

Outlier-Robust Geometric Perception: A Novel Thresholding-Based Estimator with Intra-Class Variance Maximization

Lei Sun¹

Abstract—Geometric perception problems are fundamental tasks in robotics and computer vision. In real-world applications, they often encounter the inevitable issue of outliers, preventing traditional algorithms from making correct estimates. In this paper, we present a novel general-purpose robust estimator TIVM (Thresholding with Intra-class Variance Maximization) that can collaborate with standard non-minimal solvers to efficiently reject outliers for geometric perception problems. First, we introduce the technique of intra-class variance maximization to design a dynamic 2-group thresholding method on the measurement residuals, aiming to distinctively separate inliers from outliers. Then, we develop an iterative framework that robustly optimizes the model by approaching the pure-inlier group using a multi-layered dynamic thresholding strategy as subroutine, in which a self-adaptive mechanism for layer-number tuning is further employed to minimize the user-defined parameters. We validate the proposed estimator on 3 classic geometric perception problems: rotation averaging, point cloud registration and category-level perception, and experiments show that it is robust against 70–90% of outliers and can converge typically in only 3–15 iterations, much faster than state-of-the-art robust solvers such as RANSAC, GNC and ADAPT. Furthermore, another highlight is that: our estimator can retain approximately the same level of robustness even when the inlier-noise statistics of the problem are fully unknown.

I. INTRODUCTION

Geometric spatial perception problems, which aim to estimate the geometric models regarding the state of a robot or a sensor (e.g. camera) and its spatial relationship to the environment [1], are the building blocks in robotics and computer vision, having found broad applications in motion estimation [2], object/camera localization [3], [4], 3D reconstruction [5], [6], autonomous navigation and SLAM [7].

Unfortunately, in realistic applications, geometric perception problems are often not only affected by noisy data, but also corrupted by the outliers, which are spurious measurements typically caused by wrong data association/matches, extreme noise or device malfunction, making traditional estimation methods fail to yield correct results.

In terms of handling outliers, RANSAC [8] is the most classic robust heuristic widely applied in many perception problems to remove outliers by incorporating the minimal solver into a hypothesize-and-check paradigm. Though some variants (e.g. [9]–[11]) are proposed to improve the performance, RANSAC-based estimators usually suffer from exponentially increasing computational cost w.r.t. the outlier ratio as well as the minimal subset size, inefficient for use in

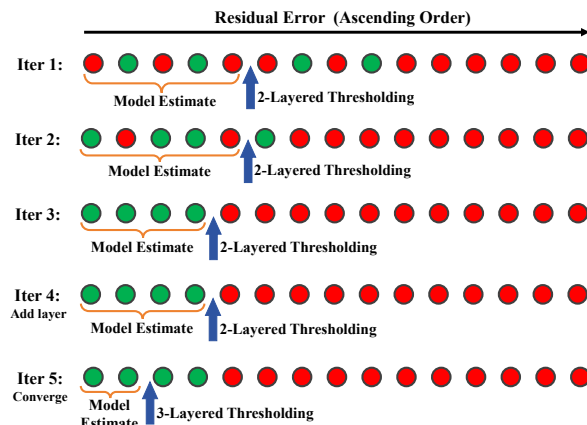


Fig. 1. Intuitive illustration on the thresholding framework of the proposed TIVM estimator in Algorithm 1.

many real-world tasks. Moreover, minimal solvers may not fit with all the problems, as will be discussed in Section II-A.

More recently, global non-minimal robust estimators such as the GNC framework [12]–[14] and ADAPT [1] are receiving increasing attention, since they can reject outliers directly in conjunction with standard non-minimal solvers that have been rapidly developed these years for multiple perception problems (e.g. [15] for 2D-3D shape reconstruction, [16] for category-level perception, and [17] for pose graph optimization) without the need of initial guess. However, although they can reject 70–80% outliers, they are generally too slow to converge, requiring at least dozens of, or even over a hundred, iterations to attain convergence. Since most non-minimal solvers are not as fast as the minimal ones, using these non-minimal robust estimators could result in long computational time.

In addition, another challenge in the geometric perception problems is that: a user-defined inlier threshold is needed to differentiate inliers from outliers during robust estimation, but determining such inlier threshold can be difficult since it usually requires users to manually collect sufficient data to estimate the noise statistics or complete an arduous work of parameter tuning. Besides, the inlier-noise statistics may vary during actual tasks (e.g. due to change of weather, long-time usage), so the original calibration could become inaccurate.

Contributions. To address these limitations, this paper proposes a novel general-purpose non-minimal robust estimation method TIVM (Thresholding with Intra-class Variance Maximization) for geometric perception, as intuitively illustrated in Fig. 1. First, we present a dynamic residual thresholding method based on intra-class variance maximization seeking to flexibly classify the measurements into a low-

*This work was not supported by any organization

¹Lei Sun is with School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, USA (e-mail: leisunjames@126.com)

magnitude group (approaching inliers) and a high-magnitude group (approaching outliers). Second, an iterative model-optimizing framework is developed upon a multi-layered dynamic thresholding method to guarantee high robustness. Moreover, a self-adaptive strategy to tune the layer number is invented to enable automatic algorithm operation without introducing additional user-defined parameters. More importantly, the resulting estimator TIVM could achieve robust estimation even without any prior knowledge of the inlier-noise statistics. We test TIVM on the rotation averaging, point cloud registration, and category-level perception problems. TIVM demonstrates 70–90% robustness against outliers no matter the noise-statistics are known or unknown, and it could converge in merely 3–15 iterations, which is significantly faster than other state-of-the-art competitors including RANSAC, GNC and ADAPT.

II. RELATED WORK

A. Naive Estimation Methods without Outliers

Minimal Solvers. Minimal solvers can make estimates with the fewest measurements possible, and its typical examples include: Horn’s triad-based method [18] for point cloud registration, and the 5-point solver [19] for two-view geometry. However, due to the lack of redundant measurements to refine the solution, minimal solvers may be easily affected by noise (e.g. as discussed in [9]). Moreover, some problems are unsuitable to be solved by minimal solvers because they have too many variables to solve (e.g. category-level perception that involves solving multiple shape parameters) and this will make the dimension (number of measurements required) of the minimal solvers too high and hence tremendously slowing down the convergence when running with RANSAC.

Non-minimal Solvers. Non-minimal solvers takes an arbitrary number of measurements as inputs (on the condition that the problem is overdetermined) to make the optimal estimation by solving least-squares optimization under the assumption of Gaussian noise. Typical examples include: Arun’s SVD method [20] for point cloud registration, Semi-Definite relaxations for 3D registration [21] and category-level perception [16], as well as Sum-of-Squares relaxations for 2D-3D shape alignment & reconstruction [13], [15]. Though non-minimal solvers can minimize the effect of noise and yield optimal estimates, they usually require relatively long runtime compared to minimal solvers (e.g. due to the relaxation methods on high-degree polynomial optimization problems). Thus, once combined with the outlier rejection frameworks (e.g. GNC, ADAPT) where plenty of iterations are needed, the time-efficiency issue appears critical.

B. Outlier-Robust Estimation Methods

Consensus Maximization. It seeks to find the model that corresponds to the largest number of measurements with residual errors below the inlier threshold in order to achieve robust estimation. RANSAC [8] is the most popular consensus maximization method that randomly samples small measurement subsets to make minimal model estimates and finds the best model enabling the largest measurement consensus

set. Additional techniques such as local optimization [9], [33] or measurement ranking [10] have been applied to improve RANSAC. But RANSAC methods’ natural exponential runtime with the growth of outlier ratio and problem dimension confines their usage in many practical situations. Branch-and-Bound [22] is another maximum consensus framework that returns optimal results by searching in the parameter space, but it also suffers from the worst-case exponential time w.r.t. problem size. Also, ADAPT [1] can reject outliers by alternating between non-minimal estimation and measurement trimming with a gradually decreasing residual threshold. But it often requires dozens to hundreds of iterations to converge, and considering the slow runtime of some non-minimal solvers, time-efficiency is still a notable issue.

M-Estimation. M-estimation employs robust cost functions to diminish the weights of outliers. Earlier local M-estimation methods [23] usually require an initial guess and then conduct iterative optimization for convergence, but they are prone to get stuck in local minimum if the initialization is bad. Graduated Non-Convexity (GNC) is presented to avoid the need of the initial guess as a global estimator, which is first used in [12] for robust point cloud registration. Then, GNC is promoted as a general-purpose estimator for diverse perception problems [13] and its accelerated version GNC-IRLS [14] is also supplemented more currently. But GNC typically needs 30–50 iterations to converge while GNC-IRLS also requires 7-20+ iterations in our experiments, thus confronted with the same runtime issue as ADAPT.

III. PRELIMINARIES FOR GEOMETRIC PERCEPTION

We first provide a general definition for geometric spatial perception problems as in [13]. For a geometric estimation problem where its model (a.k.a. variables to solve) is represented as $\mathbf{x} \in \mathcal{X}$ (here \mathcal{X} is the feasible domain), its measurements are denoted as $\mathbf{m}_i, i \in \mathcal{N} = \{1, 2, \dots, N\}$, and the residual function that measures the difference (non-negative error) between the measurements computed with current model \mathbf{x} and the actual measurements \mathbf{m}_i is formulated as $Re(\mathbf{m}_i, \mathbf{x})$ (also abbreviated as Re_i). When there exists no outlier in the measurements, this problem can be optimally solved by the following minimization formulation:

$$\min_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^N Re^2(\mathbf{m}_i, \mathbf{x}). \quad (1)$$

For example, if the estimation problem is point cloud registration, \mathbf{x} should denote the rigid transformation, \mathcal{X} should be $SE(3)$, and \mathbf{m}_i would be the putative correspondences matched between point clouds.

However, in reality, measurements \mathbf{m}_i are often corrupted by outliers, so we need a outlier-robust formulation for this estimation problem which can be represented as a consensus maximization problem ([1], [24]) such that

$$\begin{aligned} & \max_{\mathcal{I} \subset \mathcal{N}} |\mathcal{I}|, \\ & s.t. |Re(\mathbf{m}_i, \mathbf{x}^*)| \leq \gamma (\forall i \in \mathcal{I}), \end{aligned} \quad (2)$$

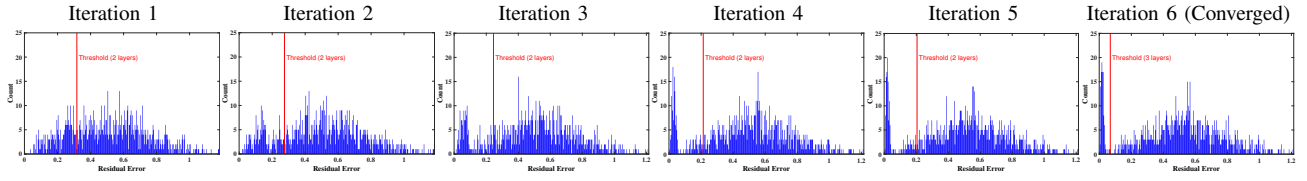


Fig. 2. Illustration of the residual-error histograms with labeled thresholds in our TIVM algorithm at different iterations.

where the goal is to find the optimal model \mathbf{x}^* that enables as many measurements \mathbf{m}_i as possible to satisfy the condition that their residual errors Re_i are lower than the user-defined inlier threshold γ (also known as the noise bound) that serves as the criterion to differentiate inliers from outliers.

For robust perception problems without inlier-noise statistics, threshold τ would be unknown, which greatly increases the difficulty of robust estimation; but fortunately, we manage to design a novel paradigm that uses maximum intra-class variance to achieve noise-statistics-free robust estimation, as rendered in the following section.

IV. METHODOLOGY

Motivation. As discussed in [1], for the optimal model \mathbf{x}^* , its measurement residuals Re_i should naturally form 2 relatively distinct groups, one with low magnitude (as inliers) and the other with high magnitude (as outliers). Thus, our idea is to design an iterative dynamic thresholding framework to gradually approximate the optimal model by achieving 2 groups of measurements with the maximum intra-group variance on their residual errors in each iteration. And thanks to this framework, one huge advantage of the proposed algorithm is that: even when the inlier-noise statistics are not given, it could still return robust (reasonably accurate) estimation results.

We provide the pseudocode: Algorithm 1 for the noise-statistics-free scenario (without known inlier threshold τ) in advance, for the convenience of the subsequent elaboration.

A. Intra-Class Variance Maximization for Dynamic Residual Thresholding

Intra-class variance maximization is an efficacious approach to classify a set of data into multiple groups in a way that the largest possible variance between classes is achieved, and its most popular application is in Otsu's image thresholding method [25]. In this work, we employ this technique to accomplish dynamic thresholding on the residual errors, trying to distinguish inliers from outliers.

A prerequisite for this technique is to create a series of consecutive *intervals* to encompass all the residual errors so as to construct a 'residual histogram', as exemplified in Fig. 2. Given that all residual errors are positive, we set the maximum residual error as the upper-bound of all intervals such that $D^{max} = \max(Re_i), i \in \mathcal{N}$, and then divide D^{max} into L consecutive small intervals where each interval has the same length of $\Delta D = \frac{D^{max}}{L}$ (typically, we set $L = 300$ in practice). Now there exist L intervals given by: $(0, \Delta D], (\Delta D, 2\Delta D], (2\Delta D, 3\Delta D], \dots, (D^{max} - \Delta D, D^{max}]$.

Then, we let n_l ($l = 1, 2, \dots, L$) denote the number of residual errors that fall into the l_{th} interval, so apparently

$\sum_{l=1}^L n_l = N$. Based on the procedures derived in [25], we normalize n_l as $p_l = \frac{n_l}{N}$, so the probability for a residual error to be lower than $k \cdot \Delta D$ can be represented as:

$$P_k = \sum_{l=1}^k p_l, \quad (3)$$

where k is an integer ranging from 1 to L . The cumulative mean of residual errors lower than $k \cdot \Delta D$ can be given by:

$$\mu_k = \sum_{l=1}^k l \cdot p_l. \quad (4)$$

Letting $k = L$, we can have the mean of all residual errors:

$$\bar{\mu} = \mu_L = \sum_{l=1}^L l \cdot p_l. \quad (5)$$

Up to now, the intra-class variance for a certain k (where $k = 1, 2, \dots, L$) can be computed as:

$$\sigma_k = \frac{(\bar{\mu}P_k - \mu_k)^2}{P_k(1 - P_k)}. \quad (6)$$

As a result, the best threshold T^* can be easily solved by:

$$T^* = \Delta D \cdot k^* = \Delta D \cdot \arg \max_{k=1,2,\dots,L} (\sigma_k). \quad (7)$$

With the best T^* , we can dynamically separate all the measurements into 2 groups based on their residual-error histogram as: a lower-magnitude group \mathcal{G}^α with residual errors no larger than T^* ($\forall i \in \mathcal{G}^\alpha, Re_i \leq T^*$), and a higher-magnitude group \mathcal{G}^β with residual errors larger than T^* ($\forall i \in \mathcal{G}^\beta, Re_i > T^*$).

However, when outliers are prevalent among measurements (e.g. $> 50\%$), grouping the measurements based on the routine above may still include outliers into the lower-magnitude group \mathcal{G}^α , hindering us from correctly sifting out the true inliers. Therefore, we need to further perform multiple layers of dynamic thresholding to enhance the inlier confidence of \mathcal{G}^α .

B. Multi-Layered Thresholding for Outlier Rejection

Once we complete the first thresholding over all measurements with threshold T_1^* to obtain group \mathcal{G}_1^α (where subscript j denotes that \mathcal{G}_j^α and T_j^* are obtained in the j_{st} layer of thresholding), we can conduct another layer of dynamic thresholding on \mathcal{G}_1^α , that is, treating all measurements in the current lower-magnitude group \mathcal{G}_1^α as the input and applying (3)–(7) again to further divide \mathcal{G}_1^α into 2 groups: \mathcal{G}_2^α and \mathcal{G}_2^β with the new threshold T_2^* estimated from (7). Thus, assuming the residual errors are computed over the optimal model \mathbf{x}^* , if we conduct m layers of such 2-group

Algorithm 1: TIVM^o (Noise-statistics-free Version)

Input: measurements $\{m_i\}_{i \in \mathcal{N}}$;
1 Set $\mathcal{C}^0 \leftarrow \mathcal{N}$, $L \leftarrow 300$, $m \leftarrow 2$ and $t_{max} \leftarrow 100$;
2 **for** $t = 1 : t_{max}$ **do**
3 Solve model x^t using the non-minimal solver with measurements in set \mathcal{C}^{t-1} ;
4 Compute the residual errors Re_i w.r.t. all the measurements m_i ($i \in \mathcal{N}$) using x^t ;
5 $D^{max} \leftarrow \max(Re_i)$ and $\Delta D \leftarrow \frac{D^{max}}{L}$;
6 Set L intervals and count n_l ($l \in \{1, 2, \dots, L\}$);
7 $\mathcal{G}_0^\alpha \leftarrow \mathcal{N}$ and $T_0^* \leftarrow L$;
8 **for** $j=1:m$ **do**
9 $\forall l \in \{1, 2, \dots, \frac{T_j^*}{\Delta D} - 1\}$, compute $p_l \leftarrow \frac{n_l}{|\mathcal{G}_{j-1}^\alpha|}$;
10 $\forall k \in \{1, 2, \dots, \frac{T_j^*}{\Delta D}\}$, update P_k , μ_k , $\bar{\mu}$ and σ_k according to (3)–(6);
11 Solve threshold T_j^* based on (7) to separate \mathcal{G}_{j-1}^α into 2 groups: \mathcal{G}_j^α and \mathcal{G}_j^β ;
12 **end**
13 **if** $checkConv$ and $\frac{|\bar{Re} - \frac{1}{N} \sum_{i=1}^N Re_i|}{\bar{Re}} \leq 10^{-3}$ **then**
14 **break**
15 **end**
16 $checkConv \leftarrow false$, $T^t \leftarrow T_m^*$ and $\mathcal{C}^t \leftarrow \mathcal{G}_m^\alpha$;
17 **if** $|T^t - T^{t-1}| \leq \Delta D$ **then**
18 $m \leftarrow m + 1$ and $checkConv \leftarrow true$;
19 $\bar{Re} = \frac{1}{N} \sum_{i=1}^N Re_i$;
20 **end**
21 **end**
22 **return** optimal model $x^* \leftarrow x^t$ and inlier set \mathcal{C}^t ;

thresholding where m is adequate, the final low-magnitude group in the last layer, denoted as \mathcal{G}_m^α , should be a pure-inlier (outlier-free) group. (This multi-layered thresholding process corresponds to *Lines 5–12* in Algorithm 1.)

But unfortunately, the model we use for estimation could not be the optimal model since \mathcal{G}_m^α is not supposed to contain pure inliers. Thus, in this circumstance, we introduce an iterative optimizing framework where in each iteration, the measurements in \mathcal{G}_m^α are repeatedly updated and adopted for non-minimal estimation and residual computing in the next iteration. In this framework, model x and group \mathcal{G}_m^α are gradually optimized to approach the optimal (pure-inlier) solutions via iterations of the multi-layered dynamic thresholding procedure as described above. (This process corresponds to *Lines 2–4 & 21* in Algorithm 1.)

C. Self-Adaptive Layer Number Tuning

Nonetheless, there exists a challenge that: it is hard for the users to preset the appropriate layer number m for dynamic thresholding. If m is too small, outliers may be included in \mathcal{G}_m^α ; if m is too large, the measurement samples for non-minimal estimation would be too sparse, which may compromise the estimation accuracy.

To resolve this issue, we design a self-adaptive mechanism for automatically tuning the layer number m . Specifically, we first start dynamic thresholding with a small value of m

Algorithm 2: TIVM (Noise-statistics Version)

Input: measurements $\{m_i\}_{i \in \mathcal{N}}$, inlier threshold τ ;
1 (*Lines 1–15* in Algorithm 1)
2 **if** $T_m^* \leq 2 \cdot \tau$ **then**
3 **break**
4 **end**
5 (*Lines 16–21* in Algorithm 1)
6 Find all the measurements whose residual errors satisfy: $Re_i \leq \tau$ and add them to the inlier set \mathcal{C}^* ;
7 Solve optimal model x^* with measurements in \mathcal{C}^* ;
8 **return** optimal model x^* and full inlier set \mathcal{C}^* ;

(e.g. $m = 2$ in practice), and when the iterative optimizing framework converges, we would examine if the current group \mathcal{G}_m^α is good enough. This could be done by checking whether the model estimated by the converged \mathcal{G}_m^α set is similar enough to the model estimated by the \mathcal{G}_{m+1}^α after one more iteration with an incremented layer number: $m + 1$. If these 2 models have close enough average residual errors (which indicates the models are similar enough), then the current converged \mathcal{G}_m^α should be a pure-inlier group. But if not, meaning that the current layer number is insufficient for separating the pure inliers, then the iteration should continue with a new incremented layer number: $m \leftarrow m + 1$. (This mechanism corresponds to *Lines 13–20* in Algorithm 1.)

So far, Algorithm 1 provides the full algorithm (that we name TIVM^o) for robust estimation in the noise-statistics-free scenario. An example illustration of the residual-error histograms at all the iteration of the proposed algorithm is shown in Fig. 2, where we can observe that: as the iteration goes on, the distribution of the residual errors changes from 1 cluster with hardly any clear separation boundary into 2 distinct groups including: an inlier group with low residual magnitude and low variance (since inliers are consistent) plus an outlier group with high residual magnitude and high variance (since outliers are random).

D. TIVM with Noise-Statistics Information

In some cases, the inlier-noise statistics information is known by the user, so we could further boost the estimation accuracy with the aid of the given inlier threshold τ , and the new version is shown in Algorithm 2 (that we name TIVM). The main structure of the algorithm remains unchanged. The difference is that: we could add another convergence condition that: when the current threshold T_m^* is smaller than $2 \cdot \tau$ (indicating that the current model is good enough to differentiate inliers from outliers in 2 sufficiently distinct clusters), we should stop the iterations and then estimate the optimal model x^* directly using the measurements with residual errors lower than τ .

Note that Algorithm 2 only serves to lower the estimation errors by finding the full inlier set using the ground-truth inlier-noise threshold τ , but theoretically it cannot change the outlier-robustness because it is determined by whether the final low-magnitude group \mathcal{G}_m^α contains outliers (which completely relies on Algorithm 1).

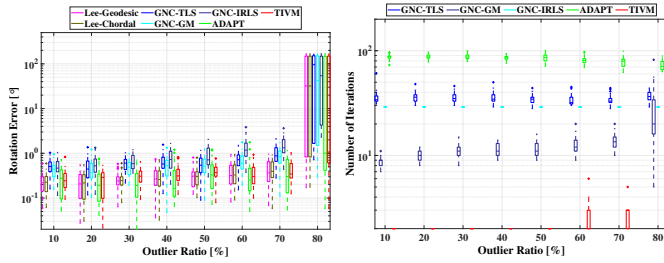


Fig. 3. Benchmarking on robust rotation averaging.

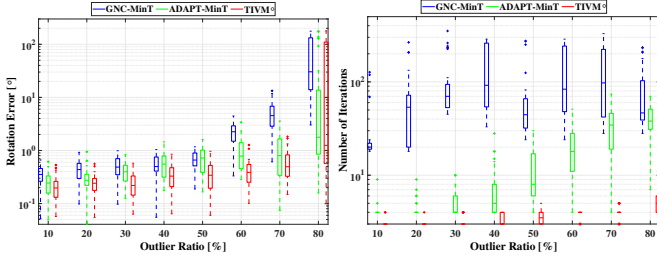


Fig. 4. Benchmarking on robust rotation averaging without inlier-noise statistics.

V. EXPERIMENTS

We evaluate our estimator TIVM (and the noise-statistics-free version TIVM^o) in 3 classic geometric perception problems: rotation averaging, point cloud registration and category-level perception. In each problem, we benchmark our estimator against state-of-the-art general-purpose robust solvers including: (i) GNC-TLS and GNC-GM [13] with control parameter $\mu = 1.4$ (according to [13]) and 100 maximum iterations, (ii) GNC-IRLS [14] with all parameters set as in source code of [13] with 100 maximum iterations, (iii) ADAPT [1] where the discount ratio is set to 0.99 as suggested in [1] with 200 maximum iterations, and (iv) other problem-specific specialized solvers. In addition, we test with unknown inlier-noise statistics where inlier threshold τ is unavailable and TIVM^o is compared against the state-of-the-art minimally-tuned solvers: GNC-MinT and ADAPT-MinT in [26] with 1000 maximum iterations. Experiments are conducted in Matlab on a laptop having an i9-12900H CPU and 32GB RAM with single thread.

A. Rotation Averaging

Robust rotation averaging aims to find the best rotation $\mathbf{R}^* \in SO(3)$ from a series of noisy rotation estimates w.r.t. the ground-truth value: $\mathbf{R}_i \in SO(3)$, $i = 1, 2, \dots, N$, which are potentially corrupted by outliers. The residual error can be defined as: $Re_i = G(\mathbf{R}_i, \mathbf{R}^*)$, where ‘ $G(\cdot, \cdot)$ ’ represents the geodesic distance [27] between rotations.

Setup. In each run, we obtain a random rotation $\mathbf{R}_{gt} \in SO(3)$ as the ground truth and then generate $N = 100$ noisy rotations \mathbf{R}_i as the input measurements such that $\mathbf{R}_i = \mathbf{R}_{gt} \text{Exp}(\eta_i \mathbf{e}_i)$ where η_i is a random isotropic Gaussian noise angle with standard deviation of $\sigma = 5^\circ$, \mathbf{e}_i is a random unit-norm 3D vector and $\text{Exp}(\cdot)$ is the exponential map. To create outliers, we replace 10–80% of the rotations in $\{\mathbf{R}_i\}_{i=1}^N$ with random rotations. For specialized solvers in benchmarking, we adopt the state-of-the-art Chordal-median and Geodesic-median methods [28]. The inlier threshold is set as $\tau = 3\sigma$.

In GNC, we use the weighted L1-chordal mean for weighted non-minimal rotation estimation; in ADAPT and TIVM, we use the L1-chordal median [28] as the non-minimal solver. The benchmarking results are based on 30 Monte Carlo runs, as shown in Fig. 3.

From Fig. 3, we see that: (1) all the solvers tolerate 70% outliers but fail at 80% outliers, (2) among general-purpose solvers, our TIVM is the most efficient solver that requires fewer than 10 iterations, GNC-IRLS is the second fastest solver mostly with 7–20 iterations, and GNC-TLS, GNC-GM and ADAPT need 30–100 iterations, and (3) TIVM has higher estimation accuracy than the 3 GNC solvers.

Unknown Noise-Statistics Tests. We perform similar tests without given inlier-noise statistics, as shown in Fig. 4. TIVM^o can still tolerate up to 70% outliers (with errors lower than 2°), shows higher accuracy than GNC-MinT and ADAPT-MinT and converges typically within 10 iterations (significantly faster than GNC-MinT and ADAPT-MinT).

B. Point Cloud Registration

Robust point cloud registration solves the optimal rigid transformation: $\mathbf{R}^* \in SO(3)$ and $\mathbf{t}^* \in \mathbb{R}^3$ that best align two 3D point clouds from putative correspondences $\{\mathbf{p}_i \leftrightarrow \mathbf{q}_i\}_{i=1}^N$ (usually established by feature matching techniques, e.g. [29]) which are corrupted by outliers. The residual error can be denoted with L2-norm as: $Re_i = \|\mathbf{R}^* \mathbf{p}_i + \mathbf{t}^* - \mathbf{q}_i\|_2$.

Setup. We adopt the *bunny* from Stanford 3D Scan Repository [30]. We first downsample it to $N = 1000$ points as the source point cloud: $\{\mathbf{p}_i\}_{i=1}^N$ and then resize it to fit in a $1 \times 1 \times 1m$ cube. In each run, we generate a random ground-truth transformation: $(\mathbf{R}_{gt} \in SO(3), \mathbf{t}_{gt} \in \mathbb{R}^3)$ to transform $\{\mathbf{p}_i\}_{i=1}^N$, and then add random Gaussian noise with standard deviation $\sigma = 0.01$ to obtain the target point cloud $\{\mathbf{q}_i\}_{i=1}^N$. Subsequently, a portion of the points in $\{\mathbf{q}_i\}_{i=1}^N$ are replaced with random 3D points near the point cloud surface to generate outliers. This setup is illustrated in Fig. 5. In terms of the specialized solvers, we use FGR [12], GORE [31], TEASER [32] and TriVoC [24]. We also include RANSAC [8] and FLO-RANSAC [33] (both with 500 maximum iterations and $p = 0.99$ confidence) as 2 state-of-the-art robust estimators for comparison. The inlier threshold is set to $\tau = 5\sigma$. We adopt Horn’s method [18] as the minimal solver for RANSAC and FLO-RANSAC, and Arun’s SVD approach [20] as the non-minimal solver. Benchmarking results over 30 Monte Carlo runs are reported in Fig. 6.

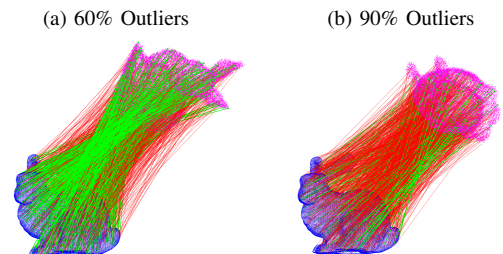


Fig. 5. Demonstration of the experimental setup of robust point cloud registration. Green and red lines denote inliers and outliers, respectively.

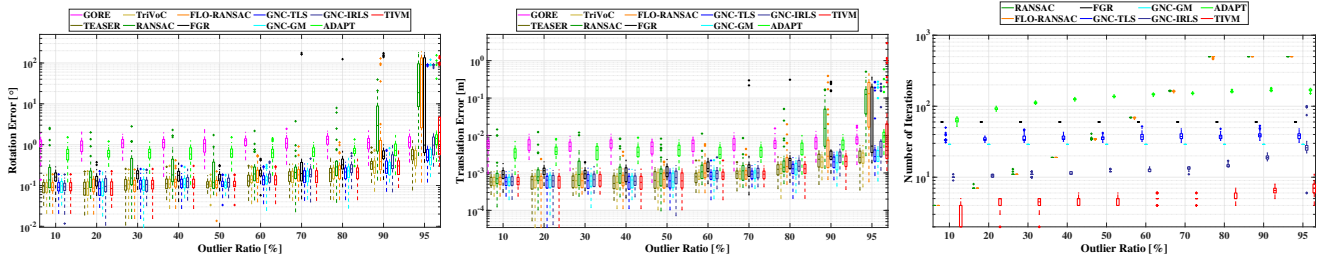


Fig. 6. Benchmarking on robust point cloud registration.

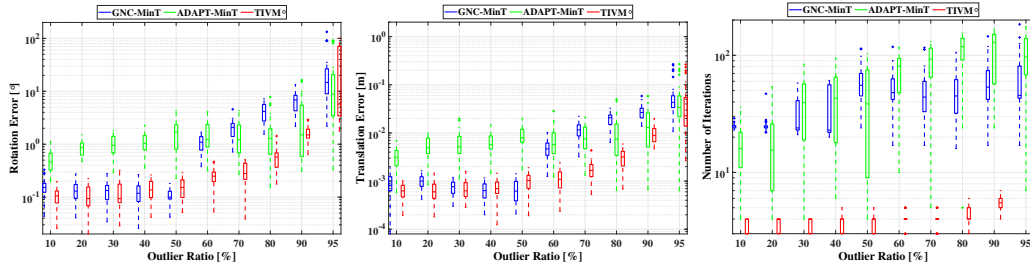


Fig. 7. Benchmarking on robust point cloud registration without inlier-noise statistics.

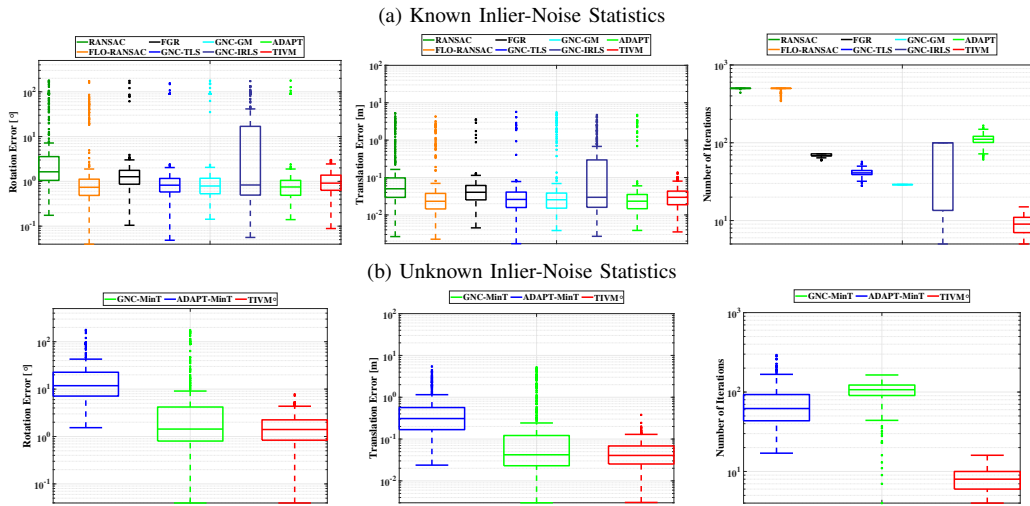


Fig. 8. Quantitative results on scan stitching problems over 3DMatch dataset [34].

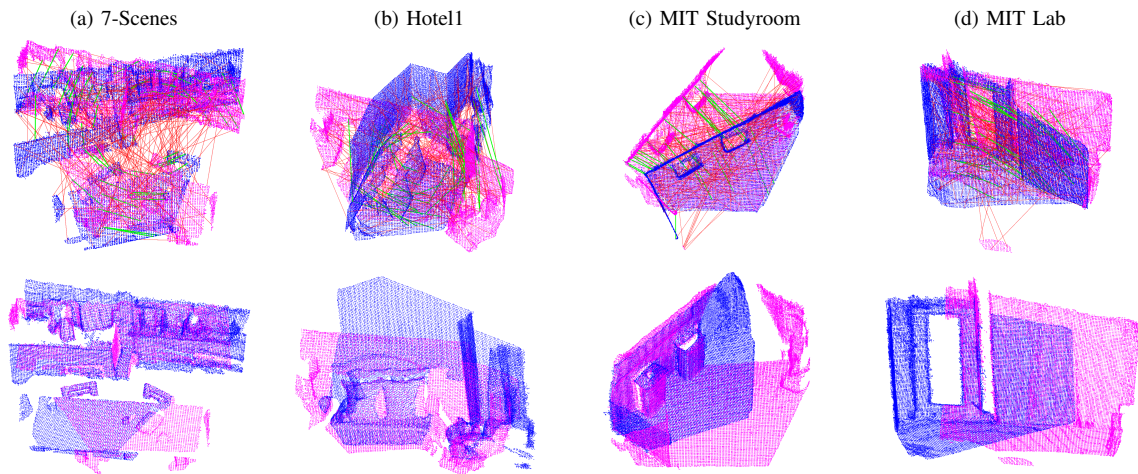


Fig. 9. Qualitative examples of scan stitching over 3DMatch dataset [34]. **First Row:** Matched correspondences, where green and red lines denote inliers and outliers, respectively. **Second Row:** Scan stitching results using TIVM.

We can observe from Fig. 6 that: (1) our TIVM is robust against 90% outliers, in line with the 3 GNC solvers and ADAPT, and outperforming FGR (that fails at 70%) and

RANSAC & FLO-RANSAC (that fail at 90%), and (2) TIVM generally converges in merely 2–10 iterations and is the fastest general-purpose robust solver.

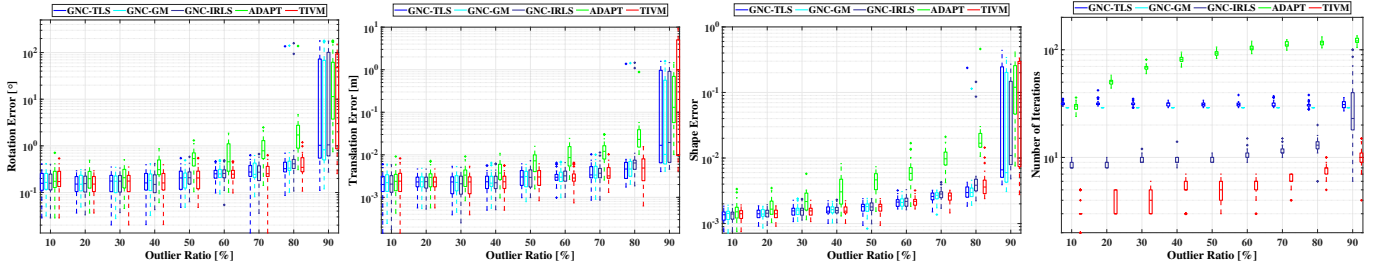


Fig. 10. Benchmarking on robust category-level perception.

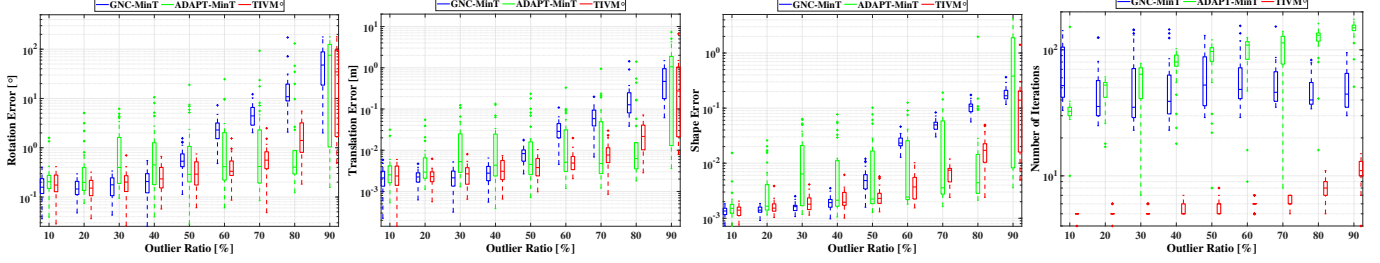


Fig. 11. Benchmarking on robust category-level perception without inlier-noise statistics.

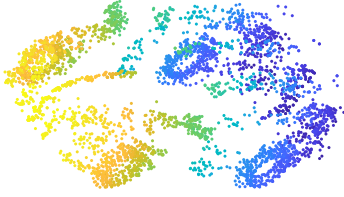


Fig. 12. Demonstration of the 3D car model from [35] (in the form of point cloud) as the ground-truth shape for robust category-level perception.

Unknown Noise-Statistics Tests. In Fig. 7, we test the noise-statistics-free cases, where we find that: TIVM^o still retains robustness and sufficient accuracy at 90% (rotation errors lower than 3° and translation errors lower than 0.02m) and also converges in about 10 iterations, up to 1 order of magnitude faster than the other 2 minimally-tuned solvers.

Real Application: Scan Stitching. To validate our solver in real-world applications, we further test TIVM along with other competitors over the scan stitching problems using the 3DMatch dataset [34]. We use the FPFH descriptor [29] to establish correspondences between RGB-D scans and only preserve the test cases with outlier ratios lower than 90% (because all general-purpose solvers fail at over 90%) and $N > 200$ (for ensuring sufficient correspondence number). The quantitative results in both known and unknown inlier-statistics scenarios are displayed in Fig. 8, and some qualitative examples are illustrated in Fig. 9. We can see that our solver shows the best outlier-robustness and the highest time-efficiency (≤ 15 iterations) in both scenarios.

C. Category-Level Perception

Robust category-level perception estimates both the pose: $(\mathbf{R}^* \in SO(3), \mathbf{t}^* \in \mathbb{R}^3)$ and the shape model: $\mathbf{c}^* = [c_1^*, c_2^*, \dots, c_K^*]^\top$ of the 3D object from correspondences between points $\{\mathbf{y}_i\}_{i=1}^N$ on the object and points on the given CAD models: $\{\mathcal{B}_k\}_{k=1}^K = \{\{\mathbf{b}_k(i)\}_{i=1}^N\}_{k=1}^K$, which are contaminated by outliers. (Readers can find more explicit problem definition in [16].) These correspondences are de-

scribed as: $\{\mathbf{y}_i \leftrightarrow \{\mathbf{b}_k(i)\}_{k=1}^K\}_{i=1}^N$, and the residual function can be represented as: $Re_i = \|\mathbf{R}^* \sum_{k=1}^K c_k^* \mathbf{b}_k(i) + \mathbf{t}^* - \mathbf{y}_i\|_2$.

Setup. We make use of the car models from the FG3DCar dataset [35] as the shape model, where we can find $K = 15$ available CAD models over $N = 256$ 3D points. In each run, we generate a random ground-truth pose: $(\mathbf{R}_{gt} \in SO(3), \mathbf{t}_{gt} \in \mathbb{R}^3)$ and a ground-truth shape model: \mathbf{c}_{gt} to transform these car models and then add Gaussian noise with standard deviation $\sigma = 0.01$ on them to obtain the object points: $\{\mathbf{y}_i\}_{i=1}^N$. We replace some of the object points with random 3D points to produce outliers. The inlier threshold is set to $\tau = 5\sigma$ for known noise-statistics tests. We apply the certifiably optimal method [16] as the non-minimal solver. Fig. 10 and 11 show the results over 30 Monte Carlo runs in 2 noise-statistics scenarios, and the entire shape model in this experiment is displayed in Fig. 12.

TIVM remains robust at 80% outliers, while all the other general-purpose non-minimal solvers yields wrong results at 80%. The iteration number of TIVM is within 3–15 at all time, which is much more efficient than the other competitors. Although GNC-IRLS can converge in 8–20 iterations with $\leq 80\%$ outliers, it gets much slower at 90%.

Unknown Noise-Statistics Tests. In Fig. 11 where the noise-statistics-free tests are performed, we can see that the robustness of TIVM^o is still 80% and its iteration number till convergence is still within 15. ADAPT-MinT seems to show a lower estimation error at 80% outliers, but it is 1–2 orders of magnitude less efficient than TIVM^o and is less accurate at the lower-outlier regime.

VI. CONCLUSION

In this paper, a general-purpose non-minimal robust estimator TIVM for robust geometric perception problems is rendered. By incorporating the multi-layered intra-class variance maximization technique with a self-adaptive layer-number tuning strategy into an iterative optimizing framework, an effective thresholding method for robust estimation

by separating inliers from outliers is proposed. The special thresholding & grouping mechanism of this estimator allows its robust performance even in the noise-statistics-free situations. Through validation experiments on various classic perception problems, the proposed estimator is shown to tolerate 70–90% outliers with high accuracy no matter the inlier-noise statistics are given or not and usually require only 3–15 iterations to reach convergence, more efficient than the other general-purpose state-of-the-art robust solvers.

The main limitation of TIVM lies in that: when the measurements in the geometric perception problem are too sparse (e.g. fewer than 10), the intra-class variance maximization over the residual-error histogram may not be sufficiently accurate to reflect the measurement distribution (since the intervals are significantly more than the data), which might compromise its performance.

Our demo code is provided at: <https://github.com/LeiSun-98/TIVM-master>.

REFERENCES

- [1] V. Tzoumas, P. Antonante, and L. Carlone, “Outlier-robust spatial perception: Hardness, general-purpose algorithms, and guarantees,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5383–5390.
- [2] J. Zhang and S. Singh, “Visual-lidar odometry and mapping: Low-drift, robust, and fast,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 2174–2181.
- [3] A. Zeng, K.-T. Yu, S. Song, D. Suo, E. Walker, A. Rodriguez, and J. Xiao, “Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge,” in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 1386–1383.
- [4] J. M. Wong, V. Kee, T. Le, S. Wagner, G.-L. Mariottini, A. Schneider, L. Hamilton, R. Chipalkatty, M. Hebert, D. M. Johnson, et al., “Segicp: Integrated deep semantic segmentation and pose estimation,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5784–5789.
- [5] G. Blais and M. D. Levine, “Registering multiview range data to create 3d computer objects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 820–824, 1995.
- [6] S. Choi, Q.-Y. Zhou, and V. Koltun, “Robust reconstruction of indoor scenes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5556–5565.
- [7] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [8] 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.
- [9] O. Chum, J. Matas, and J. Kittler, “Locally optimized ransac,” in *Joint Pattern Recognition Symposium*. Springer, 2003, pp. 236–243.
- [10] O. Chum and J. Matas, “Matching with proscac-progressive sample consensus,” in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, vol. 1. IEEE, 2005, pp. 220–226.
- [11] L. Sun, “Ransic: Fast and highly robust estimation for rotation search and point cloud registration using invariant compatibility,” *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 143–150, 2022.
- [12] Q.-Y. Zhou, J. Park, and V. Koltun, “Fast global registration,” in *European Conference on Computer Vision*. Springer, 2016, pp. 766–782.
- [13] 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.
- [14] L. Peng, C. Kümmerle, and R. Vidal, “On the convergence of irls and its variants in outlier-robust estimation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 808–17 818.
- [15] H. Yang and L. Carlone, “In perfect shape: Certifiably optimal 3d shape reconstruction from 2d landmarks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 621–630.
- [16] J. Shi, H. Yang, and L. Carlone, “Optimal pose and shape estimation for category-level 3d object perception,” *Robotics: Science and Systems (RSS)*, 2021.
- [17] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, “Se-sync: A certifiably correct algorithm for synchronization over the special euclidean group,” *The International Journal of Robotics Research*, vol. 38, no. 2-3, pp. 95–125, 2019.
- [18] B. K. Horn, “Closed-form solution of absolute orientation using unit quaternions,” *Josa a*, vol. 4, no. 4, pp. 629–642, 1987.
- [19] D. Nistér, “An efficient solution to the five-point relative pose problem,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 6, pp. 756–770, 2004.
- [20] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-squares fitting of two 3-d point sets,” *IEEE Transactions on pattern analysis and machine intelligence*, no. 5, pp. 698–700, 1987.
- [21] J. Briaies and J. Gonzalez-Jimenez, “Convex global 3d registration with lagrangian duality,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4960–4969.
- [22] H. Li, “Consensus set maximization with guaranteed global optimality for robust geometry estimation,” in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 1074–1080.
- [23] K. MacTavish and T. D. Barfoot, “At all costs: A comparison of robust cost functions for camera correspondence outliers,” in *2015 12th conference on computer and robot vision*. IEEE, 2015, pp. 62–69.
- [24] L. Sun and L. Deng, “Trivoc: Efficient voting-based consensus maximization for robust point cloud registration with extreme outlier ratios,” *IEEE Robotics and Automation Letters*, 2022.
- [25] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [26] P. Antonante, V. Tzoumas, H. Yang, and L. Carlone, “Outlier-robust estimation: Hardness, minimally tuned algorithms, and applications,” *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 281–301, 2021.
- [27] R. Hartley, J. Trunpf, Y. Dai, and H. Li, “Rotation averaging,” *International journal of computer vision*, vol. 103, no. 3, pp. 267–305, 2013.
- [28] S. H. Lee and J. Civera, “Robust single rotation averaging,” *arXiv preprint arXiv:2004.00732*, 2020.
- [29] 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.
- [30] B. Curless and M. Levoy, “A volumetric method for building complex models from range images,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996, pp. 303–312.
- [31] A. P. Bustos and T.-J. Chin, “Guaranteed outlier removal for point cloud registration with correspondences,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2868–2882, 2017.
- [32] H. Yang, J. Shi, and L. Carlone, “Teaser: Fast and certifiable point cloud registration,” *IEEE Transactions on Robotics*, 2020.
- [33] K. Lebeda, J. Matas, and O. Chum, “Fixing the locally optimized ransac—full experimental evaluation,” in *British machine vision conference*, vol. 2. Citeseer, 2012.
- [34] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, “3dmatch: Learning the matching of local 3d geometry in range scans,” in *CVPR*, vol. 1, no. 2, 2017, p. 4.
- [35] Y.-L. Lin, V. I. Morariu, W. Hsu, and L. S. Davis, “Jointly optimizing 3d model fitting and fine-grained classification,” in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*. Springer, 2014, pp. 466–480.