movements. In addition, visual object tracking methods can be classified into many classes. They are capable of being categorized as multiple-object vs. single-object tracking methods, discriminative vs. generative methods, offline methods vs. online learning methods, methods based on non-aware or context-aware. Single object tracking methods aim to track one object at each sequence, in the other hand multi-object tracking methods can track multiple targets simultaneously [7, 8]. For generative methods, the tracking operation is completed by finding the most suitable window [9, 10]. However, discriminative methods separate the target object patch from its background [11]. Some methods have been developed based on correlation filter (CF). This leads to two classes, which are correlation filter methods vs. non-correlation filter methods [12, 13]. In our study, we are interested by correlation filter methods or trackers (CFT). CFT are more accurate and have a high frame rate [14]. Many methods have been developed in this class. Additionally, CF is used in order to enhance the robustness and efficiency of the method. CFT carry out computation in the frequency domain, which leads to reduce the computational cost and improving performances of trackers. Moreover, CF is an online training technique, meaning it doesn't require offline training or testing, making it more adaptive and efficient in real-time applications.

However, CFT suffer for many drawbacks. In generally, CFT use a simple linear interpolation for updating. Based on this strategy, some problems can appear such as the reliability of tracking results, the robustness of trackers, more extra computational load, accuracy of tracking, etc. To overcome these issues, we have proposed a new method of visual object tracking based on correlation filter. We have included a deep learning network in our technique. The outcomes show that the proposed strategy outperforms various current state-of-the-art techniques. Our fundamental contributions are summarized as follows:

1. Using reliable online training method that can keep important information from the first frame.
2. Utilizing the Kalman filtering approach to gather target trajectory information so that the tracker can conduct robust tracking even when the object is occluded, deformable, or moving quickly.
3. Using a mechanism for monitoring the process's reliability during the model training, this guarantees that the target information is valid.
4. We propose a method for assessing scale variation in target tracking algorithms that takes advantage of an optimization strategy, aiming to address redundancy and fixed scale issues commonly found in current techniques.

The paper will be structured as follows; the second section will cover relevant work. We shall outline the framework of our proposed method in the third part. Simulation and results will be illustrated in the fourth section. The fifth section will describe our conclusion and future directions.

Related Work

Correlation Filter Methods

Correlation filter-based tracking has garnered significant interest in recent years because of its exceptional computational efficiency. The first algorithm was proposed by Blome et al [15] which developed a minimum output sum of square error method to learn filters for fast visual tracking. In [16], Henriques et al proposed a method that employs the kernel trick with correlation filter and ridge regression. Furthermore, a different extended version has been proposed to enhance the tracking accuracy. Yan et al. [17] propose a CF trained by multi-features such as HOG and Color name features. For the estimation of the target scale, the DSST [18] provided a separate one-dimensional scaling correlation filter that was capable of dealing with scale change efficiently and precisely. All of these methods use the cosine window approach to lessen the effect of boundaries caused by the fast Fourier transform of periodic training samples.

Deep Learning Methods

Recent research shows that deep convolution neural networks (CNN) outperform alternative algorithms in computer vision problems. Considering that deep learning seems to have the properties of requiring huge training data sets as well as high processing needs, trackers that use deep learning may be split into two types. The first one is using a pretrained convolutional neural network (CNN) as a feature extractor and combining it with other approaches. The alternate is to train the tracker using enormous data sets and entirely utilize a deep learning architecture. the Fully-convolutional Siamese network [19] (SiamFC) Identifying the degree of matching among the target and object tracking templates using an end-to-end method of learning with deep features.

Temporal Stability

One of the most extensively used mathematics-based methods for dealing with noisy video sequences is the Kalman filter. For time-dependent parameters in dynamic systems, the recursive Kalman filter is used to predict the system's next state. The performance of prediction systems is significantly influenced by the transition model of the algorithm [20], which describes the motion model of the object. Tripathi et al. [21] suggested an adaptive KF approach for estimating occluded object location based on optimizations. The scale-adaptive tracker demonstrated good results in an experimental test. With the use of a Kalman Filter and KCF, Huynh [22] created a novel tracking system that overcomes occlusion and human-crossing obstruction.

Proposed Tracker

Our new object tracker will be explained in detail in this section. First of all, the efficient online training strategy, which keeps the valuable information of the first frame target intact throughout the training, will be illustrated. Then, the simulation of object motion based on KF in order to get information about its trajectory is going to be presented.

Finally, the artificial bee colony optimization (ABC) will be used to enhance the precession of our tracker.

The framework of our new object tracker is represented by the following Fig. 1.

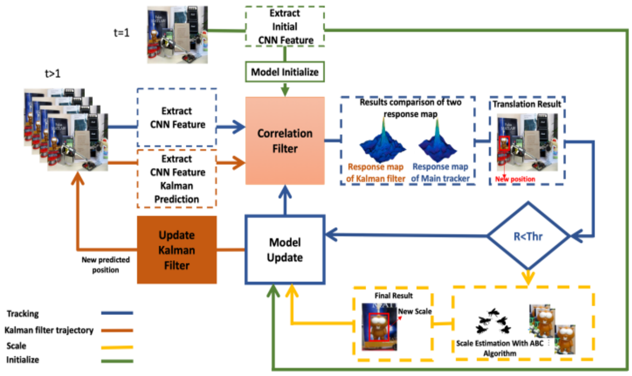

Fig. 1. The tracking technique presented requires initializing the correlation filter using target information in the initial frame. Both the Kalman filter and the correlation filter (CF) estimate the target's position in consecutive frames, while the consistency is assessed by the reliability analysis module. After achieving accurate localization, the ABC method is used to estimate the scale of the target. Next, the first frame's target information and the current accurate information are used to train the model simultaneously. CF has been updated to include "r" for tracking result reliability and the term "Thr" refers to the threshold used to measure the robustness of tracking. CNN is utilized for extracting visual information from images.

Trajectory Estimation based on Kalman Filter (KF)

KF can provide an estimation of some unknown variables based on some measurements taken overtime. KF has been very useful in different applications. Among these applications we find video tracking. KF uses simple forms and does not need a lot of computational power [20]. KF can be seen as an ideal filter, because it minimizes the difference between the real and the estimated state. Because of simplicity and efficiency of KF to represent target motion, a constant velocity of model has been used. This procedure can be divided into two stages, which are prediction and correction.

Convolution Target Location

Standard trackers based on correlation filtering [15] have a fundamental idea, which is to learn a discriminant classifier at first then search for the largest value of the relevant response graph to determine the estimated location of the target object. Furthermore, these trackers begin by training the filter with labels and samples. The target is then located in the search patch using the filter. Finally, the new object's position will be used to update the filter. The model of this process is as follows:

$$w^* = argmin \sum \left\| w\, x_{m,n} - y(m,n) \right\|^2 + \lambda \left\| w \right\|_2^2 \tag{1}$$

where x stands for samples, w for the learned correlation filter, y for the regression response with a Gaussian shape, λ is a parameter of regularization to overcome the over-fitting problem.

Deeper architectures are more sensitive to any significant appearance changes. For the purpose of extracting convolutional features from the layers conv3-4, conv4-4 and conv5-4, we have used the VGG Net [23]. VGG Net is the pre-trained convolutional neural network (CNN), which proposed in Visual Geometry Group.

Furthermore, the obtained feature vector x of size $M \times N \times D$ from the three conv-layers will be used in our tracker. Where: M stands for the width, N is height and D indicates the number of channels. Finally, the inner product is produced via a linear kernel in the Hilbert space e.g:

(2) $$w.x_{m,n} = \sum_{d=1}^{D} w^{T}_{m,n,d}.x_{m,n,d}$$

The size of the correlation filter model is $M \times N$. Circular samples $x_{m,n},(m,n) \in [0,1,...,M-1] \times [0,1,...,N-1]$ are produced using circular matrices of x_l. A two-dimensional Gaussian label function is associated to each circular sample $x_{m,n}$.

(3) $$y(m,n) = e^{-\frac{(m-M/2)^2+(n-N/2)^2}{2\sigma^2}}$$

The variables M and N are used to describe the dimensions of the convolutional feature map, while σ denotes the width of the kernel. Additionally, both x and the correlation filters w have identical sizes. The solution of the filter W^d can be obtained in the frequency domain, following the same procedure as depicted in [24], by using Fast Fourier transform (FFT) for each channel. In the frequency domain, each channel of the learnt filter is represented as follows:

(4) $$W^d = \frac{Y \odot \overline{X}^d}{\sum_{i=1}^{D} X^i \odot \overline{X}^i + \lambda}$$

Complex conjugates are shown by the bars above variables. The operator $\odot$ represents a Hadamard (element-wise) product and $d \in 1,..,D$ is the number of channels. The subsequent frame will be utilized for obtaining the ROI's convolutional features. Where the first layer is specified as $Z_t \in R^{M,N,D}$. The following formula show the calculation of first layer response map Rp_t :

(5) $$Rp_t = F^{-1}\left(\sum_{d=1}^{D} W^d \odot \overline{Z}^d\right)$$

By looking at the greatest value of the response map, we could find the predicted position of the target, which is provided by the first layer. We take a critical look at the response maps produced by a variety of different feature layers. To be more specific, we employ weights to combine the response maps made up of convolutional layers. To establish the definite target location, we commence by utilizing the coordinates of the highest value inside the factored composite response map:

(6) $$Rp_{fusion} = \sum_{l=1}^{3} a_l Rp_l$$

(7) $$(x,y) = arg_{m,n} max Rp_{fusion}(m,n)$$

(8) $$pos = (x,y)$$

The fusion weight of each layer is represented by a_l. We assume that weights of the convolution layers are 1, 0.5, and 0.2 from coarse to fine, respectively. pos is the target location calculated by our tracker.

Appearance Model Update

As described in [25], when tracking in a complex environment, the object being tracked information remains relatively consistent between consecutive frames and includes a significant amount of redundant information.

The change and redundant information lead to slow down the tracking speed and also weaken the tracker performances. To overcome this problem, we have used a strategy based on sparse updating in conjunction with a reliability analysis of target information. In order to estimate the reliability, we have exploited the following equations:

(9) $$\rho_l = 1 - \min\left(\frac{r_{mean}}{r_{max}}, 0.6\right)$$

(10) $$r_l = \min\left(\frac{\mu_l}{\mu_{max}}.\rho_l, 1\right)$$

r_{max} and r_{mean} represent the response map's maximum and average values respectively. This process allows us to obtain more robust and accurate filters. Correlation filter has been updated by a moving average approach. This updating improves performances of tracker by avoiding significant changes in the model. The updating is as follows:

(11) $$\hat{W}_l^{t} = (1-\eta)\hat{W}_l^{t-1} + \eta\hat{w}_l^{new}$$

The relevant learning rate is identified by η.

Scale estimation

Determining the intended size of the object in the image is critical. Among the applications, where the estimation scale is important, we find robotics and surveillance. In order to estimate the target position, the correlation filter for locating the target is trained by extracting depth information from the image or frame. However, the scale of the target is maintained as the same as it is in the previous frame. The training of a scale filter is therefore required in order to obtain an accurate estimation of the present target scaling.

For each frame, several scales are used to evaluate which images are most likely to be target, with the middle of each picture being used as a reference point for all scales. After obtaining the HOG features of each sample, a feature vector x^s is constructed by serially interconnecting features acquired from the retrieved s samples.

Furthermore, each $\boldsymbol{x^s}$ has d-dimensional features. Lastly, training samples of different scales in the S layer were passed through to train the scale filter w^s. The following equation has been used to minimize the error of correlation response measured against the desired correlation output g :

(12) $$\varepsilon = \left\| g - \sum_{s=1}^{d} w^s * x^s \right\|^2 + \lambda \sum_{s=1}^{d} \left\| w^s \right\|^2$$

where g is normally the predicted Gaussian response. g, w^s, and x^s have the same dimension and size. The regularization coefficient is denoted by λ. * stands for circular correlation.

The Equation 12 can be seen as a linear least squares problem. To solve it, it should be transformed to the Fourier domain. The filter that minimizes Equation 12 is as follows:

(13) $$W^{s}=\frac{\overline{G}^{s}}{\sum_{s=1}^{d}\overline{X}^{s}X^{s}+\lambda}=\frac{A^{s}}{B^{s}}$$

In this situation, the corresponding value's discrete Fourier transform is represented by capital letters. $\boldsymbol{G}$ stands for complex conjugation.

To solve the problem of redundant or fixed-scale tracking, we have used an estimation scale approach based on swarm-optimization. We have used the Artificial Bee Colony (ABC) algorithm [26], which is one of the most powerful swarm-optimization algorithms. ABC mimics the honeybee swarm's clever foraging behavior. The artificial colony bee has three types of bees: employed, onlookers and scouts bees.

The ABC method provides a random distribution of food source locations of SN solutions, where SN is the number of employed bees or onlooker bees. Each solution is a vector of d dimensions. d is the number of parameters that should be optimized. Afterwards, calculate the value of fitness function for possible solution. In our case the function of fitness is:

(14) $$y_{t}^{s}=F^{-1}\left\{\frac{\sum_{s=1}^{d}\overline{A}_{t-1}^{s}Z_{t}^{s}}{B_{t-1}^{s}+\lambda}\right\}$$

In the ABC algorithm, a scale pool S is assigned as the x_i vector for each frame of the video. Multiple targets of various scales are chosen from this pool. At frame t, ABC algorithm initializes the vector with random values where the objective function is eq 14. A_{t-1} and B_{t-1} are the filter coefficients at frame $t-1$. The different steps of our new object tracker are summarized though algorithm. 2.

Algorithm2 The Fundamental Process of Our Algorithm
1: Input: pos (1) the starting location of the object in the initial frame, VGGNet-19 pre-trained models;
2: Output: Predicted location post; updated correlation filters; updated scale model;
3: Initiate filter model using Equation (1);
4: Initiate KF models ;
5: X0 extracted CNN's features at initial frame;
6: for t=2, 3, …. do
7: Extracted CNN's features at frame t according to pos(t-1).
8: Equation (5) to calculate the response maps of filter;
9: Using the calculated pos(t) as recording, predict the target location poskalman(t) using the Kalman Filter (KF);
10: Extracted CNN's features of frame t according to poskalman(t), Calculate the response maps regarding Equation (5).;
11: Choose the maximum response map between the two and Compute the reliability information using Equations (5) and (8);
12: if r>Thr then
13: Use Equation (8) to determine the target location pos(t) at the current frame;
14: Find the best scale that maximizes Equation (14) based on ABC algorithm
15: Update scale model
16: The online training module feeds it with the extracted features X0 and Xt and updates CF;
17: else
18: Choose pos(t) that has the maximum response between the two.
19: end if
20: end for

Experiments
The evaluation of Performance

In this part, we conduct comprehensive experiments for the proposed method on a variety of benchmarks. First, we will go through the specifics of our tracker's realization and parameter configurations. Secondly, we evaluate the usefulness of each contribution to our proposed method.

The tracker we suggested was developed in MATLAB 2018b and evaluated on a desktop computer with an Intel i7-4700HQ CPU @ 2.40GHz x 8 and 8 GB RAM. In order to implement VGGNet-19 [30] convolutional feature extraction, we utilize MatConvNet, a popular MATLAB deep learning framework. Finally, we exhibit the tracking performance of our proposed tracker in comparison to various state-of-the-art trackers. We analyse the performance of our system using the OTB2013 [27], OTB2015 [28], TempleColor128 [29] and UAV123 [30] datasets, that include attributes-based evaluations, quantitative, and qualitative. To be more precise, the tracker's settings are set to the following:

The regularization term parameter is set to $\lambda=0.01$, the padding value for the region has been customized to 1.8, the learning rate is $\eta=0.01$ and with a value of 0.1 the filter σ 's kernel width is adjusted. With weights of 1, 0.5 and 0.25 correspondingly, we have utilized the features of conv5-4, conv4-4, and conv3-4 of VGGNet-19. For translation KF, assign a value of $Q=[25,10,10]$, $R=25$ to the covariances of motion and measured noise.

Finally, the control parameters in ABC scale are given in Table. 1 below.

Table 1. ABC algorithm parameters

Parameter	Value
Maximum number of iterations	30
Population size	30
Limit value	20
Acceleration coefficient	1
Dimension	1

Evaluation Criteria

We utilize the OPE criteria [31,32] to evaluate all trackers on all datasets, which includes two metrics: success rate and Precision. Precision can be defined as the Euclidean distance between the center of the estimated result box and the center of the ground truth bounding box. When comparing the performance of different trackers, a 20-pixel distance threshold is typically utilized. The degree of crossover between the two boxes determines the success rate. It is assumed that the tracking in this frame was successful when the overlap area reaches a specific threshold, such as IOU >=0.5. IOU is estimated according to the following equation:

(15) $$IOU=\frac{area\left(B_{T}\cap B_{G}\right)}{area\left(B_{T}\cup B_{G}\right)}$$

Consider B_T as representing the estimation and B_G as the ground truth. The success rate may be calculated by dividing the total number of frames monitored by the number of frames that were successfully tracked during the experiment. The area under the curve (AUC) is a widely used metric for ranking trackers in a success plot. It is calculated by measuring the area under the curve.

Evaluation on OTB2013 and OTB2015 datasets

Ablation experiments on the OTB dataset are performed in this part to evaluate the performance of each module described in this article. The DCF tracker is used as a baseline, but features are extracted using a convolution network. For the purpose of the evaluation of different components' efficiency, we built three independent trackers by integrating the baseline tracker with the ABC scale and each component: Baseline with ABC scale + RA the baseline and the reliability analysis module are combined, the baseline with ABC scale + KF is produced by combining the baseline with the Kalman Filter and the last one is baseline with ABC scale + OT signifies that the target information from both the initial frame and the present frame is used to train the updated filter. Moreover, we evaluate the performance of ABC scale technique against the traditional one. Fig. 2 and 3 the left figure shows accuracy measurements, with tracking accuracy at a 20-pixel error threshold indicated in the legend. The right figure displays the overall success rate of each tracker, represented by the legend's area under the curve (AUC).

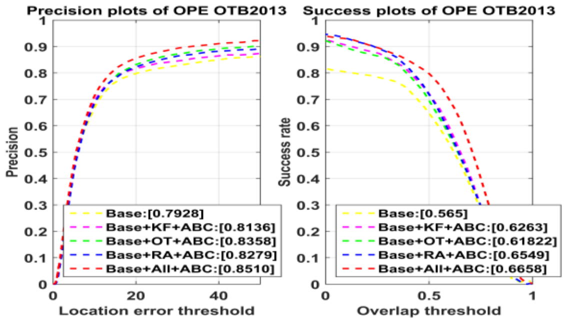

Fig. 2. One-pass evaluation on OTB-2013 shows precision and success plots, with legend displaying AUC and 20-pixel precision for each ablation.

On OTB2013 and OTB2015, Baseline With ABCscale+All achieves peak precision scores of **86.7%** and **85.1%**, respectively. Compared to the four other trackers, this reflects precision gains of 0.9%, 1.0%, 3.3%, and 7.0% on OTB2013, and 1.6%, 2.2%, 3.8%, and 5.9% on OTB2015. With ABCscale+All, Baseline attains success scores of **66.5%** (OTB2013) and **62.5%** (OTB2015), corresponding to gains of 1.1%, 3.9%, 4.7%, and 10.0% on OTB2013, and 1.5%, 2.0%, 2.3%, and 6.3% on OTB2015. The KF module contributes the least improvement among the three evaluated components. The poor precision of the benchmark tracker leads to more measurement error in the Kalman Filter process, producing mediocre prediction results. Therefore, single KF module enhances the benchmark tracker's efficiency slightly compared to the other three modules.

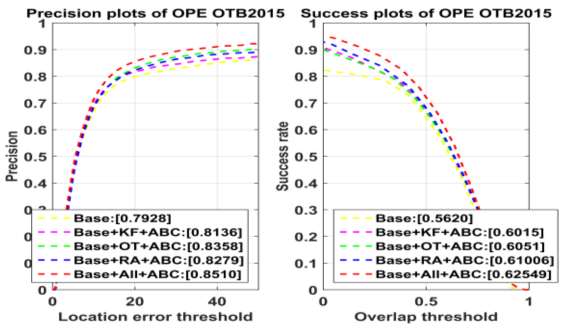

Fig. 3. One-pass evaluation on OTB-2015 over 100 sequences, with legend showing AUC and 20-pixel precision for each ablation.

Additionally, The OT module enhances tracking performance by reliably preserving initial target information, underscoring the importance of accurate object data in effective tracking. The RA module contributes most significantly, maintaining precise target representation and mitigating drift and loss across challenging scenarios such as occlusion and long sequences.

Fig. 4 and 5 show that incorporating the ABC scale improves DP rates from 85.38% to 86.7% (OTB2013) and 84.53% to 85.1% (OTB2015). Success rates also increase slightly: 63.2% to 66.58% (OTB2013) and 62.1% to 62.55% (OTB2015).

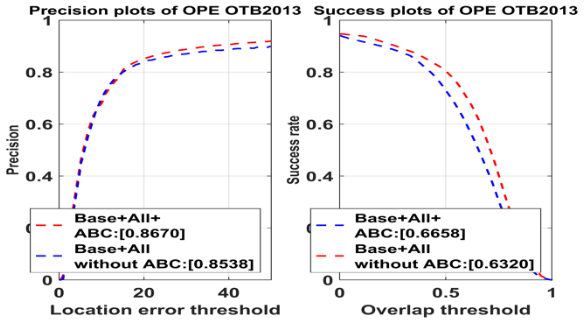

Fig. 4. One-pass evaluation on OTB-2013 over 50 sequences, with legend showing AUC and 20-pixel precision for ablations with and without the ABC approach.

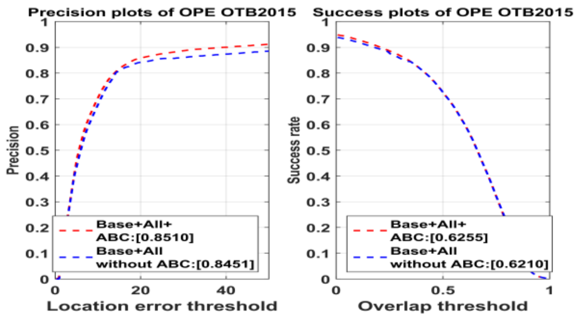

Fig. 5. One-pass evaluation on OTB-2015 over 100 sequences, with legend showing AUC and 20-pixel precision for ablations with and without the ABC approach.

Comparison with Other Trackers

We compared our tracker up against several state-of-the-art techniques, including: IFCT [33], SPCF [34], Zhao et al. [35], KCF [36], Liu et al. [37], Autotrack [38], Ad_SASTCA [39], Shu et al. [40], LADCF_HC [41], MCMCF [42]. Additionally, we performed experiments using OTB2013 and OTB2015, as well.

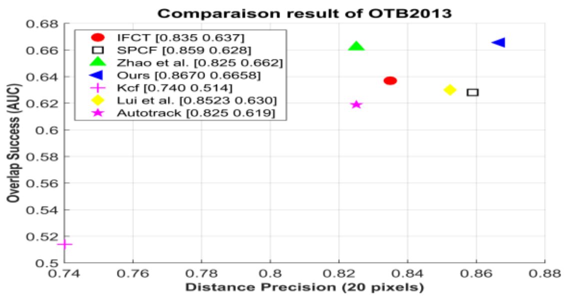

Fig. 6. Success (AUC) and precision (20-pixel threshold) comparison of the proposed tracker with six others on the OTB2013 dataset.

The proposed tracker achieves superior performance on the OTB2013 dataset, with a precision of 0.8670 and a success rate of 0.6658 shown in Fig. 6. Compared to other state-of-the-art methods, SPCF reports 0.859 / 0.628, Liu et al. 0.8523 / 0.630, IFCT 0.835 / 0.637, Zhao et al. 0.825 / 0.662, Autotrack 0.825 / 0.619, and KCF 0.740 / 0.514 in terms of precision and success, respectively. These results highlight the effectiveness of the proposed approach in maintaining accurate and robust tracking across diverse scenarios.

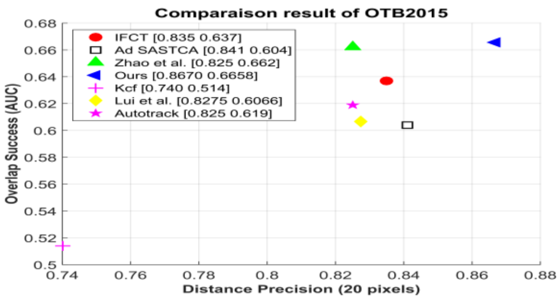

Fig. 7. Performance comparison of the proposed tracker with others on the OTB2015 dataset, based on success rates and precision.

On the OTB2015 dataset, the proposed tracker demonstrates superior performance compared to six state-of-the-art methods. It achieves the highest precision of **0.8670** and success rate (AUC) of **0.6658**. In contrast, IFCT reports 0.835 / 0.637, Ad_SASTCA 0.841 / 0.604, Zhao et al. 0.825 / 0.662, Liu et al. 0.8275 / 0.6066, Autotrack 0.825 / 0.619, and KCF 0.740 / 0.514. These results highlight the effectiveness of the proposed method in achieving both accurate and robust tracking performance under varied conditions.

Evaluation on TC128 and UAV123 datasets

The proposed tracker achieves a precision of **0.718** and a success score of **0.550** on the TC128 dataset are shown in Fig. 8. While slightly below LADCF_HC in precision, it matches its success performance and surpasses Shu et al., Autotrack, and Lui et al. This balance between accuracy and robustness demonstrates the effectiveness of our method across diverse tracking conditions.

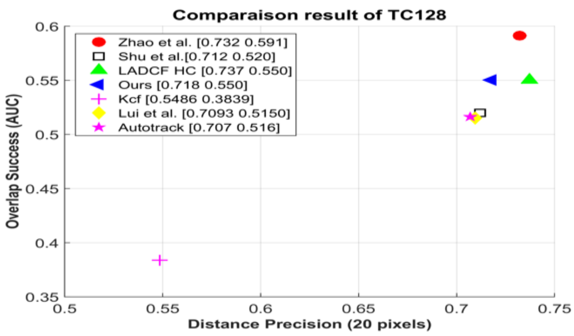

Fig. 8. Precision and success plots on the TC128 dataset, illustrating the performance of the proposed tracker against six state-of-the-art algorithms across 123 sequences

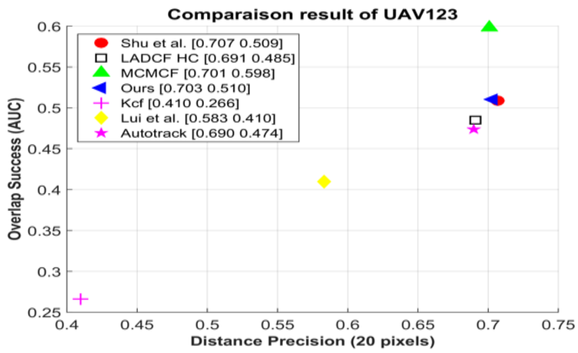

Fig. 9. Precision and success plots on the UAV123 dataset, illustrating the performance of the proposed tracker against six state-of-the-art algorithms across 123 sequences.

Fig. 9 shows the evaluation of the proposed tracker on the UAV123 dataset using precision and success metrics. It achieves a precision of **0.703** and a success score of **0.510**, outperforming several state-of-the-art methods such as LADCF_HC, Autotrack, and Lui et al. These results demonstrate the effectiveness and competitiveness of our approach in aerial visual tracking scenarios.

Conclusion

This paper presents a robust visual tracking framework incorporating a reliability analysis module, an online training and update strategy, a Kalman Filter (KF), and an ABC-based scale adaptation mechanism. The proposed method effectively addresses key challenges in visual tracking, including occlusion, drift, appearance changes, and scale variation. Experimental results across four benchmark datasets OTB2013, OTB2015, TC128, and UAV123 demonstrate the competitiveness of the proposed tracker. It ranks 1st on OTB2013 and OTB2015, also performs among the top trackers on TC128 and UAV123, surpassing several state-of-the-art methods. While performance under in-plane rotation remains a limitation, the overall results confirm the robustness and generalizability of the approach across diverse tracking scenarios.

Authors: *Ramdane TAGLOUT is a PhD student at the Department of Computer Science, University of Bouira, Algeria, LIM Laboratory, Faculty of Exact Sciences, University of Bouira, 10000, Bouira, Algeria, Information technology department, Faculty of Exact Sciences, University of Bouira, 10000, Bouira, Algeria, email:r.taglout@univ-bouira.dz, Ph.D. Bilal SAOUD is a Senior lecturer at Bouira university, Algeria, Electrical Engineering Department, Faculty of Applied Sciences, University of Bouira, 10000, Bouira, Algeria, LISEA Laboratory, Faculty of Applied Sciences, University of Bouira, 10000, Bouira, Algeria; email:bilal340@gmail.com.*

REFERENCES

[1] Agarwal, N., Chiang, C. & Sharma, A. A study on computer vision techniques for self-driving cars. International Conference On Frontier Computing. pp. 629-634 (2018).
[2] Park, S., Ko, J., Huh, J. & Kim, J. Review on generative adversarial networks: Focusing on computer vision and its applications. Electronics. 10, 1216 (2021).
[3] Le, T., Nguyen-Xuan, H. & Kim, S. Hand drawing-based daylight analysis using deep learning and augmented reality. Results In Engineering. 22 pp. 102119 (2024), https://www.sciencedirect.com/science/article/pii/S259012302400 736.
[4] Saraswat, D., Bhattacharya, P., Verma, A., Prasad, V., Tanwar, S., Sharma, G., Bokoro, P. & Sharma, R. Explainable AI for healthcare 5.0: opportunities and challenges. IEEE Access.(2022).
[5] Suresh, G., Reddy, P., Praveen, M., Arun, P. & Jayanth, K. REAL TIME ROOM AUTOMATION USING IOT DEVICE. Turkish Journal Of Computer And Mathematics Education (TURCOMAT). 14, 422-434 (2023)
[6] Taglout, R. & Saoud, B. Deep learning application on object tracking. Przeglad Elektrotechniczny. 2023 (2023)
[7] Yilmaz, A., Javed, O. & Shah, M. Object tracking: A survey. Acm Computing Surveys (CSUR). 38, 13-es (2006)
[8] Smeulders, A., Chu, D., Cucchiara, R., Calderara, S., Dehghan, A. & Shah, M. Visual tracking: An experimental survey. IEEE Transactions On Pattern Analysis And Machine Intelligence. 36, 1442-1468 (2013)
[9] Fiaz, M., Mahmood, A., Javed, S. & Jung, S. Handcrafted and deep trackers: Recent visual object tracking approaches and trends. ACM Computing Surveys (CSUR). 52, 1-44 (2019)

[10] Liu, B., Yang, L., Huang, J., Meer, P., Gong, L. & Kulikowski, C. Robust and fast collaborative tracking with two stage sparse optimization. Computer Vision–ECCV 2010: 11th European Conference On Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11. pp. 624-637(2010)
[11] Miao, Q., Xu, C., Li, F., Zuo, W. & Meng, Z. Delayed rectification of discriminative correlation filters for visual tracking. The Visual Computer. 39, 1237-1250 (2023)
[12] Xu, T., Feng, Z., Wu, X. & Kittler, J. Adaptive channel selection for robust visual object tracking with discriminative correlation filters. International Journal Of Computer Vision. 129 pp. 1359- 1375 (2021)
[13] Li, B., Yan, J., Wu, W., Zhu, Z. & Hu, X. High performance visual tracking with siamese region proposal network. Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition.pp. 8971-8980 (2018)
[14] Zhu, X., Wu, X., Xu, T., Feng, Z. & Kittler, J. Robust visual object tracking via adaptive attribute-aware discriminative correlation filters. IEEE Transactions On Multimedia. 24 pp. 301 312 (2021)
[15] Bolme, D., Beveridge, J., Draper, B. & Lui, Y. Visual object tracking using adaptive correlation filters. 2010 IEEE Computer Society Conference On Computer Vision And Pattern Recognition.pp. 2544-2550 (2010)
[16] Henriques, J., Caseiro, R., Martins, P. & Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. Computer Vision–ECCV 2012: 12th European Conference On Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part IV 12. pp. 702-715 (2012)
[17] Yan, H., Xie, M.,Wang, P., Zhang, Y. & Luo, C. Kernel correlated filtering target tracking algorithm based on multi-features fusion. IEEE Access. 7 pp. 96079-96084 (2019).
[18] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, ‘Accurate scale estimation for robust visual tracking’, in British machine vision conference, Nottingham, September 1-5, 2014, (2014).
[19] Bertinetto, L., Valmadre, J., Henriques, J., Vedaldi, A. & Torr, P. Fully-convolutional siamese networks for object tracking. European Conference On Computer Vision. pp. 850-865 (2016).
[20] Li, Q., Li, R., Ji, K. & Dai, W. Kalman filter and its application. 2015 8th International Conference On Intelligent Networks And Intelligent Systems (ICINIS). pp. 74-77 (2015).
[21] Tripathi, R., Ghosh, S. & Chandle, J. Tracking of object using optimal adaptive Kalman filter. 2016 IEEE International Conference On Engineering And Technology (ICETECH). pp. 1128-1131 (2016).
[22] Huynh, X., Choi, I. & Kim, Y. Tracking a human fast and reliably against occlusion and human-crossing. Image And Video Technology. pp. 461-472 (2015).
[23] Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. ArXiv Preprint ArXiv:1409.1556. (2014)
[24] Naresh Boddeti, V., Kanade, T. & Vijaya Kumar, B. Correlation filters for object alignment. Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition. pp. 2291-2298 (2013).
[25] Chen, H., Zhang, W. & Yan, D. Robust Visual Tracking with Reliable Object Information and Kalman Filter. Sensors. 21, 889 (2021).
[26] Karaboga, D. & Basturk, B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. Journal Of Global Optimization. 39, 459-471 (2007).
[27] Y. Wu, J. Lim, M.-H. Yang, Online object tracking: A benchmark, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2411–2418.
[28] Y. Wu, J. Lim, M.-H. Yang, Object tracking benchmark, IEEE Trans Pattern Anal Mach Intell 37 (9) (2015) 1834–1848.
[29] P. Liang, E. Blasch, H. Ling, Encoding color information for visual tracking: algorithms and benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5630–5644.
[30] M. Mueller, N. Smith, B. Ghanem, A benchmark and simulator for uav tracking,in: European Conference on Computer Vision, Springer, 2016, pp. 445–461.
[31] Smeulders, A., Chu, D., Cucchiara, R., Calderara, S., Dehghan,Transactions On Pattern Analysis And Machine Intelligence. 36, 1442-1468 (2013).
[32] M. Kristan et al., ‘The visual object tracking vot2015 challenge results’, in Proceedings of the IEEE international conference on computer vision workshops, 2015, pp. 1–23.
[33] Song, H. IOT-Oriented Visual Target Tracking and Supply Chain Art Product Design. Mobile Information Systems, 2022(1), 3773469.
[34] Lu, Z., Yang, G., Liu, D., Yang, J., Yang, Y., & Zhou, C. Selective part-based correlation filter tracking algorithm with reinforcement learning. IET Image Processing, (2022),16(4), 1208-1226.
[35] Zhao, Jianwei, Fuyuan Wei, NingNing Chen, and Zhenghua Zhou. "Spatial and long–short temporal attention correlation filters for visual tracking." IET Image Processing 16, no. 11 (2022): 3011-3024.
[36] Henriques, J., Caseiro, R., Martins, P. & Batista, J. High-Speed Tracking with Kernelized Correlation Filters. IEEE Transactions On Pattern Analysis And Machine Intelligence. 37, 583-596 (2015).
[37] Liu, J., Ye, P., Zhang, X., & Xiao, G. Real-time long-term tracking with reliability assessment and object recovery. IET Image Processing, (2021), 15(4), 918-935.
[38] Li, Y., Fu, C., Ding, F., Huang, Z., & Lu, G. AutoTrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, (2020), pp. 11923-11932.
[39] Su, Y., Liu, J., Xu, F., Zhang, X., & Zuo, Y. A Novel Anti-Drift Visual Object Tracking Algorithm Based on Sparse Response and Adaptive Spatial-Temporal Context-Aware. Remote Sensing, (2021), 13(22), 4672.
[40] Shu, Q., Lai, H., Wang, L., & Jia, Z. Multi-feature fusion target re-location tracking based on correlation filters. IEEE Access, (2021), 9, 28954-28964.
[41] Liang, M., Wu, X., Wang, Y., Zhu, Z., Cao, B., & Xu, J. Multi-model and multi-expert correlation filter for high-speed tracking. IEEE Access, (2021), 9, 52326-52335.
[42] Chen, Z., Zheng, H., Zhai, X., Zhang, K., & Xia, H. Correlation Filter of Multiple Candidates Match for Anti-Obscure Tracking in Unmanned Aerial Vehicle Scenario. Mathematics, (2022), 11(1), 163.