An Improved Kernelized Correlation Filter Algorithm for Underwater Target Tracking

Xingmei Wang [ID]**, Guoqiang Wang, Zhonghua Zhao, Yue Zhang * and Binghua Duan**

College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China; wangxingmei@hrbeu.edu.cn (X.W.); wgq0121@hrbeu.edu.cn (G.W.); zhaozhonghua@hrbeu.edu.cn (Z.Z.); 18S003089@stu.hit.edu.cn (B.D.)

* Correspondence: zhangyue_heu@163.com; Tel.: +86-180-4562-7166

Abstract: To obtain accurate underwater target tracking results, an improved kernelized correlation filter (IKCF) algorithm is proposed to track the target in forward-looking sonar image sequences. Specifically, a base sample with a dynamically continuous scale is first applied to solve the poor performance of fixed-scale filters. Then, in order to prevent the filter from drifting when the target disappears and appears again, an adaptive filter update strategy with the peak to sidelobe ratio (PSR) of the response diagram is developed to solve the following target tracking errors. Finally, the experimental results show that the proposed IKCF can obtain accurate tracking results for the underwater targets. Compared to other algorithms, the proposed IKCF has obvious superiority and effectiveness.

Keywords: kernelized correlation filter algorithm; forward looking sonar; underwater target; target tracking

---

1. Introduction

With the rapid development of the world economy, the strategic position of the ocean has become more and more important [1]. Forward-looking sonar is a device that detects oceans by sound waves. It is mainly used for underwater target location, underwater target tracking, and obstacle avoidance [2]. In particular, underwater target tracking in forward-looking sonar image sequences has gradually attracted more global attention. Therefore, numerous approaches have recently been proposed for providing exceptional underwater target tracking performance.

At present, the underwater target tracking methods are mainly the extension of optical image target tracking methods. Among them, mean-shift [3], Kalman filter (KF) [4], and particle filter (PF) [5] are the most commonly used. Lane D. M. et al. [6] introduced the optical flow motion estimation to track the target in sonar image sequences. The method is shown to work well, with a good tracking performance when objects merge, split, and change shape. In order to solve the problem of large computations of the optical flow motion estimation tracking method, Chu H. et al. [7] improved the target's alternating problem and occlusion problem using the combination of mean-shift and improved KF, through which the accuracy of target tracking is guaranteed. For underwater target tracking, Quidu et al. [8] introduced the KF for multitarget tracking in forward-looking sonar image sequences; this algorithm was designed for still target lying on a flat seafloor. On this basis, a framework using navigation data for performing robust multitarget tracking based on the KF is proposed to track an obstacle in a series of sonar images [9], but this work cannot be used for the tracking of moving obstacles. In addition, Zhang T et al. [10] introduced the PF to track the target in forward-looking sonar image sequences. The experimental results show that the PF has robustness in target tracking. Later, Li et al. [11] adopted an improved Otsu method to detect the underwater target, and then carried out

underwater target tracking through the PF algorithm combined with a multifeature adaptive fusion strategy. However, the PF suffers from a high computation cost, because it requires a larger number of particles to represent the posterior density of the object state. Although there are many improved algorithms, most of them are still based on the above algorithms. There are some limitations in the improvement of these algorithms. Correlation filter tracking as a discriminative algorithm [12] is beginning to be noticed, due to its robustness and flexibility.

Correlation filter tracking as a discriminative algorithm was first used for target tracking by Bolme D. S. in 2010 [13]. The proposed minimum output sum of squared error (MOSSE) filter first uses the target region as a positive sample in the current frame and the background region as a negative sample. The classifier is trained by a machine learning method, then the next frame with a trained classifier to find the best region as the target area. Among the discriminative tracking algorithms, the correlation filter tracking algorithm is one of the outstanding representatives. Based on the correlation filter tracking algorithm, a series of improved algorithms have been proposed. Rui Caseiro et al. [14] used a circulant structure of tracking-by-detection with kernels (CSK) to detect the correlation of adjacent frames and transform to the frequency domain to accelerate the operation, which can solve the problem of increasing the calculational amount caused by too many training samples of the filter. In addition, Danelljan Martin [15] introduced a correlation filter based on a color feature template by studying the contribution of color in the detection and tracking framework. The results showed that the color features provide excellent performance in tracking. In 2015, Henriques J. F. et al. [16] proposed a kernelized correction filters (KCF) algorithm that generates training samples by cyclic shift; the Fourier transform can be used to perform diagonalization, which reduces the memory and computational complexity by several orders of magnitude, and the kernel trick is used to train the nonlinear filter, which greatly improves the performance of the filter on target detection. However, when the target size changes in the image sequences, it will lead to filter drift in the tracking process, resulting in tracking error. In order to solve this problem, Li Y. et al. [17] proposed a novel scale adaptive kernelized correlation filter tracker with multiple feature integration; the proposed algorithm solved the scale change issue in object tracking through a multiple scales searching strategy. Later, Xu Y. et al. [18] proposed a scale calculation approach for visual object tracking, and the results demonstrated that this algorithm outperforms state-of-the-art methods while operating in real-time. Reference [19] adopted a deformable part-based correlation filter tracking approach to cope with challenging cases, like partial occlusion, deformation, or scale changes.

In these regards, to obtain more accurate tracking results, this paper presents an improved kernelized correlation filter (IKCF) to track the underwater target in forward-looking sonar image sequences. A base sample with a dynamic continuous change scale is used to predict the best location of the target. Then, an adaptive filter update strategy is proposed to update the filter using the peak to sidelobe ratio (PSR) of the response diagram. Compared to the results of KCF, CSK, MOSSE, and mean-shift, the proposed IKCF can obtain better tracking results. Moreover, it has obvious superiority and effectiveness for a target with varying scales and for when the target disappears and appears again. Therefore, the proposed IKCF has important theoretical and practical value.

2. Kernelized Correlation Filter Algorithm

The KCF is based on the correlation filter framework; it converts the solution of the filter into the problem of training binary classifier. Specifically, a sample centered on the target is selected as a positive sample and a negative sample around the target is selected to train the classifier. Then, the next frame image is detected with the classifier, and the result of the detection is used as the location of target in the next frame.

In the KCF, assuming that a number of training samples and their tag values are given as $\{(\chi_1, y_1), (\chi_2, y_2), \ldots, (\chi_i, y_i), \ldots, (\chi_n, y_n)\}$, the purpose of training is to find a function

$f(z) = w^T z$, which makes the label $f(\chi_i)$ of sample χ_i predicted by the classifier model have the minimum mean square error with its real tag y_i. It can be expressed as:

$$\min_i \sum (f(\chi_i) - y_i)^2 + \lambda \parallel w \parallel^2 \tag{1}$$

where χ_i represents the ith sample, w is the coefficient vector of the linear classifier, y_i is the label of the ith sample, and λ is the regularization parameter, which is used to prevent overfitting. A closed-form solution can be obtained by finding the partial derivative of w and equating the derivative result to zero. Therefore, w is defined as follows:

$$w = \left(X^T X + \lambda I\right)^{-1} X^T y \tag{2}$$

where the matrix X is composed of samples, $X = [\chi_1, \chi_2, \ldots \chi_i, \ldots \chi_n]$, y is the column vector formed by the label of the sample, $y = [y_1, y_2, \ldots y_i, \ldots y_n]$, and I is the identity matrix. Equation (2) to the complex domain can be expressed as:

$$w = \left(X^H X + \lambda I\right)^{-1} X^H y \tag{3}$$

where X^H is a Hermitian transpose. Then, the coefficients w of the linear regression classifier $f(x)$ are solved and the training of the linear classifier is finished.

In order to reduce the computational complexity, the KCF uses cyclic shift to construct the training sample X. The cyclic shift operation is carried out on the base sample. The base sample is a rectangular image block whose center is the target and the size is a fixed multiple of the target size. Assuming that the $n \times 1$ dimensional vector $\chi = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i, \ldots, \varepsilon_n]^T$ is used to represent the base sample, the cyclic shift operation of the base sample in one dimension is expressed as follows:

$$\{P^u \chi | u = 0, \ldots, n - 1\} \tag{4}$$

where u represents the direction and time of cyclic shifts, P is a permutation matrix, and the specific form of P is expressed as:

$$P = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5}$$

A training sample X generated by shifting one element of the base sample $\chi = [\varepsilon_1, \varepsilon_2, \ldots \varepsilon_i, \ldots \varepsilon_n]^T$ is shown as:

$$X = C(\chi) = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \\ \varepsilon_n & \varepsilon_1 & \cdots & \varepsilon_{n-1} \\ \varepsilon_{n-1} & \varepsilon_n & \cdots & \varepsilon_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_2 & \varepsilon_3 & \cdots & \varepsilon_1 \end{bmatrix} \tag{6}$$

where $C(\chi)$ represents a circular matrix generated by χ.

According to the properties of the circulant matrix, X can be diagonalized by the discrete Fourier transform. It is:

$$X = F diag(\hat{\chi}) F^H \tag{7}$$

where F is a discrete Fourier transform matrix, which is a constant value. $\hat{\chi}$ represents the discrete Fourier transform of generating vector χ of the cyclic matrix, and $diag\,(\hat{\chi})$ represents the diagonal matrix generated by diagonal elements $\hat{\chi}$.

To simplify the solving process w, Equation (7) is substituted by Equation (3) to solve the linear classifier coefficient w. w can be obtained as follows:

$$\hat{w} = \frac{\hat{\chi}^* \odot \hat{y}}{\hat{\chi}^* \odot \hat{\chi} + \lambda} \tag{8}$$

When the data are characterized by nonlinear distribution, the linear regression classifier cannot show a good performance, and a more robust nonlinear regression function is needed to be used as a classifier. The KCF extends the problem to nonlinear space, through a kernel trick, and the solution of the nonlinear problem in low dimensional space is mapped to the high dimensional kernel space. By solving the linear problem in high dimensional space, the solution of the nonlinear problem in the original low dimensional space is obtained. After mapping with the kernel function, the linear regression coefficient of linear problems in high dimensional kernel space is:

$$w = \sum_i \alpha_i \varphi\,(\chi_i) \tag{9}$$

where $\varphi\,(\chi)$ means that the base sample χ is mapped to the dual space relative to the original space of w, and α represents the coefficient of the dual space. Therefore, the linear regression function in the high dimensional kernel space can be expressed as:

$$f\,(z) = w^T \varphi\,(z) = \sum_{i=1}^{n} \alpha_i \varphi\,(\chi_i) \varphi\,(z) = \sum_{i=1}^{n} \alpha_i \kappa\,(z_i, \chi_i) \tag{10}$$

where κ refers tothe kernel function. The KCF uses the Gauss kernel by default, and the calculation process of the kernel function is:

$$\kappa^{\chi\chi'} = \kappa\,(\chi, \chi') = \exp\left(-\frac{1}{\sigma^2}\left(\parallel \chi \parallel^2 + \parallel \chi' \parallel^2 - 2F^{-1}\,(\hat{\chi}^* \odot \hat{\chi}')\right)\right) \tag{11}$$

where σ is the standard deviation, * means complex conjugation, and F^{-1} means inverse Fourier transformation.

Therefore, the solution of regression coefficient is transformed into the solution of α. The solution of ridge regression based on kernel function is:

$$\alpha = (K + \lambda I)^{-1} y \tag{12}$$

where K is a $n \times n$ kernel matrix. It can be expressed as:

$$K_{ij} = \kappa\,(\chi_i, \chi_j) \tag{13}$$

Since the Gaussian kernel function satisfies Theorem 1, the corresponding kernel matrix K can be proven to be a cyclic matrix, and then the solution of α can be simplified as follows:

$$\hat{\alpha} = \frac{\hat{y}}{\hat{\kappa}^{\chi\chi'} + \lambda} \tag{14}$$

whereˆexpresses the Discrete Fourier Transform. $\kappa^{\chi\chi}$ is the first row of kernel matrix $K = C\,(\kappa^{\chi\chi})$, which is sample χ and the value of itself calculated by the kernel function.

Theorem 1. *Given circulant data $C\,(\chi)$, the corresponding kernel matrix K is circulant if the kernel function satisfies $\kappa\,(\chi, \chi') = \kappa\,(M\chi, M\chi')$ for any permutation matrix M.*

After training α using the above method, the sample can be tested. Since all the samples to be detected Z are generated by cyclic shift of the base sample z and the training sample X is generated by cyclic shift of the base sample χ, it is easy to verify that this kernel matrix satisfies Theorem 1. The kernel matrix is:

$$K^z = C\left(\kappa^{\chi z}\right) \tag{15}$$

According to Equation (10), the response output value of all input samples can be calculated:

$$f(z) = \sum_{i=1}^{n} \alpha_i \kappa\left(z_i, \chi_i\right) = \left(K^z\right)^T \alpha \tag{16}$$

For more efficient calculations, K^z can be diagonalized by the Discrete Fourier Transform:

$$\hat{f}(z) = \hat{\kappa}^{\chi z} \odot \hat{\alpha} \tag{17}$$

Target tracking with the KCF is considered an efficient algorithm, since it obtains the fastest and highest performance among the recent top-performing algorithms. However, the KFC cannot effectively track targets with varying scales or targets that disappear and appear again. To overcome this drawback, an IKCF is proposed to track the underwater target in forward-looking sonar image sequences in this paper.

3. Improved Kernelized Correlation Filter Algorithm

In the proposed IKCF, a base sample with a dynamically continuous scale is firstly presented to solve the problem, which is the poor performance of fixed-scale filters in detecting targets with varying scales. Then, to solve the problem of large computation and filter drift caused by the filter's frequent updating, an adaptive filter update strategy based on PSR is adopted, so as to obtain a better tracking result.

3.1. Base Sample With Dynamically Continuous Scale

The traditional KCF uses a fixed-size base sample for training; when the target size changes in the tracking process, it will cause the filter training effect not to be good, resulting in tracking failure. In order to solve the problem of a poor training effect with a fixed-size base sample in the KCF, a base sample with a dynamically continuous scale is proposed.

Figure 1 is a few frames extracted from image sequences; the target marked by the red rectangle in the image (target size is 27×29) has been getting larger. Figure 2 is a response diagram corresponding to Figure 1, which is tracked by the traditional KCF at a fixed-size base sample of 5.5 times the target size.

From Figure 2, it can be seen that when the target is getting larger, the corresponding output response diagram peak increases first and then decreases. Figures 1b and 2b show that the underwater target scale is advantageous for the detection of the filter when it becomes larger. Figures 1c and 2c show that if the target scale exceeds a certain value, the detection effect of the filter will begin to decrease and the tracking accuracy will decrease.

In order to avoid the influence of the target scale change on tracking accuracy, the base sample with a dynamically continuous scale is proposed to solve this problem. Firstly, the features of the histogram of oriented gradients are extracted for base samples of a different scale. Then, using the target feature template $v = \{v_1, v_2, ..., v_n\}$ and the base sample feature $\tau = \{\tau_1, \tau_2, ..., \tau_n\}$ in the filter model to perform the kernelized correlation operations, the result of kernelized correlation operations is:

$$\kappa^{\tau v} = \kappa\left(\tau, v\right) = \exp\left(-\frac{1}{\sigma^2}\left(\parallel \tau \parallel^2 + \parallel v \parallel^2 - 2F^{-1}\left(\hat{\tau}^* \odot \hat{v}\right)\right)\right) \tag{18}$$

where F^{-1} means inverse Fourier transformation, * means complex conjugation, and $\hat{\tau}$ is the Fourier transform of the variable τ.

(a)　　　　　　　　　　　　　　　　　(b)

(c)
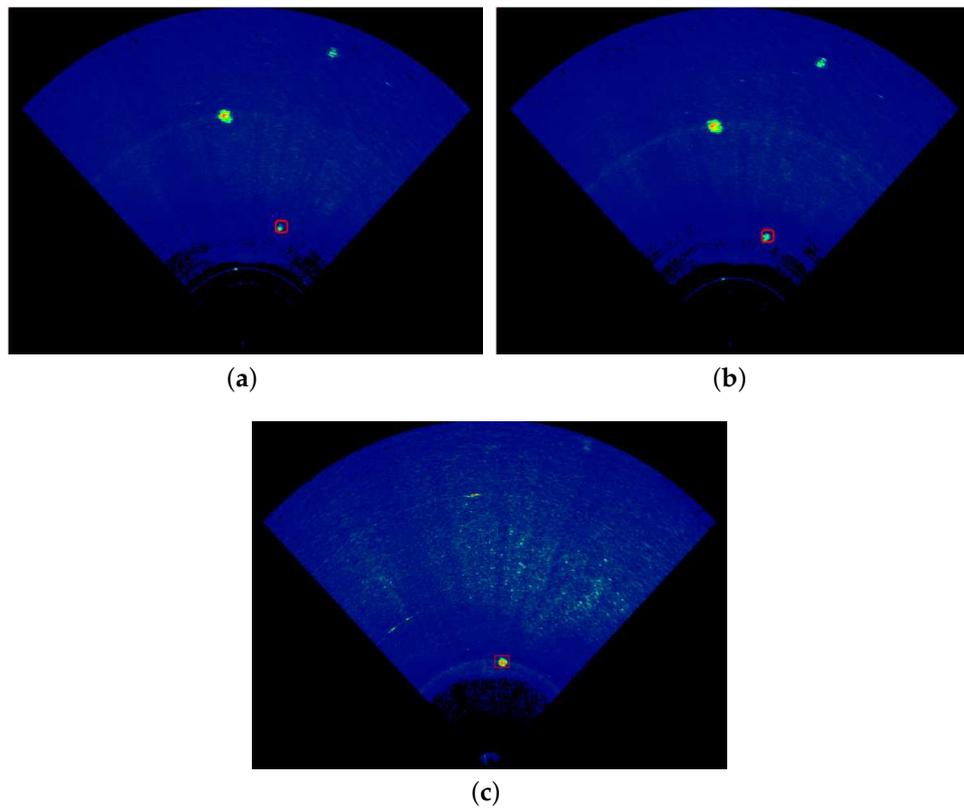
Figure 1. Tracking results at different times: (**a**) The tracking results at the 6th frame; (**b**) the tracking results at the 12th frame; (**c**) the tracking results at the 31st frame.

(a)　　　　　　　　　　　　　　　　　(b)

(c)
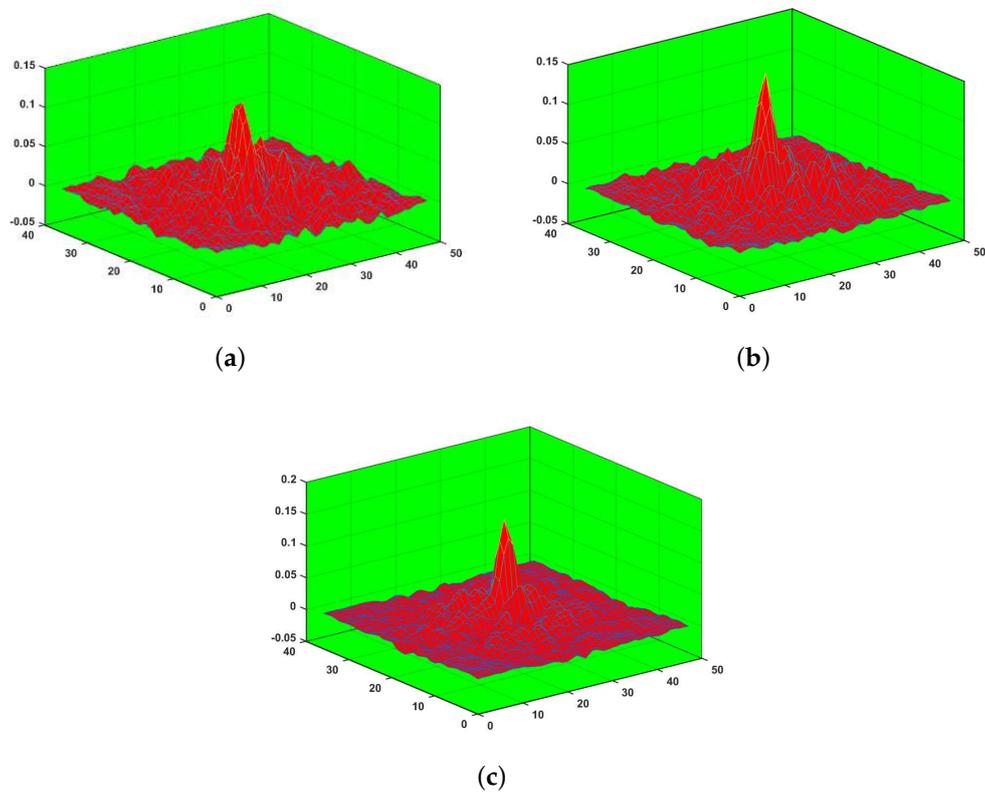
Figure 2. Response diagram of tracking results: (**a**) Response diagram of the 6th frame image; (**b**) response diagram of the 12th frame image; (**c**) response diagram of the 31st frame image.

The results of the kernelized correlation operations $\kappa^{\tau\upsilon} = \{\kappa_1{}^{\tau\upsilon}, \kappa_2{}^{\tau\upsilon}, ..., \kappa_n{}^{\tau\upsilon}\}$ and filter coefficients $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ are converted to the frequency domain, and a plurality of response diagram $f = \{f_1, f_2, \ldots, f_n\}$ in the spatial domain is obtained by Fourier inverse transformation, and the response diagram can be calculated by:

$$f_i(\tau) = F^{-1}\left(\hat{f}_i(\tau)\right) = F^{-1}\left(\hat{\kappa}^{\tau\upsilon} \odot \hat{\alpha}\right) \tag{19}$$

where $i = 1, 2, ..., n$, $\odot$ represents the point multiplication operation between elements. The optimal position of the target can be predicted based on the response diagram with maximum peak.

3.2. Adaptive Filter Update Strategy Based on PSR

The traditional KCF adopts a frame-by-frame linear updating strategy in the update of the filter model, which enhances the adaptability of the filter model to environmental changes. However, this strategy also leads to many problems. First, the model is updated frame by frame, making the update too frequent, which not only leads to a large amount of calculation, but also causes a decrease in robustness due to overfitting. Second, when the target disappears and appears again, the updating method of linear interpolation will introduce erroneous information to the filter model. In order to solve this problem, an adaptive filter update strategy based on PSR is proposed.

In the process of the filter model updating, some method should be used to make the detection tracking process aware of whether the tracking result has a large deviation, and to timely update the model according to the result. The PSR is used as the basis for judging whether the tracking result is wrong. The PSR calculation process is:

$$PSR_t = \frac{p - \mu}{\sigma} \tag{20}$$

where p is the response diagram peak value of the tth frame. The area around the peak is the sidelobe area. μ and σ are the mean and standard deviation of the sidelobe regions.

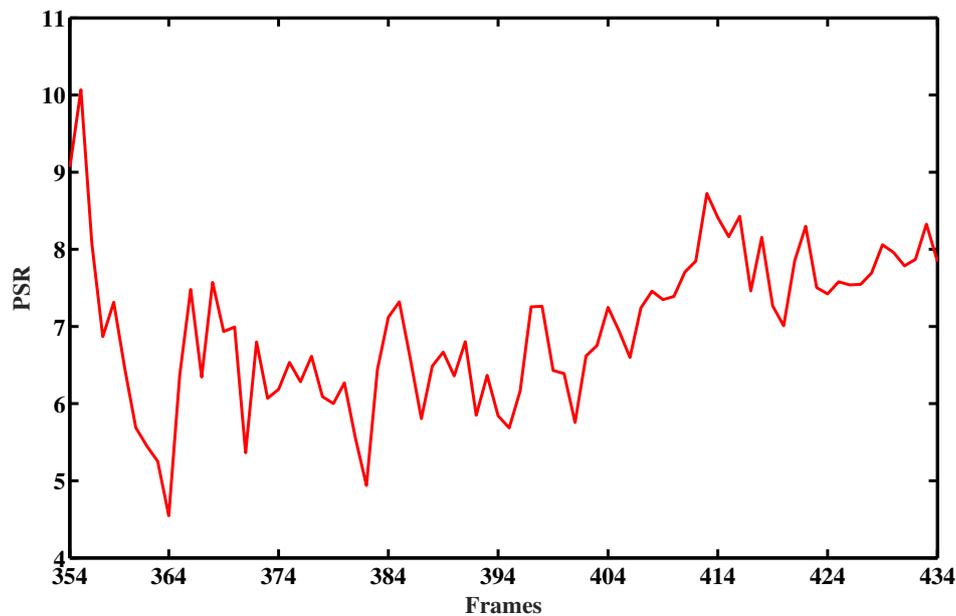Figure 3 is the PSR obtained by KCF in the image sequences.

Figure 3. The peak to sidelobe ratio (PSR) of image response diagram.

It can be seen from Figure 3 that the PSRs of the 8th, 9th, 10th, 11th, and 12th frames are generally lower, while the 45th and 68th frames are higher. It can be analyzed that the disappearance of the target appears in the 8th, 9th, 10th, 11th, and 12th frames, and the target in the 45th and 68th frames is

clear and normal, so it can be concluded that the larger the value of the PSR is, the better the tracking result. Based on this conclusion, an update threshold is set. In the tracking process, the tracking filter model is updated only when the PSR of the current frame is higher than the threshold. Otherwise, it is not updated. With this update strategy, there are several advantages. First, this is an indirect update method, which reduces the amount of computation caused by frequent frame-by-frame updates to a certain extent. Secondly, the proposed update strategy can automatically avoid updating the model under the current unfavorable situation when encountering the situation of tracking deviation, so as to prevent the model from drifting due to the introduction of erroneous information.

The filter model is mainly composed of two parts: One is the regression coefficient α, and the other is target template x in the nonlinear space. x is a feature extracted from the target region. α is obtained by the ridge regression training. The update process of the entire filter model is as follows:

$$\begin{cases} \alpha_t = (1 - \eta) \times \alpha_{t-1} + \eta \times \alpha_{t-1} \\ x_t = (1 - \eta) \times x_{t-1} + \eta \times x_{t-1} \end{cases} \tag{21}$$

where (α_t, x_t) indicates the filter model trained by the new detection results at the tth frame. (α_{t-1}, x_{t-1}) indicate the filter model at the $t - 1$th frame and η is interpolated coefficients.

## 4. Experimental Results and Analysis

This section shows numerical examples to validate the generality and effectiveness of the proposed IKCF for underwater target tracking in forward-looking sonar image sequences; the experimental data are the real forward-looking sonar data obtained in Qiandao Lake. The selected image sequences are processed by pseudocolor processing [20] in this paper. To evaluate the performance of the proposed IKCF for underwater target tracking, the normal target, the target with varying scales, and the target that disappears and appears again are selected for experimental comparison, and the proposed IKCF is compared with KCF [16], CSK [14], MOSSE [13], and mean-shift [7] to verify the effectiveness.

To demonstrate the effectiveness of the proposed IKCF for tracking the normal target, the 1st frame to the 60th frame are selected in forward-looking sonar image sequences. Figure 4 shows the tracking results of the normal target when the frame is 12th, 25th, 34th, and 51th, respectively. Center position error curves (1st frame to 60th frame) of normal target are shown in Figure 5. Location error precision curves (1st frame to 60th frame) of the normal target are shown in Figure 6. Overlap precision curves (1st frame to 60th frame) of the normal target are shown in Figure 7. The center position error [21] is the Euclidean distance between the real position of the underwater target and the position of the underwater target tracked by the algorithm, which is used to determine the accuracy of the tracking algorithm. The location error precision [16] shows the percentage of correctly tracked frames for a range of distance thresholds. The overlap precision [22] shows the percentage of correctly tracked frames for a range of overlap thresholds.

As seen from Figure 4 and 5, the mean-shift, KCF, and MOSSE have serious tracking errors in the tracking process; although the initial tracking error is relatively small, with the frame number increase, the error appears to have an upward trend. The proposed IKCF has a higher tracking accuracy and stability. As further seen in Figure 6, the ratio of the image frame number to the total frame numbers of the proposed IKCF is far superior to that of the other algorithms in the tracking result, and the center position error value of all the image frames is within 6 pixels, while that of the KCF, CSK, MOSSE, and mean-shift algorithms is within 10 pixels, 8 pixels, 14 pixels, and 20 pixels, respectively. Therefore, the proposed IKCF has a higher tracking accuracy. From Figure 7, the location error precision of the proposed IKCF is far superior to that of the other algorithms in the tracking result, and the overlap threshold of all the image frames is within 0.76, while that for the KCF, CSK, MOSSE, and mean-shift algorithms is within 0.64, 0.74, 0.48, and 0.38, respectively. Therefore, the proposed IKCF has a higher tracking accuracy.
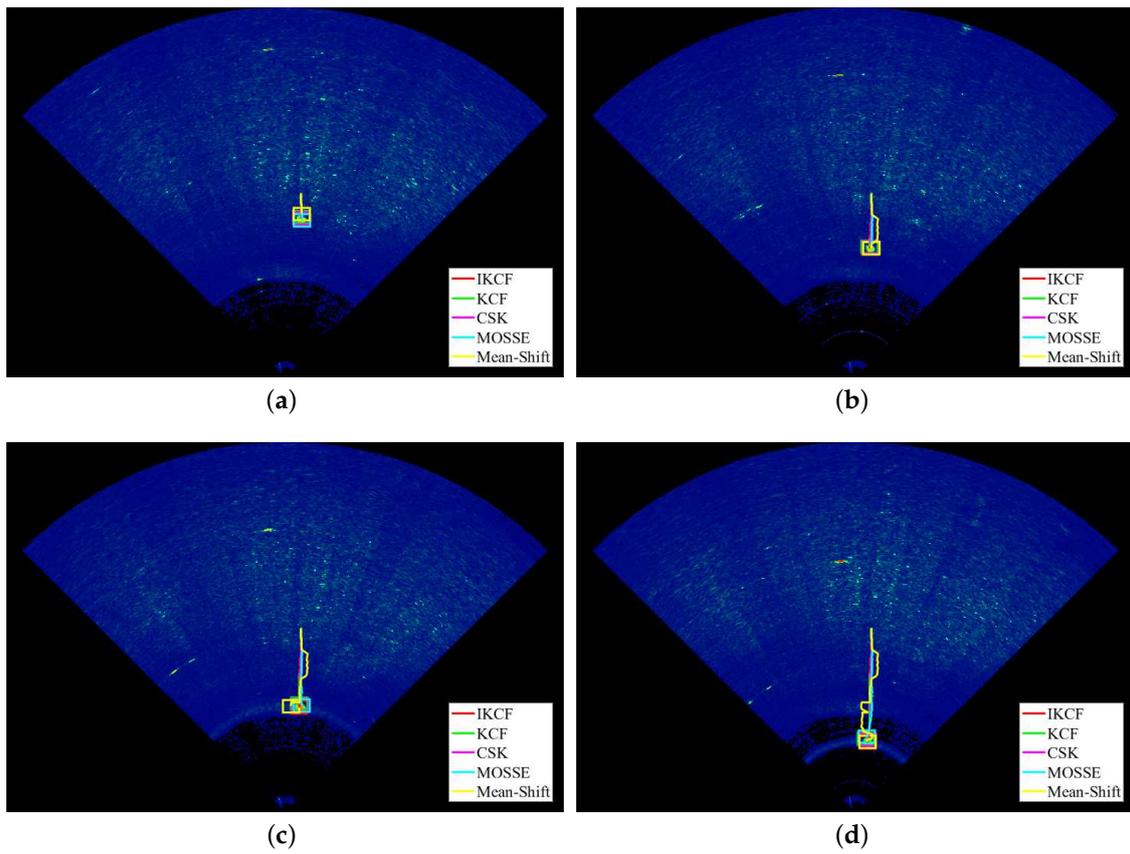
Figure 4. Tracking results for normal target: (**a**) The tracking results at the 12th frame; (**b**) the tracking results at the 25th frame; (**c**) the tracking results at the 35th frame; (**d**) the tracking results at the 51st frame.
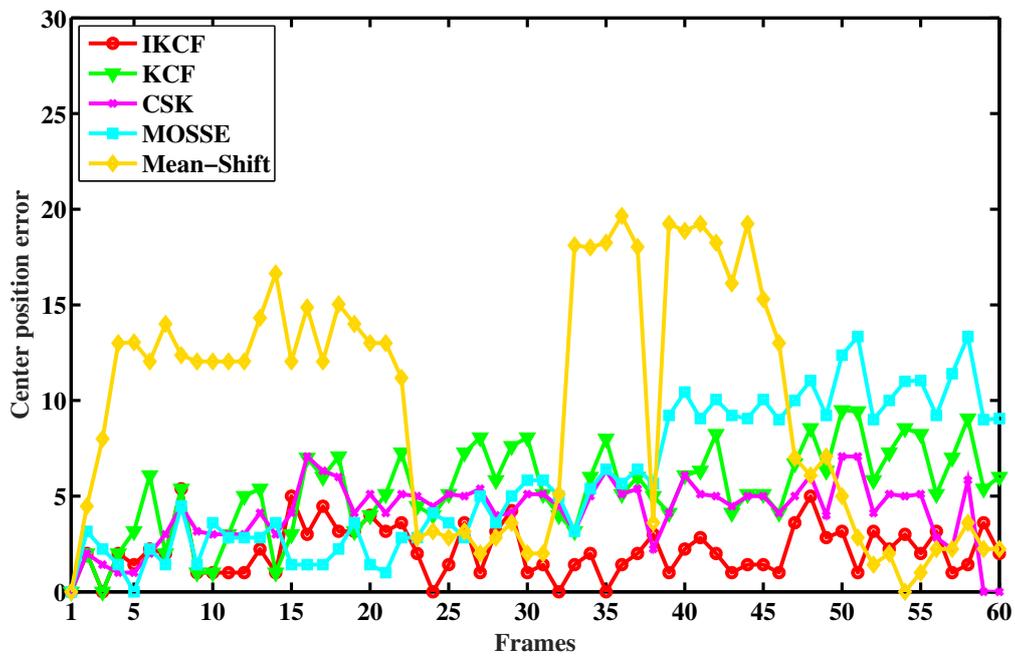
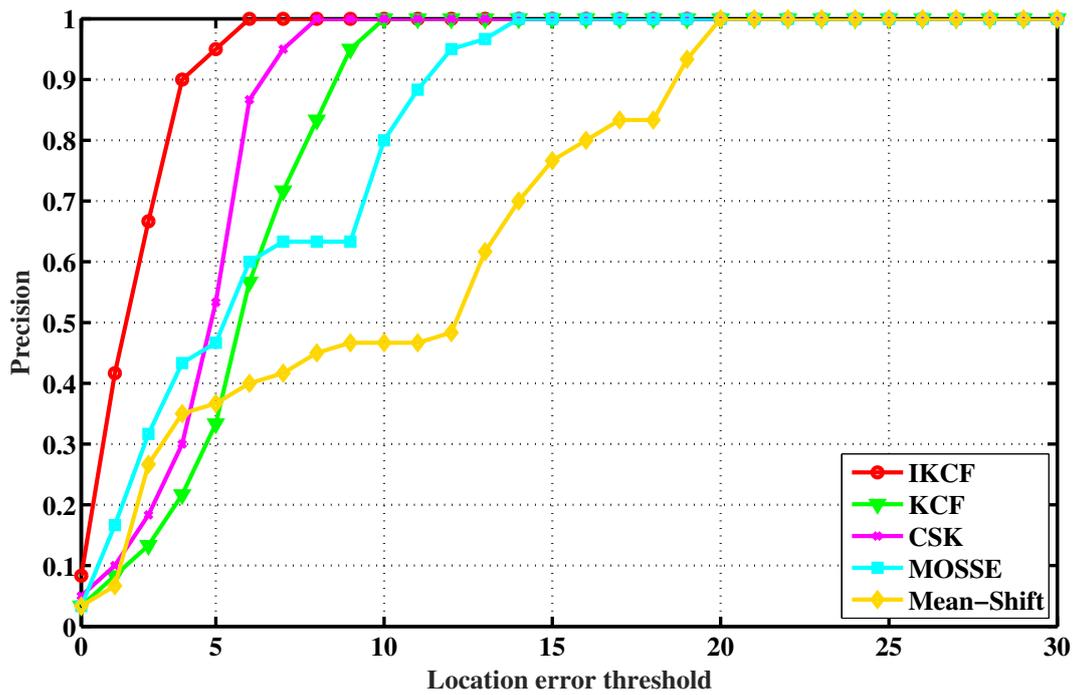**Figure 5.** Center position error curves (1st frame to 60th frame).

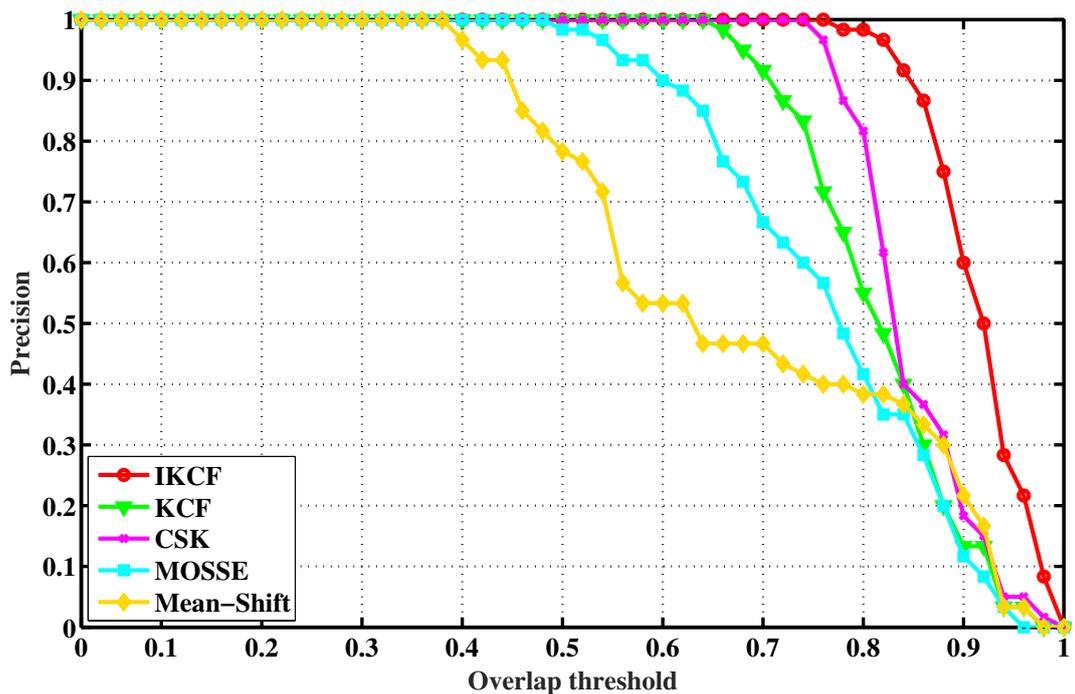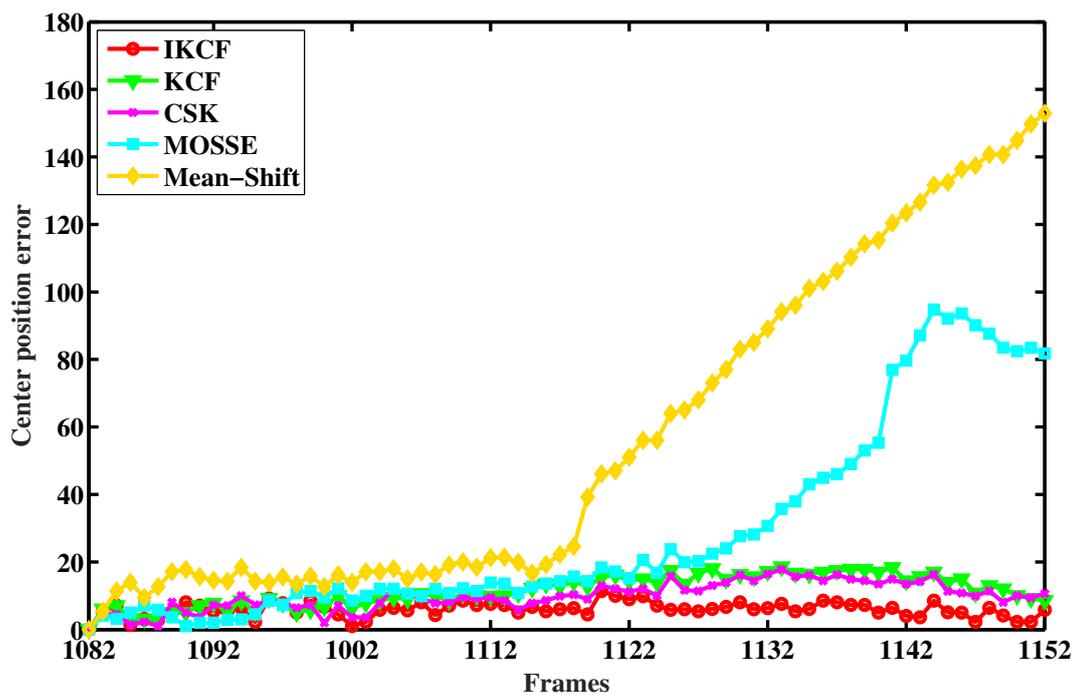**Figure 6.** Location error precision curves (1st frame to 60th frame).

Figure 7. Overlap precision curves (1st frame to 60th frame).

Figure 8 shows the tracking results of the target with varying scales when the frame is 1087th, 1093th, 1112th, and 1132th, respectively. Center position error curves (1082nd frame to 1152nd frame) of the target with varying scales are shown in Figure 9. Location error precision curves (1082nd frame to 1152nd frame) of the target with varying scales are shown in Figure 10. Overlap precision curves (1082nd frame to 1152nd frame) of the target with varying scales are shown in Figure 11.

(**a**)

(**b**)

(**c**)

(**d**)
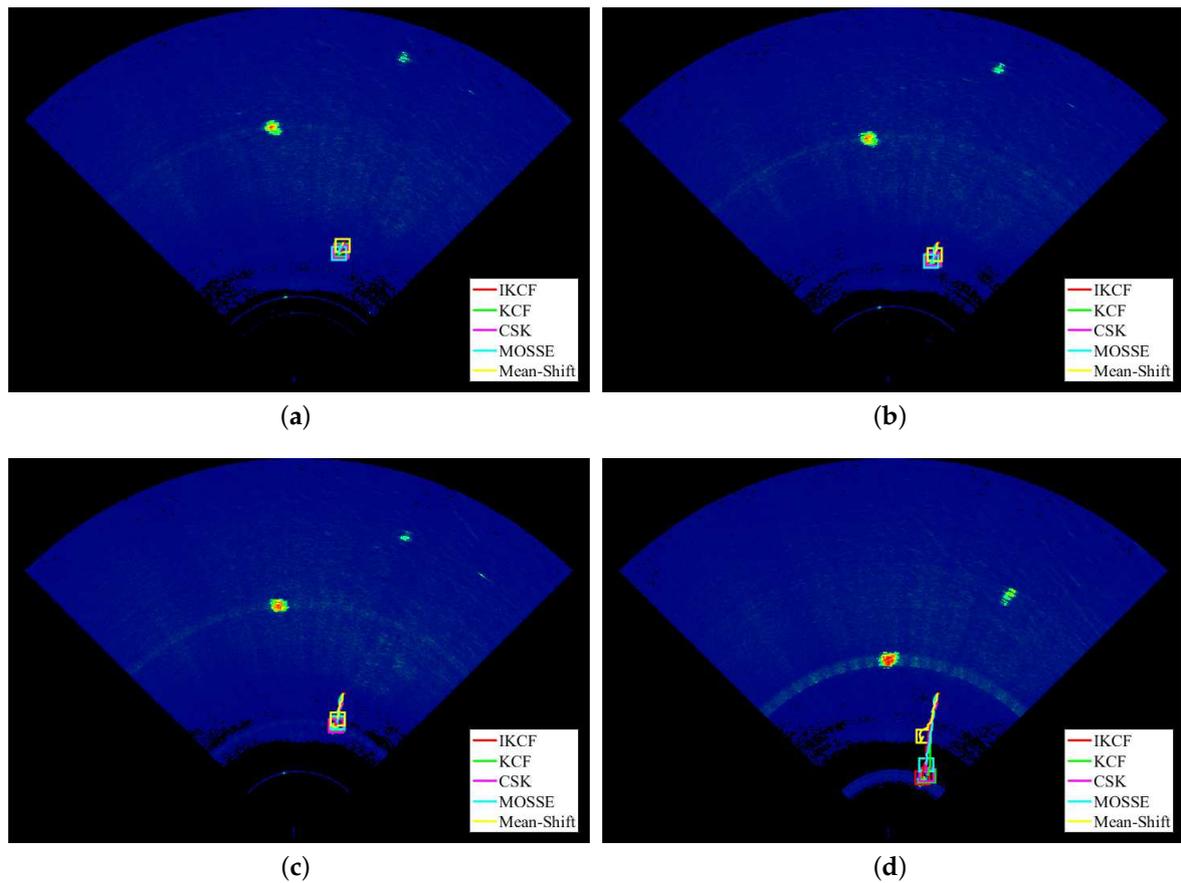
Figure 8. Tracking results for the target with varying scales: (**a**) The tracking results at the 1087th frame; (**b**) the tracking results at the 1093rd frame; (**c**) the tracking results at the 1112th frame; (**d**) the tracking results at the 1132nd frame.

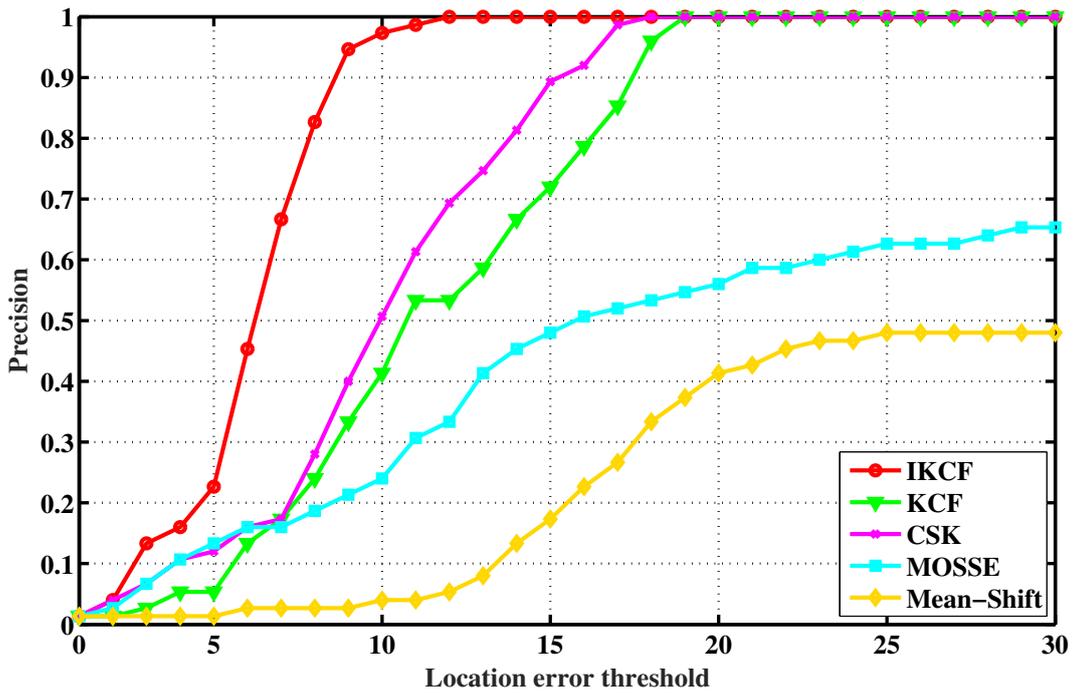**Figure 9.** Center position error curves (1082nd frame to 1152nd frame).

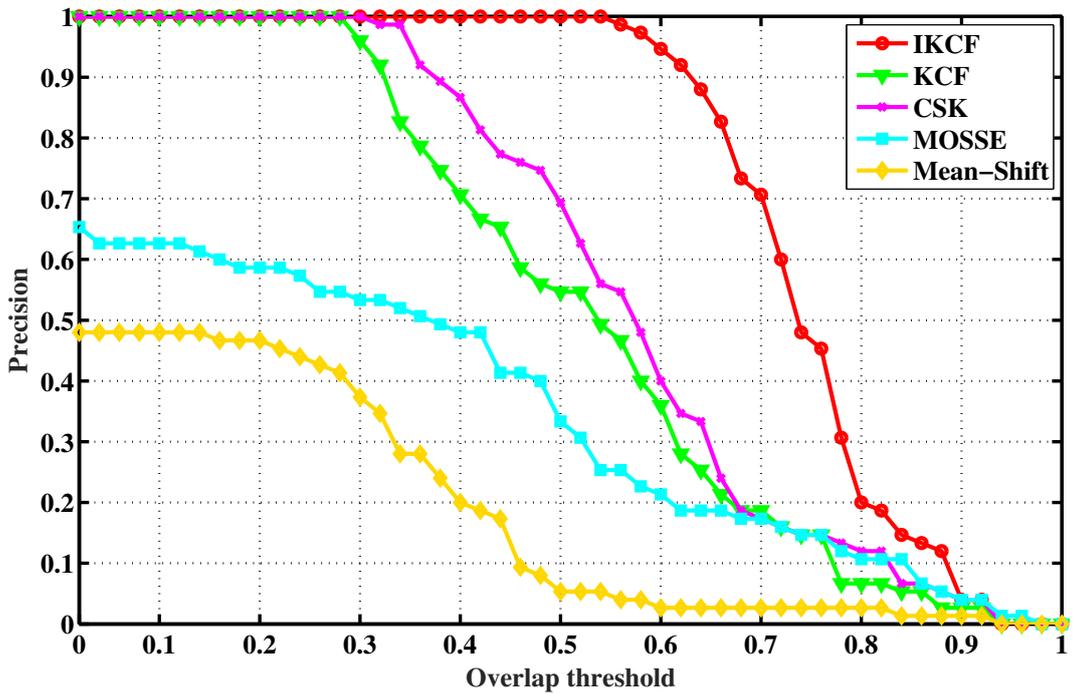**Figure 10.** Location error precision curves (1082nd frame to 1152nd frame).

Figure 11. Overlap precision curves (1082nd frame to 1152nd frame).

From Figure 8, it can be concluded that each algorithm in the first 1112 frames can track the position of the underwater target relatively correctly, while at the 1112nd frame, the mean-shift begins to deviate during the tracking process, and then at the 1132nd frame, the mean-shift and MOSSE algorithms all show tracking errors. In order to describe the underwater target tracking results more clearly, as can be further seen in Figure 9, the mean-shift and MOSSE algorithms have serious tracking errors in the tracking process. The KCF and CSK were compared with the proposed IKCF. The results showed that the proposed IKCF has a high tracking accuracy and can effectively track underwater

targets. Figure 10 shows that even if the maximum center error is 30 pixels, the ratios of the image frame number to the total frame number are only 48% and 65% in the tracking result of the mean-shift and MOSSE, respectively, for the same central position error value. The ratio of the image frame to the total frame of the proposed IKCF is higher than that of the KCF and CSK, and the central position error value of all the image frame numbers is within 12 pixels, while the KCF and CSK are within 19 pixels and 18 pixels, respectively. Therefore, the tracking accuracy of the proposed IKCF is higher when the target scale is changed in the image sequences. It is further seen from Figure 11 that the proposed IKCF is much better than the KCF, CSK, MOSSE, and mean-shift. Therefore, when the target scale changes, the proposed IKCF has stronger adaptability than other algorithms and can maintain a high tracking accuracy.

Figure 12 shows the tracking results when the target disappears and appears again when frame is 359th, 365th, 393rd, and 403rd, respectively. Center position error curves (354th frame to 434th frame) of the occluded target are shown in Figure 13. Location error precision curves (354th frame to 434th frame) of the occluded target are shown in Figure 14. Overlap precision curves (354th frame to 434th frame) of the target that disappears and appears again are shown in Figure 15.
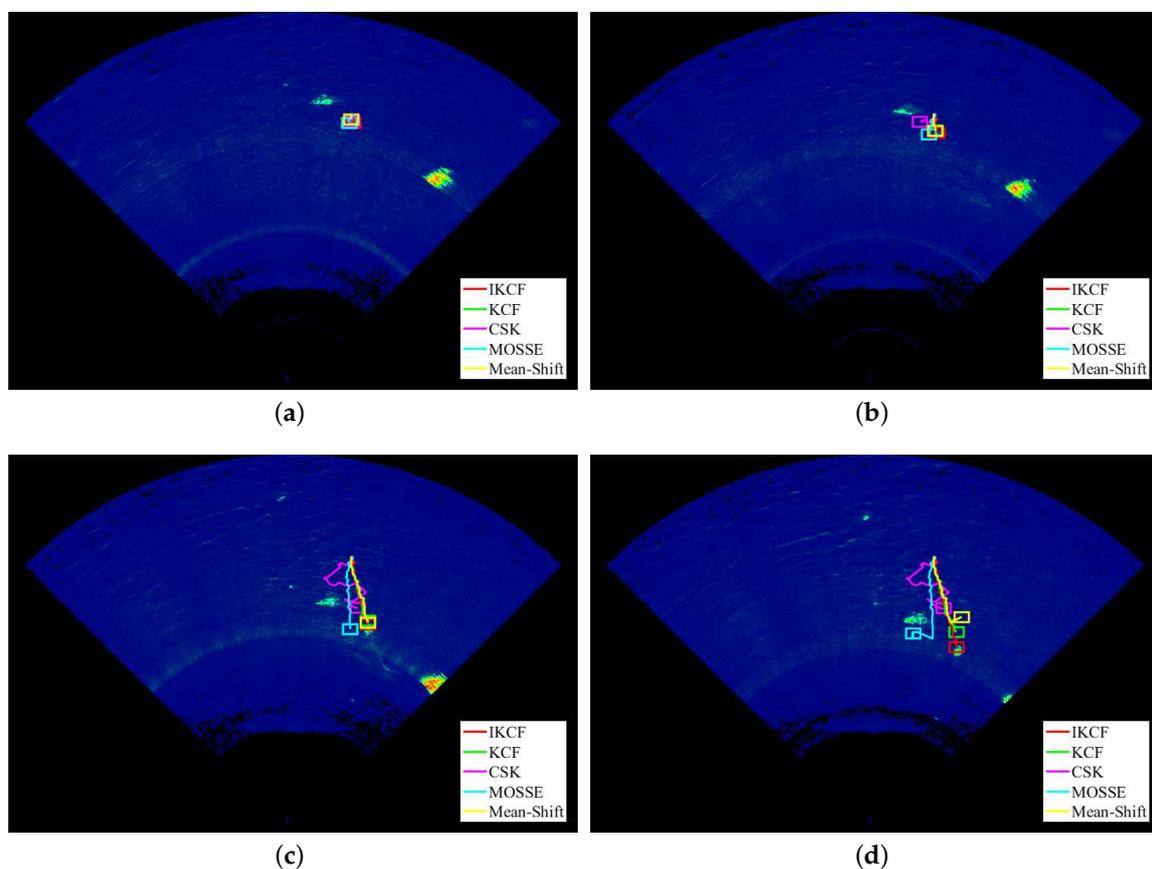
Figure 12. Tracking results for the target that disappears and appears again: (**a**) The tracking results at the 359th frame; (**b**) the tracking results at the 365th frame; (**c**) the tracking results at the 393th frame; (**d**) the tracking results at the 403rd frame.

From the tracking results in Figure 12, when the target disappears and appears again at the 365th frame, all the other algorithms could track the position of the underwater target correctly except CSK, and after the 365th frame, the tracking deviation of KCF, MOSSE, and mean-shift gradually increased. Until the 393rd frame, the MOSSE has a tracking error, at the 403rd frame, the KCF and mean-shift also have tracking errors; only the proposed IKCF still maintains the correct tracking trajectory and has a high tracking accuracy. As can be further seen in Figure 13, the central position error of the KCF,

CSK, MOSSE, and mean-shift increases as the target disappears, whereas the proposed IKCF still has a high tracking accuracy and stability, and has strong robustness. From Figures 14 and 15, regardless of the location error and the overlap threshold, the ratio of the image frame number to the total frame number cannot reach 100% of the KCF, CSK, MOSSE, and mean-shift algorithms. The proposed IKCF has a higher tracking accuracy compared with other algorithms and has a certain adaptability and effectiveness.
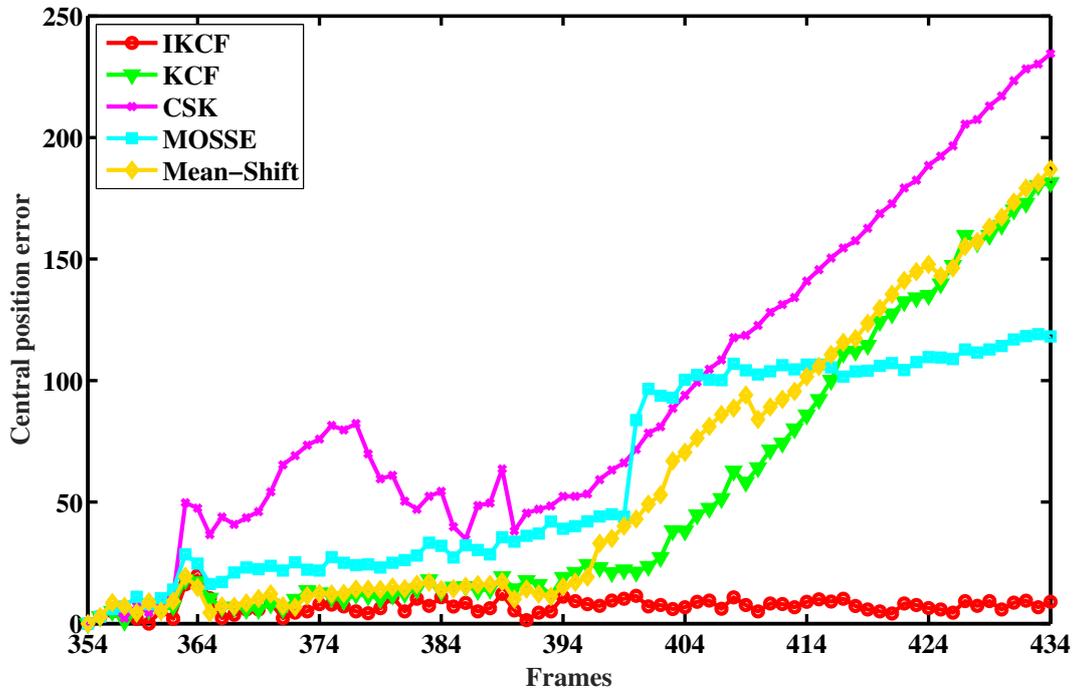

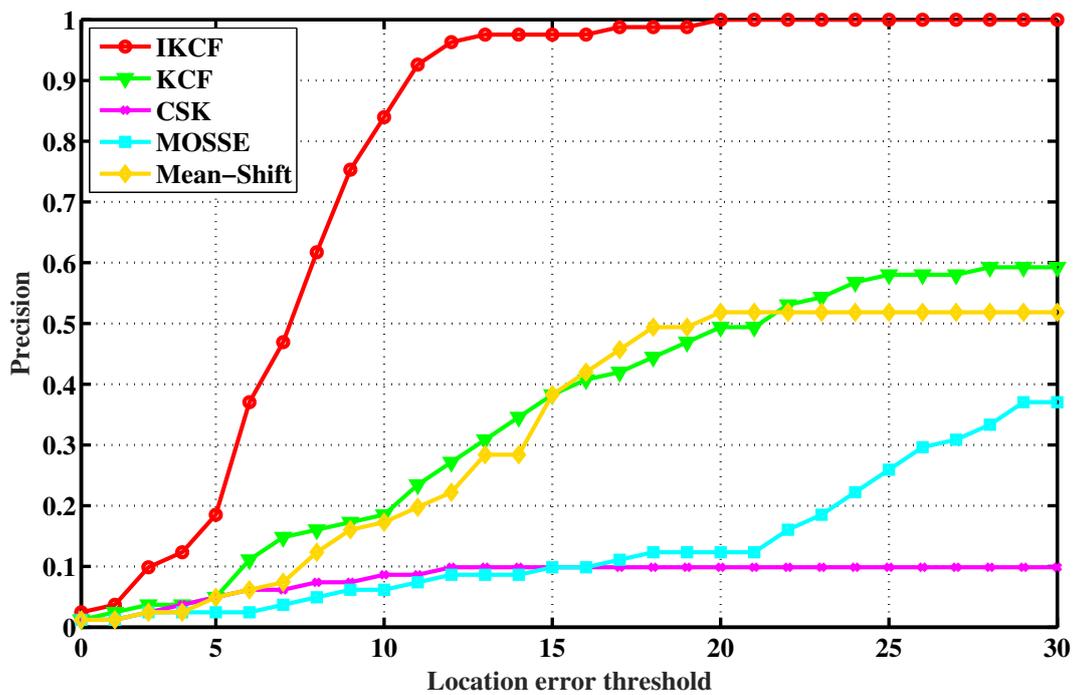**Figure 13.** Center position error curves (354th frame to 434th frame).

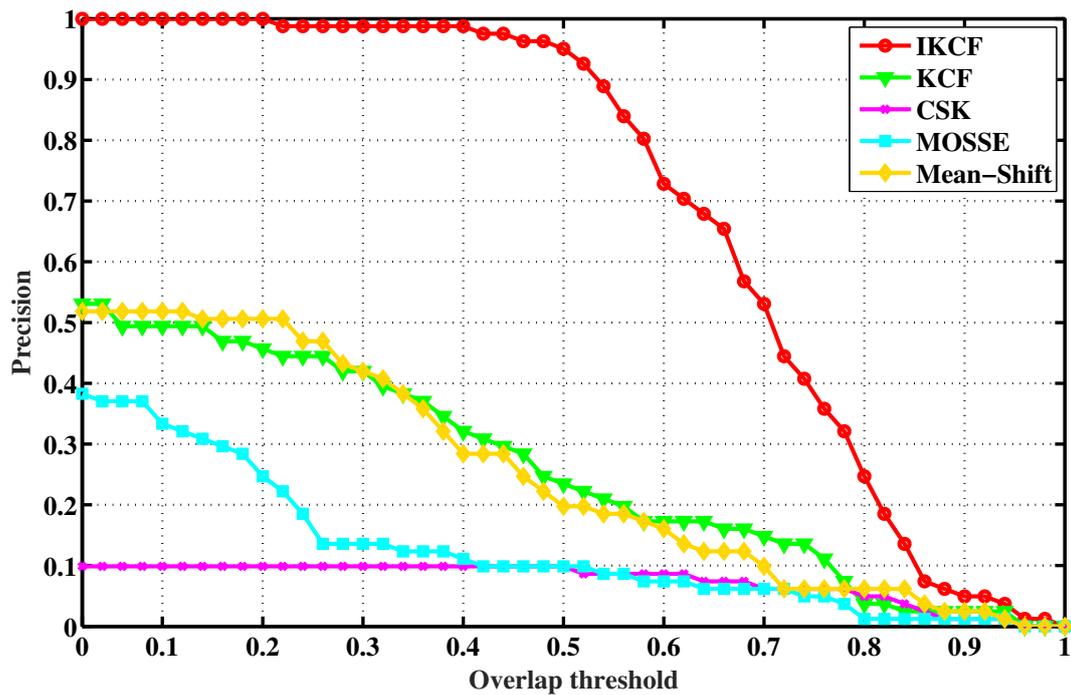**Figure 14.** Location error precision curves (354th frame to 434th frame).

Figure 15. Overlap precision curves (354th frame to 434th frame).

The traditional KCF uses a fixed-size base sample for training when tracking the target. In this way, if the target size is reduced, the background information contained in the base sample will increase, and the introduction of a large amount of background information eventually leads to poor training of the filter. If the target is enlarged, the background information contained in the base sample is correspondingly reduced, resulting in too many positive samples and too few negative samples, which is ultimately not conducive to filter training. The poor training effect of the filter means that it is likely to fail when the filter is used to detect the target in the detection process, and the failure of the detection will affect the subsequent detection and tracking process. Once the error is accumulated, it will eventually lead to tracking failure. For the base sample with a dynamically continuous scale proposed in this paper, when the target scale changes, the base sample also dynamically changes accordingly, which can ensure that the background information contained in the base sample is just right, neither too much nor too little, which in turn can guarantee the training effect of the filter.

The traditional KCF adopts a frame-by-frame linear updating strategy in the update of the filter model. However, the model is updated frame by frame, making the updates too frequent, which leads to a large amount of calculation. When the target disappears and appears again, the linear updating strategy will introduce erroneous information to the filter model, due to the accumulation of tracking error, and the error will be continuously amplified, which will eventually lead to the drift of the model and cause tracking failure. The proposed adaptive filter update strategy based on PSR can reduce the amount of computation caused by frame-by-frame linear updating to a certain extent, and the proposed method can automatically avoid updating the model under the adverse circumstances when the tracking deviation occurs, so as to prevent the model from drifting due to the introduction of the erroneous information.

Through the above verification and comparative experimental analysis, it is clear that the proposed IKCF has a better tracking accuracy in forward-looking sonar image sequences, and it still has effectiveness and adaptability, to some extent, with targets with varying scales and targets that disappear and appears again.

5. Conclusions

Considering the growing requirements of underwater target tracking, the IKCF was proposed to track targets in underwater sonar image sequences. For the first time, in order to solve the problem of the poor training effect with a fixed-size base in traditional KCF, a base sample with a dynamically continuous scale was proposed. Then, an adaptive filter update strategy based on PSR was adopted. In the process of filter model updating, the PSR was used to judge whether the model needs to be updated, which can reduce the computation and solve the filter drift problem caused by the target disappearing and appearing again. The visual and quantitative experimental results have shown that the proposed IKCF has a higher tracking accuracy and better effectiveness. Therefore, the proposed method can provide better underwater target tracking and has important theoretical and practical value.

References

1. Wang, X.; Hao, W.; Li, Q. An adaptive cultural algorithm with improved quantum-behaved particle swarm optimization for sonar image detection. *Sci. Rep.* **2017**. [CrossRef] [PubMed]
2. Song, S.; Herrmann, J.M.; Si, B.; Feng, X. Two-dimensional forward-looking sonar image registration by maximization of peripheral mutual information. *Int. J. Adv. Robot. Syst.* **2017**. [CrossRef]
3. Lei, F.; Zhang, X.; Yi, W. Real-time tracking of underwater moving target. In Proceedings of the 31st Chinese Control Conference, Hefei, China, 25–27 July 2012; pp. 3984–3988.
4. Ruiz, I.T.; Petillot, Y.; Lane, D.; Bell, J. Tracking objects in underwater multibeam sonar images. In Proceedings of the IEE Colloquium on Motion Analysis and Tracking, London, UK, 10 May 1999; pp. 1–7.
5. Liu, L.; Xu, W.; Bian, H. A LBF-associated contour tracking method for underwater targets tracking. *Proc. IEEE Oceans* **2016**. [CrossRef]
6. Lane, D.M.; Chantler, M.J.; Dai, D. Robust tracking of multiple objects in sector-scan sonar image sequences using optical flow motion estimation. *IEEE J. Ocean. Eng.* **1998**, *23*, 31–46. [CrossRef]
7. Chu, H.; Wang, K. Target tracking based on mean shift and improved kalman filtering algorithm. In Proceedings of the IEEE International Conference on Automation and Logistics, La Ferte-Bernard, France, 21–23 May 2009; pp. 808–812.
8. Quidu, I.; Bertholom, A.; Dupas, Y. Ground obstacle tracking on forward looking sonar images. *Proc. ECUA* **2010**, *14*, 8566–8577.
9. Quidu, I.; Jaulin, L.; Bertholom, A.; Dupas, Y. Robust multitarget tracking in forward-looking sonar image sequences using navigational data. *IEEE J. Ocean. Eng.* **2012**, *37*, 417–430. [CrossRef]
10. Zhang, T.; Wan, L.; Chen, Y.; Li, D. Object track in underwater sonar images. In Proceedings of the International Congress on Image and Signal Processing, Tianjin, China, 17–19 October 2009; pp. 1–4.
11. Li, M.; Ji, H.; Wang, X.; Weng, L.; Gong, Z. Underwater object detection and tracking based on multi-beam sonar image processing. In Proceedings of the IEEE International Conference on Robotics and Biomimetics, Shenzhen, China, 12–14 December 2013; pp. 1071–1076.
12. Hare, S.; Saffari, A.; Torr, P.H.S. Struck: Structured output tracking with kernels. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 263–270.
13. Bolme, D.S.; Beveridge, J.R.; Draper, B.A. Visual object tracking using adaptive correlation filters. *Comput. Vis. Pattern Recognit.* **2010**, *119*, 2544–2550. [CrossRef]

14. Rui, C.; 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; pp. 702–715.

15. Danelljan, M.; Khan, F.S.; Felsberg, M. Adaptive color attributes for real-time visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.

16. Henriques, J.F.; Caseiro, R.; Martins, P. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596. [CrossRef] [PubMed]

17. 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.

18. Xu, Y.; Wang, J.; Li, H.; Miao, Z. Patch-based scale calculation for real-time visual tracking. *IEEE Signal Process. Lett.* **2015**, *23*, 40–44. [CrossRef]

19. Akin, O.; Erdem, E.; Erdem, A. Deformable part-based tracking by coupled global and local correlation filters. *J. Vis. Commun. Image Represent.* **2016**, *38*, 763–774. [CrossRef]

20. Wang, X.; Wang, G.; Zhang, W. Pseudo-color processing of forward looking sonar image: An adaptive hot metal coding algorithm. In Proceedings of the 2018 Chinese Control And Decision Conference, Shenyang, China, 9–11 June 2018; pp. 394–399.

21. Bae, C.; Kyuchang, K.; Guang, L.; Chung, Y. A novel real time video tracking framework using adaptive discrete swarm optimization. *Expert Syst. Appl.* **2016**, *64*, 385–399. [CrossRef]

22. 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.