EasyChair Preprint

№ 8275

Multi-Frame Track-Before-Detect for Scalable Extended Target Tracking

Desheng Zhang, Wujun Li, Shixing Yang, Yingshun Wang, Chuan Zhu and Wei Yi

June 16, 2022

Multi-Frame Track-Before-Detect for Scalable Extended Target Tracking

Desheng Zhang, Wujun Li, Shixing Yang, Yingshun Wang, Chuan Zhu, Wei Yi*
School of Information and Communication Engineering
University of Electronic Science and Technology of China
Email:{uestc2016zhang, kussoyi}@gmail.com

***Abstract*—This paper mainly addresses the scalable detection and tracking of the extended target in the low signal-to-noise(SNR) environment. As the appearance and shape of the extended target are constantly varied, it is challenging to achieve robust detection and tracking. For this, a novel adaptive scale (AS) kernelized correlation filter (KCF) based on multi-frame track-before-detect (MF-TBD) framework is proposed. By embedding scaling pools into the response map to handle the scale variation and accumulating target energy overall feasible trajectories, AS-MF-TBD estimates the kinematic state and geometric shapes simultaneously. Both simulation data and real radar data are used to demonstrate the superiority of the proposed method in terms of detection performance and estimation accuracy.**

***Index Terms*—Extended target tracking; Multi-frame detect; Track-before-detect; Adaptive scale; Kernelized correlation filter; FMCW radar.**

I. Introduction

Traditional target tracking bases its assumption on the point target model. Previously, the low-resolution sensors observed targets occupying at most one resolution cell. With the development of sensor technology in recent years, it has become increasingly common to observe targets occupying multiple resolution cells. Furthermore, the extended target tracking problem is becoming increasingly important in maritime surveillance and autonomous driving applications. Specific techniques, extended target tracking (ETT) algorithms have been proposed [1]. Compared with point target tracking, ETT needs to estimate kinematic state and geometric shape simultaneously.

An essential aspect of ETT is the modeling of geometric shapes. The shapes indicate how the measurement characteristics originating from the extended target are spatially distributed around its centroid. One of the most common model is the random matrix model (RMM), which represents the geometric shapes as symmetric positive defined matrix [2], [3]. Another model is the random hypersurface model (RHM), which assumes the measurements locate the randomly scaled curves of the target contour [4], [5]. Compared with RHM, RMM is simple and less computationally burdensome, making it more practical and promising. For convenience, this paper builds upon our idea to use elliptic RMM to model the shape of the extended target.

This work was supported in part by the National Natural Science Foundation of China under Grants 61771110, U19B2017 and 61871103, in part by the Fundamental Research Funds of Central Universities under Grant ZYGX2020ZB029.

The classical ETT methods mainly contain two stages, including threshold detection and target tracking. It is common to extract the plot measurement by thresholding raw sensor data at the detection stage. Then, the detected information with sufficient intensity is passed to the following tracking stages. Next, the tracking methods estimate the kinematic states and shapes by predicting and updating poster probability density (PDF). Similar to Kalman Filter (KF) [6], joint probabilistic density function (JPDA) [7], multi-hypothesis tracking (MHT) [8] and probability hypothesis density (PHD) [9], have been generalized from point target tracking to ETT. The estimation accuracy of these methods relies on the front detection results. Unfortunately, they will fail when the target SNR is low because the information contained about the target may be discarded after thresholding detection.

Another effective method for detecting and tracking dim targets is multi-frame track-before-detect (MF-TBD). It has been widely applied in different scenarios during the last decades, such as infrared and optical [10], [11] scenarios, underwater sonar [12], [13] and radar system [14], [15]. Unlike the classic detection and tracking methods, the MF-TBD method directly processes multi-frame measurements without threshold detection. It can achieve superior detection and tracking performance in the scenario where the target SNR is low. Typical implementations of MF-TBD include the 3-dimensional matched filtering [16], dynamic programming [17]–[20], Hough transform [21] and particle filter [22]. However, most of them focus on point target model, and there are few methods for extended target model. Recently, an MF-TBD method based on pseudo spectrum proposed in [23] accumulates the intra-frame energy by point spreading function (PSF) requiring known the energy diffusion regions. Nevertheless, the PSF is still too simple to reflect the changes in target characteristics. It will suffer poor detection and tracking performance when extended targets' appearances and geometric shapes are time-varying and unknown. Another MF-TBD method proposed in [24], requires a fixed geometric shape and straight-line motions. Meanwhile, it faces the heavy computation burden. Furthermore, all the above methods based on the MF-TBD framework are not effectively detecting and tracking scalable extended targets.

This paper focuses on detecting and tracking the scalable extended targets in low SNR scenarios based on multi-frame TBD framework. For this, we propose a novel AS-MF-TBD algorithm that could estimate the geometric shapes and improve tracking accuracy. At first, we analyze the general framework of MF-TBD based on the maximum a posteriori (MAP) criterion. Then, we employ an adaptive scale searching strategy based on the traditional KCF to solve appearance varied. Next, record the similarity score between reference template and candidate templates after KCF pre-processing, and pass it as the test statistic into the MF-TBD framework. Lastly, both simulated data and real radar data are used to demonstrate the effectiveness of the proposed algorithm.

II. Models and Notations

A. Dynamic Model

Consider an extended target that moves in the 2-dimensional x-y surveillance region with nearly constant velocity (CV) model. As in [2], it is common to model its physical shape with elliptic random matrices. Let $\mathbf{x}_k = [p_{x,k}, p_{y,k}, \dot{p}_{x,k}, \dot{p}_{y,k}, l_{x,k}, l_{y,k}]^\top \in \mathbb{R}^6$ denotes the dynamic states of extended target at the k-th frame, where $^\top$ represents the matrix transpose, the $(p_{x,k}, p_{y,k})$ and $(\dot{p}_{x,k}, \dot{p}_{y,k})$ are the locations and velocities coordinates of centroid, $l_{x,k}$ and $l_{y,k}$ are the semi-major axis and semi-minor axis of the ellipse, respectively.

With the constraints of CV model, the dynamic motion of target follows the first-order Markov process given by

$$\mathbf{x}_k \left| \mathbf{x}_{k-1} \right. \sim p\left(\mathbf{x}_k \mid \mathbf{x}_{k-1}\right) = \mathcal{N}\left(\mathbf{x}_k; \mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}_{k-1}\right), \quad (1)$$

where $p(\cdot)$ represents the probability density function (PDF) of the random event $\mathbf{x}_k \mid \mathbf{x}_{k-1}$, the $\mathcal{N}(\mathbf{x}; \mu, \mathbf{\Sigma})$ denotes the Gaussian PDF evaluated at $\mathbf{x}$ with mean μ and covariance $\mathbf{\Sigma}$, the term $\mathbf{Q}_{k-1}$ is the covariance matrix for the process noise, and the state transition matrix $\mathbf{F}$ is given by

$$\mathbf{F} = \begin{bmatrix} \mathbf{E}_{2\times2} & \mathbf{A}_1 & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \mathbf{E}_{2\times2} & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{A}_2 \end{bmatrix}, \quad (2)$$

$$\mathbf{A}_1 = \begin{bmatrix} T_s & 0 \\ 0 & T_s \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} \varepsilon_x & 0 \\ 0 & \varepsilon_y \end{bmatrix}, \quad (3)$$

where $\mathbf{E}_{2\times2}$ denotes the second-order unit matrix, T_s indicates the observation interval between consecutive frames, the ε_x and ε_y are scale ratios in the x-direction and y-direction, which are used to model the changes of geometric shapes over time. After that, the dynamic state evolution of extended target during K scans can be expressed as

$$\mathbf{X}_{1:K} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K\}. \quad (4)$$

B. Measurement Model

Assume that the surveillance region is divided into $N_x \times N_y$ pixels cells based on the sensor resolution, where N_x and N_y denote the number of cells in x-direction and y-direction, respectively. As in [23], [25], [26], the statistical characteristics and energy diffusion of extended regions can be modeled by PSF with known the centroid of extended target (p_k^x, p_k^y). Taking advantage of this, the measurement $z_k^{x,y}$ recorded in the (x, y)-th cell at the k-th frame can be expressed as

$$z_k^{x,y} = \begin{cases} \left| A_k \exp\left(j\varphi_k\right) h_k^{x,y}\left(\mathbf{x}_k\right) + \omega_k^{x,y} \right|, & \text{if } (x,y) \in \xi_k \\ \left| \omega_k^{x,y} \right|, & \text{others} \end{cases}, \quad (5)$$

where $|\cdot|$ denotes the modulo operation, $\omega_k^{x,y}$ denotes the additive background noise, A_k denotes the amplitude for the centroid of extended target during the k-th frame, φ_k is the phase during the k-th frame, which is a random variable uniformly distributed in the interval $[0, 2\pi)$, the PSF $h_k^{x,y}\left(\mathbf{x}_k\right)$ takes both the locations of centroid and extended size into consideration, the two-dimension Gaussian PSF is assumed here given as

$$h_k^{x,y}\left(\mathbf{x}_k\right) \propto \exp\left(-\frac{\left(x - p_{x,k}\right)^2}{\sigma_x^2} - \frac{\left(x - p_{y,k}\right)^2}{\sigma_y^2} \right), \quad (6)$$

It is worth noting that $h_k^{x,y}\left(\mathbf{x}_k\right)$ has the peak value at the center $(p_{x,k}, p_{y,k})$ and gradually decrease both in x and y directions with the variance of σ_x^2 and σ_y^2, the ξ_k is the set of scattered points (x, y) originating from the extended target given as

$$\xi_k = \left\{ (x,y) \left| \left\| \mathbf{R}\left(\beta_k\right) \begin{bmatrix} \frac{2}{l_{x,k}} & 0 \\ 0 & \frac{2}{l_{y,k}} \end{bmatrix} \begin{bmatrix} p_{x,k} - x \\ p_{y,k} - y \end{bmatrix} \right\| \le 1 \right. \right\}, \quad (7)$$

where the $\|\cdot\|$ denotes matrix norm, the terms $2/l_{x,k}$ and $2/l_{y,k}$ are eigenvalues related to the extended size, the rotation matrix $\mathbf{R}\left(\beta_k\right)$, and the β_k denotes orientation at the k-th frame, making the major axis parallel to velocities,

$$\mathbf{R}\left(\beta_k\right) = \begin{bmatrix} \cos\beta_k & \sin\beta_k \\ -\sin\beta_k & \cos\beta_k \end{bmatrix}, \quad (8)$$

$$\beta_k = \arctan\left(\dot{p}_{y,k} / \dot{p}_{x,k}\right). \quad (9)$$

After that, the whole pixel measurement at the k-th are integrated into the set $\mathbf{z}_k = \left[z_k^{1,1}, \ldots, z_k^{x,y}, \ldots, z_k^{N_x,N_y}\right]$. Furthermore, the measurements of extended target during K scans can be expressed as

$$\mathbf{Z}_{1:K} = \{\mathbf{z}_1, \ldots, \mathbf{z}_K\}. \quad (10)$$

III. Problem Formulation

The traditional MF-TBD methods integrate the target energy along feasible trajectories during multiple consecutive frames, then declare detection results and extract trajectories when the accumulated energy exceeds the given threshold. The iterative MF-TBD framework can be derived as [27]

$$I\left(\mathbf{x}_k \mid \mathbf{Z}_{1:K}\right) = \max_{\mathbf{x}_{k-1} \in \tau(\mathbf{x}_k)} I\left(\mathbf{x}_{k-1} \mid \mathbf{Z}_{1:K}\right) + s\left(\mathbf{z}_k \mid \mathbf{x}_k\right), \quad (11)$$

where $\tau\left(\mathbf{x}_k\right)$ is used to store the state transfer relationships between frames, $I\left(\mathbf{x}_k \mid \mathbf{Z}_{1:K}\right)$ denotes the accumulated merit function, and $s\left(\mathbf{z}_k \mid \mathbf{x}_k\right)$ denotes the test statistic.

$$\psi\left(\mathbf{x}_k\right) = \arg \max_{\mathbf{x}_{k-1} \in \tau(\mathbf{x}_k)} I\left(\mathbf{x}_{k-1} \mid \mathbf{Z}_{1:K}\right), \quad (12)$$

Once the target is detected, the final trajectory, $\hat{\mathbf{X}}_{1:K} = [\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K]$ can be extracted by backtracking state transfer $\psi(\mathbf{x}_k)$ $(k = K-1, \ldots, 1)$ given as

$$\hat{\mathbf{x}}_k = \psi(\hat{\mathbf{x}}_{k+1}). \tag{13}$$

Most of the existing MF-TBD methods focus on point target tracking, which employs the log-likelihood ratio (LLR) or amplitude as the test statistic. Such methods achieve superior detection and tracking performance in the low SNR environments. Recently, MF-TBD methods have been generalized for ETT with known and fixed characteristics. The energy can be efficiently accumulated in the known regions with the help of PSF. However, these methods may suffer from performance degradation with unknown and scalable geometric shapes. This is because, MF-TBD methods require accurate dynamic model constrained or a prior information while accumulating energy. Once the characteristics of targets mismatch with the actual model, MF-TBD methods work inefficiently. In addition, it is also challenging to obtain the time-varing kinematic and scaling characteristics accurately. If we design such a test statistic that can significantly discriminate the differences between target and background clutter, and still identify different geometric shapes, the problem for ETT can provide an effective solution based on the MF-TBD framework in low SNR environment.

IV. Development of The Proposed Algorithm

A. Problem Description

As mentioned above, it is challenging to accurately track the scalable extended targets based on the existing MF-TBD methods. Fortunately, there are many classical and efficient methods to handle the changes of appearance features in the visual tracking. The kernelized correlation filter (KCF) has been introduced in many applications with real-time requirements [28]–[30]. In fact, visual tracking is relatively similar to the target tracking in radar. To deal with the scalable extended target tracking, we intend to incorporate KCF into the MF-TBD framework.

B. The Brief Review of KCF

KCF works in the Tracking-by-Detection framework. In the initial frame, given the reference template $\mathbf{s}_0$ (bounding rectangle $[p_{x,0}, p_{y,0}, l_{x,0}, l_{y,0}]$, including center $(p_{x,0}, p_{y,0})$ in x-y coordinate, width $l_{x,0}$, and height $l_{y,0}$) of the extended target, positive samples close to the given center and negative samples beyond a close distance are generated as the position set $\{\mathbf{s}_i\}_{i=1}^m$. Here $m \in \mathbb{N}$ denotes the total number of samples. In our work, we use the raw intensities in the rectangle region of radar image as feature, therefore $\mathbf{u}_i \in \mathbb{R}^d, d = l_{x,0} \times l_{y,0}$. Based on the distance of the sampled position $\mathbf{s}_i$ to $\mathbf{s}_0$, the label value y_i is associated with the corresponding feature $\mathbf{u}_i$. Given m feature-label pairs $\{\mathbf{u}_i, y_i\}_{i=1}^m$, KCF learn the discriminative classifier in high-dimensional kernel space quickly via Fast Fourier Transform (FFT), i.e. $y_i = f(\mathbf{u}_i)$. In the latter frame, the classification function $f(\cdot)$ is used to test the feature $\{\mathbf{v}_i\}_{i=1}^m$ from the candidate positions $\{\mathbf{s}_i^-\}_{i=1}^m$, which are sampled around the previous position $\hat{\mathbf{s}}_i^-$. The location $\hat{\mathbf{s}}_i^-$, where the tested featured has the maximum response, is chosen as the estimated position of the target in that frame [30], [31].

Generally, KCF loops in training, testing and update phases. In the training phase, KCF learns the discriminative classifier $f(\cdot)$, requiring to minimize the squared error between input features $\mathbf{u_i}$ and label y_i. In addition, the kernel trick maps the input samples $\mathbf{u}_i = \{u_i\}_{i=1}^m$ into the high dimensional space $\zeta(\mathbf{u}_i)$ to enhance the discriminative capability of the classifier.

If view the $\zeta(\mathbf{u}_i)$ as the new feature space and define $\mathbf{w}$ as follows:

$$\mathbf{w} = \sum_{i=1}^{m} a_i \zeta(\mathbf{u}_i), \tag{14}$$

Based on the Regularized Least-Squares criteria, the learning task of KCF is given as

$$\min_{\mathbf{w}} \sum_{i=1}^{m} (f(\mathbf{u}_i) - y_i)^2 + \lambda \|\mathbf{w}\|^2, \tag{15}$$

where λ is regularization coefficient to avoid over-fitting.

Substituting (14) into (15) and utilizing FFT technique, the coefficient can be obtained as

$$\boldsymbol{\alpha} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{Y})}{\mathcal{F}(\mathbf{K}_{uu}) + \lambda}\right), \tag{16}$$

where the symbol $\mathcal{F}^{-1}$ and $\mathcal{F}$ denote FFT and the inverse FFT respectively, $\mathbf{K}_{uu}$ denotes the kernel matrix, the subscript uu indicates that the kernel function uses the input features $\mathbf{u}$ and its own circular shifting samples as input.

In the testing phase, circular sampling for new input features $\mathbf{v}_i = \{v_i\}_{i=1}^m$, obtaining the response map of the region of interested. KCF locates the peak of response map $\mathbf{Y}$ as the estimated positions.

$$\mathbf{Y} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{K}_{uv})\mathcal{F}(\boldsymbol{\alpha})), \tag{17}$$

where the definition of $\mathbf{K}_{uv}$ is similar to $\mathbf{K}_{uu}$.

In the update phase, KCF uses a simple updating strategy based on the linear interpolation on both the referred feature $\mathbf{u}$ and the classifier's key parameter $\boldsymbol{\alpha}$. In each time step k, once the candidate $\mathbf{v_k}$ is chosen as the target, the referred feature $\mathbf{u}_k$ at time k is updated by the combination of the new feature $\mathbf{v}_k$ and the previous feature $\mathbf{u}_{k-1}$ with the learning rate η. Similarly, the parameter $\boldsymbol{\alpha}$ is updated in this way.

$$\mathbf{u}_k = \eta \times \mathbf{v}_k + (1-\eta) \times \mathbf{u}_{k-1} \tag{18}$$

$$\boldsymbol{\alpha}_k = \eta \times \boldsymbol{\alpha}_k + (1-\eta) \times \boldsymbol{\alpha}_{k-1}. \tag{19}$$

Remark 1. *The traditional KCF methods employ the fixed scale template for online tracking, the performance degrades while the target suffers from self-rotation and scale variations.*

Remark 2. *The changes of appearance result in energy diffusion in the response map $\mathbf{Y}$, which is no more needle-shaped Gaussian label. In fact, it is not always achievable to locate the target by searching for peaks in the single frame.*

C. The Proposed AS-MF-TBD Method

In order to enhance the performance of detection and tracking for scalable extended targets, we further propose an improved MF-TBD methods based on adaptive scale KCF (AS-KCF). In this subsection, we firstly predict the state transition based on kinematic constraint. Next, we obtain the similarity score based on KCF with different scaling ratios of reference template. Finally, by setting the similarity score as the test statistics, the state sequence of the extended target can be estimated accurately based on the MF-TBD framework.

1) Cell prediction based on kinematic constraint: The baseline of KCF tracking is that the target state transition within the fixed searching window inter-frame. However, its performance may degrade for the sensors with long scanning times between adjacent frames, such as radar. It is straightforward to introduce the kinematic constraints into searching the state transition for more accurate estimation. Here, we propose an effective strategy to predict the state transition region for every candidate target. Assume that the (x, y)-th pixel cell is the centroid of extended target $(p_{x,k}, p_{y,k})$ at the k-th frame. In the next frame, its corresponding candidate states of the target are predicted as

$$\Omega_{k+1} = \{p_{x,k} + T_s \times \dot{p}_{x,k} - \delta_x, p_{y,k} + T_s \times \dot{p}_{y,k} - \delta_y\}, \tag{20}$$

where Ω_{k+1} denotes the set of centroid's state transition at the $(k+1)$-th frame, δ_x and δ_y depend on the maneuvering characteristics of the extended targets in x-direction and y-direction, respectively. In the next frame, the proposed AS-MF-TBD searches for the maximum response in the predicted region Ω_{k+1}.

2) Adaptive scale searching strategy: In contrast to the traditional KCF employing the fixed sizes of reference template, we propose an adaptive scale searching strategy for tracking scalable extended targets.

Assume that there are multiple scaling ratio factors $\{b_i\}_{i=1}^{d} \in \mathbf{B}_k$ at the k-th frame, where b_i is the i-th scaling ratio factor, d is the total number of scaling, and the $\mathbf{B}_k$ denotes the scaling pool. Then, we describe how the appearance and sizes of extended target vary based on the scaling ratio factors.

As above mentioned, the scaling ratio factors $\hat{b}_i$ and state of extended target $\hat{\mathbf{x}}_k$ can be estimated simultaneously under the maximum likelihood (ML) criterion,

$$\left\{\hat{b}_i, \hat{\mathbf{x}}_k\right\} = \underset{b_i \in \mathbf{B}_k, \mathbf{x}_k \in \Omega_k}{\arg\max}\ p\left(L_k^{b_i} | \mathbf{x}_k\right), \tag{21}$$

where $p\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right)$ denotes the probability of similarity identified as extended target when the scaling ratio factor is b_i at the k-th frame. Moreover, since the kinematic information is also hidden in the $p\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right)$, if we take it into consideration for estimating and updating the scale ratio factor b_i and state $\mathbf{x}_k$, the classifier can further improve its ability to discriminate between targets and background clutters.

For convenience, we firstly define the $\lambda\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right)$ as the probability of similarity given as

$$\lambda\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right) = p\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right) p\left(\mathbf{x}_k | \mathbf{x}_{k-1}\right), \tag{22}$$

where $p\left(\mathbf{x}_k | \mathbf{x}_{k-1}\right)$ denotes the state transition PDF constrained by dynamic model. Thus, the problem in (21) can be derived as

$$\left\{\hat{b}_i, \hat{\mathbf{x}}_k\right\} = \underset{b_i \in \mathbf{B}_k, \mathbf{x}_k \in \Omega_k}{\arg\max}\ \lambda\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right) \tag{23}$$

Given the geometric shapes in the initial frame, and the geometric shapes and kinematic characteristics constantly change during movement. Thus, the problem of detecting and tracking of the scalable extended target can be regarded as jointly estimating the scaling ratio factor and kinematic state. Exactly, the $\lambda\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right)$ has different similarity probability corresponding to the different scaling ratios of the reference template $\mathbf{u}$. Only the estimated scaling ratio is the same as the actual one, the similarity probability $\lambda\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right)$ achieves its maximum.

It is worth noting that the adaptive scale KCF method achieves robust tracking of the scalable extended target at relatively high SNR. However, the performance of this method degrades dramatically in the case of low SNR, since the trained classifier $f(\cdot)$ is unable to accurately discriminate between extended target and background clutter through the single frame.

Remark 3. *The similarity probability $\lambda\left(\mathcal{L}_k^{b_i} | \mathbf{x}_k\right)$ has the same meaning as the test statistic of traditional MF-TBD methods. Moreover, it also can be seen as the more advanced test statistic that maps the raw data into the high-dimensional space through machine learning and extracts the difference between the target and background clutter.*

Therefore, we regard it as the test statistic in the MF-TBD framework for tracking scalable extended target in the low SNR environment.

3) The procedure of AS-MF-TBD: After the cell prediction and adaptive scale KCF pre-processing, it is feasible to directly accumulate the energy based on response map $\mathcal{L}_k^{b_i}$ with the different scaling ratio b_i. Then, these response maps $\left\{\mathcal{L}_k^{b_i}\right\}_{i=1}^{d}$ will be further concentrated under the ML criterion given as

$$\hat{b}_i = \max_{b_i \in \mathbf{B}_k} \mathcal{L}_k^{b_i}(\mathbf{x}_k), \tag{24}$$

where the $\hat{b}_i$ denotes the estimated scaling ratio. The procedure of concentrating different scaling ratios can be implemented in parallel. After this, for $2 \le k \le K$, the accumulated merit function with different scaling ratio can be derived as

$$I(\mathbf{x}_k | \mathcal{L}_{1:K}) = \max_{\mathbf{x}_{k-1} \in \Omega_k} \left\{ I\left(\mathbf{x}_{k-1} \middle| \mathcal{L}_k^{\hat{b}_i}\right) + \mathcal{L}_k^{\hat{b}_i}(\mathbf{x}_k) \right\}. \tag{25}$$

After multi-frame accumulation, the decision on the presence or absence of an extended target will be executed

$$\max_{\mathbf{x}_K} I(\mathbf{x}_K | \mathcal{L}_{1:K}) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{26}$$

Algorithm 1 Procedure for AS-MF-TBD Algorithm

Input: Raw measurement: $\mathbf{Z}_{1:K} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_K\}$, Detection threshold: γ, Scaling ratio pool: $\mathbf{B}_k = \{b_i\}_{i=1}^{d}$.

Output: State sequences $\mathbf{X}_{1:K} = \{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_K\}$)and scaling ratios $\hat{b}_{1:K} = \left\{\hat{b}_1, \ldots, \hat{b}_K\right\}$.

1) Integration:

for $k = 1, 2, \ldots, K$ **do**

  for $x = 1, 2, \ldots, N_x$ **do**

    for $y = 1, 2, \ldots, N_y$ **do**

      Construct the response map

        $\mathcal{L}_k(\mathbf{x}_k) = \textit{AS-KCF}(\mathbf{z}_k, \mathbf{B}_k)$

      Cell prediction

        $p_{x,k+1} = p_{x,k} + T_s \times \dot{p}_{x,k} - \delta_x$

        $p_{y,k+1} = p_{y,k} + T_s \times \dot{p}_{y,k} - \delta_y$

      Extract the scaling ratio

        $\hat{b}_i = \max_{b_i \in \mathbf{B}_k} \mathcal{L}_k^{b_i}(\mathbf{x}_k),$

      Inter-frame energy integration

        $I(\mathbf{x}_k | \mathcal{L}_{1:K}) = I\left(\mathbf{x}_{k-1} \middle| \mathcal{L}_k^{\hat{b}_i}\right) + \mathcal{L}_k^{\hat{b}_i}(\mathbf{x}_k)$

        $\psi(\mathbf{x}_k | \mathbf{x}_{k-1}) = \arg\max_{\mathbf{x}_{k-1}} I(\mathbf{x}_{k-1} | \mathcal{L}_{1:K})$

    end for

  end for

end for

2) Detection:

Judge the $I(\mathbf{x}_K | \mathcal{L}_{1:K})$ by threshold γ

$$\left\{\hat{b}_{1:K}, \hat{\mathbf{X}}_{1:K}\right\} = \arg\max_{\mathbf{x}_K} I(\mathbf{x}_K | \mathcal{L}_{1:K})$$

$$s.t. I(\mathbf{x}_K | \mathcal{L}_{1:K}) > \gamma.$$

3) Backtracking:

Search for the state $\hat{\mathbf{x}}_k, k = K-1, ..., 1,$

$$\hat{\mathbf{x}}_k = \psi(\hat{\mathbf{x}}_{k+1} | \hat{\mathbf{x}}_k).$$

where H_1 and H_0 indicate the hypothesis that the extended target exists and vice versa, the γ is the detection threshold determined by false alarm probability. If the accumulated merit function $I(\mathbf{x}_k | \mathcal{L}_{1:K})$ exceeds the threshold, backtracking is conducted. As a result, the estimated state and scaling ratio sequence of extended target are obtained as

$$\left\{\hat{b}_{1:K}, \hat{\mathbf{X}}_{1:K}\right\} = \arg\max_{\mathbf{x}_K} I(\mathbf{x}_K | \mathcal{L}_{1:K}). \tag{27}$$

Based on the above derivation, we can infer the real state sequences $\hat{\mathbf{X}}_{1:K} = \{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K\}$. The geometric shapes of the extended target need further calculation base the scaling ratio at k-th frame. The detailed procedure for the proposed AS-MF-TBD method is presented in Algorithm 1.

V. Experimental Results

In this section, the proposed AS-MF-TBD compares with log-likelihood ratio track-before-detect (LLR-TBD) [32] and Random Matrix Model for extended target tracking (RMM-ETT) [33] in simulated data and real radar data. It is worth noting that RMM-EOT uses the measurement processed by the constant false alarm rate detector (CFAR). The LLR-TBD and AS-MF-TBD uses the raw measurement data as input. LLR-TBD does not consider how to estimate the geometric shapes, we only compare its tracking and detection performance.

A. Evaluation Metrics

1) Probability of target detection (Pd): The probability that estimated positions of the centroid are within 2 resolution cells compared to the actual positions at the last frame. It is used to evaluate the detection performance.

2) Intersection-over-union (IoU): The overlap ratio between estimated regions and actual regions. It is used to evaluate the tracking performance.

$$\mathrm{IoU} = \frac{\mathbf{E} \cap \mathbf{A}}{\mathbf{E} \cup \mathbf{A}}, \tag{28}$$

where $\mathbf{E}$ and $\mathbf{A}$ denote the estimated regions and actual regions. It measures the overlap of the estimated extended target with the actual one. The larger IoU means the higher accuracy in the geometric shape of estimated extended target.

3) Root mean square error (RMSE): The average position distance difference between the valid tracks and their corresponding ground-truth trajectories at every frame. It is used to evaluate the tracking performance.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left\{(p_{x,n} - \hat{p}_{x,n})^2 + (p_{y,n} - \hat{p}_{y,n})^2\right\}}, \tag{29}$$

where N denotes the times of Monte Carlo simulations (in the simulation data) or the total number of frames captured (in the real radar data), $(\hat{p}_{x,n}, \hat{p}_{y,n})$ and $(p_{x,n}, p_{y,n})$ are the estimated and actual positions at the n-th scan, respectively.

B. Simulate Data Results

In the simulation, we assume that at most one extended target moves with CV in the 60×60 cells surveillance region. The initial dynamic state of the extended target is $\mathbf{x}_1 = [20, 20, 2.9, 2.1, 5, 4]^\top$, and the covariance matrices $\mathbf{Q}_k$ in (1) are selected by

$$\mathbf{Q}_k = \begin{bmatrix} \frac{T_s^2}{2} & 0 & 0 & 0 \\ 0 & \frac{T_s^2}{2} & 0 & 0 \\ T_s & 0 & 0 & 0 \\ 0 & T_s & 0 & 0 \\ 0 & 0 & \sigma_{l_x} & 0 \\ 0 & 0 & 0 & \sigma_{l_y} \end{bmatrix} \cdot \begin{bmatrix} \sigma_{a_x} & 0 & 0 & 0 \\ 0 & \sigma_{a_y} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{r}_k, \tag{30}$$

where $T_s = 1$ denotes the observation interval, σ_{l_x} and σ_{l_y} are the typical deviations of the major semi axis and minor semi axis respectively, σ_{a_x} and σ_{a_y} are the acceleration standard deviation along x and y directions, respectively. The term $\mathbf{r}_k$ indicates a four element random vector, whose elements are Gaussian distributed with zero mean and identity covariance matrix.

The measured intensity at each cell being Gaussian distributed as modelled in (5). The numbers of frames in a batch processing are $K = 4, 6, 8$ and the detection threshold γ for

each algorithm is chosen to guarantee a constant false alarm rate $P_{fa} = 10^{-2}$. Besides, the coefficients in (3) are set as $\varepsilon_x = \varepsilon_y = 1.08$. the regularization coefficient $\lambda = 0.01$ and the learning rate $\eta = 0.095$, the scaling pool $\mathbf{B}_k$ was set as $[0.99, 0.995, 1.0, 1.05, 1.10, 1.15]$. In the following analysis, the results are gathered by averaging over 1000 Monte Carlo realizations.

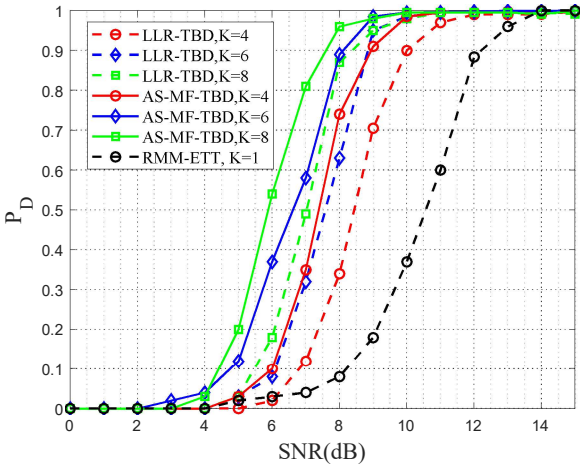

Fig. 1. The detection probability, P_D, of LLR-TBD, RMM-ETT and AS-MF-TBD at different frames.

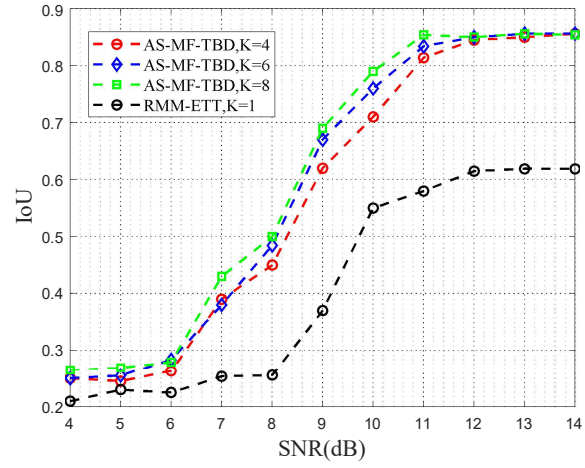

Fig. 2. The detection probability, P_D, of LLR-TBD, RMM-ETT and AS-MF-TBD at different frames.

The detection performance is shown in Fig.1, where the P_D curves of all methods against SNRs from 0dB to 14dB for comparison. As expected, the AS-MF-TBD ranks first in detection performance. In the low SNR case, the test statistic of AS-MF-TBD methods based on the response map of AS-KCF has the better capability to discriminate the target from background clutter. Despite the fact that LLR based MF-TBD takes the target energy into consideration, it requires additional prior information. In the case of changing the appearances or shape of targets constantly, the present state of targets maybe mismatched with the prior information, LLR based on MF-TBD may cause performance degradation. However, the RMM-ETT method works poorly because of the information loss after thresholding pre-processing, especially if the target SNR is low. Since LLR-TBD requires the extended region known, we compare only the RMM-ETT and proposed AS-MF-TBD here. As the number of accumulated frames increases, the better detection performance can be achieved. This is because, the non-coherent accumulation of inter frame energy enhances the characteristics of target and reduces the effect of noise.

In Fig.2, it can be seen that IoU ratio of RMM-ETT and AS-MF-TBD can approach almost the same performance while SNR is enough high. As expected, the proposed has the better extended size of estimation accuracy in low SNR.

C. Real Radar Data Results

We implemented the experiment on Texas Instrument's short range millimeter-wave radar AWR1642BOOST and ADC capture card DCA1000. The AWR1642BOOST device is an integrated single-chirp frequency modulated continuous wave (FMCW) radar sensor capable of operation in the 76-81 GHz band. The device comprises of the entire millimeter-wave radio-frequency and analog baseband signal chain for two transmitters and four receivers, as well as two customer programmable processor cores in the forms of the digital signal processor (DSP) and microcontroller (MCU).

Here, we configure it to operate in the time-division multiplexing multiple-input-multiple-output (TDM-MIMO) mode with a field of view of 120 degrees and angle resolution of 15 degrees. The starting frequency is set as 77 GHz, and bandwidth is set as 750.24 MHz. Moreover, the range of surveillance is limited to 60 m with a high resolution (0.234 m per range unit). The detailed specifications are shown in Table I.

Captured raw echo consists of 1200 consecutive frames in 48 seconds, and selects one frame at every 200 frames to reduce the computational burden. Thus, the processed raw echo by AS-MF-TBD method only consists of 60 frames. Each frame of captured echo contains four channels of complex data simultaneously, in which there are 128 repetitive chirps in each channel, and each chirp contains 256 complex samples. After 1D-FFT, 2D-FFT and 3D-FFT preprocessing, the Range-Doppler-Azimuth cube data of the moving targets are obtained.

See the experimental scenario in Fig.3, the human target moves away from 2 m to 60 m in the playground with a nearly constant velocity of 1.3 m/s.

As shown in Fig.4, the human target on the raw Range-Azimuth map occupies nearly 20 resolutions in the azimuth bins and 5 resolutions in the range bins at the 45-th frame. The geometric shape of a human target can be fitted with an ellipse. The appearances and radar cross-section (RCS) vary as the human target moved from the near zone to the far zone. It gradually drowned in the intense noise and clutter.

Next, the 60 frames of raw echo were processed by the sliding windows with $K = 6$, and the radial search ve-

Fig. 3. The experimental environment: A human target straightly moves away from AWR1642 radar with constant velocity 1.3m/s.

TABLE I
MILLIMETER-WAVE RADAR SPECIFICATIONS

Chirps Setting	
Carrier frequency	77 GHz
Bandwidth	750.24 MHz
Pulse number	128
Coherent processing number	256
Periodicity of frame	40 ms
Sensor Parameters	
Range resolution	0.234 m
Azimuth Resolution	15°
Maximum unambiguous range	50 m
Maximum angle range	±60°

locities of the human target were set as $[-3,-2,0,1,2,3]$ cells/frame, while the tangential velocities were set as $[-1,0,1]$ cells/frame. The state transition space was set as $[-2,-1,0,1,2]$ cells both in range and azimuth directions. Besides, the regularization coefficient $\lambda = 0.01$ and the learning rate $\eta = 0.075$, the scaling pool $\mathbf{B}_k$ was set as $[0.85, 0.9, 0.95, 10, 1.1, 1.2, 1.3]$.

By processing the raw echo through the methods and parameters described above, the comparison of tracking and shape estimation results are shown in Fig.5 and Fig.6. Obviously, the RMSE curves of RMM-ETT and ASCK-TBD have almost relatively small and IoU relatively high. It indicates that both have better detection and tracking performance because the SNR is high enough in Range-Azimuth in the near zone. However, the proposed AS-MF-TBD has a superior performance than RMM-ETT in the far zone. The tracking accuracy of RMM-ETT drops significantly since the strong clutter has affected the first CFAR stage. From the 40-th frame, the RMSE of the center position estimated by RRMM-ETT, which is already beyond the range threshold, indicates the RMM-ETT can't detect the target anymore.

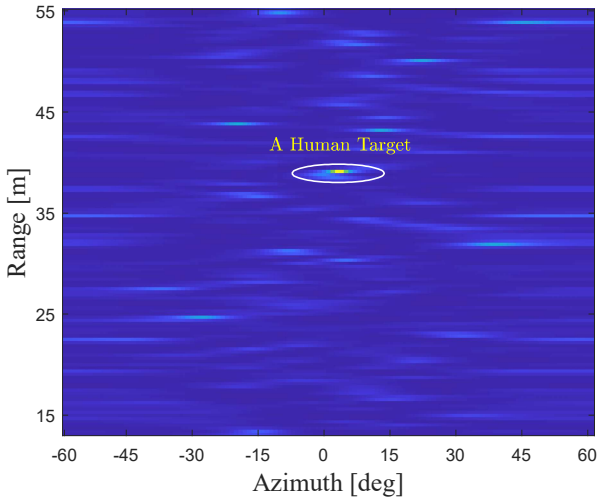

Fig. 4. The 45-th frame Range-Azimuth map in the far zones.

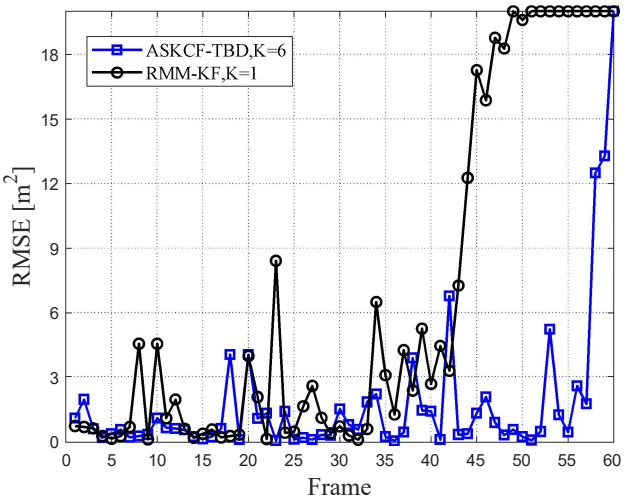

Fig. 5. The RMSE curves of RMM-ETT and AS-MF-TBD with accumulated frames $K = 6$.

VI. CONCLUSION

In our work, a novel adaptive scaled KCF is applied to track a scalable extended target based on MF-TBD framework. By setting the similarity probability of KCF with different scaling ratios as the test statistic, AS-MF-TBD helps improve the detection performance and accuracy of tracking. Compared with other methods, AS-MF-TBD shows the more vital ability to discriminate between the target and the surrounding background clutter.

REFERENCES

[1] K. Granström, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview, and applications," *Journal of Advances in Information Fusion*, vol. 12, no. 2, 2017.

[2] M. Feldmann, D. Fränken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, 2011.

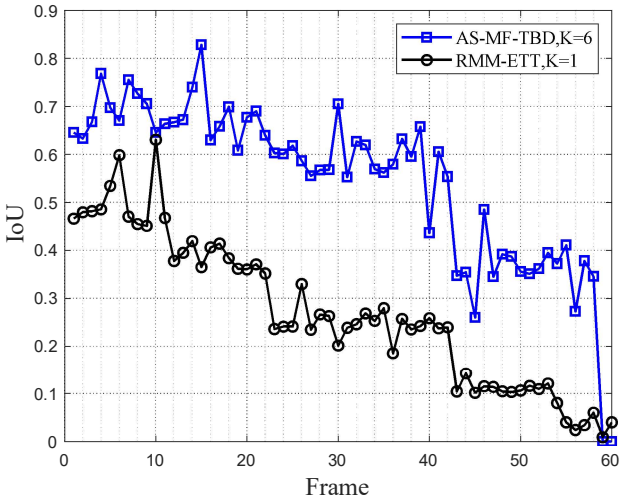

Fig. 6. The IoU curves of RMM-ETT and AS-MF-TBD with accumulated frames $K = 6$.

[3] J. Lan and X. R. Li, "Extended-object or group-target tracking using random matrix with nonlinear measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 5130–5142, 2019.

[4] M. Baum and U. D. Hanebeck, "Extended object tracking with random hypersurface models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 149–159, 2014.

[5] A. Zea, F. Faion, M. Baum, and U. D. Hanebeck, "Level-set random hypersurface models for tracking nonconvex extended objects," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2990–3007, 2016.

[6] N. Wahlström and E. Özkan, "Extended target tracking using gaussian processes," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4165–4178, 2015.

[7] Q. Li, L. Song, and Y. Zhang, "Multiple extended target tracking by truncated jpda in a clutter environment," *IET Signal Processing*, vol. 15, no. 3, pp. 207–219, 2021.

[8] D. Jiang, M. Liu, Y. Gao, and Y. Gao, "Time-matching extended target probability hypothesis density filter for multi-target tracking of high resolution radar," *Signal Processing*, vol. 157, pp. 151–160, 2019.

[9] K. Granstrom, C. Lundquist, and O. Orguner, "Extended target tracking using a gaussian-mixture phd filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3268–3286, 2012.

[10] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "A recursive moving-target-indication algorithm for optical image sequences," *IEEE transactions on aerospace and electronic systems*, vol. 26, no. 3, pp. 434–440, 1990.

[11] S. M. Tonissen and R. J. Evans, "Peformance of dynamic programming techniques for track-before-detect," *IEEE transactions on aerospace and electronic systems*, vol. 32, no. 4, pp. 1440–1451, 1996.

[12] Q. Peng, W. Li, and L. Kong, "Multi-frame track-before-detect algorithm for passive sonar system," in *2019 International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, 2019, pp. 1–6.

[13] F. Ehlers, D. Orlando, and G. Ricci, "Batch tracking algorithm for multistatic sonars," *IET Radar, Sonar & Navigation*, vol. 6, no. 8, pp. 746–752, 2012.

[14] E. Grossi, M. Lops, and L. Venturino, "A novel dynamic programming algorithm for track-before-detect in radar systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2608–2619, 2013.

[15] S. Buzzi, M. Lops, L. Venturino, and M. Ferri, "Track-before-detect procedures in a multi-target environment," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1135–1150, 2008.

[16] Y. Xiong, J.-X. Peng, M.-Y. Ding, and D.-H. Xue, "An extended track-before-detect algorithm for infrared target detection," *IEEE transactions on aerospace and electronic systems*, vol. 33, no. 3, pp. 1087–1092, 1997.

[17] E. Grossi, M. Lops, and L. Venturino, "Track-before-detect for multiframe detection with censored observations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 2032–2046, 2014.

[18] S. Buzzi, M. Lops, and L. Venturino, "Track-before-detect procedures for early detection of moving target from airborne radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 3, pp. 937–954, Jul. 2005.

[19] D. Orlando, L. Venturino, M. Lops, and G. Ricci, "Track-before-detect strategies for STAP radars," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 933–938, Feb. 2010.

[20] W. Li, W. Yi, L. Kong, and K. C. Teh, "An efficient track-before-detect for multi-prf radars with range and doppler ambiguities," *IEEE Transactions on Aerospace and Electronic Systems*, 2022.

[21] H. Im and T. Kim, "Optimization of multiframe target detection schemes," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 176–187, Jan 1999.

[22] M. Bocquel, H. Driessen, A. Bagchi, M. Bocquel, H. Driessen, and A. Bagchi, "Multitarget particle filter addressing ambiguous radar data in TBD," in *2012 IEEE Radar Conference*, May. 2012, pp. 0575–0580.

[23] G. Zhou and L. Wang, "Pseudo-spectrum based speed square filter for track-before-detect in range-doppler domain," *IEEE Transactions on Signal Processing*, vol. 67, no. 21, pp. 5596–5610, Nov. 2019.

[24] B. Yan, N. Xu, W.-B. Zhao, and L.-P. Xu, "A three-dimensional hough transform-based track-before-detect technique for detecting extended targets in strong clutter backgrounds," *Sensors*, vol. 19, no. 4, p. 881, 2019.

[25] H. Jiang, W. Yi, T. Kirubarajan, L. Kong, and X. Yang, "Multiframe radar detection of fluctuating targets using phase information," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 2, pp. 736–749, Apr. 2017.

[26] D. Zhang, W. Li, and W. Yi, "Multi-frame joint tracking and shape estimation method for weak extended targets," pp. 1–7, 2021.

[27] W. Yi, Z. Fang, W. Li, R. Hoseinnezhad, and L. Kong, "Multi-frame track-before-detect algorithm for maneuvering target tracking," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4104–4118, 2020.

[28] Y. Zhou, T. Wang, R. Hu, H. Su, Y. Liu, X. Liu, J. Suo, and H. Snoussi, "Multiple kernelized correlation filters (mkcf) for extended object tracking using x-band marine radar data," *IEEE Transactions on Signal Processing*, vol. 67, no. 14, pp. 3676–3688, 2019.

[29] 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. 3, pp. 583–596, 2014.

[30] Y. Zhou, H. Su, S. Tian, X. Liu, and J. Suo, "Multiple kernelized correlation filters based track-before-detect algorithm for tracking weak and extended target in marine radar systems," *IEEE Transactions on Aerospace and Electronic Systems*, 2022.

[31] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *European conference on computer vision*. Springer, 2014, pp. 254–265.

[32] W. Yi, H. Jiang, T. Kirubarajan, L. Kong, and X. Yang, "Track-before-detect strategies for radar detection in G0-distributed clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 5, pp. 2516–2533, Oct. 2017.

[33] M. Feldmann, D. Fränken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, 2010.