Adaptive Target Tracking With Interacting Heterogeneous Motion Models

Ki-In Na, Sunglok Choi, *Member, IEEE*, and Jong-Hwan Kim, *Fellow, IEEE*

***Abstract*— Multiple motion estimators such as an interacting multiple model (IMM) have been utilized to track target objects such as cars and pedestrians with diverse motion patterns. However, the standard IMM has limitations in combining motion models with different state definitions, so it cannot contain a complementary set of models that accurately work for all motion patterns. In this paper, we propose IMM-based adaptive target tracking with heterogeneous velocity representations and linear/curvilinear motion models. It can integrate four motion models with different state definitions and dimensions to be completely complimentary for all types of motions. We experimentally demonstrate the effectiveness of the proposed method with accuracy for various motion patterns using two types of datasets: synthetic datasets and real datasets. Experimental results show that the proposed method achieves the adaptive target tracking for diverse types of motion and also for various objects such as cars, pedestrians, and drones in the real world.**

***Index Terms*— Target tracking, interacting multiple model, heterogeneous motion models, Bayesian filtering.**

I. Introduction

ROAD objects such as cars, pedestrians, and cyclists exhibit various motion patterns, and those patterns can be dynamically changed with respect to traffic situations. It is important to select an appropriate motion model that can reliably track all surrounding objects for the safe operation of ADAS, and autonomous driving [1]–[3]. However, since each motion model has different tracking performance for different motion patterns, a single motion model is not sufficient to deal with diverse motion patterns of road objects [4]–[11]. Therefore, it is crucial to combine multiple motion models and compensate for their weaknesses for accurate and reliable object tracking.

Multiple model (MM) estimation methods utilizing multiple motion models have been studied such as generalized pseudo-Bayesian (GPB) [12], interacting multiple model (IMM) [13], and variable-structure multiple model (VSMM) [14]. IMM is widely used for tracking road objects as one of the most effective MM algorithms since it provides the best compromise between estimation accuracy and computational cost [15]–[17]. Increasing the number of motion models in IMM does not always guarantee better accuracy. Thus, it needs to compose a compact and complete set of motion models to achieve computational efficiency and tracking performance [18].

The combination of heterogeneous motion models has been investigated to positively affect MM estimation in [19], [20]. It has been demonstrated that IMM estimators applying linear (*constant velocity*; CV) and curvilinear (*coordinated turn*; CT) motion models with heterogeneous state dimensions can track moving objects in various motions (so called *state augmentation*) [21]–[26]. The tracking accuracy of IMM estimators can also be improved by integrating motion models represented in heterogeneous coordinates such as Cartesian and polar coordinates (so called *state mixing*). The previous work of the state mixing transforms all states into the same coordinates, mixes them, and then returns them to their original coordinates [27]. However, motion models do not work complementary since the tracking result is biased when all states are transformed to the same coordinates. Additionally, the previous IMMs cannot conduct state augmentation and mixing simultaneously, so motion models with heterogeneous state dimensions and coordinate representations cannot be combined.

In this light, we propose the adaptive target tracking in which motion models with heterogeneous velocity representations and state dimensions can completely complement each other. For unbiased state mixing, Q mixing is proposed to generate unbiased transition noise for states with both velocity representations in a single model. We also extend the state mixing by proposing ω mixing to apply motion models with heterogeneous state dimensions. To the best of our knowledge, the proposed method is the first IMM estimator capable of simultaneously applying multiple motion models with heterogeneous state dimensions (state augmentation) and velocity representations (state mixing) without bias. Accordingly, the proposed IMM estimator with interacting heterogeneous motion models effectively compensates for the weaknesses of each single motion model, and can reliably track objects with dynamically changing motions.

The main contributions of this paper are summarized as follows:

Manuscript received 17 September 2021; revised 16 March 2022 and 25 May 2022; accepted 6 July 2022. Date of publication 22 July 2022; date of current version 7 November 2022. This work was supported by the ICT Research and Development Program of MSIP/IITP (Development of multimodal sensor-based intelligent systems for outdoor surveillance robots) under Grant 2017-0-00306. The Associate Editor for this article was B. Fidan. *(Corresponding author: Jong-Hwan Kim.)*

Ki-In Na is with the Intelligent Robotics Research Division, ETRI, Daejeon 34129, Republic of Korea, and also with the Robotics Program, KAIST, Daejeon 34141, Republic of Korea (e-mail: kina4147@etri.re.kr).

Sunglok Choi is with the Department of Computer Science and Engineering, Seoul National University of Science and Technology (SEOULTECH), Seoul 01811, Republic of Korea (e-mail: sunglok@seoultech.ac.kr).

Jong-Hwan Kim is with the School of Electrical Engineering, KAIST, Daejeon 34141, Republic of Korea (e-mail: johkim@rit.kaist.ac.kr).

This article has supplementary downloadable material available at https://doi.org/10.1109/TITS.2022.3191814, provided by the authors.

Digital Object Identifier 10.1109/TITS.2022.3191814

- An in-depth analysis of motion models is conducted for different motion complexities (CV and CT) and different velocity representations (Cartesian and polar velocity).
- The IMM-based adaptive target tracking is proposed for unbiased interaction of motion models with heterogeneous velocity representations and state dimensions.
- The improvement of the proposed method is demonstrated in experiments using synthetic dataset for diverse motions and real dataset for multiple objects.

The rest of this paper is organized as follows. Section II summarizes the recent related works on motion models and IMM estimators with heterogeneous motion models. Section III introduces the target tracking using single motion models and state augmentation of multiple motion models with different state dimensions. Section IV proposes adaptive tracking using IMM estimators with heterogeneous velocity representations and also extends it to integrate multiple models with different-sized states. Sections V and VI illustrate experiments with synthetic datasets on four different motions and real datasets on three road objects, respectively. The concluding remarks follow in Section VII.

II. Related Works

A. Motion Models

1) Model Complexity: Object motion can be modeled in numerous motion models according to the model complexity, such as CV and CT. Previous studies have compared linear and curvilinear motion models to understand the relationship between model complexity and tracking performance. Tsogas *et al.* [4] empirically demonstrated that the more sophisticated models such as constant turn rate and constant tangential acceleration model (CTRA) accurately estimate the dynamic behavior of an object with an unscented Kalman filter (UKF). Schubert *et al.* [5] also represented that increasing model complexity to some extent helps to improve tracking performance. In their follow-up study [6], they described that the selection of an appropriate model depends on various factors such as the scenario, the observability of motion parameters, and the required prediction interval.

2) Velocity Representation: Past studies have also shown that tracking performance is dependent on the choice of state coordinate representations [7]–[11]. They generally compared the CT motion models with different velocity representations: Cartesian and polar velocities. It was demonstrated that CT models with the polar velocity representation provide better tracking performance than Cartesian velocity. Roth *et al.* [11] additionally suggested that for CT with Cartesian velocity, the sensitivity of the tracking performance to the noise parameters was reduced using UKF.

It is crucial to select the appropriate model complexity and velocity representation of the motion model according to the tracking situation, but in-depth analysis on it is still lacking despite many past studies. Moreover, it is difficult to track objects in the real environment with only a single motion model because the tracking situations are varied and complicated in real environments. Therefore, multi-motion estimators, IMMs, have been studied to combine multiple motion models with different model complexity and velocity representations.

B. IMM Estimators With Heterogeneous Motion Models

1) State Augmentation: The IMM estimator has been studied for state augmentation to simultaneously employ motion models with heterogeneous-sized states of different model complexity: simple, unbiased, and prior approaches. Simple approach [12] augments state variables with zero mean and variance in the smaller state, but this leads to biased estimation of the corresponding state variables in the larger state. Yuan *et al.* [21] proposed an unbiased approach that augments the smaller state with the mean and variance from the larger state. This approach achieves reasonable results with real data from thrusting ballistic projectiles. However, when a mode probability of a larger state is close to zero, it provides numerically unstable results because the weighted sum of covariance is nearly a singular matrix. Lopez *et al.* [22] presented a method similar to the unbiased method but it is implemented state-by-state, in contrast to [21] which is conducted directly on the state vector. Granstrom *et al.* [23] presented a prior approach that utilizes a uniform distribution derived from prior knowledge of minimum and maximum values for variance distribution of augmented state. The choice of augmentation distribution significantly affects the tracking performance of the IMM estimator.

Ou and Wang [24] employed the state augmentation to determine the optimal mode mixing strategy among simple, unbiased, and prior approaches. The optimal approach is dynamically selected based on mode probability and innovation matrix, implicitly meaning the target status. They also demonstrated that the simple and unbiased approaches perform well during CV motion, but the tracking error increases extremely when the target status is switched. Meanwhile, the prior approach utilizes a uniform distribution to keep a more significant covariance. This approach achieves robustness in sudden mode transitions but sacrifices tracking accuracy for steady status. Laneuveille [25] extended the study of performance comparisons between CT models with Cartesian velocity and polar velocity to a MM estimator. They exhibited that IMMs containing CV and CT models with polar velocity significantly improve tracking accuracy compared to IMMs containing CV and CT models with Cartesian velocity. Visina *et al.* [26] introduced the nonzero mean, white noise turn-rate (WNTR) model to handle the sharp turns quickly. They also added a CT mode to the IMM estimator with WMTRs for adaptive tracking in both slow and sharp turns based on the prior approach. Allig and Wanielik [28] presented the fusion of tracks with unequal dimensional state spaces, called heterogeneous track-to-track fusion (HT2TF), using covariance intersection. They investigated different state augmentation approaches and compared different optimization variants because state augmentation influences the optimization of the covariance intersection.

2) State Mixing: The appropriate combination of multiple motion models is required to compensate for the different weaknesses depending on model complexity and velocity representations of the motion models. Therefore, an

IMM estimator needs to fuse multiple motion models with heterogeneous state dimensions as well as heterogeneous velocity representations. Gao *et al.* [27] presented the IMM estimator mixing multiple motion models represented in heterogeneous state spaces. The model states are transformed into a common coordinate space for fusing multiple models in a linear relationship. The transformed states, however, are biased to the result of the target coordinates, so the transformation breaks the complementarity. In this regard, we propose a novel method for the IMM estimator to mix motion models with heterogeneous velocity representations to complement each other completely. Also, the proposed state mixing method is extended by including state augmentation to fuse motion models with unequal state dimensions.

III. Target Tracking With Homogeneous Velocity Representation

A. Notation

In this paper, target tracking is performed in the 2-dimensional space. The position of the target is represented as x and y in the Cartesian coordinate system and its orientation is denoted as θ with respect to the X-axis in the world coordinate system. Even though the real world is 3-dimensional, the 2-dimensional representation is common in many applications where a target is located on a plane such as ground, indoor floors, and roads. In the 2-dimensional space, the linear velocity (also known as *speed*) of the target is represented as v as a scalar and $(\dot{x}, \dot{y})$ as a Cartesian vector in the world coordinate system, respectively. If the target always moves along with its orientation without slippage, these velocity representations are related as

$$\dot{x} = v\cos\theta \quad \text{and} \quad \dot{y} = v\sin\theta. \tag{1}$$

Equation (1) is also known as *nonholonomic constraint* [29] or *ideal unicycle* kinematics. The angular velocity (also known as *turn rate*) is denoted as ω whose value is positive when the target turns counterclockwise (CCW) with the corresponding to the right-hand coordinate system. We expect that our idea and verification in the 2-dimensional representation is also effective for its extension to the 3-dimensional representation.

Our target tracking is based on discrete-time Bayesian filtering which is widely used in the area of target tracking. In this paper, the state variable is represented as $\mathbf{x}$ and time interval between state transitions is denoted as T. When the uncertainty of the state variable is described as a covariance matrix, it is denoted as P. The Bayesian filtering is composed of two steps, state prediction and correction, in conjunction with

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) \quad \text{and} \quad \mathbf{z}_k = h(\mathbf{x}_k), \tag{2}$$

where f is a transition function that predicts the current state from the previous state with a control input $\mathbf{u}$, and h is an observation function that simulates a measurement $\mathbf{z}$ from the current state. The measurement vector $\mathbf{z}$ can be defined diversely according to sensors and their perception outputs. Here, we commonly define the measurement vector as $\mathbf{z} = [x, y]^\top$ by considering GPS positioning, range-based object detection, and visual surveillance. In addition, we mainly use the unscented Kalman filter (UKF) since it can model nonlinear transition and observation functions efficiently and accurately compared to the original Kalman filter and extended Kalman filter (EKF) [30]. The UKF involves not only state variables but also sigma points to represent uncertainty of the state better. Its formulation and procedure are described in the accompanying supplementary material in detail.

B. Motion Models

We investigate four types of motion models with combinations of motion assumptions and representations, which are enumerated in Table I [8], [12]. The four motion models, CV-CC, CV-PC, CT-CC, and CT-PC, are commonly based on constant velocity, also known as white noise acceleration. The prefix in the model name indicates the velocity assumptions as constant velocity (CV-) and coordinated turn (CT-). The suffix denotes the velocity representations in Cartesian coordinates (-CC) and polar coordinates (-PC).

Firstly, the motion models are categorized as two groups with respect to their velocity assumptions. The **CV-** group is based on constant linear velocity and zero angular velocity (so called *constant velocity*; CV) [5]. Even though real targets violate this assumption (e.g. $\dot{v} \neq 0$ or $\omega \neq 0$), their linear velocity and direction are updated due to the noise considered in Bayesian filtering. For example, the linear velocity v of CV-PC will be updated due to white noise of linear acceleration whose covariance is $\sigma_{\dot{v}}^2$. Similarly, the orientation θ of CV-PC model will be updated due to $\sigma_{\dot{\omega}}^2$. It is important to assign those noise parameters large enough to take into account unmodeled physics. The orientation of CV-CC model is not defined, but it can be instantaneously derived according to $\theta = \tan^{-1}\frac{\dot{y}}{\dot{x}}$ from (1) when the target moves under the nonholonomic constraint [29]. The CV-PC is also known as *constant steering angle and velocity* (CSAV) model [5].

Next, the **CT-** group is derived under the assumption of constant linear and angular velocities (so called *coordinated turn*; CT) [9]. Since motion models in the CT group are degenerated in case of zero angular velocity ($\omega = 0$), it is common to use the CV models instead in near-zero angular velocity. However, there is no common rule or threshold to determine the near-zero to switch CT models to CV models. On the other hand, our motion model interaction automatically decides such substitution in the IMM filter framework.

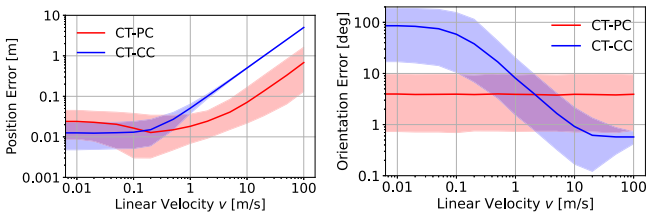
In addition, the models are also divided into another two groups with respect to their velocity representations [7]. The **-CC** group represents its target velocity as a vector $[\dot{x}, \dot{y}]^\top$ in the world *Cartesian coordinate* system. The **-PC** group represents its target velocity as a scalar v, which is the radial (or magnitude) component in the local *polar coordinate* system. Three velocity variables are connected each other as shown in (1) involving the target orientation θ. The CT-PC model is well-known as a *velocity motion* model [31] or a *constant turn rate and velocity* (CTRV) model [5].

C. Motion Model Analysis

Motion models with different velocity representations have different uncertainty propagation of target position

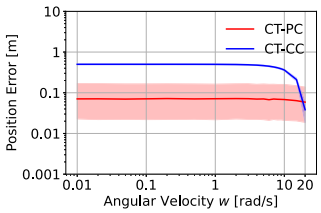
TABLE I
SINGLE MOTION MODELS (ACCORDING TO MOTION ASSUMPTIONS AND REPRESENTATIONS)

	CV-CC	CV-PC	CT-CC	CT-PC
Motion Assumptions	Constant velocity	Constant velocity	Coordinated turn	Coordinated turn
Velocity Coordinates	Cartesian coordinates	Polar coordinates	Cartesian coordinates	Polar coordinates
States Variables	$\mathbf{x} = [x\ y\ \dot{x}\ \dot{y}]^\top$	$\mathbf{x} = [x\ y\ \theta\ v]^\top$	$\mathbf{x} = [x\ y\ \dot{x}\ \dot{y}\ \omega]^\top$	$\mathbf{x} = [x\ y\ \theta\ v\ \omega]^\top$
Transition Functions $f(\mathbf{x})$	$\begin{bmatrix} x+\dot{x}T \\ y+\dot{y}T \\ \dot{x} \\ \dot{y} \end{bmatrix}$	$\begin{bmatrix} x+vT\cos\theta \\ y+vT\sin\theta \\ \theta \\ v \end{bmatrix}$	$\begin{bmatrix} x+\frac{\dot{x}}{\omega}\sin(\frac{\omega T}{2})-\frac{\dot{y}}{\omega}(1-\cos(\omega T)) \\ y+\frac{\dot{y}}{\omega}\sin(\frac{\omega T}{2})+\frac{\dot{x}}{\omega}(1-\cos(\omega T)) \\ \dot{x}\cos(\omega T)-\dot{y}\sin(\omega T) \\ \dot{x}\sin(\omega T)+\dot{y}\cos(\omega T) \\ \omega \end{bmatrix}$	$\begin{bmatrix} x+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\cos(\theta+\frac{\omega T}{2}) \\ y+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\sin(\theta+\frac{\omega T}{2}) \\ \theta+\omega T \\ v \\ \omega \end{bmatrix}$
Transition Noise Q	$\mathrm{Q} = \mathrm{G}\Sigma\mathrm{G}^\top$			
Noise Jacobian G	$\begin{bmatrix} \frac{T^2}{2} & 0 \\ 0 & \frac{T^2}{2} \\ T & 0 \\ 0 & T \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 \\ \frac{T^2}{2}\sin\theta & 0 \\ 0 & \frac{T^2}{2} \\ T & 0 \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2} & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 \\ T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 \\ \frac{T^2}{2}\sin\theta & 0 \\ 0 & \frac{T^2}{2} \\ T & 0 \\ 0 & T \end{bmatrix}$
Noise Covariance	$\Sigma = \mathrm{diag}(\sigma_{\ddot{x}}^2, \sigma_{\ddot{y}}^2)$	$\Sigma = \mathrm{diag}(\sigma_{\dot{v}}^2, \sigma_{\dot{\omega}}^2)$	$\Sigma = \mathrm{diag}(\sigma_{\ddot{x}}^2, \sigma_{\ddot{y}}^2, \sigma_{\dot{\omega}}^2)$	$\Sigma = \mathrm{diag}(\sigma_{\dot{v}}^2, \sigma_{\dot{\omega}}^2)$

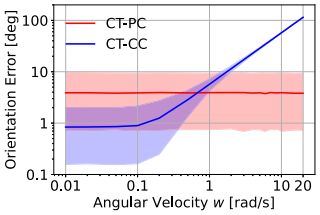


(a) Position error at $\omega = 0.1$ rad/s

(b) Orientation error at $\omega = 0.1$ rad/s



(c) Position error at $v = 10$ m/s



(d) Orientation error at $v = 10$ m/s

Fig. 1. Error and uncertainty propagation of CT-CC and CT-PC in Monte Carlo experiments (10^4 trials).

and orientation. Such differences lead to distinct position and orientation accuracies in Bayesian filtering, which is our motivation to adopt multiple motion models in the filtering. To visualize uncertainty propagation of the motion models, Monte Carlo experiments were performed with CT-CC and CT-PC models. In the experiments, state prediction was executed 10^4 times from the initial state ($\theta = \pi/4$ rad) with respect to varying linear velocity (from 0 to 100 m/s) and angular velocity (from 0.01 to 20 rad/s). The state prediction involved unbiased Gaussian noises on orientation and velocity, whose values of standard deviation were $\sigma_\theta = 0.1$ rad, $\sigma_v = 0.3$ m/s, and $\sigma_\omega = 0.1$ rad/s with $T = 0.1$ s. Fig. 1 shows the medians of position and orientation errors in logarithmic scale along with their 10 and 90 percentiles of which region represents the degree of uncertainty.

Fig. 1 (a) and (b) show position and orientation errors and uncertainty in varying linear velocity with a fixed angular velocity, $\omega = 0.1$ rad/s. In low speed ($v < 0.1$ m/s), the CT-CC model has higher position accuracy but quite worse orientation accuracy compared to CT-PC. It is evident that CT-CC has worse orientation error because its orientation calculation, $\theta = \tan^{-1}\frac{\dot{y}}{\dot{x}}$, becomes unstable when $\dot{x}$ is small. As the speed increases, CT-CC has lower position accuracy than CT-PC, but with less uncertainty. On the contrary, CT-CC provides better orientation accuracy than CT-PC in high speeds. CT-PC has the same orientation error and uncertainty at various linear velocities.

Fig. 1 (c) and (d) reveal accuracy and uncertainty in varying angular velocity with a fixed linear velocity, $v = 10$ m/s. CT-CC mostly has lower position accuracy with less uncertainty, but the accuracy becomes higher in a very high angular displacement (around $\omega T = 1.5$ rad). CT-CC, however, has higher orientation accuracy with large uncertainty in slow angular velocity (low curvature; smooth turn) and worse orientation accuracy with less uncertainty in fast angular velocity (high curvature; sharp turn). On the other hand, CT-PC has almost the same accuracy and uncertainty of position and orientation at various angular velocities.

From the above analysis on Fig. 1, we observed that motion models with the same motion assumptions but different velocity representations do not have the same accuracy (shown as mean values) and uncertainty (shown as 10 and 90 percentiles) under various motion conditions of linear and angular velocities. Our observations on motion models with different velocity representations have triggered us to focus on motion models against diverse motion types such as slow-to-fast speed and smooth-to-sharp turn.

D. IMM-UKF Tracking With State Augmentation

Two IMM estimators containing CV and CT modes with homogeneous velocity representations are designed to verify the state augmentation: IMM-AUG-CC (CV-CC and CT-CC) and IMM-AUG-PC (CV-PC and CT-PC), as shown in Table II. Since CT models include one more state variable of angular

TABLE II
IMM WITH STATE AUGMENTATION FOR HETEROGENEOUS SIZED STATES

	IMM-AUG-CC		**IMM-AUG-PC**	
State Variables	$\mathbf{x} = [x\ y\ \dot{x}\ \dot{y}\ \omega]^\top$		$\mathbf{x} = [x\ y\ \theta\ v\ \omega]^\top$	
Modes	CV-CC	CT-CC	CV-PC	CT-PC
Transition Functions $f(\mathbf{x})$	$\begin{bmatrix} x+\dot{x}T \\ y+\dot{y}T \\ \dot{x} \\ \dot{y} \\ 0 \end{bmatrix}$	$\begin{bmatrix} x+\frac{\dot{x}}{\omega}\sin(\frac{\omega T}{2})-\frac{\dot{y}}{\omega}(1-\cos(\omega T)) \\ y+\frac{\dot{y}}{\omega}\sin(\frac{\omega T}{2})+\frac{\dot{x}}{\omega}(1-\cos(\omega T)) \\ \dot{x}\cos(\omega T)-\dot{y}\sin(\omega T) \\ \dot{x}\sin(\omega T)+\dot{y}\cos(\omega T) \\ \omega \end{bmatrix}$	$\begin{bmatrix} x+vT\cos\theta \\ y+vT\sin\theta \\ \theta \\ v \\ 0 \end{bmatrix}$	$\begin{bmatrix} x+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\cos(\theta+\frac{\omega T}{2}) \\ y+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\sin(\theta+\frac{\omega T}{2}) \\ \theta+\omega T \\ v \\ \omega \end{bmatrix}$
Noise Jacobian G	$\begin{bmatrix} \frac{T^2}{2} & 0 \\ 0 & \frac{T^2}{2} \\ T & 0 \\ 0 & T \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2} & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 \\ T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 \\ \frac{T^2}{2}\sin\theta & 0 \\ 0 & \frac{T^2}{2} \\ T & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 \\ \frac{T^2}{2}\sin\theta & 0 \\ 0 & \frac{T^2}{2} \\ T & 0 \\ 0 & T \end{bmatrix}$
Noise Covariance	$\Sigma = \mathrm{diag}(\sigma_{\ddot{x}}^2, \sigma_{\ddot{y}}^2)$	$\Sigma = \mathrm{diag}(\sigma_{\ddot{x}}^2, \sigma_{\ddot{y}}^2, \sigma_{\dot{\omega}}^2)$	$\Sigma = \mathrm{diag}(\sigma_{\dot{v}}^2, \sigma_{\dot{\omega}}^2)$	$\Sigma = \mathrm{diag}(\sigma_{\dot{v}}^2, \sigma_{\dot{\omega}}^2)$

velocity ω than CV models, IMM filters must employ multiple motion models with unequal dimensional states. Therefore, we apply an unbiased approach for the state augmentation in which CV mode with the smaller dimensional state takes the mean and variance for the absent state variable ω from CT mode with a larger dimensional state when calculating the mixed state of CT mode in interaction step of IMM [21]. The state augmentation in the unbiased approach is represented as follows:

$$\mathbf{x}^{\mathrm{CT}} = \begin{bmatrix} \mathbf{x}_{\omega^c}^{\mathrm{CT}} \\ \mathbf{x}_{\omega}^{\mathrm{CT}} \end{bmatrix}, \quad \mathrm{P}^{\mathrm{CT}} = \begin{bmatrix} \mathrm{P}_{\omega^c\omega^c}^{\mathrm{CT}} & \mathrm{P}_{\omega^c\omega}^{\mathrm{CT}} \\ \mathrm{P}_{\omega\omega^c}^{\mathrm{CT}} & \mathrm{P}_{\omega\omega}^{\mathrm{CT}} \end{bmatrix}, \tag{3}$$

$$\mathbf{x}^{\mathrm{CV}} = \begin{bmatrix} \mathbf{x}_{\omega^c}^{\mathrm{CV}} \\ \mathbf{x}_{\omega}^{\mathrm{CT}} \end{bmatrix}, \quad \mathrm{P}^{\mathrm{CV}} = \begin{bmatrix} \mathrm{P}_{\omega^c\omega^c}^{\mathrm{CV}} & 0 \\ 0 & \mathrm{P}_{\omega\omega}^{\mathrm{CT}} \end{bmatrix}, \tag{4}$$

where $\mathbf{x}^{\mathrm{CT}}$ and P^{CT} are the state and covariance of CT mode, and $\mathbf{x}^{\mathrm{CV}}$ and P^{CV} are those of CV mode, respectively. $\mathbf{x}_{\omega^c}$ refers to state variables other than ω, and is common state variables in both motion models: $[x\ y\ \dot{x}\ \dot{y}]$ for IMM-AUG-CC and $[x\ y\ \theta\ v]$ for IMM-AUG-PC. The CV mode is augmented with the state ω of the CT mode when the mixed state $\tilde{\mathbf{x}}^{\mathrm{CT}}$ and covariance $\tilde{\mathrm{P}}^{\mathrm{CT}}$ of CT mode are calculated as follows:

$$\tilde{\mathbf{x}}^{\mathrm{CT}} = \mu_{\mathrm{CV|CT}}\mathbf{x}^{\mathrm{CV}} + \mu_{\mathrm{CT|CT}}\mathbf{x}^{\mathrm{CT}}, \tag{5}$$

$$\tilde{\mathrm{P}}^{\mathrm{CT}} = \mu_{\mathrm{CV|CT}}(\mathrm{P}^{\mathrm{CV}} + \delta^{\mathrm{CV}}\delta^{\mathrm{CV}\top}) + \mu_{\mathrm{CT|CT}}(\mathrm{P}^{\mathrm{CT}} + \delta^{\mathrm{CT}}\delta^{\mathrm{CT}\top}), \tag{6}$$

with

$$\delta^{\mathrm{CV}} = \mathbf{x}^{\mathrm{CV}} - \tilde{\mathbf{x}}^{\mathrm{CT}} \quad \text{and} \quad \delta^{\mathrm{CT}} = \mathbf{x}^{\mathrm{CT}} - \tilde{\mathbf{x}}^{\mathrm{CT}}, \tag{7}$$

where $\mu_{\mathrm{CT|CT}}$ and $\mu_{\mathrm{CV|CT}}$ are the mixing probabilities for CT modes. $\tilde{\mathbf{x}}^{\mathrm{CT}}$ is derived by the weighted sum of two mode states. $\tilde{\mathrm{P}}^{\mathrm{CT}}$ is calculated as the weighted sum with variance correction from δ^{CV} and δ^{CT}, which are differences between the mixed state and each mode state. The mixed state $\tilde{\mathbf{x}}^{\mathrm{CV}}$ and covariance $\tilde{\mathrm{P}}^{\mathrm{CV}}$ of CV mode are estimated in the standard approach of IMM (See supplementary material for details of IMM-UKF.).

The unbiased approach for state augmentation outperforms the simple approach in normal cases [24]. However, when mixing probability of modes with a larger state, $\mu_{\mathrm{CT|CT}}$ is close to zero, tracking results are numerically unstable because the weighted sum of covariance becomes a nearly singular matrix. In this case, we set a small value to the covariance of ω to keep it the positive definite. The prior approach [23] utilizes a uniform distribution according to the value range for the augmented state. This approach performs better in mode change or dynamic motions but cannot adequately generate covariance in regular motions. Therefore, this paper applies the unbiased approach to implement the IMM estimator for state augmentation.

IV. ADAPTIVE TRACKING WITH INTERACTING HETEROGENEOUS MOTION MODELS

Depending on the velocity representation, the motion model shows different tracking performance with changes in linear and angular velocities as described in Section III-C. In this light, we first propose an IMM estimator capable of utilizing motion models with heterogeneous velocity states to compensate for the weaknesses of their respective velocity representations. In addition, we introduce an IMM estimator that simultaneously applies both state augmentation for heterogeneous dimensional states and state mixing for heterogeneous velocity states to achieve adaptive tracking of dynamically moving objects.

A. IMM-UKF Tracking With State Mixing for Heterogeneous Velocity Representations

To apply motion models with different velocity representations to IMM, we design IMM-CV-MIX for mixing two CV models (CV-CC and CV-PC) and IMM-CT-MIX for mixing two CT models (CT-CC and CT-PC), as reported in Table III. Each mode of IMMs has velocity state variables represented in different coordinates: $(\dot{x},\ \dot{y})$ for Cartesian coordinates and $(\theta,\ v)$ for polar coordinates. Thus, we set the state vectors as $[\theta\ v\ \dot{x}\ \dot{y}]$ to redundantly contain the velocity states represented in both coordinates. The added velocity states from other coordinates are calculated by the predicted original velocity states of the modes in the prediction step. In addition, to calculate the noise covariance for the other representations, we propose Q mixing that multiplies the Jacobian J between two different velocity representations as follows:

$$\tilde{\Sigma}_{\dot{v}\dot{\omega}} = \mathrm{J}_{\mathrm{CP}}\Sigma_{\ddot{x}\ddot{y}}\mathrm{J}_{\mathrm{CP}}^\top \quad \text{and} \quad \tilde{\Sigma}_{\ddot{x}\ddot{y}} = \mathrm{J}_{\mathrm{PC}}\Sigma_{\dot{v}\dot{\omega}}\mathrm{J}_{\mathrm{PC}}^\top, \tag{8}$$

TABLE III
IMM WITH STATE MIXING FOR HETEROGENEOUS VELOCITY REPRESENTATIONS

	IMM-CV-MIX		**IMM-CT-MIX**	
State Variables	$\mathbf{x} = [x\ y\ \theta\ v\ \dot{x}\ \dot{y}]^\top$		$\mathbf{x} = [x\ y\ \theta\ v\ \dot{x}\ \dot{y}\ \omega]^\top$	
Modes	CV-CC	CV-PC	CT-CC	CT-PC
Transition Functions $f(\mathbf{x})$	$\begin{bmatrix} x+\dot{x}T \\ y+\dot{y}T \\ \tan^{-1}(\frac{\dot{y}}{\dot{x}}) \\ \sqrt{(\dot{x})^2+(\dot{y})^2} \\ \dot{x} \\ \dot{y} \end{bmatrix}$	$\begin{bmatrix} x+vT\cos\theta \\ y+vT\sin\theta \\ \theta \\ v \\ v\cos\theta \\ v\sin\theta \end{bmatrix}$	$\begin{bmatrix} x+\frac{\dot{x}}{\omega}\sin(\omega T)-\frac{\dot{y}}{\omega}(1-\cos(\omega T)) \\ y+\frac{\dot{x}}{\omega}(1-\cos(\omega T))+\frac{\dot{y}}{\omega}\sin(\omega T) \\ \tan^{-1}(\frac{\dot{x}\sin(\omega T)+\dot{y}\cos(\omega T)}{\dot{x}\cos(\omega T)-\dot{y}\sin(\omega T)}) \\ \sqrt{(\dot{x})^2+(\dot{y})^2} \\ \dot{x}\cos(\omega T)-\dot{y}\sin(\omega T) \\ \dot{x}\sin(\omega T)+\dot{y}\cos(\omega T) \\ \omega \end{bmatrix}$	$\begin{bmatrix} x+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\cos(\theta+\frac{\omega T}{2}) \\ y+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\sin(\theta+\frac{\omega T}{2}) \\ \theta+\omega T \\ v \\ v\cos(\theta+\omega T) \\ v\sin(\theta+\omega T) \\ \omega \end{bmatrix}$
Noise Jacobian G	$\begin{bmatrix} 0 & 0 & \frac{T^2}{2} & 0 \\ 0 & 0 & 0 & \frac{T^2}{2} \\ 0 & \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 & 0 \\ 0 & 0 & T & 0 \\ 0 & 0 & 0 & T \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 & 0 & 0 \\ \frac{T^2}{2}\sin\theta & 0 & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 & 0 \\ 0 & T & 0 & 0 \\ 0 & 0 & T & 0 \\ 0 & 0 & 0 & T \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & \frac{T^2}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{T^2}{2} & 0 \\ 0 & 0 & T & 0 & 0 \\ 0 & 0 & 0 & T & 0 \\ 0 & \frac{T^2}{2} & 0 & 0 & 0 \\ T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & T \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 & 0 & 0 \\ \frac{T^2}{2}\sin\theta & 0 & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 & 0 \\ 0 & 0 & T & 0 \\ 0 & 0 & 0 & T \end{bmatrix}$
Mixed Noise Covariance	$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \Sigma_{\dot{x}\dot{y}} \end{bmatrix}$	$\tilde{\Sigma} = \begin{bmatrix} \Sigma_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \tilde{\Sigma}_{\dot{x}\dot{y}} \end{bmatrix}$	$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times2} & 0 \\ \mathbf{0}_{2\times2} & \Sigma_{\dot{x}\dot{y}} & 0 \\ 0 & 0 & \sigma^2_{\dot{\omega}} \end{bmatrix}$	$\tilde{\Sigma} = \begin{bmatrix} \Sigma_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \tilde{\Sigma}_{\dot{x}\dot{y}} \end{bmatrix}$

where $\tilde{\Sigma}_{\dot{v}\dot{\omega}}$ and $\tilde{\Sigma}_{\dot{x}\dot{y}}$ are the mixed noise covariance transformed from the original noise covariances $\Sigma_{\dot{x}\dot{y}}$ and $\Sigma_{\dot{v}\dot{\omega}}$, respectively. $\mathrm{J_{CP}}$ is transformation from Cartesian to polar coordinates and $\mathrm{J_{PC}}$ is the opposite as

$$\mathrm{J_{CP}} = \begin{bmatrix} \frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}} & \frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}} \\ \frac{-\dot{y}}{\dot{x}^2+\dot{y}^2} & \frac{\dot{x}}{\dot{x}^2+\dot{y}^2} \end{bmatrix} \text{ and}$$
$$\mathrm{J_{PC}} = \begin{bmatrix} \cos\theta & -v\cos\theta \\ \sin\theta & v\cos\theta \end{bmatrix}, \tag{9}$$

and $\Sigma_{\dot{x}\dot{y}}$ and $\Sigma_{\dot{v}\dot{\omega}}$ are the original noise covariances for CC and PC groups, respectively, as

$$\Sigma_{\dot{x}\dot{y}} = \begin{bmatrix} \sigma^2_{\dot{x}} & 0 \\ 0 & \sigma^2_{\dot{y}} \end{bmatrix} \text{ and } \Sigma_{\dot{v}\dot{\omega}} = \begin{bmatrix} \sigma^2_{\dot{v}} & 0 \\ 0 & \sigma^2_{\dot{\omega}} \end{bmatrix}. \tag{10}$$

The proposed state mixing method consists of redundant velocity states and Q mixing. In IMM, each mode independently estimates the state vector and covariance matrix in prediction and correction steps, and the motion models are mixed in the interaction step. Thus, it is crucial to prevent poorly performing coordinate representations from ruining the results in the interaction step. Redundant velocity states with heterogeneous representations provide robustness to various motion changes without bias, allowing recovery from instability when one of the velocity state representations degrades. Moreover, transition covariance of the motion model with redundant velocity states is computed to maintain a complementary relationship through Q mixing.

Lastly, in combination step, an IMM estimator with different velocity representations generates combined velocity states as follows:

$$\mathbf{x} = \sum_{i\in\mathbf{M}} \mu_i \mathbf{x}_i, \tag{11}$$

$$\mathrm{P} = \sum_{i\in\mathbf{M}} \mu_i \left[\mathrm{P}_i + (\mathbf{x}-\mathbf{x}_i)(\mathbf{x}-\mathbf{x}_i)^\top\right], \tag{12}$$

$$\mathbf{x}^{\mathrm{CC}}_{\dot{x}\dot{y}} = \mu_{\mathrm{PC}}\psi_{\mathrm{PC}}(\mathbf{x}_{\theta v}) + \mu_{\mathrm{CC}}\mathbf{x}_{\dot{x}\dot{y}}, \tag{13}$$

$$\mathrm{P}^{\mathrm{CC}}_{\dot{x}\dot{y}} = \mu_{\mathrm{PC}}\mathrm{J_{PC}}\mathrm{P}_{\theta v}\mathrm{J}^\top_{\mathrm{PC}} + \mu_{\mathrm{CC}}\mathrm{P}_{\dot{x}\dot{y}}, \tag{14}$$

where $\mathbf{M}$ is a model set of IMM estimator and μ_i is the mode probability of the i-th mode. ψ_{PC} transforms the polar velocity, $\mathbf{x}_{\theta v}$ into the Cartesian velocity, $\mathbf{x}_{\dot{x}\dot{y}}$. Consequently, the proposed method combines the results of all modes and derive a result represented in one coordinate.

The IMM with CV and CT using state augmentation provides reliable tracking performance for linear and curvilinear motions. In addition, the IMM applying the state mixing of CC and PC is designed to compensate for the shortcomings of each motion model according to the velocity representation. To compensate for all the weaknesses of motion models with different model complexity and velocity representation, we propose IMM-AUG-MIX by combining the state augmentation and the state mixing. IMM-AUG-MIX can be modeled by combining two or more among the four single motion models; CV-CC, CV-PC, CT-CC, and CT-PC. As a result of comparing all combinations, the combination of the four models provides the same or better performance. Therefore, we design IMM-AUG-MIX combining all four models in Table IV.

B. IMM-UKF Tracking With State Augmentation and Mixing

The process of IMM-AUG-MIX is configured as shown in Fig. 2. The ω mixing is proposed to obtain the augmented state and covariance of CV modes from two CT modes and to calculate the mixed state $\tilde{\mathbf{x}}^{\mathrm{CT}}_i$ and covariance $\tilde{\mathrm{P}}^{\mathrm{CT}}_i$ of the i-th CT mode by mixing all modes through unbiased state augmentation in the interaction step. The augmented state $\mathbf{x}^{\mathrm{CT}}_{i,\omega}$ and covariance $\mathrm{P}^{\mathrm{CT}}_{i,\omega\omega}$ for angular velocity ω of CV modes are derived from the normalized weighted sum of CT modes with the mixing probability as follows:

$$\mathbf{x}^{\mathrm{CT}}_{i,\omega} = \sum_{j\in\mathbf{M}^{\mathrm{CT}}} \mu_{j|i}\mathbf{x}^j_\omega \Big/ \sum_{j\in\mathbf{M}^{\mathrm{CT}}} \mu_{j|i}, \tag{15}$$

$$\mathrm{P}^{\mathrm{CT}}_{i,\omega\omega} = \sum_{j\in\mathbf{M}^{\mathrm{CT}}} \mu_{j|i}\mathrm{P}^j_{\omega\omega} \Big/ \sum_{j\in\mathbf{M}^{\mathrm{CT}}} \mu_{j|i}, \tag{16}$$

where $\mu_{j|i}$ is the mixing probability of the j-th mode to estimate the mixed state of i-th mode. $\mathbf{M}^{\mathrm{CT}}$ is a set of

TABLE IV

IMM WITH STATE AUGMENTATION AND STATE MIXING FOR FOUR HETEROGENEOUS MOTION MODELS

	IMM-AUG-MIX			
State Variables	$\mathbf{x} = [x\ y\ \dot{x}\ \dot{y}\ \theta\ v\ \omega]^\top$			
Modes	CV-CC	CV-PC	CT-CC	CT-PC
Transition Functions $f(\mathbf{x})$	$\begin{bmatrix} x+\dot{x}T \\ y+\dot{y}T \\ \tan^{-1}(\frac{\dot{y}}{\dot{x}}) \\ \sqrt{(\dot{x})^2+(\dot{y})^2} \\ \dot{x} \\ \dot{y} \\ 0 \end{bmatrix}$	$\begin{bmatrix} x+vT\cos\theta \\ y+vT\sin\theta \\ \theta \\ v \\ v\cos\theta \\ v\sin\theta \\ 0 \end{bmatrix}$	$\begin{bmatrix} x+\frac{\dot{x}}{\omega}\sin(\omega T)-\frac{\dot{y}}{\omega}(1-\cos(\omega T)) \\ y+\frac{\dot{x}}{\omega}(1-\cos(\omega T))+\frac{\dot{y}}{\omega}\sin(\omega T) \\ \tan^{-1}(\frac{\dot{x}\sin(\omega T)+\dot{y}\cos(\omega T)}{\dot{x}\cos(\omega T)-\dot{y}\sin(\omega T)}) \\ \sqrt{(\dot{x})^2+(\dot{y})^2} \\ \dot{x}\cos(\omega T)-\dot{y}\sin(\omega T) \\ \dot{x}\sin(\omega T)+\dot{y}\cos(\omega T) \\ \omega \end{bmatrix}$	$\begin{bmatrix} x+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\cos(\theta+\frac{\omega T}{2}) \\ y+\frac{2v}{\omega}\sin(\frac{\omega T}{2})\sin(\theta+\frac{\omega T}{2}) \\ \theta+\omega T \\ v \\ v\cos(\theta+\omega T) \\ v\sin(\theta+\omega T) \\ \omega \end{bmatrix}$
Noise Jacobian G	$\begin{bmatrix} 0 & 0 & \frac{T^2}{2} & 0 \\ 0 & 0 & 0 & \frac{T^2}{2} \\ 0 & \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 & 0 \\ 0 & 0 & T & 0 \\ 0 & 0 & 0 & T \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 & 0 & 0 \\ \frac{T^2}{2}\sin\theta & 0 & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 & 0 \\ 0 & 0 & T & 0 \\ 0 & 0 & 0 & T \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & \frac{T^2}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{T^2}{2} & 0 \\ 0 & 0 & T & 0 & 0 \\ 0 & 0 & 0 & T & 0 \\ 0 & \frac{T^2}{2} & 0 & 0 & 0 \\ T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & T \end{bmatrix}$	$\begin{bmatrix} \frac{T^2}{2}\cos\theta & 0 & 0 & 0 \\ \frac{T^2}{2}\sin\theta & 0 & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 & 0 \\ 0 & 0 & T & 0 \\ 0 & 0 & 0 & T \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Mixed Noise Covariance	$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \tilde{\Sigma}_{\ddot{x}\ddot{y}} \end{bmatrix}$	$\tilde{\Sigma} = \begin{bmatrix} \Sigma_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \tilde{\Sigma}_{\ddot{x}\ddot{y}} \end{bmatrix}$	$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times 2} & 0 \\ \mathbf{0}_{2\times 2} & \Sigma_{\ddot{x}\ddot{y}} & 0 \\ 0 & 0 & \sigma_{\dot{\omega}}^2 \end{bmatrix}$	$\tilde{\Sigma} = \begin{bmatrix} \Sigma_{\dot{v}\dot{\omega}} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \tilde{\Sigma}_{\ddot{x}\ddot{y}} \end{bmatrix}$

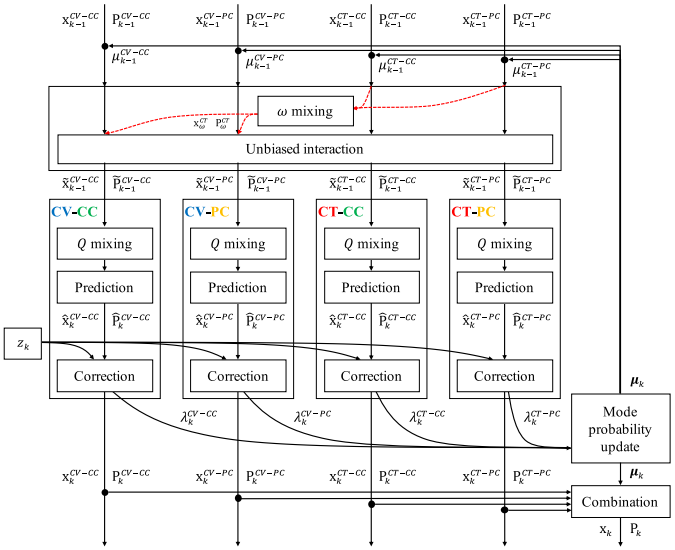


Fig. 2. Tracking process of IMM-AUG-MIX with state augmentation and state mixing of four motion models: CV-CC, CV-PC, CT-CC, and CT-PC. λ_k is the likelihood of each mode to evaluate mode probabilities.

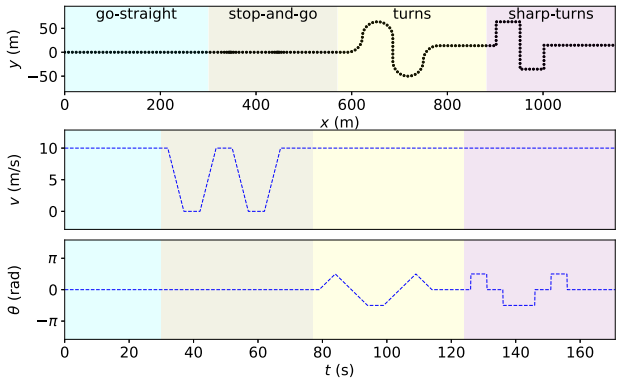


Fig. 3. Synthetic datasets for various motions in 10 m/s: `go-straight`, `stop-and-go`, `turns`, and `sharp-turns`, sequentially.

CT modes. Then, the mixed state $\tilde{\mathbf{x}}_i^{\mathrm{CT}}$ and covariance $\tilde{\mathrm{P}}_i^{\mathrm{CT}}$ of the CT mode are calculated as the weighted sum of all modes through the unbiased state augmentation [21] as follows:

$$\tilde{\mathbf{x}}_i^{\mathrm{CT}} = \sum_{p\in\mathbf{M}^{\mathrm{CV}}} \mu_{p|i}\mathbf{x}_p^{\mathrm{CV}} + \sum_{q\in\mathbf{M}^{\mathrm{CT}}} \mu_{q|i}\mathbf{x}_q^{\mathrm{CT}}, \tag{17}$$

$$\tilde{\mathrm{P}}_i^{\mathrm{CT}} = \sum_{p\in\mathbf{M}^{\mathrm{CV}}} \mu_{p|i}(\mathrm{P}_p^{\mathrm{CV}} + \delta_p^{\mathrm{CV}}\delta_p^{\mathrm{CV}\top}) + \sum_{q\in\mathbf{M}^{\mathrm{CT}}} \mu_{q|i}(\mathrm{P}_q^{\mathrm{CT}} + \delta_q^{\mathrm{CT}}\delta_q^{\mathrm{CT}\top}), \tag{18}$$

where

$$\mathbf{x}_p^{\mathrm{CV}} = \begin{bmatrix} \mathbf{x}_{p,\omega^c}^{\mathrm{CV}} \\ \mathbf{x}_{i,\omega}^{\mathrm{CT}} \end{bmatrix}, \mathrm{P}_p^{\mathrm{CV}} = \begin{bmatrix} \mathrm{P}_{p,\omega^c\omega^c}^{\mathrm{CV}} & 0 \\ 0 & \mathrm{P}_{i,\omega\omega}^{\mathrm{CT}} \end{bmatrix}, \tag{19}$$

$$\mathbf{x}_q^{\mathrm{CT}} = \begin{bmatrix} \mathbf{x}_{q,\omega^c}^{\mathrm{CT}} \\ \mathbf{x}_{q,\omega}^{\mathrm{CT}} \end{bmatrix}, \mathrm{P}_q^{\mathrm{CT}} = \begin{bmatrix} \mathrm{P}_{q,\omega^c\omega^c}^{\mathrm{CT}} & \mathrm{P}_{q,\omega^c\omega}^{\mathrm{CT}} \\ \mathrm{P}_{q,\omega\omega^c}^{\mathrm{CT}} & \mathrm{P}_{q,\omega\omega}^{\mathrm{CT}} \end{bmatrix}, \tag{20}$$

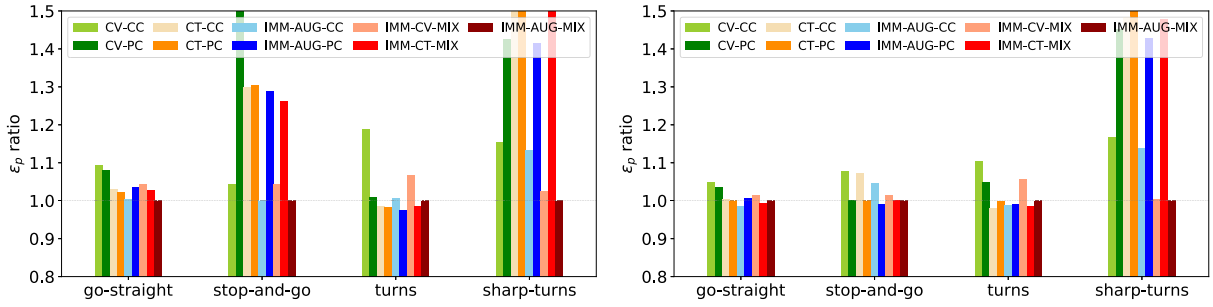
$$\delta_p^{\mathrm{CV}} = \mathbf{x}_p^{\mathrm{CV}} - \tilde{\mathbf{x}}_p^{\mathrm{CT}} \quad \text{and} \quad \delta_q^{\mathrm{CT}} = \mathbf{x}_q^{\mathrm{CT}} - \tilde{\mathbf{x}}_q^{\mathrm{CT}}. \tag{21}$$

The mixed state of CV modes, such as CV-CC and CV-PC, are fused in the standard approach of the IMM. Q mixing is also conducted for the state mixing as described in Section IV-A. Consequently, the proposed IMM estimator, IMM-AUG-MIX, can combine motion models with both heterogeneous state dimensions and heterogeneous velocity representations.

V. EXPERIMENTS WITH SYNTHETIC DATASETS

A. Experimental Setting

We compared the tracking accuracy of the motion models introduced in this paper by applying UKF: CV-CC, CV-PC, CT-CC, CT-PC, IMM-AUG-CC, IMM-AUG-PC, IMM-CV-MIX, IMM-CT-MIX, and IMM-AUG-MIX. For the performance evaluation according to different motions, we built synthetic datasets with sequential motions of `go-straight`, `stop-and-go`, `turns`, and `sharp-turns` as shown in Fig. 3. The linear velocities of synthetic datasets were set



(a) ε_p ratio for 5 m/s. (b) ε_p ratio for 10 m/s.

Fig. 4. Accuracy of 5 m/s and 10 m/s for each motion in experiments with synthetic datasets. For clear comparison, the positional errors are represented as a ratio with IMM-AUG-MIX.

TABLE V

RANK OF TRACKING ACCURACY FOR MOTIONS AND VELOCITIES (RED: 1ST, BLUE: 2ND, GREEN: 3RD)

Motion	go-straight				stop-and-go				turns				sharp-turn				Total
Velocity (m/s)	1	5	10	20	1	5	10	20	1	5	10	20	1	5	10	20	
CV-CC	8	9	9	9	2	4	9	9	9	9	9	7	3	4	4	4	108
CV-PC	5	8	8	8	7	9	5	2	5	7	7	8	6	6	6	8	105
CT-CC	3	5	5	6	6	7	8	7	4	4	1	3	9	8	8	6	90
CT-PC	1	3	3	1	9	8	3	5	1	2	5	6	8	9	9	9	82
IMM-AUG-CC	6	2	1	5	1	1	7	8	6	6	3	5	4	3	3	3	**64**
IMM-AUG-PC	4	6	6	4	8	6	1	1	2	1	4	2	5	5	5	7	**67**
IMM-CV-MIX	9	7	7	7	4	3	6	6	8	8	8	9	1	2	2	2	89
IMM-CT-MIX	2	4	2	3	5	5	4	4	3	3	2	1	7	7	7	5	**64**
IMM-AUG-MIX	7	1	4	2	3	2	2	3	7	5	6	4	2	1	1	1	**51**

based on the typical speed of pedestrians, bicycles, and cars in the real environment as 1 m/s, 5 m/s, 10 m/s, and 20 m/s. The first `go-straight` was set as a linear motion for 30 s. In the `stop-and-go`, objects reached the target speed within 5 s from a standstill and stopped from the target speed within 5 s. It also continued motion for 5 s between deceleration and acceleration. The `turns` consists of smooth left turns and smooth right turns. The angular velocity of `turns` was set at ± 0.1 rad/s for left and right turns of $\pm\pi/2$ rad. The `sharp-turns` consists of sharp left turns and sharp right turns, which were designed to rotate $\pm\pi/2$ rad in a single frame ($T = 0.1$ s). After every sharp turn, it linearly moves until the next sharp turn.

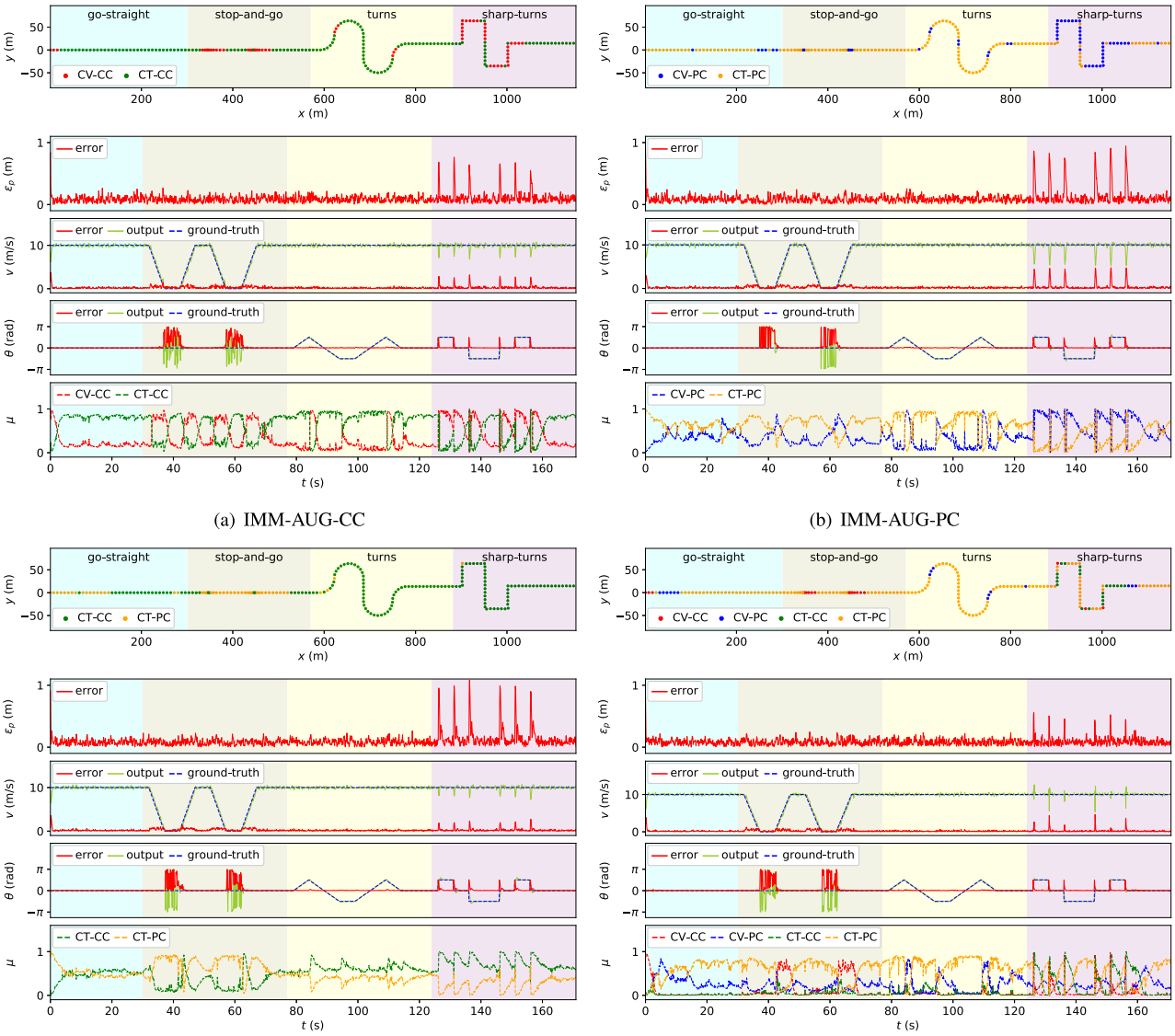
The input period was set to 10 Hz. The position noise was added through a normal distribution with a standard deviation of 0.1 m. Each motion in synthetic datasets has a linear motion for 10 s at the end of the motion, allowing the motion estimators to enter the subsequent motion in a stable status. Also, we estimated the appropriate noise variance of each single motion model through a grid search, a brute-force searching from a manually specified subset, for `go-straight` and `turns`. These experiments with synthetic datasets were performed with 100 Monte Carlo simulations for each linear velocity and motion.

B. Results and Analysis

The position error, ε_p was calculated as the root mean square error (RMSE) between the tracking result and the ground truth to evaluate the tracking accuracy. As shown in Fig. 4, the position errors of motion models were compared in proportion to the IMM-AUG-MIX. Moreover, we ranked the motion models on position error to demonstrate the robustness of motion models in Table V for each motion and target speed. A small rank-sum means that the motion model consistently achieves better tracking performance for all motions. For clear comparison, we colored the results in red, blue, and green in order, starting with the smallest rank sum in Table V. For IMM estimators, the mode probabilities were additionally visualized to confirm which modes are dominant for each motion as shown in Fig. 5.

1) Go Straight: CT-PC (8), IMM-CT-MIX (11), IMM-AUG-CC (14), and IMM-AUG-MIX (14) provide higher tracking accuracy than other models in the order, but the difference among them is not significant. The number in parentheses means the sum of ranks in the corresponding motion. The large noise covariance of the CV model, which is set for both straight and circular motion, increases the position error for `go-straight` motion. Therefore, CT models commonly show more accurate performance than CV models, when modeling a straight motion. Similarly, IMM filters with CT models produce lower position errors than those with CV models. IMM estimators mainly select CT-CC and CT-PC modes, as can be seen from the mode probabilities μ in Fig. 5. In particular, CT-CC and CT-PC of IMM-CT-MIX have similar mode probabilities because they produce equally accurate tracking results in linear motion.

2) Stop and Go: IMM estimators such as IMM-AUG-MIX (10), IMM-AUG-PC (16), and IMM-AUG-CC (17) achieve far superior accuracy to the other. CV-CC is more



(a) IMM-AUG-CC

(b) IMM-AUG-PC

(c) IMM-CT-MIX

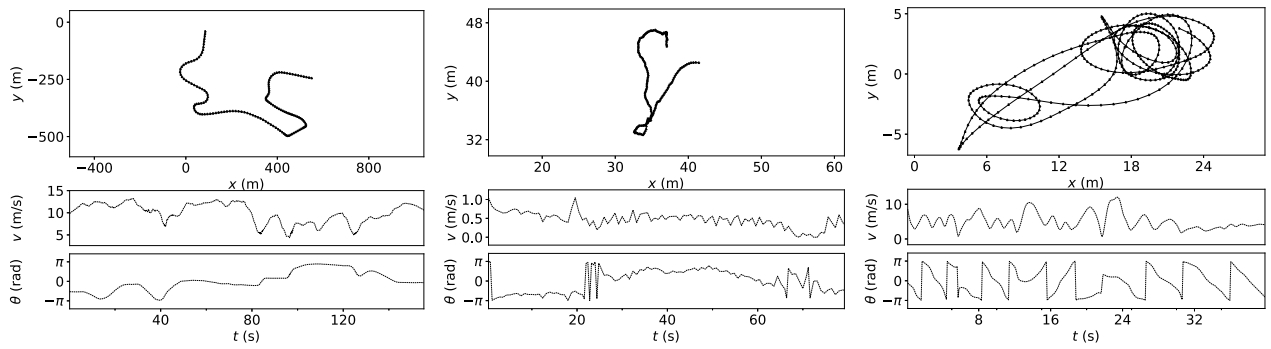
(d) IMM-AUG-MIX

Fig. 5. Tracking results of synthetic datasets for position, velocity, orientation, and mode probability over time when moving at 10 m/s.

accurate than other single motion models for lower target speeds (1 and 5 m/s). CT-PC (10 m/s) and CV-PC (20 m/s) are better for higher target speeds. However, with these results, it is not sufficient to analyze the performance of motion models for `stop-and-go` motion. `stop-and-go` consists of four motion parts: deceleration, stop, start, and acceleration. Therefore, we compared the dominance of the motion model according to detailed motions using the mode probability.

As shown in Fig. 5, for IMM-CT-MIX, models with Cartesian velocity states have more significant mode probability only when the object starts moving. For IMM-AUG-PC, when stopping and start, CV-PC has a larger mode probability than CT-PC. Specifically, the CC models can quickly react to sudden motion changes, such as the starting motion since velocity states $\dot{x}$ and $\dot{y}$ of CC models are independently separated. CV-PC is stable to the positional noise when stationary, and CT-PC tracks the object accurately in both acceleration and deceleration. These results are more evident in IMM-AUG-MIX. As shown in Fig. 5(d), CV-PC during the stop, CV-CC at the start, and CT-PC during deceleration and acceleration become dominant. This result demonstrates that each motion model has strengths in different situations, and the IMM-AUG-MIX, which combines them all, shows the most advanced tracking performance.

3) Turns: IMM-AUG-PC (9), IMM-CT-MIX (9), and CT-CC (12) achieve better tracking performance. Except for CV-related models such as CV-CC, CV-PC, and IMM-CV-MIX, the tracking accuracy of the other models is almost analogous. In other words, CT models are commonly superior to CV models similar to `go-straight`. The IMM-CV-MIX containing only CV modes exhibits a higher position error than the IMM-CT-MIX containing CT modes. Moreover, CT-CC and CT-PC are primarily selected in IMM filters. Through the results of `go-straight` and `turns` motions,



(a) Scene 2 of Car (b) Scene 2 of Pedestrian (c) Scene 2 of UAV

Fig. 6. Real objects datasets for cars, pedestrians, and UAVs. We sampled three scenes for each object to evaluate the tracking performance.

it is demonstrated that the CT model is a suitable motion model for normal smooth motions. A single motion model of CT, such as CT-CC and CT-PC, can also achieve the same performance as multi-motion estimators in such normal motions.

4) Sharp Turns: IMM-AUG-MIX (5), IMM-CV-MIX (7), and IMM-AUG-CC (13) produce much lower position errors than other motion models. CT-PC, which performs best in regular motions, has the lowest tracking accuracy. On the other hand, CV-CC performs the best among single motion models in `sharp-turns`. As shown in Fig. 5(a) and (d), CV-CC is mainly selected at abrupt turns, and as it gradually stabilizes, the mode probability of other models increases. Also, IMM-CT-MIX mostly chose CT-CC in `sharp-turns` motion. To the specific, CC mode is more adaptive to sharp orientation changes because CC mode has independent axial velocity variances. Briefly, CV and CC provide better performance than CT and PC, respectively, for a sudden large turn.

To sum up, each single motion model exhibits different tracking accuracy depending on the speed and the rate of motion change. Since a single motion model cannot reliably track an object moving in diverse motions, IMM estimators with suitable combinations achieve relatively better tracking performance. As the above results show, the mode probability of multiple motion models generally depends on the performance of single motion models for each motion. CT models such as CT-CC and CT-PC present reliable tracking accuracy in both linear and curvilinear motions, but degrade in abruptly changing motions as `stop-and-go` and `sharp-turn`. Therefore, in rapidly changing motions, CV-CC and CV-PC are mainly selected over CT models in the IMM estimator.

As shown in Table V, IMM-AUG-PC with state augmentation for motion models of polar velocity representation improves tracking performance in smooth motion, but not in sharp motion changes. On the other hand, IMM-CV-MIX with state mixing for two CV models provides more accurate target tracking for abrupt motion changes, but not for smooth motion changes. IMM-AUG-MIX with state augmentation and state mixing for all types of heterogeneous motion models shows the most reliable tracking performance for smooth motion and sharp motion changes, regardless of straight or curved motions. In other words, the proposed adaptive target tracking allows heterogeneous motion models to interact without bias, thus completely complementing the weakness of each single motion model with each other.

VI. Experiments With Real Datasets

A. Experimental Setting

The tracking performance was evaluated for real road objects with different motion characteristics, such as cars, pedestrians, and UAVs. Therefore, we employed datasets for a car, pedestrian, and UAV from the KITTI odometry dataset, New York's grand central station dataset, and UZH-FPV drone racing dataset, respectively, as shown in Fig. 6.

1) Car: KITTI odometry dataset [32] was recorded by driving around the mid-size city of Karlsruhe, in rural areas, and on highways. This dataset is mainly used to evaluate the results of monocular or stereo visual odometry, but we utilized it to evaluate a single target tracking for a car. We selected 4 scenes out of 11 sequences with ground-truth trajectories and extracted trajectories of 1500 frames in 10 Hz.

2) Pedestrian: To demonstrate the target tracking for pedestrians, we used New York's grand central station dataset [33]. This dataset includes manually annotated pedestrian trajectories in a crowded space. Moreover, it contains longer trajectories than other pedestrian datasets. We selected four non-linear pedestrian trajectories with 200 frames at 2.5 Hz.

3) UAV: UZH-FPV drone racing dataset [34] is the most aggressive visual-inertial odometry dataset for UAVs. The drones in the dataset move dynamically both indoors and outdoors, with large accelerations and rotations. We selected four scenes facing forward, including two indoor scenes and two outdoor scenes. UZH-FPV dataset was originally captured in 1000 Hz, but we converted those ground-truth trajectories to 400 frames in 10 Hz like real detection observations.

Each dataset for one object type has four scenes: one for noise covariance estimation through a grid search and three for performance evaluation. We added a normally distributed positional noise with a standard deviation of 0.01 m to the pedestrian dataset and 0.1 m to the car and drone datasets to produce real-like detection. We estimated the noise covariance of each single motion model for one scene of each object type, and the IMM estimator utilized the single motion models

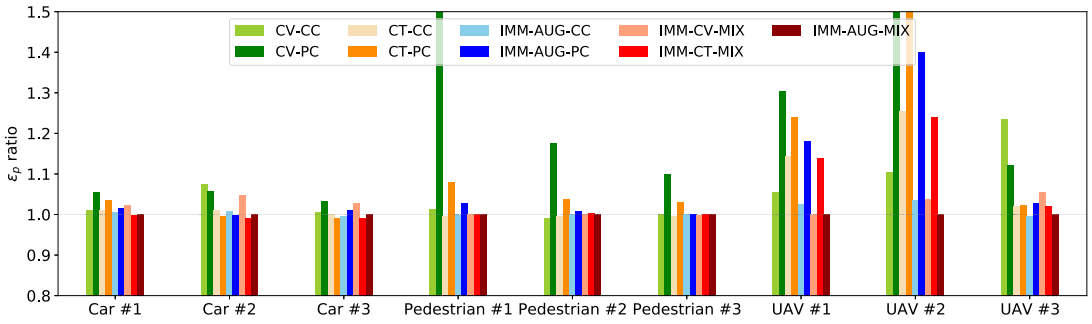


Fig. 7. Real experiment accuracy results for each pair of objects and scenes. All ε_p are normalized to the results of IMM-AUG-MIX to clearly compare performance with IMM-AUG-MIX.

TABLE VI

RANK OF TRACKING ACCURACY FOR REAL ROAD-OBJECTS (**RED**: 1ST, **BLUE**: 2ND, **GREEN**: 3RD)

Objects / Models	Car				Pedestrian				UAV				Total
	#1	#2	#3	Sum	#1	#2	#3	Sum	#1	#2	#3	Sum	
CV-CC	5	9	6	20	6	1	3	**10**	4	4	9	17	47
CV-PC	9	8	9	26	9	9	9	27	9	8	8	25	78
CT-CC	4	6	5	15	1	2	1	**4**	6	6	4	16	**35**
CT-PC	8	2	1	**11**	8	8	8	24	8	9	5	22	57
IMM-AUG-CC	3	5	3	**11**	2	5	5	**12**	3	2	1	**6**	**29**
IMM-AUG-PC	6	3	7	16	7	7	6	20	7	7	6	20	56
IMM-CV-MIX	7	7	8	22	4	4	2	**10**	1	3	7	**11**	43
IMM-CT-MIX	1	1	2	**4**	5	6	7	18	5	5	3	13	**35**
IMM-AUG-MIX	2	4	4	**10**	3	3	4	**10**	2	1	2	**5**	**25**

with these noise covariances. The real experiment was also performed with 100 Monte Carlo simulations for each pair of objects and scenes.

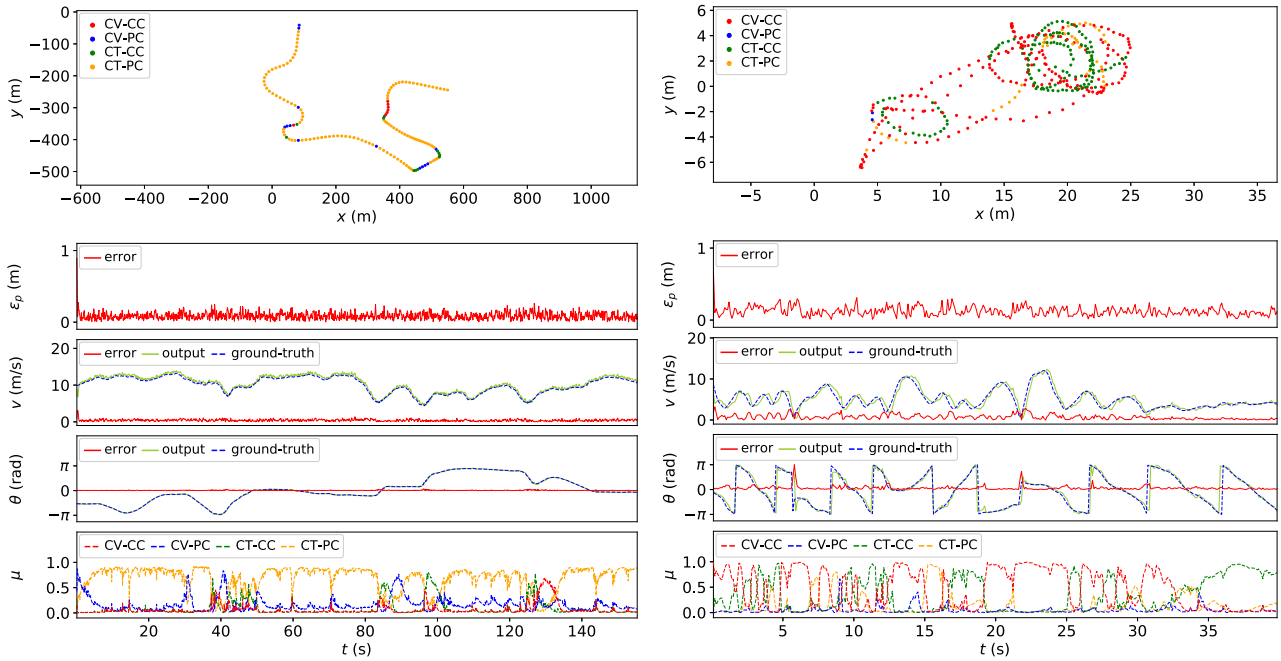
B. Results and Analysis

We compared the tracking accuracy of motion estimators by measuring the position errors ε_p for each pair of objects and scenes, as shown in Fig. 7 and in Table VI.

1) Car: The trajectories are mostly smooth straight and curved motion with few sharp turns as shown in Fig. 6(a). Therefore, CV models commonly produce larger ε_p than CT models in a single motion model comparison. Likewise, IMM-CV-MIX has lower tracking performance than IMM-CT-MIX. Except for IMM-CV-MIX, the IMM estimators have small tracking errors and no significant performance difference. As shown in Fig. 8(a), CT-PC of IMM-AUG-MIX is mostly dominant in normal motions. Other alternative motion models back up CT-PC when the orientation and velocity abruptly and largely change, as in the experimental results of synthetic datasets. Specifically, at 40 s, since the orientation is sharply changed, the CV-CC and CV-PC modes sequentially become dominant with a large mode probability. Also, since CT-CC has a lower position error than CT-PC at low speed and sharp turn, the mode probability of CT-CC becomes higher at 100 s. At 130 s, the mode probabilities of CT-CC and CV-CC sequentially increase. When the linear and angular velocities are simply decelerated, CT-CC dominates for a short time. After that, because both velocities rapidly change to acceleration, CV-CC becomes the dominant mode to stabilize unstable motion estimation. In short, CV-CC is needed to recover from unstable estimates caused by sudden starts, rapid accelerations and sharp turns. Cars move fast with smooth turns on highways, so IMM-CT-MIX is more suitable for object tracking than IMM-AUG-MIX, which requires more computation. On the other hand, cars include not only smooth motion, but also sudden motion changes such as stop-and-go and U-turns on urban roads. Therefore, IMM-AUG-MIX using state augmentation and mixing is essential for adaptive object tracking on urban roads.

2) Pedestrian: As shown in Fig. 6(b), in crowded spaces such as train stations, pedestrians usually move at low speeds, sometimes with sudden large turns or back-and-forth motions. Therefore, motion models with Cartesian velocity states provide much more accurate tracking than those with polar velocity states in such a non-linear slow motions. Notably, CV-CC and CT-CC achieve low positional errors even compared to IMM estimators. On the other hand, CV-PC and CT-PC generate high position errors. Except for IMM-AUG-PC, which contains only PC models, IMM estimators with at least one CC model do not differ significantly from each other in tracking accuracy. The proposed IMM-AUG-MIX takes second place with CV-CC and IMM-CV-MIX. However, unlike inconsistent motion in crowded space, pedestrians move smoothly and monotonously in non-crowded open spaces such as parks. In other words, pedestrians exhibit sharp and smooth movement patterns depending on the complexity of their surroundings. Therefore, the superiority of IMM-AUG-MIX for pedestrian tracking will become evident in spaces with varying degrees of openness and congestion.

3) UAV: UAVs have the most dynamic motions, rapidly moving with sharp turns. For scenes 1 and 2, CV and CC are better than CT and PC, respectively, but for scene 3, almost the



(a) Scene 2 of Car

(b) Scene 2 of UAV

Fig. 8. Real object tracking results of IMM-AUG-MIX for position, velocity, orientation, and mode probability over time.

opposite. This is because the UAV motions in scenes 1 and 2 change more dynamically than in scene 3. As shown in Fig. 8(b), CV-CC and CT-CC alternately have high mode probabilities. Specifically, the direction of rotation changes while moving rapidly at 6 s, 13 s, and 22 s. Therefore, CV-CC takes all mode probabilities to recover from unstable motion estimation quickly. From 34 s to 40 s, CT-CC prevails because the target moves at low linear and high angular velocities. IMM-AUG-MIX attains remarkable tracking performance in all UAV scenes, and its superiority is strongly demonstrated in more dynamic scenes. The dynamically changing mode probabilities of IMM-AUG-MIX in Fig. 8(b) also prove that all four modes actively interact in the IMM estimator. This completely complementary relationship enables adaptive tracking of flying drones in diverse motions and dynamic changes.

In summary, IMM-AUG-MIX achieves the lowest sum of rank, as reported in Table VI. It demonstrates the tracking reliability of IMM-AUG-MIX for a variety of objects with different motion characteristics. Similar to experiment results on synthetic datasets, multi-motion combinations of different model complexity (straight and circular motions) and velocity representations (Cartesian and polar velocities) provides evenly-enhanced performance for all object tracking. It is also confirmed that the proposed state mixing of an IMM estimator properly complements the weaknesses of each velocity representations. Consequently, IMM-AUG-MIX with the interaction of heterogeneous motion models is the most adaptable IMM estimator for tracking moving objects with diverse motions and change rates.

VII. Conclusion

We proposed an IMM-based adaptive target tracking that can integrate multiple motion models with heterogeneous velocity representations without bias and be extended to combine motion models with different model complexity. Therefore, the proposed IMM-AUG-MIX can contain multiple motion models with different state dimensions through state augmentation and different velocity representations through state mixing. Experiments with synthetic and real datasets demonstrated that an IMM estimator with heterogeneous motion models significantly improves tracking performance for moving objects with various motions and motion change rates. Moreover, an in-depth analysis of the motion models helps to understand the strengths and weaknesses of each motion model for different tracking situations. The proposed adaptive target tracking with interacting heterogeneous motion models can be applied to many automotive applications, such as localization and multi-object tracking. Also, the improvement of the proposed method will be noticeable on urban roads where various dynamic objects such as cars, pedestrians, and cyclists exist.

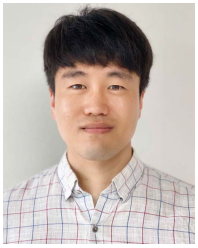
As further work, we will apply IMM-AUG-MIX to multi-target tracking in complex urban environments by simultaneously fine-tuning models for multiple objects. We expect that the proposed method can be applied to combine not only the four adopted motion models but also other motion models with heterogeneous state definitions and dimensions.

References

[1] J. Li, W. Zhan, Y. Hu, and M. Tomizuka, "Generic tracking and probabilistic prediction framework and its application in autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3634–3649, Sep. 2020.

[2] A. Kamann, J. B. Bielmeier, S. Hasirlioglu, U. T. Schwarz, and T. Brandmeier, "Object tracking based on an extended Kalman filter in high dynamic driving situations," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

[3] F. Camara *et al.*, "Pedestrian models for autonomous driving. Part I: Low-level models, from sensing to tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6131–6151, Jul. 2020.
[4] M. Tsogas, A. Polychronopoulos, and A. Amditis, "Unscented Kalman filter design for curvilinear motion models suitable for automotive safety applications," in *Proc. 7th Int. Conf. Inf. Fusion*, 2005, p. 8.
[5] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *Proc. Int. Conf. Inf. Fusion*, 2008, pp. 1–6.
[6] R. Schubert, C. Adam, M. Obst, N. Mattern, V. Leonhardt, and G. Wanielik, "Empirical evaluation of vehicular models for ego motion estimation," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 534–539.
[7] F. Gustafsson and A. J. Isaksson, "Best choice of coordinate system for tracking coordinated turns," in *Proc. 35th IEEE Conf. Decis. Control*, Dec. 1996, pp. 3145–3150.
[8] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.
[9] X. Yuan, C. Han, Z. Duan, and M. Lei, "Comparison and choice of models in tracking target with coordinated turn motion," in *Proc. 7th Int. Conf. Inf. Fusion*, 2005, p. 6.
[10] M. R. Morelande and N. J. Gordon, "Target tracking through a coordinated turn," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2005, p. 21.
[11] M. Roth, G. Hendeby, and F. Gustafsson, "EKF/UKF maneuvering target tracking using coordinated turn models with polar/Cartesian velocity," in *Proc. Int. Conf. Inf. Fusion*, 2014, pp. 1–8.
[12] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Singapore: YBS Publishing Storrs, 2011.
[13] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Control*, vol. AC-33, no. 8, pp. 780–783, Aug. 1988.
[14] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part V. Multiple-model methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1255–1321, Oct. 2005.
[15] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 103–123, Jan. 1998.
[16] S. J. Lee, Y. Motai, and H. Choii, "Tracking human motion with multichannel interacting multiple model," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1751–1763, Aug. 2013.
[17] M. Schreier, V. Willert, and J. Adamy, "Compact representation of dynamic driving environments for ADAS by parametric free space and dynamic object maps," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 367–384, Feb. 2016.
[18] X.-R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," *IEEE Trans. Autom. Control*, vol. 41, no. 4, pp. 478–493, Apr. 1996.
[19] R. Orjuela, B. Marx, J. Ragot, and D. Maquin, "State estimation of nonlinear systems based on heterogeneous multiple models: Some recent theoretical results," in *Proc. Workshop Adv. Control Diagnosis*, 2009, pp. 1–7.
[20] R. Orjuela, B. Marx, J. Ragot, and D. Maquin, "Nonlinear system identification using heterogeneous multiple models," *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 1, pp. 103–115, Mar. 2013.
[21] T. Yuan, Y. Bar-Shalom, P. Willett, E. Mozeson, S. Pollak, and D. Hardiman, "A multiple IMM estimation approach with unbiased mixing for thrusting projectiles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3250–3267, Oct. 2012.
[22] R. Lopez, P. Danès, and F. Royer, "Extending the IMM filter to heterogeneous-order state space models," in *Proc. 49th IEEE Conf. Decis. Control (CDC)*, Dec. 2010, pp. 7369–7374.
[23] K. Granström, P. Willett, and Y. Bar-Shalom, "Systematic approach to IMM mixing for unequal dimension states," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 4, pp. 2975–2986, Oct. 2015.
[24] N. Ou and S. Wang, "A new approach to unequal dimension states mixing for IMM estimator," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–4.
[25] D. Laneuville, "Polar versus Cartesian velocity models for maneuvering target tracking with IMM," in *Proc. IEEE Aerosp. Conf.*, Mar. 2013, pp. 1–15.
[26] R. Visina, Y. Bar-Shalom, and P. Willett, "Multiple-model estimators for tracking sharply maneuvering ground targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 3, pp. 1404–1414, Jun. 2018.
[27] Y. Gao, Y. Liu, X. R. Li, and V. P. Jilkov, "Multiple-model estimation with heterogeneous state representation," in *Proc. Int. Conf. Inf. Fusion*, 2015, pp. 1840–1847.
[28] C. Allig and G. Wanielik, "Unequal dimension track-to-track fusion approaches using covariance intersection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5881–5886, Jun. 2022.
[29] K. M. Lynch and F. C. Park, *Modern Robotics: Mechanics, Planning, and Control*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
[30] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
[31] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA, USA: MIT Press, 2005.
[32] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.
[33] S. Yi, H. Li, and X. Wang, "Understanding pedestrian behaviors from stationary crowd groups," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3488–3496.
[34] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are we ready for autonomous drone racing? The UZH-FPV drone racing dataset," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 6713–6719.



Ki-In Na received the B.S. degree in mechanical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2009, and the M.S. degree in robotics from KAIST, Daejeon, Republic of Korea, in 2011, where he is currently pursuing the Ph.D. degree. Since 2011, he has been a Research Scientist at the Electronics and Telecommunication Research Institute (ETRI), Daejeon. His current research interests include detection and tracking of moving objects, social navigation, human–robot interaction, and artificial intelligence for real applications.

Sunglok Choi (Member, IEEE) received the B.S. degree in mechanical and aerospace engineering from Seoul National University, Seoul, Republic of Korea, in 2006, and the M.S. and Ph.D. degrees in robotics from KAIST, Daejeon, Republic of Korea, in 2008 and 2019, respectively. From 2008 to 2021, he was a Research Scientist at the Electronics and Telecommunication Research Institute (ETRI), Daejeon. Since 2021, he has been with the Department of Computer Science and Engineering, Seoul National University of Science and Technology (SEOULTECH). His research interests include autonomous navigation, 3D computer vision, and robust regression.

Jong-Hwan Kim (Fellow, IEEE) received the Ph.D. degree in electronics engineering from Seoul National University, Seoul, Republic of Korea, in 1987. Since 1988, he has been with the School of Electrical Engineering, KAIST, Daejeon, Republic of Korea, where he is currently leading the Robot Intelligence Technology Laboratory as a KT Endowed Chair Professor. He is also the Director for KoYoung–KAIST AI Joint Research Center and Machine Intelligence and Robotics Multisponsored Research and Education Platform. He has authored five books, ten edited books, and around 450 refereed articles in technical journals and conference proceedings. His current research interests include intelligence technology, machine intelligence learning, and AI robots.