# Robust Interacting Multiple Model Filter Based on Student's *t*-Distribution for Heavy-Tailed Measurement Noises

**Dong Li** [1,*] and **Jie Sun** [2]

[1]   Unit 94, PLA 91550, Dalian 116023, China
[2]   Unit 93, PLA 91550, Dalian 116023, China; sunshine_scholar@163.com
*   Correspondence: nudtlidong@163.com; Tel.: +86-0411-8085-8444

**Abstract:** In maneuvering target tracking applications, the performance of the traditional interacting multiple model (IMM) filter deteriorates seriously under heavy-tailed measurement noises which are induced by outliers. A robust IMM filter utilizing Student's *t*-distribution is proposed to handle the heavy-tailed measurement noises in this paper. The measurement noises are treated as Student's *t*-distribution, whose degrees of freedom (dof) and scale matrix are assumed to be governed by gamma and inverse Wishart distributions, respectively. The mixing distributions of the target state, dof, and scale matrix are achieved through the interacting strategy of IMM filter. These mixing distributions are used for the initialization of time prediction. The posterior distributions of the target state, dof, and scale matrix conditioned on each mode are obtained by employing variational Bayesian approach. Then, the target state, dof, and scale matrix parameters are jointly estimated. A variational method is also given to estimate the mode probability. The unscented transform is utilized to solve the nonlinear estimation problem. Simulation results show that the proposed filter improves the estimation accuracy of target state and mode probability over existing filters under heavy-tailed measurement noises.

**Keywords:** IMM; filter; robust; heavy-tailed measurement noises; Student's *t*-distribution; variational Bayesian

## 1. Introduction

The Kalman filter is widely used for target tracking due to the low computational complexity for real time processing and the statistic optimality under a linear state space model with Gaussian noises. Because of the model uncertainty of target maneuvering and the heavy-tailed measurement noises deduced by outliers, the performance of Kalman filter may break down dramatically, and even the convergence cannot be guaranteed. Addressing model uncertainty and investigating robust filter against heavy-tailed measurement noises are two important research topics in the target tracking field.

Model uncertainty in maneuvering target tracking exhibits that the motion model of target is time varying and that the switching time between different models is unknown. Due to the model uncertainty, the model used by Kalman filter may not match the actual dynamics so that the filter may gradually diverge. Multiple-model (MM) methods are considered as the mainstream approach to maneuvering target tracking under model uncertainty [1]. These methods establish a finite set of models to describe the different motion patterns for maneuvering target, and achieve the ultimate estimate of state by a certain combination of the estimates from each model. In the literature, the MM methods can be categorized into two different types: fixed-structure MM (FSMM) methods which use a fixed model set at all times and variable-structure MM (VSMM) methods in which the model set is not

fixed. The generalized pseudo-Bayesian estimator of order $n$ (GPBn) filter and the interacting multiple model (IMM) filter [2] are the typical FSMM methods. Among them, the IMM filter is considered to be the best compromise between complexity and performance, and has been successfully applied in a large number of tracking applications [3–8]. In IMM filter, several sub-filters operate in parallel and cooperate with each other through an interacting strategy, leading to improved performance of estimation. For the FSMM methods, if the model set used does not match the set of actual models of target movement, the performance can be degraded. To cope with this problem, several VSMM methods have been proposed, such as likely mode set (LMS) filter [9], expected-mode augmentation (EMA) filter [10], equivalent-model augmentation (EqMA) filter [11], and hybrid grid multiple model (HGMM) filter [12]. These filters use different model-set adaptation schemes to adjust the model set. For example, the EqMA filter augments the model set by a new variable model depending directly on the history information from the recent models, while the HGMM filter adapts the model set by the hybrid grid scheme which consists of a fixed coarse grid and an adaptive fine one. In recent years, some researchers have proposed Sequential Monte Carlo (SMC) methods [13,14] for dealing with the problem of model uncertainty in target tracking. These methods maintain several particle filters (PFs) in parallel and fuse the results of PFs to provide the posterior probability of model and the posterior distribution of state. In addition, the SMC method in [14] uses a scale parameter to detect the model change and does not require the mode transition probabilities. This strategy of handling model change is quite different from the above mentioned FSMM and VSMM methods.

Besides the model uncertainty, the outlier measurements existing in real tracking scenarios should also be handled. Outliers can induce heavy-tailed non-Gaussian noises, and the performance of Kalman filter may be seriously degraded since its underlying optimality criterion is minimum mean square error, which is very sensitive to heavy-tailed noises. A large number of robust filters have been proposed to improve the robustness against non-Gaussian noises. The Huber-based filter [15–19] is a classical robust method based on a combined minimum $l_1$ and $l_2$ norm estimation. The maximum correntropy Kalman filter (MCKF) [20–24] is another effective way to suppress the impact of non-Gaussian noises. It uses the robust maximum correntropy criterion (MCC) as the optimality criterion. Simulation experiments show that the MCKF outperforms Huber-based filter in estimation accuracy [20,22–24]. However, neither Huber-based filter nor MCKF takes advantage of the inherent heavy-tailed feature of the noise distribution. Thus, their estimation precision is limited. The Student's $t$-distribution is robust to outliers since it has heavier tail compared with Gaussian distribution. Based on this idea, many robust filters utilizing Student's $t$-distribution has been proposed [25–30]. The Student's $t$ filter [25,26] is derived by modeling the posterior distribution of the state, the heavy-tailed process and measurement noises as Student's $t$-distributions. The filter recursions are designed according to the framework of Bayesian filter. This filter may induce significant bias since an approximated method of moment matching is employed to prevent the growth of the degrees of freedom (dof) of Student's $t$-distribution. The variational Bayesian (VB) based Student's $t$ filter [27–30] represents the Student's $t$-distribution of noises as an infinite mixture of Gaussians, and employs VB approach to jointly estimate the state and the unknown parameters of the Student's $t$-distribution. This filter can achieve higher estimation accuracy than the Student's $t$ filter [29,30].

In maneuvering target tracking, the problems of model uncertainty and heavy-tailed measurement noises could exist simultaneously. Investigating robust IMM filter to cope with both model uncertainty and heavy-tailed noises is meaningful. As far as we know, Shen et al. [31] considered utilizing Student's $t$-distribution to improve the robustness of IMM filter for the first time. They modeled the heavy-tailed measurement noises as Student's $t$-distribution and employed IMM and VB approaches to estimate the target state, the probability of motion mode and the parameters of noises. However, the dof of Student's $t$-distribution, which is determined by the estimated parameters of an auxiliary random variable, becomes very large after several filtering steps. Then, the Student's $t$-distribution converges to Gaussian distribution and loses the heavy-tailed property. Therefore, the robustness of this filter against heavy-tailed measurement noises is limited. In addition, the scale matrix of Student's $t$-distribution

in this filter is assumed to be known exactly. However, this premise is usually unavailable since the scale matrix is unknown or time varying in many actual applications. Furthermore, this filter is based on linear systems and should be generalized to nonlinear systems which are more common in maneuvering target tracking.

Motivated by the above discussions, we propose a new robust IMM filter to better handle the heavy-tailed measurement noises and the system nonlinearity. The posterior distribution of target state is approximated by Gaussian distribution, and the measurement noises are modeled as Student's *t*-distribution. The unknown dof and scale matrix of the Student's *t*-distribution are assumed to obey gamma and inverse Wishart distributions, respectively. Then, the state and the parameters of gamma and inverse Wishart distributions for each mode are interacted via IMM mixing, and they are jointly estimated by using VB approach. An approximation method is given to derive the mode-conditioned predicted likelihood which is used for calculating the mode probability. The unscented transform (UT) is employed to tackle the system nonlinearity. The growth of dof estimates is prevented and the estimation accuracy of state is improved over existing filters under heavy-tailed measurement noises as shown in our simulation example.

## 2. System Model and Assumptions

Consider the following jump Markov system

$$\begin{cases} x_k = F_{k-1}(x_{k-1}, r_k) + w_{k-1}(r_k) \\ z_k = H_k(x_k, r_k) + \varepsilon_k \end{cases}$$

where $k$ is the scan time index, $x_k \in \mathrm{R}^n$ and $z_k \in \mathrm{R}^m$ are the state and measurement vectors respectively, $r_k$ is the mode of target movement taking value in a finite set $\{1, 2, \cdots, M\}$, $M$ is the total number of dynamic models, $F_{k-1}(x_{k-1}, r_k)$ and $H_k(x_k, r_k)$ are the state transition function and measurement function based on the mode $r_k$ respectively, $w_{k-1}(r_k)$ is the process noise vector based on the mode $r_k$, and $\varepsilon_k$ is the measurement noise vector. The mode $r_k$ is assumed to be a homogeneous Markov chain with the transition probability $\pi_{ij} = p(r_k = j | r_{k-1} = i)$, and $w_{k-1}(r_k)$ is assumed to be uncorrelated zero-mean Gaussian noises with covariance matrix $Q_{k-1}(r_k)$. Throughout this paper, we abbreviate $F_{k-1}(x_{k-1}, r_k = i)$, $H_k(x_k, r_k = i)$, $w_{k-1}(r_k = i)$ and $Q_{k-1}(r_k = i)$ by $F_{k-1}^i(x_{k-1})$, $H_k^i(x_k)$, $w_{k-1}^i$ and $Q_{k-1}^i$, respectively. Then, the transition probability density function (PDF) of state is given by $p(x_k | x_{k-1}, r_k = i) = \mathrm{N}(x_k; F_{k-1}^i(x_{k-1}), Q_{k-1}^i)$, where $\mathrm{N}(\cdot; m, P)$ denotes a Gaussian PDF with mean $m$ and covariance matrix $P$.

When measurement outliers occur, the distribution of measurement noises has heavy-tailed non-Gaussian characteristic. The performance of traditional IMM filter based on Gaussian noise assumption may dramatically degrade since it is very sensitive to heavy-tailed noises. Since the Student's t-distribution has heavier tail and is more robust to outliers than Gaussian distribution, we model the measurement noises as Student's t-distribution, i.e., $\varepsilon_k \sim \mathrm{Std}(\varepsilon_k; 0, R_k, v_k)$, where $\mathrm{Std}(\cdot; \mu, R, v)$ stands for a Student's $t$ PDF with mean $\mu$, scale matrix $R$ and dof $v$. The Student's t-distribution can be represented by an infinite mixture of Gaussian distributions as [32]

$$\mathrm{Std}(x; \mu, R, v) = \int_0^\infty \mathrm{N}(x; \mu, R/\lambda) \mathrm{G}(\lambda; v/2, v/2) d\lambda$$

where $\lambda$ is an auxiliary random variable, $\mathrm{G}(\cdot; a, b)$ denotes a Gamma PDF with shape parameter $a$ and rate parameter $b$. The expression for $\mathrm{G}(\cdot; a, b)$ is

$$\mathrm{G}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \quad x \geq 0 \tag{1}$$

where $\Gamma(a) = \int_0^{+\infty} \exp(-t)t^{a-1}dt$ is the Gamma function. Then, the likelihood PDF conditioned on the mode $r_k = i$ can be written as

$$p(z_k|x_k, r_k = i) = \begin{array}{l} \text{Std}(z_k; H_k^i(x_k), R_k, v_k) \\ \int_0^\infty N(z_k; H_k^i(x_k), R_k/\lambda_k)G(\lambda_k; v_k/2, v_k/2)d\lambda_k \end{array} \tag{2}$$

The tail thickness of Student's *t*-distribution is controlled by the dof. When the dof decreases, the tail becomes thicker. When the dof goes to infinity, the Student's *t*-distribution approaches a Gaussian distribution. Shen et al. [31] assumed that the posterior distribution of the auxiliary variable $\lambda_k$ is a Gamma distribution, i.e., $p(\lambda_k|r_k = i, z_{1:k}) = G(\lambda_k; \hat{\alpha}_k^i, \hat{\beta}_k^i)$ and used the VB approach to estimate the parameters $\hat{\alpha}_k^i$ and $\hat{\beta}_k^i$. However, the estimates of both $\hat{\alpha}_k^i$ and $\hat{\beta}_k^i$ which determine the dof of Student's *t*-distribution become large after several filtering steps. Thus, the distribution of measurement noises loses the heavy-tailed property and the robustness of the filter cannot be guaranteed. To prevent the growth of the dof estimates, we assume that the posterior distribution of the dof $v_k$ is a Gamma distribution, i.e., $p(v_k|r_k = i, z_{1:k}) = G(v_k; \hat{a}_k^i, \hat{b}_k^i)$, and estimate the parameters $\hat{a}_k^i$ and $\hat{b}_k^i$ instead. Furthermore, in order to deal with the uncertainty of scale matrix $R_k$, we choose inverse Wishart distribution as the posterior distribution of $R_k$, i.e., $p(R_k|r_k = i, z_{1:k}) = IW(R_k; \hat{t}_k^i, T_k^i)$, where $IW(\cdot; t, T)$ stands for an inverse Wishart PDF with degree $t$ and inverse scale matrix $T$. The PDF of inverse Wishart distribution is defined by

$$IW(R; t, T) = \frac{\det(T)^{0.5t}}{2^{0.5tm}\Gamma_m(0.5t)}\det(R)^{-0.5(t+m+1)}\exp(-0.5\text{tr}(TR^{-1})) \tag{3}$$

where $\Gamma_m(0.5t) = \pi^{0.25m(m-1)}\prod_{i=0}^{m-1}\Gamma(0.5(t-i))$ Then, the unknown statistic of the scale matrix $R_k$ can be determined by estimating the parameters $\hat{t}_k^i$ and $T_k^i$.

## 3. Design of Robust IMM Filter

### 3.1. Model Interaction

We assume that the mode-conditioned posterior distributions of the state $x_{k-1}$, the scale matrix $R_{k-1}$ and the dof $v_{k-1}$ at time $k-1$ are Gaussian, inverse Wishart, and Gamma distributions, respectively, and that they are mutually independent. Then, the joint mode-conditioned posterior PDF of $x_{k-1}$, $R_{k-1}$ and $v_{k-1}$ can be expressed as

$$\begin{aligned} p(x_{k-1}, R_{k-1}, v_{k-1}|r_{k-1} = j, z_{1:k-1}) &= p(x_{k-1}|r_{k-1} = j, z_{1:k-1})p(R_{k-1}|r_{k-1} = j, z_{1:k-1})p(v_{k-1}|r_{k-1} = j, z_{1:k-1}) \\ &= N(x_{k-1}; \hat{x}_{k-1}^j, P_{k-1}^j)IW(R_{k-1}; \hat{t}_{k-1}^j, T_{k-1}^j)G(v_{k-1}; \hat{a}_{k-1}^j, \hat{b}_{k-1}^j) \end{aligned}$$

Let $u_k^i = p(r_k = i|z_{1:k})$ and $u_{k|k-1}^i = p(r_k = i|z_{1:k-1})$ denote the posterior and predicted mode probabilities respectively, and $u_{k-1}^{ij} = p(r_{k-1} = i|r_k = j, z_{1:k-1})$ denote the mixing probability, we have

$$\begin{aligned} u_{k|k-1}^j &= \sum_{i=1}^M p(r_k = j|r_{k-1} = i, z_{1:k-1})p(r_{k-1} = i|z_{1:k-1}) \\ &= \sum_{i=1}^M \pi_{ij}u_{k-1}^i \end{aligned} \tag{4}$$

$$\begin{aligned} u_{k-1}^{ij} &= \frac{p(r_k=j|r_{k-1}=i,z_{1:k-1})p(r_{k-1}=i|z_{1:k-1})}{p(r_k=j|z_{1:k-1})} \\ &= \pi_{ij}u_{k-1}^i/u_{k|k-1}^j \end{aligned} \tag{5}$$

The mixing PDF of $x_{k-1}$, $R_{k-1}$ and $v_{k-1}$ is given by

$$
\begin{aligned}
p(x_{k-1}, R_{k-1}, v_{k-1}|r_k = i, z_{1:k-1}) &= \sum_{j=1}^{M} p(x_{k-1}, R_{k-1}, v_{k-1}|r_{k-1} = j, z_{1:k-1})p(r_{k-1} = j|r_k = i, z_{1:k-1}) \\
&= \sum_{j=1}^{M} u_{k-1}^{ji} N(x_{k-1}; \hat{x}_{k-1}^{j}, P_{k-1}^{j}) IW(R_{k-1}; \hat{t}_{k-1}^{j}, T_{k-1}^{j}) G(v_{k-1}; \hat{a}_{k-1}^{j}, \hat{b}_{k-1}^{j})
\end{aligned}
\tag{6}
$$

We approximate the above summed PDF in the right side of (6) by a single one as

$$
p(x_{k-1}, R_{k-1}, v_{k-1}|r_k = i, z_{1:k-1}) = N(x_{k-1}; \hat{x}_{k-1}^{0i}, P_{k-1}^{0i}) IW(R_{k-1}; \hat{t}_{k-1}^{0i}, T_{k-1}^{0i}) G(v_{k-1}; \hat{a}_{k-1}^{0i}, \hat{b}_{k-1}^{0i})
\tag{7}
$$

By matching the first and the second-order moments between (6) and (7) for the state distribution, we obtain the mixing state $\hat{x}_{k-1}^{0i}$ and the covariance matrix $P_{k-1}^{0i}$ as

$$
\hat{x}_{k-1}^{0i} = \sum_{j=1}^{M} u_{k-1}^{ji} \hat{x}_{k-1}^{j}
\tag{8}
$$

$$
P_{k-1}^{0i} = \sum_{j=1}^{M} u_{k-1}^{ji} (P_{k-1}^{j} + (\hat{x}_{k-1}^{j} - \hat{x}_{k-1}^{0i})(\hat{x}_{k-1}^{j} - \hat{x}_{k-1}^{0i})^{T})
\tag{9}
$$

However, the mixing parameters $\hat{t}_{k-1}^{0i}$ and $T_{k-1}^{0i}$ cannot be computed analytically by matching the first two moments of inverse Wishart distributions between (6) and (7) due to the complexity of second-order moment. We adopt the method of minimizing weighted Kullback–Leibler (KL) divergence in [33] to overcome this difficulty and derive that

$$
\hat{t}_{k-1}^{0i} = \sum_{j=1}^{M} u_{k-1}^{ji} \hat{t}_{k-1}^{j}
\tag{10}
$$

$$
T_{k-1}^{0i} = \sum_{j=1}^{M} u_{k-1}^{ji} T_{k-1}^{j}
\tag{11}
$$

For the Gamma distribution, we also use the method of matching the first two moments like the state. According to (6), the conditional mean and covariance of $v_{k-1}$ are given by

$$
\begin{aligned}
E(v_{k-1}|r_k = i, z_{1:k-1}) &= \int v_{k-1} p(v_{k-1}|r_k = i, z_{1:k-1})dv_{k-1} \\
&= \int v_{k-1} \sum_{j=1}^{M} u_{k-1}^{ji} G(v_{k-1}; \hat{a}_{k-1}^{j}, \hat{b}_{k-1}^{j})dv_{k-1} \\
&= \sum_{j=1}^{M} u_{k-1}^{ji} \int v_{k-1} G(v_{k-1}; \hat{a}_{k-1}^{j}, \hat{b}_{k-1}^{j})dv_{k-1}
\end{aligned}
$$

$$
\begin{aligned}
Var(v_{k-1}|r_k = i, z_{1:k-1}) &= \int (v_{k-1} - E(v_{k-1}|r_k = i, z_{1:k-1}))^2 p(v_{k-1}|r_k = i, z_{1:k-1})dv_{k-1} \\
&= \int (v_{k-1} - E(v_{k-1}|r_k = i, z_{1:k-1}))^2 \sum_{j=1}^{M} u_{k-1}^{ji} G(v_{k-1}; \hat{a}_{k-1}^{j}, \hat{b}_{k-1}^{j})dv_{k-1} \\
&= \sum_{j=1}^{M} u_{k-1}^{ji} \int (v_{k-1} - E(v_{k-1}|r_k = i, z_{1:k-1}))^2 G(v_{k-1}; \hat{a}_{k-1}^{j}, \hat{b}_{k-1}^{j})dv_{k-1}
\end{aligned}
$$

Since the mean and covariance of a Gamma PDF $G(\cdot; a, b)$ are $a/b$ and $a/b^2$ respectively, we have

$$
E(v_{k-1}|r_k = i, z_{1:k-1}) = \sum_{j=1}^{M} u_{k-1}^{ji} \hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j}
\tag{12}
$$

$$\text{var}(v_{k-1}|r_k = i, z_{1:k-1}) =$$
$$\sum_{j=1}^{M} u_{k-1}^{ji} \int \{(v_{k-1} - \hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j})^2 + (\hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j} - \text{E}(v_{k-1}|r_k = i, z_{1:k-1}))^2\} \text{G}(v_{k-1}; \hat{a}_{k-1}^{j}, \hat{b}_{k-1}^{j}) dv_{k-1}$$
$$= \sum_{j=1}^{M} u_{k-1}^{ji} (\hat{a}_{k-1}^{j}/(\hat{b}_{k-1}^{j})^2 + (\hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j} - \sum_{j=1}^{M} u_{k-1}^{ji} \hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j})^2) \tag{13}$$

On the other hand, according to (7), we have

$$\text{E}(v_{k-1}|r_k = i, z_{1:k-1}) = \hat{a}_{k-1}^{0i}/\hat{b}_{k-1}^{0i}, \tag{14}$$

$$\text{Var}(v_{k-1}|r_k = i, z_{1:k-1}) = \hat{a}_{k-1}^{0i}/(\hat{b}_{k-1}^{0i})^2 \tag{15}$$

By combining (12)–(15), the mixing parameters $\hat{a}_{k-1}^{0i}$ and $\hat{b}_{k-1}^{0i}$ for the Gamma distribution can be calculated as

$$\hat{a}_{k-1}^{0i} = \frac{(\sum_{j=1}^{M} u_{k-1}^{ji} \hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j})^2}{\sum_{j=1}^{M} u_{k-1}^{ji} (\hat{a}_{k-1}^{j}/(\hat{b}_{k-1}^{j})^2 + (\hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j} - \sum_{j=1}^{M} u_{k-1}^{ji} \hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j})^2)}, \tag{16}$$

$$\hat{b}_{k-1}^{0i} = \frac{\sum_{j=1}^{M} u_{k-1}^{ji} \hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j}}{\sum_{j=1}^{M} u_{k-1}^{ji} (\hat{a}_{k-1}^{j}/(\hat{b}_{k-1}^{j})^2 + (\hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j} - \sum_{j=1}^{M} u_{k-1}^{ji} \hat{a}_{k-1}^{j}/\hat{b}_{k-1}^{j})^2)}. \tag{17}$$

### 3.2. Time Prediction

The time prediction step is to derive the mode-conditioned predicted PDF $p(x_k, R_k, v_k|r_k = i, z_{1:k-1})$ at time $k$ by using the mixing PDF $p(x_{k-1}, R_{k-1}, v_{k-1}|r_k = i, z_{1:k-1})$ and the dynamical models at time $k-1$.

The mode-conditioned predicted PDF can be calculated by the following Chapman–Kolmogorov equation:

$$p(x_k, R_k, v_k|r_k = i, z_{1:k-1}) =$$
$$\int p(x_{k-1}, R_{k-1}, v_{k-1}|r_k = i, z_{1:k-1}) p(x_k, R_k, v_k|x_{k-1}, R_{k-1}, v_{k-1}, r_k = i) dx_{k-1} dR_{k-1} dv_{k-1} \tag{18}$$

We assume that the dynamical models of the state, the scale matrix and the dof are independent, i.e.,

$$p(x_k, R_k, v_k|x_{k-1}, R_{k-1}, v_{k-1}, r_k) = p(x_k|x_{k-1}, r_k) p(R_k|R_{k-1}, r_k) p(v_k|v_{k-1}, r_k). \tag{19}$$

According to (7), (18) and (19), we have

$$p(x_k, R_k, v_k|r_k = i, z_{1:k-1}) = \int \text{N}(x_k; F_{k-1}^{i}(x_{k-1}), Q_{k-1}^{i}) \text{N}(x_{k-1}; \hat{x}_{k-1}^{0i}, P_{k-1}^{0i}) dx_{k-1} \times$$
$$\int p(R_k|R_{k-1}, r_k) \text{IW}(R_{k-1}; \hat{t}_{k-1}^{0i}, T_{k-1}^{0i}) dR_{k-1} \int p(v_k|v_{k-1}, r_k) \text{G}(v_{k-1}; \hat{a}_{k-1}^{0i}, \hat{b}_{k-1}^{0i}) dv_{k-1} \tag{20}$$

We can see that the mode-conditioned predicted distributions of the state, the scale matrix and the dof are also independent. These distributions are assumed to be Gaussian, inverse Wishart and Gamma distributions, respectively, which have the same form of PDF as the posterior distributions. Then, the mode-conditioned predicted PDF can be represented by

$$p(x_k, R_k, v_k|r_k = i, z_{1:k-1}) = \text{N}(x_k; \hat{x}_{k|k-1}^{i}, P_{k|k-1}^{i}) \text{IW}(R_k; \hat{t}_{k|k-1}^{i}, T_{k|k-1}^{i}) \text{G}(v_k; \hat{a}_{k|k-1}^{i}, \hat{b}_{k|k-1}^{i}). \tag{21}$$

By matching the first and the second-order moments between (20) and (21) for the state, we obtain the predicted mean and the covariance matrix as

$$\hat{x}^i_{k|k-1} = \int F^i_{k-1}(x_{k-1}) \mathrm{N}(x_{k-1}; \hat{x}^{0i}_{k-1}, P^{0i}_{k-1}) dx_{k-1}, \tag{22}$$

$$P^i_{k|k-1} = \int (F^i_{k-1}(x_{k-1}) - \hat{x}^i_{k|k-1})(F^i_{k-1}(x_{k-1}) - \hat{x}^i_{k|k-1})^{\mathrm{T}} \mathrm{N}(x_{k-1}; \hat{x}^{0i}_{k-1}, P^{0i}_{k-1}) dx_{k-1} + Q^i_{k-1} \tag{23}$$

However, it is hard to compute the predicted parameters of inverse Wishart and Gamma distributions since the dynamical models of the scale matrix and the dof are usually unknown in practice. Here, a heuristic dynamical model is chosen as in [34] by introducing a forgetting factor $\rho \in (0, 1]$, which indicates the extent of parameter fluctuation. Then, the predicted parameters are given by

$$T^i_{k|k-1} = \rho T^{0i}_{k-1}, \tag{24}$$

$$\hat{t}^i_{k|k-1} = \rho(\hat{t}^{0i}_{k-1} - m - 1) + m + 1, \tag{25}$$

$$\hat{a}^i_{k|k-1} = \rho \hat{a}^{0i}_{k-1}, \tag{26}$$

$$\hat{b}^i_{k|k-1} = \rho \hat{b}^{0i}_{k-1} \tag{27}$$

When $\rho = 1$, the parameters are kept stationary. When $\rho$ is close to 0, the parameters are high time-fluctuation.

### 3.3. Measurement Update

We define $\Theta = \{x_k, R_k, \lambda_k, v_k\}$. The measurement update step is to compute the mode-conditioned posterior PDF $p(\Theta|z_{1:k}, r_k = i)$ and the posterior mode probability $u^i_k$ when a new measurement $z_k$ is collected at time $k$. However, deriving the mode-conditioned posterior PDF is analytically intractable. We employ the VB approach [35] to obtain an approximated solution.

According to the VB approach, the mode-conditioned posterior PDF is approximated by a free form factored PDF:

$$p(\Theta|z_{1:k}, r_k = i) \approx q^i(x_k)q^i(R_k)q^i(\lambda_k)q^i(v_k)$$

The posterior PDFs $q^i(x_k)$, $q^i(R_k)$, $q^i(\lambda_k)$ and $q^i(v_k)$ are calculated by minimizing the KL divergence as

$$\min_{q^i(x_k),q^i(R_k),q^i(\lambda_k),q^i(v_k)} \mathrm{KL}(q^i(x_k)q^i(R_k)q^i(\lambda_k)q^i(v_k)\|p(\Theta|z_{1:k}, r_k = i)),$$

where $\mathrm{KL}(q(\cdot)\|p(\cdot)) \triangleq \int q(x) \ln(q(x)/p(x)) dx$ is the KL divergence. The solutions of the above optimal problem satisfy

$$\ln q^i(x_k) = \mathrm{E}_{q^i(R_k)q^i(\lambda_k)q^i(v_k)} \ln p(\Theta, z_{1:k}|r_k = i) + c_x, \tag{28}$$

$$\ln q^i(R_k) = \mathrm{E}_{q^i(x_k)q^i(\lambda_k)q^i(v_k)} \ln p(\Theta, z_{1:k}|r_k = i) + c_R, \tag{29}$$

$$\ln q^i(\lambda_k) = \mathrm{E}_{q^i(x_k)q^i(R_k)q^i(v_k)} \ln p(\Theta, z_{1:k}|r_k = i) + c_\lambda, \tag{30}$$

$$\ln q^i(v_k) = \mathrm{E}_{q^i(x_k)q^i(R_k)q^i(\lambda_k)} \ln p(\Theta, z_{1:k}|r_k = i) + c_v \tag{31}$$

where $c_x$, $c_R$, $c_\lambda$ and $c_v$ are constants independent of $x_k$, $R_k$, $\lambda_k$ and $v_k$, respectively.

According to (2) and (21), we have

$$
\begin{aligned}
p(\Theta, z_{1:k}|r_k = i) \ &= p(z_{1:k-1}|r_k = i)p(\Theta, z_k|r_k = i, z_{1:k-1}) \\
&= p(z_{1:k-1}|r_k = i)p(x_k|z_{1:k-1}, r_k = i)p(z_k|x_k, R_k, \lambda_k, r_k = i) \\
&\quad \times p(R_k|z_{1:k-1}, r_k = i)p(\lambda_k|v_k)p(v_k|z_{1:k-1}, r_k = i) \\
&= p(z_{1:k-1}|r_k = i)\mathrm{N}(x_k; \hat{x}^i_{k|k-1}, P^i_{k|k-1})\mathrm{N}(z_k; H^i_k(x_k), R_k/\lambda_k) \\
&\quad \times \mathrm{IW}(R_k; \hat{t}^i_{k|k-1}, T^i_{k|k-1})\mathrm{G}(\lambda_k; v_k/2, v_k/2)\mathrm{G}(v_k; \hat{a}^i_{k|k-1}, \hat{b}^i_{k|k-1})
\end{aligned}
\tag{32}
$$

Substituting (1), (3) and the expression of Gaussian PDF into (32), we obtain the detailed expression of $p(\Theta, z_{1:k}|r_k = i)$

$$
\begin{aligned}
\ln p(\Theta, z_{1:k}|r_k = i) =\ &-0.5(x_k - \hat{x}^i_{k|k-1})^{\mathrm{T}}(P^i_{k|k-1})^{-1}(x_k - \hat{x}^i_{k|k-1}) - \\
&0.5\lambda_k(z_k - H^i_k(x_k))^{\mathrm{T}}R_k^{-1}(z_k - H^i_k(x_k)) + \\
&0.5(m + v_k - 2)\ln \lambda_k - 0.5v_k\lambda_k + (\hat{a}^i_{k|k-1} - 1)\ln v_k - \hat{b}^i_{k|k-1}v_k + 0.5v_k\ln(0.5v_k) \\
&- \ln \Gamma(0.5v_k) - 0.5(\hat{t}^i_{k|k-1} + m + 2)\ln \det(R_k) - 0.5\mathrm{tr}(T^i_{k|k-1}R_k^{-1}) + c
\end{aligned}
\tag{33}
$$

where $c$ denotes a constant with respect to $\Theta$.

Substituting (33) into (28) yields

$$
\ln q^i(x_k) = -0.5(x_k - \hat{x}^i_{k|k-1})^{\mathrm{T}}(P^i_{k|k-1})^{-1}(x_k - \hat{x}^i_{k|k-1}) - 0.5(z_k - H^i_k(x_k))^{\mathrm{T}}(\widetilde{R}_k)^{-1}(z_k - H^i_k(x_k)) + c_x,
$$

where $\widetilde{R}_k$ is given by

$$
\widetilde{R}_k = (\mathrm{E}_{q^i(\lambda_k)}\lambda_k)^{-1}(\mathrm{E}_{q^i(R_k)}R_k^{-1})^{-1},
\tag{34}
$$

where $\mathrm{E}_{q^i(\cdot)}$ denotes the expectation with respect to the PDF $q^i(\cdot)$. $\widetilde{R}_k$ can be deemed as modified covariance matrix of measurements taking into account the uncertainty of both and $R_k$ and $\lambda_k$. Then,

$$
q^i(x_k) \propto \mathrm{N}(x_k; \hat{x}^i_{k|k-1}, P^i_{k|k-1})\mathrm{N}(z_k; H^i_k(x_k), \widetilde{R}_k).
$$

According to the construction of nonlinear Gaussian filter in [36], $q^i(x_k)$ can be approximated by a Gaussian PDF, i.e.,

$$
q^i(x_k) \approx \mathrm{N}(x_k; \hat{x}^i_k, P^i_k),
\tag{35}
$$

where

$$
\hat{x}^i_k = \hat{x}^i_{k|k-1} + P^i_{xz,k|k-1}(P^i_{zz,k|k-1})^{-1}(z_k - \hat{z}^i_{k|k-1}),
\tag{36}
$$

$$
\hat{z}^i_{k|k-1} = \int H^i_k(x_k)\mathrm{N}(x_k; \hat{x}^i_{k|k-1}, P^i_{k|k-1})\mathrm{d}x_k,
\tag{37}
$$

$$
P^i_{zz,k|k-1} = \int (H^i_k(x_k) - \hat{z}^i_{k|k-1})(H^i_k(x_k) - \hat{z}^i_{k|k-1})^{\mathrm{T}}\mathrm{N}(x_k; \hat{x}^i_{k|k-1}, P^i_{k|k-1})\mathrm{d}x_k + \widetilde{R}_k,
\tag{38}
$$

$$
P^i_{xz,k|k-1} = \int (x_k - \hat{x}^i_{k|k-1})(H^i_k(x_k) - \hat{z}^i_{k|k-1})^{\mathrm{T}}\mathrm{N}(x_k, \hat{x}^i_{k|k-1}, P^i_{k|k-1})\mathrm{d}x_k,
\tag{39}
$$

$$
P^i_k = P^i_{k|k-1} - P^i_{xz,k|k-1}(P^i_{zz,k|k-1})^{-1}(P^i_{xz,k|k-1})^{\mathrm{T}}
\tag{40}
$$

Substituting (33) into (29) yields

$$
\ln q^i(R_k) = -0.5\mathrm{tr}(\{(\mathrm{E}_{q^i(\lambda_k)}\lambda_k)A^i_k + T^i_{k|k-1}\}R_k^{-1}) - 0.5(\hat{t}^i_{k|k-1} + m + 2)\ln \det(R_k) + c_R,
$$

where

$$
A^i_k = \mathrm{E}_{q^i(x_k)}(z_k - H^i_k(x_k))(z_k - H^i_k(x_k))^{\mathrm{T}}
\tag{41}
$$

According to (3), we can deduce that $q^i(R_k)$ is an inverse Wishart PDF with the following expression:

$$q^i(R_k) = \text{IW}(R_k; \hat{t}_k^i, T_k^i), \tag{42}$$

where

$$\hat{t}_k^i = \hat{t}_{k|k-1}^i + 1, \tag{43}$$

$$T_k^i = (\text{E}_{q^i(\lambda_k)} \lambda_k) A_k^i + T_{k|k-1}^i \tag{44}$$

Substituting (33) into (30) yields

$$\ln q^i(\lambda_k) = -0.5(\text{tr}((\text{E}_{q^i(R_k)} R_k^{-1}) A_k^i) + \text{E}_{q^i(v_k)} v_k) \lambda_k + 0.5(m + \text{E}_{q^i(v_k)} v_k - 2) \ln \lambda_k + c_\lambda.$$

According to (1), $q^i(\lambda_k)$ is a Gamma PDF as

$$q^i(\lambda_k) = \text{G}(\lambda_k; \hat{\alpha}_k^i, \hat{\beta}_k^i), \tag{45}$$

where

$$\hat{\alpha}_k^i = 0.5(m + \text{E}_{q^i(v_k)} v_k), \tag{46}$$

$$\hat{\beta}_k^i = 0.5(\text{tr}((\text{E}_{q^i(R_k)} R_k^{-1}) A_k^i)) + \text{E}_{q^i(v_k)} v_k \tag{47}$$

Substituting (33) into (31), and using the Stirling's approximation $\ln \Gamma(0.5 v_k) \approx 0.5(v_k - 1) \ln 0.5 v_k - 0.5 v_k$, we obtain

$$\ln q^i(v_k) = (\hat{a}_{k|k-1}^i - 0.5) \ln v_k + (0.5 \text{E}_{q^i(\lambda_k)} \ln \lambda_k - 0.5 \text{E}_{q^i(\lambda_k)} \lambda_k - \hat{b}_{k|k-1}^i + 0.5) v_k + c_v$$

According to (1), $q^i(v_k)$ is also a Gamma PDF as

$$q^i(v_k) = \text{G}(v_k; \hat{a}_k^i, \hat{b}_k^i), \tag{48}$$

where

$$\hat{a}_k^i = \hat{a}_{k|k-1}^i + 0.5, \tag{49}$$

$$\hat{b}_k^i = \hat{b}_{k|k-1}^i - 0.5 \text{E}_{q^i(\lambda_k)} \ln \lambda_k + 0.5 \text{E}_{q^i(\lambda_k)} \lambda_k - 0.5 \tag{50}$$

The unknown expectations in (34), (41), (44), (46), (47) and (50) are given by

$$\text{E}_{q^i(\lambda_k)} \lambda_k = \hat{\alpha}_k^i / \hat{\beta}_k^i, \tag{51}$$

$$\text{E}_{q^i(R_k)} R_k^{-1} = (\hat{t}_k^i - m - 1)(T_k^i)^{-1}, \tag{52}$$

$$\begin{aligned} A_k^i &= \int (z_k - H_k^i(x_k))(z_k - H_k^i(x_k))^{\text{T}} \text{N}(x_k; \hat{x}_k^i, P_k^i) dx_k \\ &= (z_k - \hat{z}_k^i)(z_k - \hat{z}_k^i)^{\text{T}} + \int (H_k^i(x_k) - \hat{z}_k^i)(H_k^i(x_k) - \hat{z}_k^i)^{\text{T}} \text{N}(x_k; \hat{x}_k^i, P_k^i) dx_k \end{aligned} \tag{53}$$

$$\hat{z}_k^i = \int H_k^i(x_k) \text{N}(x_k; \hat{x}_k^i, P_k^i) dx_k, \tag{54}$$

$$\text{E}_{q^i(v_k)} v_k = \hat{a}_k^i / \hat{b}_k^i, \tag{55}$$

$$\text{E}_{q^i(\lambda_k)} \ln \lambda_k = \psi(\hat{\alpha}_k^i) - \ln \hat{\beta}_k^i \tag{56}$$

where $\psi(x)$ is the digamma function.

Combining (36)–(40), (43), (44), (46), (47) and (49)–(56), the fixed point iteration algorithm can be employed to acquire the unknown quantities $\hat{x}_k^i$, $P_k^i$, $\hat{t}_k^i$, $T_k^i$, $\hat{\alpha}_k^i$, $\hat{\beta}_k^i$, $\hat{a}_k^i$, and $\hat{b}_k^i$. The iteration is terminated when the variations of these quantities are small enough.

The posterior mode probability can be calculated by

$$
\begin{aligned}
u_k^i &= \frac{p(r_k=i|z_{1:k-1})p(z_k|r_k=i,z_{1:k-1})}{p(z_k|z_{1:k-1})} \\
&= \frac{u_{k|k-1}^i p(z_k|r_k=i,z_{1:k-1})}{\sum_{i=1}^{M} u_{k|k-1}^i p(z_k|r_k=i,z_{1:k-1})}
\end{aligned}
\tag{57}
$$

In (57), the predicted mode probability $u_{k|k-1}^i$ is given by (4), and the predicted likelihood $p(z_k|r_k = i, z_{1:k-1})$ can be derived by

$$
p(z_k|r_k = i, z_{1:k-1}) = \int p(\Theta, z_k|r_k = i, z_{1:k-1})d\Theta
$$

However, the above integration is computationally infeasible since $x_k$, $R_k$, $\lambda_k$ and $v_k$ are coupled in $p(\Theta, z_k|r_k = i, z_{1:k-1})$ according to (32). We use the similar variational approach in [37] to derive an approximated predicted likelihood. The predicted likelihood can be rewritten as

$$
\begin{aligned}
\ln p(z_k|r_k = i, z_{1:k-1}) = \;& \int q^i(x_k)q^i(R_k)q^i(\lambda_k)q^i(v_k) \ln \frac{p(\Theta,z_k|r_k=i,z_{1:k-1})}{q^i(x_k)q^i(R_k)q^i(\lambda_k)q^i(v_k)} dx_k dR_k d\lambda_k dv_k \\
&+ \mathrm{KL}(q^i(x_k)q^i(R_k)q^i(\lambda_k)q^i(v_k)\|p(\Theta|r_k = i, z_{1:k}))
\end{aligned}
$$

Since the KL divergence term in the right side of the above equation is minimized by VB approach, the predicted likelihood can be approximated by

$$
\ln p(z_k|r_k = i, z_{1:k-1}) \approx \int q^i(x_k)q^i(R_k)q^i(\lambda_k)q^i(v_k) \ln \frac{p(\Theta, z_k|r_k = i, z_{1:k-1})}{q^i(x_k)q^i(R_k)q^i(\lambda_k)q^i(v_k)} dx_k dR_k d\lambda_k dv_k
\tag{58}
$$

Substituting (35), (42), (45), (48) and the expression of $p(\Theta, z_k|r_k = i, z_{1:k-1})$ in (32) into (58) yields

$$
\begin{aligned}
\ln p(z_k|r_k = i, z_{1:k-1}) \approx \;& 0.5\{\ln \det(P_k^i) - \ln \det(P_{k|k-1}^i) + n - m \ln \pi + m\psi(\hat{\alpha}_k^i) \\
& - m \ln \hat{\beta}_k^i - (\hat{\alpha}_k^i/\hat{\beta}_k^i)\mathrm{tr}\{(\hat{t}_k^i - m - 1)(T_k^i)^{-1}A_k^i\} + \hat{t}_{k|k-1}^i \ln \det(T_{k|k-1}^i) - \hat{t}_k^i \ln \det(T_k^i) \\
& + 2 \ln \Gamma_m(0.5\hat{t}_k^i) - 2 \ln \Gamma_m(0.5\hat{t}_{k|k-1}^i) - \mathrm{tr}((P_{k|k-1}^i)^{-1}\{P_k^i + (\hat{x}_k^i - \hat{x}_{k|k-1}^i)(\hat{x}_k^i - \hat{x}_{k|k-1}^i)^{\mathrm{T}}\}) \\
& + \mathrm{tr}((\hat{t}_k^i - m - 1)(T_k^i - T_{k|k-1}^i)(T_k^i)^{-1}) + 2\hat{a}_{k|k-1}^i \ln \hat{b}_{k|k-1}^i - 2(\hat{a}_k^i - 1) \ln \hat{b}_k^i + 2 \ln \Gamma(\hat{a}_k^i) \\
& - 2 \ln \Gamma(\hat{a}_{k|k-1}^i) + 2(\hat{b}_k^i - \hat{b}_{k|k-1}^i)\hat{a}_k^i/\hat{b}_k^i - 2\psi(\hat{a}_k^i)\}
\end{aligned}
\tag{59}
$$

The final estimate of the state is given by a probabilistically weighted average of all the mode-conditioned estimates, i.e.,

$$
\hat{x}_k = \sum_{i=1}^{M} u_k^i \hat{x}_k^i,
\tag{60}
$$

and the corresponding covariance matrix is calculated as

$$
P_k = \sum_{i=1}^{M} u_k^i (P_k^i + (\hat{x}_k^i - \hat{x}_k)(\hat{x}_k^i - \hat{x}_k)^{\mathrm{T}}).
\tag{61}
$$

### 3.4. Approximated Gaussian Integrations Based on Unscented Transform

Due to the nonlinearity of the state transition function $F_{k-1}^i(x_{k-1})$ and the measurement function $H_k^i(x_k)$, the Gaussian integrations in (22), (23), (37)–(39), (53) and (54) cannot be computed analytically. In this paper, the UT [38] is employed to calculate the Gaussian integrations approximately.

For the Gaussian integrations in (22) and (23), $2n + 1$ sigma points are generated from the mixing state $\hat{x}_{k-1}^{0i}$ and the covariance matrix $P_{k-1}^{0i}$ as

$$
\begin{cases}
\mathcal{X}_{k-1,0}^{0i} = \hat{x}_{k-1}^{0i} \\
\mathcal{X}_{k-1,p}^{0i} = \hat{x}_{k-1}^{0i} + \left( \sqrt{(n+\lambda)P_{k-1}^{0i}} \right)_p & p = 1, 2, \cdots, n \\
\mathcal{X}_{k-1,p}^{0i} = \hat{x}_{k-1}^{0i} - \left( \sqrt{(n+\lambda)P_{k-1}^{0i}} \right)_p & p = n+1, n+2, \cdots, 2n
\end{cases} ,
$$

where $\left( \sqrt{P} \right)_p$ denotes the $p$-th column of the matrix square root of $P$, $\lambda = \alpha^2(n + \kappa) - n$ is a scaling parameter, $\alpha$ controls the divergence of the sigma points and is usually set to a small positive value (e.g., 0.01), $\kappa$ is a secondary scaling parameter which is usually set to zero. Then, the predicted state $\hat{x}_{k|k-1}^{i}$ and the covariance matrix $P_{k|k-1}^{i}$ are calculated as

$$
\hat{x}_{k|k-1}^{i} = \sum_{p=0}^{2n} w_{m,p} F_{k-1}^{i}(\mathcal{X}_{k-1,p}^{0i}), \tag{62}
$$

$$
P_{k|k-1}^{i} = \sum_{p=0}^{2n} w_{c,p} \left( F_{k-1}^{i}(\mathcal{X}_{k-1,p}^{0i}) - \hat{x}_{k|k-1}^{i} \right) \left( F_{k-1}^{i}(\mathcal{X}_{k-1,p}^{0i}) - \hat{x}_{k|k-1}^{i} \right)^{\mathrm{T}} + Q_{k-1}^{i}, \tag{63}
$$

where the weights of sigma points are given by

$$
\begin{cases}
w_{m,0} = \lambda / (n + \lambda) \\
w_{c,0} = \lambda / (n + \lambda) + (1 - \alpha^2 + \beta) \\
w_{m,p} = w_{c,p} = 0.5 / (n + \lambda) & p = 1, 2, \cdots, 2n
\end{cases} ,
$$

where $\beta$ is used to incorporate the prior information of the distribution. $\beta = 2$ is optimal under Gaussian distributions.

For the Gaussian integrations in (37)–(39), the sigma points generated from the predicted state $\hat{x}_{k|k-1}^{i}$ and the covariance matrix $P_{k|k-1}^{i}$ are given by

$$
\begin{cases}
\mathcal{X}_{k|k-1,0}^{i} = \hat{x}_{k|k-1}^{i} \\
\mathcal{X}_{k|k-1,p}^{i} = \hat{x}_{k|k-1}^{i} + \left( \sqrt{(n+\lambda)P_{k|k-1}^{i}} \right)_p & p = 1, 2, \cdots, n \\
\mathcal{X}_{k|k-1,p}^{i} = \hat{x}_{k|k-1}^{i} - \left( \sqrt{(n+\lambda)P_{k|k-1}^{i}} \right)_p & p = n+1, n+2, \cdots, 2n
\end{cases}
$$

Then, the predicted measurement mean $\hat{z}_{k|k-1}^{i}$, the covariance matrix $P_{zz,k|k-1}^{i}$ and the cross-covariance matrix $P_{xz,k|k-1}^{i}$ are calculated as

$$
\hat{z}_{k|k-1}^{i} = \sum_{p=0}^{2n} w_{m,p} H_k^i(\mathcal{X}_{k|k-1,p}^{i}), \tag{64}
$$

$$
P_{zz,k|k-1}^{i} = \sum_{p=0}^{2n} w_{c,p} \left( H_k^i(\mathcal{X}_{k|k-1,p}^{i}) - \hat{z}_{k|k-1}^{i} \right) \left( H_k^i(\mathcal{X}_{k|k-1,p}^{i}) - \hat{z}_{k|k-1}^{i} \right)^{\mathrm{T}} + \widetilde{R}_k, \tag{65}
$$

$$
P_{xz,k|k-1}^{i} = \sum_{p=0}^{2n} w_{c,p} \left( \mathcal{X}_{k|k-1,p}^{i} - \hat{x}_{k|k-1}^{i} \right) \left( H_k^i(\mathcal{X}_{k|k-1,p}^{i}) - \hat{z}_{k|k-1}^{i} \right)^{\mathrm{T}} \tag{66}
$$

We generate the sigma points from the posterior state $\hat{x}_k^i$ and the covariance matrix $P_k^i$ as

$$
\begin{cases}
\mathcal{X}_{k,0}^i = \hat{x}_k^i \\
\mathcal{X}_{k,p}^i = \hat{x}_k^i + \left( \sqrt{(n+\lambda)P_k^i} \right)_p & p = 1, 2, \cdots, n \\
\mathcal{X}_{k,p}^i = \hat{x}_k^i - \left( \sqrt{(n+\lambda)P_k^i} \right)_p & p = n+1, n+2, \cdots, 2n
\end{cases}
.
$$

Then, the Gaussian integrations in (53) and (54) are calculated as

$$
\hat{z}_k^i = \sum_{p=0}^{2n} w_{m,p} H_k^i (\mathcal{X}_{k,p}^i), \tag{67}
$$

$$
A_k^i = (z_k - \hat{z}_k^i)(z_k - \hat{z}_k^i)^{\mathrm{T}} + \sum_{p=0}^{2n} w_{c,p} (H_k^i(\mathcal{X}_{k,p}^i) - \hat{z}_k^i)(H_k^i(\mathcal{X}_{k,p}^i) - \hat{z}_k^i)^{\mathrm{T}} \tag{68}
$$

With the above derivations, the proposed robust IMM filter is summarized as follows:

**Step 1:** Choose initial estimates $\hat{x}_0^i$, $P_0^i$, $u_0^i$, $\hat{a}_0^i$, $\hat{b}_0^i$, $\hat{t}_0^i$ and $T_0^i$ for each mode $i$ Set $\pi_{ij}$ and $\rho$. Let $k = 1$

**Step 2:** Calculate the predicted mode probability $u_{k|k-1}^j$ and the mixing mode probability $u_{k-1}^{ij}$ using (4) and (5). Then, calculate the mixing quantities $\hat{x}_{k-1}^{0i}$, $P_{k-1}^{0i}$, $\hat{t}_{k-1}^{0i}$, $T_{k-1}^{0i}$, $\hat{a}_{k-1}^{0i}$ and $\hat{b}_{k-1}^{0i}$ for each mode $i$ using (8)–(11), (16) and (17).

**Step 3:** Calculate the predicted quantities $\hat{x}_{k|k-1}^i$, $P_{k|k-1}^i$, $\hat{t}_{k|k-1}^i$, $T_{k|k-1}^i$, $\hat{a}_{k|k-1}^i$ and $\hat{b}_{k|k-1}^i$ for each mode $i$ using (62), (63) and (24)–(27).

**Step 4:** Update the posterior quantities $\hat{x}_k^i$, $P_k^i$, $\hat{t}_k^i$, $T_k^i$, $\hat{a}_k^i$ and $\hat{b}_k^i$ for each mode $i$ through fixed point iterations based on VB approach as follows:

   Update $\hat{t}_k^i$ and $\hat{a}_k^i$ using (43) and (49), and initialize $\hat{x}_k^i = \hat{x}_{k|k-1}^i$, $P_k^i = P_{k|k-1}^i$, $T_k^i = T_{k|k-1}^i$, $\hat{b}_k^i = \hat{b}_{k|k-1}^i$.
   Repeat
       Update $A_k^i$ using (67) and (68).
       Update $\hat{a}_k^i$ and $\hat{\beta}_k^i$ using (46), (47), (52) and (55).
       Update $T_k^i$ and $\hat{b}_k^i$ using (44), (50), (51) and (56).
       Update $\hat{x}_k^i$ and $P_k^i$ using (34), (36), (40), (51), (52) and (64)–(66).
   Until converged

**Step 5:** Update the posterior mode probability $u_k^i$ using (57) and (59).

**Step 6:** Calculate $\hat{x}_k$ and $P_k$ using (60) and (61). Let $k + 1 \to k$, and return to Step2.

## 4. Simulation Example

This section presents a two-dimensional maneuvering target tracking scenario with a period of 200 s to demonstrate the performance of the proposed filter. A maneuvering target moves following two models: constant velocity (CV) model and coordinated turn (CT) model. The modes representing the CV model and CT model at time $k$ are denoted by $r_k = 1$ and $r_k = 2$, respectively. The state of the CV model is $x_k = (x_k, y_k, \dot{x}_k, \dot{y}_k)^{\mathrm{T}}$ including the components of the position $(x_k, y_k)$ and the velocity $(\dot{x}_k, \dot{y}_k)$. The dynamics of the CV model is given by

$$
x_k = \begin{pmatrix} \mathbf{I}_2 & \Delta t \mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{I}_2 \end{pmatrix} x_{k-1} + w_{k-1}^1,
$$

where $\Delta t$ is the sampling period. The covariance matrix of the process noises $w_{k-1}^1$ is given by

$$
Q_{k-1}^1 = q_x^1 \begin{pmatrix} (\Delta t)^3 \mathbf{I}_2 / 3 & (\Delta t)^2 \mathbf{I}_2 / 2 \\ (\Delta t)^2 \mathbf{I}_2 / 2 & \Delta t \mathbf{I}_2 \end{pmatrix},
$$

where $q_x^1$ is the power spectral density. The state of the CT model is $\boldsymbol{x}_k = (x_k, y_k, \dot{x}_k, \dot{y}_k, \omega_k)^{\mathrm{T}}$, where $\omega_k$ is the turn rate. The dynamics of the CT model is given by

$$
\boldsymbol{x}_k = \begin{pmatrix}
1 & 0 & \sin(\omega_{k-1}\Delta t)/\omega_{k-1} & (\cos(\omega_{k-1}\Delta t) - 1)/\omega_{k-1} & 0 \\
0 & 1 & (1 - \cos(\omega_{k-1}\Delta t))/\omega_{k-1} & \sin(\omega_{k-1}\Delta t)/\omega_{k-1} & 0 \\
0 & 0 & \cos(\omega_{k-1}\Delta t) & -\sin(\omega_{k-1}\Delta t) & 0 \\
0 & 0 & \sin(\omega_{k-1}\Delta t) & \cos(\omega_{k-1}\Delta t) & 0 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix} \boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1}^2,
$$

and the covariance matrix of the process noises $\boldsymbol{w}_{k-1}^2$ is given by

$$
\boldsymbol{Q}_{k-1}^2 = \begin{pmatrix}
q_x^2 (\Delta t)^3 \mathbf{I}_2/3 & q_x^2 (\Delta t)^2 \mathbf{I}_2/2 & \mathbf{0} \\
q_x^2 (\Delta t)^2 \mathbf{I}_2/2 & q_x^2 \Delta t \mathbf{I}_2 & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & q_\omega
\end{pmatrix},
$$

where $q_x^2$ and $q_\omega$ are the power spectral densities corresponding to $(x_k, y_k, \dot{x}_k, \dot{y}_k)$ and $\omega_k$, respectively.

A sensor located at $(x_s, y_s)$ collects noisy range and azimuth measurements of the target according to the equation

$$
\boldsymbol{z}_k = \begin{pmatrix}
\sqrt{(x_k - x_s)^2 + (y_k - y_s)^2} \\
\arctan((y_k - y_s)/(x_k - x_s))
\end{pmatrix} + \boldsymbol{\varepsilon}_k.
$$

The sampling period is $\Delta t = 1$s. The target executes a $-4°/$s coordinated turn from 0 s to 70 s, moves at a constant velocity from 71 s to 120 s, and executes a $4°/$s coordinated turn from 121 s to 200 s. The power spectral densities are set to $q_x^1 = q_x^2 = 0.01$ m$^2$s$^{-3}$ and $q_\omega = 2 \times 10^{-6}$ rad$^2$s$^{-3}$. The true initial state of target is $\boldsymbol{x}_0 = (0\ \text{m},\ 0\ \text{m},\ 10\ \text{m/s},\ 10\ \text{m/s},\ -4°/\text{s})^{\mathrm{T}}$, and the location of the sensor is $(x_s, y_s) = (-400\ \text{m},\ -400\ \text{m})$. The true track of the target and the sensor location are shown in Figure 1.
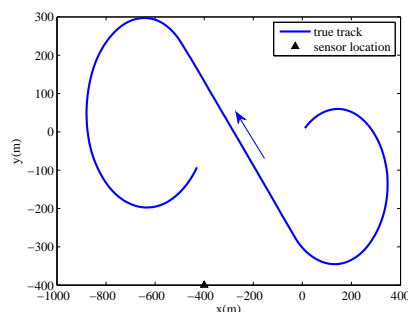


**Figure 1.** True track of target and sensor location.

To illustrate the performance of the proposed filter, we consider the following four cases under different conditions of measurement noises:

Case A: Gaussian noises

$$
\boldsymbol{\varepsilon}_k \sim \mathrm{N}(\boldsymbol{\varepsilon}_k; \mathbf{0}, \overline{\boldsymbol{R}}),
$$

where $\overline{\boldsymbol{R}} = \mathrm{diag}((10\,\text{m})^2, (0.2°)^2)$ is the known nominal covariance matrix.

Case B: Gaussian noises with time varying covariance matrix

$$
\boldsymbol{\varepsilon}_k \sim \mathrm{N}(\boldsymbol{\varepsilon}_k; \mathbf{0}, (0.1 + 0.05\cos(2\pi k/100))\overline{\boldsymbol{R}}).
$$

This case is used to simulate the situation that the actual covariance matrix is deviated from the nominal one $\overline{\boldsymbol{R}}$.

Case C: Contaminated Gaussian noises

$$\varepsilon_k \sim (1 - \delta)\mathrm{N}(\varepsilon_k; \mathbf{0}, \overline{\mathbf{R}}) + \delta\mathrm{N}(\varepsilon_k; \mathbf{0}, 100\overline{\mathbf{R}}),$$

where $\mathrm{N}(\varepsilon_k; \mathbf{0}, \overline{\mathbf{R}})$ can be seen as a normal distribution, $\mathrm{N}(\varepsilon_k; \mathbf{0}, 100\overline{\mathbf{R}})$ can be considered as a perturbing distribution with much larger covariance matrix due to outliers, and $\delta \in [0, 1]$ is a perturbing parameter that represents the extent of the perturbation. This case is used to simulate the corrupted measurements by outliers. We set $\delta = 0.1$.

Case D: Contaminated Gaussian noises with time varying covariance matrix

$$\varepsilon_k \sim (1 - \delta)\mathrm{N}(\varepsilon_k; \mathbf{0}, (0.1 + 0.05\cos(2\pi k/100))\overline{\mathbf{R}}) + \delta\mathrm{N}(\varepsilon_k; \mathbf{0}, 100\overline{\mathbf{R}}).$$

Both outliers and covariance mistuning are simultaneously simulated in this case. $\delta$ is also set to 0.1.

Four existing filters including VB based Student's *t* filter (VBStdF) [30] utilizing only single CT model, IMM filter (IMMF) [1], VB based IMM filter (IMMVBF) estimating the unknown covariance matrix of measurement noises [33], and VB based IMM filter modeling measurement noises as Student's *t*-distribution (IMMVBStdF) [31] are compared with the proposed filter. 1000 Monte Carlo (MC) runs are carried out for each case of measurement noises above. All the filters are implemented in MATLAB on an Intel i7 3.6GHz processor.

The initial estimates of the degree and the inverse scale matrix with respect to the inverse Wishart distribution in VBStdF, IMMVBF and the proposed filter are chosen as $\hat{t}_0 = 7$ and $\mathbf{T}_0 = (\hat{t}_0 - m - 1)\overline{\mathbf{R}}$. All the initial estimates of the Gamma distribution parameters in VBStdF, IMMVBStdF and the proposed filter are set to 0.5. For all the VB approach based filters, $\rho = 1 - \exp(-4)$ is used, and the VB iteration is terminated when the difference of position estimates between two adjacent iterations is less than 1e-6. The mode probabilities are initialized as $u_0^1 = 0.1$ and $u_0^2 = 0.9$, and the mode transition probability is set to $\pi_{ij} = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}$ for all the IMM type filters. The initial covariance matrix of the state estimate is set to $\mathbf{P}_0 = \mathrm{diag}(100\,\mathrm{m}^2,\ 100\,\mathrm{m}^2,\ 1\,(\mathrm{m/s})^2,\ 1\,(\mathrm{m/s})^2,\ 1 \times 10^{-5}\,(\mathrm{rad/s})^2)$, and the initial estimate of the state $\hat{x}_0$ is chosen randomly from the Gaussian distribution $\mathrm{N}(\cdot; x_0, \mathbf{P}_0)$ for all the filters.

The root mean square errors (RMSEs) and average RMSEs (ARMSEs) of position, velocity and turn rate are used to evaluate the estimation accuracy of the filters. The RMSE and ARMSE of position are defined as

$$\mathrm{RMSE}_{\mathrm{pos}}(k) = \sqrt{\frac{1}{P}\sum_{i=1}^{P}\left((\hat{x}_k^{(i)} - x_k^{(i)})^2 + (\hat{y}_k^{(i)} - y_k^{(i)})^2\right)},$$

$$\mathrm{ARMSE}_{\mathrm{pos}} = \frac{1}{T}\sum_{k=1}^{T}\mathrm{RMSE}_{\mathrm{pos}}(k).$$

where $P$ is the total count of MC runs, $(x_k^{(i)}, y_k^{(i)})$ and $(\hat{x}_k^{(i)}, \hat{y}_k^{(i)})$ are true position and estimated position for the *i*-th MC run, and $T$ is the total count of samples. The RMSEs and ARMSEs of velocity and turn rate are defined similar.

The RMSEs of position, velocity and turn rate in four cases are shown in Figures 2–5, and the corresponding ARMSEs are listed in Table 1. The overall performance of VBStdF in estimation accuracy is the worst among the filters for all cases. We do not plot the entire RMSE curves of position for VBStdF to show other curves distinctly. The largest RMSE of position for VBStdF under four cases are 53.03 m, 50.19 m, 99.82 m and 98.30 m, respectively. The reason for the poor performance of VBStdF is that the single-model based filter is less adaptive to target maneuvering compared with multiple-model based filter.
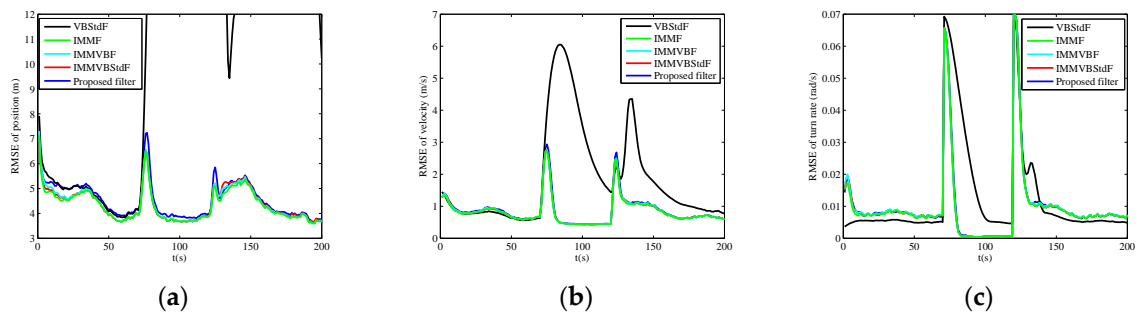
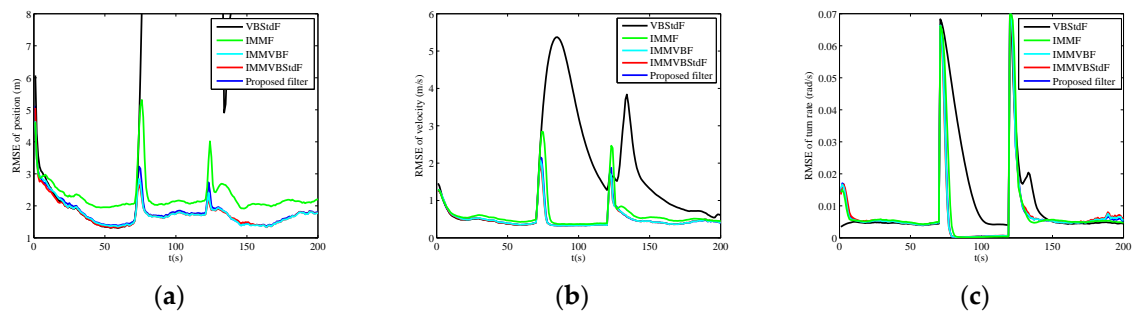**Figure 2.** Root mean square errors (RMSEs) versus time in Case A. (**a**) Position; (**b**) Velocity; (**c**) Turn rate.



**Figure 3.** RMSEs versus time in Case B. (**a**) Position; (**b**) Velocity; (**c**) Turn rate.
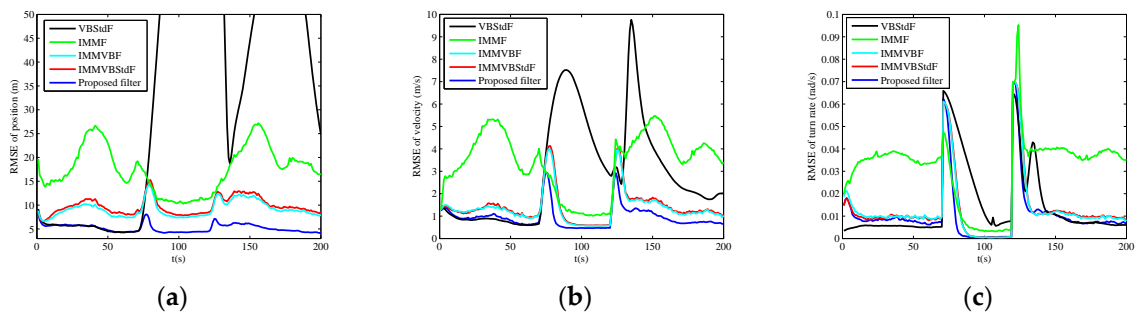


**Figure 4.** RMSEs versus time in Case C. (**a**) Position; (**b**) Velocity; (**c**) Turn rate.
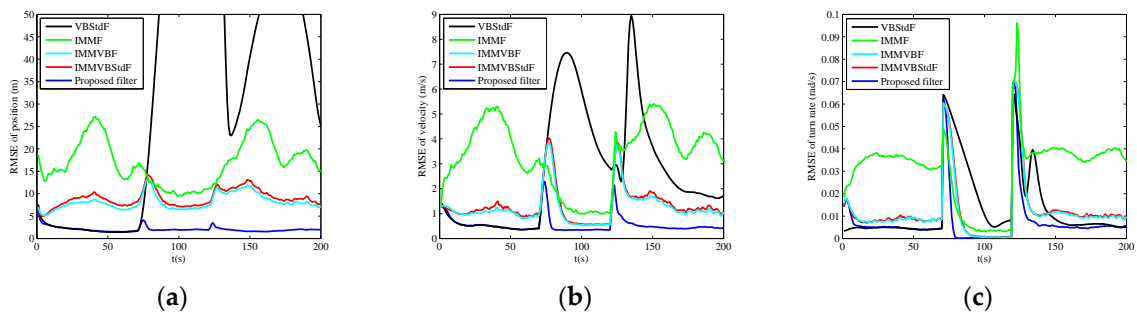


**Figure 5.** RMSEs versus time in Case D. (**a**) Position; (**b**) Velocity; (**c**) Turn rate.

**Table 1.** Average root mean square errors of filters in four cases.

|  | Case | VBStdF | IMMF | IMMVBF | IMMVBStdF | Proposed Filter |
|---|---|---|---|---|---|---|
| **Position (m)** | A | 19.9747 | 4.3292 | 4.3755 | 4.3812 | 4.5475 |
|  | B | 17.1240 | 2.3064 | 1.7662 | 1.7485 | 1.8055 |
|  | C | 39.8238 | 17.5279 | 9.1497 | 9.7692 | 5.1772 |
|  | D | 37.3586 | 16.9337 | 8.1700 | 8.9868 | 2.0456 |
| **Velocity (m/s)** | A | 1.8595 | 0.8287 | 0.8410 | 0.8361 | 0.8627 |
|  | B | 1.5341 | 0.6204 | 0.5097 | 0.5018 | 0.5136 |
|  | C | 2.9977 | 3.3905 | 1.3687 | 1.4143 | 0.9405 |
|  | D | 2.7279 | 3.3061 | 1.2562 | 1.3375 | 0.5493 |
| **Turn rate (rad/s)** | A | 0.0125 | 0.0098 | 0.0100 | 0.0099 | 0.0101 |
|  | B | 0.0111 | 0.0077 | 0.0071 | 0.0072 | 0.0071 |
|  | C | 0.0149 | 0.0316 | 0.0130 | 0.0130 | 0.0106 |
|  | D | 0.0138 | 0.0309 | 0.0122 | 0.0125 | 0.0073 |

In Case A, the estimation accuracy of the proposed filter is lower than IMMF and IMMVBF. The reason is that the assumed Student's *t*-distribution for measurement noises deviates from the actual Gaussian distribution. The performance of the proposed filter in estimation accuracy is also poorer than IMMVBStdF. That is because the scale matrix of Student's *t*-distribution is required to be estimated additionally for the proposed filter, while the scale matrix is known exactly for IMMVBStdF. However, the accuracy degradation of the proposed filter is not obvious compared with IMMF, IMMVBF, and IMMVBStdF.

In Case B, IMMF has larger RMSEs than IMMVBF, IMMVBStdF, and the proposed filter since IMMF utilizes mistuned covariance matrix of measurement noises, while the IMMVBF and the proposed filters can learn the covariance matrix adaptively. The IMMVBStdF seems to have the ability of capturing the unknown covariance matrix either. The performance of the proposed filter in estimation accuracy is also slightly worse than IMMVBF and IMMVBStdF as Case A.

In Case C, the proposed filter outperforms other filters obviously in estimation accuracy. IMMF and IMMVBF are based on the assumption of Gaussian noises so that they cannot cope with heavy-tailed measurement noises. Figure 6 shows the estimates of the dof parameters $\hat{\alpha}_k^1$, $\hat{\beta}_k^1$, $\hat{\alpha}_k^2$ and $\hat{\beta}_k^2$ versus time for IMMVBStdF and the proposed filter in one MC run. We can see that the estimates of dof parameters for IMMVBStdF become large after several time steps. Therefore, the distribution of measurement noises converges to a Gaussian distribution and loses the heavy-tailed property. The estimates of dof parameters for the proposed filter maintain small values overall, except some high peaks for $\hat{\beta}_k^1$ and $\hat{\beta}_k^2$ curves caused by outliers. Therefore, the proposed filter is more robust against heavy-tailed measurement noises than IMMVBStdF. In addition, we can see that the IMMVBF and IMMVBStdF are more robust than IMMF.

The results under Case D are similar to Case C.

The 1000 MC runs averaged estimates of mode probabilities for CV model are shown in Figure 7. In Case A, we can see that the estimation results are nearly the same for all the filters. In Case B, the accuracy of IMMF estimates is lower than other filters and the response to mode changes for IMMF is more lagging than other filters. The reason is also that the covariance matrix of the measurement noises utilized by IMMF is mistuned. In Case C and D, the proposed filter achieves more accuracy estimates of mode probabilities than other filters, indicating that the proposed filter can also improve the estimation accuracy of mode probability under heavy-tailed measurement noises.
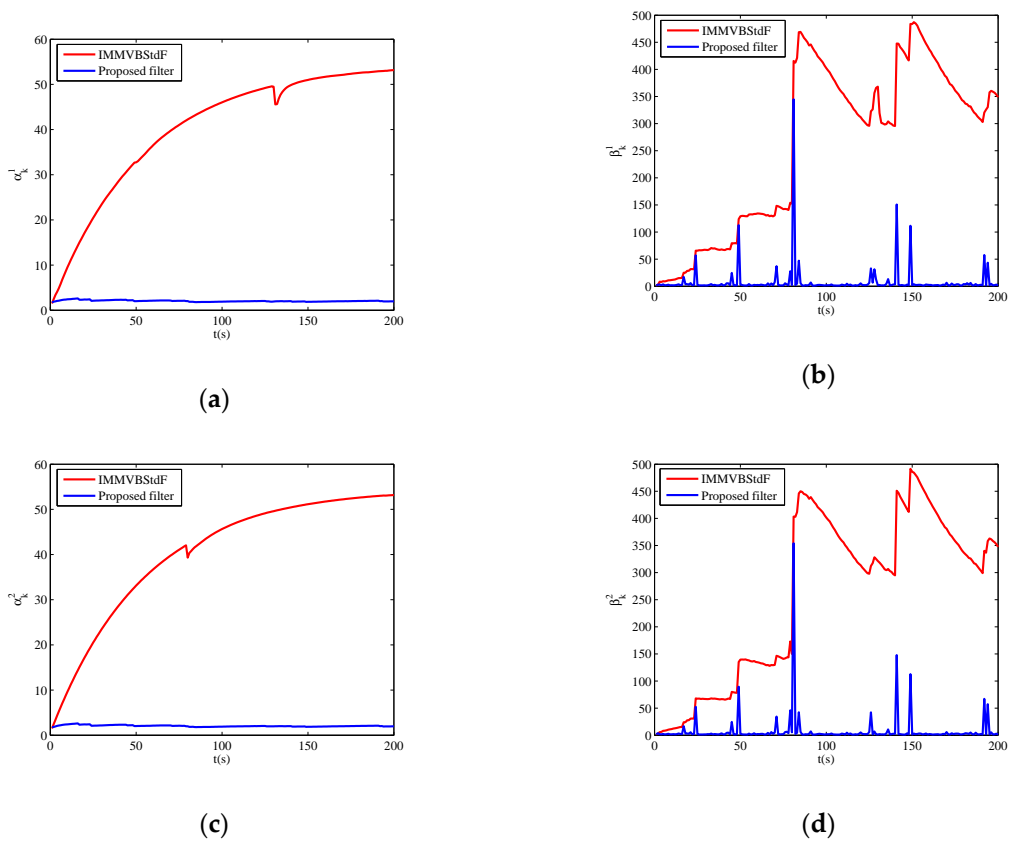
(**a**)



(**b**)



(**c**)



(**d**)

**Figure 6.** Estimates of degrees of freedom parameters versus time. (**a**) $\hat{\alpha}_k^1$; (**b**) $\hat{\beta}_k^1$; (**c**) $\hat{\alpha}_k^2$; (**d**) $\hat{\beta}_k^2$.



(**a**)
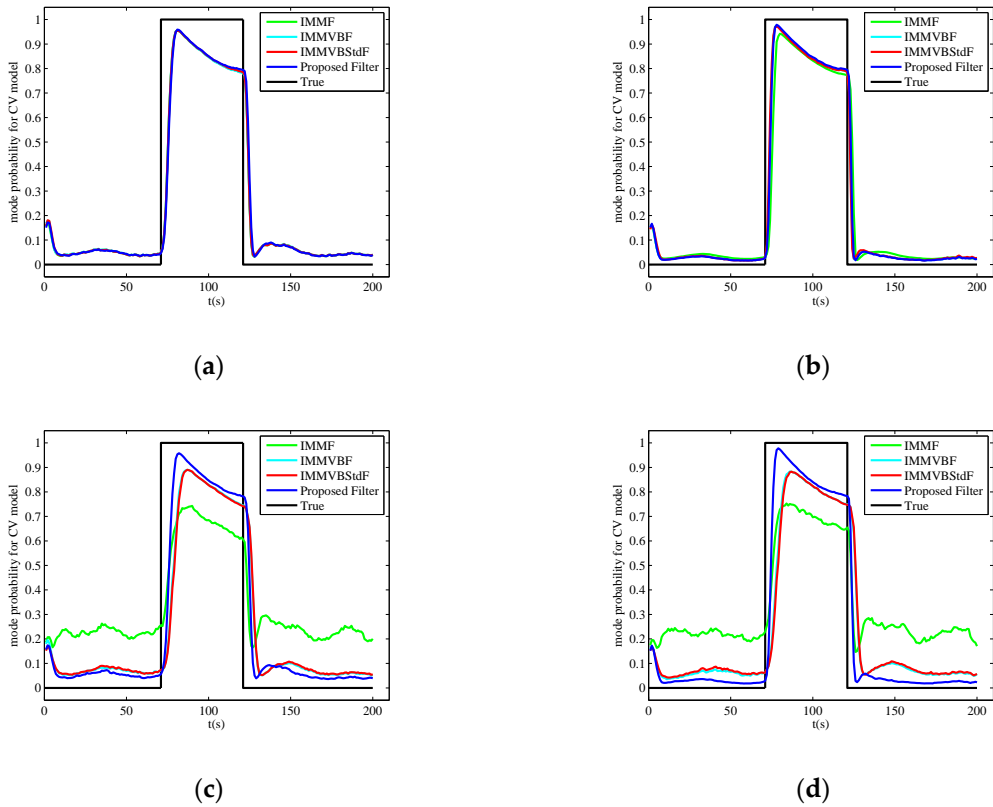


(**b**)



(**c**)



(**d**)

**Figure 7.** 1000 Monte Carlo runs averaged estimates of mode probabilities for constant velocity model in four cases. (**a**) Case A; (**b**) Case B; (**c**) Case C; (**d**) Case D.

We use the number of floating points operations (flops) [39] to measure the computational complexity of filter. The computational complexity of VBStdF, IMMF, IMMVBF, IMMVBStdF and the proposed filter are listed in Table 2. We can see that the proposed filter has the highest computational complexity among the filters. According to Table 2, the computational complexity of proposed filter is higher than VBStdF mainly because of multiple model operations, and is higher than IMMF mainly because of fixed point iterations. The proposed filter requires more flops compared with IMMVBF and IMMVBStdF since the proposed filter involves more parameters to be estimated in the filter recursions.

**Table 2.** Computational complexity of filters, where $m$ denotes the dimension of measurements, $n$ denotes the dimension of target state, $M$ denotes the number of models for interacting multiple model type filters, and $N$ denotes the number of fixed point iterations for variational Bayesian approach based filters.

| Filter | Number of Flops |
|---|---|
| VBStdF | $(4m^3 + 2n^3/3 + 14m^2n + 20mn^2 + 10m^2 + 8n^2 + 15mn + 4m + n + 15)N + 31n^3/3 + m^2 + 12n^2 + 5n$ |
| IMMF | $(4n^2 + 3n + 5)M^2 + (4m^3/3 + 32n^3/3 + 8m^2n + 12mn^2 + 5m^2 + 19n^2 + 9mn + 4m + 4n + 6)M + 3n$ |
| IMMVBF | $(m^3 + 2n^3/3 + 14m^2n + 20mn^2 + 9m^2 + 8n^2 + 15mn + 3m + n + 3)MN + (2m^2 + 4n^2 + 3n + 7)M^2 + (8m^3/3 + 14n^3 + 15n^2 + 3m + 7n + 33)M + 3n$ |
| IMMVBStdF | $(3m^3 + 2n^3/3 + 14m^2n + 20mn^2 + 6m^2 + 8n^2 + 15mn + 4m + n + 1)MN + (4n^2 + 3n + 13)M^2 + (14n^3 + 15n^2 + 7n + 39)M + 3n$ |
| Proposed filter | $(4m^3 + 2n^3/3 + 14m^2n + 20mn^2 + 10m^2 + 8n^2 + 15mn + 4m + n + 15)MN + (2m^2 + 4n^2 + 3n + 15)M^2 + (8m^3/3 + 14n^3 + 15n^2 + 3m + 7n + 68)M + 3n$ |

We further test the computational cost of filters for the above tracking scenario. The 1000 MC runs averaged computation time in four cases are shown in Table 3. It can be seen that the proposed filter expends more computational cost than other filters, demonstrating that the proposed filter has higher computational complexity than other filters.

**Table 3.** Computation time of filters in four cases.

| Case | VBStdF | IMMF | IMMVBF | IMMVBStdF | Proposed Filter |
|---|---|---|---|---|---|
| A | 0.3900 s | 0.1092 s | 0.5304 s | 0.4992 s | 0.8736 s |
| B | 0.3588 s | 0.1092 s | 0.4836 s | 0.4836 s | 0.8424 s |
| C | 0.5148 s | 0.1092 s | 0.4992 s | 0.4524 s | 1.1076 s |
| D | 0.4836 s | 0.1092 s | 0.4524 s | 0.4368 s | 0.9516 s |

## 5. Conclusions

We propose a robust IMM filter against heavy-tailed measurement noises for maneuvering target tracking. The heavy-tailed measurement noises are treated as Student's *t*-distribution, and the unknown dof and scale matrix are assumed to be governed by Gamma and inverse Wishart distributions, respectively. Then, the filter recursions for the target state and the parameters of Gamma and inverse Wishart distributions are designed in the IMM framework. In the model interaction step, the mixing distributions of the state, dof, and scale matrix are achieved by using the methods of matching the first two moments and minimizing weighted KL divergence. In the measurement update step, the state and the unknown parameters are jointly estimated by employing the VB approach. The problem of system nonlinearity is solved by utilizing UT to compute the Gaussian integrations approximately. Simulation results show that the proposed filter outperforms other related filters in terms of the estimation accuracy for the state and the mode probability under heavy-tailed measurement noises.

The main contributions of this paper are summarized as follows. Firstly, a robust IMM filter is proposed to properly handle both model uncertainty and outlier measurements. Secondly, by

treating the unknown dof of Student's *t*-distribution of heavy-tailed measurement noises as a Gamma distributed random variable, the heavy-tailed property of Student's *t*-distribution is maintained and the robustness is improved over the existing filter. Finally, the problem of system nonlinearity is addressed by using UT technology.

## References

1. Li, X.R.; Jilkov, V.P. Survey of maneuvering target tracking. Part V. Multiple-model methods. *IEEE Trans. Aerosp. Electr. Syst.* **2005**, *41*, 1255–1321. [CrossRef]
2. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*; John Wiley & Sons: New York, NY, USA, 2001; pp. 441–466.
3. Achutegui, K.; Martino, L.; Rodas, J.; Escudero, C.J.; Míguez, J. A multi-model particle filtering algorithm for indoor tracking of mobile terminals using RSS data. In Proceedings of the 2009 IEEE Control Applications, (CCA) & Intelligent Control, (ISIC), Saint Petersburg, Russia, 8–10 July 2009; pp. 1702–1707. [CrossRef]
4. Liang, H.; Kang, F. Tracking UUV based on interacting multiple model unscented particle filter with multi-sensor information fusion. *Optik* **2015**, *126*, 5067–5073. [CrossRef]
5. Goff, G.M.; Black, J.T.; Beck, J.A. Tracking maneuvering spacecraft with filter-through approaches using interacting multiple models. *Acta Astronaut.* **2015**, *114*, 152–163. [CrossRef]
6. Lu, K.; Zhou, R. Sensor fusion of Gaussian mixtures for ballistic target tracking in the re-entry phase. *Sensors* **2016**, *16*, 1289. [CrossRef] [PubMed]
7. Menegaz, H.M.T.; Battistini, S. Switching multiple model filter for boost-phase missile tracking. *IEEE Trans. Aerosp. Electr. Syst.* **2018**, *54*, 2547–2553. [CrossRef]
8. Zhang, Z.; Gao, C.; Jing, W. Infrared measurement and composite tracking algorithm for air-breathing hypersonic vehicles. *J. Electron. Imaging* **2018**, *27*, 023023. [CrossRef]
9. Li, X.R.; Zhang, Y. Multiple-model estimation with variable structure. Part V: Likely-model set algorithm. *IEEE Trans. Aerosp. Electr. Syst.* **2000**, *36*, 448–466. [CrossRef]
10. Li, X.R.; Jilkov, V.P.; Ru, J. Multiple-model estimation with variable structure. Part VI: Expected-mode augmentation. *IEEE Trans. Aerosp. Electr. Syst.* **2005**, *41*, 853–867. [CrossRef]
11. Lan, J.; Li, X.R. Equivalent-model augmentation for variable-structure multiple-model estimation. *IEEE Trans. Aerosp. Electr. Syst.* **2013**, *49*, 2615–2630. [CrossRef]
12. Xu, L.; Li, X.R.; Duan, Z. Hybrid grid multiple-model estimation with application to maneuvering target tracking. *IEEE Trans. Aerosp. Electr. Syst.* **2016**, *52*, 122–136. [CrossRef]
13. Urteaga, I.; Bugallo, M.F.; Djurić, P.M. Sequential Monte Carlo methods under model uncertainty. In Proceedings of the 2016 IEEE Statistical Signal Processing Workshop (SSP), Palma de Mallorca, Spain, 26–29 June 2016. [CrossRef]
14. Martino, L.; Read, J.; Elvira, V.; Louzada, F. Cooperative parallel particle filters for online model selection and applications to urban mobility. *Digit. Signal Process.* **2017**, *60*, 172–185. [CrossRef]
15. Koch, K.R.; Yang, Y. Robust Kalman filter for rank deficient observation models. *J. Geodesy* **1998**, *72*, 436–441. [CrossRef]
16. Petrus, P. Robust Huber adaptive filter. *IEEE Trans. Signal Process.* **1999**, *47*, 1129–1133. [CrossRef]
17. Chang, L.; Hu, B.; Chang, G.; Li, A. Huber-based novel robust unscented Kalman filter. *IET Sci. Meas. Technol.* **2012**, *6*, 502–509. [CrossRef]
18. Li, K.; Chang, L.; Hu, B. A variational Bayesian-based unscented Kalman filter with both adaptivity and robustness. *IEEE Sens. J.* **2016**, *16*, 6966–6976. [CrossRef]
19. Huang, J.; Yan, B.; Hu, S. Centralized fusion of unscented Kalman filter based on Huber robust method for nonlinear moving target tracking. *Math. Prob. Eng.* **2015**, *2015*, 291913. [CrossRef]

20. Wang, Y.; Zheng, W. Robust information filter based on maximum correntropy criterion. *J. Guid. Control Dyn.* **2016**, *39*, 1124–1129. [CrossRef]

21. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy Kalman filter. *Automatica* **2017**, *76*, 70–77. [CrossRef]

22. Hou, B.; He, Z.; Zhou, X.; Zhou, H.; Li, D.; Wang, J. Maximum correntropy criterion Kalman filter for α-Jerk tracking model with non-Gaussian noise. *Entropy* **2017**, *19*, 648. [CrossRef]

23. Hou, B.; He, Z.; Li, D.; Zhou, H.; Wang, J. Maximum correntropy unscented Kalman filter for ballistic missile navigation system based on SINS/CNS deeply integrated mode. *Sensors* **2018**, *18*, 1724. [CrossRef]

24. Wang, G.; Gao, Z.; Zhang, Y.; Ma, B. Adaptive maximum correntropy Gaussian filter based on variational Bayes. *Sensors* **2018**, *18*, 1960. [CrossRef] [PubMed]

25. Roth, M.; Özkan, E.; Gustafsson, F. A student's t filter for heavy tailed process and measurement noise. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 5770–5774. [CrossRef]

26. Huang, Y.; Zhang, Y.; Li, N.; Chambers, J. Robust student's t based nonlinear filter and smoother. *IEEE Trans. Aerosp. Electr. Syst.* **2016**, *52*, 2586–2596. [CrossRef]

27. Zhu, H.; Leung, H.; He, Z. A variational Bayesian approach to robust sensor fusion based on Student's t-distribution. *Inf. Sci.* **2013**, *221*, 201–214. [CrossRef]

28. Xu, D.; Shen, C.; Shen, F. A robust particle filtering algorithm with non-Gaussian measurement noise using Student's t-distribution. *IEEE Signal Process. Lett.* **2014**, *21*, 30–34. [CrossRef]

29. Huang, Y.; Zhang, Y.; Li, N.; Wu, Z.; Chambers, J.A. A novel robust student's t-based Kalman filter. *IEEE Trans. Aerosp. Electr. Syst.* **2017**, *53*, 1545–1554. [CrossRef]

30. Huang, Y.; Zhang, Y.; Xu, B.; Wu, Z.; Chambers, J. A new outlier-robust student's t based Gaussian approximate filter for cooperative localization. *IEEE/ASME Trans. Mechatron.* **2017**, *22*, 2380–2386. [CrossRef]

31. Shen, C.; Xu, D.; Huang, W.; Shen, F. An interacting multiple model approach for state estimation with non-Gaussian noise using a variational Bayesian method. *Asian J. Control* **2015**, *17*, 1424–1434. [CrossRef]

32. Peel, D.; Mclachlan, G.J. Robust mixture modelling using the t distribution. *Stat. Comput.* **2000**, *10*, 339–348. [CrossRef]

33. Li, W.; Jia, Y. Kullback-Leibler divergence for interacting multiple model estimation with random matrices. *IET Signal Process.* **2016**, *10*, 12–18. [CrossRef]

34. Särkkä, S.; Nummenmaa, A. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Trans. Autom. Control* **2009**, *54*, 596–600. [CrossRef]

35. Šmídl, V.; Quinn, A. *The Variational Bayes Method in Signal Processing*; Springer: New York, NY, USA, 2006.

36. Ito, K.; Xiong, K. Gaussian filters for nonlinear filtering problems. *IEEE Trans. Autom. Control* **2000**, *45*, 910–927. [CrossRef]

37. Li, W.; Jia, Y. Adaptive filtering for jump Markov systems with unknown noise covariance. *IET Control Theory Appl.* **2013**, *7*, 1765–1772. [CrossRef]

38. Wan, E.A.; Merwe, R.V.D. The unscented Kalman filter for nonlinear estimation. In Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, Lake Louise, AB, Canada, 4 October 2000; pp. 153–158. [CrossRef]

39. Zhang, Z.; Hao, Y.; Wu, X. Complexity analysis of three deterministic sampling nonlinear filtering algorithms. *J. Harbin Inst. Technol.* **2013**, *45*, 111–115. [CrossRef]