Visual Object Tracking using Surface Fitting for Scale and Rotation Estimation

Yuhao Wang and Jun Ma*
School of Physics and Optoelectronic Engineering
Taiyuan University of Technology
Taiyuan, 030600, China
[e-mail: zymajun@126.com, 2447015939@qq.com]
*Corresponding author: Jun Ma

Received June 27, 2020; revised September 23, 2021; accepted January 11, 2021; published May 31, 2021

Abstract

Since correlation filter appeared in the field of object tracking, it plays an increasingly vital role due to its excellent performance. Although many sophisticated trackers have been successfully applied to track the object accurately, very few of them attaches importance to the scale and rotation estimation. In order to address the above limitation, we propose a novel method combined with Fourier-Mellin transform and confidence evaluation strategy for robust object tracking. In the first place, we construct a correlation filter to locate the target object precisely. Then, a log-polar technique is used in the Fourier-Mellin transform to cope with the rotation and scale changes. In order to achieve subpixel accuracy, we come up with an efficient surface fitting mechanism to obtain the optimal calculation result. In addition, we introduce a confidence evaluation strategy modeled on the output response, which can decrease the impact of image noise and perform as a criterion to evaluate the target model stability. Experimental experiments on OTB100 demonstrate that the proposed algorithm achieves superior capability in success plots and precision plots of OPE, which is 10.8% points and 8.6% points than those of KCF. Besides, our method performs favorably against the others in terms of SRE and TRE validation schemes, which shows the superiority of our proposed algorithm in scale and rotation evaluation.

Keywords: Object Tracking, Fourier-Mellin Transform, Confidence Evaluation, Surface Fitting

http://doi.org/10.3837/tiis.2021.05.009 **ISSN : 1976-7277**

1. Introduction

Object tracking is one of the most heated topics in computer vision and pattern recognition, which has already been widely utilized in various application fields, such as behavior recognition, human-computer interaction, surveillance and unmanned driving [1]. Although it has scored remarkable progress in the past decades, there still remains a series of challenges due to the environmental variation, especially when the target is heavily obscured or undergoes scale and rotation changes.

Recently, a variety of tracking methods have been proposed, and these methods can be separated into two main categories: one is the generative approach and the other is the discriminative approach. The generative tracking algorithm usually constructs an appearance model to depict the target and looks for the object regions with best matching scores [2-4]. Unlike the generative model, the discriminative model [5-8] trains a classifier which can effectively distinguish between the object and the background. Correlation filters as the discriminative approach have achieved a great success in the object tracking community due to its rapid speed of calculation and robust tracking performance in recent years.

Since Bolme et al. introduce MOSSE [9] into visual object tracking field, correlation filter algorithms have attracted great attention of researchers due to their splendid tracking performance. Henriques et al. introduce the cyclic matrix for object tracking in CSK [10] and then improve the KCF [11] tracker by extending the single-channel grayscale feature to a HOG-based multidimensional feature, which greatly improves the robustness of the tracking. Although the above algorithms achieve great success, few people pay attention to how to estimate scale and rotation changes efficiently and accurately. For the purpose of coping with the scale changes of the target, Zhang et al. [12] propose a scale update strategy by using the generated confidence map in STC. Aiming at the problem of fixed template size in kernel correlation filter, Li and Zhu [13] utilize the scale pool method to achieve adaptive tracking of target in SAMF. Danelljan et al. [14] raise the DSST tracking algorithm which trains the classifier on the scale pyramid to figure out the problem of scale estimation. Unfortunately, despite all these methods achieve excellent tracking accuracy, they rarely take the rotation changes of the target into account. In addition, they handle the scale changes through learning discriminative CF based on a scale pyramid representation, which is limited and insensitive.

In the visual tracking scene, model drift may appear due to the accumulation of errors, which may lead to tracking failure. In order to improve the robustness of tracking, many researchers pay attention to the deep convolutional neural networks (CNN) [15-17] method to boost the tracking performance. The HCF [18] extracts the CNN features which can accurately locate the target and contain more semantic information. The is an On the basis of SRDCF [19], DeepSRDCF [20] replaces hand-crafted features with CNN features and achieves good tracking results. ECO [21] extracts a very comprehensive feature (CNN+HOG+CN) and introduces a factorized convolution approach to reduce model parameters. Nam et al. [22] propose the MDNet algorithm which uses large-scale video with annotated boxes to train CNN to obtain a general feature. SiamFC introduced by Bertinetto [23, 24] construct a fully convolutional network between the regions and the exemplar image which utilizes the mask of the first frame to match the subsequent frames to calculate the pixel-level score map. Although convolutional neural network can extract rich information features and achieve excellent tracking effect, due to the complexity of deep neural network structure, it puts forward higher requirements for GPU and other hardware.

It is a common situation that object target emerges scale and rotation changes during the tracking process. Yin [25] utilize a scale adaptive method to improve the tracking performance

and reduce the computational costs. Besides, in the past studies, more consideration has been given to the estimation of the scale changes while few of them have been made to analyze the rotation motion of the target. Therefore, how to achieve robust object tracking when the target has rotation is still a challenging research problem. Furthermore, many algorithms do not make reliable judgments on the tracking results and the results of each frame are used to renew the model even though the tracking result is prone to drift. Although LMCF [26] put forward an approach to prevent model drift through appraising the confidence of the tracking result, it can not comprehensively evaluate scale-rotation and translation changes. In the paper [27, 28], although the scale estimation problems can be solved by a scale pyramid representation and the model drift is overcome by a high-confidence judgement mechanism, the method is limited to solve the rotation estimation problems. Based on the above discussion we can see that these methods have both advantages and disadvantages. **Table 1** summarizes several typical object tracking methods.

Table 1. Comparison of characteristics of several object tracking algorithms

Type	Trackers	Scale	Rotation	Confidence	GPU
Correlation filter	KCF	☒	☒	☒	☒
	DSST	✓	☒	☒	☒
	LMCF	✓	☒	✓	☒
Deep Learning	ECO	✓	☒	☒	✓
	MDNet	✓	☒	☒	✓
	SiamFC	✓	☒	☒	✓

To address the above limtations, we propose a valid visual object tracking algorithm FMCS to address the challenging problem of scale and rotation changes. As shown in **Fig. 1**, we can see that the tracking target undergoes a posture transformation during movement, and the direction of the target on the image plane can be predicted according to the direction information of the rotated bounding box. We know that the Fourier-Mellin transform can convert the scale and rotation variations of the original image into pure translation on a log-polar image, which has been successfully applied in the field of image registration. Inspired by this idea, we employ a correlation filter to predict the translation motion and estimate the scale and rotation angle changes by Fourier-Mellin transform in log-polar coordinates. Specifically, it is well known that Fourier-Mellin transform is an extended phase correlation algorithm and the phase correlation algorithm can only measure the displacement of the whole pixel. When the target translation is not an integer, the peak energy obtained by phase correlation will diffuse to the surrounding adjacent pixels, which will form many small burrs around the main peak and reduce the accuracy of translation parameter estimation [26]. Therefore, in response to the above problem, we propose a quadric surface fitting for the points around the registration position of the whole pixel. Besides, in order to establish a stable and robust rotation and scale parameter estimation model, we exploit a confidence evaluation strategy to acquire a joint optimization result of translation, scale and rotation changes. Extensive experiments demonstrate that our algorithm FMCS has an outstanding performance in complex environment.

In this paper, we present a novel efficient scale and rotate estimation algorithm FMCS. It can predict the position, scale and rotation angle of the target in the image at a speed of about 25 fps on CPU. We take advantage of the scale and rotation changes factor to update the target

appearance model. In addition, making use of the APCE standard, we propose a novel confidence evaluation strategy to evaluate the variation factors for purpose of getting a better tracking model. Our algorithm consists of three parts: (1) scale and rotation angle estimation, (2) confidence evaluation strategy, and (3) quadric surface fitting. The contributions made in this article can be summarized in three points:

1) We develope a robust tracker which combines correlation filter and Fourier-Mellin transform. The advantages of the tracker is introduced to improve the expressive ability of translation, scale and rotation estimation.
2) A confidence evaluation strategy is developed to judge the tracking performance by comprehensively evaluating the results of translation, rotation and scale changes. We comprehensively evaluate and update the tracking model through the changes of the three indicators. This method can solve the problems of tracking failure and model drift to a certain extent.
3) We introduce a surface fitting mechanism to improve algorithm accuracy. Extensive experiments on OTB100 show that our method achieves superior performance.

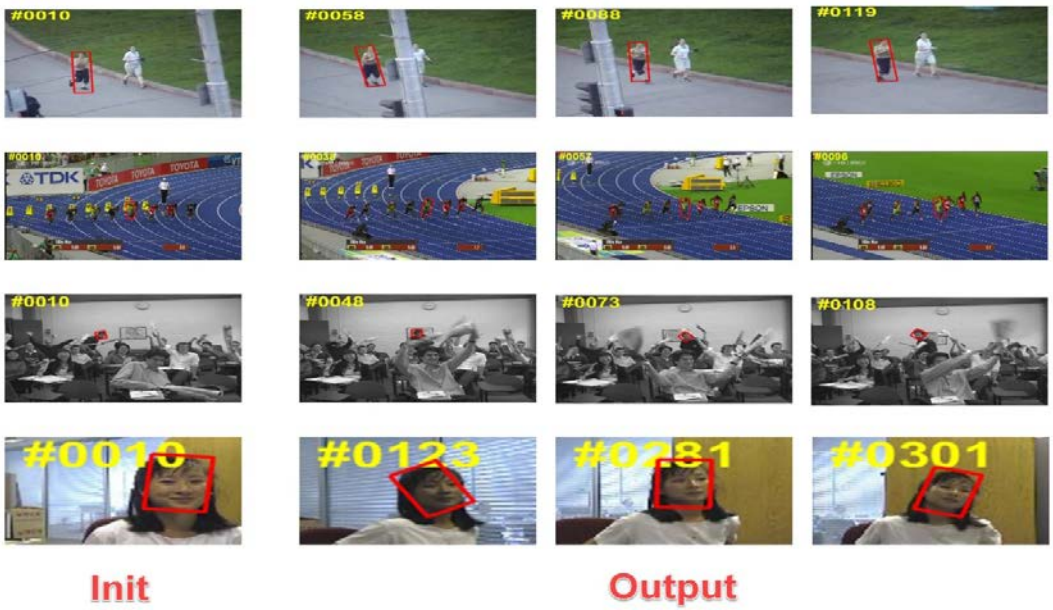

Fig. 1. We propose the scale and rotate estimation algorithm, the representative patches for correlation and updating of sequences is shown above.

2. Our Approach

In order to achieve target tracking with rotation and scale adaptability, a novel ensemble algorithm is proposed in this paper to estimate the confidence of the tracking result and the displacement, scale and rotation parameters of the target. Simultaneously, a surface fitting mechanism is used to improve algorithm accuracy. The specific content of the algorithm is shown in **Fig. 2**.

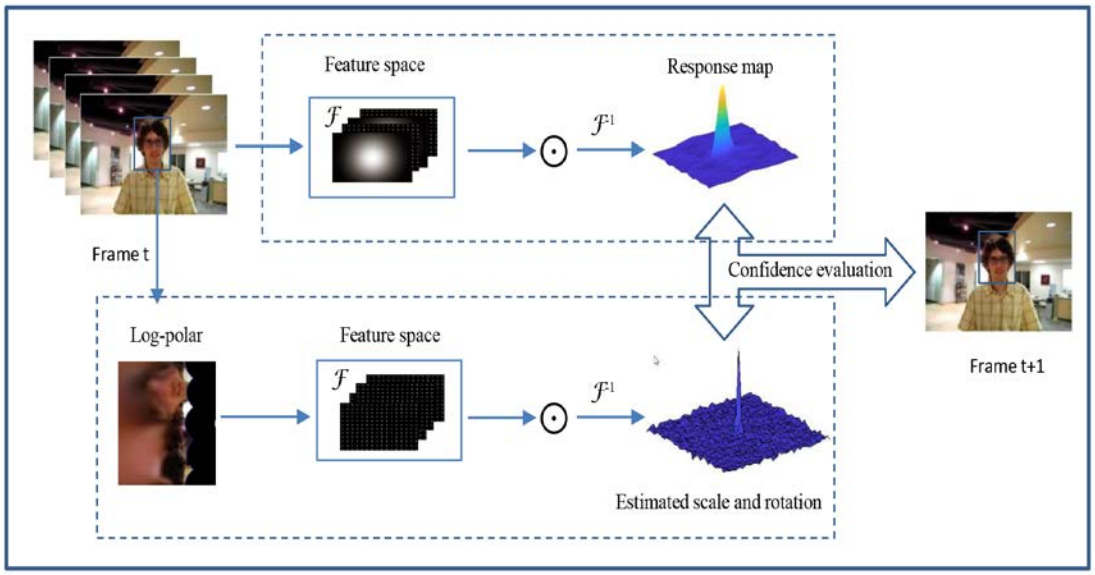

Fig. 2. Overview of the proposed algorithm

In the first place, the first frame of the video is adopted to get initial parameters of the previous target position, scale and rotation angle. Then calculate the target position, scale and rotation parameters of the next frame through the correlation filter and Fourier-Mellin transform. In order to achieve higher accuracy, a quadric surface fitting method is used to get the sub-pixel position. Finally, the confidence evaluation strategy is proposed to achieve the best comprehensive performance under the influence of three factors: translation scale and rotation transformation. If the strategy is effective, update the model of the target. **Fig. 3** gives a flow diagram to show the process of our entire paperwork.

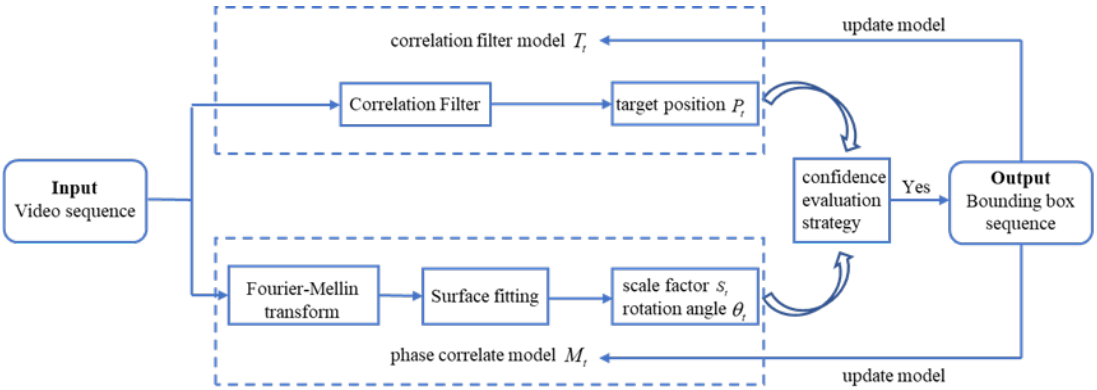

Fig. 3. The flow diagram of our paperwork

2.1 Translation Estimation by Kernelized Correlation Filters

The correlation tracking provides an elegant framework for efficient and effective object tracking. In this paper, the kernel correlation filtering (KCF) architecture is adopted. The idea of this algorithm is to construct a circular matrix by cyclically sampling the image blocks to enrich the target samples. Then, based on the characteristics of the cyclic matrix, the solution

of the problem is transformed into the discrete Fourier transform domain by using the diagonalization of the cyclic matrix. This process greatly reduces the computational complexity and increases the training speed.

The sample training uses the ridge regression function to train the classifier. The trained classifier is used to judge the correlation between the newly input image and the previous frame image. The peak position of the target response can be considered as the possible location of the predicted target. The sample training process is a ridge regression problem. The purpose of training is to find the function $f(z) = w^T z$ that minimizes the squared error. The filter w can be expressed as follows:

$$w = \min_w \sum_i (f(x_i) - y_i)^2 + \lambda \| w \|^2 \tag{1}$$

The kernel correlation filter is to introduce a kernel function to solve the ridge regression of the kernel space. The feature space is mapped to a higher dimensional space by a nonlinear mapping function $\phi(x)$, so that the mapped samples are linearly separable in the new space. At this point, the filter w can be defined as

$$w = \sum_i \alpha_i \varphi(x_i) \tag{2}$$

Therefore, the optimization problem is transformed into finding the optimal vector α. We introduce a kernel function and rewrite object function as follows:

$$f(z) = w^T \varphi(z) = \sum_{i=1}^{n} \alpha_i \kappa(z, x_i) \tag{3}$$

We can solve the problem by

$$\alpha = (K + \lambda I)^{-1} y \tag{4}$$

Where K is the construction of the cyclic matrix, and I is the identity matrix.

Using the properties of the circulant matrix, α is solved by the discrete Fourier transform.

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \tag{5}$$

Where $\hat{k}^{xx}$ is the kernel correlation of x and itself.

In order to track the target quickly and determine the location of the current frame, we introduce the kernel matrix of the detection

$$K^z = C(k^{xz}) \tag{6}$$

Then the response is calculated by

$$\hat{f}(z) = \hat{k}^{xz} \odot \hat{\alpha} \tag{7}$$

Where $\hat{k}^{xz}$ is the kernel correlation of z and the base sample x. According to (7), the new position of current frame target locates in the maximal value of output response map.

2.2 Scale and Rotation Estimation by Fourier-Mellin Transform

Phase correlation is a typical transform domain-based image registration method which enables accurate registration between images that only exist in translational motion. Later, Reddy et al. [29] propose the Fourier-Mellin algorithm based on the phase correlation. The algorithm transforms the Cartesian coordinate system into a log-polar coordinate system where the rotation and scale changes can be viewed as the pure translational moving along the axis. In the target tracking process, the rotation and scale parameters of the target can be estimated by log-polar transformation of the original image. The Fourier-Mellin transform convert the scale and rotation changes on the original image into the frequency domain, which

is different from DSST and other scale-based pyramid estimation methods. This method can realize the scale and rotation parameters estimation of continuous space.

Suppose there is an image $f_1(x, y)$, and the image $f_2(x, y)$ is obtained by rotation θ_0, scale factor α, and translation (x_0, y_0). The transformed image $f_2(x, y)$ can be expressed as:

$$f_2(x,y) = f_1\left(\frac{x\cos\theta_0 + y\sin\theta_0 - x_0}{\alpha}, \frac{-x\sin\theta_0 + y\cos\theta_0 - y_0}{\alpha}\right) \tag{8}$$

Their Fourier transforms of them are related by

$$F_2(\xi,\eta) = \frac{e^{-j2\pi(\xi x_0+\eta y_0)}}{\alpha^2} \times F_1\left(\frac{\xi\cos\theta_0 + \eta\sin\theta_0}{\alpha}, \frac{-\xi\sin\theta_0 + \eta\cos\theta_0}{\alpha}\right) \tag{9}$$

Ignoring $\left|\frac{e^{-j2\pi(\xi x_0+\eta y_0)}}{\alpha^2}\right|$,and letting M_1 and M_2 be the Fourier magnitude spectra of two images. Then Fourier magnitude spectra between the two only has a relationship of rotation and scale change. We can get the following expression

$$M_2(\xi,\eta) = M_1\left(\frac{\xi\cos\theta_0 + \eta\sin\theta_0}{\alpha}, \frac{-\xi\sin\theta_0 + \eta\cos\theta_0}{\alpha}\right) \tag{10}$$

Take the logarithm of the Fourier magnitude spectra and then perform polar transformation, the above expression can be expressed as

$$M_2(\log\rho,\theta) = M_1(\log\rho - \log\alpha, \theta - \theta_0) \tag{11}$$

Then, the equation is rewritten as follow,

$$M_2(\phi,\theta) = M_1(\phi - d, \theta - \theta_0) \tag{12}$$

Where $\phi = \log\rho, d = \log\alpha$

Finally, the rotation and scale changes between the images are transformed into a translational motion on the distance and angle axes of the log-polar coordinate images.

The process of Fourier-Mellin transformation is as follows:

1) Given an image patch to be detected, and then we convert it into log-polar.

2) By extracting the hog features of the patch, the detection area model ψ_t can be learned with numerous training samples efficiently through fast Fourier transform (FFT).

3) The phase correlation between the previous frame model ψ_{t-1} and the detection region model ψ_t is performed, and then search the position of the peak pulse.

The ideal result has only one peak position, and the other positions are 0. However, in actual experiments, the inverse transformation of the normalized cross power spectrum phase shows that it not only has a correlation peak, but also has some uncorrelated peaks in the three-dimensional image. The position of the peak represents the translation parameter, and the energy of the peak reflects the correlation between the two images.

When the Fourier-Mellin transform method is used for image registration, the calculated translation amount can only be estimated to integer pixels. In practical applications, the displacement value of an image is generally a decimal pixel, that is, a sub-pixel. When the offset is an integer, the pulse function is unimodal; when the offset is a sub-pixel, the pulse function is multimodal, with the highest peak accompanied by several small peaks. It can be determined that the actual offset is not at the position of any one peak, but between some of the highest peaks. In view of the above problems, we propose an algorithm that combines the

Fourier-Mellin transform and surface fitting to further improve the detection accuracy during target tracking.

2.3 Sub-pixel Registration Based on Surface Fitting Method

Surface fitting is a technique often used in graphic imaging. When doing experiments or measurements, some discrete data points are often obtained. These discrete data points often have errors and are not completely accurate, which affects the analysis of experimental results. Normally, we need to use the surface fitting technology to find the specific parameters of the function based on discrete data points. Then, the experimental data is estimated based on the fitted surfaces to obtain more accurate data. The surface fitting method has the advantages of strong robustness, high accuracy and high calculation efficiency, and it has been widely used in practical applications.

In order to solve the problem of low detection accuracy of the phase correlation method, a quadric surface fitting method is used in this study. Taking the peak position as the center, the neighborhood data is fitted and the exact sub-pixel matching position is obtained by solving the extreme point. For this reason, after applying the Fourier-Mellin transform to the entire pixel-level registration of the image, a 4×4 matrix is selected centering on the position of the pulse peak, and the 16 pixels are used as the measured data to fit the quadric surface. The general equation of a quadric is

$$f(x, y) = ax^2 + bxy + cy^2 + dx + ey + f \tag{13}$$

Where a, b, c, d, e, f are the coefficients to be measured of the equation, which can be solved by the method of least squares. The least square method is the most commonly used method in surface fitting.

Using the least squares approximation theory, the sum of the squared deviations between the estimated data and the measured data is

$$E_r = \sum (z - f(x, y))^2 \tag{14}$$

Where z is the measured data and E_r is the sum of the squares of all deviations. To minimize E_r, we find the partial derivatives of E_r about a, b, c, d, e, f, and make their partial derivatives equal to 0.

Then, the solution of (14) can be used to find the coefficients to be measured of the equation. The peak position of the fitted quadric surface is the actual position of the pulse peak.

2.4 Confidence Evaluation Strategy

This section mainly describes the confidence evaluation strategy. We judge the tracking results by comprehensively evaluating the results of translation, rotation and scale change. By comparing the confidence of the tracking results multiple times and finding the optimal value, the tracking results are more accurate and robust.

During the tracking process, the target may be occluded, rotated, and scale changed. The main task of the tracking algorithm is to robustly estimate the position of the target in these challenging situations. In the process of the tracking, it is very meaningful to accurately judge the current tracking status of the target. The larger the peak value of the response map, the better the tracking result. Conversely, if the peak value is low, the whole response map may emerge many spurious peaks and continue to oscillate.

According to [30], the APCE is considered as the best standard for describing the degree of fluctuation in response map. The APCE is expressed as

$$APCE = \frac{\left|F_{\max} - F_{\min}\right|^2}{mean(\sum_{x,y}(F_{x,y} - F_{\min})^2)} \tag{15}$$

In this case, we propose our confidence evaluation strategy.

$$\kappa = 0.002 APCE_t + 0.8 F_{\rho(\max)} \tag{16}$$

Where κ is confidence index, $APCE_t$ is the fluctuation value of translation module, $F_{\rho(\max)}$ is the phase correlation peak value.

The κ value in the tracking process is used as the tracking confidence threshold. By comparing the κ value results in the tracking process, the position with the smallest target fluctuation and the highest correlation is found as the optimal result. We briefly introduce the outline of our method in **Algorithm FMCS**.

Algorithm FMCS: Proposed tracking algorithm**.**
Input: Status of the target $S_{t-1} = (P_{t-1}, S_{t-1}, \theta_{t-1})$ Current image I_t Previous target position P_{t-1}, scale factor S_{t-1}, and rotation angle θ_{t-1} Corrected kernel correlation filter model T_{t-1} Phase correlate model M_{t-1}
Step 1: Estimate the target position P_t' using the kernel correlation filter T_{t-1}. Step 2: Compute the $F_{\rho(\max)}$ using the phase correlate model M_{t-1} Step 3: Surface fitting to get the sub-pixel position, estimate the scale factor S_t' and rotation angle θ_t' Step 4: Compute the APCE and confidence index κ if $\kappa_n < \kappa_{n+1}$ and $n<5$, $S_t \leftarrow (P_{t-1}, S_t', \theta_t')$ end
Output: Status of the target $S_t = (P_t, S_t, \theta_t)$ Target position P_t, scale factor S_t and rotation angle θ_t Updated corrected kernel correlation filter T_{t-1} and phase correlate model M_{t-1}

3. Experimental Classification Results and Analysis

3.1 Implementation Details

The simulation experiments in this article were performed in a computer environment with an Intel i5-3210 2.5GHz CPU and 8GB RAM, and all the methods were implemented in matlab2018a.

3.2 Datasets

In order to fully verify the accuracy and robustness of the proposed algorithm FMCS, OTB100 [31] was selected as the test dataset in this experiment, which contains 100 videos. The

OTB100 dataset divides the video sequence motion scene into 11 attributes, and the results are shown in **Table 2**. The details are as follows: DEF, OCC, FM, IV, SV, LR, OPR, MB, BC, OV and IPR. Through these factors we can have a more comprehensive and reliable evaluation of the tracking algorithm tested.

Table 2. Annotated Sequence Attributes in the Performance Evaluation

Attributes	Description
DEF	Deformation – Non-rigid object deformation
OCC	Occlusion – The target is partially or fully occluded
FM	Fast Motion – The motion of the ground truth is larger than tm pixels (tm=20)
IV	Illumination Variation –The illumination in the target region is significantly changed
SV	Scale Variation –The ratio of the bounding boxes of the first frame
LR	Low Resolution – The number of pixels inside the ground-truth bounding box is less than tr (tr=400)
OPR	Out-of-Plane Rotation –The target rotates out of the image plane
MB	Motion Blur –The target region is blurred due to the motion of target or camera
BC	Background Clutters – The background near the target has similar color or texture as the target
OV	Out-of-View – Some portion of the target leaves the view
IPR	In-Plane Rotation – The target rotates in the image plane

3.3 Tracking Algorithm Evaluation Index

We evaluate the proposed method on OTB100 and all the tracking methods are evaluated by two methods. In experiments, the precision plot and the success plot are used to evaluate the trackers, (i) Precision Plot, which is measured by calculating the distance between the target frame and the center of the tracking frame. The accuracy of target tracking is evaluated by the size of the Euclidean distance. The smaller the Euclidean distance, the higher the accuracy. (ii) Success Plot, the success rate of target tracking is evaluated by tracking the overlap rate of the bounding boxes. Calculate the number of successful frames with an overlap rate greater than a given threshold in the experiment. The overlap rate is expressed as $OS = \frac{|R_a \cap R_b|}{|R_a \cup R_b|}$. In the formula, the bounding box is obtained by the target tracking algorithm and the real bounding box is given by the dataset. When the OS of a frame is larger than the set threshold, the target can be considered correct. The general threshold is set to 0.5.

To make a comprehensive comparison of tracking algorithms, we use three standards to acquire an accuracy plot and a success plot: OPE, TRE and SRE. The OPE initializes the first frame with the position of the target in the ground-truth, and then runs the tracking algorithm to get the average accuracy and success rate. However, it does not assess the impact of initialization on tracker performance. In order to analyze a tracker's temporal and spatial robustness, the TRE and the SRE are introduced. The image sequence is scrambled in time (temporally, starting from different frames) and spatially (spatially, different bounding boxes), and then the algorithm is fully evaluated. We report the results in **Fig. 5**.

3.4 Analysis of Experimental Results

In order to prove the algorithm FMCS is effective, we carry out three experiments including different versions evaluation, overall performance evaluation and qualitative evaluation. Different versions evaluation is to measure whether each version algorithm can promote the performance of baseline algorithm. Overall performance evaluation is to compare the

performance of our method with others by exploiting OPE, TRE, SRE. Qualitative evaluation is to performance of our algorithm with other state-of-art trackers on different challenging sequences.

3.4.1 Different Versions Evaluation

In order to show the effect of the proposed algorithm, we implement different versions of FMCS on OTB2015. We denote FMCS without surface fitting as FMCS-NS, without confidence evaluation strategy as FMCS-NE and with neither of these two as FMCS-N2. The tracking results are summarized in **Table 3**.

Table 3. The tracking results of different versions of FMCS

Trackers	Confidence evaluation	Surface fitting	Success	Precision
FMCS	Yes	Yes	0.585	0.783
FMCS-NS	Yes	No	0.561	0.781
FMCS-NE	No	Yes	0.579	0.776
FMCS-N2	No	No	0.553	0.773

As can be seen in **Table 3**, the algorithm FMCS wo proposed shows the best tracking accuracy and robustness while FMCS-NS performs second and FMCS-NE performs third. Without the confidence evaluation strategy, FMCS-NE performs poorly because the algorithm can not find the best position of the object because of the noise effects. Without the surface fitting, FMCS-NS gets worse performance due to the integer-level parameters can only get a rough position, which cannot meet the high-precision positioning of the image for detection. Without both of these two strategies, FMCS-N2 is the worst performer in all evaluation metrics. Although the proposed confidence evaluation strategy and surface fitting strategy increase the detection time, compared with the FMCS-N2, FMCS improves the tracking performance observably according to the experimental results. As shown in **Fig. 4**, we evaluate the performance of the proposed algorithm and the precision score and success score of FMCS obtains the most rewarding performance, indicating that the additional strategies are effective.

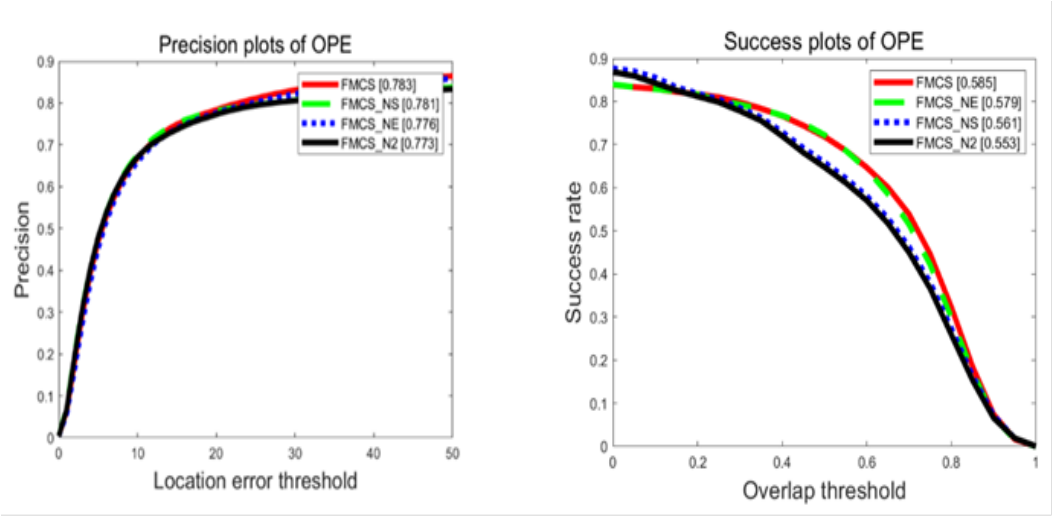

Fig. 4. Precision and Success plots of different versions of FMCS

3.4.2 Overall Performance Evaluation

We evaluate overall performance of the algorithm FMCS on the benchmark OTB100 with comparisons to six state-of-the-art trackers, including KCF [11], SAMF [13] ,DSST [14], LMCF [26] , Staple [32] and LCT [33]. Among them, DSST and SAMF can deal with the scale changes. LMCF and LCT have their own confidence evaluation and re-detection strategies. KCF, DSST, SAMF, Staple, LMCF, LCT are developed based on correlation filter trackers.

Fig. 5 is a comprehensive evaluation of all tracking results of the algorithm on the OTB100. From it we can see that our tracker achieves a promising performance. Compared to the baseline algorithm KCF, FMCS tracker achieves an obvious improvement (8.6% in precision rate and 10.8% in success rate). Additionally, compared with DSST and SAMF, our method not only deals with the scale changes but also reaches accurate rotation angle. Therefore, our FMCS method gain the excellent performance both in the precision rate and success rate, which demonstrates that our confidence evaluation strategy and surface fitting strategy are effective.

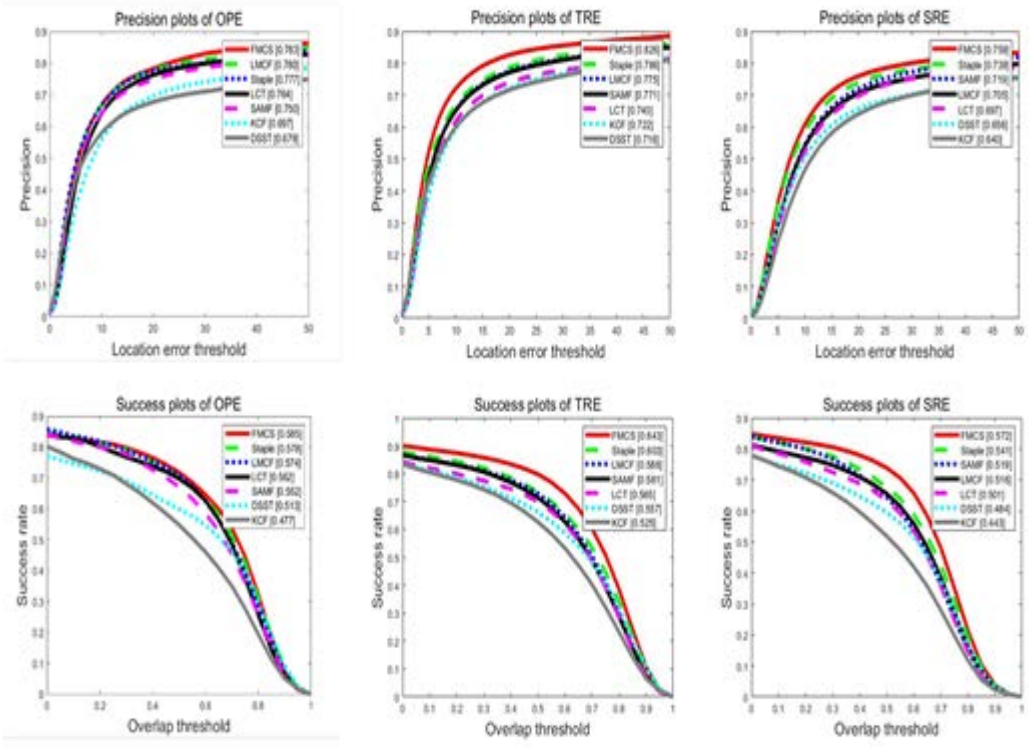

Fig. 5. Precision and Success plots over all 100 sequences in the OTB100. The evaluated trackers are Staple, LMCF, LCT, SAMF, KCF, DSST.

The videos in OTB100 are annotated with eleven different attributes to describe the various challenges in the object tracking problem. Results on these attributes can be used to evaluate the advantages and disadvantages of trackers in various aspects. **Table 4** and **Table 5** show the comparison of FMCS with other top six tracking algorithms on these eleven attributes.

Table 4. Precision scores of FMCS and other six state-of-the-art on eleven attributes. Best: bold

Attribute	FMCS	Staple	LMCF	LCT	SAMF	DSST	KCF
Fast motion(39)	**75.5**	69.7	73.0	68.1	65.4	55.2	62.1
Background clutter(32)	77.6	77.3	**82.8**	74.3	69.9	71.3	72.2
Motion blur(29)	**76.3**	70.7	73.0	69.9	65.5	56.7	60.1
Deformation(44)	68.7	**73.5**	71.7	68.7	67.2	52.7	61.7
Illumination variation(38)	74.3	79.3	**80.1**	75.0	71.6	72.3	71.6
In-plan rotation(52)	**78.6**	76.2	74.5	78.1	71.4	68.3	69.8
Low resolution(9)	58.1	63.1	67.9	53.7	**68.5**	56.7	56.0
Occlusion(48)	**74.3**	72.1	73.6	67.9	72.2	58.9	63.2
Out of view(14)	58.2	66.1	**69.3**	59.2	62.8	48.1	50.1
Scale variation(64)	**74.2**	71.0	71.5	67.9	69.6	62.8	63.3
Out-of-plane rotation(63)	**78.3**	73.0	76.0	74.6	73.9	64.4	67.7

Table 5. Success scores of FMCS and other six state-of-the-art on eleven attributes. Best: bold

Attribute	FMCS	Staple	LMCF	LCT	SAMF	DSST	KCF
Fast motion(39)	**57.7**	53.7	55.1	53.4	50.7	44.7	45.9
Background clutter(32)	57.2	58.3	**60.7**	55.5	53.1	53.3	50.1
Motion blur(29)	56.4	54.6	56.1	53.3	52.5	46.9	45.9
Deformation(44)	48.6	54.5	51.7	49.4	**49.9**	41.0	43.5
Illumination variation(38)	**57.0**	60.3	60.3	**57.0**	53.5	56.4	48.6
In-plan rotation(52)	**57.1**	54.6	53.5	55.4	51.3	49.7	46.7
Low resolution(9)	**46.6**	41.8	45.0	35.4	43.0	38.3	30.7
Occlusion(48)	**55.4**	54.5	54.1	50.6	53.8	44.9	44.5
Out of view(14)	**49.9**	48.1	53.9	45.2	48.0	38.6	39.3
Scale variation(64)	**54.2**	51.4	51.3	48.5	48.8	46.2	39.4
Out-of-plane rotation(63)	**49.9**	48.1	53.9	45.2	48.0	38.6	39.3

Considering precision rate and success rate, the FMCS achieves the best performance in most of the attributes. Thanks to confidence evaluation strategy, our FMCS algorithm can track the objects under Occlusion. When the objects undergo Occlusion, the performance of our tracker has been improved by 1.3% as compared with the LMCF. With the accurate scale and rotation estimation mechanism, our FMCS tracker performs well when the object is under Scale variation, Out-of-plane rotation and In-plan rotation. Besides, attributing the scale and rotation estimation mechanism, our FMCS method can effectively deal with the Scale variation, In-plan rotation and Out-of-plan rotation challenges.

3.4.3 Qualitative Evaluation

In order to acquire an intuitive comparison, we evaluate the tracking algorithms on eight sequences that have serious scale and rotation changes. **Fig. 6** demonstrates the results generated by seven state-of-the-art trackers (LCT, DSST, Staple, LMCF, KCF and SAMF).

Fig. 6. Tracking results on several challenging sequences. From left to right and top to down are Skiing, Toy, Dragonbaby, Jogging-2, Freeman3, Freeman4, Shaking, Skater.

For *Toy* sequence, the task is to track a moving object in the hand with scale variation and fast motion, and our FMCS and Staple methods acquire relatively high precision. For *Dragonbaby* sequence, DSST and KCF algorithms tend to drift at frame 32 owing to the fast motion and rotation and gradually other algorithms start to keep up with the target's movement. Finally, all trackers except FMCS fail to locate the object at frame 96, and only our FMCS can successfully accomplish the whole tracking task. As shown in *Jogging-2* sequence, the girl is completely occluded by the pillar at frame 45, and some trackers lose the target. However, when the object reappears in the scene, only FMCS and LMCF can recover from drifting. Although our algorithm uses the APCE idea mentioned in the LMCF, we have proposed our own confidence evaluation strategy during the target tracking process. Through the above strategy, we find the best status of the templated object and record the most reliable model to deal with the challenging problems.

From the *Freeman3* and *Freeman4* sequences, it can be observed that our FMCS algorithm is able to handle scale and rotation changes. Despite DSST and SAMF are devoted to handle scale variation, they all have a poor performance owing to the limited range of scale space. In the *Skiing*, *Shaking* and *Skater* sequences, many trackers are less discriminative of the target that have evident camera motion. In contrast, our FMCS approach adaptively fuses the response confidence of two independent models, which strengthens the robustness of our tracker and increases the tracking precision.

4. Conclusion

In this paper, we proposed a novel effective algorithm FMCS to estimate the scale and rotation changes for robust object tracking. Furthermore, a confidence evaluation strategy is proposed for preventing model drift and in order to improve the precision of tracking results, we fit the second-order polynomial to the best matching neighborhood. The maximum value of each polynomial is the best value of the sub-pixel accuracy of the object position. We compare four different versions of FMCS and compare it with other 6 mainstream algorithms in the OTB100. The results show that our method using OPE performs well with precision plot of 78.3% and success plot of 58.5%, which achieves outstanding performance and verifies the effectiveness of the proposed algorithm in dealing with scale and rotation estimation. Our method mainly focuses on an effective evaluation of the scale and rotation transformation of the tracking target. On the other hand, in spite of the advantages of rotating the bounding box, it is computationally very laborious to estimate it. To attain this, in the future, we will pay more attention to get better rotation angle with increased computational efficiency.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (U1631133).

References

[1] W. A. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 7, pp. 1442-1468, 2014. Article (CrossRef Link)

[2] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust l1 tracker using accelerated proximal gradient approach," in *Proc. of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1830-1837, 2012. Article (CrossRef Link)

[3] X. Jia, H. Lu, and M. H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proc. of IEEE Conference on computer vision and pattern recognition*, pp. 1822-1829, 2012. Article (CrossRef Link)

[4] W. Zhong, W. Lu, and M. H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proc. of IEEE Conference on Computer vision and pattern recognition*, pp. 1838-1845, 2012. Article (CrossRef Link)

[5] C. Ma, J. B. Huang, X. Yang, and M. H. Yang, "Adaptive correlation filters with long-term and short-term memory for object tracking," in *Proc. of International Journal of Computer Vision*, pp. 771-796, 2018. Article (CrossRef Link)

[6] T. Zhang, S. Liu, N. Ahuja, M. H. Yang, and B. Ghanem, "Robust visual tracking via consistent low-rank sparse learning," in *Proc. of International Journal of Computer Vision*, pp. 171-190, 2017. Article (CrossRef Link)

[7] W. Zuo, X. Wu, L. Lin, L. Zhang, and M. H. Yang, "Learning support correlation filters for visual tracking," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1158-1172, 2018. Article (CrossRef Link)

[8] N. Wang, J. Shi, D. Y. Yeung, and J. Jia, "Understanding and diagnosing visual tracking systems," in *Proc. of IEEE International Conference on Computer Vision*, pp. 3101-3109, 2015. Article (CrossRef Link)

[9] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2544-2550, 2010. Aritcle (CrossRef Link)

[10] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. of European conference on computer vision*, pp. 702-715, 2012. Article (CrossRef Link)

[11] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE transactions on pattern analysis and machine intelligence* vol. 37, no. 5, pp.583-596, 2015. Article (CrossRef Link)
[12] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M. H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Proc. of European Conference on Computer Vision*, pp. 127-141, 2014. Article (CrossRef Link)
[13] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. of European Conference on Computer Vision*, pp. 254-265, 2014. Article (CrossRef Link)
[14] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. of British Machine Vision Conference*, 2014. Article (CrossRef Link)
[15] L. Bertinetto, J. Valmadre, and J. F. Henriques, "Fully-convolutional siamese networks for object tracking," in *Proc. of European Conference on Computer Vision*, pp. 850-865, 2016. Article (CrossRef Link)
[16] M. Danelljan, C. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Transactions Pattern Analysis Machine Intelligence*, vol. 39, pp. 1561-1575, 2016. Article (CrossRef Link)
[17] S. Hong, T. You, S. Kwak, and B. Han, "Online tracking by learning discriminative saliency map with convolutional neural network," in *Proc. of 2015 International conference on machine learning* pp. 597-606. Article (CrossRef Link)
[18] C. Ma, J. B. Huang, X. Yang, and M. H. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. of the IEEE International Conference on Computer Vision*, pp. 3074-3082, 2015. Article (CrossRef Link)
[19] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. of IEEE International Conference on Computer Vision*, pp. 4310-4318. Article (CrossRef Link)
[20] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. of IEEE International Conference on Computer Vision Workshop*, pp. 621-629, 2015. Article (CrossRef Link)
[21] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6931-6939, 2017. Article (CrossRef Link)
[22] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4293-4302, 2016. Article (CrossRef Link)
[23] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. of European Conference on Computer Vision*, pp. 850-865, 2016. Article (CrossRef Link)
[24] B. X. Chen and J. K. Tstoso, "Fast visual object tracking with rotated bounding boxed," *arXiv:1907.03892*, 2019. Article (CrossRef Link)
[25] X. Yin, G. Liu, and X. Ma, "Fast Scale Estimation Method in Object Tracking," *IEEE Access*, vol. 8, pp. 31057-31068, 2020. Article (CrossRef Link)
[26] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4021-4029, 2017. Article (CrossRef Link)
[27] W. Wang, C. Liu, B. Xu, L. Li, W. Chen, and Y. Tian, "Robust Visual Tracking Based on Fusional Multi-Correlation-Filters with a High-Confidence Judgement Mechanism," *Applied Sciences*, vol. 10, no. 6, 2020. Article (CrossRef Link)
[28] M. Y. Abbass, K. C. Kwon, N. Kim, S. A. Abdelwahab, F. E. Abe El-Samie, and A. A. M. Khalaf, "Efficient object tracking using hierarchical convolutional features model and correlation filters," *The Visual Computer*, pp. 1-12, 2020. Article (CrossRef Link)
[29] H. Wang, J. Zhao, J. Song, Z. Pan, and X. Jiang, "A New Rapid-Precision Position Measurement Method for a Linear Motor Mover Based on a 1-D EPCA," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7485-7494, 2018. Article (CrossRef Link)
[30] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, pp. 1266-1271, 1996. Article (CrossRef Link)

[31] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834-1848, 2015. Article (CrossRef Link)
[32] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1401-1409, 2016. Article (CrossRef Link)
[33] C. Ma, X. Yang, C. Zhang, and M. H. Yang, "Long-term correlation tracking," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5388-5396, 2015. Article (CrossRef Link)

Yuhao Wang is a master student in Taiyuan University of Technology, whose main research direction is computer vision target recognition and tracking.

Jun Ma received the B.Sc., M.Sc., and Ph.D. degrees from Taiyuan University of Technology, China. She is involved in new sensor technology, control science and engineering.