Gaussian Mixture Midway-Merge for Object SLAM With Pose Ambiguity

Jae Hyung Jung, *Graduate Student Member, IEEE*, and Chan Gook Park, *Member, IEEE*

***Abstract*—In this letter, we propose a novel method to merge a Gaussian mixture on matrix Lie groups and present its application for a simultaneous localization and mapping problem with symmetric objects. The key idea is to predetermine the weighted mean called a *midway point* and merge Gaussian mixture components at the associated tangent space. Through this rule, the covariance matrix captures the original density more accurately, and the need for the back-projection is spared when compared to the conventional merge. We highlight the midway-merge by numerically evaluating dissimilarity metrics of density functions before and after the merge on the rotational group. Furthermore, we experimentally discover that the rotational error of symmetric objects follows heavy-tailed behavior and formulate the Gaussian sum filter to model it by a Gaussian mixture noise. The effectiveness of our approach is validated through virtual and real-world datasets.**

***Index Terms*—Gaussian mixture, matrix Lie group, object detection, SLAM.**

I. Introduction

A GAUSSIAN mixture (GM) is prevalent in engineering problems such as state estimation and target tracking due to its ability to model multiple hypotheses. A radar altimeter exhibits a GM noise characteristic in vegetated areas by multiple returns from the ground and vegetation canopy [1], [2]. The posterior intensity is represented by a GM in the probability hypothesis density filter [3] for the efficient update recursion. Especially in robotics research, GM is a versatile tool to perceive the surrounding real-world environment. In simultaneous localization and mapping (SLAM) problems, multiple loop closure and data association hypotheses can be encoded in GM distributions [4], [5]. It is common to model the map as a mixture of local Gaussian distributions for better registration [6], [7]. Multiple pose hypotheses of symmetric objects are modeled in a GM distribution [8], [9].

Manuscript received 7 August 2022; accepted 16 November 2022. Date of publication 24 November 2022; date of current version 9 December 2022. This letter was recommended for publication by Associate Editor T. A. Vidal-Calleja and Editor J. Civera upon evaluation of the reviewers' comments. This work was supported by the National Research Foundation of Korea funded by the Ministry of Science and ICT, the Republic of Korea, under Grant NRF-2022R1A2C2012166. *(Corresponding author: Chan Gook Park.)*

Jae Hyung Jung is with the Department of Aerospace Engineering, Automation and Systems Research Institute, Seoul National University, Seoul 08826, Republic of Korea (e-mail: lastflowers@snu.ac.kr).

Chan Gook Park is with the Navigation and Electronic System Laboratory, Department of Aerospace Engineering, Institute of Advanced Aerospace Technology, Seoul National University, Seoul 08826, Republic of Korea (e-mail: chanpark@snu.ac.kr).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2022.3224665, provided by the authors.

Digital Object Identifier 10.1109/LRA.2022.3224665

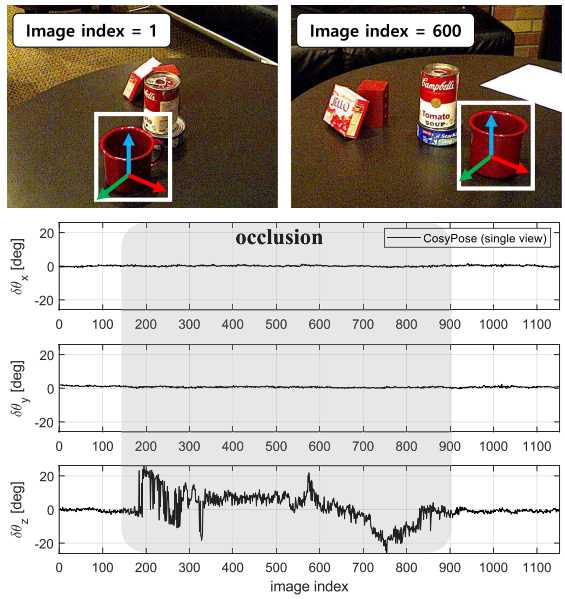

Fig. 1. Rotational error estimated by a 6 DOF pose detector [10] of a mug in the YCB-Video dataset [11]. The symmetric z-axis exhibits heavy-tailed noise distribution due to self-occlusion.

The Bayes rule plays a central role in estimation problems, and the Bayesian filter is solved analytically in a linear Gaussian system. However, once a GM is introduced in either process or measurement distributions, the Bayesian recursion exponentially increases the number of hypotheses [12], [13]. To maintain a tractable size of a state dimension, Gaussian components should be merged or pruned with the minimum information loss. On the other hand, in the aspect of modeling a state-space representation, a matrix Lie group is a natural tool to deal with the underlying geometry, for instance, describing the three-dimensional position and attitude of a rigid body [14], [15]. Filtering on matrix Lie groups [16], [17] has shown promising consistency and accuracy over the conventional parameterization in a vector space.

However, dealing with a GM on matrix Lie groups is not straightforward. A question arises in how to define GMs and a reduction procedure on matrix Lie groups. To this end, we propose to merge a GM at a common tangent space that is called a *midway point*, a weighted matrix of mean matrices. Through this merge rule, we prove that the approximation error is less than the conventional merge by Ćesić et al. [18] when deriving the merged covariance matrix. Furthermore, our approach eliminates the need for the back-projection leading to a lighter computational burden over the conventional merge.

We present a promising application example in object SLAM with pose ambiguity. Since a pioneering work of object-based SLAM [19], these systems have shown promising results over conventional low-level SLAM by capturing semantically as well as geometrically meaningful information. However, estimating the 6 degrees of freedom (DOF) pose of a symmetric object with occlusion is still very challenging. For instance, we experimentally discover that the rotational error along the symmetric axis of a mug behaves as a heavy-tailed distribution, as shown in Figs. 1 and 5. A standard Gaussian distribution cannot capture this behavior and properly weight these outliers. To tackle this, we fit the noise distribution by a GM and formulate a Kalman filter with the proposed GM merge method using a 6 DOF pose detector as a sensor. The main contribution of this letter is as follows.

- We propose a novel GM merge on matrix Lie groups that is called *Gaussian mixture midway-merge*, where probability density functions (PDF) are transformed at the tangent space of the predetermined midway point, then merged.
- The strength of our merge method is validated through the numerical evaluation of PDF dissimilarity metrics on the special orthogonal group $SO(3)$.
- We formulate a Kalman filter with the proposed merge for object SLAM with pose ambiguity and demonstrate its effectiveness in a Monte-Carlo simulation, photo-realistic simulator, and real-world dataset. We open our implementation[1] for the benefit of the community.

II. Related Works

A. Gaussian Mixture Merge

A GM distribution approximates any arbitrary PDFs but leads to exponentially increasing hypotheses in Bayesian recursions. To restrain the inflation, GM reduction includes two main steps: component selection and merging. The goal of the selection is to pick the most similar densities among components. Williams and Maybeck [20] proposed the integral square difference (ISD) as a dissimilarity measure. Runnalls [21] derived an upper bound for the Kullback-Leibler divergence (KLD) between PDFs before and after the moment-preserving merge. Assa and Plataniotis [22] adopted the Wasserstein distance to preserve the geometric shape of GMs. However, previous works were limited to PDFs in a vector space. We focus on a GM merge on matrix Lie groups.

The most relevant work to ours is [18]. Built upon the Gaussian PDF on matrix Lie groups [23], [24], they transform one of the densities in a pair of Gaussian components to the other tangent space, then merge the components as done in a vector space. After this step, the density should be back-projected to matrix Lie groups for a zero-mean at the new associated tangent space. However, their approach assumed that the distance between mean matrices in a GM should be small to make a valid PDF after the merge, and the back-projection involves two matrix multiplications per a merge. In contrast, we directly merge components at the midway point.

[1][Online]. Available: https://github.com/lastflowers/midway

B. Object SLAM

In the context of object SLAM, object poses, as well as robot poses, are jointly estimated. Primitive shapes are employed to model objects such as spheres [25], ellipsoids [26], cuboids [27], and cylinders [28]. Others use an accurate CAD model to formulate likelihoods [19]. In the aspect of estimator consistency, [29], [30] proved that unobservable bases do not depend on a linearization point in a framework of the invariant extended Kalman filter (IEKF). However, previous approaches did not explicitly consider ambiguous object measurements.

To deal with multiple hypotheses due to the ambiguity, PoseRBPF [31] represents the marginalized pose distribution by the augmented autoencoder [32], but their method only includes a single object for a detector that ignores the correlation between objects. Fu et al. [8] utilized a max-mixture [4] for multimodal pose estimates in the back-end implementation, but the approximation is vulnerable to bad initialization. To overcome the sensitivity in initialization, Lu et al. [9] proposed a heuristic technique to re-initialize a hypothesis, but their noise assumption for a mug did not reflect the real-world noise characteristic. Merrill et al. [33] introduced the prior keypoint heatmap as an input to a deep neural network, but they did not explicitly consider the noise behavior of symmetric objects. We focus on representing the actual noise distribution of a symmetric object and formulate a Kalman filter to account explicitly for its uncertainty.

III. Preliminaries

A. Matrix Lie Group

A matrix Lie group G is a *group* as well as a *smooth manifold* where its elements are matrices. A Lie algebra identified on the identity matrix consists of a vector space $\mathfrak{g}$ with the Lie bracket. The hat operator $(\cdot)^\wedge$ transforms an element in $\mathfrak{g}$ to a vector element and the inverse mapping is designated as $(\cdot)^\vee$. Elements in both structures are related through the matrix exponential and logarithm map,[2]

$$\exp(x^\wedge) = X, \quad \ln(X)^\vee = x \tag{1}$$

where $x \in \mathbb{R}^N$, $x^\wedge \in \mathfrak{g}$, and $X \in G$. Closed-form expressions for certain matrix Lie groups can be found in [14].

We use an approximated *Baker-Campbell-Hausdorff* (BCH) formula throughout this letter,

$$\ln\left(\exp(x_1^\wedge)\exp(x_2^\wedge)\right)^\vee \approx x_1 + J_l(-x_1)^{-1}x_2 \tag{2}$$

where the higher-order term $O(\|x_2\|^2)$ is assumed to be 0, $\|x\|^2 = x^T x$ is a 2-norm, $J_l(x) = \sum_{n=0}^{\infty} \mathrm{ad}(x)^n/(n+1)!$, and $\mathrm{ad}(x)$ is an adjoint of a Lie algebra. Based on (2), a useful equation is obtained

$$\ln\left(\exp(-x_1^\wedge)\exp((x_1+x_2)^\wedge)\right)^\vee \approx J_l(-x_1)x_2 \tag{3}$$

if x_2 is small [18].

B. Gaussian Distribution on Matrix Lie Groups

A Gaussian distribution on matrix Lie groups is defined through a vector element at the identity [23],

$$X = \exp(-\xi^\wedge)\hat{X}. \tag{4}$$

[2]The matrix logarithm is one-to-many, but it is uniquely defined, for instance, if the magnitude of the rotational part in SO(3) is $\|\phi\| < \pi$.

We follow the right-invariant error convention [16], $\hat{X} \in G$ is a mean matrix, and $\xi \sim N(0, P)$, a Gaussian distribution with a zero-mean and covariance P. Assuming that ξ is concentrated at the identity and by changing the coordinate $dX = |J_l(\xi)| d\xi$,

$$\begin{aligned} 1 &= \int_{\mathbb{R}^N} (2\pi)^{-\frac{N}{2}} |P|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\xi^T P^{-1}\xi\right) d\xi \\ &= \int_G \eta \exp\left(-\frac{1}{2}\ln(\hat{X}X^{-1})^{\vee^T} P^{-1} \ln(\hat{X}X^{-1})^{\vee}\right) dX \end{aligned} \quad (5)$$

where $\eta = (2\pi)^{-\frac{N}{2}} |J_l(\xi) P J_l(\xi)^T|^{-\frac{1}{2}}$. We denote the last line of (5) as $X \sim N_G(\hat{X}, P)$. As introduced in [18], a GM on matrix Lie groups is expressed as

$$\sum_i w_i N_G(\hat{X}_i, P_i) \quad (6)$$

where a weight of the ith component w_i satisfies $\sum_i w_i = 1$.

C. Moment-Preserving Merge in a Vector Space

It is straightforward to merge two Gaussian distributions in a vector space by preserving the first and second-moment [21]. Suppose that we are given a GM with two components,

$$w_1^* N(\hat{x}_1, P_1) + w_2^* N(\hat{x}_2, P_2) \quad (7)$$

where w_1^* and w_2^* are normalized weights. Then, their moment-preserving merged distribution is $N(\hat{x}_m, P_m)$ where

$$\begin{aligned} \hat{x}_m &= w_1^* \hat{x}_1 + w_2^* \hat{x}_2 \\ P_m &= w_1^* P_1 + w_2^* P_2 + w_1^* w_2^* (\hat{x}_1 - \hat{x}_2)(\hat{x}_1 - \hat{x}_2)^T. \end{aligned} \quad (8)$$

IV. Gaussian Mixture Merge

We review the uncertainty transformation in [18] and reveal a limitation that would yield information loss with increasing distance between means of GM components. Then, we introduce a *Gaussian mixture midway-merge* to mitigate this.

A. Uncertainty at the Transformed Mean

To merge a GM on matrix Lie groups, it is straightforward to merge them at the same mean. Assume that we wish to express $N_G(\hat{X}_i, P_i)$ at some common point $\hat{X}_c$. Using notations in Section III, we can write the ith random vector as $\xi_i = \ln(\hat{X}_i X^{-1})^{\vee}$. Substituting $X = \exp(-\xi_c^{\wedge})\hat{X}_c$,

$$\begin{aligned} \xi_i &= \ln\left(\hat{X}_i \hat{X}_c^{-1} \exp(\xi_c^{\wedge})\right)^{\vee} = \ln\left(\exp(\Delta x_i^{\wedge}) \exp(\xi_c^{\wedge})\right)^{\vee} \\ &\overset{(3)}{\approx} J_l(\Delta x_i)(\xi_c + \Delta x_i) \end{aligned} \quad (9)$$

if Δx_i is small where $\exp(\Delta x_i^{\wedge}) := \hat{X}_i \hat{X}_c^{-1}$. Then, substituting (9) to the first line of (5), we obtain

$$\begin{aligned} 1 \approx \int_{\mathbb{R}^N} &(2\pi)^{-\frac{N}{2}} \left|\bar{P}_i\right|^{-\frac{1}{2}} \\ &\times \exp\left(-\frac{1}{2}(\xi_c + \Delta x_i)^T \bar{P}_i^{-1} (\xi_c + \Delta x_i)\right) d\xi_c \end{aligned} \quad (10)$$

where we used $d\xi_i = |J_l(\Delta x_i)| d\xi_c$ and the new covariance is $\bar{P}_i = J_l^{-1}(\Delta x_i) P_i J_l^{-T}(\Delta x_i)$. Therefore, the random vector at the new mean follows, $\xi_c \sim N(-\Delta x_i, \bar{P}_i)$ [18].

It is clear that as the distance between mean matrices $\|\Delta x_i\|$ increases, (10) is no longer a valid PDF, hence $\bar{P}_i$ cannot capture the correct uncertainty. Also, being a valid density in the line of (5) requires back-projection to make a zero-mean. To mitigate the assumption and remove the additional computation, we merge GM components at a *midway point* in the following section.

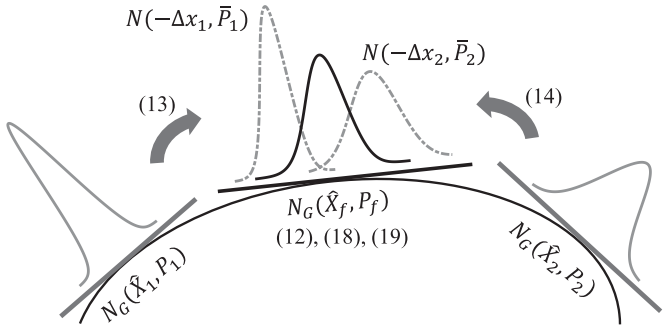

Fig. 2. Schematic illustration of the proposed merge with the corresponding equations at each step.

B. The Midway-Merge

Given two components in a GM on matrix Lie groups,

$$w_1 N_G(\hat{X}_1, P_1) + w_2 N_G(\hat{X}_2, P_2) \quad (11)$$

the key idea is to merge them at the fused mean,

$$\hat{X}_f = \left(\hat{X}_2 \hat{X}_1^{-1}\right)^{w_2^*} \hat{X}_1 \quad (12)$$

where $w_2^* = w_2/(w_1 + w_2)$ is a normalized weight. Eq. (12) is analogous to linear interpolation in a Lie algebra [14]. We call $\hat{X}_f$ as a *midway point* as it is placed between $\hat{X}_1$ and $\hat{X}_2$ as schematically seen from Fig. 2. If we express $N_G(\hat{X}_1, P_1)$ at the midway point using (10),

$$\begin{aligned} \xi_f &\sim N\left(-\Delta x_1, \bar{P}_1\right) \\ \Delta x_1 &= \ln\left(\hat{X}_1 \hat{X}_f^{-1}\right)^{\vee} \\ \bar{P}_1 &= J_l^{-1}(\Delta x_1) P_1 J_l^{-T}(\Delta x_1) \end{aligned} \quad (13)$$

where we assume that $O(\|\Delta x_1\|^2) = 0$. Likewise, we transform the second component to the midway point,

$$\begin{aligned} \xi_f &\sim N\left(-\Delta x_2, \bar{P}_2\right) \\ \Delta x_2 &= \ln\left(\hat{X}_2 \hat{X}_f^{-1}\right)^{\vee} \\ \bar{P}_2 &= J_l^{-1}(\Delta x_2) P_2 J_l^{-T}(\Delta x_2). \end{aligned} \quad (14)$$

Again, we assume that $O(\|\Delta x_2\|^2) = 0$.

We merge the transformed distributions in (13) and (14) at the tangent space on $\hat{X}_f$ by preserving moments as introduced in Section III-C,

$$\begin{aligned} \Delta x_f &= -w_1^* \Delta x_1 - w_2^* \Delta x_2 \\ &= -w_1^* \ln\left(\hat{X}_1 \hat{X}_f^{-1}\right)^{\vee} - w_2^* \ln\left(\hat{X}_2 \hat{X}_f^{-1}\right)^{\vee}. \end{aligned} \quad (15)$$

We define the distance between $\hat{X}_1$ and $\hat{X}_2$ as

$$\exp(\Delta x^{\wedge}) := \hat{X}_2 \hat{X}_1^{-1}. \quad (16)$$

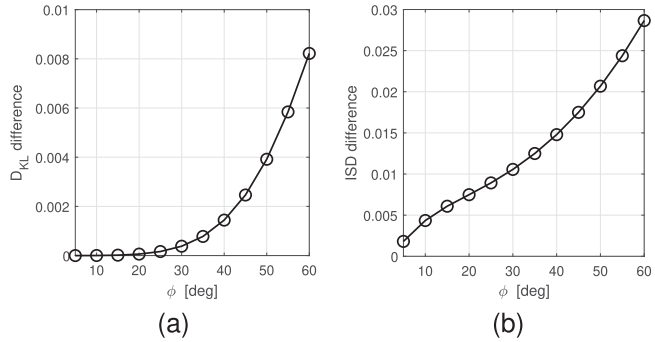

Fig. 3. The differences of the (a) Kullback-Leibler divergence and (b) integral square distance between the proposed and previous method [18].

Substituting (12) and (16) to (15) yields

$$\begin{aligned}\Delta x_f = & -w_1^* \ln\left(\exp\left(-w_2^* \Delta x^\wedge\right)\right)^\vee \\ & - w_2^* \ln\left(\exp\left(\Delta x^\wedge\right) \exp\left(-w_2^* \Delta x^\wedge\right)\right)^\vee \\ = & \, w_1^* w_2^* \Delta x - w_2^*(1-w_2^*)\Delta x \\ = & \, 0. \end{aligned} \tag{17}$$

This implies that the merged distribution at the tangent space of $\hat{X}_f$ is a valid distribution as a line of (5). Using (8), the merged covariance is

$$\begin{aligned} P_f &= w_1^* \bar{P}_1 + w_2^* \bar{P}_2 + w_1^* w_2^* (\Delta x_2 - \Delta x_1)(\Delta x_2 - \Delta x_1)^T \\ &= w_1^* \bar{P}_1 + w_2^* \bar{P}_2 + w_1^* w_2^* \Delta x \Delta x^T, \end{aligned} \tag{18}$$

and the weight is

$$w_f = w_1 + w_2. \tag{19}$$

Fig. 2 illustrates the overall procedure of the *Gaussian mixture midway-merge* in which each component is transformed to the same mean and merged in a vector space. It is remarkable to note that our approach does not require any back-projection to matrix Lie groups by virtue of (17) that was needed in [18]. This spares computing the adjoint when merging the covariance.

C. Approximation Error Analysis

We have assumed that both $\|\Delta x_1\|^2$ and $\|\Delta x_2\|^2$ are zero when deriving the merge method. This approximation is actually less significant than the assumption (in our notation, $\|\Delta x\|^2 = 0$) in the previous method [18].

Theorem: Given $\|\Delta x_1\|$, $\|\Delta x_2\|$, and $\|\Delta x\|$ in (13), (14), and (16), respectively, then, $\|\Delta x_1\|^2 + \|\Delta x_2\|^2 \leq \|\Delta x\|^2$

Proof: From the definition we can see that the distance between $\hat{X}_1$ and $\hat{X}_f$ is

$$\Delta x_1 = \ln\left(\exp(-w_2^* \Delta x^\wedge)\right)^\vee = -w_2^* \, \Delta x. \tag{20}$$

Likewise, the distance between $\hat{X}_2$ and $\hat{X}_f$ is

$$\Delta x_2 = \ln\left(\exp((1-w_2^*)\Delta x^\wedge)\right)^\vee = (1-w_2^*) \, \Delta x. \tag{21}$$

Therefore,

$$\begin{aligned} \|\Delta x_1\|^2 + \|\Delta x_2\|^2 &= (w_1^{*2} + w_2^{*2})\|\Delta x\|^2 \\ &\leq \|\Delta x\|^2 \end{aligned} \tag{22}$$

since $0 \leq w_1^* \leq 1$, $0 \leq w_2^* \leq 1$ and $w_1^* + w_2^* = 1$. ■

It is remarkable to note that the approximation error in the Gaussian mixture midway-merge is always less than the conventional method [18] except when $w_1^* = 1$ or $w_2^* = 1$. In these extreme cases, a merge is not required. Based on (22), our approach can estimate the merged covariance matrix more accurately when merging Gaussian distributions on matrix Lie groups.

V. Gaussian Merge on SO(3)

The objective of this test is to investigate dissimilarity between densities before $p(X)$ and after the merge $q(X)$ with increasing distance of mean matrices. Suppose that we are given densities on $SO(3)$ such that

$$\begin{aligned} p(X) = & \underbrace{0.5}_{w_1^*} N_G \left(\underbrace{\exp([0,0,0]^{T^\wedge})}_{\hat{X}_1}, \underbrace{5^2 I_3}_{P_1} \right) \\ & + \underbrace{0.5}_{w_2^*} N_G \left(\underbrace{\exp([\phi,-\phi,\phi]^{T^\wedge})}_{\hat{X}_2}, \underbrace{10^2 I_3}_{P_2} \right) \end{aligned} \tag{23}$$

$$q(X) = N_G\left(\hat{X}_f, P_f\right) \tag{24}$$

where the angle, ϕ is increased from 5 to 60 deg with the interval of 5 deg. The KLD and ISD are defined as follows

$$\begin{aligned} D_{KL}(p \,||\, q_\star) &= \int_{SO(3)} p(X) \ln\left(\frac{p(X)}{q_\star(X)}\right) dX \\ ISD(p,\, q_\star) &= \int_{SO(3)} \left(p(X) - q_\star(X)\right)^2 dX \end{aligned} \tag{25}$$

where $p(X)$ is defined in (23), and $q_\star(X)$ either could be the proposed merge $q_1(X)$ or the previous method $q_2(X)$ [18].

Since a direct numerical integration on $SO(3)$ is intractable, we numerically integrate the KLD and ISD on $\mathfrak{so}(3)$ with a 0.3 deg interval using the BCH formula up to the 5th order terms. By definition of KLD,

$$\begin{aligned} & D_{KL}(p \,||\, q) = \\ & \int_{\mathfrak{so}(3)} \left(\eta_1 \exp\left(-\frac{1}{2}\xi^T P_1^{-1} \xi\right) + \eta_2 \exp\left(-\frac{1}{2}\xi_2^T P_2^{-1} \xi_2\right) \right) \\ & \times \ln\left(\frac{\eta_1 \exp\left(-\frac{1}{2}\xi^T P_1^{-1}\xi\right) + \eta_2 \exp\left(-\frac{1}{2}\xi_2^T P_2^{-1}\xi_2\right)}{\eta_f \exp\left(-\frac{1}{2}\xi_f^T P_f^{-1}\xi_f\right)} \right) d\xi \end{aligned} \tag{26}$$

where ξ, ξ_2, and ξ_f are random vectors on $\hat{X}_1$, $\hat{X}_2$, and $\hat{X}_f$, respectively. The normalizers are

$$\begin{aligned} \eta_1 &= w_1^* (2\pi)^{-\frac{3}{2}} |P_1|^{-\frac{1}{2}}, \\ \eta_2 &= w_2^* (2\pi)^{-\frac{3}{2}} |P_2|^{-\frac{1}{2}} \frac{|J_l(\xi)|}{|J_l(\xi_2)|}, \\ \eta_f &= (2\pi)^{-\frac{3}{2}} |P_f|^{-\frac{1}{2}} \frac{|J_l(\xi)|}{|J_l(\xi_f)|}. \end{aligned} \tag{27}$$

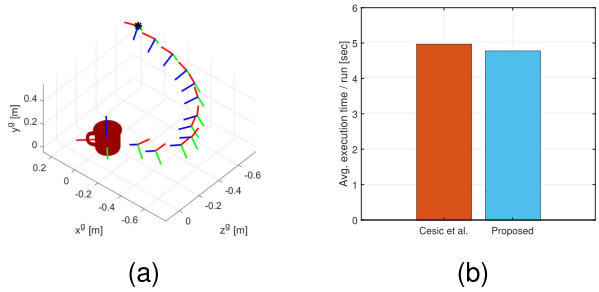

Fig. 4. (a) Virtual trajectory with a mug where the asterisk marks the first camera pose. (b) Average execution time of a single run in 100 Monte-Carlo simulations.

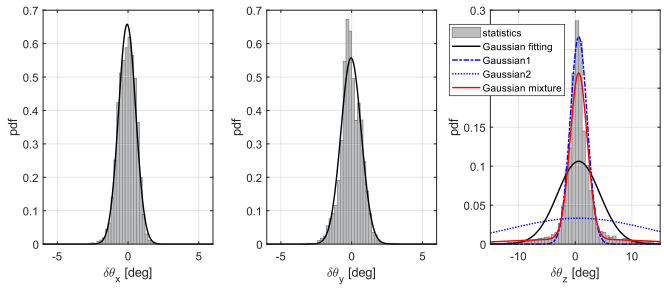

Fig. 5. Rotational error histogram of the self-occluded mug estimated by CosyPose in the YCB-Video dataset. The Gaussian mixture captures the heavy-tailed density, while the Gaussian fitting fails to account for it.

Algorithm 1: GM-IEKF with Midway-Merge.

Input: $\{w^{(h)},\ X^{(h)},\ P^{(h)}\}_{h=1}^{N_h^-},\ \{Y_j\}_{j=1}^{M},\ u_m$
Output: $\{w^{(h)},\ X^{(h)},\ P^{(h)}\}_{h=1}^{N_h}$
// prediction for every hypothesis
for $h = 1$ to N_h^- **do**
  $(X^{(h)},\ P^{(h)}) \leftarrow \texttt{Prediction}\,(X^{(h)},\ P^{(h)},\ u_m)$
end for
// update for every hypothesis
for $h = 1$ to N_h^- **do**
  for $i = 1$ to 2 **do**
    $w^{(h+)} \leftarrow \texttt{weightUpdate}\,(w^{(h)}, X^{(h)}, P^{(h)}, R_i)$
    $(X^{(h+)},\ P^{(h+)}) \leftarrow$ $\texttt{Update}\,(X^{(h)},\ P^{(h)},\ \{Y_j\}_{j=1}^{M},\ R_i)$
  end for
end for
$N_h^+ = 2N_h^-$
while $N_h^+ \geq N_h$ **do**
  $\texttt{midwayMerge}\,(w^{(h)},\ X^{(h)},\ P^{(h)})$
end while

Using the BCH formula ξ_2 is approximated as follows

$$\begin{aligned}\xi_2 &= \ln(\hat{X}_2 X^{-1})^\vee = \ln(\hat{X}_2 \hat{X}_1^{-1} \hat{X}_1 X^{-1})^\vee \\ &= \ln\left(\exp(\Delta x^\wedge)\exp(\xi^\wedge)\right)^\vee \\ &\approx J_l(\xi)^{-1}\Delta x + \xi + \frac{1}{12}\Delta x^\wedge \Delta x^\wedge \xi - \frac{1}{24}\xi^\wedge \Delta x^\wedge \Delta x^\wedge \xi \\ &\quad + \frac{1}{120}\Delta x^\wedge \xi^\wedge \Delta x^\wedge \xi^\wedge \Delta x + \frac{1}{120}\xi^\wedge \Delta x^\wedge \xi^\wedge \Delta x^\wedge \xi. \quad (28)\end{aligned}$$

Likewise, ξ_f is approximated as analogous to ξ_2. We solve (26) by the Euler method with $\Delta\xi_i = 0.3\,\text{deg}$ interval,

$$\sum_{i=1}^{1200^3} p(\xi_i) \ln\left(\frac{p(\xi_i)}{q(\xi_i)}\right) \Delta\xi_i. \quad (29)$$

Fig. 3 shows the differences of each dissimilarity measure: $D_{KL}(p\,||\,q_2) - D_{KL}(p\,||\,q_1)$ and $ISD(p,\,q_2) - ISD(p,\,q_1)$. Since the differences are larger than zero at all ϕ, our method more accurately captures the original density $p(X)$ than the conventional method. As expected from our error analysis in Section IV-C, the differences become larger when ϕ increases. This is originated from the violation of the assumption that $\|\Delta x\|^2 \approx 0$.

TABLE I
ROTATIONAL ROOT MEAN SQUARE ERROR AND AVERAGED NORMALIZED ESTIMATION ERROR SQUARED IN 100 MONTE-CARLO RUNS

Noise level	Method	Rotation robot [deg]	Rotation object [deg]	ANEES
$\sqrt{10^0\,Q}$	GM ($\mathcal{T}_L$)	0.919	0.885	1.023
	GM (*midway*)	0.919	0.885	1.023
$\sqrt{10^1\,Q}$	GM ($\mathcal{T}_L$)	1.540	1.246	1.026
	GM (*midway*)	1.540	1.246	1.026
$\sqrt{10^2\,Q}$	GM ($\mathcal{T}_L$)	2.677	2.021	1.036
	GM (*midway*)	**2.676**	**2.020**	1.036
$\sqrt{10^3\,Q}$	GM ($\mathcal{T}_L$)	4.564	3.312	1.083
	GM (*midway*)	**4.561**	**3.309**	**1.082**
$\sqrt{10^4\,Q}$	GM ($\mathcal{T}_L$)	7.413	5.096	1.449
	GM (*midway*)	**7.400**	**5.079**	**1.443**
$\sqrt{10^5\,Q}$	GM ($\mathcal{T}_L$)	10.646	6.502	1.890
	GM (*midway*)	**10.582**	**6.396**	**1.863**

VI. OBJECT SLAM WITH SYMMETRY

We consider a SLAM problem with a symmetric object as a promising example of the midway-merge. Specifically, we are interested in jointly estimating robot and object poses parameterized by

$$X = (T_R, T_1, \ldots, T_M) \quad (30)$$

where T_R represents a robot pose and $\{T_j\}_{j=1}^{M}$ are object poses in $SE(3)$. The robot pose is driven by the odometry model,

$$\frac{d}{dt}T_R(t) = T_R(t)\,(u(t) + n_w(t))^\wedge \quad (31)$$

where $u(t) \in \mathbb{R}^6$ is the true body velocity, and $n_w(t) \in \mathbb{R}^6$ is a zero-mean white Gaussian noise with $E[n_w(t)\,n_w(\tau)^T] = Q\,\delta(t-\tau)$. We assume that objects are static $d/dt\,T_j = 0$. An onboard sensor (RGB sensor with a deep neural network) measures a relative pose of the jth object at a discrete time,

$$Y_j(t_k) = T_R(t_k)^{-1}\,T_j(t_k)\exp(n_j(t_k)^\wedge). \quad (32)$$

The measurement noise is a white GM that is uncorrelated to n_w such that

$$n_j(t_k) \sim \alpha_1\,N(0, R_1) + \alpha_2\,N(0, R_2). \quad (33)$$

We investigate around 10 k images of a mug in the YCB-Video dataset [11] when its handle is self-occluded. The rotational error histogram by CosyPose [10] with a single view is shown in Fig. 5. It is clearly seen that the symmetric z-axis exhibits heavy-tailed

TABLE II
ROOT MEAN SQUARE ERROR OF THE ROBOT AND MUG POSE IN THE YCB-VIDEO DATASET

Sequence	Method	Robot position [cm]	Robot rotation [deg]	Object position [cm]	Object rotation [deg]	Relative position [cm]	Relative rotation [deg]
0000	ORB-SLAM3 [34]	1.46	1.52	-	-	-	-
	CosyPose [10]	-	-	-	-	0.77	2.89
	RIEKF [30]	1.69	2.15	1.83	1.64	0.81	1.96
	GM ($\mathcal{T}_L$)	1.62	2.03	1.77	1.80	0.82	1.84
	GM (*midway*)	1.62	2.03	1.75	1.80	0.82	1.84
0007	ORB-SLAM3	3.70	2.63	-	-	-	-
	CosyPose	-	-	-	-	0.93	4.30
	RIEKF	3.57	3.15	3.52	2.07	1.06	2.43
	GM ($\mathcal{T}_L$)	3.53	2.26	3.48	1.87	0.98	1.67
	GM (*midway*)	3.50	2.25	3.46	1.86	0.98	1.67
0022	ORB-SLAM3	3.11	3.63	-	-	-	-
	CosyPose	-	-	-	-	0.90	7.75
	RIEKF	2.77	3.90	2.25	3.24	0.90	3.48
	GM ($\mathcal{T}_L$)	2.73	1.80	2.30	2.64	0.83	2.68
	GM (*midway*)	2.82	1.82	2.36	2.65	0.83	2.69
0027	ORB-SLAM3	1.61	2.94	-	-	-	-
	CosyPose	-	-	-	-	1.15	2.86
	RIEKF	1.66	2.97	1.86	1.09	1.14	2.28
	GM ($\mathcal{T}_L$)	1.51	2.23	1.72	1.10	1.12	2.38
	GM (*midway*)	1.53	2.23	1.74	1.10	1.12	2.38
0033	ORB-SLAM3	1.37	1.95	-	-	-	-
	CosyPose	-	-	-	-	1.02	4.01
	RIEKF	1.35	2.79	1.75	3.61	1.03	2.38
	GM ($\mathcal{T}_L$)	1.52	1.63	1.85	2.53	0.95	2.49
	GM (*midway*)	1.43	1.63	1.78	2.53	0.95	2.49
0039	ORB-SLAM3	2.01	1.99	-	-	-	-
	CosyPose	-	-	-	-	0.83	1.44
	RIEKF	1.82	1.46	1.48	1.38	0.84	0.82
	GM ($\mathcal{T}_L$)	1.93	1.03	1.67	0.93	0.83	0.67
	GM (*midway*)	2.02	1.08	1.64	0.99	0.83	0.67
Mean	ORB-SLAM3	2.21	2.44	-	-	-	-
	CosyPose	-	-	-	-	0.93	3.88
	RIEKF	2.14	2.74	2.12	2.17	0.96	2.23
	GM ($\mathcal{T}_L$)	2.14	1.83	2.13	1.81	0.92	1.96
	GM (*midway*)	2.15	1.84	2.12	1.82	0.92	1.96

TABLE III
AVERAGE EXECUTION TIME IN MILLISEC. FOR GM MERGES PER FRAME

Method	0000	0007	0022	0027	0033	0039	Mean
GM ($\mathcal{T}_L$)	3.10	3.95	4.93	3.97	6.70	6.72	4.90
GM (*midway*)	**2.92**	**3.52**	**4.28**	**3.48**	**5.90**	**5.91**	**4.34**

behavior. We encode this by two GM components in n_j for a simple but effective representation.

We implement the Gaussian sum filter where each hypothesis is an extended Kalman filter parameterized by the right-invariant error based on [29].Weights of each hypothesis are recursively updated according to the Bayes rule [12]. By introducing the multimodal noise, it is evident that the number of hypotheses exponentially increases. Unless otherwise noted, we merge two Gaussian components sequentially that have the largest and the smallest weight until the number of hypotheses reaches the predefined threshold N_h. Also, we merge the largest and second-largest weights of Gaussian components to summarize the estimated quantity. Algorithm 1 shows the Gaussian mixture invariant extended Kalman filter (*GM-IEKF*) with the midway-merge. We implement it in MATLAB with Intel i5-7600 CPU at 3.50 GHz.

VII. EXPERIMENTS

Throughout this section, we designate GM-IEKF with the proposed midway-merge as GM(*midway*), while GM-IEKF with the conventional merge anchored at the tangent space of the larger weight [18] as GM($\mathcal{T}_L$). We define pose error as

$$\epsilon = \ln(\hat{T}\, T^{-1})^{\vee} \tag{34}$$

by the convention in (4). In the simulation part, we present the averaged normalized estimation error squared (ANEES) to measure estimator consistency. At a specific time, ANEES is defined as

$$\text{ANEES} = \frac{1}{N\, N_s} \sum_{i=1}^{N_s} \epsilon_i^T P_i^{-1} \epsilon_i \tag{35}$$

where $N = 6 + 6M$ is a dimension of the state, $N_s = 100$ is the number of simulative runs, and P is the error covariance. Therefore, ANEES should be the unity if the covariance fully explains the actual error.

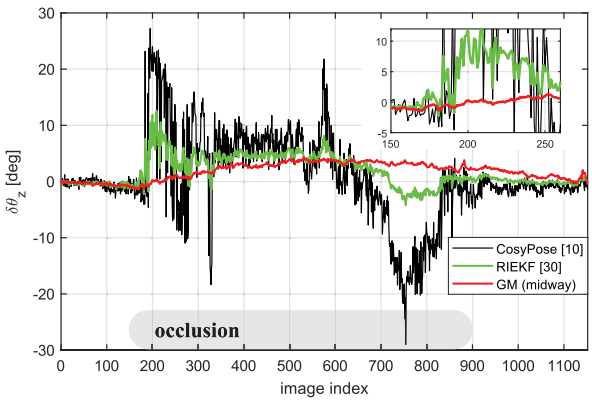

Fig. 6. The rotational error of the symmetric axis of the mug in the 0022 sequence. Our filtering method successfully mitigates large errors by virtue of prior information and the GM noise modeling.

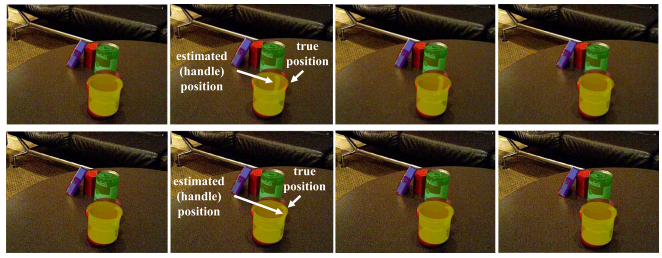

Fig. 7. Temporally consecutive images (image index: from 206 to 209) that include the occluded mug in the 0022 sequence. We render measurements (top row) and filtered pose (bottom row) on top of the image. The measurement clearly exhibits a large error along the symmetric axis.

A. *Monte-Carlo Simulation*

We generate a virtual trajectory shown in Fig. 4(a), odometry readings, and pose detection based on a sensor specification. The odometry noises are set as $0.026\,\mathrm{deg/s}/\sqrt{\mathrm{Hz}}$ and $0.002\,\mathrm{m/s}/\sqrt{\mathrm{Hz}}$, respectively. A virtual detector outputs a 6 DOF pose of the mug where only the rotation z-axis gives the GM noise, that is

$$\begin{aligned} \alpha_1 &= 0.7, \\ \sqrt{R_1} &= \mathrm{diag}(4^\circ,\ 4^\circ,\ 4^\circ,\ 0.01\mathrm{m},\ 0.01\mathrm{m},\ 0.01\mathrm{m}), \\ \alpha_2 &= 0.3, \\ \sqrt{R_2} &= \mathrm{diag}(4^\circ,\ 4^\circ,\ 12^\circ,\ 0.01\mathrm{m},\ 0.01\mathrm{m},\ 0.01\mathrm{m}). \end{aligned} \tag{36}$$

To expose merge methods in an extreme case, we merge the most distant components after the filter update. Given a single object, we set $N_h = 4$ in our simulation.

Table I reports averaged rotational root mean square error (RMSE) and ANEES in 100 runs. We increase the process noise deviation $\sqrt{Q}$ to simulate low-grade sensors. The proposed merge decreases the estimation error further and outputs more consistent estimates as $\sqrt{Q}$ increases. The more uncertain the process model is, the more opportunities multimodal a posteriori occurs. That is, $\|\Delta x\|^2$ in (16) becomes larger. Furthermore, as our method directly merges distributions at the predetermined midway point, it spares computing the adjoint to back-project the merged distribution. This reduces execution time, as shown in Fig. 4(b). We also provide a comparison result with a longer moving distance in a photo-realistic simulator. Please see our supplementary video for further details.

B. *Real-World Dataset*

In the YCB-Video dataset [11], we validate the object-based SLAM formulated by GM-IEKF with the midway-merge method. The goal of this test is to investigate the efficacy of the midway-merge and how estimation error decreases with respect to the state-of-the-art 6 DOF pose detector, object SLAM (RIEKF [30]) and low-level SLAM (ORB-SLAM3 [34] with RGB-D). Among possible candidates such as CosyPose [10] and PoseRBPF [31], we use CosyPose with a single view as a sensor trained by the authors because of its superior performance. We select data sequences that contain a mug in scenes that possess heavy-tailed noise distribution along the symmetric axis as analyzed in Fig. 5. The noise characteristic of CosyPose is predictable since the authors introduce *the symmetric distance* in the training loss function. This allows the network to train symmetric objects, but at the same time, this admits estimation errors due to the symmetry. We model this error by a GM distribution in the filter update.

Objects other than the mug in scenes are also modeled as heavy-tailed distribution for a generalization of the proposed approach. Given multiple objects, we set $N_h = 12$. Please note that we exclude objects that output very unstable estimates, such as bricks and bowls. Since odometry measurements are unavailable in the dataset, we use a constant velocity model.

Table II reports RMSE of the robot (camera) and mug pose, and robot-mug relative pose. GM-IEKF with two merge methods outputs almost identical estimation error. We interpret this is due to the low noise level Q (slow and smooth motion) as analyzed in Table I. However, the midway-merge consistently reduces execution time with respect to the conventional merge [18] as summarized in Table III by sparing the back-projection procedure. The execution time for GM(*midway*) corresponds to the time to process the 11 line in Algorithm 1 per frame, while the midway-merge is replaced by the conventional merge in GM($\mathcal{T}_L$). The back-projection involves two dense matrix multiplication in a state dimension N. In the YCB-Video dataset, $N = 36$ depending on the number of objects, and Table III indicates that 13 times GM merge on average per frame is not trivial.

In contrast, filtered pose by our approach GM(*midway*) reduces 49.5% of relative rotation error when compared to CosyPose. We observe the significant rotational error reduction in the sequence 0022 where the mug handle is occluded in most scenes. Figs. 6 and 7 highlight this that the pose detector suffers from the large deviation when the handle is occluded, while our approach mitigates this by perceiving the prior pose. Please see our supplementary video for further visualization. RIEKF [30], fed by the identical 6 DOF pose as ours, also improves pose error when compared to CosyPose thanks to the prior information. However, the method cannot explicitly address the noise due to symmetry as opposed to ours, which leads to a large rotation error as highlighted in Fig. 6. Lastly, GM(*midway*) improves the robot localization accuracy with respect to ORB-SLAM3 (RGB-D) in most sequences, even without using depth measurements. This implies that objects possess rich information for localization and mapping.

VIII. Conclusion

We have proposed the *Gaussian mixture midway-merge* that merges Gaussian distributions on matrix Lie groups. Specifically, our approach computes the merged mean and transforms the covariance matrices at the corresponding tangent space. This simple but powerful technique decreases information loss when the distance between mean matrices increases and computation time by sparing the adjoint. As a promising example, we formulate GM-IEKF, the Gaussian sum filter with the proposed merge for a symmetric object SLAM problem. A thorough Monte-Carlo simulation and demonstration on real-world as well as synthetic datasets reveal that the midway-merge has a lighter computational burden and a nice property when the noise level increases when compared to the conventional merge. We believe that our method has great potential in state estimation on matrix Lie groups that deals with Gaussian mixtures. Our future work includes evaluating the presented approach with diverse symmetric objects in a long-range scenario. We would like to generalize our GM noise model to explain a symmetric object such as a bowl without any distinguished textures.

References

[1] T. Schon, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2279–2289, Jul. 2005.

[2] J. Park, Y.-G. Park, and C. G. Park, "Parameter estimation of radar noise model for terrain referenced navigation using a new EM initialization method," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 1, pp. 107–112, Feb. 2020.

[3] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[4] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *Int. J. Robot. Res.*, vol. 32, no. 7, pp. 826–840, 2013.

[5] K. J. Doherty, D. P. Baxter, E. Schneeweiss, and J. J. Leonard, "Probabilistic data association via mixture models for robust semantic SLAM," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 1098–1104.

[6] C. O'Meadhra, W. Tabib, and N. Michael, "Variable resolution occupancy mapping using Gaussian mixture models," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 2015–2022, Apr. 2019.

[7] H. Huang, H. Ye, Y. Sun, and M. Liu, "GMMLoc: Structure consistent visual localization with Gaussian mixture models," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 5043–5050, Oct. 2020.

[8] J. Fu, Q. Huang, K. Doherty, Y. Wang, and J. J. Leonard, "A multi-hypothesis approach to pose Ambiguity in object-based SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 7639–7646.

[9] Z. Lu, Q. Huang, K. Doherty, and J. J. Leonard, "Consensus-informed optimization over mixtures for Ambiguity-aware object SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 5432–5439.

[10] Y. Labbé, J. Carpentier, M. Aubry, and J. Sivic, "CosyPose: Consistent multi-view multi-object 6D pose estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 574–591.

[11] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes," in *Proc. Robot.: Sci. Syst.*, 2018, vol. 14.

[12] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Autom. Control*, vol. 17, no. 4, pp. 439–448, Aug. 1972.

[13] C. Qian, C. Song, S. Li, Q. Chen, and J. Guo, "Algorithm of Gaussian sum filter based on SGQF for nonlinear non-Gaussian models," *Int. J. Control, Automat. Syst.*, vol. 19, no. 8, pp. 2830–2841, 2021.

[14] T. D. Barfoot, *State Estimation for Robotics*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[15] M. Brossard, A. Barrau, P. Chauchat, and S. Bonnabel, "Associating uncertainty to extended poses for on lie group IMU preintegration with rotating Earth," *IEEE Trans. Robot.*, vol. 38, no. 2, pp. 998–1015, Apr. 2022.

[16] A. Barrau and S. Bonnabel, "The invariant extended Kalman filter as a stable observer," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1797–1812, Apr. 2017.

[17] A. Barrau and S. Bonnabel, "Invariant particle filtering with application to localization," in *Proc. IEEE 53rd Conf. Decis. Control*, 2014, pp. 5599–5605.

[18] J. Ćesić, I. Marković, and I. Petrović, "Mixture reduction on matrix lie groups," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1719–1723, Nov. 2017.

[19] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. Kelly, and A. J. Davison, "SLAM : Simultaneous localisation and mapping at the level of objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1352–1359.

[20] J. L. Williams and P. S. Maybeck, "Cost-function-based Gaussian mixture reduction for target tracking," in *Proc. 6th Int. Conf. Inf. Fusion*, 2003, pp. 1047–1054.

[21] A. R. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 989–999, Jul. 2007.

[22] A. Assa and K. N. Plataniotis, "Wasserstein-distance-based Gaussian mixture reduction," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1465–1469, Oct. 2018.

[23] T. D. Barfoot and P. T. Furgale, "Associating uncertainty with three-dimensional poses for use in estimation problems," *IEEE Trans. Robot.*, vol. 30, no. 3, pp. 679–693, Jun. 2014.

[24] G. Bourmaud, R. Mégret, M. Arnaudon, and A. Giremus, "Continuous-discrete extended Kalman filter on matrix lie groups using concentrated Gaussian distributions," *J. Math. Imag. Vis.*, vol. 51, no. 1, pp. 209–228, 2015.

[25] D. Frost, V. Prisacariu, and D. Murray, "Recovering stable scale in monocular SLAM using object-supplemented bundle adjustment," *IEEE Trans. Robot.*, vol. 34, no. 3, pp. 736–747, Jun. 2018.

[26] L. Nicholson, M. Milford, and N. Sünderhauf, "QuadricSLAM: Dual quadrics from object detections as landmarks in object-oriented SLAM," *IEEE Robot. Automat. Lett.*, vol. 4, no. 1, pp. 1–8, Jan. 2019.

[27] S. Yang and S. Scherer, "CubeSLAM: Monocular 3-D object SLAM," *IEEE Trans. Robot.*, vol. 35, no. 4, pp. 925–938, Aug. 2019.

[28] X. Liu et al., "Large-scale autonomous flight with real-time semantic SLAM under dense forest canopy," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 5512–5519, Apr. 2022.

[29] J. H. Jung and C. G. Park, "Object-based visual-inertial navigation system on matrix lie group," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 9499–9505.

[30] Y. Song, Z. Zhang, J. Wu, Y. Wang, L. Zhao, and S. Huang, "A right invariant extended Kalman filter for object based SLAM," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 1316–1323, Apr. 2022.

[31] X. Deng, A. Mousavian, Y. Xiang, F. Xia, T. Bretl, and D. Fox, "PoseRBPF: A rao–blackwellized particle filter for 6-D object pose tracking," *IEEE Trans. Robot.*, vol. 37, no. 5, pp. 1328–1342, Oct. 2021.

[32] M. Sundermeyer, Z.-C. Marton, M. Durner, M. Brucker, and R. Triebel, "Implicit 3D orientation learning for 6D object detection from RGB images," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 699–715.

[33] N. Merrill et al., "Symmetry and uncertainty-aware object SLAM for 6DoF object pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 14901–14910.

[34] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.