

RESEARCH ARTICLE

A practical evaluation of correlation filter-based object trackers with new features

Islam Mohamed^{1,2}, Ibrahim Elhenawy¹, Ahmed W. Sallam², Andrew Gatt³, Ahmad Salah^{1,4*}

1 Faculty of Computers and Informatics, Zagazig University, Zagazig, Egypt, **2** Technical Research and Development Center, Cairo, Egypt, **3** Science Department, Eno River Academy, Hillsborough, NC, United States of America, **4** Department of Information Technology, CAS-Ibri, University of Technology and Applied Sciences, Muscat, Oman

* ahmad@hnu.edu.cn**OPEN ACCESS**

Citation: Mohamed I, Elhenawy I, Sallam AW, Gatt A, Salah A (2022) A practical evaluation of correlation filter-based object trackers with new features. PLoS ONE 17(8): e0273022. <https://doi.org/10.1371/journal.pone.0273022>

Editor: Mohamed Hammad, Menoufia University, EGYPT

Received: December 24, 2021

Accepted: August 1, 2022

Published: August 25, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0273022>

Copyright: © 2022 Mohamed et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The relevant data for this study are available here: http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html.

Funding: The author(s) received no specific funding for this work.

Abstract

Visual object tracking is a critical problem in the field of computer vision. The visual object tracker methods can be divided into Correlation Filters (CF) and non-correlation filters trackers. The main advantage of CF-based trackers is that they have an accepted real-time tracking response. In this article, we will focus on CF-based trackers, due to their key role in online applications such as an Unmanned Aerial Vehicle (UAV), through two contributions. In the first contribution, we proposed a set of new video sequences to address two uncovered issues of the existing standard datasets. The first issue is to create two video sequence that is difficult to be tracked by a human being for the movement of the Amoeba under the microscope; these two proposed video sequences include a new feature that combined background clutter and occlusion features in a unique way; we called it *hard-to-follow-by-human*. The second issue is to increase the difficulty of the existing sequences by increasing the displacement of the tracked object. Then, we proposed a thorough, practical evaluation of eight CF-base trackers, with the top performance, on the existing sequence features such as out-of-view, background clutters, and fast motion. The evaluation utilized the well-known OTB-2013 dataset as well as the proposed video sequences. The overall assessment of the eight trackers on the standard evaluation metrics, e.g., precision and success rates, revealed that the Large Displacement Estimation of Similarity transformation (LDES) tracker is the best CF-based tracker among the trackers of comparison. On the contrary, with a deeper analysis, the results of the proposed video sequences show an average performance of the LDES tracker among the other trackers. The eight trackers failed to capture the moving objects in every frame of the proposed Amoeba movement video sequences while the same trackers managed to capture the object in almost every frame of the sequences of the standard dataset. These results outline the need to improve the CF-based object trackers to be able to process sequences with the proposed feature (i.e., *hard-to-follow-by-human*).

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

Object tracking is a critical and difficult task. Thus, object tracking has become a popular topic of study in recent years. There are many uses for visual object tracking such as visual surveillance [1], video understanding [2], robotics [3], and unmanned aerial vehicle-based monitoring [4]. Object tracking has a significant impact in unmanned aerial vehicle missions, for example autonomous landing [5], tracking of flying vehicles [6], and target following [7], given a target in an axis-aligned [8–11] or rotated bounding box [12]. After the target bounding box has been defined in the beginning frame, the tracker must identify the patch that is the most similar in the next frames. The state is usually depicted as a bounding box enclosing the goal. Under adverse environmental conditions, most current trackers are still vulnerable. Occlusion, camera motion, clutter, similar appearance, and illumination effect are all problems faced by visual object trackers.

The tracking task is a mix of classification and estimation problems [13]. The first task is to use classification to provide a reliable location of the target. The following procedure is to calculate an accurate target location, which is frequently defined by the use of a bounding box. Visual object tracking techniques that are currently in use include discriminative correlation filters [14, 15], deep neural networks [16, 17], Siamese networks [18–20], and classic methods [21, 22], as shown in Fig 1. In this paper, we concentrate on CF-based trackers due to their computational efficiency and their real-time response. For instance, CF-based methods are the dominant object tracking technique in UAV tracking, which requires a fast tracker. Recently, the correlation filter has become one of the most common techniques in visual tracking for real-time applications. The correlation filter's effectiveness is primarily owing to its use of circulant matrices to perform potentially complex computation in the frequency domain instead of the spatial domain, to speed up the processing.

Recently, Correlation Filter Trackers (CFTs) have accomplished notable successes in terms of the reported accuracy rate on the standard datasets [23]. In the following study, we will provide a quick overview on some pertinent CFTs to illustrate this success. MOSSE tracker [24] is the oldest filter-based correlation method, which trains the filter with only grayscale samples. The CSK tracker [25] adds a kernel trick to the CF formula. With circulant shifted samples, the filter coefficients may be effectively optimized in the frequency domain. According to CSK [25], the KCF tracker [14] takes advantage of multi-channel HOG descriptors [26] to improve the representation of features capability as well as tracking efficiency. Likewise, in order to establish reliable tracking in color videos, color names (CN) features are added [27]. IBCCF

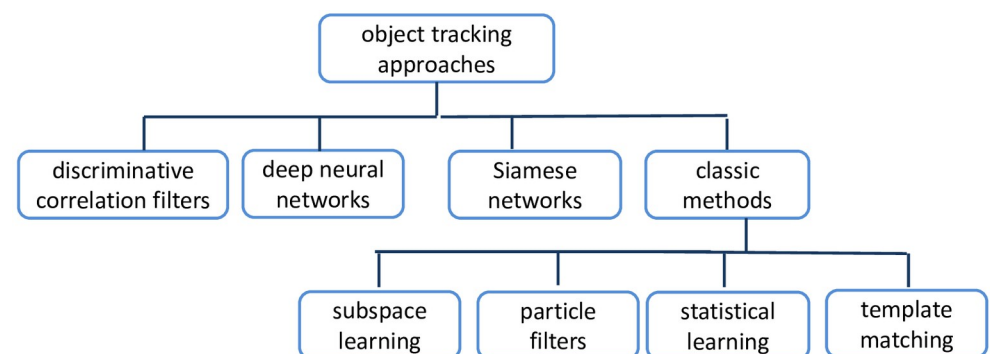


Fig 1. A taxonomy of object tracking methods.

<https://doi.org/10.1371/journal.pone.0273022.g001>

[28], DSST [29], and SAMF [30] trackers use multi-scale searching techniques to deal with the scale adaptation issue. The authors in [31] proposed using dual filters, one for translation and another filter for scale variation.

The field of visual object tracking is a well-established field with several standard object tracking datasets (e.g., OTB-2013 [8], OTB-2015 [32], and VOT-2015 [33]) and successful CF-based trackers (e.g., MOSSE tracker [24]) and non-CF-based trackers [34]. The CF-based trackers are known for their low computational requirements and fast response time (i.e., locating the target in an image very fast). In this paper, we will focus only on CF-based visual trackers. The literature includes a very limited practical assessment of the CF-based visual tracker, especially for recent methods. Besides, The existing standard video sequences used for the practical evaluation the object trackers contain objects that are easy to be tracked by human beings. In addition, some of the standard video sequence datasets include objects that move with large displacement value; it is called fast motion feature. Video sequences with fast motion feature include a fast moving object but in the same direction. Thus, the performance of CF-based trackers should be evaluated when there is a large displacement in both object speed and direction. These are the major shortcoming of the CF-based object tracker which motivated the current work.

In this work, the need for exploring the performance of the CF-based object trackers on more difficult video sequences relative to the existing standard dataset motivated this work. In this context, we proposed two new video sequences for the movement of the Amoeba under the microscope. These two proposed video sequences have a new feature; they are difficult to track by human beings. We call this feature *hard-to-follow-by-human*, as the objects in the existing standard video sequences can be easily tracked by a human being. In addition, we proposed to extend the existing *fast motion* feature to increase the object displacement in different scenarios; we called this feature *very fast motion*. In addition, the lack of practical evaluation of the recent CF-based object trackers motivated this work as well. Thus, we provide a comprehensive practical evaluation of recent eight CF-based trackers, namely, LDES, BICF, CPCF, IBRI, MRCF, ARCF, DRCF, and TB_BICF. These eight trackers were evaluated using a variety of well-known benchmarks (i.e., OTB-2013) [8, 32] and the proposed video sequences as well. Then, the performance of each tracker was justified to outline their points of weakness and strength. Finally, we employed a success and precision plots [8]. The main contributions of this work can be summarized as follows:

1. We proposed a set of new video sequences with new attributes, namely, *hard-to-follow-by-human* and *very fast motion*. To our knowledge, it is the first time to propose evaluating object trackers with a video sequence which is hard to follow by human's eyes. The proposed video sequences and the results and trackers' source codes are publicly available (<https://github.com/elmesedi/A-practical-evaluation-of-correlation-filter-based-object-Trackers/>).
2. We thoroughly practically evaluated eight recent CF-based trackers through a set of experiments, and the obtained results outlined the merits and demerits of these eight trackers.

The remainder of this article is structured as follows. The background is introduced in Section 2 and early work is reviewed. In Section 4, we will explained the evaluation metrics to evaluate the trackers. The dataset formulation is described in Section 3. Section 5 illustrates the assessment of the suggested trackers and the results of the experiment. Finally, we conclude the proposed work in Section 6.

2 Background and related work

2.1 Background

Many people are probably acquainted with the fundamentals of the correlation's principles, as they exist in statistics and probability. In pattern recognition, correlation is used to assess how similar or dissimilar a test target is to the training images. On the other hand, simple correlation is only effective when the test target and training images are similar. There is a demand for a wide range of approaches to enhance the fundamental correlation as well as obtain qualities for example flexibility with variances or distortions in reality (e.g., scale changes, image rotations, changes in light, etc.). The idea of correlation pattern recognition is explained in detail, as shown in Fig 2. There are two sub-images in Fig 2; a sub-image for testing with a variety of patterns, a test image t , and a template that we are searching for, a reference image r . In Fig 2, we search for the letter "S". Assume the two sub-images are binary, with white areas having a value of 0 and black regions having a value of 1. The sub-image for testing $t[x, y]$ and the template/reference image $r[x, y]$ are correlated as follows. We attempted aligning the template's top-left corner on the image's top-left corner, as shown in Fig 3.

Fig 3 includes one large image and one small sub-image. The small sub-image represents the patten, i.e., the object to be tracked, and the large image represents the image for testing, an image with many patterns. The small sub-image appears at the top-left corner of the large image; it is the image with kiwi fruit. The two images in Fig 3 are multiplied, pixel by pixel, and the product array's values are added to find the template's correlation value with the image for position of the two with relation to one another. Then, the values of correlation are calculated again by moving the template to all potential centerings with respect to the image. The correlation values are then determined again by moving the template to all feasible centerings with respect to the image. As a result, we can locate the desired targets by searching for peaks in the correlation output and determining if the peaks are large enough to detect the template's presence.

The simple correlation process is represented by Eq 1. We need to consider complicated correlation filters because the simple technique only works if the image contains perfect duplicates of the template image and no other items that have a similar appearance to the template. In Fig 2, for instance, the letter "O" and the letter "C" have a significant correlation, and for the letter "O" a simple correlation provides a high correlation result, which is inappropriate. Another drawback of the simple correlation process is that it can be extremely sensitive to

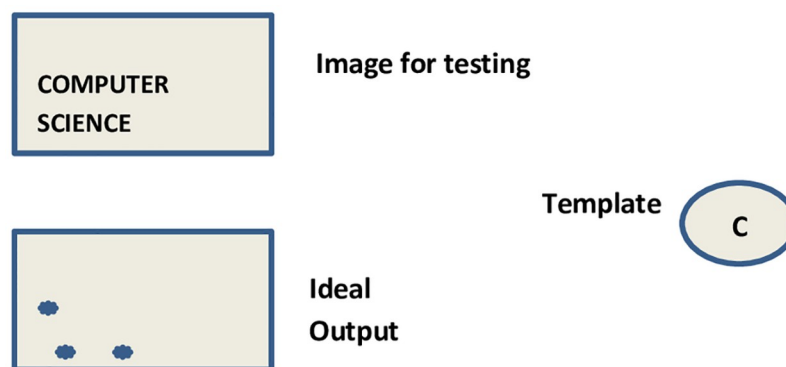


Fig 2. The concept of image correlation.

<https://doi.org/10.1371/journal.pone.0273022.g002>



Fig 3. Correlation pattern recognition.

<https://doi.org/10.1371/journal.pone.0273022.g003>

noise.

$$c[x, y] = \sum_i \sum_k t[k, 1] r[k + x, 1 + y] \quad (1)$$

Correlation outputs can lead to incorrect decisions if this randomness is not explicitly addressed. The simple correlation approach fails if the template finds massive modifications in appearance or shape in the target scene, often referred to as distortions, which could be caused by changes in lighting or viewing geometry (for example, scale changes, rotations, etc.). A good tracking system must be capable of dealing with such expected variation [35].

2.2 Related work

In 2010, for the first time, the correlation filter approach was employed in the object tracking. After nearly a decade of development, correlation filter tracking methods have reached maturity. Learning from gray-scale images can generate the minimum output sum of squared errors (MOSSE) [24]. For the first time, the correlation filter is applied to the tracking field. This filter is simple to compute and track the object rapidly, but it cannot predict accuracy when the appearance of the object changes.

In 2012, Henriques et al. introduced circulant structure tracking with kernels (CSK) [25]. MOSSE is a very reliable and efficient tracking system, and it has a low computational complexity that can process a large number of frames per second. In [36], the authors compared MOSSE and DCFs, where the performance of MOSSE was found to be better. On the other hand, MOSSE has several drawbacks; it cannot ensure that it will track an object accurately

when its appearance changes. In addition, MOSSE does not consider the target's historical data, instead, it updates the filter parameters with the previous and current frames only [37]. Thus, MOSSE tracker's robustness is low as well as its accuracy. Consequently, due to the accumulate error of MOSSE, the drifting problem increases which can not be reversed, and as a result the object will be lost; this is called anti-drift ability, which is low for the MOSSE method [37].

The MOSSE filter was improved by CSK [25] by introducing kernel approaches. The main limitation of CSK that it is limited to a single channel. Thus, the kernel correlation filter (KCF) [14] modified channel characteristics to multi-channel features and provides CN features for tracking. The CN feature [27] enhances the filter's capability to discriminate. In spite of the filter's ability to adapt to fast motion, out-of-view, and rotation, it still has to be enhanced. It is reported that KCF is sensitive to moving backgrounds, occlusion, and noise [38]. Of note, KCF utilized a constant size of the tracking bounding box. Its performance is limited when the tracked object size changes; in other words, KCF does not have the scale-invariant property [39].

To handle the multi-scale variation problem, Danelljan et al. [40] suggested a discriminative scale space tracker (DSST) based on the feature pyramid, as well as an enhanced method (fDSST) [29]. The authors proposed utilizing sub-grid interpolation of correlation values; this significantly reduced the method's processing time and the target search space with no effect on the proposed method's accuracy or robustness. One limitation of the DSST method is its poor performance to decide the target size when the object size changes frequently [41]. Regrading fDSST, the only improvement over DSST is the tracking speed [42].

By using phase correlation, Li et al. [43] estimated the target's scale and rotation by transforming the sample into log polar space. LDES [43] is a new tracker based on a correlation filter with a new method for estimating robust similarity transformations on massive displacements. LDES tracker divides the search space of the object into two sub-spaces. LDES performs well with a video sequence containing rigid objects, but its performance decays with occlusion [44].

The BICF method [45] is proposed to develop a new bidirectional incongruity-aware correlation filter. The bidirectional incongruity error, that is dependent on the response, is integrated into the correlation filter. BICF has the ability to quickly learn to adjust in looks, as well as suppression of inconsistent errors. Then, TB_BICF is proposed as a variant of BICF in [46]. The TB_BICF proposed a novel bidirectional incongruity-aware correlation filter using the nature of tracking reversibility. Two drawbacks of the BICF and TB_BICF are reported in [46]. First, BICF could not track the object when it went out of view and reappeared again. Second, the BICF's performance is poor when the video sequence has an occlusion feature.

DRCF [47] proposes using a dual regularization method to dynamically regularize the filter that performs the correlation process within the core of the filter that produces ridge regression. DRCF proposed using a dual regulation strategy to reduce the regression correlation of the operation's border effect. DRCF utilizes a saliency-aware regularizer to handle any possible change in the size of the tracked object using an effective saliency detection technique. The main drawbacks of DRCF are discussed in [47]. First, rapid illumination effects the robustness of the DRCF tracker. Second, the fast scale change has a negative effect on the accuracy of the saliency detection.

IBRI tracker [48] was motivated by the fact that most of the existing methods combine the information of only two successive video frames; this approach limits the usefulness of the information, as using more frames increases the information and improves the object's boundary robustness. In addition, some methods utilized historical data from the previous video frames but without removing the noise of these frames. The authors of IBRI proposed a new

aerial object tracking method that takes advantage of disruptor-aware interval-based response inconsistency. Thus, IBRI tracker performance is good even when scale variation is presented.

CPCF [49] introduces a new method for pursuing dynamic consistency in correlation filters. CPCF utilizes HoG [26] and CN [27] features. CPCF utilized a consistency map to measure the level of consistency for each two successive video frames. This is preformed in a dynamic manner to extract rich temporal information.

MRCF [50] is a new tracking approach that has been proposed depending on a correlation filter with a multi-regularizers. The MRCF method is aware of the channel reliability. Thus, it can accomplish an automatic channel weight distribution through learning the distribution of each channel weight. It is mainly designed to be used as a UAV tracker.

To date, CPCF, MRCF, and IBRI trackers are not well evaluated, as they are recent methods. Thus, there are no reported drawbacks to date for this tracker. This motivated the present work to evaluate these methods alongside other trackers with good performance.

3 Dataset formulation

Object tracking benchmark datasets have been proposed to standardize the task of assessments of visual object tracking methods. They include video sequences with a variety of target categories, as well as varied time lengths and difficult attributes. These datasets provide a wide range of video sequences, frames, attributes, and classes (or clusters). One of the most common benchmarks is the OTB-2013 dataset [8], which consists of 50 difficult video sequences labelled with 11 attributes. The sequences in the OTB-2013 datasets present several tracking challenges such as occlusion, deformation, and background clutters.

The summary of the proposed video sequences is listed in Table 1. First, modification of the existing video sequence was proposed by increasing the object displacement in consecutive frames. In other words, we proposed to modify the fast motion feature into the very fast motion feature. Thus, we selected two sequences from the OTB-2013 dataset already have the FM feature, namely, Tiger1, Surfer, and one sequence lacking the FM feature, i.e., Football1. This will help us to evaluate the performance of the trackers by increasing the motion speed for both the fast and slow motion objects.

In the following text, all the original sequence are from OTB-2103, unless otherwise is mentioned. The first proposed video sequence is based on Tiger1 sequence; it is called “Tiger1_VFM_1”. The original Tiger1 sequence already has the FM property, and we proposed increasing the object displacement, speed, by removing the last ten frames out of every 50 frames. Thus, the proposed sequence includes a sudden change in the object’s motion speed relative to the original Tiger1 sequence. Subsequently, the object’s motion speed return to

Table 1. The proposed video sequences.

Sequence Name	Sequence description	Attributes
Tiger1_VFM_1	Drop the last ten frames from every 50 frames.	OPR, IV, DEF, MB, OCC, FM, VFM, IPR
Tiger1_VFM_2	Drop ten frames when the object reach to the end of the view.	IV, MB, OCC, DEF, IPR, FM, VFM, OPR
Surfer_VFM	Drop the last 5 frames of every 20 frames, from the first 40 frames only.	SV, FM, VMF, IPR, OPR, LR
Football1_Modf	Drop the first frame from every 10 frames.	IPR, OPR, BC, FM
AMB1	A clear object is occluded by a semi-opaque object.	OPR, BC, DEF, FM,
AMB2	A blurred object is occluded by a semi-opaque object.	OPR, BC, DEF, FM

<https://doi.org/10.1371/journal.pone.0273022.t001>

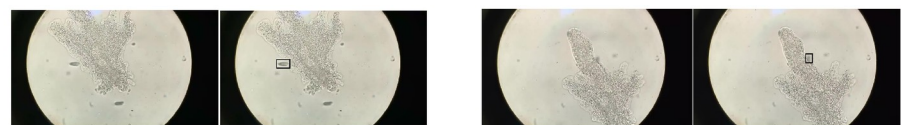
normal in the next 40 frames and so on. We aim to explore whether the tracker can still detect a sudden change in the object's displacement.

In the second proposed sequence, the change will be observed in the motion speed as well as direction. The original Tiger1 sequence includes a predictable path of the tiger object. In the proposed sequence, a sudden change was posited in both the object's motion speed and direction by dropping ten frames when the tiger object reaches the end of the view. That results in reversing the direction of the object path. We call this proposed sequence "Tiger1_VFM_2". The third proposed sequence is similar to the first proposed sequence apart from two main differences. First, the original sequence is the Surfer. Second, rate of changing the motion speed is higher, as we proposed dropping the last five frames out of each 20 frames. We called this sequence "Surfer_VFM".

The fourth proposed sequence, and the last proposed modified sequence of the OTB-2103 dataset, is based on a sequence that does not have the FM feature. The aim of this selection is to evaluate the effect of increasing the speed of a slow moving object. The original sequence is the Football1; we proposed dropping the first frame from every 10 frames which is called "Football1_Modf".

The fifth and sixth proposed video sequences, in Table 1, are new sequences are not included in any existing dataset. The proposed sequences are extracted from a video capturing the Amoeba movement under a microscope. There is an intuitive impression that the sequences of the standard dataset can be easily tracked by a human beings. Thus, we aim to explore the performance of the CF-based trackers when the sequence has this property. In this context, we proposed extracting the two sequences from a video for life under the microscope of amoeba (https://www.youtube.com/watch?v=ZpAbk_xaew). The proposed two sequences based on this video are called AMB1 and AMB2, where AMB stands for Amoeba. AMB1 sequence consists of 20 frames from seconds 1:37 to 1:40, and AMB2 consists of 20 frames from seconds 1:25 to 1:27 of the aforementioned video. A sample frame of each of the two sequences are illustrated in Fig 4. The two sequences have the following features: out-of-plane rotation, deformation, and background clutters attributes, and occlusion.

The sequence AMB1, Fig 4a, is easier to detect than AMB2, Fig 4b, as the object in AMB1 is clear, while the object in AMB2 is blurred and the object's color resembles the background color. Occlusion is classified as inter-occlusion, background occlusion, and self-occlusion [51]. The proposed AMB1 and AMB2 push the limit of the classes of the occlusion feature. The intuition of proposing the AMB1 and AMB2 sequences is that the moving amoeba is entered into a larger object which should occult the amoeba, but as the object is not fully opaque, the amoeba object is not fully occult, and hard to be tracked by human's eyes. In other words, the proposed sequences, AMB1 and AMB2, combine two different features which are occlusion and background clutters. The merger of these two features results in a new feature called *semi-opaque occlusion*. In this feature, the occluding object can have a transparency level that makes it hard to track the object but not impossible for human's eyes. In other words, the occlusion



(a) The AMB1 sequence.

(b) The AMB2 sequence.

Fig 4. The proposed amoeba video sequences, clear image to the left and annoated image to the right. a) AMB1 b) AMB2.

<https://doi.org/10.1371/journal.pone.0273022.g004>

feature was binary; we proposed to make this feature with several levels of transparency, rather than restricting its identity to an occlusion or otherwise.

4 Evaluation metrics

We evaluated the trackers of interest using a familiar benchmark datasets [8]. The benchmark datasets contains 50 challenging image sequences with 11 attributes: out-of-view (OV), fast motion (FM), scale variation (SV), low resolution (LR), deformation (DEF), illumination variation (IV), background clutters (BC), motion blur (MB), occlusion (OCC), out-of-plane rotation (OPR), and in-plane rotation (IPR). Each attribute provides a unique challenge to object tracking. Many attributes can be assigned to one sequence. The benchmark OTB-2013 uses success rate and the precision for quantitative evaluation. Furthermore, it evaluates the robustness of object tracking in two ways.

Precision plot

The center location error is a metric commonly used for assessing tracking precision. It calculates the difference between the tracker's predicted center X_k^T and the ground-truth box center X_k^G in the k -th frame. It is commonly referred to as RMSE (root mean square error):

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N \|X_k^G - X_k^T\|^2} \quad (2)$$

where N represents the total number of frames.

The average center location error across all frames of a particular sequence is then utilized to describe the overall efficiency of that sequence. The output location can be uncertain when the tracker misses the target, and the average error number might not always be useful to predict tracking efficiency [52]. The precision plot [25, 52] is a new way of evaluating global tracking performance. It displays the frames percentage in which the predicted position is located inside a specific range of the ground truth. As a representative precision score for each tracker, we choose the threshold score = 20 pixels [52].

Success plot

The bounding box overlap another assessment measure, given ground truth box b_a , where the tracked box b_t and the overlap score(S) is calculated as follows:

$$S = \frac{|b_t \cap b_a|}{|b_t \cup b_a|} \quad (3)$$

where \cap and \cup represent the intersection and union of two different parts, respectively, the region's pixel count is represented by $|\cdot|$. We count how many successful frames have an overlap S that exceeds the defined threshold t_o to assess a series of frames' performance. The success plot shows the ratios of successful frames at the thresholds varied from 0 to 1. For tracker evaluation, using a single value for the success rate at a certain threshold (for example, $t_o = 0.6$) could not be accurate or representative. Rather, we rate the tracking algorithms based on the area under the curve (AUC) Of every success plot.

Robustness evaluation

The standard technique of assessing trackers is to execute them through a test sequence that begins with the ground truth location in the first frame and ends with producing the average

success rate or accuracy. This is called a one-pass evaluation (OPE). A tracker, however, may be sensitive to initialization and its efficiency might decrease or significantly increase when initializing differently at a different starting frame. Thus, OTB-2013 [8] recommends two methods for evaluating the robustness of a tracker to initialization, by temporally perturbing the initialization (starting at various frames) and spatially (starting by various bounding boxes). Spatial Robustness Evaluation (SRE) and Temporal Robustness Evaluation (TRE) are the terms for these tests. SRE and TRE are better than the traditional OPE approach, which only initializes every tracker on the initial frame; in TRE, each tracker is started on a separate frame (i.e., with a temporal spread); whereas a noisy bounding box in the SRE tracker is initialized.

5 Experiments and analysis

5.1 Setup

The proposed experiments were carried out on OTB-2013 [8] benchmark and our proposed video sequences in order to assess the benefits of the eight CF-based tracking algorithms. The experiments were performed on a desktop PC with Intel(R) Core(TM) i5-2410M CPU and RAM of 10GB using MATLAB 2017. In the proposed assessments of the CF-based trackers, we chose eight trackers with presently available source codes, namely, DRCF [47], BICF [45], TB_BICF [46], LDES [43], IBRI [48], MRCF [50], CPCF [49], and ARCF_H [53]. The utilized video sequences include a set of original five video sequences from the OTB-13 dataset, namely, Tiger1, Surfer1, Football1, DragonBaby, Ironman, and Box video sequences. In addition, we utilized four modified video sequences, namely, Football_Modf, Tiger1_VFM_1, Tiger1_VFM_2, and Surfer_VFM. Besides, two new video sequences, namely, AMB1 and AMB2.

We used the resultant tracking information produced by OTB-2013 benchmark to compare the experimental results of the source codes of the eight CF trackers. Occlusion (OCC), motion blur (MB), background clutters (BC), fast motion (FM), and illumination variation were the features considered when comparing these trackers.

The results considering comparing the CF-trackers qualitative and quantitative analysis. The qualitative comparison includes frames with boundary boxes of the different trackers. The quantitative analysis includes three metrics, namely, 1) the precision and success plots of the OPE, TRE, and SRE, 2) the number of frame per second (fps), and 3) the number of frames where the tracker managed to capture the object. OPE is a standard assessment approach in which on each sequence is processed by the trackers only once. For TRE, it executes trackers on 20 sub-sequences separated by the parent sequence with various lengths, and SRE assesses trackers by starting with scaled ground or significantly modified truth bounding boxes in the first frame.

5.2 Discussion

First, for the qualitative results, Figs 5–7 show different performance of the eight trackers on different video sequences. Considering Figs 5 and 6 (http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html), these two figures show that the proposed modified video sequences are more difficult for the tracker than the original video sequences. For instance, the performance of the trackers on OTB-2013 original sequence Tiger1 in Fig 5a is much better than the performance of the same trackers on the proposed modified sequence in Fig 5b and 5c. Similarly, this behaviour can be visually seen by comparing the original sequence in Fig 5d to the modified sequence in Fig 5e. Besides, the objects in the frames of the Football1 sequence in Fig 6b are almost captured by all of the eight trackers, while the proposed modified Football1 sequence (i.e., Football1_Modf), in Fig 6c, the performance of the trackers slightly degraded.

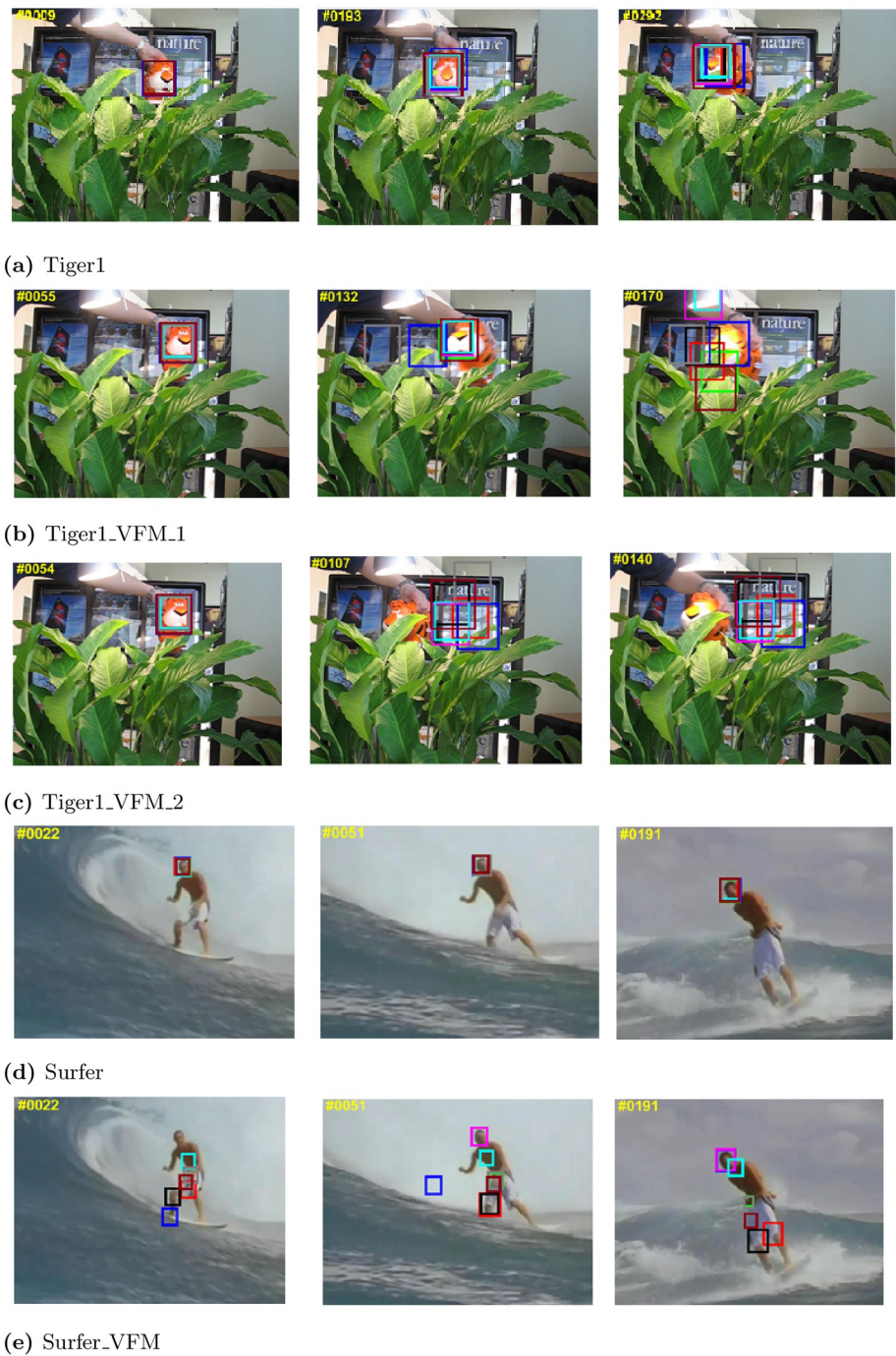


Fig 5. Qualitative assessment of the tested trackers on the a) Tiger1, b) Tiger1_VFM_1, c) Tiger1_VFM_2, d) Surfer, and e) Surfer_VFM sequences [8].

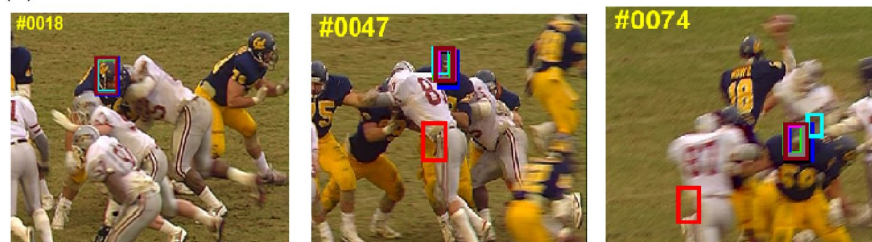
<https://doi.org/10.1371/journal.pone.0273022.g005>



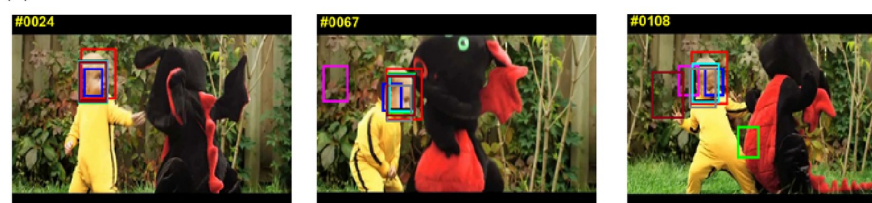
(a) Ironman



(b) Football1



(c) Football1_Modf



(d) DragonBaby

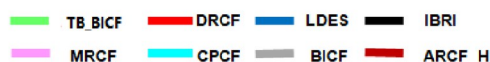


Fig 6. Qualitative assessment of the tested trackers on the a) Ironman, b) Football1, c) Football1_Modf and d) DragonBaby video sequences [8].

<https://doi.org/10.1371/journal.pone.0273022.g006>

Second, we proposed using three metrics for the quantitative evaluation. Using the first metric, the performance of the eight trackers are evaluated in terms of the throughput, i.e., speed of the trackers to process an input video sequence. A number of factors influence the computational performance of CF-based trackers. These factors include the target object's bounding box size, the bounding box for searching area, the number of iterations, and the

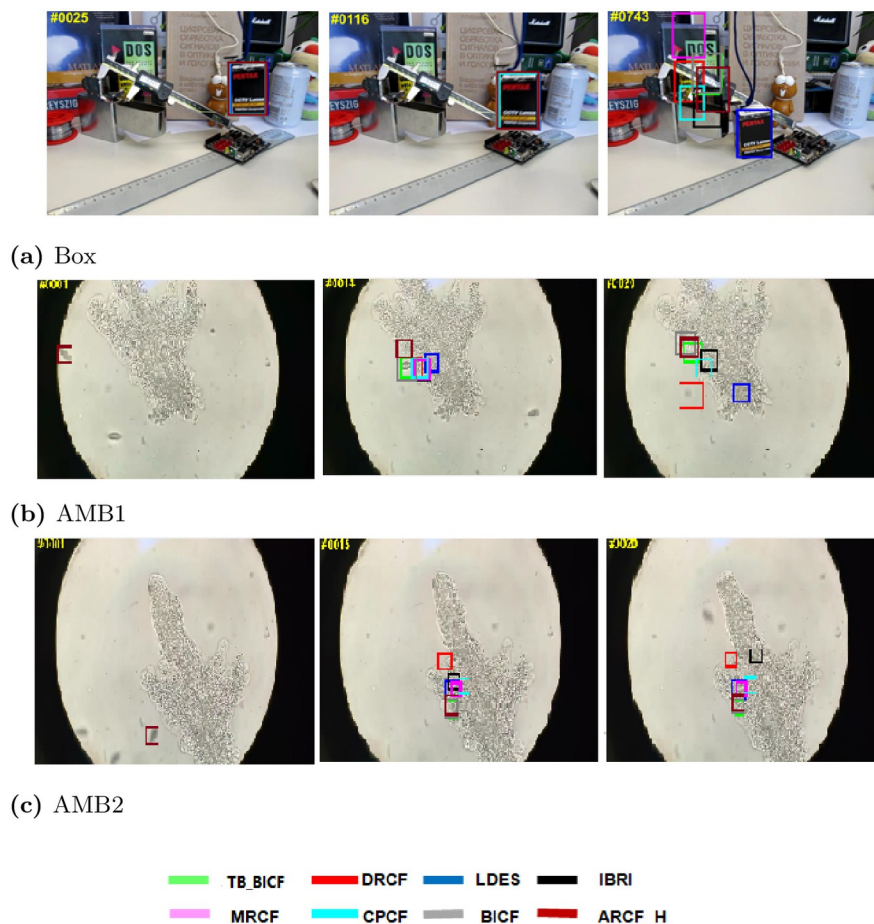


Fig 7. Qualitative assessment of the tested trackers on the a) Box [8], b) AMB1 and c) AMB2 sequences.

<https://doi.org/10.1371/journal.pone.0273022.g007>

number of features per sequence. Table 2 lists the the eight trackers' average frame per second results for each tracker where the running times can be classified into three groups, namely, 1) slow, 2) fast, and 3) normal. The slow group of the largest running times trackers includes LDES and IBRI. The slowest performance of the LDES tracker can be linked to the idea, which is proposed by LDES, of using two boundary boxes to track the object of interest. This proposed idea of using two boundary boxes increases the running time drastically. The slow running time of IBRI tracker can be linked to the extended use of historical frames to track the object. The main motivation of IBRI tracker was that most of the existing methods combine the information of only two successive video frames; thus, IBRI extended the number of used frames at the cost of the running time. The fast group of the trackers includes the CPCF tracker only with almost 12 frames per second. The fast performance the CPCF tracker can be justified by utilizing two fast to compute two features, HoG and CN; computing those two features

Table 2. The average fps for the eight trackers.

Visual Tracking Methods	BICF	TB_BICF	CPCF	DRCF	IBRI	ARCF_H	LDES	MRCF
fps	8.48	9.93	12.18	9.65	6.27	10.25	3.41	10.46

<https://doi.org/10.1371/journal.pone.0273022.t002>

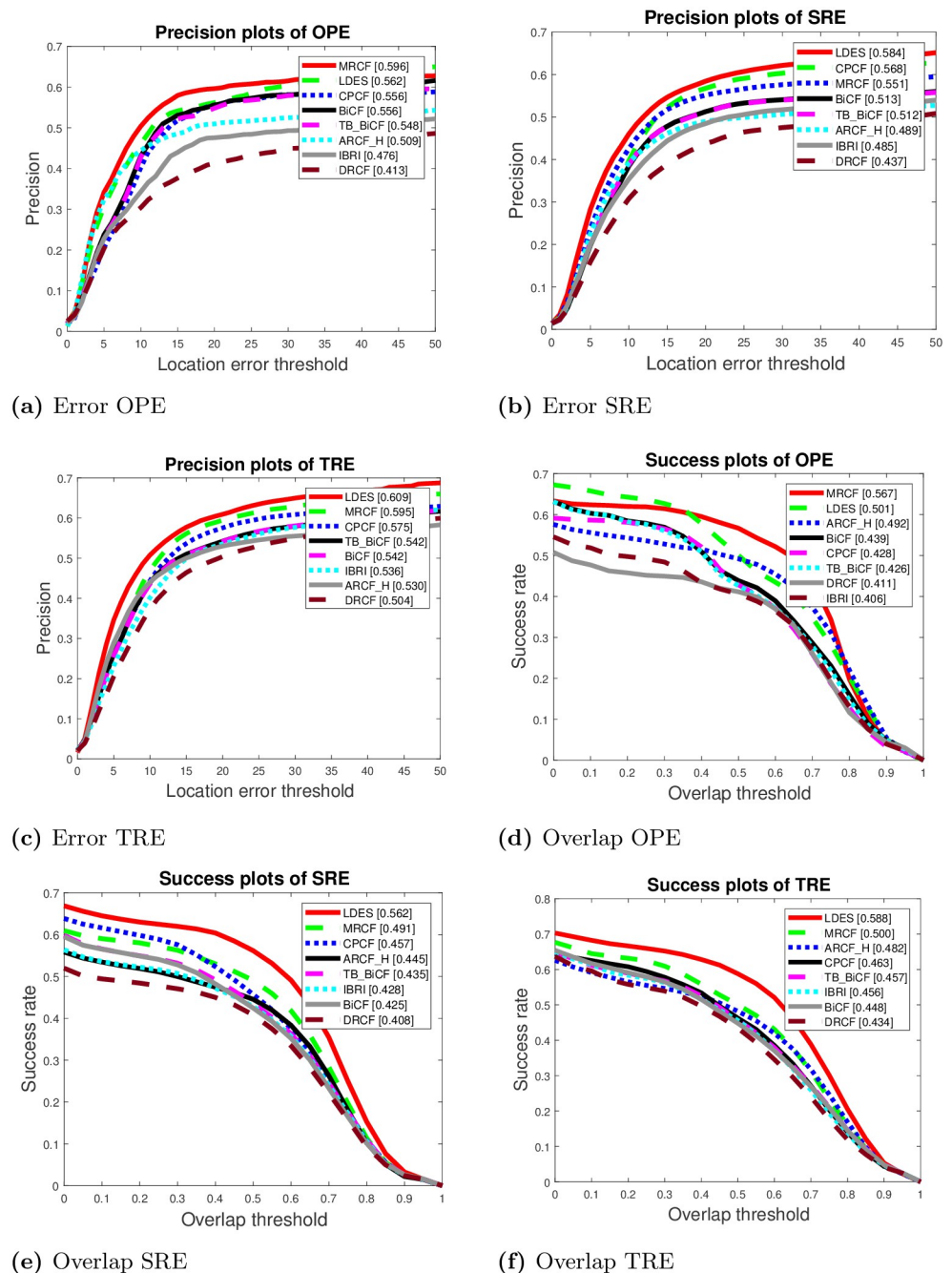


Fig 8. Success and precision plots of error and overlap for OPE, SRE, and TRE.

<https://doi.org/10.1371/journal.pone.0273022.g008>

includes low computational requirements. The third group includes the remaining other six trackers which are close to each other in terms of the fps metric.

The second quantitative metric includes precision and success plots of OPE, TRE, and SRE. Fig 8 illustrates the precision and success plots of OPE, TRE, and SRE metrics for the eight trackers. Obviously, the LDES tracker has the best precision and success rates for the three metrics except for OPE success rate, as the best tracker was MRCF. Thus, the overall results show that superior results of LDES followed by MRCF while the worst performance yielded by DRCF tracker. Besides, Fig 9 depicts the success plots for 11 different attributes of the utilized

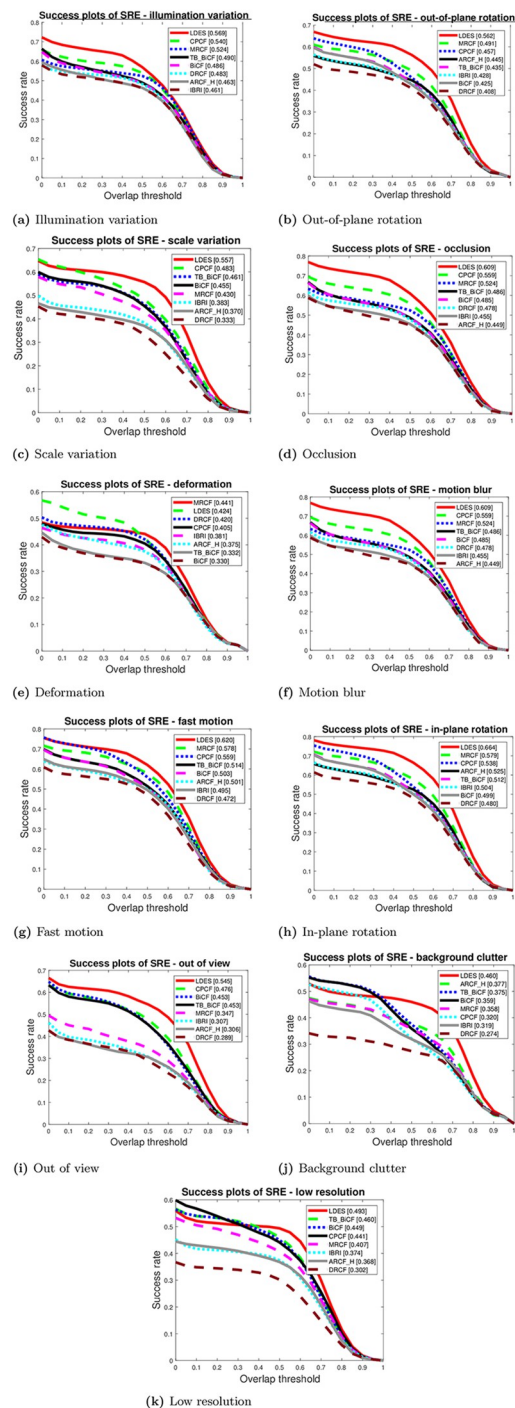


Fig 9. Attribute-based success plots of SRE (illumination variation, out-of-plane rotation, scale variation, occlusion, deformation, and motion blur). (a) Illumination variation. (b) Out-of-plane rotation. (c) Scale variation. (d) Occlusion. (e) Deformation. (f) Motion blur. (g) Fast motion. (h) In-plane rotation. (i) Out of view. (j) Background clutter. (k) Low resolution.

<https://doi.org/10.1371/journal.pone.0273022.g009>

Table 3. The percentage of frames the trackers managed to capture the object from the entire sequence.

	ARCF_H	BICF	TB_BICF	CPCF	LDES	MRCF	DRCF	IBRI
Tiger1(354)	346(97.7%)	346(97.7%)	346(97.7%)	354(100%)	330(93.2%)	354(100%)	354(100%)	354(100%)
Tiger1_VFM_1(294)	281(95.6%)	84(28.6%)	124(42.2%)	125(42.5%)	78(26.5%)	125(42.5%)	277(94.2%)	275(93.5%)
Tiger1_VFM_2(324)	256(79.8%)	64(19.8%)	64(19.8%)	278(85.8%)	64(19.8%)	280(86.4%)	65(20%)	276(85.2%)
Surfer(376)	376(100%)	376(100%)	376(100%)	376(100%)	376(100%)	376(100%)	376(100%)	376(100%)
Surfer_VFM(366)	15(4.1%)	15(4.1%)	15(4.1%)	15(4.1%)	306(83.6%)	355(97%)	15(4.1%)	15(4.1%)
Football1(74)	74(100%)	74(100%)	74(100%)	73(98.6%)	74(100%)	74(100%)	55(74.3%)	74(100%)
Football1_Modf(67)	74(100%)	74(100%)	74(100%)	73(98.6%)	74(100%)	74(100%)	23(34.3%)	74(100%)
AMB1(20)	6(30%)	18(90%)	18(90%)	13(65%)	12(60%)	12(60%)	16(80%)	13(65%)
AMB2(20)	4(20%)	4(20%)	4(20%)	5(25%)	4(20%)	4(20%)	4(20%)	4(20%)

<https://doi.org/10.1371/journal.pone.0273022.t003>

datasets. In ten of these attributes, the LDES tracker again has the best performance followed by MRCF and ARCF_H while the only attribute the LDES has the second place was the deformation attribute.

The third quantitative metric is the summary of the number of frames captured by the trackers for the video sequences listed in Table 1. Table 3 lists the percentage of successfully captured framed. The overall results of Table 3 show that the eight trackers managed to track the objects in almost all of the frames of the original OTB-2013 video sequences, namely, Tiger1, Surfer, and Football1. This can be linked to the fact that the performance of the trackers are tuned according to the standard datasets. On the contrary, all of the trackers failed to track the objects of the proposed modified sequences in all of the frames of the modified sequences when the number of dropped frames are large, five or more frames.

The results of Table 3 outline the degrade in the eight trackers' performance when comparing the original video sequence of Tiger1 and Surfer at one side with the modified sequences Tiger1_VFM_1, Tiger1_VFM_2, and Surfer_VFM at the other side. For instance, the LDES, the best tracker of the previously discussed metrics, managed to capture 93.2% of the original OTB-2013 Tiger1 video sequence frames, which contains 354 frames. This performances, of LDES tracker, degraded to 26.5% and 19.8% for the proposed modified sequences of Tiger1_VFM_1 and Tiger1_VFM_2, respectively. The LDES tracker managed to capture the object in 330 frames out of 354 frames of the original Tiger1 video sequence while the LDES tracker managed to captured the object only in 78 frames out of 294 frames of the proposed modified video sequence Tiger1_VFM_1. The performance of the eight trackers degraded on the modified sequences but with varies ratios based on the tracker parameters, e.g., bounding box size, for searching area in the frame, the number of considered previous frames. Of note, the same results can be found for the Surfer sequence and its modified sequence as well. To study the effect of the number of dropped frames on the performance deterioration, we modified the Tiger1, Surfer, and Football1 video sequence of OTB-2013 by dropping 10, 5, and 1 consecutive frames, respectively. As the Football1 is modified by dropping one frame each ten frames, the trackers' performance is almost fixed, almost no effect is noticed. Thus, the recommended number of dropped frames should be five or more consecutive frames, as the trackers' performance deteriorated for the modified sequences of Surfer and Tiger1 only. Finally, the intervals of the missing frames of each sequences for each tracker are listed in Table 4, where N/A means the tracker did not miss the object in any frame.

In Table 3, the results of AMB1 and AMB2 video sequence are consistence with the motivation of the proposed feature, i.e., *hard-to-follow-by-human*, as the more difficult to follow the object by human's naked eyes, the worst the performance of the tracker. The AMB2 video

Table 4. The frame ranges of CF-based trackers when failing to capture objects.

	ARCF_H	BICF	TB_BICF	CPCF	LDES	MRCF	DRCF	IBRI
Tiger1(354)	(347–354)	(347–354)	(347–354)	N/A	(60–64) (67–71) (94–110)	N/A	N/A	N/A
Tiger1_VFM_1(294)	(51–56) (131–139)	(51–57) (91–294)	(51–57) (131–294)	(51–56) (131–294)	(51–57) (85–294)	(51–56) (131–294)	(51–56) (131–142)	(51–58) (131–142)
Tiger1_VFM_2(324)	(51–56) (71–125) (316–324)	(51–57) (71–324)	(51–57) (71–324)	(51–56) (71–95) (108–125)	(51–57) (71–324)	(51–56) (71–94) (108–124)	(51–56) (71–324)	(51–59) (71–94) (108–125)
Surfer(376)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Surfer_VFM(366)	(16–366)	(16–366)	(16–366)	(16–366)	(16–76)	(16–26) (366)	(16–366)	(16–366)
Football1(74)	N/A	N/A	N/A	(74)	N/A	N/A	(26–28) (58–74)	N/A
Football1_Modf(67)	N/A	N/A	N/A	(74)	N/A	N/A	(24–67)	N/A
AMB1(20)	(7–20)	(18,19)	(19,20)	(14–20)	(13–20)	(13–20)	(17–20)	(14–20)
AMB2(20)	(5–20)	(5–20)	(5–20)	(6–20)	(5–20)	(5–20)	(4) (6–20)	(5–20)

<https://doi.org/10.1371/journal.pone.0273022.t004>

sequence is more difficult to follow for human's eyes than AMB1; this finding is emphasized by results of Table 3.

The results of Figs 8 and 9 are the standard curves to evaluate the eight trackers. These two figures show that the LDES tracker has the overall best performance. On the contrary, with a deeper analysis of the results in Tables 3 and 4 of the proposed modified video sequences, the results reveal that the LDES tracker is not the best tracker. For instance, Table 3 shows that the best trackers of the proposed modified sequences Tiger1_VFM_1 and Tiger1_VFM_1 are ARCF_H and MRCF, respectively. The LDES tracker was the worst tracker in terms of object capturing on the proposed modified sequence Tiger1_VFM_1; this is exactly the opposite to the results of the standard dataset of Figs 8 and 9. Despite the average performance of the BICF and CPCF trackers relative to the LDES tracker's performance in Figs 8 and 9, the best trackers were BICF and CPCF for the proposed AMB1 and AMB2 video sequences, respectively. For the proposed video sequences, the LDES tracker performance was average in comparison to the other trackers. These results emphasize that there is a need to test the CF-based trackers on video sequences rather than the sequences of the standard datasets.

6 Conclusion

The CF-based trackers are the most suitable visual object tracker type for real-time applications such as AUV. In this paper, we proposed to provide a practical evaluation of the most recent well-performing CF-based trackers. The lack of such practical evaluations motivated this work. In this context, we carefully selected eight CF-based trackers based on their performance on the state-of-the-art datasets (e.g., OTB-2013 and OTB-2015). Those eight trackers are theoretically discussed in terms of their merit and demerit. Then, we propose to practically evaluate those trackers by utilizing the OTB-2013 dataset and proposing a set of new video sequences with new features which are not been included in the standard object tracking datasets before, namely, *hard-to-follow-by-human* and *very fast motion*. The proposed feature *hard-to-follow-by-human* merges both the background clutters and occlusions features in a unique manner. As it can be concluded from the new feature name, the video sequence includes an object which is difficult, but not impossible, for a human to track. In addition, we

proposed modifying a number of existing video sequences of OTB-2013 with the fast motion feature; the modification includes increasing the speed of motion by dropping more frames with different scenarios. Considering the OTB-2013 dataset and the new video sequences, the eight trackers are thoroughly evaluated on different metrics, including precision and success rate for quantitative assessment. Finally, all of the eight CF-trackers failed to capture the object in every frame in the proposed sequences while the same trackers managed to capture the object in almost every frame of the sequences of the standard dataset. The performance of the eight trackers was very poor on the video sequence with the proposed features *hard-to-follow-by-human*. The future directions include improving the CF-based filter to manage to better capture objects in sequences with the proposed *hard-to-follow-by-human* feature and comparing other object tracking datasets, as this study was limited only to the OTB-2013 dataset.

Supporting information

S1 File.
(ZIP)

Author Contributions

Conceptualization: Ahmad Salah.

Data curation: Ahmed W. Sallam, Andrew Gatt.

Investigation: Andrew Gatt.

Methodology: Islam Mohamed, Andrew Gatt, Ahmad Salah.

Project administration: Ahmad Salah.

Software: Islam Mohamed.

Supervision: Ibrahim Elhenawy, Ahmad Salah.

Validation: Ibrahim Elhenawy, Ahmad Salah.

Visualization: Islam Mohamed, Ibrahim Elhenawy, Ahmed W. Sallam.

Writing – original draft: Islam Mohamed, Ibrahim Elhenawy, Ahmed W. Sallam.

Writing – review & editing: Islam Mohamed, Ibrahim Elhenawy, Ahmed W. Sallam, Ahmad Salah.

References

1. Xing J, Ai H, Lao S. Multiple Human Tracking Based on Multi-view Upper-Body Detection and Discriminative Learning. In: 2010 20th International Conference on Pattern Recognition. IEEE; 2010.
2. Renoust B, Le DD, Satoh S. Visual Analytics of Political Networks From Face-Tracking of News Video. IEEE Transactions on Multimedia. 2016; 18(11):2184–2195. <https://doi.org/10.1109/TMM.2016.2614224>
3. L Liu HA J Xing, Ruan X. Hand posture recognition using finger geometric feature. IEEE. 2012;.
4. Mueller M, Smith N, Ghanem B. A Benchmark and Simulator for UAV Tracking. In: Computer Vision—ECCV 2016. Springer International Publishing; 2016. p. 445–461.
5. Lin S, Garratt MA, Lambert AJ. Monocular vision-based real-time target recognition and tracking for autonomously landing an UAV in a cluttered shipboard environment. Autonomous Robots. 2016; 41(4):881–901. <https://doi.org/10.1007/s10514-016-9564-2>
6. Fu C, Carrio A, Olivares-Mendez MA, Suarez-Fernandez R, Campoy P. Robust real-time vision-based aircraft tracking from Unmanned Aerial Vehicles. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2014.

7. Cheng H, Lin L, Zheng Z, Guan Y, Liu Z. An autonomous vision-based target tracking system for rotorcraft unmanned aerial vehicles. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE; 2017.
8. Wu Y, Lim J, Yang MH. Online Object Tracking: A Benchmark. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2013.
9. Liang P, Blasch E, Ling H. Encoding Color Information for Visual Tracking: Algorithms and Benchmark. *IEEE Transactions on Image Processing*. 2015; 24(12):5630–5644. <https://doi.org/10.1109/TIP.2015.2482905> PMID: 26415202
10. Valmadre J, Bertinetto L, Henriques JF, Tao R, Vedaldi A, Smeulders AW, et al. Long-term tracking in the wild: A benchmark. In: Proceedings of the European conference on computer vision (ECCV); 2018. p. 670–685.
11. Muller M, Bibi A, Giancola S, Alsubaihi S, Ghanem B. Trackingnet: A large-scale dataset and benchmark for object tracking in the wild. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 300–317.
12. Kristan M, Matas J, Leonardis A, Vojř T, Pflugfelder R, Fernandez G, et al. A novel performance evaluation methodology for single-target trackers. *IEEE transactions on pattern analysis and machine intelligence*. 2016; 38(11):2137–2155. <https://doi.org/10.1109/TPAMI.2016.2516982> PMID: 26766217
13. Danelljan M, Bhat G, Khan FS, Felsberg M. Atom: Accurate tracking by overlap maximization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 4660–4669.
14. Henriques JF, Caseiro R, Martins P, Batista J. High-Speed Tracking with Kernelized Correlation Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015; 37(3):583–596. <https://doi.org/10.1109/TPAMI.2014.2345390> PMID: 26353263
15. Zhang J, Yuan T, He Y, Wang J. A background-aware correlation filter with adaptive saliency-aware regularization for visual tracking. *Neural Computing and Applications*. 2022; p. 1–18.
16. Qi Y, Zhang S, Qin L, Yao H, Huang Q, Lim J, et al. Hedged Deep Tracking. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016.
17. Liu B, Chang X, Yuan D, Yang Y. HCDL-SRCF tracker: Learning an adaptively multi-feature fuse tracker in spatial regularized correlation filters framework. *Knowledge-Based Systems*. 2022; 238:107913. <https://doi.org/10.1016/j.knosys.2021.107913>
18. Zhang T, Xu C, Yang MH. Robust Structural Sparse Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019; 41(2):473–486. <https://doi.org/10.1109/TPAMI.2018.2797082> PMID: 29994599
19. Wang Q, Teng Z, Xing J, Gao J, Hu W, Maybank S. Learning Attentions: Residual Attentional Siamese Network for High Performance Online Visual Tracking. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE; 2018.
20. Li B, Yan J, Wu W, Zhu Z, Hu X. High Performance Visual Tracking with Siamese Region Proposal Network. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE; 2018.
21. Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*. 2002; 50(2):174–188. <https://doi.org/10.1109/78.978374>
22. Avidan S. Support vector tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2004; 26(8):1064–1072. <https://doi.org/10.1109/TPAMI.2004.53> PMID: 15641735
23. Mueller M, Smith N, Ghanem B. Context-aware correlation filter tracking. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1396–1404.
24. Bolme D, Beveridge JR, Draper BA, Lui YM. Visual object tracking using adaptive correlation filters. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE; 2010.
25. Henriques JF, Caseiro R, Martins P, Batista J. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. In: *Computer Vision—ECCV 2012*. Springer Berlin Heidelberg; 2012. p. 702–715.
26. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE; 2005.
27. Danelljan M, Khan FS, Felsberg M, Weijer JVD. Adaptive Color Attributes for Real-Time Visual Tracking. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2014.
28. Li F, Yao Y, Li P, Zhang D, Zuo W, Yang MH. Integrating Boundary and Center Correlation Filters for Visual Tracking with Aspect Ratio Variation. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). IEEE; 2017.
29. Danelljan M, Häger G, Khan FS, Felsberg M. Accurate Scale Estimation for Robust Visual Tracking. In: Proceedings of the British Machine Vision Conference 2014. British Machine Vision Association; 2014.

30. Zhang H, Guo H, Liang C, Yan X, Liu J, Weng J. Grading Tai Chi Performance in Competition with RGBD Sensors. In: Computer Vision—ECCV 2014 Workshops. Springer International Publishing; 2015. p. 3–13.
31. Xu L, Gao M, Liu Z, Li Q, Jeon G. Accelerated duality-aware correlation filters for visual tracking. *Neural Computing and Applications*. 2022; p. 1–16.
32. Wu Y, Lim J, Yang MH. Object Tracking Benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015; 37(9):1834–1848. <https://doi.org/10.1109/TPAMI.2014.2388226> PMID: 26353130
33. Kristan M, Matas J, Leonardis A, Felsberg M, Cehovin L, Fernandez G, et al. The visual object tracking vot2015 challenge results. In: Proceedings of the IEEE international conference on computer vision workshops; 2015. p. 1–23.
34. Zhan J, Zhao H, Zheng P, Wu H, Wang L. Salient superpixel visual tracking with graph model and iterative segmentation. *Cognitive Computation*. 2021; 13(4):821–832. <https://doi.org/10.1007/s12559-019-09662-y>
35. Kumar BVKV, Mahalanobis A, Juday RD. Correlation Pattern Recognition. Cambridge University Press; 2005.
36. Felsberg M. Visual tracking: Tracking in scenes containing multiple moving objects. In: *Advanced Methods and Deep Learning in Computer Vision*. Elsevier; 2022. p. 305–336.
37. Han K. Image object tracking based on temporal context and MOSSE. *Cluster Computing*. 2017; 20(2):1259–1269. <https://doi.org/10.1007/s10586-017-0800-0>
38. Rani AS, Maik V, Chithravathi B. Robust object tracking using kernalized correlation filters (KCF) and Kalman predictive estimates. In: 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE; 2017. p. 587–591.
39. Ma J, Luo H, Hui B, Chang Z. Robust scale adaptive tracking by combining correlation filters with sequential Monte Carlo. *Sensors*. 2017; 17(3):512. <https://doi.org/10.3390/s17030512> PMID: 28273840
40. Danelljan M, Hager G, Khan FS, Felsberg M. Discriminative Scale Space Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017; 39(8):1561–1575. <https://doi.org/10.1109/TPAMI.2016.2609928> PMID: 27654137
41. Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr PH. Staple: Complementary learners for real-time tracking. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1401–1409.
42. Gao L, Liu B, Fu P, Xu M, Li J. Visual tracking via dynamic saliency discriminative correlation filter. *Applied Intelligence*. 2022; 52(6):5897–5911. <https://doi.org/10.1007/s10489-021-02260-2>
43. Li Y, Zhu J, Hoi SC, Song W, Wang Z, Liu H. Robust estimation of similarity transformation for visual object tracking. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33; 2019. p. 8666–8673.
44. Abbasi S, Rezaeian M. Visual object tracking using similarity transformation and adaptive optical flow. *Multimedia Tools and Applications*. 2021; 80(24):33455–33473. <https://doi.org/10.1007/s11042-021-11344-7>
45. Lin F, Fu C, He Y, Guo F, Tang Q. BiCF: Learning Bidirectional Incongruity-Aware Correlation Filter for Efficient UAV Object Tracking. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2020.
46. Lin F, Fu C, He Y, Guo F, Tang Q. Learning Temporary Block-Based Bidirectional Incongruity-Aware Correlation Filters for Efficient UAV Object Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*. 2020; 31(6): 2160–2174. <https://doi.org/10.1109/tcsvt.2020.3023440>
47. Fu C, Xu J, Lin F, Guo F, Liu T, Zhang Z. Object Saliency-Aware Dual Regularized Correlation Filter for Real-Time Aerial Tracking. *IEEE Transactions on Geoscience and Remote Sensing*. 2020; 58(12):8940–8951. <https://doi.org/10.1109/TGRS.2020.2992301>
48. Fu C, Ye J, Xu J, He Y, Lin F. Disruptor-Aware Interval-Based Response Inconsistency for Correlation Filters in Real-Time Aerial Tracking. *IEEE Transactions on Geoscience and Remote Sensing*. 2020; p. 1–13.
49. Fu C, Yang X, Li F, Xu J, Liu C, Lu P. Learning Consistency Pursued Correlation Filters for Real-Time UAV Tracking. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020)*, Las Vegas, USA. 2020.
50. Junjie Ye FL Changhong Fu, Lu G. Multi-Regularized Correlation Filter for UAV Tracking and Self-Localization. *IEEE*. 2021.
51. Lee BY, Liew LH, Cheah WS, Wang YC. Occlusion handling in videos object tracking: A survey. In: *IOP conference series: earth and environmental science*. vol. 18. IOP Publishing; 2014. p. 012020.

52. Babenko B, Yang MH, Belongie S. Robust Object Tracking with Online Multiple Instance Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2011; 33(8):1619–1632. <https://doi.org/10.1109/TPAMI.2010.226> PMID: 21173445
53. Huang Z, Fu C, Li Y, Lin F, Lu P. Learning aberrance repressed correlation filters for real-time UAV tracking. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*; 2019. p. 2891–2900.