2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Distribution Estimation for Global Data Association via Approximate Bayesian Inference

Yixuan Jia[1], Mason B. Peterson[1], Qingyuan Li[1], Yulun Tian[2], Jonathan P. How[1]

***Abstract*— Global data association is an essential prerequisite for robot operation in environments seen at different times or by different robots. Repetitive or symmetric data creates significant challenges for existing methods, which typically rely on maximum likelihood estimation or maximum consensus to produce a *single* set of associations. However, in these ambiguous scenarios, the distribution of solutions to global data association problems is often highly multimodal, and such single-solution approaches frequently fail. In this work, we introduce a data association framework that leverages approximate Bayesian inference to capture multiple solution modes to the data association problem, thereby avoiding premature commitment to a single solution under ambiguity. Our approach represents hypothetical solutions as particles that evolve via deterministic or randomized updates, naturally parallelizable on GPUs, to cover the modes of the underlying solution distribution. Simulated and real-world experiments with highly ambiguous data show that our method correctly estimates the distribution over transformations when registering point clouds or object maps. Code is available at: https://github.com/mit-acl/mmda.**

I. INTRODUCTION

Data association is essential in many robotic applications, enabling key perception technologies such as dynamic object tracking [1]–[3] and simultaneous localization and mapping (SLAM) [4]–[6]. In these scenarios, robots must recognize when an object or feature they are currently observing is the same as something they (or another robot) may have seen from a different perspective. Without correct data association, the environment representation may be inconsistent, leading to undesirable behaviors in downstream tasks (e.g., incorrect associations in loop closure detection can lead to dramatically distorted maps [6]).

This work considers *global* data association, which aims to find pairwise correspondences for registering two point clouds [7], object-level maps [8], or scene graphs [9] *without an initial guess*. While significant progress has been made towards point-to-point data association [10]–[12], most existing methods formulate data association using maximum likelihood estimation or maximum consensus, returning a *single* solution representing the most likely relative pose estimate or most consistent set of data associations. However, both approaches can fail in the presence of ambiguous data, where multiple likely solutions coexist and cannot be distinguished without additional observations. Such ambiguities are common in human environments, which often contain symmetrical or repetitive structures. In these scenarios, the success of the solver may be determined purely by the optimization initialization or the realization of measurement noise. For example, consider a robot trying to localize itself from a reference map within the staircase shown in Fig. 1. The repetitive nature of each floor of the staircase results in many likely sets of object associations, yielding a multimodal distribution over the robot pose estimate. If the robot were to choose the most likely option, it would believe that it was two floors above its actual location.

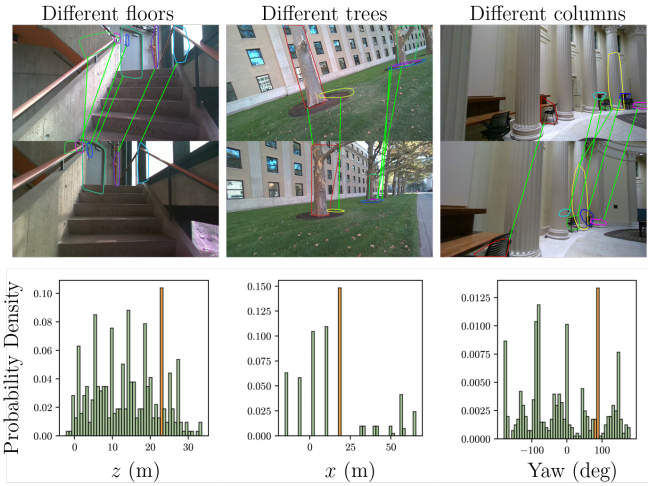

Fig. 1: Symmetric and repetitive structures are common in human environments and induce perceptual aliasing. To localize, robots must explicitly model the uncertainty from ambiguities inherent in those environments. The top row shows several such scenarios, where the visualized object associations are perceptually similar but incorrect, necessitating a multimodal representation. The bottom row shows the multimodal distribution generated by our proposed method. The orange mode is incorrect and corresponds to the ambiguous associations shown in the top row.

Currently, there are mainly three applicable approaches to avoid incorrect estimations in ambiguous scenarios: (1) outlier rejection, (2) probabilistic data association, and (3) probabilistic pose estimation. Existing outlier rejection methods [12]–[15] typically assume that all inlier measurements are mutually consistent while all outliers are uncorrelated. As a result, they do not work well when measurements form multiple consistent clusters and the resulting solution distribution has multiple modes. On the other hand, probabilistic data association methods [4], [5], [16], [17] model the distributions of associations in a factor graph by introducing new measurement models and factors. However, these methods primarily address local ambiguities rather than

This work is supported by ARL DCIST under Cooperative Agreement Number W911NF-17-2-0181.

[1]Massachusetts Institute of Technology, Cambridge, MA 02139, USA. {yixuany, masonbp, andyli27, jhow}@mit.edu.
[2]Robotics Department, University of Michigan, MI 48109, USA. yulunt@umich.edu.

global ambiguities that are inherent in highly symmetric or repetitive environments. Another line of work augments existing methods such as ICP with probabilistic methods [18], [19] to estimate uncertainty in pose estimation. However, as shown in our experiments, such approaches struggle to estimate highly non-uniform distributions.

Contributions. In this work, we address the limitations of the aforementioned work by proposing novel algorithms for *multimodal* global data association. Our core insight is to view data association through the lens of approximate Bayesian inference, by interpreting the former's objective function as log likelihood defined over sets of candidate associations. Building on this insight, we develop deterministic and randomized multimodal global data association algorithms that connect with two well-established Bayesian inference frameworks. Specifically, our approach models hypothetical solutions as a set of particles that are updated using approximate-inference-based update rules. We show that the proposed method can handle difficulties specific to the optimization formulation of global data association and directly benefit from modern GPU-parallelized optimization solvers. Experimentally, we perform extensive comparison studies against existing approaches and show the proposed approach can capture both highly peaked and highly uniform solution distributions.

II. Related Works

Data Association In robotics, data association refers to the task of matching observations of the environment across agents or time. Specifically, we address methods that consider distributions over solutions to the data association problem, rather than methods that commit to a single hypothesis (i.e., a single most likely pose estimate or a single set of data associations). Although recent work has begun to address multimodal data association problems, most of these efforts concentrate on SLAM backend design, particularly in the context of factor graph optimization. For example, work on probabilistic data association [4], [5], [16] focuses on representing distributions of associations in the factor graph by introducing new measurement models. However, this line of work often assumes good initial guesses provided by the frontend and usually focuses on extracting the most likely solution instead of estimating the full distribution. Another approach uses Multi-Hypothesis Tracking (MHT) [2], [3], [17] to track multiple possible solutions in data association problems, but MHT is typically computationally expensive due to the exponential growth of the hypothesis tree.

The most relevant line of work to ours is [18], [19]. In [18], Maken et al. develop a variant of Iterative Closest Point (ICP), termed Bayesian ICP, which uses stochastic gradient Langevin dynamics [20] to obtain samples from the posterior distributions of relative poses between point clouds. Because Bayesian ICP is computationally expensive and not parallelized, the authors then propose Stein ICP [19], which employs Stein variational gradient descent (SVGD) to approximate the posterior distribution more efficiently, achieving better empirical performance. Although SVGD helps to discover different modes, the quality of the generated distribution is sensitive to initialization and parameter tuning.

Bayesian Inference in Robotics Sampling from, or approximating, posteriors defined by unnormalized densities is central in statistics, with two main approaches: MCMC and variational inference (VI). In robotics, SVGD, which is a VI method, has been used for state estimation and motion planning [19], [21]–[23]; motivated by these successes, we develop an SVGD-based variant and evaluate it on real-world global data association. We also develop a variant based on Langevin dynamics, which has been studied for both escaping local minima in stochastic optimization and approximating Bayesian posteriors [24], [25], and show it is effective for posterior approximation in global data association.

III. Preliminaries

Given two ordered lists of points $\mathcal{S}, \mathcal{T}$, we can form a list of potential associations. For example, if each point in $\mathcal{S}$ can be associated with each point in $\mathcal{T}$, we can obtain $|\mathcal{S}| \times |\mathcal{T}|$ potential associations. The association between point $s \in \mathcal{S}$ and point $t \in \mathcal{T}$ is denoted as (s, t). To ease the burden of notation, we will use the index of a point interchangeably with its coordinate, e.g., s will denote both the s-th point in $\mathcal{S}$ as well as the coordinate of the s-th point in $\mathcal{S}$ expressed in a local reference frame (e.g. robot's sensor frame). We use $\|\cdot\|_2$ to denote the 2-norm in Euclidean space.

CLIPPER Given n potential associations (e.g., $n = |\mathcal{S}| \times |\mathcal{T}|$), we can form a symmetric affinity matrix $M \in [0,1]^{n \times n}$. Let $i = (s_i, t_i)$ and $j = (s_j, t_j)$, CLIPPER [26] then defines $M[i,j] = e^{-d((s_i,t_i),(s_j,t_j))^2/(2\sigma^2)}$ if $d((s_i,t_i),(s_j,t_j)) < \varepsilon$, and $M[i,j] = 0$ otherwise, where $d((s_i,t_i),(s_j,t_j)) = |\|s_i - s_j\|_2 - \|t_i - t_j\|_2|$. Intuitively, $d(\cdot,\cdot)$ measures the violation of geometric consistency, since a consistent set of associations should preserve the relative distances of points. Then the problem of finding the maximum consistent set of associations can be formulated as follows:

Problem 1.

$$\max_{u \in \{0,1\}^n} \frac{u^\top M u}{u^\top u}$$
$$\text{subject to} \quad u_i u_j = 0 \ \text{ if } \ M[i,j] = 0, \ \forall i, j.$$

The matrix M can be viewed as the sum of a weighted adjacency matrix and the identity matrix, where the weighted adjacency matrix defines a graph called the *consistency graph*, wherein nodes represent potential associations and edges connect nodes only if they are consistent. In the case where M is binary, Problem 1 is equivalent to finding the maximum clique in the consistency graph, which is expensive to obtain for large graphs. Therefore, CLIPPER solves a relaxed problem, described as follows. Define another matrix M_d such that $M_d[i,j] = M[i,j]$ if $M[i,j] \neq 0$, and $M_d[i,j] = -d$ otherwise, where $d \in \mathbb{R}_+$. Intuitively, if two associations are deemed inconsistent, the corresponding entry in M_d will incur penalties if selected. Then, inspired by [27], CLIPPER solves an optimization problem of the form:

Problem 2.

$$\max_{u \in \mathbb{R}^n_+} \quad u^T M_d u$$
$$\text{subject to} \quad \|u\|_2 \leq 1.$$

The optima lie on the boundary of $\|u\|_2 \leq 1$ [26], so we can equivalently write $u \in \mathbb{R}^n_+ \cap \mathbb{S}^{n-1}$ and remove the inequality constraint. Moreover, as remarked in [26], when $d \geq n$, the (local) optima of Problem 2 correspond to the (local) optima of the original Problem 1. The local optima correspond to maximal cliques in the consistency graph while the global optimum corresponds to the maximum clique.

Stein Variational Gradient Descent SVGD aims to approximate a target distribution with an unnormalized density function $p : \mathbb{R}^n \to \mathbb{R}$ using a set of particles $X = \{x_i\}_{i=1}^N,\ x_i \in \mathbb{R}^n\ \forall i$. Denoting the distribution represented by the set of particles X as q_X, the goal is to find X such that $\mathrm{KL}(q_X \parallel p)$ is minimized. Suppose we would like to iteratively update the particles via a map $T(x)$ that takes the form $x + \alpha\phi(x)$ where $\alpha > 0$ is the step size. The optimal perturbation direction $\phi(\cdot)$ is $\max_{\|\phi\|_{\mathcal{H}} \leq 1} \left\{ -\frac{d}{d\alpha} \mathrm{KL}(T q_X \parallel p)\big|_{\alpha=0} \right\}$, where $\mathcal{H}$ is a reproducing kernel Hilbert space [28]. Intuitively, $\phi(\cdot)$ is the functional gradient of the KL divergence with respect to the particles. A key result from [29] shows that $\phi(\cdot)$ takes a closed-form expression:

$$\phi(x_i) = \frac{1}{N} \sum_{j=1}^{N} \nabla \log p(x_j)\, k(x_i, x_j) + \nabla_{x_j} k(x_i, x_j), \quad (1)$$

where $k(\cdot, \cdot)$ is a kernel specified by the user. Intuitively, SVGD works by pushing the particles to high-density regions via the $\nabla \log p(\cdot)$ term, while pulling them apart from each other to encourage discovering different modes via the repulsive term $\nabla k(\cdot, \cdot)$. Note that the repulsive term is deterministic, which makes SVGD a deterministic algorithm. With only a single particle, SVGD reduces to exact maximum a posteriori (MAP) optimization [28].

Langevin Dynamics Given an unnormalized density function $p : \mathbb{R}^n \to \mathbb{R}$, Langevin dynamics (LD) can be applied to draw samples from the distribution defined by $p(\cdot)$ by iteratively updating the particle [28]:

$$x_{t+1} \leftarrow x_t + \alpha \nabla \log p(x_t) + 2\sqrt{\alpha}\xi_t,\ \xi_t \sim \mathcal{N}(0, 1). \quad (2)$$

Intuitively, the second term on the right side encourages particles to cover regions with higher density, and the third term encourages particles to spread out so that more modes can be discovered. Unlike SVGD, LD is a stochastic algorithm due to the random noise term. Moreover, the computational complexity is reduced by removing the $O(n^2)$ computation needed to evaluate the kernel terms in SVGD. This update can be viewed as a discrete version of an Itô diffusion $dx_t = -\nabla \log p(x_t) dt + 2dB_t$ where B_t is a standard Brownian motion.

IV. Proposed Method

In this section, we introduce our proposed methods. The general strategy is to view Problem 2 from a probabilistic perspective in which the objective becomes drawing samples from a target distribution. Then, the approximate Bayesian inference techniques introduced in the previous section can be applied to approximate the target distribution.

Referring to Problem 2, let $F_d(u) = u^T M_d u$. The data association problem can be viewed as obtaining samples from the posterior distribution (assuming a uniform prior):

$$p(u|\mathcal{S}, \mathcal{T}) \propto p(\mathcal{S}, \mathcal{T}|u) \propto \exp(F_d(u)),\ u \in \mathbb{R}^n_+ \cap \mathbb{S}^{n-1}. \quad (3)$$

This type of reformulation of an optimization problem into an inference problem is common in trajectory optimization and control literature [30], [31] . Since $p(u|\mathcal{S}, \mathcal{T})$ rarely admits a closed-form expression, our goal is to approximate $p(u|\mathcal{S}, \mathcal{T})$ with a distribution $q(u)$ that is tractable for computation.

In this work, we propose two variants of our method: **Stein CLIPPER** based on SVGD and **Langevin CLIPPER** based on Langevin dynamics. Both algorithms update particles in parallel, enabling efficient implementation on GPUs.

Stein CLIPPER Recall the update direction in SVGD given by Eq. (1). To compute the update, we need the gradient of the log posterior (i.e. the score) as well as a kernel function. Since $p(u|\mathcal{S}, \mathcal{T}) \propto \exp(F_d(u))$, $\log p(u|\mathcal{S}, \mathcal{T}) \propto F_d(u)$. Because the score is independent of the normalization constant, we have:

$$\nabla_u \log p(u|\mathcal{S}, \mathcal{T}) = \nabla_u F_d(u) = 2u^T M_d. \quad (4)$$

Note that we adopt the convention that the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix [32]. We choose the widely used radial basis function (RBF) kernel [33], which takes the form $k(x, x') = e^{-\|x - x'\|_2^2 / (2\sigma_k^2)}$.

See Algorithm 1 for the overall Stein CLIPPER approach. Line 1 initializes n_p particles from $\mathrm{Uniform}([0, 1]^n)$ that each represents a possible set of associations (i.e. u in Problem 2). We use a homotopy method [26] to gradually increase d (see Problem 1). We initialize Δd to be the maximum eigenvalue of M and increase d by Δd for every outer iteration, until $d \geq n$, at which point the optima of Problem 1 should be recovered (see Problem 2 and the comments immediately after it). Empirically, we find setting $\Delta d = \lambda_1(M)$ offers good trade-offs between speed and convergence without using a problem-specific parameter. This homotopy step is important for both the original CLIPPER (to avoid being trapped in local minima) and Stein CLIPPER (to prevent particle degeneracy). The inner loop (Lines 6 - 10) consists of the main optimization steps. Intuitively, for each fixed d, we solve Problem 2 via gradient ascent where the gradient is computed via Eq. (1) and adjusted by the adaptive gradient algorithm AdaGrad [34]. AdaGrad adjusts and rescales the learning rate for each particle automatically. Empirically, we find AdaGrad very important for returning diverse particles and preventing particle degeneracy. After the update, we perform a projection step to project each particle back to $\mathbb{R}^n_+ \cap \mathbb{S}^{n-1}$. Finally, a clique is extracted from each particle by taking the associations corresponding to their top $\hat{\omega} = \mathrm{round}(u^\top M_d u)$ entries as done in [26].

Langevin CLIPPER Recall that the update step from Langevin dynamics is given by Eq. (2). Thus we only need

Algorithm 1 Stein CLIPPER

Require: Affinity matrix $M \in [0,1]^{n \times n}$, step size α, kernel $k(\cdot,\ \cdot)$

1: $\theta \leftarrow \text{rand}(n_p, n)$ ▷ initialize uniformly in $[0,1]$
2: $d \leftarrow 0$
3: $\Delta d \leftarrow \lambda_1(M)$ ▷ $\lambda_1(M)$: maximum eigenvalue of M
4: **while** $d < n$ **do**
5: $\quad d \leftarrow d + \Delta d$
6: $\quad$ **while** max iterations not reached **do**
7: $\quad\quad M_d \leftarrow M - dC$
8: $\quad\quad$ Compute $\phi(\theta_i)$ via Eq. (1) for each $i \in [n_p]$
9: $\quad\quad \theta \leftarrow \theta + \alpha \cdot \text{AdaGrad}(\phi(\theta))$
10: $\quad\quad \theta \leftarrow \max(\theta/\|\theta\|,\ 0)$
11: Extract a clique from each θ_i by taking the top $\hat{\omega}_i = \text{round}({\theta_i}^\top M_d \theta_i)$ entries of θ_i

Algorithm 2 Langevin CLIPPER

Require: Affinity matrix $M \in [0,1]^{n \times n}$, step size α

1: $\theta \leftarrow \text{rand}(n_p, n)$ ▷ initialize uniformly in $[0,1]$
2: $M_d \leftarrow M - nC$
3: **while** max iterations not reached **do**
4: $\quad$ Compute $\phi(\theta_i)$ via Eq. (2) for each $i \in [n_p]$
5: $\quad \theta \leftarrow \theta + \alpha \cdot \text{AdaGrad}(\phi(\theta))$
6: $\quad \theta \leftarrow \max(\theta/\|\theta\|,\ 0)$
7: Extract a clique from each θ_i by taking the top $\hat{\omega}_i = \text{round}({\theta_i}^\top M_d \theta_i)$ entries of θ_i

the expression for $\nabla_u \log p(u|\mathcal{S},\mathcal{T})$ to perform the update, which was derived in Eq. (4).

The overall Langevin CLIPPER algorithm is recorded in Algorithm 2. Note that d is directly set to n (Line 2) instead of gradually being increased via the homotopy method. We find that Langevin CLIPPER is able to reliably capture the full distribution without the homotopy step, avoiding the particle degeneracy observed with SVGD. We hypothesize that this is due to a rigorously motivated relative noise scale with respect to the score ($\sqrt{\alpha}$ vs. α) [25], which automatically adjusts the noise scale to prevent overshooting while still being able to spread out the particles enough to escape the basin of local minima. In contrast, the kernel term in SVGD has to be manually selected and properly tuned. Lines 3 - 7 iteratively update the particles by moving them in the direction given by Eq. (2). Again, AdaGrad helps rescale the step size for each particle and a clique is extracted by taking the top $\hat{\omega}_i$ entries of θ_i for each $i \in \{1,\dots,n_p\}$.

Remark 1 Theoretically, when the affinity matrix M is binary, the modes of the distribution obtained from the proposed methods correspond to the local optima of Problem 2, which correspond to maximal cliques in the consistency graph. Enumerating all maximal cliques has a worst-case run time $O(3^{n/3})$ [35], which is intractable for most data association applications (e.g., associating $|\mathcal{S}| = 40$ points to $|\mathcal{T}| = 40$ points would result in $n = 40^2 = 1600$). The proposed methods enable the approximate enumeration of maximal cliques in a much more tractable manner, as will be shown in Section VI-C.

Fig. 2: Visualizations of some experimental setups. The bottom-row middle image shows the setup used for maximal clique enumeration (Section VI-C), where the object map consists of bushes. The remaining images show inspiration for the simulation setups (Section V): **Circle** (top-left), **Two Lines of Points** (bottom-left), and **Office** (top-right). The bottom-right image shows the objects used for point cloud registration (Section V-C).

V. Simulation Experiments

Our simulation experiments consist of two types of setups: (1) **Simulated Object Map Registration** presented in Section V-B, where the goal is to evaluate the proposed methods under perfect sensing (no noise or outlier objects). These maps are motivated by real-world examples, as shown in Fig. 2. (2) **Point Cloud Registration** presented in Section V-C, where we show that the proposed methods can also be applied to point cloud registration tasks as studied in [19].

A. Experiment Setup

Baseline Methods We compare with two baselines:

- TCAFF [2], which generates multiple hypotheses sequentially by removing correspondences obtained in previous runs and then re-solving the problem.
- Stein ICP [19], which employs SVGD to perform mini-batch gradient descent on translational and rotational parts of the particles independently.

We also attempted to use Bayesian ICP [18], but it consistently diverged on our simulated object map examples.

Ground Truth Distribution Generation In [19], ground truth pose distributions are generated by running standard ICP from different initial guesses sampled from a small range ($\pm 1, \sim \pm 10\,\text{deg}$), which may not capture the distributions often encountered in object-level map association problems, where the transformations can spread all over $\text{SE}(3)$. For example, consider the circle example shown in Fig. 2 (top left). We run ICP from 1M initial poses by uniformly sampling rotations from $\text{SO}(3)$ and translations in the bounding box defined by the extent of the map; the obtained yaw distribution is shown in Fig. 3 (yellow). We observe that the distribution is almost uniform. However, the peaks of the yaw distribution should only appear at multiples of $45\,\text{deg}$. One approach would be to enumerate all maximal cliques using Bron–Kerbosch (BK) [35]; however, this is intractable for all but the smallest of problems. Instead, we use the RANSAC-based alignment method described in [36] to generate initial proposals by running it 1M times and keeping at most 5000 proposals. We then refine the accepted proposals with ICP

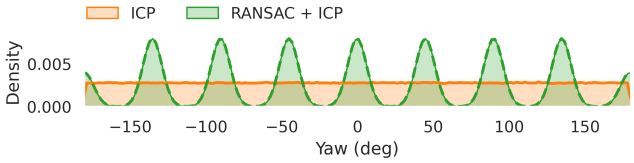

Fig. 3: Visualization of Distribution on the Circle Example Generated by ICP vs. RANSAC proposals + ICP.

[37]. The result is shown in Fig. 3 (green). As we can see from the plot, this method cleanly captures the modes. The downside of this method is that it takes minutes to generate enough proposals to approximate the distribution. However, this is sufficient for the purpose of approximating the ground truth distribution offline.

Metric To capture the discrepancy between distributions over $\mathrm{SE}(3)$, we use three different metrics: the Energy Distance (ED) [38], Maximum Mean Discrepancy (MMD) [39], and Wasserstein-1 distance (W1) [40]. For all three, we use Euclidean distance as the distance function on translation and chordal distance [41] as the distance function on rotation.

Implementation and Parameters Since Stein ICP requires upper and lower bounds to sample initial poses from, for object map experiments, we set all the bounds on translations to be 0.5m away from the bounding box enclosing the object maps. For the object point cloud examples, we sample translations from the unit cube, since the point clouds are rescaled to be within the unit cube. Rotations are sampled from $[-180\,\mathrm{deg}, 180\,\mathrm{deg})$ to cover all possibilities. We use 1000 particles and 1000 iterations for both Stein ICP and the proposed methods to ensure the generated distributions are representative of the methods' capabilities. For TCAFF, since it discovers different solutions sequentially, we only generated 100 solutions to avoid long run times. The learning rates of Stein ICP, Stein CLIPPER, and Langevin CLIPPER are 0.01, 0.001, and 1.0 respectively, which were obtained by tuning them on the two lines of points example. The AdaGrad parameters for the proposed methods are both set to 0.9. The kernel bandwidth of Stein CLIPPER is set to 0.005. The σ and ε for generating the affinity matrix M are set to $(0.1\mathrm{m}, 0.2\mathrm{m})$ for the object point cloud examples and $(0.4\mathrm{m}, 0.6\mathrm{m})$ for everything else. Both proposed methods are implemented with PyTorch [42].

B. Simulated Object Maps

We report distribution similarity and runtime metrics in Table I. These values reflect the mean and standard deviation of running each method 10 times.

Circle The first simulated object map is a circle of eight points lying on the xy plane, motivated by Fig. 2 (top left). The yaw distributions obtained from all methods are shown in Fig. 6 (left). TCAFF and the proposed methods successfully capture this discrete symmetry while Stein ICP struggles. Quantitative results are recorded in Table I (first block). TCAFF and the proposed methods have low rotation errors, matching the qualitative results.

Two Lines of Points Motivated by Fig. 2 (bottom left), this configuration tests the ability of the proposed methods to capture a translational distribution with discrete modes. Langevin CLIPPER and TCAFF are better at capturing the translation distribution exhibited in this example.

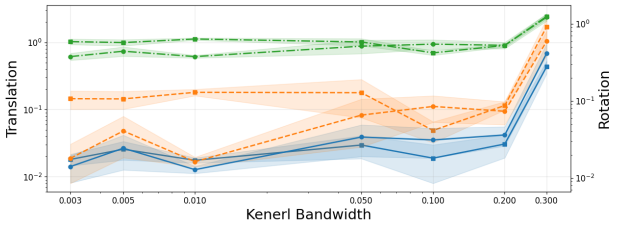

Fig. 4: Ablations of `Stein CLIPPER` over kernel bandwidth on the office example. We note that it fails to return solutions for bandwidth < 0.003 or ≥ 0.5.

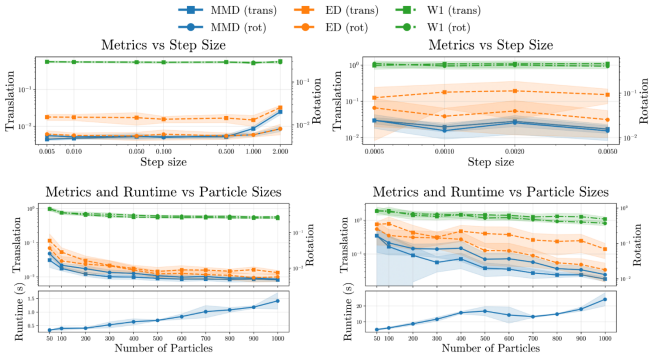

Fig. 5: Ablations of Langevin CLIPPER (left) and Stein CLIPPER (right) over particle sizes and step sizes on the office example. We can see that Langevin CLIPPER is less sensitive to the parameters and can work with a larger range of step sizes.

Office The third simulated object map is an office, shown on the right of the top row in Fig. 2, where the object meshes are taken from ModelNet [43]. We built two U-shaped pods, each consisting of 6 sets of workstations (including a chair, desk, keyboard, and monitor). This setup is motivated by the real-world office setup in Fig. 2 (top middle). The quantitative results are recorded in the third block in Table I. As we can see, the proposed methods obtain the best results overall, although Stein CLIPPER exhibits longer runtime. TCAFF's runtime greatly increases in this example, which is expected due to the increase in the number of objects and the sequential computation of TCAFF.

C. Object Point Clouds

To investigate the performance of the proposed methods on registering object point clouds, we choose three object scans from the Google Scanned Objects dataset [44], shown in Fig. 2 (bottom right). This set of experiments is motivated by the experiment setup in Stein ICP [19]. The point clouds are rescaled to a unit cube and then downsampled with voxel size 0.05m. Noise is injected from $\mathcal{N}(0, 0.025)$. We use FPFH [45] features to generate putative associations for TCAFF and the proposed methods. A visualization of the yaw distributions for the bowl is provided in Fig. 6 (right). We observe that Stein ICP and Langevin CLIPPER are better at capturing the distribution which is more uniform than those in the object map examples. The quantitative results are recorded in the lower half of Table I. We observe

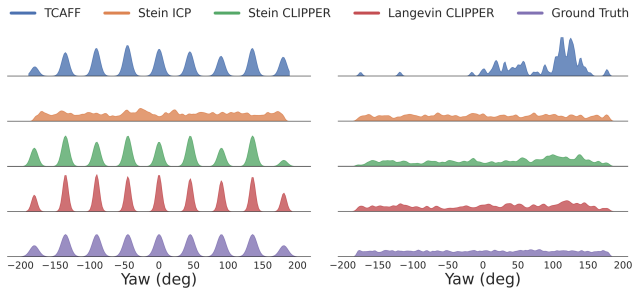

Fig. 6: Comparisons of Yaw Distributions on Circle (left) and Bowl (right). The proposed methods can handle both highly-peaked and uniform distributions.

that Langevin CLIPPER performs best overall, which shows the ability of Langevin CLIPPER to capture both highly-peaked distributions and uniform distributions. Stein ICP performs better on this set of experiments than on object map registrations, as it was developed and evaluated on similar scenarios. Stein CLIPPER performs worse on this set of experiments, which we believe is due to the kernel bandwidth. In practice, one could use different bandwidths for different experiments, but we keep the bandwidth the same to test generalizability.

D. Ablation Study

An ablation study of the sensitivity of the proposed methods to the particle size and learning rate is performed, and the result is recorded in Fig. 5. We find Langevin CLIPPER is less sensitive to the parameters compared to Stein CLIPPER. In addition, an ablation of Stein CLIPPER over the kernel bandwidth is performed and the result is reported in Fig. 4.

VI. Real World Experiments

To evaluate the capabilities of the proposed methods in scenarios with real sensor noise, we perform real-world object map registration experiments. Note that we use the *same parameters* for proposed methods as we used for simulated object maps throughout this section.

A. 3RScan Dataset

We evaluate the proposed methods using the 3RScan dataset [46], which contains scans of indoor scenes captured with RGB-D cameras. This dataset tests whether the proposed methods can be applied to general scenes, irrespective of the multimodal nature of the underlying data association distribution, as the 3RScan scenes may not necessarily elicit multimodal association solutions. We randomly select 100 scenes from 3RScan, where 2 scans are randomly selected from each scene for registration. We use the same ground truth generation as in the simulation experiments. The results averaged over all 100 scenes are recorded in Table II. An interesting observation is that TCAFF performs better on this dataset than it does in the simulation study, despite having a much longer runtime than the other methods. We hypothesize this is due to the lack of ambiguous data in 3RScan, in which case TCAFF becomes similar to a standard data association method. Langevin CLIPPER still performs the best overall,

TABLE I: Simulation Results

(Best in **green**; second-best in **yellow**)

	Metric	TCAFF-100	Stein ICP	Stein CLIPPER	Langevin CLIPPER
Object Map Registration					
Circle	Runtime (s)	0.04 ± 0.00	3.46 ± 0.22	1.77 ± 0.03	0.24 ± 0.01
	MMD (trans)	2.29 ± 0.15	90.81 ± 1.74	0.00 ± 0.00	0.00 ± 0.00
	MMD (rot)	0.19 ± 0.04	3.88 ± 0.22	1.19 ± 0.28	0.06 ± 0.02
	ED (trans)	6.49 ± 0.18	84.07 ± 3.05	0.00 ± 0.00	0.00 ± 0.00
	ED (rot)	0.59 ± 0.03	8.06 ± 0.24	2.12 ± 0.57	0.10 ± 0.04
	W1 (trans)	13.81 ± 0.41	152.74 ± 3.92	0.00 ± 0.00	0.00 ± 0.00
	W1 (rot)	12.23 ± 0.39	78.86 ± 1.38	23.07 ± 3.12	4.79 ± 0.78
Two Lines	Runtime (s)	0.97 ± 0.05	3.65 ± 0.21	3.10 ± 0.08	0.33 ± 0.02
	MMD (trans)	1.94 ± 0.06	1.73 ± 0.47	12.74 ± 5.37	1.63 ± 0.11
	MMD (rot)	7.58 ± 0.19	6.26 ± 0.15	4.68 ± 1.07	2.85 ± 0.14
	ED (trans)	7.33 ± 0.44	21.19 ± 3.28	33.50 ± 14.32	4.66 ± 0.38
	ED (rot)	5.88 ± 0.17	7.77 ± 0.25	4.24 ± 0.82	2.39 ± 0.10
	W1 (trans)	78.58 ± 1.23	117.49 ± 5.50	127.85 ± 21.88	62.94 ± 1.28
	W1 (rot)	52.46 ± 1.04	97.25 ± 0.86	41.70 ± 3.88	31.75 ± 0.86
Office	Runtime (s)	42.70 ± 0.88	3.70 ± 0.27	17.14 ± 0.30	1.19 ± 0.01
	MMD (trans)	33.13 ± 1.76	2.52 ± 0.13	2.10 ± 0.37	0.77 ± 0.09
	MMD (rot)	17.14 ± 1.24	3.16 ± 0.36	1.67 ± 0.66	0.57 ± 0.03
	ED (trans)	61.71 ± 2.54	25.10 ± 0.83	12.36 ± 7.59	1.48 ± 0.38
	ED (rot)	20.16 ± 2.29	4.52 ± 0.46	2.60 ± 1.44	0.63 ± 0.15
	W1 (trans)	167.05 ± 2.55	120.30 ± 1.62	92.50 ± 15.90	53.63 ± 2.01
	W1 (rot)	85.00 ± 3.78	52.21 ± 1.45	39.85 ± 6.54	28.99 ± 1.96
Point Cloud Registration					
Bowl	Runtime (s)	119.27 ± 4.00	4.10 ± 0.21	31.23 ± 0.81	0.83 ± 0.01
	MMD (trans)	8.66 ± 0.11	5.25 ± 0.52	10.12 ± 3.57	3.49 ± 0.25
	MMD (rot)	9.47 ± 0.31	0.88 ± 0.13	1.26 ± 0.28	0.83 ± 0.07
	ED (trans)	5.56 ± 0.12	2.64 ± 0.34	5.58 ± 1.87	1.74 ± 0.15
	ED (rot)	18.22 ± 0.93	4.74 ± 0.09	5.93 ± 1.52	1.38 ± 0.22
	W1 (trans)	21.25 ± 0.07	18.80 ± 0.37	22.60 ± 2.35	17.14 ± 0.25
	W1 (rot)	91.03 ± 0.91	55.03 ± 0.71	56.80 ± 3.78	43.05 ± 0.94
Plate	Runtime (s)	75.56 ± 1.74	4.15 ± 0.22	25.65 ± 1.62	0.69 ± 0.02
	MMD (trans)	9.45 ± 0.13	9.64 ± 0.56	19.93 ± 11.65	4.10 ± 0.44
	MMD (rot)	9.15 ± 0.28	0.85 ± 0.13	2.57 ± 0.59	0.81 ± 0.20
	ED (trans)	4.53 ± 0.08	4.55 ± 0.37	9.12 ± 5.47	1.81 ± 0.18
	ED (rot)	23.37 ± 1.17	0.47 ± 0.08	6.82 ± 3.17	1.61 ± 0.47
	W1 (trans)	20.74 ± 0.10	21.07 ± 0.27	28.89 ± 5.93	16.96 ± 0.32
	W1 (rot)	79.83 ± 1.28	27.33 ± 0.89	51.20 ± 7.99	32.98 ± 2.64
Pitcher	Runtime (s)	212.77 ± 3.51	4.35 ± 0.20	31.76 ± 0.15	1.04 ± 0.01
	MMD (trans)	6.95 ± 0.15	6.22 ± 0.75	55.54 ± 8.50	2.90 ± 0.25
	MMD (rot)	6.50 ± 0.34	1.20 ± 0.40	3.41 ± 0.38	0.51 ± 0.05
	ED (trans)	3.31 ± 0.11	2.85 ± 0.39	40.61 ± 5.81	1.22 ± 0.11
	ED (rot)	35.40 ± 0.85	2.84 ± 1.04	23.60 ± 3.69	3.15 ± 0.54
	W1 (trans)	20.72 ± 0.11	19.59 ± 0.57	48.18 ± 2.66	16.98 ± 0.22
	W1 (rot)	116.70 ± 0.93	54.09 ± 5.11	85.16 ± 4.24	47.89 ± 1.75

All values except runtimes are multiplied by 10^2 for readability.

demonstrating its effectiveness in general data association problems, regardless of whether the solution distribution is unimodal or multimodal.

B. Real World Object Map Registration

To demonstrate the importance of understanding uncertainty in ambiguous association problems, we apply our method to several real-world scenarios in which ROMAN [8], a state-of-the-art object map association method, commits to the wrong mode of the solution distribution. Initial object maps are built online using ROMAN, which processes RGBD images from a RealSense D455 stereo camera with Kimera-VIO [47] for odometry. We directly use the affinity matrix M provided by ROMAN, which incorporates gravity direction and object semantic similarity

TABLE II: 3RScan Dataset Results

Metric	TCAFF-100	Stein ICP	Stein CLIPPER	Langevin CLIPPER
Runtime (s)	13.42 ± 26.15	3.52 ± 0.15	2.61 ± 1.45	0.37 ± 0.13
MMD (trans)	2.45 ± 1.58	1.64 ± 1.59	3.09 ± 3.87	1.16 ± 0.98
MMD (rot)	2.37 ± 1.66	2.02 ± 2.46	2.66 ± 3.98	1.20 ± 1.32
ED (trans)	6.40 ± 6.63	7.63 ± 10.62	38.50 ± 54.70	4.12 ± 5.83
ED (rot)	4.11 ± 3.66	4.33 ± 5.45	9.80 ± 15.21	2.57 ± 2.99
W1 (trans)	63.52 ± 22.55	68.24 ± 30.77	104.40 ± 63.26	51.08 ± 23.24
W1 (rot)	58.86 ± 9.22	59.78 ± 15.08	65.28 ± 20.65	46.01 ± 9.46

All values except runtimes are multiplied by 10^2 for readability.

priors, but fails to uniquely disambiguate between repetitive objects. Data is recorded from the three ambiguous environments shown in Fig. 1: a staircase, a line of trees, and a circular library room. We task each algorithm with registering two object maps: one complete reference map and a smaller second map (e.g., a single floor of the staircase). For each experiment, we investigate an ambiguous axis of the estimated pose. We visualize the estimated pose distribution found by the proposed methods, the pose found by ROMAN, and a manually annotated ground truth pose for qualitative comparison. We find that Langevin CLIPPER successfully represents the multimodal distribution while Stein CLIPPER often collapses to a few modes, likely due to its sensitivity to the kernel bandwidth.

Stairs Significant elevation (z-axis) ambiguity is present in this example, which is shown in Fig. 1 (left) with distributions generated by proposed methods visualized in Fig. 7 (left). Stein CLIPPER fails to capture the full distribution and ROMAN commits to an incorrect mode. Meanwhile, Langevin CLIPPER recovers a peak at elevations corresponding to each floor of the stairway, capturing the uncertainty of the solution.

Trees In this example, shown in Fig. 1 (middle), a single-file line of tall trees is mapped, inducing ambiguity in the pose estimate's x-axis. Stein CLIPPER and Langevin CLIPPER both capture many potential modes at regular intervals, although Langevin CLIPPER finds more modes.

Library Our final example, shown in Fig. 1 (right), exhibits yaw ambiguity. The library room is circular, with four doors and columns between each door. Stein CLIPPER and ROMAN both commit to the wrong mode, while Langevin CLIPPER finds the four modes at -90, 0, 90, and 180 deg.

C. Enumerating Maximal Cliques

We set up a small example to investigate how well the proposed methods match the theoretical result discussed in Remark 1, where enumerating all maximal cliques is possible using the BK [35] algorithm. The setup of the experiment is shown in Fig. 2 (bottom middle), where two object maps are generated from two different views of the triangular arrangement of bushes. We run BK to generate all the maximal cliques of the consistency graph, by rounding edge weights above 0.5 to 1 and others to 0. The result is shown in Fig. 8. We can see that both proposed methods match BK well, although the resulting distribution is less concentrated around the modes (i.e., smaller density), which is expected since the theoretical analysis is asymptotic.

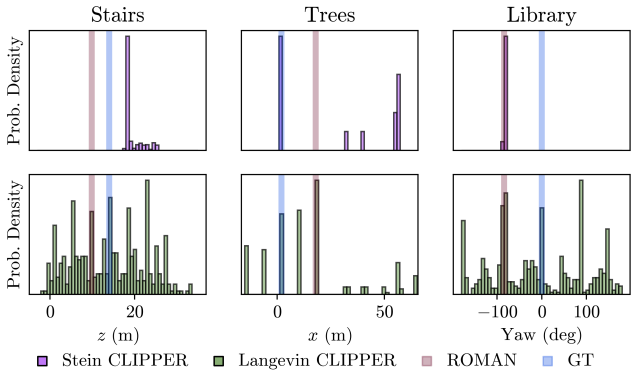

Fig. 7: Pose distributions from Stein CLIPPER and Langevin CLIPPER for the staircase, trees, and library shown in Fig. 1. For each case, the axis containing repetitive geometry is shown. Langevin CLIPPER successfully finds modes at different floors of the staircase, at different trees in the line, and around the four quadrants of the library. The single estimate found by ROMAN and the ground truth (GT) pose are also shown.

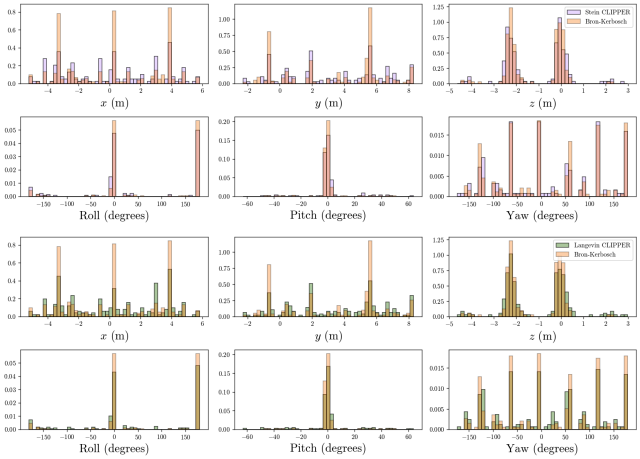

Fig. 8: Visualization of the theoretical distribution (Bron–Kerbosch) versus the actual distribution generated by the proposed methods.

VII. Conclusion

In this work, we proposed and investigated two methods for capturing multimodal distributions of solutions to global data association problems, leveraging approximate Bayesian inference techniques. Future work includes integrating the resulting multimodal pose estimate distributions into loop closures in a multimodal SLAM backend [17], [48] for improved downstream trajectory estimates. A current limitation is runtime, which must be improved to support online multi-robot SLAM.

References

[1] M. Strecke and J. Stuckler, "Em-fusion: Dynamic object-level SLAM with probabilistic data association," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5865–5874.

[2] M. B. Peterson, P. C. Lusk, A. Avila, and J. P. How, "TCAFF: Temporal consistency for robot frame alignment," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 15 821–15 827.

[3] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 2003.

[4] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, "Probabilistic data association for semantic SLAM," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 1722–1729.

[5] K. J. Doherty, D. P. Baxter, E. Schneeweiss, and J. J. Leonard, "Probabilistic data association via mixture models for robust semantic SLAM," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1098–1104.

[6] Y. Tian, Y. Chang, F. H. Arias, C. Nieto-Granda, J. P. How, and L. Carlone, "Kimera-multi: Robust, distributed, dense metric-semantic slam for multi-robot systems," *IEEE Transactions on Robotics*, vol. 38, no. 4, 2022.

[7] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3D point cloud map," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4802–4809.

[8] M. B. Peterson, Y. X. Jia, Y. Tian, A. Thomas, and J. P. How, "RO-MAN: Open-Set Object Map Alignment for Robust View-Invariant Global Localization," in *Robotics: Science and Systems (RSS)*, 2025.

[9] Y. Chang, N. Hughes, A. Ray, and L. Carlone, "Hydra-multi: Collaborative online construction of 3D scene graphs with multi-robot teams," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 10 995–11 002.

[10] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[11] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 314–333, 2020.

[12] J. G. Mangelson, D. Dominic, R. M. Eustice, and R. Vasudevan, "Pairwise consistent measurement set maximization for robust multi-robot map merging," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 2916–2923.

[13] N. Sünderhauf and P. Protzel, "Switchable constraints for robust pose graph SLAM," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 1879–1884.

[14] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Robust pose-graph loop-closures with expectation-maximization," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 556–563.

[15] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1127–1134, 2020.

[16] B. Mu, S.-Y. Liu, L. Paull, J. Leonard, and J. P. How, "SLAM with objects using a nonparametric pose graph," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4602–4609.

[17] M. Hsiao and M. Kaess, "Mh-iSAM2: Multi-hypothesis iSAM using Bayes tree and hypo-tree," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1274–1280.

[18] F. A. Maken, F. Ramos, and L. Ott, "Estimating motion uncertainty with Bayesian ICP," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 8602–8608.

[19] ——, "Stein ICP for uncertainty estimation in point cloud matching," *IEEE robotics and automation letters*, vol. 7, no. 2, pp. 1063–1070, 2021.

[20] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient langevin dynamics," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 681–688.

[21] F. A. Maken, F. Ramos, and L. Ott, "Stein particle filter for nonlinear, non-gaussian state estimation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5421–5428, 2022.

[22] K. Koide, S. Oishi, M. Yokozuka, and A. Banno, "MegaParticles: Range-based 6-DoF Monte Carlo Localization with GPU-Accelerated Stein Particle Filter," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 1738–1744.

[23] A. Lambert, A. Fishman, D. Fox, B. Boots, and F. Ramos, "Stein variational model predictive control," *arXiv preprint arXiv:2011.07641*, 2020.

[24] S. B. Gelfand and S. K. Mitter, "Recursive stochastic algorithms for global optimization in R^d," *SIAM Journal on Control and Optimization*, vol. 29, no. 5, pp. 999–1018, 1991.

[25] G. O. Roberts and R. L. Tweedie, "Exponential convergence of Langevin distributions and their discrete approximations," 1996.

[26] P. C. Lusk, K. Fathian, and J. P. How, "CLIPPER: A graph-theoretic framework for robust data association," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 828–13 834.

[27] M. T. Belachew and N. Gillis, "Solving the maximum clique problem with symmetric rank-one non-negative matrix approximation," *Journal of Optimization Theory and Applications*, vol. 173, no. 1, pp. 279–296, 2017.

[28] Q. Liu, "Stein variational gradient descent as gradient flow," *Advances in neural information processing systems*, vol. 30, 2017.

[29] Q. Liu and D. Wang, "Stein variational gradient descent: A general purpose bayesian inference algorithm," *Advances in neural information processing systems*, vol. 29, 2016.

[30] M. Toussaint, "Robot trajectory optimization using approximate inference," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 1049–1056.

[31] M. Okada and T. Taniguchi, "Variational inference mpc for bayesian model-based reinforcement learning," in *Conference on robot learning*. PMLR, 2020, pp. 258–272.

[32] J. R. Munkres, *Analysis on manifolds*. CRC Press, 2018.

[33] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American mathematical society*, vol. 68, no. 3, pp. 337–404, 1950.

[34] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.

[35] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Communications of the ACM*, vol. 16, no. 9, pp. 575–577, 1973.

[36] D. Aiger, N. J. Mitra, and D. Cohen-Or, "4-points congruent sets for robust pairwise surface registration," in *ACM SIGGRAPH 2008 papers*, 2008, pp. 1–10.

[37] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[38] M. L. Rizzo and G. J. Székely, "Energy distance," *wiley interdisciplinary reviews: Computational statistics*, vol. 8, no. 1, pp. 27–38, 2016.

[39] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *The journal of machine learning research*, vol. 13, no. 1, pp. 723–773, 2012.

[40] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," *Advances in neural information processing systems*, vol. 26, 2013.

[41] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation averaging," *International journal of computer vision*, vol. 103, no. 3, pp. 267–305, 2013.

[42] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[43] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1912–1920.

[44] L. Downs, A. Francis, N. Koenig, B. Kinman, R. Hickman, K. Reymann, T. B. McHugh, and V. Vanhoucke, "Google scanned objects: A high-quality dataset of 3D scanned household items," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2553–2560.

[45] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 3212–3217.

[46] J. Wald, A. Avetisyan, N. Navab, F. Tombari, and M. Niessner, "Rio: 3D object instance re-localization in changing indoor environments," in *Proceedings IEEE International Conference on Computer Vision (ICCV)*, 2019.

[47] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: an open-source library for real-time metric-semantic localization and mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1689–1696.

[48] K. J. Doherty, Z. Lu, K. Singh, and J. J. Leonard, "Discrete-continuous smoothing and mapping," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 395–12 402, 2022.