Object Tracking Incorporating Transfer Learning into an Unscented Kalman Filter

Omar Alotaibi and Brian L. Mark
Dept. of Electrical and Computer Engineering
George Mason University
Fairfax, Virginia, U.S.A.
oalotaib@gmu.edu, bmark@gmu.edu

***Abstract*—We present a novel algorithm designed to address the challenges posed by mismatched intensity of the noise in sensors performing object tracking. Our objective is to enhance the accuracy of estimation in the tracking domain, particularly in scenarios where reliable measurements are difficult to obtain due to environmental conditions affecting a specific sensor. To accomplish this, we propose a framework that integrates transfer learning techniques into an unscented Kalman filter (UKF). We introduce an additional step to model and learn the parameters of predicted observations in a learning domain at each time step. By incorporating the learned knowledge from the learning domain into the filtering process of the tracking domain, our approach demonstrates significant improvements in tracking accuracy. Through extensive simulations, we validate the effectiveness of our proposed algorithm in terms of tracking accuracy, comparing its performance to that of the traditional isolated UKF.**

***Index Terms*—Object tracking, transfer learning, Unscented Kalman filter.**

I. Introduction

It is widely assumed in the field of object tracking that all sensors used to track an object operate under the same environmental conditions. However, in practice, this assumption may not always be valid because each sensor may encounter its own environmental conditions, resulting in variations in the accuracy of estimating an object's state across sensors. Recent research in the field of object tracking has investigated the use of machine learning techniques, particularly transfer learning, which shows promise in improving estimation accuracy by leveraging knowledge from tasks or domains. Transfer learning, for example, has been successfully used to address sparsity issues in filtering, as demonstrated in [8], [4]. Transfer learning was used in [2] to combine two Kalman filters and improve estimation accuracy performance. Furthermore, in [11], Bayesian models have been used in transfer learning approaches for probabilistic graphical models, such as sharing a Gaussian prior.

This paper proposes an algorithm that incorporates the transfer learning technique into the unscented Kalman filter (UKF) framework, with the objective of improving the estimation accuracy of object state tracking in cases where sensors observe measurements with mismatched noise intensity due to technical or environmental factors among the sensors under consideration. In order to achieve this, we perform an additional stage in which we model a predicted observation density in a learning domain and learn the density parameters, such as mean and covariance. The parameters that have been learned are transferred and used in a tracking domain process. Our simulation results show that incorporating transfer learning into the UKF framework improves object state estimation accuracy significantly when compared to the traditional isolated UKF under the same conditions. These results indicate that the proposed algorithm has the potential to improve the performance accuracy of tracking object state in real-world scenarios.

The paper is organized as follows. In Section II, we provide a brief overview of the state space representation and traditional Bayesian filtering approach. We also describe the problem formulation for object tracking using two sensors with mismatched intensity levels of measurement noise. Section III introduces the framework for incorporating transfer learning into the Bayesian filtering approach, with an emphasis on knowledge utilization and transfer between the learning and tracking domains. Section IV discusses the proposed algorithm for integrating transfer learning into the UKF in the learning and tracking domains. Simulation results for the proposed algorithm are presented in Section V. The paper is concluded in Section VI.

II. System Model

A. State Space Representation

A discrete-time state space representation (SSR) is a mathematical framework utilized for describing the dynamics of a general dynamical system as follows:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1} \tag{1}$$
$$\mathbf{z}_k = h_k(\mathbf{x}_k) + \mathbf{w}_k, \tag{2}$$

where $k \in \mathbb{N}$ denotes the discrete time step. The vector $\mathbf{x}_k \in \mathbb{R}^{n_\mathbf{x}}$ represents the state of the system at time step k with a dimension of $n_\mathbf{x}$. The function $f_k \in \mathbb{R}^{n_\mathbf{x}}$ is the dynamic transition function, which describes how the state evolves over time. The vector $\mathbf{z}_k \in \mathbb{R}^{n_\mathbf{z}}$ represents the measurements obtained from the system at time step k with a dimension of $n_\mathbf{z}$. The function $h_k \in \mathbb{R}^{n_\mathbf{z}}$ is the measurement function, which relates the state to the measurements. The variables $\mathbf{v}_{k-1} \in \mathbb{R}^{n_\mathbf{v}}$ and $\mathbf{w}_k \in \mathbb{R}^{n_\mathbf{w}}$ correspond to the state process noise and measurement noise with dimensions of $n_\mathbf{v}$ and $n_\mathbf{w}$, respectively, representing the uncertainties and disturbances in the system [5]. The state and measurement noises are assumed to be independent and identically distributed (i.i.d) zero-mean Gaussian random vector sequences characterized by

$$\mathbf{v}_{k-1} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}_\mathbf{v}), \quad \mathbf{w}_k \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}_\mathbf{w}), \tag{3}$$

respectively, where $\mathbf{Q}_\mathbf{v}$ and $\mathbf{Q}_\mathbf{w}$ denote the covariances.

B. Bayesian Filter Approach

In the Bayesian filter approach, the posterior probability density function (PDF) of the state is constructed by incorporating all available statistical information and the sequence of observations. The extraction of the estimated state from the posterior PDF provides an optimal solution for addressing the estimation problem [9], [3]. The state and measurement models described by (1) and (2) are Gaussian due to the assumption of Gaussian noise sources as given by (3). This Gaussian process assumption leads to a posterior density for the state that is also Gaussian, characterized by its mean and covariance. Consequently, the recursive nature of the Bayesian filter approach primarily revolves around the update of means and covariances of conditional densities over time and measurements. The two main steps of the Bayesian filter approach, namely the prediction and update steps, can be reformulated under the assumption of a Gaussian process.

C. Tracking Model with Mismatched Noise Intensity of Sensors

We consider the problem of estimating an unknown object state $\mathbf{x}_k$ at time step k, assuming a state transition model defined as

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1} \implies p(\mathbf{x}_k \mid \mathbf{x}_{k-1}), \tag{4}$$

where $f(\cdot)$ denotes the state transition function, which models the transition of the unknown state at each time step, and $\mathbf{w}_{k-1}$ is a process noise that represents the error process in the transition model, which is assumed to be zero-mean additive white Gaussian noise (AWGN) with covariance $\mathbf{Q}_\mathbf{v}$. At each time step, a sensor observes measurements $\mathbf{z}_k$ that are associated with the desired object. The observable measurements are modeled via

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{w}_k \implies p(\mathbf{z}_k \mid \mathbf{x}_k), \tag{5}$$

where $h(\cdot)$ provides a representation of the measurement parameters of the unknown state parameters. The measurement noise $\mathbf{w}_k$ is assumed to be Gaussian with a covariance of $\mathbf{Q}_\mathbf{w} = I_\mathbf{w}\, \mathbf{B}_\mathbf{w}$, where $I_\mathbf{w}$ denotes the noise intensity.

In our tracking formulation scenario, the task involves multiple sensors tracking the same object. However, the observed measurements have mismatched intensities $I_\mathbf{w}$ among all sensors due to various environmental or technical factors that affect the observable measurements, leading to a reduction in estimation accuracy. To address this issue, we focus on a specific case of two sensors, learning and tracking the same moving object as illustrated in Figure 1, where both sensors are affected by the noise intensities, denoted as $I^\star_\mathbf{w}$ and $I_\mathbf{w}$, correspondingly.

III. Bayesian Transfer Learning Filter Approach

To enhance estimation accuracy in domains characterized by unreliable observed information, our framework incorporates Bayesian transfer learning. Bayesian transfer learning is employed to model the joint distribution between a learning domain and a tracking domain [7], [11]. This approach is particularly advantageous in scenarios where observed measurements are limited due to adverse conditions, such as high levels of noise.

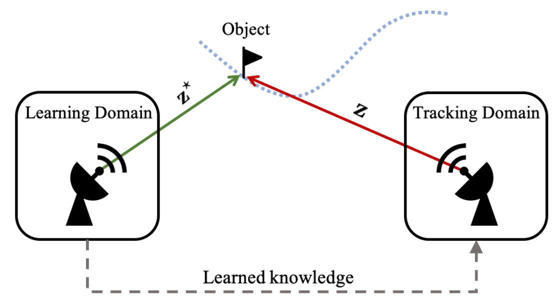

Fig. 1. A graphical illustration of knowledge transfer between two sensors tracking the same moving object.

A. Learning Domain

The learning domain comprises an unknown state variable set $\mathbf{x}^\star = \{\mathbf{x}^\star_1, \cdots, \mathbf{x}^\star_K\}$ and an observable measurement variable set $\mathbf{z}^\star = \{\mathbf{z}^\star_1, \cdots, \mathbf{z}^\star_K\}$. Unlike the conventional Bayesian approach, the predicted observation variable set $\boldsymbol{\eta}^\star = \{\boldsymbol{\eta}^\star_2, \cdots, \boldsymbol{\eta}^\star_{K+1}\}$ is required to be introduced in Bayesian transfer learning. The overall posterior density of the state object and predicted observation $p(\mathbf{x}^\star_k, \boldsymbol{\eta}^\star_{k+1} \mid \mathbf{z}^\star_{1:k})$, given the measurements set up to time step k, denoted as $\mathbf{z}^\star_{1:k} = \{\mathbf{z}^\star_1, \mathbf{z}^\star_2, \ldots, \mathbf{z}^\star_k\}$, can be estimated by using two posterior densities, as follows:

$$p(\mathbf{x}^\star_k, \boldsymbol{\eta}^\star_{k+1} \mid \mathbf{z}^\star_{1:k}) \propto\ p(\mathbf{x}^\star_k \mid \boldsymbol{\eta}^\star_{k+1}, \mathbf{z}^\star_{1:k})\ p(\boldsymbol{\eta}^\star_{k+1} \mid \mathbf{z}^\star_{1:k}). \tag{6}$$

Since the object state $\mathbf{x}^\star_k$ at the current time step k is conditionally independent of the predicted observation variable $\boldsymbol{\eta}^\star_{k+1}$ for the next time step $k+1$, the conditional probability $p(\mathbf{x}^\star_k \mid \boldsymbol{\eta}^\star_{k+1}, \mathbf{z}^\star_{1:k})$ in (6) can be simplified to $p(\mathbf{x}^\star_k \mid \mathbf{z}^\star_{1:k})$. The posterior density of the object state $p(\mathbf{x}^\star_k \mid \mathbf{z}^\star_{1:k})$ can be expressed as

$$p(\mathbf{x}^\star_k \mid \mathbf{z}^\star_{1:k}) \propto p(\mathbf{z}^\star_k \mid \mathbf{x}^\star_k)\, p(\mathbf{x}^\star_k \mid \mathbf{x}^\star_{k-1})\, p(\mathbf{x}^\star_{k-1} \mid \mathbf{z}^\star_{1:k-1}), \tag{7}$$

where $p(\mathbf{z}^\star_k \mid \mathbf{x}^\star_k)$ represents the measurement likelihood function, and $p(\mathbf{x}^\star_k \mid \mathbf{x}^\star_{k-1})$ denotes the predicted state for the next time step, computed using the transition PDF given in (4). Upon obtaining the estimated object state $\mathbf{x}^\star_k$ from (7), the posterior density of the predicted observation $p(\boldsymbol{\eta}^\star_{k+1} \mid \mathbf{z}^\star_{1:k})$ as described in (6) can be estimated by

$$p(\boldsymbol{\eta}^\star_{k+1} \mid \mathbf{z}^\star_{1:k}) \propto p(\boldsymbol{\eta}^\star_{k+1} \mid \mathbf{x}^\star_{k+1})\ p(\mathbf{x}^\star_{k+1} \mid \mathbf{x}^\star_k)\ p(\mathbf{x}^\star_k \mid \mathbf{z}^\star_{1:k}), \tag{8}$$

where $p(\boldsymbol{\eta}^\star_{k+1} \mid \mathbf{x}^\star_{k+1})$ follows the measurement density model described in (5).

B. Tracking Domain

The learned predicted observations $\boldsymbol{\eta}^\star_{2:k}$ are simultaneously transferred and used as prior knowledge for estimating a set of object state variables $\mathbf{x} = \{\mathbf{x}_1, ..., \mathbf{x}_K\}$ under different conditions in the tracking domain given a set of measurement variables $\mathbf{z} = \{\mathbf{z}_1, ..., \mathbf{z}_K\}$. By integrating the Bayesian transfer learning into the framework, the overall posterior density of the object state given the learned predicted observations $\boldsymbol{\eta}^\star_{2:k}$ and the observed measurements $\mathbf{z}_{1:k}$ up to time step k is estimated as follows:

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k}, \boldsymbol{\eta}^\star_{2:k}) \propto p(\mathbf{z}_k \mid \mathbf{x}_k, \mathbf{z}_{1:k-1}, \boldsymbol{\eta}^\star_{2:k})\ p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1}, \boldsymbol{\eta}^\star_{2:k}). \tag{9}$$

The transfer learning framework's state posterior density, denoted as $p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1}, \boldsymbol{\eta}^\star_{2:k})$, is computed via Bayes' rule as

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1}, \boldsymbol{\eta}^\star_{2:k}) \propto p(\boldsymbol{\eta}^\star_k \mid \mathbf{x}_k, \mathbf{z}_{1:k-1}, \boldsymbol{\eta}^\star_{2:k-1}) \, p(\mathbf{x}_k \mid \mathbf{x}_{k-1}). \tag{10}$$

Under the assumption that the measurement at the current time step k is conditionally independent of all previous measurements, given the current state $\mathbf{x}_k$, and that the measurements $\mathbf{z}$ and the transferred learned predicted observations $\boldsymbol{\eta}^\star$ are independent, conditioned on the current state $\mathbf{x}_k$, the overall state posterior can be simplified and obtained by substituting the transfer learning state posterior from (10) into (9):

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k}, \boldsymbol{\eta}^\star_{2:k}) \propto p(\mathbf{z}_k \mid \mathbf{x}_k) \, p(\boldsymbol{\eta}^\star_k \mid \mathbf{x}_k) \, p(\mathbf{x}_k \mid \mathbf{x}_{k-1}). \tag{11}$$

The estimation of the overall posterior, as described in (11), involves two distinct likelihood functions: the transferred predicted observation likelihood function $p(\boldsymbol{\eta}^\star_k \mid \mathbf{x}_k)$ and the measurement likelihood function $p(\mathbf{z}_k \mid \mathbf{x}_k)$. By incorporating the transferred knowledge of the predicted observation, the overall posterior density can be estimated, which leads to enhanced accuracy in the object state estimation process.

IV. Transfer Learning for UKF

The UKF [6] uses a set of sigma points to approximate a nonlinear function, unlike the extended Kalman filter (EKF) which relies on linearization. In situations where analytical Jacobians of the transition and measurement model functions are unavailable, this feature gives the UKF an advantage over the EKF [10]. By integrating the Bayesian transfer learning approach in Section III into the UKF framework, knowledge can be transferred from the learning domain to the tracking domain and used as a prior to be incorporated in the estimation process to improve the accuracy of object state estimation.

A. Learning Domain

By applying the unscented transformation (UT) method [6], the set of $2n_{\mathbf{x}} + 1$ sigma points are defined as:

$$\boldsymbol{\mathcal{X}}^\star_{0,k-1|k-1} = \hat{\mathbf{x}}^\star_{k-1|k-1} \tag{12}$$

$$\boldsymbol{\mathcal{X}}^\star_{j,k-1|k-1} = \hat{\mathbf{x}}^\star_{k-1|k-1} + \sqrt{(n_{\mathbf{x}} + \lambda)\mathbf{P}^\star_{j,k-1|k-1}} \tag{13}$$

$$\boldsymbol{\mathcal{X}}^\star_{n_{\mathbf{x}}+j,k-1|k-1} = \hat{\mathbf{x}}^\star_{k-1|k-1} - \sqrt{(n_{\mathbf{x}} + \lambda)\mathbf{P}^\star_{j,k-1|k-1}}, \tag{14}$$

where $j = 1, \ldots, n_{\mathbf{x}}$ and $\lambda = \alpha^2 \, (n_{\mathbf{x}} + \kappa) - n_{\mathbf{x}}$ is a scaling parameter with $10^{-4} \le \alpha \le 1$. The value of κ is defined by $\kappa = 3 - n_{\mathbf{x}}$, as suggested in [6]. Additionally, $\mathbf{P}^\star_{j,k-1|k-1}$ refers to the ith column of the covariance matrix. The associated weights of these sigma points are obtained as:

$$W^\star_{0,k} = \frac{\lambda}{n_{\mathbf{x}} + \lambda}, \quad W^\star_{j,k} = \frac{1}{2n_{\mathbf{x}} + 2\lambda}, \tag{15}$$

for $j = 1, 2, \ldots, 2n_{\mathbf{x}}$.

• **Prediction Step:** The mean $\hat{\mathbf{x}}^\star_{k|k-1}$ and the associated covariance $\mathbf{P}^\star_{k|k-1}$ for the predictive density can be estimated as:

$$\hat{\mathbf{x}}^\star_{k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W^\star_{j,k} \, f\left(\boldsymbol{\mathcal{X}}^\star_{j,k-1|k-1}\right) \tag{16}$$

$$\mathbf{P}^\star_{k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W^\star_{j,k} \left[f\left(\boldsymbol{\mathcal{X}}^\star_{j,k-1|k-1}\right) - \hat{\mathbf{x}}^\star_{k|k-1}\right] \left[f\left(\boldsymbol{\mathcal{X}}^\star_{j,k-1|k-1}\right) - \hat{\mathbf{x}}^\star_{k|k-1}\right]^T + \mathbf{Q}^\star_{\mathbf{v}}. \tag{17}$$

By applying the nonlinear transition model function to the sigma points, the predicted sigma points $\boldsymbol{\mathcal{X}}^\star_{j,k|k-1}$ can be obtained as

$$\boldsymbol{\mathcal{X}}^\star_{j,k|k-1} = f\left(\boldsymbol{\mathcal{X}}^\star_{j,k-1|k-1}\right). \tag{18}$$

• **Update Step:** The predicted measurement mean $\hat{\mathbf{z}}^\star_{k|k-1}$ and the associated covariance $\mathbf{P}^\star_{\mathbf{zz},k|k-1}$ can be computed as

$$\hat{\mathbf{z}}^\star_{k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W^\star_{j,k} \, h\left(\boldsymbol{\mathcal{X}}^\star_{j,k|k-1}\right) \tag{19}$$

$$\mathbf{P}^\star_{\mathbf{zz},k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W^\star_{j,k} \left[h\left(\boldsymbol{\mathcal{X}}^\star_{j,k|k-1}\right) - \hat{\mathbf{z}}^\star_{k|k-1}\right] \left[h\left(\boldsymbol{\mathcal{X}}^\star_{j,k|k-1}\right) - \hat{\mathbf{z}}^\star_{k|k-1}\right]^T + \mathbf{Q}^\star_{\mathbf{w}}. \tag{20}$$

The joint conditional density's cross-covariance can be computed as follows:

$$\mathbf{P}^\star_{\mathbf{xz},k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W^\star_{j,k} \left[\boldsymbol{\mathcal{X}}^\star_{j,k|k-1} - \hat{\mathbf{x}}^\star_{k|k-1}\right] \left[h\left(\boldsymbol{\mathcal{X}}^\star_{j,k|k-1}\right) - \hat{\mathbf{z}}^\star_{k|k-1}\right]^T. \tag{21}$$

After observing the new measurement, the posterior density of the object state can be estimated with the mean $\hat{\mathbf{x}}^\star_{k|k}$ and covariance $\mathbf{P}^\star_{k|k}$, computed as follows:

$$\hat{\mathbf{x}}^\star_{k|k} = \hat{\mathbf{x}}^\star_{k|k-1} + \mathbf{K}^\star_k(\mathbf{z}^\star_k - \hat{\mathbf{z}}^\star_{k|k-1}) \tag{22}$$

$$\mathbf{P}^\star_{k|k} = \mathbf{P}^\star_{k|k-1} - \mathbf{K}^\star_k \, \mathbf{P}^\star_{\mathbf{zz},k|k-1} \, (\mathbf{K}^\star_k)^T, \tag{23}$$

where the Kalman gain $\mathbf{K}^\star_k$ is computed as

$$\mathbf{K}^\star_k = \mathbf{P}^\star_{\mathbf{xz},k|k-1} \, (\mathbf{P}^\star_{\mathbf{zz},k|k-1})^{-1}. \tag{24}$$

• **Learning Step:** As the predicted observation variable set $\boldsymbol{\eta}^\star$ is introduced in the Bayesian transfer learning, the posterior density of the predicted observation $p(\boldsymbol{\eta}^\star_{k+1} \mid \mathbf{z}^\star_{1:k})$ in (8) is a Gaussian vector $\boldsymbol{\eta}^\star_{k+1} \sim \mathcal{N}(\hat{\boldsymbol{\eta}}^\star_{k+1|k}, \mathbf{P}^\star_{\boldsymbol{\eta\eta},k+1|k})$ with mean

$\hat{\boldsymbol{\eta}}^{\star}_{k+1|k}$ and associated covariance $\mathbf{P}^{\star}_{\boldsymbol{\eta}\boldsymbol{\eta},k+1|k}$ learned using following the procedure:

$$\boldsymbol{\mathcal{X}}^{\star}_{0,k|k} = \hat{\mathbf{x}}^{\star}_{k|k} \tag{25}$$

$$\boldsymbol{\mathcal{X}}^{\star}_{j,k|k} = \hat{\mathbf{x}}^{\star}_{k|k} + \sqrt{(n_{\mathbf{x}}+\lambda)\mathbf{P}^{\star}_{j,k|k}} \tag{26}$$

$$\boldsymbol{\mathcal{X}}^{\star}_{n_{\mathbf{x}}+j,k|k} = \hat{\mathbf{x}}^{\star}_{k|k} - \sqrt{(n_{\mathbf{x}}+\lambda)\mathbf{P}^{\star}_{j,k|k}} \tag{27}$$

$$\boldsymbol{\mathcal{X}}^{\star}_{j,k+1|k} = f\left(\boldsymbol{\mathcal{X}}^{\star}_{j,k|k}\right) \tag{28}$$

$$\hat{\boldsymbol{\eta}}^{\star}_{k+1|k} = \sum_{j=0}^{2n_{\mathbf{x}}} W^{\star}_{j,k}\ h\left(\boldsymbol{\mathcal{X}}^{\star}_{j,k+1|k}\right), \tag{29}$$

where $j = 1, 2, ..., n_{\mathbf{x}}$ and

$$\mathbf{P}^{\star}_{\boldsymbol{\eta}\boldsymbol{\eta},k+1|k} = \sum_{j=0}^{2n_{\mathbf{x}}} W^{\star}_{j,k}\ \left[h\left(\boldsymbol{\mathcal{X}}^{\star}_{j,k+1|k}\right) - \hat{\boldsymbol{\eta}}^{\star}_{k+1|k}\right] \left[h\left(\boldsymbol{\mathcal{X}}^{\star}_{j,k+1|k}\right) - \hat{\boldsymbol{\eta}}^{\star}_{k+1|k}\right]^T + \mathbf{Q}^{\star}_{\mathbf{w}}. \tag{30}$$

The posterior density of the predicted observation $p(\boldsymbol{\eta}^{\star}_{k+1} \mid \mathbf{z}^{\star}_{1:k})$ is Gaussian with mean and covariance estimated in (29) and (30), respectively. The learned parameters $\hat{\boldsymbol{\eta}}^{\star}_{k+1|k}$ and $\mathbf{P}^{\star}_{\boldsymbol{\eta}\boldsymbol{\eta},k+1|k}$ are transferred to the tracking domain and utilized as a prior for improving the tracking performance.

B. Tracking Domain

Equivalently to the learning domain, the set of $2n_{\mathbf{x}}+1$ sigma points and their associated weights in the tracking domain can be drawn and computed as follows:

$$\boldsymbol{\mathcal{X}}_{0,k-1|k-1} = \hat{\mathbf{x}}_{k-1|k-1} \tag{31}$$

$$\boldsymbol{\mathcal{X}}_{j,k-1|k-1} = \hat{\mathbf{x}}_{k-1|k-1} + \sqrt{(n_{\mathbf{x}}+\lambda)\mathbf{P}_{j,k-1|k-1}} \tag{32}$$

$$\boldsymbol{\mathcal{X}}_{n_{\mathbf{x}}+j,k-1|k-1} = \hat{\mathbf{x}}_{k-1|k-1} - \sqrt{(n_{\mathbf{x}}+\lambda)\mathbf{P}_{j,k-1|k-1}} \tag{33}$$

$$W_{0,k} = \frac{\lambda}{n_{\mathbf{x}}+\lambda}, \quad W_{j,k} = \frac{1}{2n_{\mathbf{x}}+2\lambda}, \tag{34}$$

where $j = 1, 2, ..., n_{\mathbf{x}}$.

• **Prediction Step:** The predictive density mean $\hat{\mathbf{x}}_{k|k-1}$ and covariance $\mathbf{P}_{k|k-1}$ can be estimated as follows:

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k}\ f\left(\boldsymbol{\mathcal{X}}_{j,k-1|k-1}\right) \tag{35}$$

$$\mathbf{P}_{k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k}\ \left[f\left(\boldsymbol{\mathcal{X}}_{j,k-1|k-1}\right) - \hat{\mathbf{x}}_{k|k-1}\right] \left[f\left(\boldsymbol{\mathcal{X}}_{j,k-1|k-1}\right) - \hat{\mathbf{x}}_{k|k-1}\right]^T + \mathbf{Q}_{\mathbf{v}}. \tag{36}$$

The drawn sigma points in (33) can be transitioned using the nonlinear transition function as

$$\boldsymbol{\mathcal{X}}_{j,k|k-1} = f\left(\boldsymbol{\mathcal{X}}_{j,k-1|k-1}\right) \tag{37}$$

• **Update Step:** The predicted object state density in the tracking domain is updated using two likelihood function by integrating the transfer learning, as represented in (11), into the UKF framework. In the initial stage, the transferred predicted observation likelihood function $p(\boldsymbol{\eta}^{\star}_{k} \mid \mathbf{x}_k)$ is employed to update the predicted object state density. This update process incorporates the transferred learned parameters, $\hat{\boldsymbol{\eta}}^{\star}_{k|k-1}$ and $\mathbf{P}^{\star}_{\boldsymbol{\eta}\boldsymbol{\eta},k|k-1}$, which were learned in the previous time step $k-1$ in the learning domain, into the process. The predicted mean $\hat{\boldsymbol{\eta}}_{k|k-1}$ and covariance $\mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta},k|k-1}$ for the transferred learned observation can be computed as

$$\hat{\boldsymbol{\eta}}_{k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k}\ h\left(\boldsymbol{\mathcal{X}}_{j,k|k-1}\right) \tag{38}$$

$$\mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta},k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k}\ \left[h\left(\boldsymbol{\mathcal{X}}_{j,k|k-1}\right) - \hat{\boldsymbol{\eta}}_{k|k-1}\right] \left[h\left(\boldsymbol{\mathcal{X}}_{j,k|k-1}\right) - \hat{\boldsymbol{\eta}}_{k|k-1}\right]^T + \mathbf{P}^{\star}_{\boldsymbol{\eta}\boldsymbol{\eta},k|k-1}. \tag{39}$$

The cross-covariance, which characterizes the relationship between the state and the transferred learned parameter, is defined as

$$\mathbf{P}_{\mathbf{x}\boldsymbol{\eta},k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k}\ \left[\boldsymbol{\mathcal{X}}_{j,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right] \left[h\left(\boldsymbol{\mathcal{X}}_{j,k|k-1}\right) - \hat{\boldsymbol{\eta}}_{k|k-1}\right]^T. \tag{40}$$

Given the transferred learned parameter $\boldsymbol{\eta}^{\star}_{k|k-1}$, the posterior density of the state in the transfer learning framework, denoted as $p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1}, \boldsymbol{\eta}^{\star}_{2:k}) \sim \mathcal{N}(\hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1}, \mathbf{P}^{\boldsymbol{\eta}}_{k|k-1})$, is estimated with a mean $\hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1}$ and a covariance $\mathbf{P}^{\boldsymbol{\eta}}_{k|k-1}$ as

$$\hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}^{\boldsymbol{\eta}}_k\ (\hat{\boldsymbol{\eta}}^{\star}_{k|k-1} - \hat{\boldsymbol{\eta}}_{k|k-1}) \tag{41}$$

$$\mathbf{P}^{\boldsymbol{\eta}}_{k|k-1} = \mathbf{P}_{k|k-1} - \mathbf{K}^{\boldsymbol{\eta}}_k\ \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta},k|k-1}\ (\mathbf{K}^{\boldsymbol{\eta}}_k)^T, \tag{42}$$

where the Kalman gain of the transfer learning framework $\mathbf{K}^{\boldsymbol{\eta}}_k$ is given by

$$\mathbf{K}^{\boldsymbol{\eta}}_k = \mathbf{P}_{\mathbf{x}\boldsymbol{\eta},k|k-1}\ (\mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta},k|k-1})^{-1} \tag{43}$$

Based on the estimated mean and covariance of the transfer learning state posterior density computed in (41) and (42) respectively, the set of $2n_{\mathbf{x}}+1$ sigma points will be redrawn as:

$$\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{0,k|k-1} = \hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1} \tag{44}$$

$$\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{j,k|k-1} = \hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1} + \sqrt{(n_{\mathbf{x}}+\lambda)\mathbf{P}^{\boldsymbol{\eta}}_{j,k|k-1}} \tag{45}$$

$$\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{n_{\mathbf{x}}+j,k|k-1} = \hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1} - \sqrt{(n_{\mathbf{x}}+\lambda)\mathbf{P}^{\boldsymbol{\eta}}_{j,k|k-1}}, \tag{46}$$

where $j = 1, 2, ..., n_{\mathbf{x}}$. The estimated measurement mean and covariance can be calculated as follows:

$$\hat{\mathbf{z}}_{k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k}\ h\left(\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{j,k|k-1}\right) \tag{47}$$

$$\mathbf{P}_{\mathbf{z}\mathbf{z},k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k}\ \left[h\left(\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{j,k|k-1}\right) - \hat{\mathbf{z}}_{k|k-1}\right] \left[h\left(\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{j,k|k-1}\right) - \hat{\mathbf{z}}_{k|k-1}\right]^T + \mathbf{Q}_{\mathbf{w}}. \tag{48}$$

The cross-covariance matrix, which expresses the relationship between state and measurement in the joint density, is computed as follows:

$$\mathbf{P}_{\mathbf{xz},k|k-1} = \sum_{j=0}^{2n_{\mathbf{x}}} W_{j,k} \left[\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{j,k|k-1} - \hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1}\right] \left[h\left(\boldsymbol{\mathcal{X}}^{\boldsymbol{\eta}}_{j,k|k-1}\right) - \hat{\mathbf{z}}_{k|k-1}\right]^T. \tag{49}$$

Similarly to the learning domain, upon observing the measurement $\mathbf{z}_k$, the estimated mean and covariance of the posterior density are computed as

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}^{\boldsymbol{\eta}}_{k|k-1} + \mathbf{K}_k \ (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \tag{50}$$

$$\mathbf{P}_{k|k} = \mathbf{P}^{\boldsymbol{\eta}}_{k|k-1} - \mathbf{K}_k \ \mathbf{P}_{\mathbf{zz},k|k-1} \ (\mathbf{K}_k)^T, \tag{51}$$

where $\mathbf{K}_k$ denotes the Kalman gain defined by

$$\mathbf{K}_k = \mathbf{P}_{\mathbf{xz},k|k-1} \ (\mathbf{P}_{\mathbf{zz},k|k-1})^{-1}. \tag{52}$$

The posterior density of the state, conditioned on the observed measurement and the transferred learned parameters, denoted as $p(\mathbf{x}_k \mid \mathbf{z}_{1:k}, \boldsymbol{\eta}^\star_{2:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$, given in (11), is modeled as a Gaussian density. The mean and covariance of this posterior density are estimated via (50) and (51), respectively.

V. Simulation Results

We consider the problem of estimating the unknown state vector $\mathbf{x}_k = [x_k, \dot{x}_k, y_k, \dot{y}_k, \Omega_k]^T$ associated with a single object undergoing constant velocity motion in a two-dimensional trajectory. The state vector comprises the Cartesian coordinates (x_k, y_k), the object's velocity $(\dot{x}_k, \dot{y}_k)$, and the turn rate Ω_k. Our approach adopts a state transition model based on a nonlinear process model employed in [1], which is given by

$$\mathbf{x}_k = \begin{bmatrix} 1 & \frac{\sin(\Omega_{k-1}T_s)}{\Omega_{k-1}} & 0 & -\left(\frac{1-\cos(\Omega_{k-1}T_s)}{\Omega_{k-1}}\right) & 0 \\ 0 & \cos(\Omega_{k-1}T_s) & 0 & -\sin(\Omega_{k-1}T_s) & 0 \\ 0 & \frac{1-\cos(\Omega_{k-1}T_s)}{\Omega_{k-1}} & 1 & \frac{\sin(\Omega_{k-1}T_s)}{\Omega_{k-1}} & 0 \\ 0 & \sin(\Omega_{k-1}T_s) & 0 & \cos(\Omega_{k-1}T_s) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{v}_{k-1}, \tag{53}$$

where the process noise $\mathbf{v}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_\mathbf{v})$ with

$$\mathbf{Q}_\mathbf{v} = \begin{bmatrix} q_1\frac{T_s^4}{4} & q_1\frac{T_s^3}{2} & 0 & 0 & 0 \\ q_1\frac{T_s^3}{2} & q_1 T_s^2 & 0 & 0 & 0 \\ 0 & 0 & q_1\frac{T_s^4}{4} & q_1\frac{T_s^3}{2} & 0 \\ 0 & 0 & q_1\frac{T_s^3}{2} & q_1 T_s^2 & 0 \\ 0 & 0 & 0 & 0 & q_2 T_s \end{bmatrix}. \tag{54}$$

In order to ensure a fair comparison, we make the assumption that both the learning and tracking domains possess identical error process noise covariances, i.e., $\mathbf{Q}^\star_\mathbf{v} = \mathbf{Q}_\mathbf{v}$.

The measurement vector at time step k, denoted by $\mathbf{z}_k$, consists of the object's range r_k and angle ζ_k. The measurements obtained from the sensors in both the learning and tracking domains are affected by Gaussian noise, which are modeled by the following equations:

$$\mathbf{z}^\star_k = h(\mathbf{x}_k) + \mathbf{w}^\star_k, \qquad \mathbf{w}^\star_k \sim \mathcal{N}(0, \mathbf{Q}^\star_\mathbf{w}), \tag{55}$$

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{w}_k, \qquad \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_\mathbf{w}), \tag{56}$$

where $\mathcal{N}(0, \mathbf{Q}^\star_\mathbf{w})$ and $\mathcal{N}(0, \mathbf{Q}_\mathbf{w})$ denote zero-mean Gaussian noise with covariance matrices for the learning and tracking domains defined by $\mathbf{Q}^\star_\mathbf{w}$ and $\mathbf{Q}_\mathbf{w}$, respectively. Their corresponding intensities are represented by $I^\star_\mathbf{w}$ and $I_\mathbf{w}$. In both the learning and tracking domains, a common matrix $\mathbf{B}_\mathbf{w} = \mathrm{diag}[\sigma_r^2, \sigma_\zeta^2]$ is used to ensure a fair comparison. The only differing parameters between the two domains are the noise intensities. For the simulations in this section, the object follows the nonlinear transition model for a duration of $K = 100$ time steps. The initial parameters of the object are set as $\mathbf{x}_0 = [1000 \text{ m}, 300 \text{ m/s}, 1000 \text{ m}, 0 \text{ m/s}, -3^\circ/\text{s}]^T$. The obtained results in this section are based on averaging $10{,}000$ iterations of Monte Carlo (MC) simulations, utilizing the parameter settings outlined in Table I.

TABLE I
Simulation Parameters Settings

Parameter	Value
$n_\mathbf{x}$	5
$n_\mathbf{z}$	2
T_s	1 s
K	100
MC	$10{,}000$
α	1
κ	-2
q_1	$0.1 \text{ m}^2/\text{s}^4$
q_2	$1.75 \times 10^{-2} \text{ rad}^2/\text{s}^3$
σ_r	10 m
σ_ζ	$\sqrt{10} \times 10^{-3}$ rad
$I^\star_\mathbf{w}$	1
$I_\mathbf{w}$	$0.5 \to 6$

We investigate the performance of the proposed transfer learning-unscented Kalman filter (TL-UKF) algorithm in a scenario involving a single object with relatively high maneuverability. The object's trajectory is depicted in Figure 2. The algorithm is evaluated by computing the Root Mean-Squared Error (RMSE) in the tracking domain under different noise intensity values, denoted as $I_\mathbf{w} = \{0.5, 1.5, 4\}$, while keeping the noise intensity in the learning domain fixed at $I^\star_\mathbf{w} = 1$. The performance results in terms of RMSE for the proposed algorithm compared to the isolated UKF are presented in Figure 3. For instance, under $I_\mathbf{w} = 4$ and at $k = 30$, the proposed algorithm achieves an RMSE of 16.3 m, whereas the isolated UKF achieves an RMSE of 18.53 m. These results indicate that the proposed algorithm improves the accuracy of object tracking compared to the isolated UKF. However, it is worth noting that the proposed algorithm exhibits poorer performance at certain time steps, specifically $k = \{8, 45, 62\}$. This can be attributed to the use of the suggested choice of $\kappa = 3 - n_\mathbf{x}$, which in this scenario is a negative value. The proposed algorithm has less stability compared to the isolated UKF due to the reliance on transfer learning, particularly in the context of predicted observations. The choice of the negative value of κ can lead to inaccurate approximations when applying the UT via (25)-(30) in the learning step, further reducing stability.

The overall RMSE results under varying values of noise intensity in the tracking domain are plotted in Figure 4, with the shaded areas representing the interquartile range of $10{,}000$ MC iterations. Despite the limitations faced by the TL-UKF algorithm in tracking high maneuverability scenarios, the overall performance demonstrates a significant improvement

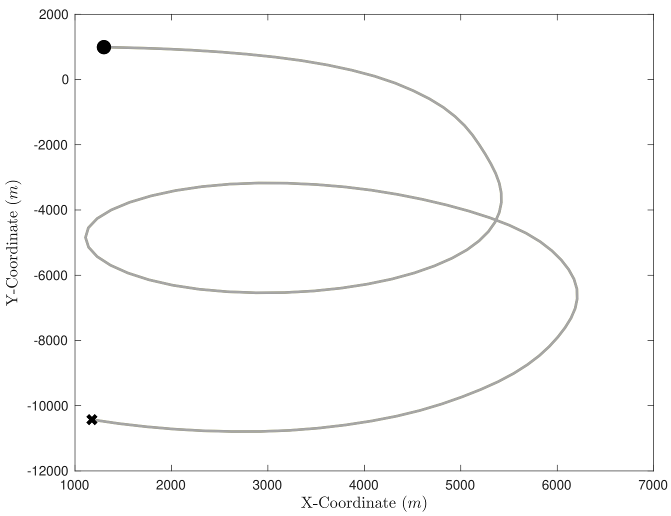

Fig. 2. High maneuverability trajectory, where ● indicates the initial point.

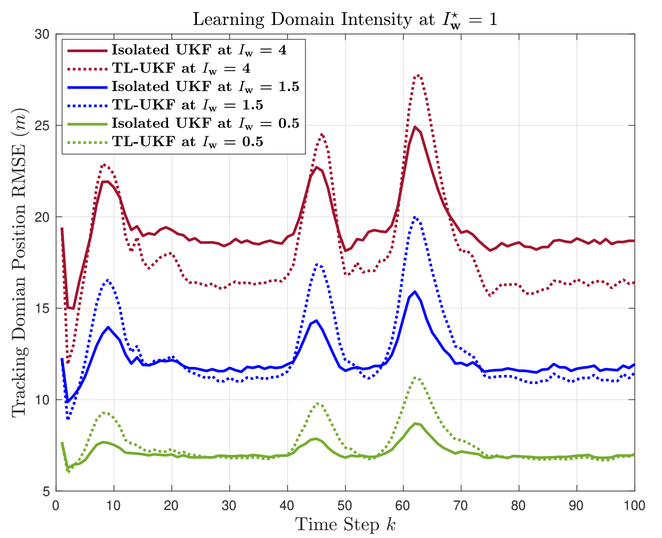

Fig. 3. RMSE curves of TL-UKF and isolated UKF high maneuverability trajectory scenario.

in the accuracy of estimation compared to the isolated UKF. For instance, under a noise intensity of $I_{\mathbf{w}} = 6$, the proposed algorithm achieves an RMSE of 20.62 m, while the isolated UKF achieves a similar RMSE of 20.53 m with a lower noise intensity of $I_{\mathbf{w}} = 4.5$. This indicates that the proposed algorithm is capable of tracking the object with comparable accuracy, even when faced with higher noise intensity that is approximately 1.5 greater.

VI. Conclusion

We proposed a tracking algorithm that incorporated transfer learning with an unscented Kalman filter. Our simulation results using a relatively high maneuverability trajectory showed the significant improvement in overall tracking accuracy that can be achieved compared to an isolated UKF. However, in

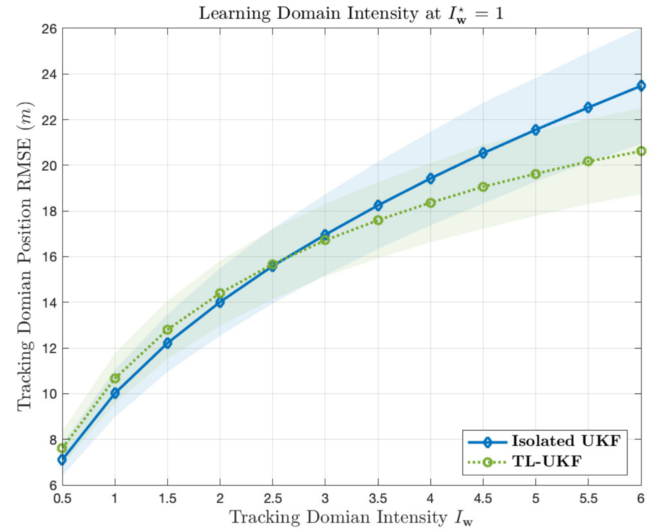

Fig. 4. Overall performance comparison of TL-UKF and isolated UKF algorithms for high maneuverability trajectory, with shaded areas representing the interquartile range.

this scenario where the proposed algorithm faces less numerical stability, inaccurately learned predicted observations in the learning domain can adversely affect the performance of the proposed algorithm at certain time steps. We are investigating this issue further in ongoing work.

References

[1] Y. Bar-Shalom and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY: John Wiley & Sons, Inc., 2001.

[2] C. Foley and A. Quinn, "Fully probabilistic design for knowledge transfer in a pair of Kalman filters," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 487–490, 2017.

[3] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F - Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, April 1993.

[4] E. Grolman, A. Bar, B. Shapira, L. Rokach, and A. Dayan, "Utilizing transfer learning for in-domain collaborative filtering," *Knowledge-Based Systems*, vol. 107, pp. 70–82, 2016.

[5] Y. Ho and R. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, October 1964.

[6] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Proceedings on Signal Processing, Sensor Fusion, and Target Recognition VI*, vol. 3068, 1997, pp. 182–193.

[7] A. Karbalayghareh, X. Qian, and E. R. Dougherty, "Optimal Bayesian transfer learning," *IEEE Transactions on Signal Processing*, vol. 66, no. 14, pp. 3724–3739, 2018.

[8] B. Li, Q. Yang, and X. Xue, "Transfer learning for collaborative filtering via a rating-matrix generative model," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 617–624.

[9] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.

[10] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proceedings of the IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, Oct. 2000, pp. 153–158.

[11] J. Xuan, J. Lu, and G. Zhang, "Bayesian transfer learning: An overview of probabilistic graphical models for transfer learning," *arXiv preprint arXiv:2109.13233*, 2021.