

Image Moment-Based Extended Object Tracking for Complex Motions

Gang Yao, Ryan Saltus, and Ashwin Dani

Abstract—A novel image moment-based model for shape estimation and tracking of an extended target moving with a complex trajectory is presented. The proposed extended object tracking algorithm is based on multiple noisy measurement points sampled from the target at each time step. The shape of the object, approximated by an ellipse, is estimated using a combination of image moments. Dynamic models of image moments for constant velocity and coordinated turn motions are mathematically derived. An unscented Kalman filter - interacting multiple model (UKF-IMM) method is used to track the object and estimate its shape. A likelihood function based on average log-likelihood is derived for the IMM filter. Simulation results of the proposed UKF-IMM algorithm with the image moment-based models are presented that show the estimation of the shape of the object moving in a complex trajectory. The intersection over union (IoU), and the root mean square errors (RMSEs) of the position and velocity of the centroid of the ellipse are used as metrics. The comparison results of the proposed algorithm with a benchmark algorithm from literature based on the IoU and RMSE metrics are presented.

Keywords—Extended object tracking, shape estimation, dynamic models of image moments, log-likelihood for filtering

I. INTRODUCTION

Object tracking can be defined as the problem of estimating the trajectory of a moving object based on the measurements from a particular sensor. In the context of different applications and the particular characteristics of the sensors, many tracking algorithms have been proposed [1]–[4]. Visual tracking algorithms are based on dense illuminance-based measurements, which use the silhouette, color and texture information to represent and track the target [5]–[10]. However, when measurements are sparse and only have the position information, extended object tracking (EOT) algorithms are used to solve this problem.

Extended object tracking is a very important and rapidly developing area [11]–[14], with applications in robotics and autonomous driving [15]. EOT can also be used as an additional step to online structure from motion (SfM) [16], [17], simultaneous localization and mapping (SLAM) [18], or for robotic harvesting [19], [20] problems to estimate the

shape of the object from feature points. The classical point-based tracking methods only estimate the kinematic states of the target, with the assumption that there is only one measurement point from the target at each time step. With the increased resolution, modern sensors such as phased array radars and laser range finders are capable of giving more than one point measurement from an observed target at a single time instance [15], [18], [21]. Sensors such as 3D cameras, e.g., Microsoft Kinect camera, give a collection of depth points as measurements along with RGB images. The multiple measurements from a target can be used to estimate and track not only the position and velocity of the centroid but also its spatial extent. Combined target tracking and shape estimation is commonly referred to as an extended object tracking problem. Comprehensive overviews of extended object tracking are shown in [11], [12].

A variety of geometric primitives are used to approximate the shape or volume of extended objects, e.g., an extended object can be modeled by a stick [22], a rectangle [23], or a spline [24]. The ellipse is widely used to approximate the spatial extent of the object in EOT [25]–[29]. When approximating the extended object as an ellipse is not suitable or inaccurate, some simple or compound primitives and models besides ellipse are proposed. For example, the spline and stick models are used to track the elongated deformable object, where the object's length is much longer than its width [22], [24]. Objects with irregular shapes are approximated by a combination of multiple ellipses [14], [30]. Fourier series expansion [28] is used for modeling star-convex shaped objects and level-set model [31] is applied for non-convex shaped objects.

In this paper, an ellipse is used to roughly estimate the target shape from multiple unassociated noisy measurements at each time step. The estimated ellipse provides the kinematic (position and velocity of the centroid) and spatial extent (orientation and size) information, which is useful for real world applications. Two widely used models to represent targets with spatial extent are elliptic random hyper-surface model (RHM) [28] and random matrix model (RMM) [29]. In both models, the true shape of the object is approximated as an ellipse. The RHM model assumes each measurement source lies on a scaled version of the true ellipse describing the object. The extent of the object is represented by the entries from the Cholesky decomposition of the symmetric positive definite (SPD) matrix [28]. In [32], multiplicative noise terms in the measurement equation are used to model the spatial distribution of the measurements, and a second order extended Kalman filter (EKF) is derived for closed form recursive measurement update. In RMM, the shape of the

This work was supported by a Space Technology Research Institutes grant (number 80NSSC19K1076) from NASA's Space Technology Research Grants Program.

Gang Yao, Ryan Saltus, and Ashwin Dani are with the Department of Electrical and Computer Engineering, University of Connecticut Storrs, Connecticut, USA. e-mail: gang.yao@uconn.edu, ryan.saltus@uconn.edu, ashwin.dani@uconn.edu Earlier versions of this paper were presented at the 2018 Annual American Control Conference (ACC) and 2017 20th International Conference on Information Fusion (Fusion) and were published in their Proceedings <https://ieeexplore.ieee.org/document/8431874> and <https://ieeexplore.ieee.org/document/8009883>.

target object is represented by using a SPD random matrix. The SPD matrix and the centroid of the object are used as state variables, which are estimated by using a corresponding filter. Multiple improvements over the RMM model are presented in literature [25]–[27]. The situation when the measurement noise is comparable to the extent of the target and cannot be neglected is considered in [27]. Considering the target will change size and shape abruptly, especially while maneuvering, the rotation matrix or scaling factor is multiplied on both sides of the SPD matrix, and the corresponding filters are derived in [25], [26]. Comparisons between the RHM and the RMM are illustrated in [33].

The dynamic model for a moving extended object describes how the target's kinematic parameters and extent evolve over time. It is useful and necessary in certain scenarios, such as tracking with a fast maneuvering target, large sampling period or gating techniques [15], [34]. For tracking a point object, the dynamic model of the kinematic parameters fully describes the state changes of the object. However, for an extended object, the dynamic model of the spatial extent is also important, especially when the target conducts maneuvering motion or the shape (especially orientation) of the extended target changes abruptly. For tracking an extended object using RMM, there is no explicit dynamic model and the update for the extent is based on simple heuristics which increase the extent's covariance, while keeping the expected value constant [29]. An alternative to the heuristic update is to use a Wishart distribution to approximate the transition density of the spatial extent [26], [29], [35]. The prediction update of extended targets within the RMM framework is explored by multiplying the rotation matrix on both sides of the SPD matrix in [25], [26]. In [26], comprehensive comparison results between four dynamic models are presented. For tracking an elliptic extended object using RHM, the covariance matrix of the uncertainty of the object's shape parameters is increased at each time step to capture the variations in the shape [28]. The dynamic model of the ellipse is based on the assumption that the rotation angle of the ellipse is the same as the direction of the velocity of the centroid [28].

Image moments have found a wide use in tracking, visual servoing and pattern recognition [36]–[39]. Hu's moments [40], which are invariant under translation, rotation and scaling of the object, are widely investigated in pattern recognition.

In this paper, a novel representation to describe an ellipse, using image moments to approximate an extended object, is presented. Dynamic models of image moments are presented, which are used to represent an extended object moving in uniform and coordinated turn (CT) motions. The image moment-based RHM is used with the interacting multiple model (IMM) approach [34], [41]–[43] for tracking targets undergoing complex trajectories. The likelihood function based on average log-likelihood is derived for the IMM. An unscented Kalman filter (UKF) is used to estimate the states of each individual model of the UKF-IMM filter.

The contributions of the paper are briefly summarized as follows:

- The minimal, complete, and non-ambiguous representation of an elliptic object, based on image moments, is

derived for extended object tracking. A new measurement model with the explicit noise term for image moment-based ellipse representation is derived. The UKF-IMM filter is adapted based on the multiple dynamic models and the newly derived measurement model.

- A novel method for calculating the average log-likelihood of the image moment-based RHM, is presented for the UKF-IMM filter. In order to estimate the model probability consistently, the calculation of the average log-likelihood function by unscented transformation is proposed.
- Results of the UKF-IMM filter with the image moment-based RHM and the new measurement model are presented and compared with a benchmark algorithm in [25] to validate the performance of the proposed approach. Results of Monte Carlo runs with initial conditions generated from predefined Gaussian distributions are shown. The image moment-based EOT results on a sequence of image data for car tracking are also presented.

Compared to our previous work in [44], [45], this paper presents (1) a detailed literature review, (2) the derivation of the transition matrix for the image moments' dynamics when the object is undergoing a CT motion, (3) a new measurement model with the consideration of the measurement noise term (including mean and covariance), and the corresponding filtering process, and (4) simulation results with thorough evaluation and comparison with a state-of-the-art algorithm and simulation results on real image-data.

The rest of the paper is organized as follows. In Section II, the image moment-based RHM is proposed to approximate an elliptic object, and its dynamic models and an implicit measurement model are analytically derived. In Section III, the Bayesian inference of the position, velocity and extent of the object from the noisy measurement points is illustrated. Since the dynamic models and the measurement model are nonlinear, UKF is applied to estimate the extended object. For tracking the moving target switching between maneuvering and non-maneuvering motions, the UKF-IMM algorithm is presented in Section IV that uses derived image moment-based RHM. The algorithm for the calculation of the likelihood function by using the average log-likelihood function and unscented transformation is also proposed. In Section V, the proposed image moment-based RHM with its dynamic models is evaluated in three tests: (1) static scenario for validating the measurement model; (2) constant velocity (CV) and CT motion for validating the dynamic models; (3) complex trajectories for validating the UKF-IMM algorithm with the proposed image moment-based RHM. Estimation results of image moment-based RHM are compared with the RMM model within the IMM framework [25]. The estimation results show that the proposed model provides comparable and accurate results. Conclusions and future work are given in Section VI. To improve legibility, the sub indices, such as the time step k and the measurement number l , will be dropped unless needed in the following.

II. IMAGE MOMENT-BASED RANDOM HYPERSURFACE MODEL

A. Representation of the Ellipse using Image Moments

In this section, a generalized representation of the ellipse using image moments is presented. The $(i+j)$ th moment of an object m_{ij} in a 2D plane is defined by [39]

$$m_{ij} = \iint_R x^i y^j dx dy, \quad \forall i, j \in \mathbb{N} \quad (1)$$

where R is the surface of the object and $\mathbb{N}$ is the set of natural numbers and the point inside the surface of the object is $[x, y]^T$. The centered moment is defined as [39]

$$\eta_{ij} = \iint_R h(\bar{x}, \bar{y}) dx dy \quad (2)$$

where $h(\bar{x}, \bar{y}) = (\bar{x})^i (\bar{y})^j$, $\bar{x} = x - x_c$, $\bar{y} = y - y_c$ and $[x_c, y_c]^T$ is the centroid of the object.

Any point on the surface of the object is represented as a point located on the boundary of the scaled ellipse. The general equation of a family of ellipses in terms of semi-major, and semi-minor axes, centroid, and orientation is given by

$$\frac{(x - x_c + \alpha(y - y_c))^2}{a_1^2(1 + \alpha^2)} + \frac{(y - y_c - \alpha(x - x_c))^2}{a_2^2(1 + \alpha^2)} - s^2 = 0 \quad (3)$$

where a_1 and a_2 are its semi-major and semi-minor axes, respectively, θ is the orientation of the ellipse, $\alpha = \tan \theta$, and s is a scale factor. The point $[x, y]^T$ inside the ellipse is represented by varying s from 0 to 1 in (3). Rewriting (3) as follows

$$\frac{a_1^2 \alpha^2 + a_2^2}{a_1^2 a_2^2 (1 + \alpha^2)} \bar{x}^2 + \frac{\alpha^2 a_2^2 + a_1^2}{a_1^2 a_2^2 (1 + \alpha^2)} \bar{y}^2 + \frac{a_2^2 - a_1^2}{a_1^2 a_2^2} \frac{2\alpha}{1 + \alpha^2} \bar{x} \bar{y} = s^2 \quad (4)$$

Consider normalized centered moments $n_{11} = \frac{\eta_{11}}{A}$, $n_{02} = \frac{\eta_{02}}{A}$, $n_{20} = \frac{\eta_{20}}{A}$, where A is the area of the ellipse, η_{11} , η_{02} , and η_{20} are centered moments. The following relationships between parameters of ellipse a_1 , a_2 , α , and the normalized centered image moments (n_{20} , n_{02} , n_{11}) are derived [39]

$$\begin{aligned} a_1^2 &= 2 \left(n_{02} + n_{20} + \sqrt{(n_{20} - n_{02})^2 + 4n_{11}^2} \right) \\ a_2^2 &= 2 \left(n_{02} + n_{20} - \sqrt{(n_{20} - n_{02})^2 + 4n_{11}^2} \right) \\ \alpha &= \frac{1}{2n_{11}} \left(n_{02} - n_{20} + \sqrt{(n_{20} - n_{02})^2 + 4n_{11}^2} \right) \end{aligned} \quad (5)$$

The following expression is obtained by substituting (5) into (4) as

$$\frac{4n_{02}}{a_1^2 a_2^2} \bar{x}^2 + \frac{4n_{20}}{a_1^2 a_2^2} \bar{y}^2 - \frac{8n_{11}}{a_1^2 a_2^2} \bar{x} \bar{y} = s^2 \quad (6)$$

The area of ellipse, A , is written in normalized centered moments n_{ij} and parameters a_1 , and a_2 as follows [39]

$$A = \pi a_1 a_2 = 4\pi \sqrt{n_{20} n_{02} - n_{11}^2} \quad (7)$$

The shape of the ellipse in (6) is represented only by image moments by using (7) as follows

$$\frac{n_{02}}{4(n_{20} n_{02} - n_{11}^2)} \bar{x}^2 + \frac{n_{20}}{4(n_{20} n_{02} - n_{11}^2)} \bar{y}^2 - \frac{2n_{11}}{4(n_{20} n_{02} - n_{11}^2)} \bar{x} \bar{y} = s^2 \quad (8)$$

Let $\mathbf{p} = [\mathbf{p}_{\text{IM}}^T, \mathbf{p}_{\text{pos}}^T]^T$, where the shape of the ellipse is represented by $\mathbf{p}_{\text{IM}} = [n_{11}, n_{20}, n_{02}]^T$ and the location of the centroid of the ellipse is represented by $\mathbf{p}_{\text{pos}} = [x_c, y_c]^T$. An ellipse is expressed using a minimal, complete, and non-ambiguous representation of parameters $\mathbf{p}$, in the following form

$$\begin{aligned} g(x, y, \mathbf{p}) &= \frac{n_{02}}{4(n_{20} n_{02} - n_{11}^2)} \bar{x}^2 + \frac{n_{20}}{4(n_{20} n_{02} - n_{11}^2)} \bar{y}^2 \\ &\quad - \frac{2n_{11}}{4(n_{20} n_{02} - n_{11}^2)} \bar{x} \bar{y} - s^2 = 0 \end{aligned} \quad (9)$$

B. Dynamic Motion Models

In order to derive the differential equation for n_{ij} , the time derivative of the centered moment, $\dot{\eta}_{ij}$ is derived first. The time derivative of centered moment $\dot{\eta}_{ij}$ is obtained from the time derivative of the contour of the ellipse as [39]

$$\dot{\eta}_{ij} = \oint_C h(x, y) \mathbf{v}^T \bar{\mathbf{n}} dl \quad (10)$$

where C is the contour of the ellipse, $\mathbf{v} = [\dot{x}, \dot{y}]^T$ is the velocity of the contour point $\mathbf{x} = [x, y]^T$, $\bar{\mathbf{n}}$ is the unitary vector normal to C at point $\mathbf{x}$, and dl is an infinitesimal element of C . If C is piece-wise continuous, and vector $h(x, y)\dot{\mathbf{x}}$ is tangent to R and continuously differentiable $\forall \mathbf{x} \in R$, Green's theorem can be used to represent (10) as [39]

$$\dot{\eta}_{ij} = \iint_R \text{div}[h(x, y)\mathbf{v}] dx dy \quad (11)$$

Using the CV and CT models, specific differential equations of $\dot{\eta}_{ij}$ are derived for each case.

1) *Linear Motion Model*: When an elliptical object is moving linearly, each point inside the ellipse at time t obeys $\mathbf{v} = \mathbf{v}_0 + \mathbf{a}t$, where $\mathbf{v}_0 \in \mathbb{R}^2$ is the initial velocity and $\mathbf{a} \in \mathbb{R}^2$ is the acceleration. The centered moment of the ellipse μ_{ij} is calculated by substituting $h(x, y) = (x - x_c)^i (y - y_c)^j$ in (11) as

$$\dot{\eta}_{ij} = \iint_R \left[\frac{\partial h}{\partial x} \dot{x} + \frac{\partial h}{\partial y} \dot{y} + h(x, y) \left(\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} \right) \right] dx dy \quad (12)$$

Since $\frac{\partial h}{\partial x}$ and $\frac{\partial h}{\partial y}$ are odd functions and R is symmetric with respect to the centroid, the state space representation of the normalized centered moments of the ellipse $\mathbf{p}_{\text{IM}} = [n_{11}, n_{20}, n_{02}]^T$ is

$$\dot{\mathbf{p}}_{\text{IM}} = 0 \quad (13)$$

The state at discrete time k is given by $\mathbf{p}_k = [\mathbf{p}_{\text{IM},k}^T, \mathbf{p}_{\text{CV},k}^T]^T$, where $\mathbf{p}_{\text{IM},k}$ is a component of the state related to image moments, $\mathbf{p}_{\text{CV},k} = [x_{c,k}, \dot{x}_{c,k}, y_{c,k}, \dot{y}_{c,k}]^T$ is the vector

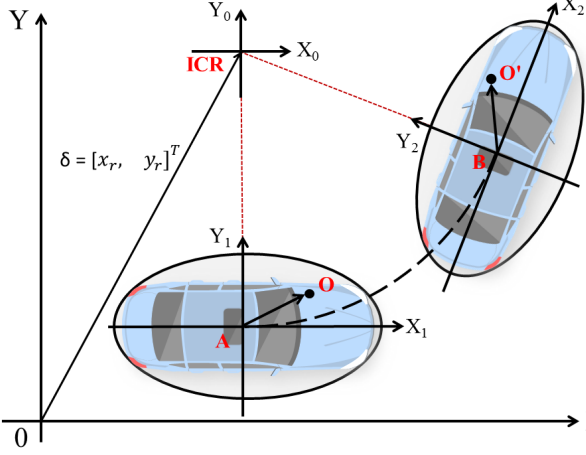


Figure 1: Coordinated turn model of the elliptic extended object.

that includes the position and velocity of the centroid of the extended object. The discretized state equation is given as follows

$$\mathbf{p}_{k+1} = \mathbf{F}_{CV}\mathbf{p}_k + \mathbf{w}_k \quad (14)$$

where the state transition matrix $\mathbf{F}_{CV} = \text{diag}(\mathbf{I}_{3 \times 3}, \mathbf{Q})$, with $\mathbf{Q} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}$, and $\mathbf{w}_k$ is the zero-mean Gaussian noise with covariance matrix $\mathbf{C}_{CV,k} = \text{diag}(\mathbf{C}_{IM,k}, \mathbf{C}_k^w, \mathbf{C}_k^w)$, where $\mathbf{C}_{IM,k} \in \mathbb{R}^{3 \times 3}$ is the diagonal noise covariance for the image moments and $\mathbf{C}_k^w = \begin{bmatrix} \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^2 & T \end{bmatrix} q$, where q is the power spectral density. Notice that the discretized white noise acceleration model is used for the state vector $\mathbf{p}_{CV,k}^T$, which is the same as the dynamic model for point based tracking. Other kinematic models for point based tracking can also be used for the state vector $\mathbf{p}_{CV,k}^T$ [42], [46].

2) *Coordinated Turn Motion Model*: The CT model, characterized by constant turning rate and constant speed, is commonly used in tracking applications [15], [42]. An elliptic extended object executing a coordinated turn is shown in Fig. 1. For any point $O(x, y)$ that belongs to the ellipse moving with a CT motion, the motion model of the point is represented as follows

$$\begin{aligned} \dot{x} &= -\omega(y - y_r) \\ \dot{y} &= \omega(x - x_r) \end{aligned} \quad (15)$$

where ω is the turning rate and $\delta = [x_r, y_r]^T$ is the displacement between the origin of the reference frame XY and the origin of the reference frame X_0Y_0 . The origin of the reference frame X_0Y_0 is the instantaneous center of rotation (ICR) of the object.

The differential equation of the centered moments of the ellipse when the object is undergoing CT motion is given by

substituting (15) into (12) as

$$\begin{aligned} \dot{\eta}_{ij} &= \omega \iint_R \left[\frac{\partial h}{\partial x}(y_r - y) + \frac{\partial h}{\partial y}(x - x_r) \right] dx dy \\ &= \omega \iint_R \left[\left(\frac{\partial h}{\partial y}x - \frac{\partial h}{\partial x}y \right) + \left(-\frac{\partial h}{\partial y}x_r + \frac{\partial h}{\partial x}y_r \right) \right] dx dy \end{aligned} \quad (16)$$

The dynamic models of the normalized centered moments of the ellipse are calculated using (16) as

$$\begin{aligned} \dot{n}_{11} &= \omega(n_{20} - n_{02}) \\ \dot{n}_{20} &= -2\omega n_{11} \\ \dot{n}_{02} &= 2\omega n_{11} \end{aligned} \quad (17)$$

The state space representation of the normalized centered moments of the ellipse $\mathbf{p}_{IM} = [n_{11}, n_{20}, n_{02}]^T$ is

$$\dot{\mathbf{p}}_{IM} = \begin{bmatrix} 0 & \omega & -\omega \\ -2\omega & 0 & 0 \\ 2\omega & 0 & 0 \end{bmatrix} \mathbf{p}_{IM} \quad (18)$$

and the solution to the state space in (18) is

$$\mathbf{p}_{IM}(t) = \mathbf{M}(t, t_0)\mathbf{p}_{IM}(t_0) \quad (19)$$

where $\mathbf{M}(t, t_0) = \begin{bmatrix} \cos 2\theta & \frac{1}{2}\sin 2\theta & -\frac{1}{2}\sin 2\theta \\ -\sin 2\theta & \cos^2 \theta & \sin^2 \theta \\ \sin 2\theta & \sin^2 \theta & \cos^2 \theta \end{bmatrix}$ is the transition matrix from t_0 to t , with $\theta = \omega(t - t_0)$. The derivation of the transition matrix is shown in Appendix A.

At each time step k , the complete state to be tracked is $\mathbf{p}_k = [\mathbf{p}_{IM,k}^T, \mathbf{p}_{CT,k}^T]^T$, where $\mathbf{p}_{IM,k}$ is a component of the state corresponding to the image moments, and $\mathbf{p}_{CT,k} = [x_{c,k}, \dot{x}_{c,k}, y_{c,k}, \dot{y}_{c,k}, \omega_k]^T$ is a vector that includes the position and velocity of the centroid, and the turning rate of the extended object. The state equation is given as follows

$$\mathbf{p}_{k+1} = \mathbf{F}_{CT}\mathbf{p}_k + \mathbf{\Gamma}\mathbf{w}_k \quad (20)$$

where the state transition matrix $\mathbf{F}_{CT} = \text{diag}(\mathbf{M}, \mathbf{E})$, $\mathbf{M} = \begin{bmatrix} \cos 2\omega_k T & \frac{1}{2}\sin 2\omega_k T & -\frac{1}{2}\sin 2\omega_k T \\ -\sin 2\omega_k T & \cos^2 \omega_k T & \sin^2 \omega_k T \\ \sin 2\omega_k T & \sin^2 \omega_k T & \cos^2 \omega_k T \end{bmatrix}$ is obtained from (19),

T is the sampling period, $\mathbf{E} = \begin{bmatrix} 1 & \frac{\sin \omega_k T}{\omega_k} & 0 & -\frac{1 - \cos \omega_k T}{\omega_k} & 0 \\ 0 & \cos \omega_k T & 0 & -\sin \omega_k T & 0 \\ 0 & \frac{1 - \cos \omega_k T}{\omega_k} & 1 & \frac{\sin \omega_k T}{\omega_k} & 0 \\ 0 & \sin \omega_k T & 0 & \cos \omega_k T & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, and $\mathbf{\Gamma} = \text{diag}(\mathbf{I}_{3 \times 3}, \mathbf{\Gamma}_{CT})$ with $\mathbf{\Gamma}_{CT} = \begin{bmatrix} \frac{1}{2}T^2 & 0 & 0 \\ 0 & T & 0 \\ 0 & \frac{1}{2}T^2 & 0 \\ 0 & 0 & T \\ 0 & 0 & T \end{bmatrix}$, and

$\mathbf{w}_k \in \mathbb{R}^{6 \times 1}$ is a zero-mean Gaussian noise vector. Notice that this model is piece-wise continuous.

C. Measurement Model

Assuming there is the uniformly spaced measurement $\bar{\mathbf{z}} = [x, y]^T$ without sensor noise, (9) maps the unknown state $\mathbf{p}$ to the pseudo-measurement 0 with the squared scale term $s^2 \sim \mathcal{U}(0, 1)$. Similar to the RHM in [28], the squared scale factor s^2 introduced in (9) represents a set of measurement points uniformly distributed inside the target. When the measurement point is located on the centroid of the ellipse, the value of s^2 equals to 0. When the measurement point is located on the edge of the ellipse, the value of s^2 equals to 1. The scaling factor s is approximated to be Gaussian

distributed with mean $2/3$ and variance $1/18$ [47]. Consider the real measurement $\mathbf{z} = [\tilde{x}, \tilde{y}]^T$ of the unknown true measurement $\bar{\mathbf{z}} = [x, y]^T$ in the presence of additive white Gaussian noise $\boldsymbol{\nu} = [\nu_x, \nu_y]^T$, where $\nu_x \sim \mathcal{N}(0, \sigma_x^2)$ and $\nu_y \sim \mathcal{N}(0, \sigma_y^2)$, the real measurement $\mathbf{z}$ is expressed as $\mathbf{z} = \bar{\mathbf{z}} + \boldsymbol{\nu}$. To find the relationship between the state vector $\mathbf{p}$ and the real measurement $\mathbf{z} = [\tilde{x}, \tilde{y}]^T$, the measurement model is derived by substituting $\mathbf{z}$ in (9). The following expression is obtained

$$g(\bar{\mathbf{z}}, \mathbf{p}) = g(\mathbf{z}, \mathbf{p}) - f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p}) = v \quad (21)$$

where v is the pseudo-measurement with the true value of 0 and $f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})$ is a polynomial related to the white noise $\boldsymbol{\nu}$, which has the mean

$$E[f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})] = \rho(n_{02}\sigma_x^2 + n_{20}\sigma_y^2) \quad (22)$$

and covariance

$$C_{f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})} = \rho^2 \left\{ 2n_{02}^2\sigma_x^4 + 2n_{20}^2\sigma_y^4 + 4n_{11}^2\sigma_x^2\sigma_y^2 + 4[n_{02}(\tilde{x} - x_g) - n_{11}(\tilde{y} - y_g)]^2\sigma_x^2 + 4[n_{20}(\tilde{y} - y_g) - n_{11}(\tilde{x} - x_g)]^2\sigma_y^2 \right\} \quad (23)$$

where $\rho = 1/4(n_{20}n_{02} - n_{11}^2)$. The derivation of $f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})$ and its first two moments are shown in Appendix B. Since the measurement model is highly nonlinear, the UKF presented in the next section is used to estimate the state vector $\mathbf{p}$.

III. UKF FOR EXTENDED OBJECT TRACKING USING IMAGE MOMENT-BASED RHM

On the basis of the dynamic motion models and the measurement model, a recursive Bayesian state estimator for tracking the elliptic extended object is derived. At each time step, several measurement points from the object are received. The task of the Bayesian state estimator is to perform backward inference, inferring the true state parameters from the measurement points. The measurement points at each time step k are denoted as $\mathbf{Z}_k = \{\mathbf{z}_{k,l}\}_{l=1}^{L_k}$, assuming there are L_k measurements and each measurement point is $\mathbf{z}_{k,l} = [x, y]^T$. The state vector up to time step k , when all the measurements are incorporated, is denoted as $\mathbf{p}_k$. Suppose that the posterior probability density function (pdf) $p(\mathbf{p}_{k-1} | \mathbf{Z}_{k-1})$ at time step $k-1$ is available, the prediction $p(\mathbf{p}_k | \mathbf{Z}_{k-1})$ for time step k is given by the Chapman-Kolmogorov equation as [34]

$$p(\mathbf{p}_k | \mathbf{Z}_{k-1}) = \int p(\mathbf{p}_k | \mathbf{p}_{k-1})p(\mathbf{p}_{k-1} | \mathbf{Z}_{k-1})d\mathbf{p}_{k-1} \quad (24)$$

where the state vector evolves by the conditional density function $p(\mathbf{p}_k | \mathbf{p}_{k-1})$. The conditional density function $p(\mathbf{p}_k | \mathbf{p}_{k-1})$ is derived based on different dynamic models in Subsection II-B.

Assuming the measurements $\mathbf{Z}_k = \{\mathbf{z}_{k,l}\}_{l=1}^{L_k}$ at time k are independent, the prediction $p(\mathbf{p}_k | \mathbf{z}_{k,l})$ is updated recursively via Bayes' rule as [28]

$$p(\mathbf{p}_k | \mathbf{z}_{k,l}) \propto p(\mathbf{z}_{k,l} | \mathbf{p}_k)p(\mathbf{p}_k | \mathbf{z}_{k,l-1}) \quad (25)$$

where $p(\mathbf{p}_k | \mathbf{z}_{k,0}) = p(\mathbf{p}_k | \mathbf{Z}_{k-1})$ and $p(\mathbf{p}_k | \mathbf{Z}_k) = p(\mathbf{p}_k | \mathbf{z}_{k,L_k})$.

When the target is moving with uniform motion (CV model, which is a linear system), its states $\mathbf{p}_{k|k-1}$ and covariance $C_{k|k-1}$ are predicted based on the dynamic model (14) as

$$\mathbf{p}_{k|k-1} = \mathbf{F}_{CV}\mathbf{p}_k \quad (26)$$

$$C_{k|k-1} = \mathbf{F}_{CV}\mathbf{p}_k\mathbf{F}_{CV}^T + \mathbf{C}_{CV} \quad (27)$$

However, the proposed image moment-based RHM and its CT model are nonlinear. When the system is nonlinear, a linearization method such as the EKF will introduce large errors in the true posterior mean and covariance [48]. UKF addresses this problem by using unscented transformation (UT), which doesn't require the calculations of the Jacobian and Hessian matrices. The UT sigma point selection scheme results in approximations that are accurate to the third order for Gaussian inputs for all nonlinearities and has the same order of the overall number of computations as the EKF [48]. When the state variables in $\mathbf{p} \in \mathbb{R}^{M \times 1}$ with mean $\bar{\mathbf{p}}$ and covariance $C_{\mathbf{p}}$ are propagating through a nonlinear function $\mathbf{g} = \psi(\mathbf{p})$, such as (19) or (21), the mean $\bar{\mathbf{g}}$ and covariance $C_{\mathbf{g}}$ of $\mathbf{g}$ are approximated by generating the UT sigma points $\mathcal{X}_i$ as [48]

$$\bar{\mathbf{g}} = \sum_{i=0}^{2M} W_i^{(m)} \mathcal{Y}_i \quad (28)$$

$$C_{\mathbf{g}} = \sum_{i=0}^{2M} W_i^{(C)} \{\mathcal{Y}_i - \bar{\mathbf{g}}\} \{\mathcal{Y}_i - \bar{\mathbf{g}}\} \quad (29)$$

where $\mathcal{Y}_i = \psi(\mathcal{X}_i)$. The sigma points $\mathcal{X}_i$ and the weights $W_i^{(m)}$ and $W_i^{(C)}$ are calculated by [48]

$$\mathcal{X}_0 = \bar{\mathbf{p}}$$

$$\mathcal{X}_i = \bar{\mathbf{p}} + \left(\sqrt{(M + \lambda)C_{\mathbf{p}}} \right)_i \quad i = 1, \dots, M$$

$$\mathcal{X}_i = \bar{\mathbf{p}} - \left(\sqrt{(M + \lambda)C_{\mathbf{p}}} \right)_i \quad i = M + 1, \dots, 2M$$

$$W_0^{(m)} = \lambda / (M + \lambda)$$

$$W_0^{(C)} = \lambda / (M + \lambda) + (1 - \gamma^2 + \beta)$$

$$W_i^{(m)} = W_i^{(C)} = 1 / [2(M + \lambda)] \quad i = 1, \dots, 2M \quad (30)$$

where λ is the scaling parameter as $\lambda = \gamma^2(M + \kappa) - M$, γ is the parameter that determines the spread of the sigma points around the mean $\bar{\mathbf{p}}$, κ is the secondary scaling parameter that is usually set to 0 and β is the parameter that incorporates the prior knowledge of the distribution of $\mathbf{p}$. The UKF for the image moment-based RHM is illustrated in Algorithm 1.

IV. TRACKING EXTENDED TARGET WITH IMM

In this section, the proposed image moment-based random hypersurface model is embedded within the IMM approach for tracking an extended target undergoing complex trajectories. When the extended target is switching between maneuvering and non-maneuvering motions, its kinematic state and spatial extent may change abruptly. Multiple model approaches, such as the IMM, are effective at tracking a target with complex trajectories, especially with a high maneuvering index [42], [43].

Algorithm 1: UKF with sequential processing of measurements.

Set the time steps N ;
Set the initial state vector $\mathbf{p}_0$ and covariance C_0 ;
for $k=1$ **to** N **do**
 case CV model do
 Predict state $\mathbf{p}_{k|k-1}$ as in (26);
 Predict covariance $C_{k|k-1}$ as in (27);
 case CT model do
 Augment the state vector
 $\mathbf{p}_{k-1}^a = [(\mathbf{p}_{k-1})^T, \mathbf{w}_k^T]^T$;
 Calculate sigma points using (30);
 Predict states based on (20) with sigma points;
 Use (28), (29) to calculate the mean and covariance of the state vector $\mathbf{p}_{k|k-1}$;
Obtain the measurement points $\mathbf{Z}_k = \{\mathbf{z}_{k,l}\}_{l=1}^{L_k}$ at time step k ;
 for $l=1$ **to** L_k **do**
 Calculate the mean and covariance of $f(\mathbf{z}_{k,l}, \boldsymbol{\nu}, \mathbf{p}_{k|k-1,l})$ using (22) and (23);
 Augment the state vector $\mathbf{p}_{k|k-1,l}^a = [(\mathbf{p}_{k|k-1,l})^T, f(\mathbf{z}_{k,l}, \boldsymbol{\nu}, \mathbf{p}_{k|k-1,l}), s]^T$;
 Calculate sigma points using (30);
 Calculate pseudo-measurement $v_{k,l}$ based on (21) for measurement point $\mathbf{z}_{k,l}$;
 Use (28), (29) to calculate the mean and covariance of the $v_{k,l}$;
 Update state vector $\mathbf{p}_{k,l}$;
 end for

The IMM approach assumes the target obeys one of a finite number of motion models and identifies the beginning and the end of the motion models by updating the model probabilities. The adaptation via model probability update helps the IMM approach keep the estimation errors consistently low, both during maneuvers as well as non-maneuver intervals. Methods for using IMM when the dimensions of the state vectors do not match are presented in [41].

The image moment-based model with the derived dynamic motion models, such as the CV motion model and the CT motion model in Section II, are integrated in an IMM framework. Since the dynamic motion model and the measurement model are nonlinear, the UKF-IMM algorithm is used. The dimensions of the states for CV and CT models are not the same. During the mixing in the IMM, the state vector of the CV model is appended with a zero component [42]. The flowchart of the UKF-IMM algorithm is shown in Fig. 2, where $\mu_{k-1}^{i|j}$ is the mixing probability, $p_{i|j}$ is the Markov chain transition matrix between the i th and j th models and Λ_k^j is the likelihood function corresponding to the j th model. There are multiple measurement points at each time step, the sequential approach is adopted for UKF and the likelihood function is generated based on the measurement model.

At each time step k , assuming there are L_k measurements $\mathbf{Z}_k = \{\mathbf{z}_{k,l}\}_{l=1}^{L_k}$. The pseudo-measurement variable $v_{k,l}$ is

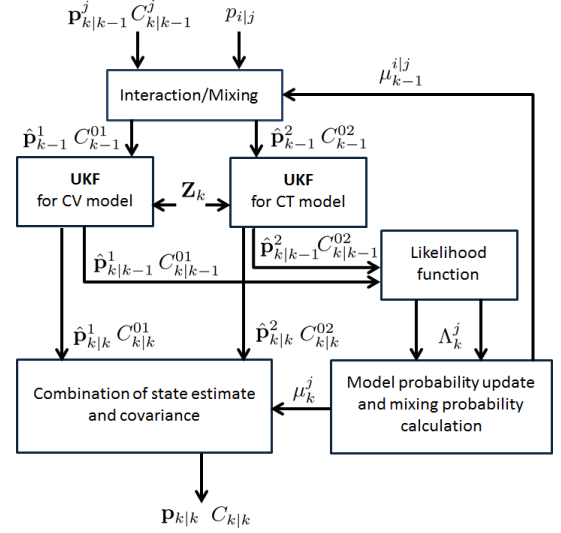


Figure 2: Flowchart of UKF-IMM framework.

generated for each measurement $\mathbf{z}_{k,l}$, based on the predicted state vector $\mathbf{p}_{k|k-1}^j$, covariance $C_{k|k-1}^j$ and the measurement model in (21). The mean and the covariance of the pseudo-measurement variable $v_{k,l}$ are obtained via UT. The log-likelihood function based on the pseudo-measurement variable $v_{k,l}$ is

$$\log \Lambda_k^j = \sum_{l=1}^{L_k} \left[-\frac{(0 - \phi_{v,l})^2}{2\sigma_{v,l}^2} - \log \left(\sqrt{2\pi\sigma_{v,l}^2} \right) \right] \quad (31)$$

where $\phi_{v,l}$ and $\sigma_{v,l}^2$ are the mean and covariance of the pseudo-measurement $v_{k,l}$, generated for each measurement point $\mathbf{z}_{k,l}$. In many cases, the likelihood Λ_k^j becomes extremely small and the model probability is inaccurate. To avoid this issue, the average log-likelihood $\log \bar{\Lambda}_k^j$ is given by

$$\log \bar{\Lambda}_k^j = \frac{1}{L_k} \log \Lambda_k^j \quad (32)$$

and

$$\bar{\Lambda}_k^j = \exp \left(\log \bar{\Lambda}_k^j \right) \quad (33)$$

which is the value of the measurement likelihood between 0 and 1. This measurement likelihood is used in the IMM filter. The details of the calculation of the measurement likelihood are shown in Algorithm 2.

V. SIMULATION RESULTS

In this section, several simulation tests are conducted to evaluate the performance of the proposed image moment-based extended object tracking. To validate the measurement model in (21), the shape of the static object is estimated at different noise levels in the first simulation. Then the tracking of the extended targets moving with CV motion and CT motion are demonstrated. The CV model in (14) and the CT model in (20) are used and validated for these cases. At last, targets with complex trajectories are simulated. In this case, the UKF-IMM algorithm with the CV model and the CT model is applied. The improved RMM in combination with the IMM approach

Algorithm 2: Calculation of the measurement likelihood $\bar{\Lambda}_k^j$ corresponding to the j th model by unscented transformation.

Obtain the predicted state vector $\mathbf{p}_{k|k-1}^j$ and covariance $C_{k|k-1}^j$ of model j ;

Obtain the measurement points $\mathbf{Z}_k = \{\mathbf{z}_{k,l}\}_{l=1}^{L_k}$ at time step k ;

for $l=1$ **to** L_k **do**

Calculate the mean and covariance of $f(\mathbf{z}_{k,l}, \boldsymbol{\nu}, \mathbf{p}_{k|k-1})$ using (22) and (23);

Augment the state vector

$\mathbf{p}_{k|k-1,l}^a = \left[\left(\mathbf{p}_{k|k-1,l}^j \right)^T, f(\mathbf{z}_{k,l}, \boldsymbol{\nu}, \mathbf{p}_{k|k-1}), s \right]^T$;

Calculate sigma points $\mathcal{X}$ using (30);

Propagate sigma points $\mathcal{X}$ through the measurement model in (21);

Use (28), (29) to calculate the mean and covariance of the pseudo-measurement $v_{k,l}$;

Sum the values of the log-likelihood function using (31);

Calculate the value of the average log-likelihood function using (32) and the measurement likelihood using (33);

in [25] is implemented as a benchmark comparison for our proposed image moments based random hypersurface model.

The intersection over union (IoU) is used as the metric to evaluate the proposed algorithm. The IoU is defined as the area of the intersection of the estimated shape and the true shape divided by the union of the two shapes [24]

$$\text{IoU} = \frac{\text{area}(\mathbf{p}) \cap \text{area}(\hat{\mathbf{p}})}{\text{area}(\mathbf{p}) \cup \text{area}(\hat{\mathbf{p}})} \quad (34)$$

where $\mathbf{p}$ is the true state vector and $\hat{\mathbf{p}}$ is the estimated state vector. IoU is between 0 and 1, where the value 1 corresponds to a perfect match between the estimated area and the ground-truth. Additionally, the root mean squared errors (RMSEs) of the estimated position and velocity of the centroid $[x_c, y_c]^T$ of the extended target are also evaluated, which are defined as [42]

$$\text{RMSE} = \sqrt{\frac{1}{I} \sum_{i=1}^I \xi_i^2} \quad (35)$$

where I is the number of Monte Carlo runs, ξ_i is the error of the estimation from the i th run. For the RMSE of the position, $\xi_{i,p} \triangleq (\hat{x}_c - x_c)^2 + (\hat{y}_c - y_c)^2$, where $[\hat{x}_c, \hat{y}_c]^T$ is the estimated centroid of the extended target and $[x_c, y_c]^T$ is the ground-truth. Similarly, for the RMSE of the velocity, the estimation error is defined as $\xi_{i,v} \triangleq (\hat{\dot{x}}_c - \dot{x}_c)^2 + (\hat{\dot{y}}_c - \dot{y}_c)^2$, where $[\hat{\dot{x}}_c, \hat{\dot{y}}_c]^T$ is the estimated velocity of the centroid and $[\dot{x}_c, \dot{y}_c]^T$ is the ground-truth. The filter tuning is based on the Monte Carlo method. The parameters include the initial kinematic states, the initial shape states and the process noise covariance. One-point initialization in [42] is used for tuning the kinematic parameters. The shape parameters and

the process noise covariance are tuned by performing a parameter sweep with Monte Carlo runs. The parameters with the smallest errors are chosen for validation and comparison.

A. Static Extended Object

The major and minor axes of the elliptic target are set to 3cm and 2cm, respectively. The simulation is performed by uniformly sampling 200 points from the static extended objects. Three different levels of additive white Gaussian noise with variances $\text{diag}(0.1^2, 0.1^2)$ (low), $\text{diag}(0.5^2, 0.5^2)$ (medium) and $\text{diag}(1, 1)$ (high, in cm^2), are used to generate the noisy measurements.

UKF is used for estimating the state given noisy measurements uniformly sampled from ellipse-shaped extended objects. The prior of the shape is specified as a circle with radius of 1cm. The centroid of the ellipse is initialized as random variables with Gaussian distribution of mean $[0.1, 0.5]^T$ and covariance $\text{diag}(0.01, 0.25)$ (with position in cm). The estimation results of one particular run for the ellipse-shaped object are shown in Figs. 3(a), 3(b), 3(c). Fig. 5 shows the mean values of IoU over 100 Monte Carlo runs with different numbers of measurements. With the increases in covariance of the measurement noise, the proposed image moment-based model also gives a shape close to the actual shape of the targets.

B. Linear Motion

In this subsection, an extended object with elliptical shape moving with a constant velocity is simulated. The major and minor axes are set to 30cm and 20cm, respectively. The extended object originates at position $[0, 80]^T \text{cm}$ and moves with a constant velocity of $[4, 2]^T \text{cm/s}$ for 60 seconds. The measurements are sampled from the target every 10 seconds. At each time step k , uniformly sampled measurements from the objects are generated and the number of the measurement points is generated with Poisson distribution of mean 50.

For UKF implementation, the prior of the shape variables is initialized as $\mathbf{p}_{\text{IM}} = [n_{11}, n_{20}, n_{02}]^T = [2, 30, 20]^T$ with covariance $\text{diag}(10, 2, 1)$. The kinematic states are initialized as random variables with Gaussian distribution of mean $\mathbf{p}_{\text{CV},k} = [x_{c,k}, \dot{x}_{c,k}, y_{c,k}, \dot{y}_{c,k}]^T = [0, 4, 80, 2]^T$ and covariance $\text{diag}(3^2, 0.18, 3^2, 0.18)$ (with position in cm and velocity in cm/s). The additive white Gaussian noise variance is selected as $\text{diag}(4^2, 4^2)$ for each point measurement. The power spectral density of the process noise covariance in the CV model in (14) is set as $q = 0.01$ and $\mathbf{C}_{\text{IM}} = \text{diag}(0.01, 0.01, 0.01)$. The tracking results for the ellipse-shaped extended object are shown in Fig. 4. The mean value of the RMSE of the position over 100 Monte Carlo runs is 1.40cm, and the mean value of the RMSE of the velocity over 100 Monte Carlo runs is 1.15cm/s. The mean value of the IoU is 0.81.

The RMM algorithm in [25] is also tested in this case. The kinematic process noise is $q = 0.1$ in (14) and the extension agility is $\delta = 100$. The rotation angle of the rotation matrix for the SPD matrix is set as $\theta = 0^\circ$. The mean value of the RMSE of the position over 100 Monte Carlo runs is 1.24cm,

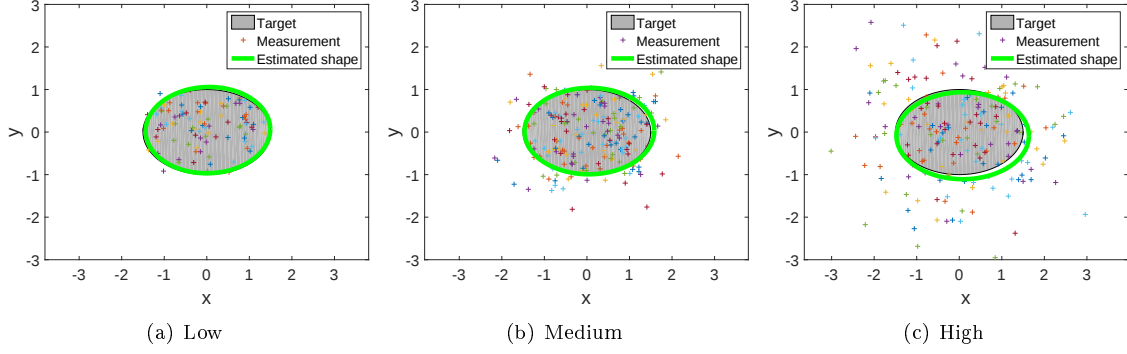


Figure 3: Estimation of the shape of an extended target with different measurement noise levels in a particular run (200 measurements); estimation of the shape is drawn in green and the ground truth is drawn in gray.

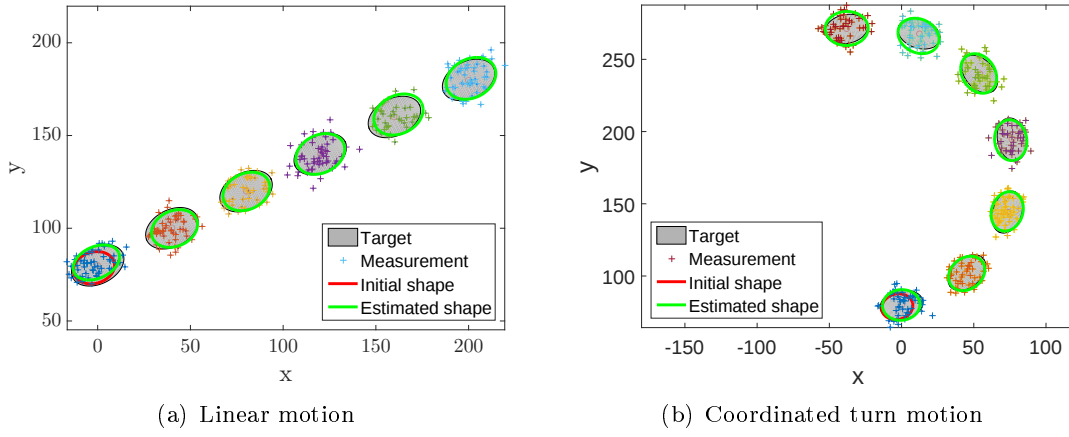


Figure 4: Tracking of the extended objects during CV model and CT model in a particular run; initial shape is shown in red, estimation of the shape is drawn in green and the ground truth is drawn in gray: (a) The ellipse with a constant velocity; (b) The ellipse executes a $3^\circ/s$ coordinated turn.

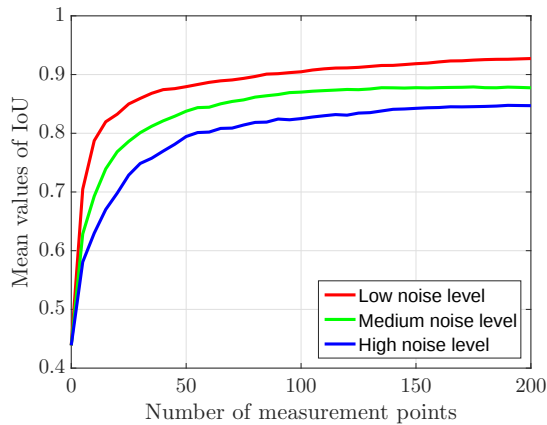


Figure 5: The mean values of IoU over 100 Monte Carlo runs with different number of measurements.

and the mean value of the RMSE of the velocity over 100 Monte Carlo runs is 1.04 cm/s . The mean value of the IoU of the ellipse is 0.78. The image moment-based model has comparable accuracy with the RMM in terms of the RMSE

and IoU metrics.

C. Coordinated Turn Motion

The extended object undergoing CT motion is simulated in this case. The extended object located at $[0, 80]^T \text{ cm}$ moves with velocity $[5, 1]^T \text{ cm/s}$ at time $t = 0$, and executes a $3^\circ/s$ coordinated turn for 60 seconds. The sampling interval is 10 seconds. At each time step, noisy measurement points are uniformly generated from the extent of the target and the number of measurements is generated with Poisson distribution of mean 50. White Gaussian noise with variance of $\text{diag}(4^2, 4^2)$ (with position in cm and velocity in cm/s) is added to each measurement.

The extended object executing a coordinated turn is estimated based on the dynamic model (20). The prior of the shape variables is initialized as $\mathbf{p}_{\text{IM}} = [n_{11}, n_{20}, n_{02}]^T = [1, 30, 20]^T$ with covariance $\text{diag}(10, 5, 0.1)$. The kinematic states are initialized as random variables with Gaussian distribution. The ground truth is set as its mean and the covariance matrix is $\text{diag}(3^2, 0.18, 3^2, 0.18, (2^\circ/s)^2)$. The tracking results for the ellipse-shaped object are shown in Fig. 4(b).

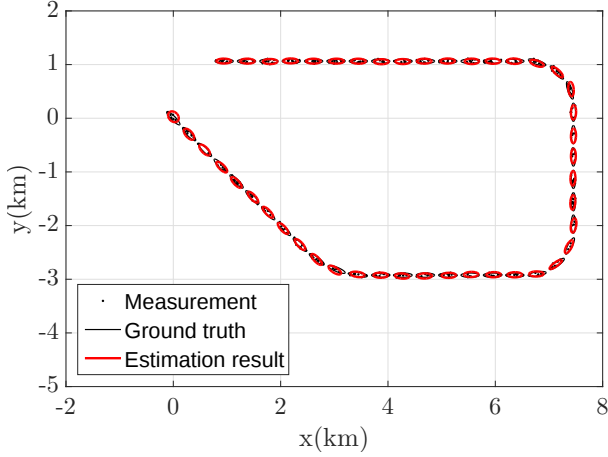


Figure 6: The trajectory, measurements and one running example of the simulation. Estimation results are shown for every 30 seconds.

The mean value of the RMSE of the position over 100 Monte Carlo runs is 1.67cm. The mean value of the RMSE of the velocity over 100 Monte Carlo runs is 0.32cm/s, and the mean value of the IoU of the ellipse is 0.81.

The RMM algorithm in [25] is also tested in this case. The kinematic process noise is set as $q = 0.1$ in (14) and the extension agility is set as $\delta = 10$. The rotation angle of the rotation matrix for the SPD matrix is set as $\theta = 30^\circ$. The mean value of the RMSE of the position over 100 Monte Carlo runs is 1.29cm, and the mean value of the RMSE of the velocity over 100 Monte Carlo runs is 1.27cm/s. The mean value of the IoU of the ellipse is 0.71. The image moment-based model, which provides a dynamic model for the shape of an extended object undergoing a coordinated turn, estimates the position and velocity, as well as the orientation and extent of the target very accurately.

D. Complex trajectory

The image moment-based RHM is embedded in the IMM framework. The target initially located at the origin moves with a constant velocity of 50km/h. The target first executes a 45° coordinated turn at 260 seconds with a turning rate of $0.46^\circ/s$ for 100 seconds, then executes two 90° coordinated turns at 570 seconds and 830 seconds with a turning rate of $0.90^\circ/s$ for 100 seconds. The trajectory is shown in Fig. 6. Similar trajectories are also used in [25], [27], [32]. The major and minor axes of the elliptic target are set to 340m and 80m, respectively. The number of the measurements in each scan is generated based on a Poisson distribution with mean of 10, and the measurement points are uniformly distributed inside the ellipse. The variance of the additive white Gaussian noise is $\text{diag}(10^2, 10^2)$ (with position in m and velocity in m/s), and the sampling time is 10s.

The proposed image moment-based RHM with the UKF-IMM filter combines the CV model in (14) and the CT model in (20). The power spectral density q for the process noise covariance in the CV model in (14) is set as 0.01 and $\mathbf{C}_{\text{IM}} = \text{diag}(1, 1, 1)$. For the CT model in (20),

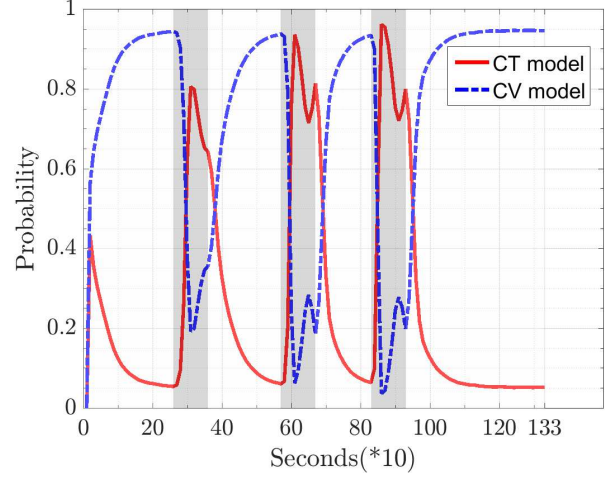


Figure 7: Model probability of the UKF-IMM filter for the image moment-based RHM. The gray areas indicate the ground truth of the maneuvering intervals.

$\mathbf{w}_k = [0.0001, 0.0001, 0.0001, 0.0001, 0.0001, (0.02^\circ)^2]^T$. The extent of the target is initialized as a circle of radius 100m. The parameters of the initial circle (i.e., center and the radius) or initial ellipse (i.e., center, semi-major and semi-minor axes lengths, and the orientation) are chosen by observing the measurement data in the first frame, such that the area of the circle or ellipse is the same as or smaller than the smallest enclosing area of the measurement points. The initial image moment states are computed based on the initial shape of the target. The initial covariance of the image moments is tuned based on the Monte Carlo runs. The kinematic states (position and velocity in each dimension) are initialized as random variables with Gaussian distribution. Its mean is the ground truth and the covariance matrix is set as $\text{diag}(100, 2)$. The turning rate for the CT model is initialized as a random variable with distribution $\mathcal{N}(0, 1)$. The initial probability μ_0^j of the two models in the IMM filter is equal and the Markov chain transition matrix is selected to be $p_{i|j} = \begin{bmatrix} 0.99 & 0.01 \\ 0.10 & 0.90 \end{bmatrix}$. The model probability of the proposed algorithm is shown in Fig. 7.

The proposed image moment-based RHM with UKF-IMM algorithm is compared with the image moment-based RHM with CT model and the RMM with IMM algorithm in [25]. The image moment-based RHM with CT model is using the same parameters as the CT model in the RHM with UKF-IMM method. The RMM-IMM algorithm uses two models. The model with a high kinematic process noise ($q = 0.1$ in (14)) and a high extension agility ($\delta = 10$) accounts for abrupt changes in shape and orientation during maneuvers, and another model with low kinematic noise ($q = 0.01$ in (14)) and a low extension agility ($\delta = 30$) accounts for the non-maneuvers. The rotation angles of the rotation matrices for the two SPD matrices are set as $\theta = 0^\circ$ and $\theta = 9^\circ$ [25].

The simulation results over 1000 Monte Carlo runs are shown in Fig. 8. The average IoU value of the image moment-based RHM with UKF-IMM method is 0.67, the average IoU

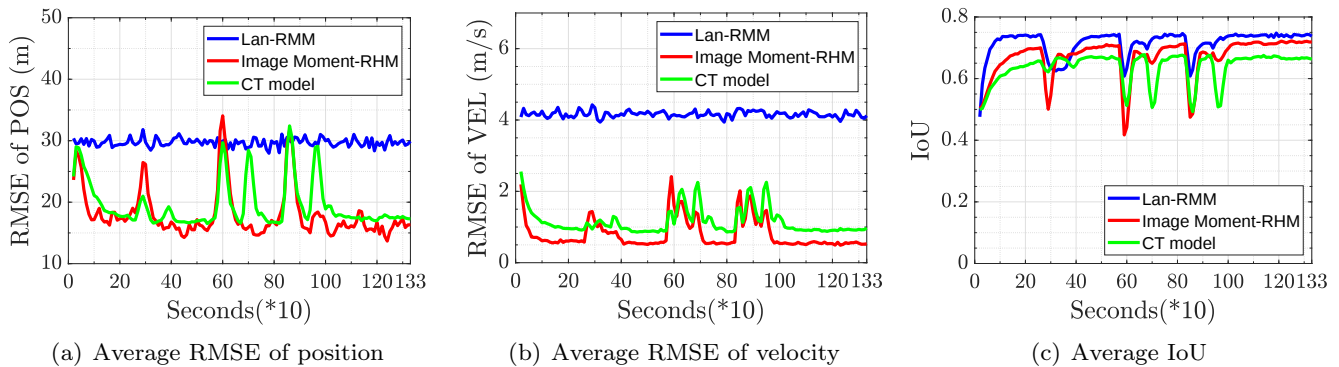


Figure 8: Simulation results of the proposed image moment-based RHM using UKF-IMM method, the image moment-based RHM using CT model and the RMM in [25] over 1000 Monte Carlo runs: (a) The average RMSE of the position of the centroid; (b) The average RMSE of the velocity of the centroid; (c) The average IoU.

value of the image moment-based RHM only using CT model is 0.64, and the average IoU value of the RMM is 0.71. The proposed image moment-based RHM with UKF-IMM method has lower RMSE values both for position and velocity, and higher average IoU value compared with the estimation results of the RHM with the CT model. The proposed image moment-based RHM has significantly lower RMSE values both for position and velocity, and comparable average IoU value to the RMM in [25]. The centroid of the ellipse estimated by the proposed image moment-based RHM is closer to the ground-truth than the centroid of the ellipse estimated by RMM. The target is a flattened ellipse, which has an eccentricity of 0.97. There are very few measurement points located around the tip area of the ellipse, and the proposed image moment-based RHM is not as sensitive as the RMM to these measurements. The estimated ellipse by the RMM is more flattened compared with the proposed image moment-based RHM. Besides, the rotation angles need to be provided beforehand for the RMM in [25], whereas the image moment-based RHM estimates the turning rate automatically during the estimation process. In real applications, it's impractical to obtain the turning rates in advance.

E. Simulation using Image-data

Another scenario where a car is driving through an intersection is simulated. This simulation is based on the real trajectory generated from a video clip. A short video clip from the Stanford drone dataset [49] is used, which shows a moving car from a bird's eye view. The video is captured with a 4k camera mounted on a quadcopter platform (a 3DR solo) which is hovering above an intersection on a university campus at an altitude of approximately 80 meters. The video clip contains 431 frames with an image size of 1422 by 1945 pixels, and has been undistorted and stabilized [49]. The ground truth is manually labeled at each frame and the measurement points are uniformly generated inside the bounding box of the ground truth. The number of measurements in each frame is generated based on a Poisson distribution with mean of 10. The sensor

noise is Gaussian white noise with variance $\text{diag}(10^2, 10^2)$ (with position in pixels and velocity in pixels/frame). In Fig. 9, the first top-view scene of the moving car is shown and 11 snapshots of the estimation results out of the 431 frames are plotted on the same figure. The car is switching between linear motions and rotational motions, which are approximated by CV motions and CT motions, respectively.

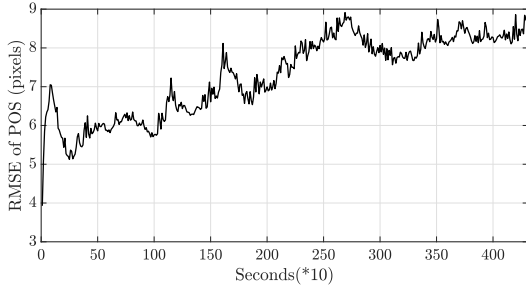
The CV model in (14) and the CT model in (20) with the UKF-IMM filter are applied to track the moving car. The power spectral density q for the process noise covariance in the CV model in (14) is set as 0.1 and $\mathbf{C}_{\text{IM}} = \text{diag}(0.01, 0.01, 0.01)$. For the CT model in (20), $\mathbf{w}_k = [0.01, 0.01, 0.01, 0.01, 0.01, (0.1^\circ)^2]^T$. The kinematic states (position and velocity in each dimension) are initialized as random variables with Gaussian distribution. Its mean is the ground truth and the covariance matrix is set as $\text{diag}(1, 0.02, 4, 0.08)$. The turning rate for the CT model is initialized as a random variable with distribution $\mathcal{N}(0, 1)$. The initial probability μ_0^j of the two models in the IMM filter is equal and the Markov chain transition matrix is selected to be $p_{ij} = \begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix}$. The proposed algorithm runs with 1000 Monte Carlo runs and the estimation results are shown in Fig. 10. The mean value of the RMSE of the centroid position over 1000 Monte Carlo runs is 7.28 pixels. The mean value of the IoU over 1000 Monte Carlo runs is 0.70. The ground truth is approximated as the smallest enclosing ellipse of the bounding box.

F. Discussion

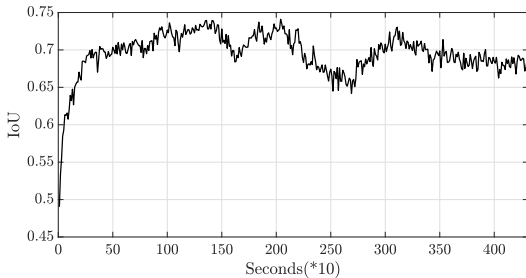
The proposed image moment-based RHM and its measurement and dynamic models are validated in the simulations of the static target, the target with the CV motion and the target with the CT motion. As the noise levels are increased, the size of the estimated elliptical shape doesn't increase. When the target is moving with the CV motion or the CT motion, the proposed algorithm predicts the position and velocity of the moving target, as well as the spatial extent and orientation. To estimate the target with complex trajectory,



Figure 9: Illustration of the estimation results in a particular run by the proposed image moment-based RHM. The target is initialized as a circle (green circle) with a radius of 20 pixels. The estimated result (yellow ellipse), the measurements (red crosses) and the ground truth (blue box) are shown for every 40 frames.



(a) Average RMSE of position



(b) Average IoU

Figure 10: Tracking results of the proposed image moment-based RHM algorithm over 1000 Monte Carlo runs.

the proposed image moment-based RHM is embedded within the IMM framework. The proposed average measurement log-likelihood function estimates the model probability accurately and consistently. The RMSE values of the position and velocity of the target's centroid are lower than the results from the compared RMM algorithm. The state of the RMM is the centroid's kinematic parameters and the random matrix, which is updated based on the mean and spread matrix of the measurement points [25]. The proposed RHM uses the centroid's kinematic parameters and the three image moments as the state variables, which are updated sequentially based on each individual measurement point. Later improvements will be focused on the block processing of measurements, and estimating the kinematic states and extent in two separate steps. The accurate dynamic model has the advantage in certain scenarios, especially when predicting the location of a target undergoing fast motion or when predicting with a relatively low sampling frequency.

VI. CONCLUSION

In this paper, the minimal, complete, and non-ambiguous representation of an elliptic object is modeled based on image moments for extended object tracking. The measurement model and the dynamic models of the image moments for CV motion and CT motion are analytically derived. The UKF and its combination with the IMM approach is applied for estimating the position, velocity and spatial extent based on the noisy measurement points uniformly generated from the extended target. The proposed image moment-based RHM and its filters are validated and evaluated in different simulation scenarios. The evaluation results show that the proposed model and its inference provide accurate estimations of the position, velocity and extent of the targets. The proposed image moment-based RHM method will be tested on real experiment data in applications, such as autonomous driving and air-drone following and tracking, and it will be embedded into other Bayesian based methods for multiple extended objects tracking as a part of future research.

APPENDIX A

TRANSITION MATRIX OF THE COORDINATED TURN MOTION

$$\dot{\mathbf{p}}_{\text{IM}} = \mathbf{B}\mathbf{p}_{\text{IM}} \quad (36)$$

where $\mathbf{B} = \begin{bmatrix} 0 & \omega & -\omega \\ -2\omega & 0 & 0 \\ 2\omega & 0 & 0 \end{bmatrix}$. The solution to this linear time-invariant state space equation (36) is

$$\mathbf{p}_{\text{IM}}(t) = e^{\mathbf{B}\tau} \mathbf{p}_{\text{IM}}(t_0) \quad (37)$$

in which $\tau = t - t_0$.

The interpolation polynomial method [42] is used to get the transition matrix of the dynamic equation $f(\lambda) = e^{\lambda\tau}$. Firstly, by solving $|\lambda\mathbf{I} - \mathbf{B}| = \lambda(\lambda^2 + 4\omega^2) = 0$, the eigenvalues of the matrix $\mathbf{B}$ are calculated as $\lambda_1 = 0$, $\lambda_2 = 2\omega j$ and $\lambda_3 = -2\omega j$. Then, a polynomial $g(\lambda) = \sum_{k=0}^2 g_k \lambda^k$ is found, which is equal to $f(\lambda) = e^{\lambda\tau}$ on the spectrum of $\mathbf{B}$, that is

$$\frac{\partial^j}{\partial \lambda^j} g(\lambda)|_{\lambda=\lambda_i} = \frac{\partial^j}{\partial \lambda^j} f(\lambda)|_{\lambda=\lambda_i} \quad (38)$$

in which $i = 1, \dots, 3$ and $j = 0$. The polynomial $g(\lambda)$ is calculated as

$$g(\lambda) = 1 + \frac{\sin(2\omega\tau)}{2\omega}\lambda + \frac{\sin^2(\omega\tau)}{2\omega^2}\lambda^2 \quad (39)$$

Then, $f(\mathbf{B}) = e^{\mathbf{B}\tau}$ is calculated by making it equal to $g(\mathbf{B})$.

The transition matrix $f(\mathbf{B}) = e^{\mathbf{B}\tau}$ is calculated as $e^{\mathbf{B}\tau} = \begin{bmatrix} \cos 2\theta & \frac{1}{2}\sin 2\theta & -\frac{1}{2}\sin 2\theta \\ -\sin 2\theta & \cos^2 \theta & \sin^2 \theta \\ \sin 2\theta & \sin^2 \theta & \cos^2 \theta \end{bmatrix}$, where $\theta = \omega\tau$.

APPENDIX B

DERIVATION AND MOMENT MATCHING OF THE MEASUREMENT MODEL NOISE TERM

The real measurement $\mathbf{z} = [\tilde{x}, \tilde{y}]^T$ of the unknown true measurement $\bar{\mathbf{z}} = [x, y]^T$ is expressed as $\mathbf{z} = \bar{\mathbf{z}} + \boldsymbol{\nu}$, where $\boldsymbol{\nu} = [\nu_x, \nu_y]^T$ is the additive white Gaussian noise with $\nu_x \sim \mathcal{N}(0, \sigma_x^2)$, $\nu_y \sim \mathcal{N}(0, \sigma_y^2)$. Replacing the unknown true measurement $\bar{\mathbf{z}}$ with the real measurement $\mathbf{z} = \bar{\mathbf{z}} + \boldsymbol{\nu}$ in (9) and separating the terms as

$$g(\mathbf{z}, \mathbf{p}) = g(\bar{\mathbf{z}}, \mathbf{p}) - f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p}) \quad (40)$$

where $f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})$ is the polynomial containing the white noise terms as

$$f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p}) = \rho \left[\nu_x^2 n_{02} + \nu_y^2 n_{20} + 2\nu_x \nu_y n_{11} + 2(n_{02}\nu_x - n_{11}\nu_y)(\tilde{x} - x_c) + 2(n_{20}\nu_y - n_{11}\nu_x)(\tilde{y} - y_c) \right] \quad (41)$$

where $\rho = 1/4(n_{20}n_{02} - n_{11}^2)$. The polynomial $f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})$ is approximated as a random variable with Gaussian distribution, which has the same mean and covariance as $f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})$ by moment matching. The closed-form expression of the first two moments of $f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})$ are

$$E[f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})] = \rho \left[n_{02}\sigma_x^2 + n_{20}\sigma_y^2 \right] \quad (42)$$

$$E[f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})^2] = \rho^2 \left\{ 3n_{02}^2\sigma_x^4 + 3n_{20}^2\sigma_y^4 + (2n_{02}n_{20} + 4n_{11}^2)\sigma_x^2\sigma_y^2 + 4[n_{02}(x - x_c) - n_{11}(y - y_c)]^2\sigma_x^2 + 4[n_{20}(y - y_c) - n_{11}(x - x_c)]^2\sigma_y^2 \right\} \quad (43)$$

The covariance of $f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})$ is derived as

$$\begin{aligned} C_{f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})} &= E[f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})^2] - E[f(\mathbf{z}, \boldsymbol{\nu}, \mathbf{p})]^2 \\ &= \rho^2 \left\{ 2n_{02}^2\sigma_x^4 + 2n_{20}^2\sigma_y^4 + 4n_{11}^2\sigma_x^2\sigma_y^2 + 4[n_{02}(x - x_c) - n_{11}(y - y_c)]^2\sigma_x^2 + 4[n_{20}(y - y_c) - n_{11}(x - x_c)]^2\sigma_y^2 \right\} \end{aligned} \quad (44)$$

REFERENCES

- [1] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4021–4029.
- [2] B. Zhong, H. Yao, S. Chen, R. Ji, T.-J. Chin, and H. Wang, "Visual tracking via weakly supervised learning from multiple imperfect oracles," *Pattern Recognition*, vol. 47, no. 3, pp. 1395–1410, 2014.
- [3] B. Zhong, B. Bai, J. Li, Y. Zhang, and Y. Fu, "Hierarchical tracking by reinforcement learning-based searching and coarse-to-fine verifying," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2331–2341, 2018.
- [4] Q. Zhou, B. Zhong, Y. Zhang, J. Li, and Y. Fu, "Deep alignment network based multi-person tracking with occlusion and motion reasoning," *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1183–1194, 2018.
- [5] Y. Han, C. Deng, B. Zhao, and B. Zhao, "Spatial-temporal context-aware tracking," *IEEE Signal Processing Letters*, vol. 26, no. 3, pp. 500–504, 2019.
- [6] Y. Han, C. Deng, B. Zhao, and D. Tao, "State-aware anti-drift object tracking," *IEEE Transactions on Image Processing*, 2019.
- [7] G. Liu, "Robust visual tracking via smooth manifold kernel sparse learning," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 2949–2963, 2018.
- [8] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1396–1404.
- [9] G. Yao and A. Dani, "Visual tracking using sparse coding and earth mover's distance," *Frontiers in Robotics and AI*, 2018.
- [10] G. Yao, M. Williams, and A. Dani, "Gyro-aided visual tracking using iterative Earth Mover's Distance," in *19th Conference on Information Fusion (FUSION)*, 2016, pp. 2317–2323.
- [11] 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.
- [12] L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. K. Pang, and S. Godsill, "Overview of Bayesian sequential monte carlo methods for group and extended object tracking," *Digital Signal Processing*, vol. 25, pp. 1–16, 2014.
- [13] Y. Bar-Shalom, H. M. Shertukde, and K. R. Pattipati, "Precision target tracking for small extended objects," *Optical Engineering*, vol. 29, no. 2, pp. 121–127, 1990.
- [14] K. Granström, P. Willett, and Y. Bar-Shalom, "An extended target tracking model with multiple random matrices and unified kinematics," in *18th International Conference on Information Fusion (Fusion)*, 2015, pp. 1007–1014.
- [15] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Tracking of extended objects with high-resolution doppler radar," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1341–1353, 2016.
- [16] D. Chwa, A. P. Dani, and W. E. Dixon, "Range and motion estimation of a monocular camera using static and moving objects," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 4, pp. 1174–1183, 2016.
- [17] A. P. Dani, N. R. Fischer, and W. E. Dixon, "Single camera structure and motion," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 238–243, 2012.
- [18] J. Yang, A. Dani, S.-J. Chung, and S. Hutchinson, "Vision-based localization and robot-centric mapping in riverine environments," *Journal of Field Robotics*, vol. 34, no. 3, pp. 429–450, 2017.
- [19] S. Mehta and T. Burks, "Vision-based control of robotic manipulator for citrus harvesting," *Computers and Electronics in Agriculture*, vol. 102, pp. 146–158, 2014.
- [20] S. S. Mehta, W. MacKunis, and T. F. Burks, "Robust visual servo control in the presence of fruit motion for robotic citrus harvesting," *Computers and Electronics in Agriculture*, vol. 123, pp. 362–375, 2016.
- [21] M. Baum, F. Faion, and U. D. Hanebeck, "Tracking ground moving extended objects using RGB-D data," in *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2012, pp. 186–191.
- [22] J. Vermaak, N. Ikoma, and S. J. Godsill, "Sequential monte carlo framework for extended object tracking," *IEEE Proceedings-Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 353–363, 2005.
- [23] K. Granström, S. Reuter, D. Meissner, and A. Scheel, "A multiple model PHD approach to tracking of cars under an assumed rectangular shape," in *17th International Conference on Information Fusion (FUSION)*, 2014, pp. 1–8.

- [24] A. Zea, F. Faion, and U. D. Hanebeck, "Tracking elongated extended objects using splines," in *19th International Conference on Information Fusion (FUSION)*, 2016, pp. 612–619.
- [25] J. Lan and X. R. Li, "Tracking of extended object or target group using random matrix: new model and approach," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2973–2989, 2016.
- [26] K. Granström and U. Orguner, "New prediction for extended targets with random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1577–1589, 2014.
- [27] M. Feldmann, D. Franken, 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.
- [28] 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.
- [29] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, 2008.
- [30] P. Zong and M. Barbary, "Improved multi-bernoulli filter for extended stealth targets tracking based on sub-random matrices," *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1428–1447, 2016.
- [31] 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.
- [32] S. Yang and M. Baum, "Second-order extended Kalman filter for extended object and group tracking," in *19th International Conference on Information Fusion (FUSION)*, 2016, pp. 1178–1184.
- [33] M. Baum, M. Feldmann, D. Fränken, U. D. Hanebeck, and W. Koch, "Extended object and group tracking: A comparison of random matrices and random hypersurface models," in *GI Jahrestagung (2)*. Citeseer, 2010, pp. 904–906.
- [34] Y. Bar Shalom, P. K. Willett, and X. Tian, *Tracking and data fusion*. YBS publishing, 2011.
- [35] J. Lan and X. R. Li, "Tracking of extended object or target group using random matrix-part I: New model and approach," in *15th Conference on Information Fusion (FUSION)*, 2012, pp. 2177–2184.
- [36] R. Spica, P. R. Giordano, and F. Chaumette, "Plane estimation by active vision from point features and image moments," in *IEEE International Conference on Robotics and Automation*, 2015, pp. 6003–6010.
- [37] A. Dani, G. Panahandeh, S.-J. Chung, and S. Hutchinson, "Image moments for higher-level feature based navigation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2013, pp. 602–609.
- [38] O. Tahri and F. Chaumette, "Point-based and region-based image moments for visual servoing of planar objects," *IEEE Transactions on Robotics*, vol. 21, no. 6, pp. 1116–1127, 2005.
- [39] F. Chaumette, "Image moments: a general and useful set of features for visual servoing," *IEEE Transactions on Robotics*, vol. 20, no. 4, pp. 713–723, 2004.
- [40] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.
- [41] K. Granstrom, P. Willett, and Y. Bar-Shalom, "Systematic approach to IMM mixing for unequal dimension states," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 4, pp. 2975–2986, 2015.
- [42] Y. Bar Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.
- [43] T. Kirubarajan and Y. Bar-Shalom, "Kalman filter versus IMM estimator: when do we need the latter?" *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1452–1457, 2003.
- [44] G. Yao, K. Hunte, and A. Dani, "Image moment-based object tracking and shape estimation for complex motions," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 5819–5824.
- [45] G. Yao and A. Dani, "Image moment-based random hypersurface model for extended object tracking," in *20th Conference on Information Fusion (FUSION)*, 2017, pp. 1–7.
- [46] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. part I. Dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.
- [47] N. Wahlström and E. Özkan, "Extended target tracking using Gaussian processes," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4165–4178, 2015.
- [48] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.
- [49] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *European conference on computer vision*. Springer, 2016, pp. 549–565.