electronics

Article

Context-Aware and Occlusion Handling Mechanism for Online Visual Object Tracking

Khizer Mehmood [1], Abdul Jalil [1], Ahmad Ali [2], Baber Khan [1], Maria Murad [1], Wasim Ullah Khan [3,*] and Yigang He [3]

[1] Department of Electrical Engineering, International Islamic University, Islamabad 44000, Pakistan; khizer.mehmood@iiu.edu.pk (K.M.); abdul.jalil@iiu.edu.pk (A.J.); baber.khan@iiu.edu.pk (B.K.); maria.murad@iiu.edu.pk (M.M.)
[2] Department of Software Engineering, Bahria University, Islamabad 44000, Pakistan; ahmadali1655@hotmail.com
[3] School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China; yghe1221@whu.edu.cn
* Correspondence: kwasim814@whu.edu.cn

Abstract: Object tracking is still an intriguing task as the target undergoes significant appearance changes due to illumination, fast motion, occlusion and shape deformation. Background clutter and numerous other environmental factors are other major constraints which remain a riveting challenge to develop a robust and effective tracking algorithm. In the present study, an adaptive Spatio-temporal context (STC)-based algorithm for online tracking is proposed by combining the context-aware formulation, Kalman filter, and adaptive model learning rate. For the enhancement of seminal STC-based tracking performance, different contributions were made in the proposed study. Firstly, a context-aware formulation was incorporated in the STC framework to make it computationally less expensive while achieving better performance. Afterwards, accurate tracking was made by employing the Kalman filter when the target undergoes occlusion. Finally, an adaptive update scheme was incorporated in the model to make it more robust by coping with the changes of the environment. The state of an object in the tracking process depends on the maximum value of the response map between consecutive frames. Then, Kalman filter prediction can be updated as an object position in the next frame. The average difference between consecutive frames is used to update the target model adaptively. Experimental results on image sequences taken from Template Color (TC)-128, OTB2013, and OTB2015 datasets indicate that the proposed algorithm performs better than various algorithms, both qualitatively and quantitatively.

Keywords: spatio-temporal context; target tracking; Kalman filter

Citation: Mehmood, K.; Jalil, A.; Ali, A.; Khan, B.; Murad, M.; Khan, W.U.; He, Y. Context-Aware and Occlusion Handling Mechanism for Online Visual Object Tracking. *Electronics* **2021**, *10*, 43. https://doi.org/10.3390/electronics10010043

Received: 22 November 2020
Accepted: 24 December 2020
Published: 29 December 2020

1. Introduction

Visual Object Tracking (VOT) is an active research topic in computer vision and machine learning due to extensive applications in areas including gesture recognition [1], sports analysis [2], visual surveillance [3], medical diagnosis [4], autonomous vehicles [5,6] and radar navigation systems [7–9]. Various factors such as partial or full occlusion, background clutter, illumination variation, deformation and other factors in the environment complicate a tracking problem [10–12]. Tracking methods are categorized as generative [13] and discriminative methods [14]. Generative tracking methods focus on constructing an appearance model for target representation and search regions with high scores as results. Discriminative tracking methods treat object tracking as a classification problem by distinguishing the target from its background. Both types of tracking approaches are widely referred to in literature and have their own pros and cons in various scenarios. Generative trackers perform better analysis in case of availability of small training data. However, these trackers only consider object similarity which leads to loss of useful information

around the target that might drift the tracker when the target undergoes occlusion or scale variation. However, discriminative trackers perform better analysis in the case of large training data. However, these trackers cannot adapt adequately when the appearance of target changes, due to which, tracking is affected when the target changes its shape or size during motion [15].

1.1. Related Work

With recent advancement in visual tracking, various competitive methods have been proposed for target tracking. Zhang et al. [16] proposed network padding, stride and respective field size-based network architecture for Siamese trackers. Rahman et al. [17] proposed a Siamese network-based tracker, which utilizes an attention module inside the feature refine network to discriminate between the target and background. Zhang et al. [18] proposed a tracking method which constructs a correlation filter learning model by using handcrafted features extracted from a convolutional neural network and uses hierarchical peak to side lobe ratio (PSR) for activation of the classifier. Dai et al. [19] presented adaptive regularization in a correlation filter which can learn and update the target model according to appearance variations during tracking. Javed et al. [20] proposed a deep correlation filter-based tracking method, by utilizing both forward and backward tracking information between the regression target and response map. Despite the fact that deep learning-based methods achieve favorable results, the complexity of these methods is still higher with the requirement of offline training.

Zhang et al. [21] proposed a fast algorithm which effectively uses Spatio-temporal context (STC) information for online tracking by signifying Spatio-temporal relationships between the target and its local contexts in a Bayesian framework. The tracking problem is resolved by maximizing the confidence map which uses target location prior information. Tian et al. [22] proposed an enhanced STC tracker to address occlusion through the incorporation of a patch-based occlusion detection mechanism in the STC framework. Chen et al. [23] proposed an improved STC tracker to address occlusion by incorporating a Kalman filter for prediction of the target location in case of occlusion. Munir et al. [24] proposed a modified STC tracker to address occlusion by incorporating a Kalman filter for prediction of target location in case of occlusion and implemented it for a real-time eye tracking application. Cui et al. [25] proposed an amended STC tracker to address limitation of full occlusion. They incorporated an occlusion detection mechanism which consists of three stages during which motion and template update information is stored and used when the target is occluded. Yang et al. [26] proposed an enhanced STC tracker to address occlusion by incorporating a PSR-based occlusion feedback mechanism for the model and scale update in the STC framework. Yang et al. [27] proposed an improved STC tracker to address occlusion through incorporating a Kalman filter for prediction of target location and uses Euclidean distance to detect occlusion. Zhang et al. [28] proposed a motion aware correlation filter (MACF) which predicts position and scale of the target in the next frame by utilizing instantaneous motion estimation.

Lu et al. [29] proposed RetinaTrack, an efficient joint model for detection and tracking which modifies single stage RetinaNet to instance level embedding training. Henriques et al. [30] proposed a tracking by detection framework with a kernel trick and histogram of oriented gradients feature to track the object. Ahmed et al. [31] proposed a real-time correlation-based tracking framework by utilizing open loop control strategy, so that the target is always at the center of frame. Moreover, a video stabilization method was incorporated to eliminate the vibration at low computational cost. Ma et al. [32] proposed a long-term correlation filter tracker (LCT) which decomposed the tracking problem into estimation of translation and scale, and redetects the target by online training of a random fern classifier. Masood et al. [33] proposed tracking framework which uses a maximum average correlation height (MACH) filter for detection and proximal gradient algorithm-based particle filter for tracking.

Zhou et al. [34] proposed an STC learning algorithm with multichannel features and an improved adaptive scheme for scale by using a histogram of oriented gradients feature along with color naming and using kernel methods in the STC framework to improve tracking performance. Khan et al. [35] proposed an improved tracking algorithm based on LCT. They incorporated the Kalman filter in the LCT framework for occlusion handling and PSR of the response map for occlusion detection. Ali et al. [36] proposed a tracking algorithm that combines the mean-shift tracker, Kalman filter, and correlation filter heuristically. It updates the template based on the change in the appearance model of the target and computes similarity for each forthcoming frame based on the current frame similarity value.

Mueller et al. [37] proposed a context-aware framework for correlation filter trackers by reformulating the original optimization problem for single and multidimensional features in both primal and dual domains. Qi et al. [38] proposed an improved STC algorithm through incorporation of a context-aware correlation filter in STC framework. Zhang et al. [39] proposed an improved STC algorithm by incorporating color naming and histogram of oriented gradients features in the STC framework, along with improved scale strategy and adaptive model update scheme. Shin et al. [40] proposed an improved KCF-based tracking algorithm. They incorporated module for detection of tracking failure, mechanism for re-tracking in multiple search windows and analysis of motion vectors for deciding the search window in the KCF framework. Based on literature presented it can be concluded that significant modifications have been made in the STC algorithm in terms of model updates, incorporation of occlusion detection and handling mechanisms, utilization of contextual information, fusion of various cues and features such as histogram of oriented gradients feature and color naming, combined with deep learning techniques and incorporation of adaptive learning rate mechanisms.

The STC algorithm proposed by Zhang et al. [21] utilizes fast Fourier transform for detection. Subsequently, context information around the target plays a vital role in object tracking. The basic idea of STC is to use background information around the target area in consecutive frames. The target model is updated based on spatial context information. However, STC cannot deal effectively when the model is updated on inaccurate measurements due to occlusions, background clutter and fast motion. Context-aware formulation can be efficiently applied to deal with background clutter issues. The maximum value of the response map can be used to detect occlusions. Afterwards, the Kalman filter can be applied for occlusion handling. The model update can also be related to the motion of the target; the STC model is updated on a fixed learning rate, making it vulnerable to target motion. On the basis of target motion, the tracking model should be updated adaptively.

1.2. Our Contributions

In this paper, an improved spatio-temporal context-based tracking algorithm is proposed. It combines s context aware formulation, Kalman filter and average difference between consecutive frame-based adaptive learning rate mechanism with STC. Our approach utilizes correlation filter-based context aware formulation making it effective at utilizing the context information while making it computationally less expensive. In addition, the Kalman filter is fused in a tracking framework for occlusion handling. Moreover, an adaptive learning rate mechanism is incorporated to update the model according to change in the environment. Experimental results have been presented on de facto standard videos to show the efficacy of the proposed ideas with various state-of-the-art tracking methods.

1.3. Paper Outline

The rest of the article is organized as follows—a brief explanation of Spatio-temporal context tracking and correlation filtering is given in Section 2. Section 3 defines Context-aware tracking, Kalman filter, occlusion detection mechanism and adaptive model learning rate by an explanation of the proposed method for online tracking. Experiment and

Performance analysis is discussed in Sections 4 and 5. Section 6 provides Experimental results. Section 7 provides discussion and Section 8 concludes the article.

2. The Principle of Spatio-Temporal Context and Correlation Filter Tracking

2.1. STC Based Tracking

In visual object tracking, the target is characterized by objects around the target present in the current frame. The area which is present around the target is called context. In the context around the target, various temporal and spatial relationships exist in continuous frames. STC tracking algorithm is based on a Bayesian framework to accurately find the target location on the basis of background knowledge. It formulates the task of finding the object center by maximizing the confidence map in every frame. Every current frame target location is represented by x^* with its features defined as $X^c = \{y(i) = (I(i), i) | i \in \Omega_c(x^*)\}$ where $I(i)$ is the image grey scale value at location i while $\Omega_c(x^*)$ is the context around target center x^*. It is shown in Figure 1.

Figure 1. Graphical representation of spatial relationship between object and its context.

Confidence map of target location is described in (1).

$$\begin{aligned} y(x) &= P(x|j) = \sum_{y(i) \in X^c} P(x, y(i)|j) \\ &= \sum_{y(i) \in X^c} P(x, y(i)|j) P(y(i)|j) \end{aligned} \tag{1}$$

where j is the target, $P(y(i)|j)$ is context prior model that represents the features of context appearance. $P(x, y(i)|j)$ is spatial context model that formulates spatial relation between object location and its information of context. It is used in identifying and resolves various uncertainties for different image measurements. The goal in this tracking problem is to train the spatial context model $P(x, y(i)|j)$.

2.1.1. Confidence Map

Confidence map function $y(x)$ is presented in (2).

$$y(x) = P(x|j) = re^{-\left|\frac{x-x^*}{\alpha}\right|^{\xi}} \tag{2}$$

where r is normalization constant, α is scale parameter while ξ is shape parameter. The problem of location ambiguity occurs frequently in object tracking. Appropriate selection of the shape parameter can resolve this problem and is helpful in the learning spatial context model. Setting $\xi > 1$ results in over-smoothing of the confidence map near the center, thereby increasing location ambiguities. However, if $\xi < 1$ it generates a sharp peak

response due to which few positions are activated while learning spatial context. Due to these issues STC uses $\xi = 1$.

2.1.2. Context Prior Model

To learn the spatial context model, the context prior model needs to be calculated first. It is modeled by using an image intensity function to represent target appearance along with Gaussian weighted function mentioned in (3) and (4).

$$P(y(i)|j) = I(i)\ \omega_\gamma(i - x^*) \tag{3}$$

$$\omega_\gamma = de^{-|\frac{x-x^*}{\sigma^2}|^2} \tag{4}$$

where d is normalization constant which restricts (4) to range between 0–1 and σ is scale representation. The closer the context location i is to the currently target location x^*, larger weight should be set to predict target location in the next frame.

2.1.3. Learning Spatial Context Model

Spatial context model is defined by conditional probability function is presented in (5).

$$P(x, y(i)|j) = h^{sc}\ (x - i) \tag{5}$$

Solving (5) for spatial context.

$$= h^{sc}(x - i)\ I(i)\ \omega_\gamma(i - x^*)$$

$$= h^{sc}(x) \otimes (I(x)\ \omega_\gamma(x - x^*)) \tag{6}$$

where $\otimes$ is a convolution operator in (6). For improving calculation speed fast Fourier transform (FFT) is used and calculated as presented in (7).

$$F(y(x)) = F(h^{sc}(x)) \odot F(I(x)\omega_\gamma(x - x^*)) \tag{7}$$

where F is FFT operation and $\odot$ denotes element wise multiplication. Solving (7) for spatial context model.

$$h^{sc}(x) = F^{-1}\left(\frac{F\left(re^{-|\frac{x-x^*}{\alpha}|\ \xi}\right)}{F((I(x)\omega_\gamma(x - x^*)))}\right) \tag{8}$$

where F^{-1} denotes inverse FFT in (8). The spatial context model h^{sc} learns relative spatial relations between different pixels in the Bayesian framework.

2.1.4. Model Update

In the STC model, the tracking is considered as a detection task. The target is initialized in position at the first frame. At the tth frame, the STC model $H^{stc}_{t+1}(x)$ can be updated by using the spatial context model $h^{sc}_t(x)$. Then, the target center position x^*_{t+1} of the $(t + 1)$ frame can be attained by computing the extreme of the confidence map given in (9).

$$x^*_{t+1} = arg_{x \in \Omega_c(x^*_t)} max\ y_{t+1}(x) \tag{9}$$

The confidence map $y_{t+1}(x)$ at $t + 1$ frame can be calculated as described in (10).

$$y_{t+1}(x) = F^{-1}\left(F\left(H^{stc}_{t+1}(x)\right) \odot F(I_{t+1}(x)\omega_\gamma(x - x^*_t))\right) \tag{10}$$

Here, H^{stc}_{t+1} derives from spatial context h^{sc}_t and is able to reduce noise caused by abrupt appearance changes of I_{t+1}. The STC model can be updated as mentioned in (11).

$$H^{stc}_{t+1} = (1 - \rho)\ H^{stc}_t + \rho h^{sc}_t \tag{11}$$

where ρ is the learning rate and h_t^{sc} is the spatial context model computed in (8).

2.2. Correlation Filter Tracking

Correlation filters use sampling methods to discriminate the target position from the region of interest in consecutive frames at low computational cost. It models all possible translations of the target in the search window as circular shifts and concatenates them to form a square matrix A_0. It facilitates in computing the Fourier domain solution to the ridge regression problem given in (12).

$$\min_w \|A_0 w - y\|_2^2 + \lambda_1 \|w\|_2^2 \tag{12}$$

In (12), the learned correlation filter is denoted by vector w. Square matrix A_0 contains all circular shifts of image patch and regression target y is vectorized image of 2D Gaussian. Let $x(j)$ be the jth component of vector x and its conjugate is x^*. Then, its Fourier transform $F^H x$ is $\hat{x}$. (12) can be solved by using (13).

$$X = Fdiag(\hat{x})F^H \text{ and } X^T = Fdiag(\hat{x}^*)F^H \tag{13}$$

The convex in (12) is complex and has a unique global minimum. Equating its gradient to zero leads to a closed form solution of the filter as given in (14).

$$w = \left(A_0^T A_0 + \lambda_1 \, I\right)^{-1} A_0^T y \tag{14}$$

As A_0 is circulant, (14) can be diagonalized and its solution in Fourier domain is given in (15).

$$\hat{w} = \frac{\hat{a}_0^* \odot \hat{y}}{\hat{a}_0^* \odot \hat{a}_0 + \lambda_1} \tag{15}$$

The location of the target is the same as the location of maximum response when (15) is convolved with a search window for the next frame. The detection formula is given in (16).

$$r_p(w, \mathrm{Z}) = \mathrm{Z}w \leftrightarrow \hat{r}_p \odot \hat{w} \tag{16}$$

where Z is the search window circulant matrix.

3. Proposed Solution

In this section, the proposed tracker is introduced in detail. First, a context-aware object tracking model is investigated. Second, a Kalman filter-based motion estimation model is discussed. Third, the average difference of a consecutive frames-based model update scheme is presented. Finally, the tracker will be discussed in Algorithm 1. Figure 2 shows the flowchart of the proposed algorithm.

3.1. Context-Aware Tracking Framework

Information of context around the target elevates the tracking performance. Therefore, it is added in the solution of the context-aware correlation filter as given in (17).

$$\min_w \|A_0 w - y\|_2^2 + \lambda_1 \|w\|_2^2 + \lambda_2 \sum_{i=1}^{k} \|A_i w\|_2^2 \tag{17}$$

It should be noted that there are other possible choices for incorporating the context term. However, it leads to constrained convex optimization requiring an iterative solution, which is quite slow. When the position for the current frame is computed by STC, the filter w is trained and the background term A_i is as small as possible. The objective function can be rewritten by forming a new data matrix $B \in \mathbb{R}^{(k+1)n \times n}$ which consists of target and context patches as given in (18).

$$f_p(w, B) = \|Bw - \overline{y}\|_2^2 + \lambda_1 \|w\|_2^2 \tag{18}$$

where $B = \begin{bmatrix} A_0 \\ \sqrt{\lambda_2} A_1 \\ \vdots \\ \sqrt{\lambda_2} A_k \end{bmatrix}$ and $\overline{y} = \begin{bmatrix} y \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

Similar to the correlation filter, the function in (18) is convex and minimized by setting the gradient to zero. It is presented in (19).

$$w = \left(B^T B + \lambda_1 \, I\right)^{-1} B^T \overline{y} \tag{19}$$

Similar to (12), using (13) to determine Fourier domain closed form solution as described in (20).

$$\hat{w} = \frac{\hat{a}_0^* \odot \hat{y}}{\hat{a}_0^* \odot \hat{a}_0 + \lambda_1 + \lambda_2 \sum_{i=1}^{k} \hat{a}_i^* \odot \hat{a}_i} \tag{20}$$

The target window and its position are updated according to (20). Based on target position, the confidence map and STC model in (9) and (11) are updated.

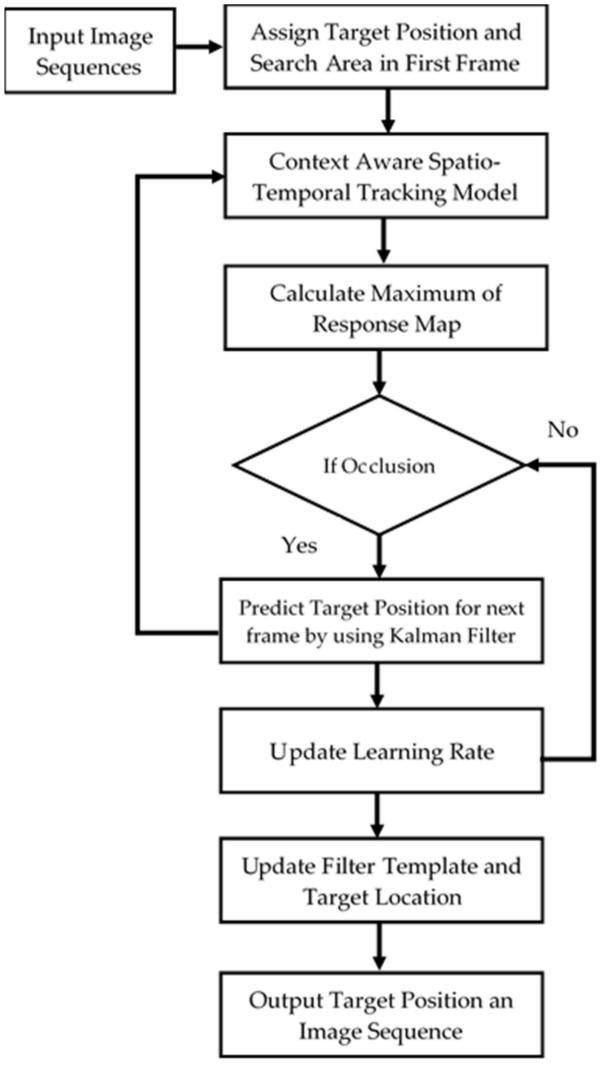

Figure 2. Implementation process of context-aware formulation, spatio-temporal context, adaptive learning rate and Kalman filtering in the tracking algorithm.

3.2. Kalman Filter-Based Motion Estimation Model

The Kalman Filter is an optimal filter which minimizes difference between true states and estimated states. It consists of four processes which are (1) Initial guess of state vector and state error covariance, (2) Forward time step propagation of the state vector and state error covariance, (3) Estimation of the Kalman gain based on state error covariance and measurement noise covariance, (4) Update state vector and state error covariance based on estimated output and Kalman gain [41]. A constant velocity motion model is used due to its simplicity and effectiveness in describing motion of the target. It consists of two stages which are prediction and correction.

3.2.1. Kalman Filter Prediction

During this state, uncertainty about the target is determined by both state and covariance prediction. The current system state can predict position based on the previous state. Similarly, covariance is calculated multiplying the covariance matrix from the previous iteration by state transition matrix and adding process noise Q. The prediction equations are described in (21) and (22).

$$X_t = AX_{t-1} + Bu_{t-1} \tag{21}$$

where X_t is the state target vector, A is the state transition matrix and Bu_{t-1} is noise.

$$S_t = AS_{t-1}A^T + Q \tag{22}$$

where S_t is the predicted error covariance and Q is the covariance of the process noise.

3.2.2. Kalman Filter Correction

The position of the target obtained from STC is used as a measurement value Y_t. By combining it with the predicted result, the Kalman gain can be calculated as described in (23).

$$K_{t-1} = S_{t-1}H^T \left(HS_{t-1}H^T + R\right)^{-1} \tag{23}$$

where R is the measurement noise covariance. The estimate is updated by combining it with the old estimate and the measurement as given in (24).

$$X_{t+1} = X_t + K_{t-1}(Y_t - HX_t) \tag{24}$$

The difference $(Y_t - HX_t)$ is called measurement innovation or residual. It reflects discrepancy between the predicted measurement HX_t and actual measurement Y_t. Error covariance is calculated by using (25).

$$S_{t+1} = (I - K_tH)\, S_t \tag{25}$$

where S_{t+1} is the updated error covariance, H is matrix related to measurement of the state and K_t is the updated Kalman gain.

3.3. Occlusion Detection

When the target undergoes occlusion, then the STC model is updated incorrectly thereby losing the target. In order to detect occlusion, maximum value of target map is used which changes its value with the situation of the target state. If the target is occluded, then the value of response map is small. However, when the target reappears then its value increases. The value of the response map determines whether the target is tracked by STC or by Kalman filter. For a given input image sequence; first the confidence map is calculated in frequency domain, then Spatio-temporal model is learned for tracking. If the target is severely occluded, then for next frame the Kalman filter will predict the position and update the STC using a feedback loop. The filter template for context-aware is updated accordingly. Kalman filter prediction can be updated as observation of target position marked for next frame.

3.4. Adaptive Learning Rate

During object tracking, target motion changes in each frame of the image sequence. Therefore, it is necessary to update the target model correctly. In STC, the learning rate is fixed, making it evitable to different appearances in the environment. So, to make it adaptive, an average difference of two consecutive frames-based mechanisms is incorporated [39]. It is given in (26).

$$\text{er} = \frac{\sum_{i,j}^{M,N} \left| I_{ij}^{n} - I_{ij}^{n-1} \right|}{M * N} \tag{26}$$

where I_{ij} is the pixel value and $M * N$ is the size of image. Learning rate is adjusted as given in (27).

$$\rho = \begin{cases} 0.005, & \text{er} < 1.2 \\ 0.075, & 1.2 \leq \text{er} < 10 \\ 0.1, & \text{er} > 10 \end{cases} \tag{27}$$

Value of learning rate ρ is assigned on the basis of er by using (27).

Algorithm 1: Proposed Tracker at time step t

Input: Image Sequence of n Frames. Position of Target at First Frame.
Output: Target Position for each frame in Image Sequence.
for frame 1 to n frames.
(1). Calculate context prior model using (**3**).
(2). Calculate confidence map using (**10**).
(3). Calculate target center.
(4). Calculate maximum of response map.
(5). **if** response map < threshold
(6). new position = Kalman prediction
(7). **end**
(8). Estimate position for next frame using (**21**).
(9). Estimate error covariance using (**22**).
(10). Calculate Kalman gain using (**23**).
(11). Update estimate via measurement using (**24**).
(12). Update error covariance using (**25**).
(13). Calculate average difference between consecutive frames using (**26**).
(14). Adjust learning rate using (**27**).
(15). Update filter template using (**20**).
(16). Update context prior model on Kalman prediction using (**3**).
(17). Update spatial context model using (**8**).
(18). Update Spatio-temporal context model using (**11**).
(19). Draw rectangle on target in each frame.
End

4. Experiments

To verify the performance of the proposed tracker both qualitatively and quantitatively, it is tested on several image sequences with complex conditions such as occlusion, illumination variation, deformation and clutter background. The proposed method is implemented in MATLAB 2016a. The experimental setup is Intel Core i3 2.30 GHz CPU with 4GB RAM.

4.1. Evaluation Criteria

Two criteria were used to evaluate the algorithm. Those are the center location error (CLE) and distance precision rate (DPR). The CLE is defined as the Euclidean distance calculated by tracking algorithm and ground truth of target. The calculation formula is given in (28).

$$\text{CLE} = \sqrt{\left(x_i - x_{gt}\right)^2 + \left(y_i - y_{gt}\right)^2} \tag{28}$$

where $(x_{i,}, y_i)$ are tracker positions and (x_{gt}, y_{gt}) are ground truth values.

Distance precision rate (DPR) is the percentage of frames at threshold of 20 pixels of estimated location distance and ground truth.

4.2. Dataset

TC-128 [42], OTB2013 [43], and OTB2015 [44] are used for tracking experiments. Eight image sequences are used in this experiment. We evaluate the proposed tracker in comparison with $MOSSE_{CA}$ and DCF_{CA} [37], STC [21] and MACF [28] trackers both qualitatively and quantitatively.

5. Performance Analysis

DPR comparison is given in Table 1. In sequences (Cardark, Cup, Jogging-1, Juice, and Man) proposed, the tracker outperforms $MOSSE_{CA}$, STC, MACF, and DCF_{CA}. In sequences (Carchasing_ce3, and Plate_ce2) all tracking methods have similar performance. Sequence Busstation_ce2 has slightly less precision value. However, the proposed has a higher mean value than other tracking methods.

Table 1. Distance precision rate at threshold of 20 pixels.

Sequence	Proposed	STC	MACF	$MOSSE_{CA}$	DCF_{CA}
Busstation_ce2	0.878	0.194	1	0.820	0.886
Carchasing_ce3	1	1	1	1	1
Cardark	1	1	1	1	1
Cup	1	1	1	0.452	1
Jogging-1	0.996	0.228	0.231	0.231	0.231
Juice	1	1	1	1	1
Man	1	1	1	1	1
Plate_ce2	1	1	1	1	1
Mean Precision	0.984	0.803	0.904	0.813	0.890

The average center location error comparison is given in Table 2. In sequences (Busstation_ce2, Cup, Jogging-1, and Man), the proposed tracker outperforms STC, $MOSSE_{CA}$, MACF, and DCF_{CA}. In sequences (Carchasing_ce3, Cardark, Juice, and Plate_ce2), the proposed tracker has slightly higher error value. However, the proposed has the lowest mean error than other tracking methods.

Table 2. Average center location error.

Sequence	Proposed	STC	MACF	$MOSSE_{CA}$	DCF_{CA}
Busstation_ce2	10.86	78.25	3.58	14.50	9.71
Carchasing_ce3	3.90	3.55	2.39	2.61	3.05
Cardark	4.09	2.83	1.67	3.15	5.11
Cup	4.63	4.84	3.11	95.87	3.85
Jogging-1	8.40	5010	94.93	115.98	89.44
Juice	4.63	5.08	0.91	3.71	1.92
Man	1.32	1.49	1.73	1.72	2.23
Plate_ce2	2.58	2.34	1.62	1.77	1.83
Mean Error	5.05	638.55	13.74	29.91	14.64

The precision plots are shown in Figure 3. These plots provide frame-by-frame precision in entire image sequences. Since precision gives the mean value of an entire image sequence, it is a possibility that the tracker might get drift for a few frames but then again tracks the target correctly. Therefore, these plots were presented to show the efficacy of the tracking method. Various challenges were present in sequences such as occlusion, illumination variations, background clutter, etc. In sequences (Carchasing_ce3, Cardark, Cup, Jogging-1, Juice, Man, and Plate_ce2), the proposed tracker has the highest precision

in the entire sequence. In sequence Busstation_ce2, the proposed tracker has slightly lower precision.

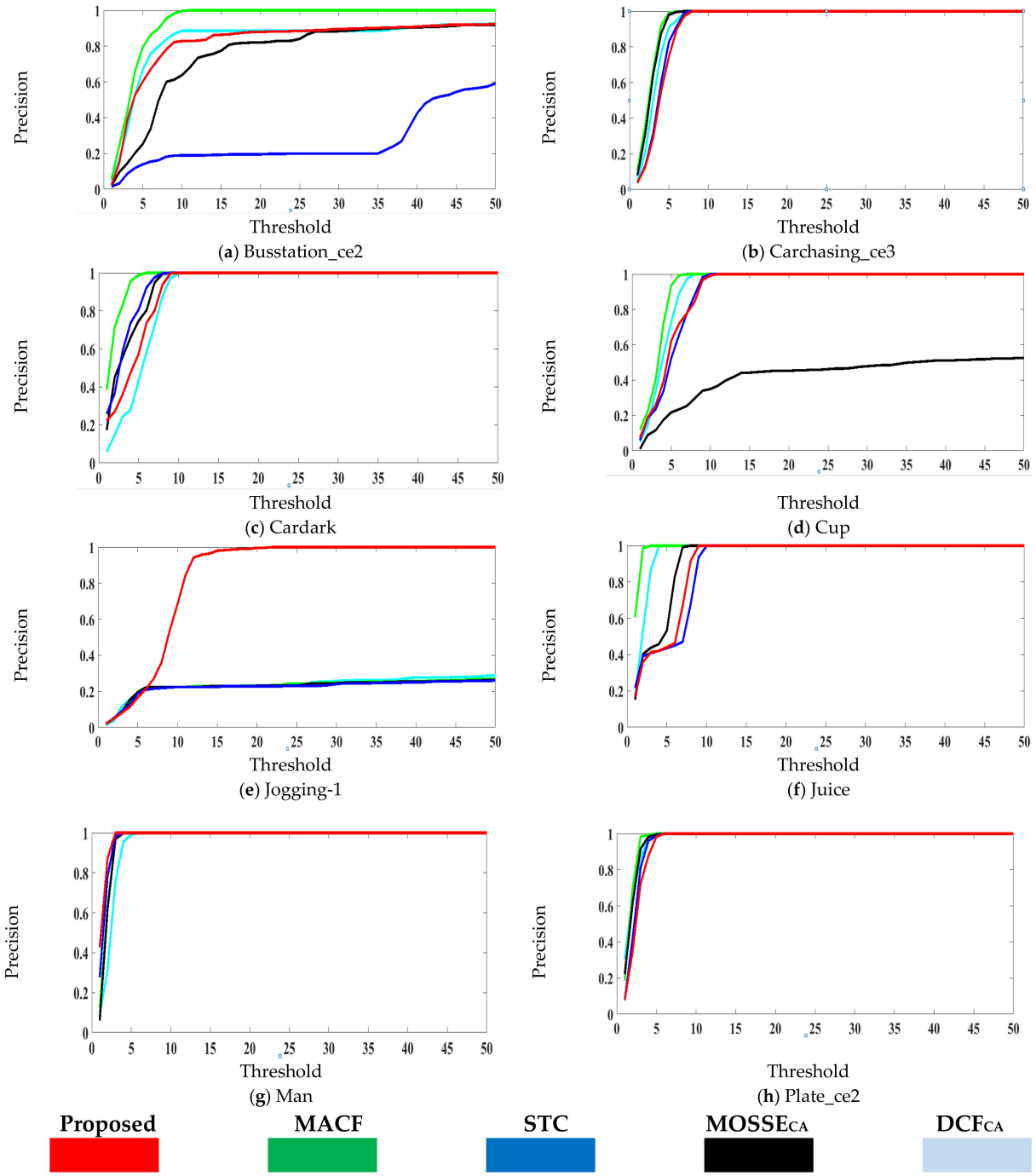

Figure 3. Precision plots comparison of proposed and other state-of-the-art trackers on selected image sequences taken from TC-128, OTB2013 and OTB2015 datasets.

The location error plots are shown in Figure 4. These plots provide frame-by-frame error in entire image sequences. Since the average center location gives mean error of entire image sequence, it is a possibility that tracker might get drift for few frames but then again tracks the target correctly. Therefore, these plots were presented to show the effectiveness of the tracking method. Various challenges were present in sequences such as occlusion, illumination variations, deformation, etc. In sequences (Busstation_ce2,

Cup, Jogging-1, and Man), the proposed tracker has the lowest error in entire sequence. In sequences (Carchasing_ce3, Cardark, Juice, and Plate_ce2), the proposed tracker has slightly higher error.

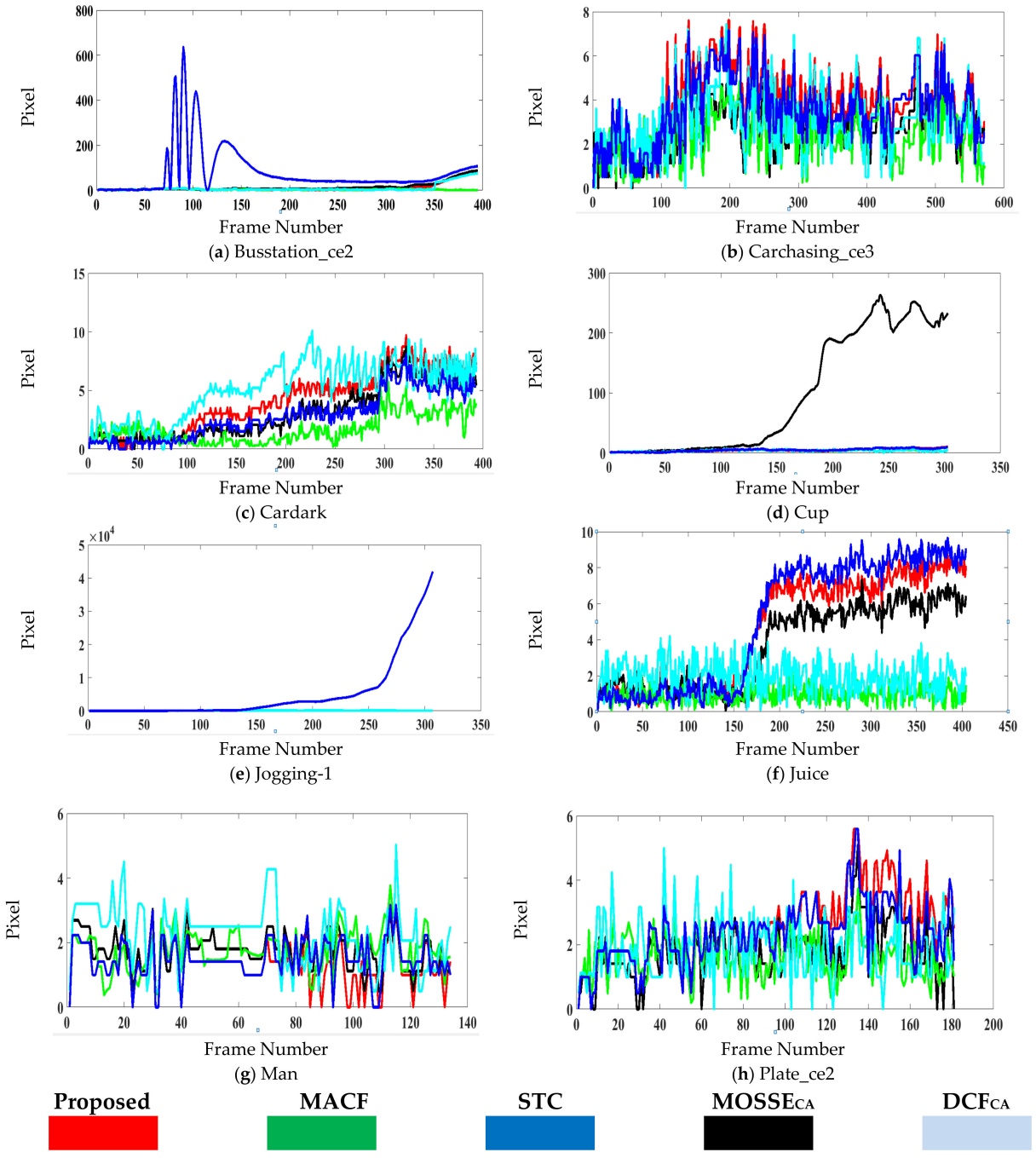

Figure 4. Centre location error (in pixels) comparison of proposed and other state-of-the-art trackers on selected image sequences taken from TC-128, OTB2013 and OTB2015 datasets.

6. Experimental Results

Qualitative results of proposed tracking with four state-of-the-art trackers over eight image sequences is shown in Figure 5. It involves various challenges such as partial or full occlusions, illumination variations, background clutter, etc. DCF_{CA} and $MOSSE_{CA}$ contains similar tracking components as our approach, i.e., correlation filtering and context aware formulation. However, correlation filter in $MOSSE_{CA}$ and DCF_{CA} is not robust to

blur motion in (cup), illumination variations in (man, cardark) and occlusions in (jogging-1, busstation_ce2). In (carchasing_ce3, plate_ce2) where the target undergoes scale variations, both $MOSSE_{CA}$ and DCF_{CA} have similar performance with the proposed in tracking of the target. With the joint use of an instantaneous motion model and Kalman filter in discriminative scale space tracking frame, MACF performs better on various challenging sequences. However, MACF tends to drift when the target undergoes occlusion and fails to recover from tracking failures (jogging-1). Although STC can estimate scale, it does not perform well in motion blur (juice) and scale variations (cup). This is because STC only uses intensity features and estimates scale from a response map of single translation filter. Moreover, it does not deal effectively with occlusion (jogging-1, busstation_ce2) as there is no occlusion handling mechanism present to deal with this issue. Moreover, its target model is updated on a fixed learning rate, making it vulnerable to the background environment.

Figure 5. Qualitative comparison of proposed with four state-of-the-art trackers on selected image sequences taken from TC-128, OTB2013 and OTB2015 datasets.

The proposed tracker performs well in all these challenging sequences. This performance can be attributed to three reasons. First, context-aware formulation in STC framework is incorporated, making it less sensitive to illumination variation (cardark, man) and motion blur (juice, man, cup). Second, incorporation of occlusion detection based on the response map and occlusion handling using Kalman filter makes it effective towards partial or full occlusion (jogging-1, busstation_ce2). Third, fusion of adaptive learning rate in the model update of the tracking model is effective in dealing with scale variation and fast motion (plate_ce2).

7. Discussion

We discuss several observations from experimental and quantitative analysis. First, context aware formulation in the correlation filter outperforms trackers without this formulation. This can be attributed to the fact that correlation filters regress all circular shifts of the target appearance model. Second, trackers with occlusion detection and handling modules outperforms trackers without these modules. This can be attributed to the fact that occlusion detection and handling mechanism does not lead the tracker to drift. Third, trackers with an adaptive learning rate mechanism perform better than a fix learning rate. It is because the tracking model copes with the changes in environment.

8. Conclusions

In the present article, an adaptive Spatio-temporal context (STC)-based algorithm for online tracking is presented, which combines the context-aware formulation, Kalman filter, response map-based occlusion detection, and average difference based adaptive model update in the STC framework. The algorithm performs better in scenarios such as full occlusion, illumination variation, deformation, and background clutter in comparison to various algorithms with the achievement of efficient performance in datasets. Even though the tracker has achieved the desired performance, the target may be lost in some cases like motion blur, fast motion, and scale variation. The problem can be resolved through the establishment of neural network-based algorithms [7–9] to improve robustness and tracking accuracy.

Author Contributions: K.M. and A.A. conceived the main idea. K.M. designed the framework under the supervision of A.A. and A.J. K.M. and B.K. validate the results. M.M., W.U.K. and Y.H. provided suggestions in manuscript writing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 51977153, 51977161, 51577046, State Key Program of National Natural Science Foundation of China under Grant No. 51637004, National Key Research and Development Plan "important scientific instruments and equipment development" Grant No. 2016YFF010220, Equipment research project in advance Grant No. 41402040301.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare that they have no conflict of interest to report regarding the present study.

References

1. Cao, S.; Wang, X. Real-time dynamic gesture recognition and hand servo tracking using PTZ camera. *Multimed. Tools Appl.* **2019**, *78*, 27403–27424. [CrossRef]
2. Santhosh, P.K.; Kaarthick, B. An Automated Player Detection and Tracking in Basketball Game. *Comput. Mater. Contin.* **2019**, *58*, 625–639.
3. Oh, S.H.; Javed, S.; Jung, S.K. Foreground Object Detection and Tracking for Visual Surveillance System: A Hybrid Approach. In Proceedings of the 11th International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 16–18 December 2013; pp. 13–18.
4. Zhou, W.; Wu, C.; Yu, X.; Gao, Y.; Du, W. Automatic fovea center localization in retinal images using saliency-guided object discovery and feature extraction. *J. Med. Imaging Health Inform.* **2017**, *7*, 1070–1077. [CrossRef]
5. Kuramoto, A.; Aldibaja, M.A.; Yanase, R.; Kameyama, J.; Yoneda, K.; Suganuma, N. Mono-Camera based 3D Object Tracking Strategy for Autonomous Vehicles. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 459–464.
6. Muresan, M.P.; Giosan, I.; Nedevschi, S. Stabilization and Validation of 3D Object Position Using Multimodal Sensor Fusion and Semantic Segmentation. *Sensors* **2020**, *20*, 1110. [CrossRef] [PubMed]
7. Kazimierski, W. Proposal of neural approach to maritime radar and automatic identification system tracks association. *IET Radar Sonar Navig.* **2017**, *1*, 729–735. [CrossRef]
8. Stateczny, A. Neural manoeuvre detection of the tracked target in ARPA systems. *IFAC Proc. Vol.* **2002**, *34*, 209–214. [CrossRef]
9. Kazimierski, W.; Zaniewicz, G.; Stateczny, A. Verification of multiple model neural tracking filter with ship's radar. In Proceedings of the 13th International Radar Symposium (IRS), Warsaw, Poland, 23–25 May 2012; pp. 549–553.
10. Ali, A.; Jalil, A.; Niu, J.; Zhao, X.; Rathore, S.; Ahmed, J.; Iftikhar, M.A. Visual object tracking—Classical and contemporary approaches. *Front. Comput. Sci.* **2016**, *10*, 167–188. [CrossRef]
11. Fiaz, M.; Mahmood, A.; Javed, S.; Jung, S.K. Handcrafted and deep trackers: Recent visual object tracking approaches and trends. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–44. [CrossRef]
12. Fiaz, M.; Javed, S.; Mahmood, A.; Jung, S.K.M. Comparative Study of ECO and CFNet Trackers in Noisy Environment. *arXiv* **2018**, arXiv:1801.09360.
13. Biresaw, T.A.; Cavallaro, A.; Regazzoni, C.S. Tracker-Level Fusion for Robust Bayesian Visual Tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 776–789. [CrossRef]

14. Sun, X.; Yao, H.; Zhang, S.; Li, D. Non-Rigid Object Contour Tracking via a Novel Supervised Level Set Model. *IEEE Trans. Image Process.* **2015**, *24*, 3386–3399.
15. Jang, S.I.; Choi, K.; Toh, K.A.; Teoh, A.B.J.; Kim, J. Object tracking based on an online learning network with total error rate minimization. *Pattern Recognit.* **2015**, *48*, 126–139. [CrossRef]
16. Zhang, Z.; Peng, H. Deeper and wider siamese networks for real time visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 4591–4600.
17. Rahman, M.M.; Ahmed, M.R.; Laishram, L.; Kim, S.H.; Jung, S.K. Siamese High-Level Feature Refine Network for Visual Object Tracking. *Electronics* **2020**, *9*, 1918. [CrossRef]
18. Zhang, J.; Jin, X.; Sun, J.; Wang, J.; Li, K. Dual model learning combined with multiple feature selection for accurate visual tracking. *IEEE Access* **2019**, *7*, 43956–43969. [CrossRef]
19. Dai, K.; Wang, D.; Lu, H.; Sun, C.; Li, J. Visual tracking via adaptive spatially regularized correlation filters. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 4670–4679.
20. Javed, S.; Zhang, X.; Seneviratne, L.; Dias, J.; Werghi, N. Deep Bidirectional Correlation Filters for Visual Object Tracking. In Proceedings of the IEEE 23rd International Conference on Information Fusion (FUSION), Rustenburg, South Africa, 6–9 July 2020; pp. 1–8.
21. 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 (ECCV), Zurich, Switzerland, 6–7 September 2014; pp. 127–141.
22. Tian, J.; Zhou, Y. Real-time patch-based tracking with occlusion handling. In Proceedings of the International Conference on Neural Information Processing, Kuching, Malaysia, 3–6 November 2014; pp. 210–217.
23. Panqiao, C.; Mengzhao, Y. STC Tracking Algorithm Based on Kalman Filter. In Proceedings of the 4th International Conference on Machinery, Materials and Computing Technology, Hangzhou, China, 23–24 January 2016; pp. 1916–1920.
24. Munir, F.; Minhas, F.; Jalil, A.; Jeon, M. Real time eye tracking using Kalman extended spatio-temporal context learning. In Proceedings of the Second International Workshop on Pattern Recognition, Singapore, 1–3 May 2017; p. 104431.
25. Cui, Z.; Yang, J.; Jiang, S.; Li, J.; Gu, Y. Robust spatio-temporal context for infrared target tracking. *Infrared Phys. Technol.* **2018**, *91*, 263–277. [CrossRef]
26. Yang, X.; Zhu, S.; Zhou, D.; Zhang, Y. An improved target tracking algorithm based on spatio-temporal context under occlusions. *Multidim. Syst. Sign Process.* **2020**, *31*, 329–344. [CrossRef]
27. Yang, H.; Wang, J.; Miao, Y.; Yang, Y.; Zhao, Z.; Wang, Z.; Sun, Q.; Wu, D.O. Combining Spatio-Temporal Context and Kalman Filtering for Visual Tracking. *Mathematics* **2019**, *7*, 1059. [CrossRef]
28. Zhang, Y.; Yang, Y.; Zhou, W.; Shi, L.; Li, D. Motion-Aware Correlation Filters for Online Visual Tracking. *Sensors* **2018**, *18*, 3937. [CrossRef]
29. Lu, Z.; Rathod, V.; Votel, R.; Huang, J. RetinaTrack: Online Single Stage Joint Detection and Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 14656–14666.
30. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [CrossRef]
31. Ahmed, J.; Ali, A.; Khan, A. Stabilized Active Camera Tracking System. *J. Real-Time Image Process.* **2016**, *11*, 315–324. [CrossRef]
32. Ma, C.; Yang, X.; Zhang, C.; Yang, M.H. Long-term correlation tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5388–5396.
33. Masood, H.; Rehman, S.; Khan, A.; Riaz, F.; Hassan, A.; Abbas, M. Approximate Proximal Gradient-Based Correlation Filter for Target Tracking in Videos: A Unified Approach. *Arab. J. Sci. Eng.* **2019**, *44*, 9363–9380. [CrossRef]
34. Zhou, X.; Liu, X.; Yang, C.; Jiang, A.; Yan, B. Multi-channel features spatio-temporal context learning for visual tracking. *IEEE Access* **2017**, *5*, 12856–12864. [CrossRef]
35. Khan, B.; Ali, A.; Jalil, A.; Mehmood, K.; Murad, M.; Awan, H. AFAM-PEC: Adaptive Failure Avoidance Tracking Mechanism Using Prediction-Estimation Collaboration. *IEEE Access* **2020**, *8*, 149077–149092. [CrossRef]
36. Ali, A.; Jalil, A.; Ahmed, J.; Iftikhar, M.A.; Hussain, M. Correlation, Kalman filter and adaptive fast mean shift based heuristic approach for robust visual tracking. *Signal Image Video Process.* **2015**, *9*, 1567–1585. [CrossRef]
37. Mueller, M.; Smith, N.; Ghanem, B. Context-Aware Correlation Filter Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1387–1395.
38. Qi, F.; Hao, Z.; Lu, Z. Spatio-Temporal Context Tracking Algorithm Based on Correlation Filtering. *J. Phys. Conf. Ser.* **2019**, *1213*, 1–7.
39. Zhang, Y.; Wang, L.; Qin, J. Adaptive spatio-temporal context learning for visual tracking. *Imaging Sci. J.* **2019**, *67*, 136–147. [CrossRef]
40. Shin, J.; Kim, H.; Kim, D.; Paik, J. Fast and Robust Object Tracking Using Tracking Failure Detection in Kernelized Correlation Filter. *Appl. Sci.* **2020**, *10*, 713. [CrossRef]
41. Zekavat, R.; Buehrer, R.M. An Introduction to Kalman Filtering Implementation for Localization and Tracking Applications. In *Handbook of Position Location: Theory, Practice, and Advances*, 2nd ed.; Wiley Online Library: Hoboken, NJ, USA, 2018; pp. 143–195.
42. Liang, P.; Blasch, E.; Ling, H. Encoding color information for visual tracking: Algorithms and benchmark. *IEEE Trans. Image Process.* **2015**, *24*, 5630–5644. [CrossRef]

43. Wu, Y.; Lim, J.; Yang, M.H. Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2411–2418.
44. Wu, Y.; Lim, J.; Yang, M.H. Object tracking benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1834–1848. [CrossRef]