information

MDPI

Article

Online Learning of Discriminative Correlation Filter Bank for Visual Tracking

Jian Wei and Feng Liu *

Jiangsu Province Key Lab on Image Processing and Image Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; tdweijian@njupt.edu.cn
* Correspondence: liuf@njupt.edu.cn; Tel.: +86-025-8586-6736

Received: 4 February 2018; Accepted: 7 March 2018; Published: 9 March 2018

Abstract: Accurate visual tracking is a challenging research topic in the field of computer vision. The challenge emanates from various issues, such as target deformation, background clutter, scale variations, and occlusion. In this setting, discriminative correlation filter (DCF)-based trackers have demonstrated excellent performance in terms of speed. However, existing correlation filter-based trackers cannot handle major changes in appearance due to severe occlusions, which eventually result in the development of a bounding box for target drift tracking. In this study, we use a set of DCFs called discriminative correlation filter bank (DCFB) for visual tracking to address the key causes of object occlusion and drift in a tracking-by-detection framework. In this work, we treat thxe current location of the target frame as the center, extract several samples around the target, and perform online learning of DCFB. The sliding window then extracts numerous samples within a large radius of the area where the object in the next frame is previously located. These samples are used for the DCFB to perform correlation operation in the Fourier domain to estimate the location of the new object; the coordinates of the largest correlation scores indicate the position of the new target. The DCFB is updated according to the location of the new target. Experimental results on the quantitative and qualitative evaluations on the challenging benchmark sequences show that the proposed framework improves tracking performance compared with several state-of-the-art trackers.

Keywords: correlation score; visual tracking; discriminative correlation filter bank; occlusion

1. Introduction

In recent years, numerous visual object tracking (VOT) algorithms have been developed to overcome the limitations of VOT; these methods provide technical support for practical applications; this topic is clearly a popular research direction in the field of computer vision, and it has important applications in various areas, such as intelligent surveillance systems, human–computer interaction, autonomous driving, unmanned aerial vehicle (UAV) monitoring, video indexing, and intelligent traffic monitoring [1]. VOT is primarily used to estimate the position of a target in every frame of each video sequence. Although a major breakthrough has been made in theoretical research, the design of a robust tracking system encounters numerous difficulties in practical complex scenarios, such as illumination variation, scale variation, occlusion, deformation, motion blur, rapid motion, in-plane rotation, out-of-plane rotation, out-of-view condition, background clutter, and low resolution. The existing algorithm based on discriminative correlation filters (DCFs) causes the bounding box to deviate from the target when encountered with partial or full occlusion in complex scenarios. In this study, we focus on the challenge posed by occlusions to object tracking. We address this limitation by learning the discriminative correlation filter bank (DCFB) through the extracted samples of the current frame and accurately estimating the location of the target in the next frame. Object tracking algorithms are generally categorized as either in generation or discriminant mode. Generative trackers [2–4]

perform tracking by searching for patches most similar to the target and have an effective appearance model. Conversely, discriminative trackers [5–9] perform tracking by separating the target from the background. In recent years, the existing DCF-based trackers [10–29] have demonstrated superior performance in terms of speed on the OTB100 dataset [30]. These trackers are primarily used to learn a DCF for locating the target in a new frame by means of the coordinates of the maximum correlation response. An online update is then performed on the basis of the new location. The popularity of correlation filters for object tracking is due to several important attributes [24,31]. First, the conversion of the correlation operations in the time domain into element-wise multiplication in the Fourier domain effectively avoids the convolution operation and reduces the time overhead. Correlation filter tracking has achieved remarkable results in terms of this principle. Second, the correlation filters can take advantage of the cyclic shift version of the sample for training. Third, the correlation filters will consider the target context information to have more discriminative power than the individual target appearance model. These advantages make DCF more suitable for visual tracking. Nevertheless, existing DCF-based tracking algorithms have two major limitations. First, a single tracker is prone to drift in the case of severe interferences, such as deformation, background clutter, scale variation, and occlusion. Moreover, a single tracker cannot easily recover, which eventually results in tracking failure. Second, the use of all samples to learn a single DCF cannot effectively handle a major appearance change and thus cannot deal with partial or full occlusion well. We use a set of DCFs called DCFB for visual tracking to address the key causes of object occlusion and drift in a tracking-by-detection framework. Figure 1 shows the general overview of our proposed tracking framework. By extracting several samples and treating the current position of the target as the center α for the radius, the extracted samples are sufficient to cover the target itself for training the DCFB. In the training step, each sample corresponds to a DCF in the DCFB, and the correlation operation is performed in the frequency domain.The coordinates of the largest correlation scores indicate the position of the target. The sliding window extracts numerous samples within a large radius γ of the area where the object in the next frame was previously located. These samples are used for the previously trained DCFB to perform the correlation operation in the Fourier domain to estimate the new location of the object. The sample corresponding to the maximum response value is the target to be tracked.

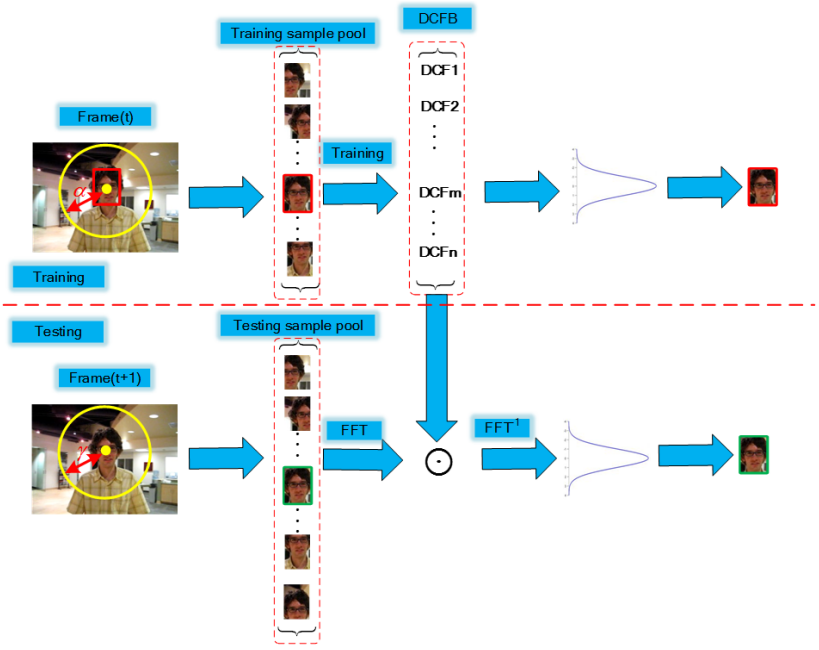

Figure 1. Overview of the proposed DCFB-based visual tracking algorithm. The operator $\odot$ is the Hadamard product.

The contribution of this study are as follows. First, we design a DCFB-based tracker with a set of DCFs. Second, we treat the current location of the target as the center, extract several samples around the target to ensure that the sample is sufficient to cover the target, and then perform online learning of DCFB. Third, our algorithm is compared with several state-of-the-art correlation filter-based trackers in terms of quantitative and qualitative evaluations on the OTB100 dataset.

The rest of this paper is organized as follows. Section 2 presents a summary of related studies. Section 3 describes our proposed tracking framework. Sections 3.1 provides the baseline tracker. Sections 3.2 and 3.3 present the DCFB framework and the DCFB tracking algorithm, respectively. Section 4 provides the experimental evaluations and analysis. Section 5 concludes the paper.

2. Related Works

Research on correlation filter-based algorithms has been extensively conducted in recent years for the development of VOT, and various correlation filter-based trackers are continuously being used in the frequency domain for real-time visual tracking mainly due to the low computational overhead. For these correlation filter-based tracking algorithms, the desired correlation output is designed as a Gaussian distribution function whose peak coordinates indicate the center of the tracking target, whereas a region other than the target produces a low response. Bolme et al. [10] developed the first correlation filter-based tracking algorithm. This algorithm learns an adaptive correlation filter using the minimum output sum of squared error (MOSSE) for object tracking. This MOSSE filter is intended to minimize the total squared error between the desired and the actual correlation outputs on grayscale samples and can be effectively calculated by fast fourier transformation (FFT) and pointwise operations in the frequency domain. It has achieved real-time performance with a speed of several hundreds frames per second.

The circulant structure with kernel (CSK) tracking algorithm [13] was proposed to adopt the kernel trick to improve the efficiency of the correlation filter-based tracker. This algorithm uses illumination intensity features for visual tracking. The kernelized correlation filters (KCF) tracking algorithm [14] adopts the histogram of oriented gradients (HOG) feature instead of the illumination intensity feature to further improve the performance of the tracker. The KCF tracking algorithm [14] achieves excellent results in terms of speed on the OTB-2013 dataset [32]. For such algorithms, negative samples are used to enhance the discriminative ability of the track-by-detection framework while investigating the structure of the circular matrix to improve the tracking efficiency. Previous studies [12,16–18,26,29,33–38] have developed the KCF algorithm to improve tracking performance. However, the KCF algorithm cannot achieve online multiscale detection and updating. Discriminative Scale Space Tracking (DSST) [11] and Scale Adaptive with Multiple Features (SAMF) [33] algorithms address the shortcomings of KCF.

In VOT2014 [39], DSST and SAMF ranked first and second, respectively. The DSST algorithm regards object tracking as two independent problems in terms of translation and scale estimations. First, the HOG feature is used to train a translation correlation filter, which is responsible for detecting object center translation and then to train another scale correlation filter to detect the object scale change. fDSST [40] is an extended version of DSST. Danelljan et al. [40] performed feature compression and scale filter acceleration through principal component analysis (PCA) dimensionality while tracking performance and frame rate have been significantly improved. In the work, we adopt the DSST algorithm as our baseline. Readers are referred to [11] for additional details about the DSST algorithm.

The part-based tracking methods [17,29,41,42] use correlation filter to show superior performance in dealing with occlusion problems during tracking. However, target response adaptation tracking algorithm [23] also achieves an amazing tracking performance in dealing with the occlusion scenarios challenge. In [22], the authors first proposed such a tracking framework that the loss function consisting of the target appearance model and training sample weight is optimized for solving occlusions. Thereby, this adaptively reduces the number of corrupted samples for accurate tracking. In our work, the proposed tracking framework is different from the above-mentioned algorithms, and embodies

the following aspects: First, the target blocks used for training correlation filter are different in the same frame. In our method, the sample instance instead of the target part to train the tracker; second, the boundary effect brought by the sample cyclic shift makes the discrimination ability of the correlation filter become weak and the detected target position is inaccurate in the next frame. Our approach does not use approximate cyclic shift samples; third, the target positioning method is different in the next frame. Our method is to find the maximum correlation response of the target sample rather than the maximum of the joint confidence map of the target parts. A single correlation filter often has poor discriminative ability under significant occlusion. Therefore, the proposed DCFB tracking framework effectively solves the occlusion problems in tracking.

The recent advancement of the performance of the DCF-based tracking algorithm is driven by the reduction of boundary effects [24,25,43] and the adoption of deep features [20,26–28]. When the target moves rapidly and occlusion, the error samples produced by the boundary effect will cause the correlation filter to be weakly discriminated, which results in tracking failure. Danelljan et al. [20] recently introduced a continuous-domain formulation of the DCF called continuous convolution operators tracking (C-COT), which enables integration of multiresolution deep-feature maps, leading to top performance in the VOT2016 challenge [44]. The C-COT model is extremely complex that it sacrifices the real-time capabilities of tracking in exchange for performance standards. Efficient Convolution Operators (ECO) [28] algorithm is the accelerated version of C-COT that optimizes the three aspects of model size, sample set size, and update strategy to achieve acceleration; the tracking speed increases by 20 times compared to that of C-COT. The existing ECO algorithm is the best correlation filter-based tracking algorithm in terms of performance.

Early DCF-based tracking methods [10–29] demonstrate excellent performance in terms of speed. However, correlation filter-based tracking methods with deep features [20,26–28,36,45] have been demonstrated to achieve remarkable performance. The low-level resolution of the deep feature is high, and the high-level resolution has complete semantic information; the discrimination and invariance of the feature are strong; thus, the tracking performance is evidently improved. The combination of correlation filters and convolutional neural networks (CNN) provides a research opportunity for improving tracking performance. In this work, we develop a novel DCF-based discriminative tracking algorithm that performs tracking efficiently and effectively.

3. Our Tracking Framework

In this section, we present a visual tracking algorithm based on DCFB. In contrast with existing DCF-based trackers [10–29] that independently learn a correlation filter on a set of observed sample patches, the proposed DCFB-based tracking algorithm is calculated with an equal number of observed samples, which are fully optimized to each target pixel in the frequency domain and improve tracking performance. Our proposed tracking algorithm can also effectively address object occlusion, which is a problem encountered in visual tracking. The proposed framework will be discussed in three parts. The employed baseline tracker and DCFB are presented in Sections 3.1 and 3.2, respectively, and the DCFB-based tracking algorithm is discussed in Section 3.3.

3.1. Baseline Approach

The correlation filter h is optimized by minimizing the following equation:

$$\varepsilon = \left\| \sum_{l=1}^{d} h^l * f^l - g \right\|^2 + \lambda \sum_{l=1}^{d} \left\| h^l \right\|^2, \tag{1}$$

where d is the number of feature dimensions, g is the Gaussian function label, f is the training example, and λ is the regularization term coefficient. The closed solution of Equation (1) is as follows:

$$H^l = \frac{\overline{G}F^l}{\sum_{k=1}^{d} \overline{F^k}F^k + \lambda}. \tag{2}$$

The updated plan is as follows:

$$\begin{aligned} A_t^l &= (1-\eta)A_{t-1}^l + \eta\overline{G_t}F_t^l, \\ B_t &= (1-\eta)B_{t-1} + \eta \sum_{k=1}^{d} \overline{F_t^k}F_t^k \end{aligned}, \tag{3}$$

where η is a learning rate parameter. The new position of the target is estimated by the maximum correlation score y on the candidate patch z in a new frame. The maximum correlation score y is computed as

$$y = \mathcal{F}^{-1}\left\{ \frac{\sum_{l=1}^{d} \overline{A^l}Z^l}{B+\lambda} \right\}. \tag{4}$$

3.2. DCFB

Most DCF-based trackers traditionally trains a DCF on a set of observed sample patches in the first frame. The position of the target is estimated in the sequential frames, and the DCF is updated according to the new position of the target. However, once the target is partially or fully occluded in these occlusion scenarios, a single correlation filter cannot easily manage the major changes in the appearance of consecutive frames, which prevents the accurate updating of the trained correlation filter. The accumulated error then causes the tracking bounding box to drift the target. We propose a novel method with a group of trained DCFs called DCFB to address the key causes of object occlusion encountered in visual tracking. Figure 1 illustrates the entire tracking procedure.

Similar to related studies [11,40,46,47], the DCFB filter can be designed as an objective function for N training image blocks in a frame, which can be expressed as

$$h = \arg\min_h \frac{1}{N} \sum_{i=1}^{N} \left(\left\| \sum_{k=1}^{K} f_i^k \otimes h^k - g_i \right\|_2^2 \right) + \lambda \sum_{k=1}^{K} \left\| h^k \right\|_2^2, \tag{5}$$

where N is the number of training samples in a frame, K is the number of feature channels, g_i is the desired correlation output associated with the training example f_i , and λ is the regularization term coefficient. The operator $\otimes$ is the time-domain correlation operation. A expression in the frequency domain can be formulated as

$$H = \arg\min \frac{1}{N} \sum_{i=1}^{N} \left(\left\| \sum_{k=1}^{K} F_i^k \odot \bar{H}^k - G_i \right\|_2^2 \right) + \lambda \sum_{k=1}^{K} \left\| H^k \right\|_2^2, \tag{6}$$

where F, H, and G denote the Fourier transforms of f, h, and g, respectively. The operator $\odot$ is the Hadamard product, and $\bar{H}$ is the complex conjugate of H.

This optimization problem can be solved efficiently in the frequency domain where it has the following closed-form expression:

$$H = \frac{1}{N} \sum_{i=1}^{N} \frac{\bar{G}_i F_i}{\sum_{k=1}^{K} \bar{F}_i^k F_i^k + \lambda}. \tag{7}$$

The online update program uses Equation (3). The correlation score y at the candidate sample z is calculated as follows:

$$y = \mathcal{F}^{-1}\left\{\bar{H} \odot Z\right\}. \tag{8}$$

The new position of the target is estimated by the maximum correlation score y in a new frame.

The main idea of the DCFB-based tracking framework is to use a series of DCF to perform the correlation operation with the same number of candidate samples in the frequency domain to find the sample with the maximum correlation response because the center of the sample is the new location of the target. Numerous candidate samples were obtained from the next frame in order to capture effective the appearance of object target. Obviously, when target is partially occluded, the effective appearance of the remaining visible samples can still provide reliable cues for tracking.

In the work, we adopt the DSST algorithm as our baseline tracker. During tracking, the baseline tracker of each candidate sample has a response map, which is to say, each baseline tracker (DSSTs) in the correlation filter bank will have a correlation response output, the correlation response of the remaining visible sample when the occlusion occurs is the maximum value, and the correlation response of other baseline trackers (DSSTs) value is relatively small, which can help predict the object position by searching for the location of the maximal value of the map. Based on the newly detected target position, we update and learn the baseline tracker again to locate the target in the next frame. Due to the different search radius used to crop the samples in the current frame and the next frame, the samples used for learning and testing are different. This is explained in Section 3.3. The parameter values of the baseline trackers (DSSTs) in the correlation filter bank are the same, otherwise, the tracking system model becomes very complicated in the training and updating phases, which directly affects the performance of the tracker, and even increases the time overhead. Thus, the parameters are fixed during tracking for all sequences.

3.3. Tracking Algorithm

In this section, we discuss the DCFB-based algorithm for visual tracking in detail.

At the training stage, $\ell_t(x_0)$ is used to denote the location of target at the t-th frame, to crop out a set of image patches within a search radius α centering at object location $\ell_t(x_0)$ with Equation (9), and the cropped image patches are then used to train the DCFB. The desired output of the candidate samples is the Gaussian distribution function when the DCFB and the dense samples perform the correlation operation in the frequency domain. The maximum score of the correlation output corresponding to the coordinates indicates the location of the target $\ell_t(x_0)$. Figure 2 shows an overview of the training:

$$X^{\alpha} = \left\{x : \|\ell_t(x) - \ell_t(x_0)\| < \alpha\right\}. \tag{9}$$

The location of the target is estimated according as follows:

$$x_0 = \arg\max_x(\max(y_1), \max(y_2), ..., \max(y_n)). \tag{10}$$

At the testing stage, the new position of the target at the $(t+1)$-th frame is estimated by using the DCFB that has been trained in the previous frame.

The position of the target $\ell_t(x_0)$ of the previous frame is taken as the center to extract the samples within a large search radius γ using Equation (11). We then search for the sample of the maximum confidence map by using the already-trained DCFB with Equation (12). Figure 3 shows an overview of the testing:

$$X^{\gamma} = \left\{x : \|\ell_{t+1}(x) - \ell_t(x_0)\| < \gamma\right\}, \gamma \gg \alpha, \tag{11}$$

$$x^* = \arg\max_x(\max(y_1), \max(y_2), ..., \max(y_n)). \tag{12}$$

A new position of the target $\ell_{t+1}(x^*)$ is thus detected in the $(t+1)$-th frame. Depending on the new location of the target $\ell_{t+1}(x^*)$ at the $(t+1)$-th frame , the tracking process will repeat the same training and testing procedures.

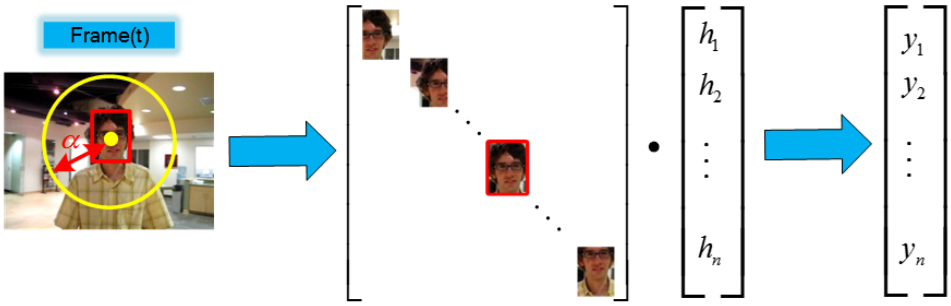

Figure 2. Overview of the training process.

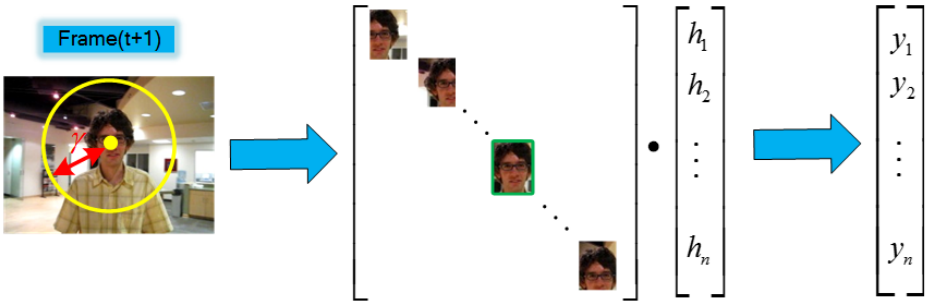

Figure 3. Overview of the testing process.

Algorithm 1 presents a summary of our tracking method. The algorithm consists of two parts, namely, tracking and updating. During the tracking, the optimal candidate patch is obtained by applying the maximum confidence function. During the updating, the detected new location of the target and the correlation filter model are updated.

Algorithm 1 Our tracking method.

Input:
1. Testing sample set $X^\gamma = \{x : \|\ell_{t+1}(x) - \ell_t(x_0)\| < \gamma\}$.
2. Previous position of the target $\ell_t(x_0)$.

Output:
1. The sample with the maximum confidence as in Equation (12).
2. Current position of the target $\ell_{t+1}(x^*)$.

Tracking:
1. Crop out a set of candidate samples using Equation (11).
2. Find the sample with the maximum confidence as in Equation (12).
3. Set $\ell_{t+1}(x^*)$ to the new location of the target.

Updating:
1. Update target location.
2. Update correlation filter model.

4. Experiments

We evaluate the proposed tracking method with comparison to state-of-the-art trackers on the OTB100 dataset [30], which contains 100 problematic image sequences. The details of implementation

are presented in Section 4.1. This section mainly includes experimental settings, comparison of trackers, and evaluation criteria. We then present the experimental results and analysis in Section 4.2.

4.1. Implementation

Our algorithm is implemented through MATLAB R2017a software platform (MathWorks, Natick, MA, USA). All experiments are performed on an Intel 4 core i5-3470 3.20GHz CPU with 12 GB RAM. The parameters are fixed during tracking for all sequences. The regularization parameter λ is set to 0.01, and the learning rate η is set to 0.025, which is equivalent to that of the DSST [11]. We use a search radius $\alpha = 4$ to crop the training samples at the current frame and $\gamma = 30$ to crop the testing samples in the next frame. We set a large search radius γ to ensure that the effective feature of the target are obtained when the target appears again for accurate localization. We use the HOG feature for image representation. The feature vector is extracted using a cell size of 4×4, and the number of orientation bins are set to 9.

The performance of the proposed tracking algorithm is evaluated by comparing it with four sophisticated tracking algorithms, namely, DSST [11], MEEM [7], STC [19], and KCF [14]. The results on the performance of these trackers are either provided by the authors on their websites or obtained through their raw codes with the default setting.

We present the results by using the one-pass evaluation (OPE) with precision and success plots. The precision plot metric measures the rate of frame within a certain threshold. We report the precision plot values at a threshold of 20 pixels for all the trackers. The success plot metric measures the overlap ratio between the tracking bounding box and the ground-truth bounding box. This metric is defined as $\frac{area(B_T \cap B_G)}{area(B_T \cup B_G)}$, where B_T and B_G are the tracking bounding box and the ground-truth bounding box, respectively. If the score is greater than 0.5, then the tracking is successful; a higher score indicates better accuracy. We provide the success plot values at a threshold of 0.5 for all the trackers. The precision and success plots present the mean results over the OTB100 dataset. Finally, we provide the results of the quantitative and qualitative evaluations of the proposed tracking algorithm and the tracking algorithms compared.

4.2. Experimental Results and Analysis

We performed quantitative and qualitative evaluations on the OTB100 dataset. We present the evaluation results, which were obtained from 100 sequences and occlusion attribute sequences by means of OPE with precision and success plots, respectively. The details are described below.

4.2.1. Quantitative Evaluation

Figure 4 depicts the overall evaluation results of the proposed tracking algorithm and the other four trackers compared. Among the trackers, the proposed tracking algorithm exhibits excellent performance in terms of distance precision (DP) and overlap success.

As shown in Figure 5, our approach performs favorably on DP and overlap success in terms of occlusion video attributes annotated in the OTB100 dataset. Our tracking algorithm generally outperforms the comparison trackers in terms of robustness and performance mainly because the large number of extracted candidate samples can traverse the appearance changes of the target to capture the target effectively. Existing DCF-based trackers are generally used to train a DCF on a set of observed sample patches with multifeature channels. However, once the target is partially or fully occluded in these occlusion scenarios, a single correlation filter can hardly manage consecutive frames with major appearance changes, which prevents the trained correlation filter from being accurately updated. The accumulated error then causes the tracking bounding box to drift the target. The proposed tracking algorithm uses DCFB to perform correlation operation in the Fourier domain with dense prediction samples. Therefore, the position of the target can be correctly identified.

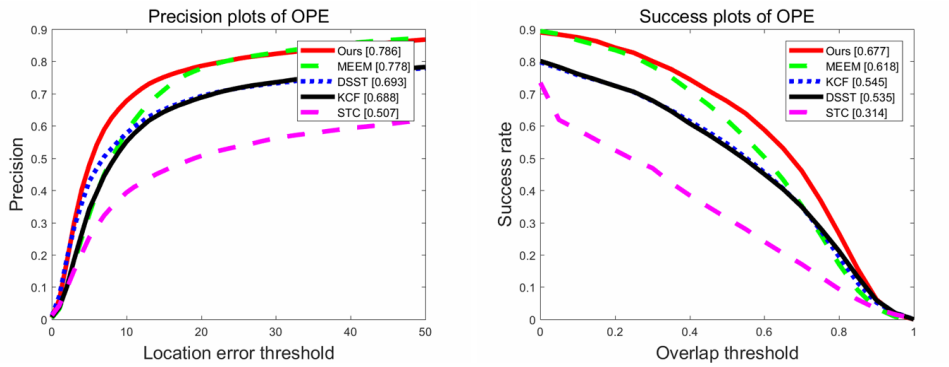

Figure 4. Precision and success plots over all sequences using OPE on the OTB100 dataset. The average score on DP at 20 pixels for each tracker is shown in the legend of the precision plot. The legend of the success plot contains the area-under-the-curve (AUC) score for each tracker.

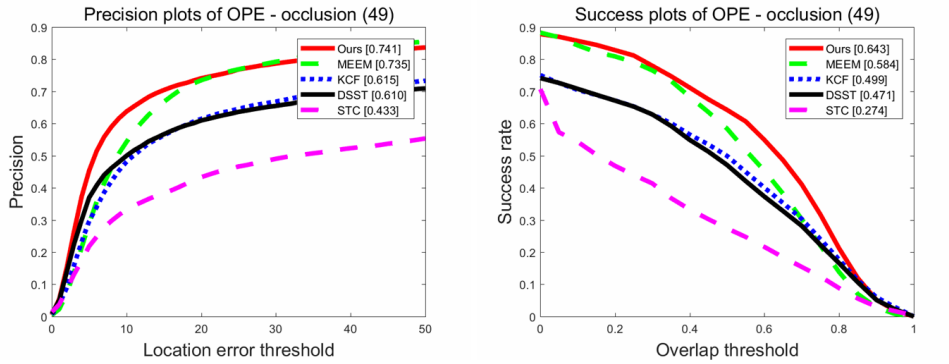

Figure 5. Precision and success plots of OPE during occlusion. The legend contains the AUC score for each tracker.

Tables 1 and 2 present the mean results on our proposed tracking algorithm and comparison trackers on the 100 benchmark sequences and occlusion sequences, respectively. Our tracking algorithm outperforms the state-of-the-art algorithms in terms of tracking occlusion. Considering that this study mainly aims to address occlusion, our proposed tracking algorithm is considered to have relatively robust performance during tracking. However, the speed is a drawback.

Table 1. Comparison with state-of-the-art trackers on the OTB100 dataset. The results are presented in terms of mean overlap precision (OP) (%) at an overlap threshold of 0.5, DP (%) at a threshold of 20 pixels, and fps. The optimal results are highlighted in bold.

Metrics	Ours	DSST	MEEM	KCF	STC
Mean OP	**67.7**	53.5	61.8	54.5	31.4
Mean DP	**78.6**	69.3	77.8	68.8	50.7
Mean fps	1.2	15	24	107	**148**

Table 2. Comparison with state-of-the-art trackers on the occlusion sequences. The results are presented in terms of mean OP (%) at an overlap threshold of 0.5, DP (%) at a threshold of 20 pixels, and fps. The optimal results are highlighted in bold.

Metrics	Ours	DSST	MEEM	KCF	STC
Mean OP	**64.3**	47.1	58.4	49.9	27.4
Mean DP	**74.1**	61	73.5	61.5	43.3
Mean fps	1.2	15	24	107	**148**

4.2.2. Qualitative Evaluation

We present a qualitative evaluation of our proposed approach in comparison with four state-of-the-art trackers (i.e., DSST [11], MEEM [7], STC [19], KCF [14]) from the literature in terms of eight challenges of occlusion sequences in Figure 6. Tracking results show that our tracking algorithm exhibits relatively robust performance in these partially or fully occluded sequences.

The tracking bounding box of the DSST and STC algorithms drift, which result in tracking failure when target objects undergo significant appearance change due to heavy occlusion (Basketball and Jogging1) and rapid motion (coke and Soccer). The proposed approach generally demonstrates robustness in these occlusion sequences and performs well in tracking objects throughout the sequence.

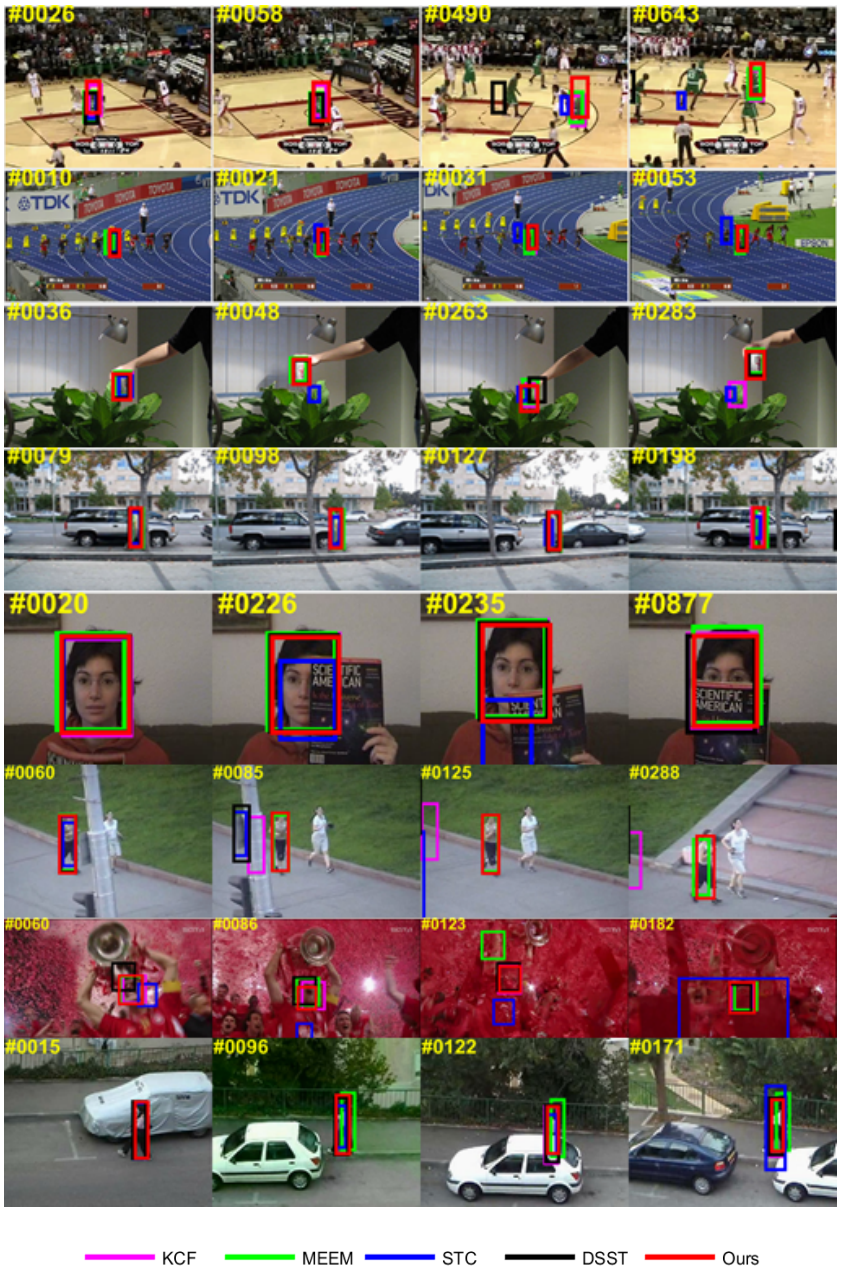

Figure 6. Qualitative evaluation of our approach in comparison with four state-of-the-art trackers. Tracking results for eight challenges of occlusion sequences (*Basketball*, *Bolt*, *coke*, *David3*, *Faceocc1*, *Jogging1*, *Soccer*, *and Woman*) from the OTB100 dataset are shown. Our approach outperforms the state-of-the-art trackers in these occlusion scenarios.

4.2.3. Experiment Analysis

The experiment in this work focused only on the benchmark OTB100 dataset [30]. Although the proposed approach demonstrates robustness in these sequences, it was not tested on the VOT2016 datasets [44]; this gap will be addressed in future work. The correlation filter bank tracking framework designed in our work is mainly that a part of the baseline trackers can effectively obtain the feature of the target visible parts when the occlusion occurs, so as to achieve the purpose of accurate tracking. The baseline trackers in the correlation filter bank are the same; otherwise, a large number of different parameters will be configured in the online training and updating stage, which will directly affect the tracking performance and increase the time overhead. Therefore, the parameters are fixed during tracking for all sequences. After estimating the position of the target in the test frame, the baseline tracker is again trained and updated according to the new target position. Thus, the samples of retraining the baselines are not the same. Our proposed approach does not address scale variation and requires further study in terms of model updates. This study aims to achieve effective object tracking in the context of occlusion challenging scenarios. Once the target is fully occluded, it will lead to inaccurate estimation of the position and cause drift problem, which is also a difficult problem to be solved in long-time tracking. In order to solve the problem of recovery in the case of tracking failure, the next step will be to study the occlusion detection model. Moreover, to ensure that the target is within the bounding box between consecutive frames, a large number of candidate samples are obtained with a large search radius, resulting in relatively heavy computational overhead and eventually sacrificing the speed in exchange for tracking performance.

5. Conclusions

In this study, we present the framework of the DCFB-based tracking algorithm that incorporates multiple correlation filters into the training and testing stages. Experiments on the OTB100 dataset show that our approach improves tracking performance in contrast with several state-of-the-art trackers. Our proposed algorithm performs well in terms of DP and overlap success in the context of occlusion scenarios (Figure 5). Our work also has several limitations in terms of scale variation and real-time tracking. Incorporating an advanced scale estimation approach [11] into our current tracking framework can address scale variation throughout the tracking process. The main reason for slow tracking is that dense candidate samples obtained within a large search radius for multiple correlation operations in the frequency domain results in heavy computational overhead when the test frame comes. In future work, we will study how to select important samples from dense candidate samples in reducing the computational overhead to improve tracking speed. Another research direction is to investigate efficient feature fusion strategies used for the framework of the DCFB for visual tracking.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China under Grant 61501260 and Grant 61471201, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20130867, in part by the Jiangsu Province Higher Education Institutions Natural Science Research Key Grant under Project 13KJA510004, in part by the Peak Of Six Talents in Jiangsu Province under Grant RLD201402, in part by the Natural Science Foundation of NJUPT under Grant NY214031, in part by the 1311 Talent Program of NJUPT, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Author Contributions: Jian Wei proposed the original idea, built the simulation model and completed the manuscript. Feng Liu modified and refined the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Smeulders, A.W.M.; Chu, D.M.; Cucchiara, R.; Calderara, S.; Dehghan, A.; Shah, M. Visual Tracking: An Experimental Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1442–1468.
2. He, S.; Yang, Q.; Lau, R.W.H.; Wang, J.; Yang, M.H. Visual Tracking via Locality Sensitive Histograms. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2427–2434.
3. Jia, X.; Lu, H.; Yang, M.H. Visual tracking via adaptive structural local sparse appearance model. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1822–1829.
4. Sevilla-Lara, L.; Learned-Miller, E. Distribution fields for tracking. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1910–1917.
5. Babenko, B.; Yang, M.H.; Belongie, S. Visual tracking with online Multiple Instance Learning. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 983–990.
6. Kalal, Z.; Mikolajczyk, K.; Matas, J. Tracking-Learning-Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1409–1422.
7. Zhang, J.; Ma, S.; Sclaroff, S. MEEM: Robust tracking via multiple experts using entropy minimization. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 188–203.
8. Avidan, S. Support vector tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1064–1072.
9. Avidan, S. Ensemble Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 261–271.
10. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.
11. Danelljan, M.; H*ä*ger, G.; Khan, F.; Felsberg, M. Accurate scale estimation for robust visual tracking. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014; BMVA Press: Durham, UK, 2014.
12. Danelljan, M.; Khan, F.S.; Felsberg, M.; Van de Weijer, J. Adaptive Color Attributes for Real-Time Visual Tracking. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.
13. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 702–715.
14. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-Speed Tracking with Kernelized Correlation Filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596.
15. Hong, Z.; Chen, Z.; Wang, C.; Mei, X.; Prokhorov, D.; Tao, D. MUlti-Store Tracker (MUSTer): A cognitive psychology inspired approach to object tracking. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 749–758.
16. Li, Y.; Zhu, J.; Hoi, S.C.H. Reliable Patch Trackers: Robust visual tracking by exploiting reliable patches. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 353–361.
17. Liu, T.; Wang, G.; Yang, Q. Real-time part-based visual tracking via adaptive correlation filters. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4902–4912.
18. Ma, C.; Yang, X.; Zhang, C.; Yang, M.H. Long-term correlation tracking. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5388–5396.
19. Zhang, K.; Zhang, L.; Liu, Q.; Zhang, D.; Yang, M.H. Fast visual tracking via dense spatio-temporal context learning. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 127–141.

20. Danelljan, M.; Robinson, A.; Khan, F.S.; Felsberg, M. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 472–488.
21. Galoogahi, H.K.; Sim, T.; Lucey, S. Multi-channel Correlation Filters. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 3072–3079.
22. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Adaptive Decontamination of the Training Set: A Unified Formulation for Discriminative Visual Tracking. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 1430–1438.
23. Bibi, A.; Mueller, M.; Ghanem, B. Target response adaptation for correlation filter tracking. In Proceedings of the European Conference on Computer Vision, Seattle, WA, USA, 27–30 June 2016; Springer: Cham, Switzerland, 2016; pp. 419–433.
24. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Learning Spatially Regularized Correlation Filters for Visual Tracking. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4310–4318.
25. Galoogahi, H.K.; Sim, T.; Lucey, S. Correlation filters with limited boundaries. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4630–4638.
26. Ma, C.; Huang, J.B.; Yang, X.; Yang, M.H. Hierarchical Convolutional Features for Visual Tracking. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 3074–3082.
27. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Convolutional Features for Correlation Filter Based Visual Tracking. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 7–13 December 2015; pp. 621–629.
28. Danelljan, M.; Bhat, G.; Khan, F.S.; Felsberg, M. ECO: Efficient Convolution Operators for Tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6931–6939.
29. Liu, S.; Zhang, T.; Cao, X.; Xu, C. Structural Correlation Filter for Robust Visual Tracking. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 4312–4320.
30. Wu, Y.; Lim, J.; Yang, M.H. Object Tracking Benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1834–1848.
31. Ma, C.; Xu, Y.; Ni, B.; Yang, X. When Correlation Filters Meet Convolutional Neural Networks for Visual Tracking. *IEEE Signal Process. Lett.* **2016**, *23*, 1454–1458.
32. Wu, Y.; Lim, J.; Yang, M.H. Online Object Tracking: A Benchmark. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2411–2418.
33. Li, Y.; Zhu, J. A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 254–265.
34. Mueller, M.; Smith, N.; Ghanem, B. Context-Aware Correlation Filter Tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1387–1395.
35. Zhang, T.; Xu, C.; Yang, M.H. Multi-task Correlation Particle Filter for Robust Object Tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4819–4827.
36. Choi, J.; Chang, H.J.; Yun, S.; Fischer, T.; Demiris, Y.; Choi, J.Y. Attentional Correlation Filter Network for Adaptive Visual Tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4828–4837.
37. Wang, M.; Liu, Y.; Huang, Z. Large Margin Object Tracking with Circulant Feature Maps. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4800–4808.
38. Chen, S.; Liu, B.; Chen, C.W. A Structural Coupled-Layer Tracking Method Based on Correlation Filters. In Proceedings of the International Conference on Multimedia Modeling, Reykjavik, Iceland, 4–6 January 2017; Springer: Cham, Switzerland, 2017; pp. 65–76.

39. Kristan, M.; Roman, P.; Jiri, M.; Luka, Č.; Georg, N.; Tomáš, V.; Gustavo, F.; Alan, L.; Aleksandar, D.; Alfredo, P.; et al. The Visual Object Tracking VOT2014 Challenge Results. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 191–217.
40. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Discriminative Scale Space Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1561–1575.
41. Fan, H.; Xiang, J. Robust Visual Tracking via Local-Global Correlation Filter. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, CA, USA, 4–9 February 2017; pp. 4025–4031.
42. Lukežič, A.; L.Čehovin.; Kristan, M. Deformable Parts Correlation Filters for Robust Visual Tracking. *IEEE Trans. Cybern.* **2018**, *PP*, 1–13.
43. Galoogahi, H.K.; Fagg, A.; Lucey, S. Learning Background-Aware Correlation Filters for Visual Tracking. *arXiv* **2017**, arXiv:1703.04590.
44. Kristan, M.; Roman, P.; Jiri, M.; Luka, Č.; Georg, N.; Tomáš, V.; Gustavo, F.; Alan, L.; Aleksandar, D.; Alfredo, P.; et al. The Visual Object Tracking VOT2016 Challenge Results. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 777–823.
45. Valmadre, J.; Bertinetto, L.; Henriques, J.; Vedaldi, A.; Torr, P.H.S. End-to-End Representation Learning for Correlation Filter Based Tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5000–5008.
46. Boddeti, V.N.; Kanade, T.; Kumar, B.V.K.V. Correlation Filters for Object Alignment. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2291–2298.
47. Babenko, B.; Yang, M.H.; Belongie, S. Robust object tracking with online multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1619–1632.