

Robust Point Cloud Registration with Geometry-based Transformation Invariant Descriptor

Jianjie Lin, Markus Rickert, Long Wen, Yingbai Hu and Alois Knoll

Abstract—This work presents a novel method for point registration in 3D space. The proposed algorithm utilizes transformation-invariant geometry information to estimate the pose of objects based on correspondences between points in two sets. Conventional methods use geometry descriptors to find these correspondences, which can result in a large number of outliers. Most existing algorithms are error-prone when outliers are present. Instead of formulating point registration as a non-convex optimization problem, we propose an intuitive method that filters out spurious correspondences. This is achieved by evaluating three different geometry-based transformation-invariant descriptors for outlier removal. We construct fully connected graphs with the proposed descriptors on correspondences, and convert the outlier removal problem into a subgraph isomorphism problem that is solved using a binary clustering approach. The resulting inlier clustering is used to estimate the transformation between the two point sets. The effectiveness of the proposed approach is evaluated on standard 3D data and the 3DMatch scan matching dataset, and compared against existing state-of-the-art methods. Results show that our method effectively reduces outliers and performs similarly to these methods.

I. INTRODUCTION

Point cloud registration is a core operation in computer vision and robotics that computes the pose transformation between two-point sets. It has applications in MRI/CAT scan alignment [1], robot manipulation [2] (Fig. 1), object recognition and localization, 3D reconstruction, etc. A standard approach to point registration without assumed correspondence is based on the iterative closest point (ICP) algorithm [3], [4] by alternatively finding correspondence and solving correspondence-based problems. However, the success of ICP relies heavily on a good initial alignment. The correspondence-based approach is another prevalent paradigm approach. It estimates the transformation T between two point sets by giving the putative correspondences. The transformation T can be analytically solved [5], [6], [7], [8] if the putative correspondences are all correct. In practice, most correspondence point sets contain a large fraction of outliers. Even state-of-the-art local descriptors such as FPFH [9] and some deep learning-based methods [10] have difficulties in producing acceptable outlier rates in real-world problems and inevitably produce mismatches. As a result, this leads to a poor estimation of these solvers. RANSAC [11] is probably the most widely used algorithm for estimating transformations assuming the presence of outliers. The aim of RANSAC is to maximize the Consensus

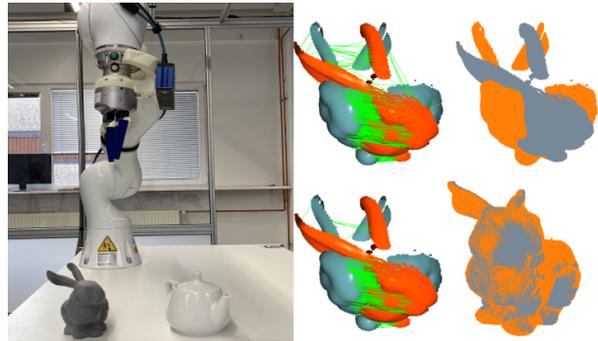


Fig. 1: The picture on the left shows the scanned scenes, and the one on the right illustrates an excerpt of bunny alignment and outlier removal.

set by randomly sampling the smallest subset and estimating the transformations using least squares. The transform with maximum consensus is returned as the solution. However, the sampling phase of RANSAC would be inefficient given a large fraction of outliers. The global method [12], [13], [14] is an alternative algorithm that guarantees robustness. Branch-and-bound (B&B) algorithms are used to optimize the parameters of the objective function. However, a general weakness of global algorithms is their high computational cost, especially on data with large size N and high outliers.

Contribution: In our previous work, we used the rotation-invariant descriptor [15] to address the object classification problem. In this work, we further explore the use of geometry-based transformation-invariant descriptors, including the **Triangle Angle Invariant (TAInI)**, **Triangle Dot Product Invariant (TDoI)**, **Triangle Area Invariant (TAInI)**, and **Triangle Dot Product Invariant (TDoI)**, for robust point cloud registration by removing outliers. We construct two fully connected graphs using the proposed descriptors on the putative correspondences and convert the outlier removal into the subgraph isomorphism problem, aiming to find two subgraphs from these two fully connected graphs that have the same topology (isomorphism). The subgraph isomorphism problem is proven to be NP-complete based on the reduction of the clique problem. In this work, we present a novel approach to reformulate the subgraph matching problem as a clustering approach to enhance the performance of outlier removal, where one cluster is identified as outliers and the other is denoted as inliers. The remaining two subgraphs contain only inliers, which exhibit the same topology and can be transformed into each other with the estimated transformation matrix. Based on our mathematical

Jianjie Lin, Markus Rickert, Long Wen, Yingbai Hu, Alois Knoll are with Robotics, Artificial Intelligence and Real-Time Systems, School of Computation, Information and Technology, Technical University of Munich, Munich, Germany {jianjie.lin, rickert, wenl, yingbai.hu, knoll}@in.tum.de

proofs and experimental observations, the proposed geometry transformation invariant descriptor effectively removes spurious correspondences from \mathcal{H} , resulting in a subset of true inliers, denoted as \mathcal{H}' . Some previous works, such as GORE [16], formulate the outlier removal as an optimization problem (maximization Consensus set). However, our proposed transformation invariant descriptor uses the geometry information for outlier removal without any additional effort, as the geometry information is inherently available from the point cloud. We present our technique as an efficient preprocessing step for robust point cloud registration. The remaining correspondences are further fed into the transformation estimation process. There are many off-the-shelf estimators that can be used to achieve this goal. We evaluated the presented transformation-invariant registration algorithm on multiple datasets, including Standard 3D data [17] and 3DMatchData [18]. Extensive experiments demonstrate that the presented approach matches or exceeds the accuracy of state-of-the-art robust registration pipelines.

II. RELATED WORK

6D pose estimation is a well-established and extensively studied topic in computer vision, and point registration is a common technique used to determine the spatial relationship between two sets of 3D points. Correspondence-based [13], [19], [20] and correspondence-free [21], [22], [23], [24] are two popular paradigms for registering 3D point clouds. In this work, we mainly focus on the correspondence-based approach.

The correspondence-based approach involves pre-matching the two point clouds using feature descriptors [25], [9] to generate putative correspondences, followed by applying an estimator to compute the transformation. Closed-form solutions [5], [6], [7], [8] exist for finding the optimal solution when the correspondences are correct (outlier-free) and the points are affected by isotropic zero-mean Gaussian noise. However, even state-of-the-art feature descriptors cannot guarantee all correct matches. Local geometrical descriptors such as the Fast Point Feature Histogram (FPFH) [9] induce a large fraction of outliers if the point clouds are noisy or have different distributions. Even advanced deep learning-based feature descriptors such as 3DSmoothNet [10] cannot guarantee sufficient correspondences and are not generalizable to new datasets. Moreover, outliers are common in real-world applications, making robust registration methods essential.

Random sample consensus (RANSAC) is a popular approach against outliers, widely used in vision and robotics applications. RANSAC is efficient and can be formulated as an optimization problem by maximizing the consensus set. However, as the outlier ratio increases, RANSAC's convergence becomes slow, and its accuracy decreases since it is challenging to sample the true inliers (good consensus set). Global methods are a variant of RANSAC for improving robustness. Yang et al. proposed Globally Optimal ICP (GO-ICP) [13], which uses a branch-and-bound framework to improve robustness, but the runtime of B&B increases

exponentially with the input size, making it unrealistic for real-world applications. Zhou et al. proposed Fast Global Registration (FGR) [20], which significantly improves GO-ICP by using a scaled Geman-McClure estimator to describe the error function and leveraging graduated non-convexity (GNC) to solve the resulting non-convex optimization. However, FGR exhibits poor performance when the outlier ratio increases. Yang et al. proposed the Teaser++ algorithm [19], which reformulates robust registration as a Truncated Least Squares Registration problem (TLS) and applies GNC to solve the non-convex optimization problem. Teaser++ is highly robust to a large fraction of outliers. However, the practical maximum clique algorithm [26] used in Teaser++ is not robust when evaluated with real datasets. Parra et al. proposed Guaranteed Outlier Removal (GORE)[16] as a data preprocessing step for point registration by directly removing outliers and resulting in a high true inliers consensus set. GORE uses the global algorithm (B&B) to tackle the problem. Parra et al. [26] also proposed a practical maximum clique (PMC) algorithm to find pairwise-consistent correspondences in 3D registration. GORE and PMC are similar in spirit to our proposed approach, which solves the problem with a global optimization algorithm. We propose a new approach for robust 6D pose estimation based on geometry-based transformation invariant descriptors. However, instead of directly removing outliers as in GORE, we use the putative correspondences to build transformation invariant descriptors and then cluster them to achieve outlier removal. This approach enables us to take advantage of any estimator for inferring transformation as the backend.

III. PROBLEM FORMULATION

The correspondence-based point registration problem can be formulated as a nonlinear least squares problem, which is given by the following equation:

$$\min_{\mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathbb{R}^3} \sum_{i=1}^N \|\mathbf{p}_i - \mathbf{R}\mathbf{q}_i - \mathbf{t}\|^2, \quad (1)$$

Here, $(\mathbf{p}_i, \mathbf{q}_i), i = 1, \dots, N$ are given putative correspondences, and ϵ_i is a zero-mean Gaussian noise with isotropic covariance $\sigma_i^2 \mathbf{I}_3$. In the outlier-free case, where the putative correspondences are correct, equation (1) has a closed form [8] and the estimation of rotation and translation can be decoupled. However, in practical applications, the outlier-free case is non-existent as putative correspondences can be incorrect due to erroneous keypoint matching (e.g., FPFH matching [9] or other feature-based matching approaches). Such mismatches result in a large fraction of correspondences being outliers. The nonlinear least squares problem is not robust to outliers and a single outlier can significantly affect the estimation. Therefore, this work proposes to utilize geometry-based transformation invariant descriptors to remove outliers. Mathematically, this approach aims to maximize the cardinality of true inliers from the putative correspondences.

IV. ALGORITHM

The putative correspondences $(\mathbf{p}_i, \mathbf{q}_i)$ obey the following generative model:

$$\mathbf{q}_i = \mathbf{R}\mathbf{p}_i + \mathbf{t} + \mathbf{o}_i + \boldsymbol{\epsilon}_i, \quad (2)$$

where \mathbf{o}_i is zero if $(\mathbf{q}_i, \mathbf{p}_i)$ is an inlier, otherwise \mathbf{o}_i is arbitrary (outlier). The noise $\boldsymbol{\epsilon}_i$ is assumed to be bounded with $|\epsilon_i| \leq \delta$. The relative position between two corresponding pairs can be formulated as

$$\mathbf{q}_j - \mathbf{q}_i = \mathbf{R}(\mathbf{p}_j - \mathbf{p}_i) + \tilde{\boldsymbol{\delta}}_{ij} + \tilde{\boldsymbol{\epsilon}}_{ij}, \quad (3)$$

with $\tilde{\boldsymbol{\delta}}_{ij} = \mathbf{o}_j - \mathbf{o}_i$ and $\tilde{\boldsymbol{\epsilon}}_{ij} = \boldsymbol{\epsilon}_j - \boldsymbol{\epsilon}_i$. We can further define $\hat{\mathbf{q}}_{ij} = \mathbf{q}_j - \mathbf{q}_i$ and $\hat{\mathbf{p}}_{ij} = \mathbf{p}_j - \mathbf{p}_i$. The relative location of pairs $(\hat{\mathbf{q}}_{ij}, \hat{\mathbf{p}}_{ij})$ depends only on the rotation but its distance is invariant to \mathbf{R} and \mathbf{t} , which can be described as

$$\|\hat{\mathbf{q}}_{ij}\| = \|\mathbf{R}\hat{\mathbf{p}}_{ij} + \tilde{\boldsymbol{\delta}}_{ij} + \tilde{\boldsymbol{\epsilon}}_{ij}\|. \quad (4)$$

Note that $\tilde{\boldsymbol{\delta}}_{ij} = \mathbf{0}$ for the case, where the corresponding pair $(\mathbf{q}_i, \mathbf{p}_i)$ is an inlier. The eqn. (4) holds the following inequality by applying $\|\boldsymbol{\epsilon}_{ij}\| \leq \delta_{ij}$ and the triangle inequality:

$$\|\mathbf{R}\hat{\mathbf{p}}_{ij}\| - \delta_{ij} \leq \|\mathbf{R}\hat{\mathbf{p}}_{ij} + \boldsymbol{\epsilon}_{ij}\| \leq \|\mathbf{R}\hat{\mathbf{p}}_{ij}\| + \delta_{ij}, \quad (5)$$

hence we can rewrite (4) as

$$\|\hat{\mathbf{q}}_{ij}\| = \|\mathbf{R}\hat{\mathbf{p}}_{ij}\| + \tilde{\delta}_{ij} + \bar{c}_{ij} = \|\hat{\mathbf{p}}_{ij}\| + \tilde{\delta}_{ij} + \bar{c}_{ij}, \quad (6)$$

with $|\bar{c}_{ij}| \leq \delta_{ij}$, and $\tilde{\delta}_{ij} = 0$ if both i and j are inliers or is an arbitrary scalar otherwise. The equation (6) shows that the geometry information between corresponding pairs can be exploited for the outlier removal if some specific conditions are fulfilled. However, the standalone information of distance is too sparse for removal and can be sensitive to the predefined threshold value. In this work, we propose three different geometry-based transformation invariant descriptors, which implicitly contains the distance preservation property and also other geometry invariant attribute.

A. Geometry-based Transformation-invariant Descriptors

We define the relative position between arbitrary triple corresponding pairs with respect to $(\mathbf{q}_i, \mathbf{p}_i)$ as $\Omega_{ijk} = \{(\hat{\mathbf{q}}_{ij}, \hat{\mathbf{p}}_{ij}), (\hat{\mathbf{q}}_{ik}, \hat{\mathbf{p}}_{ik})\}$.

1) **Triangle Angle Invariance (TAnI)**: The dot product of Ω_{ijk} can be computed as

$$\begin{aligned} \langle \hat{\mathbf{q}}_{ij}, \hat{\mathbf{q}}_{ik} \rangle &= \langle \mathbf{R}\hat{\mathbf{p}}_{ij} + \tilde{\boldsymbol{\delta}}_{ij} + \tilde{\boldsymbol{\epsilon}}_{ij}, \mathbf{R}\hat{\mathbf{p}}_{ik} + \tilde{\boldsymbol{\delta}}_{ik} + \tilde{\boldsymbol{\epsilon}}_{ik} \rangle \\ &= \hat{\mathbf{p}}_{ij}^T \hat{\mathbf{p}}_{ik} + \hat{\mathbf{p}}_{ij}^T \mathbf{R}^T (\tilde{\boldsymbol{\delta}}_{ik} + \tilde{\boldsymbol{\epsilon}}_{ik}) + \hat{\mathbf{p}}_{ik}^T \mathbf{R}^T (\tilde{\boldsymbol{\delta}}_{ij} + \tilde{\boldsymbol{\epsilon}}_{ij}) \\ &\quad + (\tilde{\boldsymbol{\delta}}_{ij} + \tilde{\boldsymbol{\epsilon}}_{ij})^T (\tilde{\boldsymbol{\delta}}_{ik} + \tilde{\boldsymbol{\epsilon}}_{ik}) \end{aligned} \quad (7)$$

If corresponding pair $(\mathbf{q}_h, \mathbf{p}_h)$, $h \in \{i, j, k\}$ are inliers, then $\tilde{\boldsymbol{\delta}}_{ij}, \tilde{\boldsymbol{\delta}}_{ik}$ are equal to zero, the eqn. (7) can be simplified as

$$\hat{\mathbf{q}}_{ij}^T \hat{\mathbf{q}}_{ik} = \hat{\mathbf{p}}_{ij}^T \hat{\mathbf{p}}_{ik} + \hat{\mathbf{p}}_{ij}^T \mathbf{R}^T \tilde{\boldsymbol{\epsilon}}_{ik} + \hat{\mathbf{p}}_{ik}^T \mathbf{R}^T \tilde{\boldsymbol{\epsilon}}_{ij} + \tilde{\boldsymbol{\epsilon}}_{ij}^T \tilde{\boldsymbol{\epsilon}}_{ik} \quad (8)$$

We normalized the point cloud to $[0, 1]^3$. Since the noise $|\epsilon_i|$ is bounded, it can be confirmed that $\|\hat{\mathbf{p}}_{ik}^T \mathbf{R}^T \tilde{\boldsymbol{\epsilon}}_{ij}\|$ is also bounded. Therefore, it leads to

$$\hat{\mathbf{q}}_{ij}^T \hat{\mathbf{q}}_{ik} = \hat{\mathbf{p}}_{ij}^T \hat{\mathbf{p}}_{ik} + \Delta_i, \quad (9)$$

where Δ_i is bounded as well. Specifically, when $\epsilon = 0$ (noise-free), it can be confirmed that the angle between relative position of triple corresponding pairs is invariant regarding rotation and translation. The angle can be computed as $\theta_{p_{ijk}} = \arccos\left(\frac{\hat{\mathbf{p}}_{ij}^T \hat{\mathbf{p}}_{ik}}{\|\hat{\mathbf{p}}_{ij}\| \|\hat{\mathbf{p}}_{ik}\|}\right)$, and $\theta_{q_{ijk}} = \arccos\left(\frac{\hat{\mathbf{q}}_{ij}^T \hat{\mathbf{q}}_{ik}}{\|\hat{\mathbf{q}}_{ij}\| \|\hat{\mathbf{q}}_{ik}\|}\right)$. The operation of \arccos and $\|\cdot\|$ requires an amount of computational resource.

2) **Triangle Dot product Invariance (TDoI)**: To reduce the computational complexity of TAnI, we introduce another geometry based descriptor which requires only dot product. The dot product preserves the triple corresponding pair triangle geometry information, which is utilized to search for the Consensus set. The descriptor of TDoI is defined as

$$\Gamma_{q_{ijk}} = \hat{\mathbf{q}}_{ij}^T \hat{\mathbf{q}}_{ik} + \hat{\mathbf{q}}_{ji}^T \hat{\mathbf{q}}_{jk} + \hat{\mathbf{q}}_{kj}^T \hat{\mathbf{q}}_{ki} \quad (10a)$$

$$\Gamma_{p_{ijk}} = \hat{\mathbf{p}}_{ij}^T \hat{\mathbf{p}}_{ik} + \hat{\mathbf{p}}_{ji}^T \hat{\mathbf{p}}_{jk} + \hat{\mathbf{p}}_{kj}^T \hat{\mathbf{p}}_{ki}. \quad (10b)$$

The descriptor TDoI is much more efficient than TAnI. But it can be sensible to noisy dataset. We further study the area of triangle, which is more robust than the TAnI and TDoI in the experiments, since it contains the information of angle, and distance information.

3) **Triangle Area Invariance (TAri)**: The area of triangle built on the triple corresponding pairs, denoted as Δ_{p_i} and Δ_{q_i} , respectively is computed as

$$\Delta_{q_{ijk}} = \frac{\|\hat{\mathbf{q}}_{ij} \times \hat{\mathbf{q}}_{ik}\|}{2}, \quad \Delta_{p_{ijk}} = \frac{\|\hat{\mathbf{p}}_{ij} \times \hat{\mathbf{p}}_{ik}\|}{2}, \quad (11)$$

since $\hat{\mathbf{q}}_{ik}$ and $\hat{\mathbf{p}}_{ik}$ are translation invariant. It is straightforward to prove that the triangle area is preserved if triple corresponding pairs are inliers. Mathematically speaking, in the case of inliers, the following equation is satisfied:

$$\hat{\mathbf{q}}_{ij} \times \hat{\mathbf{q}}_{ik} = (\mathbf{R}\hat{\mathbf{p}}_{ij} + \hat{\boldsymbol{\epsilon}}_{i+1}) \times (\mathbf{R}\hat{\mathbf{p}}_{ik} + \hat{\boldsymbol{\epsilon}}_i) \quad (12)$$

$$\begin{aligned} \|\hat{\mathbf{q}}_{ij} \times \hat{\mathbf{q}}_{ik}\| &= \|\mathbf{R}\hat{\mathbf{p}}_{ij} + \hat{\boldsymbol{\epsilon}}_{i+1}\| \|\mathbf{R}\hat{\mathbf{p}}_{ik} + \hat{\boldsymbol{\epsilon}}_i\| \sin(\theta_{q_i}) \\ &= \|\hat{\mathbf{p}}_{ij}\| \|\hat{\mathbf{p}}_{ik}\| \sin(\theta_{q_i}), \quad \text{if } |\hat{\epsilon}_i| \leq \delta. \end{aligned} \quad (13)$$

As proven from the previous section, the angle is preserved if these triple corresponding pairs are inliers, then

$$\begin{aligned} \|\hat{\mathbf{q}}_{ij} \times \hat{\mathbf{q}}_{ik}\| &= \|\hat{\mathbf{p}}_{ij}\| \|\hat{\mathbf{p}}_{ik}\| \sin(\theta_{q_i}) \\ &= \|\hat{\mathbf{p}}_{ij}\| \|\hat{\mathbf{p}}_{ik}\| \sin(\theta_{p_i}) = \|\hat{\mathbf{p}}_{ij} \times \hat{\mathbf{p}}_{ik}\|. \end{aligned} \quad (14)$$

Therefore, the triangle area is preserved ($\Delta_{q_{ijk}} \approx \Delta_{p_{ijk}}$) in case of $|\hat{\epsilon}_i| \leq \delta$ and three corresponding pairs are denoted as inlier pairs. It is known that the cross product requires a lot of computing resources. In the experiment section, we will discuss the efficiency of these three geometry based transformation descriptors in terms of time in Table III.

B. Robust registration problem formulation

We can formulate the robust registration problem as search for a Consensus set by satisfying the geometry-based transformation invariance conditions:

$$\text{Card} \left(\text{F}_{\text{cls}} \left(\text{Set} \left\{ \sum_i^{|\mathcal{N}|} \mathbb{1}(\text{pMetric}(\beta_{ijk}), c^2), \forall i \neq j \neq k \right\} \right) \right), \quad (15)$$

where $|N|$ is the number of corresponding pairs, $\mathbb{1}$ is the indicator function, and returns one if the pMetric value is smaller than the predefined threshold value. For simplicity, we define $F_1 = \mathbb{1}(\text{pMetric}(\beta_{ijk}, c^2))$. We sum the indicator results at index i and add the sum result into a set. The function F_{cls} is used to decide if the corresponding pairs are inliers, which will be discussed in detail in Section IV-C. For simplification, we reformulate (15) as

$$\text{Card}\left(F_{\text{cls}}\left(\text{Set}\left\{F_{\text{count},i}, \forall i \neq j \neq k\right\}\right)\right), \quad (16)$$

where $F_{\text{count},i} = \sum_i^{|N|} F_1$. The variable of $\text{pMetric}(\beta_{ijk})$ is defined as the norm the triangle angle difference $\left\|\left(\theta_{q_{ijk}} - \theta_{p_{ijk}}\right)\right\|^2$ by using the TANl, and the difference of $\left\|\left(\Gamma_{q_{ijk}} - \Gamma_{p_{ijk}}\right)\right\|^2$ with the TDoI. By applying the TARl, this value becomes $\left\|\left(\Delta_{q_{ijk}} - \Delta_{p_{ijk}}\right)\right\|^2$.

1) *Build Full Graph:* In general, all corresponding pairs can be described as a corresponding-pair full graph using the Kronecker product \otimes . Let $G = (V, E)$, $H = (V', E')$ be graphs, find out a subgraph $\{G_0 = (V_0, E_0) \mid V_0 \subseteq V, E_0 \subseteq E \cap (V_0 \times V_0)\}$ and $\{H_0 = (V'_0, E'_0) \mid V'_0 \subseteq V', E'_0 \subseteq E' \cap (V'_0 \times V'_0)\}$ such that $G_0 \cong H_0$, which means the G_0 and H_0 has the same topology. The proof of subgraph isomorphism being NP-complete. The relative position between two locations is denoted as the edge and can be computed as $Q = I \otimes p - p \otimes I$, removing N zero rows, which contains the itself. To this end, we split the Q into a three-dimensional matrix as $\overline{Q} = [\overline{Q}_0 \cdots \overline{Q}_{N-1}] \in \mathbb{R}^{N \times (N-1) \times 3}$, where $\overline{Q}_i = [p_i - p_0, \dots, p_i - p_{N-1}]^T$.

2) *The computational complexity:* The pairwise metrics function $\text{pMetric}(\cdot)$ defined in (16) is used to compute the metric between each two relative position. The number of triple corresponding pairs $|C|$ is upper-bounded by $\frac{N \times (N-1) \times (N-2)}{2}$.

3) *Discussion the possibility of inlier corresponding pair:* In equation (16), the function F_1 determines whether the pair (p_i, q_i) should be considered an inlier, by comparing its corresponding metric value $\text{pMetric}(\beta_{ijk})$ with a predefined threshold value. Table I shows that if (p_i, q_i) is an outlier, then $\text{pMetric}(\beta_{ijk})$ will always be greater than the threshold value, resulting in a small value for $F_{\text{count},i}$; ideally, this value should be zero. On the other hand, if (p_i, q_i) is an inlier, the value of F_1 will be one only if the other two corresponding pairs are also inliers. As a result, $F_{\text{count},i}$ will be a large value n , as triangles can be formed with other inlier pairs. For instance, if we have N putative corresponding pairs with an inlier ratio of δ , then in the ideal case, if (p_i, q_i) is an inlier, n should equal $\frac{(\delta N - 1)(\delta N - 2)}{2}$, based on Table I. In the end, the value of $F_{\text{count},i}$ can be used to prune edges inside the fully connected graphs G and H . The resulting subgraphs, G_0 and H_0 , can be transformed into each other using the estimated transformation.

C. Clustering for inliers and outliers

1) *Limitations of using single threshold:* The function $F_{\text{count},i}$ is a useful tool for classifying corresponding pairs

TABLE I: Case study to determine the value of F_1 for the corresponding pairs (p_i, q_i) , where T denotes that the pair is an inlier, F denotes that it is an outlier, and * can represent either a T or an F.

	inlier/outlier	F_1
$(p_i, q_i), (p_j, q_j), (p_k, q_k)$	TTT	1
$(p_i, q_i), (p_j, q_j), (p_k, q_k)$	T*F	0
$(p_i, q_i), (p_j, q_j), (p_k, q_k)$	TF*	0
$(p_i, q_i), (p_j, q_j), (p_k, q_k)$	F**	0

as inliers or outliers. One intuitive approach is to consider a pair an inlier if its $F_{\text{count},i}$ value is greater than a predefined hyper-parameter. However, this parameter can be difficult to tune and may need to be adjusted for different dataset distributions. To address this issue, we propose the use of clustering techniques. The goal of consensus maximization is to identify inliers and outliers, so we can interpret F_{cls} as a means of clustering corresponding pairs into inlier and outlier clusters, where the inlier clusters have a significantly larger $F_{\text{count},i}$ than the outlier clusters.

2) *Clustering for the inliers and outliers:* Instead of directly optimizing the non-convex nonlinear problem (16), we can treat it as a standard binary clustering problem, which can be solved using state-of-the-art clustering approaches such as K-means with two clusters. K-means is a popular clustering algorithm that partitions a set of observations (x_1, x_2, \dots, x_n) , where $x_i \in \mathbb{R}^d$, into $k \in n$ clusters, denoted as $S = S_0, \dots, S_k$, with the goal of minimizing the variance within each cluster. In the context of clustering outliers and inliers with $k = 2$, each observation x_i corresponds to a single value of $F_{\text{count},i}$ in the triple corresponding pair containing the corresponding pair (p_i, q_i) , and has dimension $d = 1$. We can classify the smaller cluster mean as the outlier.

D. Estimation rotation matrix

After the outlier removal, an estimator has been applied to infer the transformation between two point clouds. For further improving the robustness against the outlier, we randomly sample a subset of the Card, which contains a large fraction of the true inliers. In this work, we directly used the work from Kabsch–Umeyama algorithm [5], [6], [7], [8], which is a method for minimizing the root-mean-square deviation (RMSD) of the point corresponding pairs to find the optimal translation, rotation, and scaling, so that two sets of points can be aligned. As mentioned before, any estimator that can be applied to infer the transformation is suitable for our approach. We also demonstrate in V-C that our proposed transformation invariant descriptor can improve the performance of the existing algorithm.

The set of corresponding pair is defined as $\mathcal{P} = (p_1, \dots, p_n), p_i \in \mathbb{R}^3, \mathcal{Q} = (q_1, \dots, q_n), q_i \in \mathbb{R}^3$. The problem formulation is defined in (1). The translation t is computed with

$$t = \mu_p - R\mu_q, \quad (17)$$

Algorithm 1 Geometry based transformation invariant Point registration with clustering algorithm

Input: N putative corresponding pairs $(\mathbf{p}_i, \mathbf{q}_i)$, threshold c , H

- 1: Built fully connected corresponding pair Graph
- 2: Compute the F_1 for each $(\mathbf{p}_i, \mathbf{q}_i)$, and $F_{\text{count},i}$
- 3: Clustering the inliers/outliers with K-Means
- 4: **for** $h = 1 : H$ **do**
- 5: Randomly select a subset of inliers
- 6: Estimate the Rotation matrix, and translation vector
- 7: Compute the RMSE for the whole inlier sets
- 8: save the minimal RMSE, and its corresponding \mathbf{T}_{op^*}
- 9: **end for**
- 10: **return** \mathbf{T}_{op^*}

where μ_p and μ_q are centroids of point sets \mathcal{P} and \mathcal{Q} , respectively: $\mu_p = \frac{1}{n} \sum_{i=1}^n \mathbf{p}_i$, $\mu_q = \frac{1}{n} \sum_{i=1}^n \mathbf{q}_i$. Therefore the rotation and translation can be decoupled, and the equation (1) can be reformulated as

$$E_2 = \min \sum_{i=0}^N ((\mathbf{p}_i - \mu_p) - \mathbf{R}(\mathbf{q}_i - \mu_q))^2 \quad (18)$$

The equation (18) depends only on the rotation, and by expanding this equation, we can have

$$E_2 = \min \sum_{i=0}^N (\overline{\mathbf{p}}_i^T \overline{\mathbf{p}}_i + \overline{\mathbf{q}}_i^T \overline{\mathbf{p}}_i - 2\mathbf{R}\overline{\mathbf{q}}_i \overline{\mathbf{p}}_i^T) \quad (19)$$

with $\overline{\mathbf{p}}_i = (\mathbf{p}_i - \mu_p)$ and $\overline{\mathbf{q}}_i = (\mathbf{q}_i - \mu_q)$. Therefore, to minimize E_2 is equivalent to maximize

$$F = \sum_{i=0}^N (\mathbf{R}\overline{\mathbf{q}}_i \overline{\mathbf{p}}_i^T) = \text{trace} \left(\sum_{i=0}^N (\mathbf{R}\overline{\mathbf{q}}_i \overline{\mathbf{p}}_i^T) \right) = \text{trace}(\mathbf{R}\mathbf{A})$$

where $\mathbf{A} = \sum_{i=0}^N (\overline{\mathbf{q}}_i \overline{\mathbf{p}}_i^T)$. This is known as rotation search in computer vision or Wahba's problem in aerospace [27].

Lemma 1: For any positive definitive matrix \mathbf{A} and any orthogonal matrix \mathbf{B} , the following constraint is satisfied

$$\text{trace}(\mathbf{A}) \geq \text{trace}(\mathbf{B}\mathbf{A}) \quad (20)$$

Lemma 2: Let $\mathbf{R}_{D \times D}$ be an unknown orthonormal matrix and $\mathbf{A}_{D \times D}$ be a known real square matrix. Let $\mathbf{U}\mathbf{S}\mathbf{S}^T$ be a Singular Value Decomposition (SVD) of \mathbf{A} , where $\mathbf{U}\mathbf{U}^T = \mathbf{V}\mathbf{V}^T = \mathbf{I}$ and $\mathbf{S}\mathbf{S} = \text{d}(s_i)$ with $s_1 \geq s_2 \geq \dots \geq s_D \geq 0$. Then the optimal rotation matrix \mathbf{R} that maximizes $\text{tr}(\mathbf{R}\mathbf{A})$ is $\mathbf{R} = \mathbf{V}\mathbf{C}\mathbf{U}^T$, where $\mathbf{C} = \text{d}(1, 1, \dots, 1, \det(\mathbf{V}\mathbf{U}^T))$

The translation vector \mathbf{t} is updated using (17). We summarize the algorithm in 1. In general, the corresponding pair is given or computed by using the FPFH matching.

V. EXPERIMENTS

We evaluated the proposed approach against other state-of-the-art registration algorithms (FGR [20], RANSAC-1k (implemented in Open3D [28]), TEASER++ [19]) on a Stanford 3D Scanning Repository [17] and 3DMatch Dataset [18] regarding rotation error, translation error, and time. The algorithm presented in this paper is labeled as TAnIPR, TDoIPR, and TArIPR, respectively. Three different experiments were conducted. The tested point sets were downsampled for

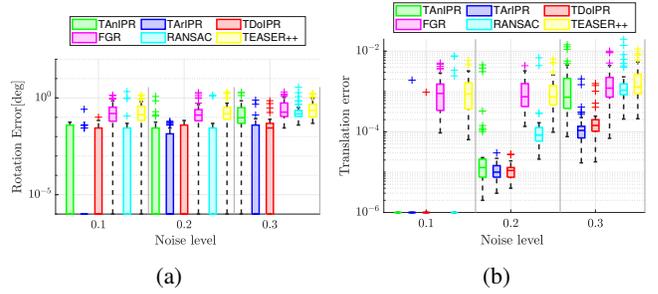


Fig. 2: Benchmark results for all algorithms on the Bunny dataset with three levels of Gaussian noise by ranging the outlier from 10%-90%. The corresponding point sets are founded using the FPFH descriptors.

TABLE II: The benchmark results for the proposed invariant descriptors on three point sets were evaluated with respect to the confusion matrix metrics: TPR (true positive rate), FPR (false positive rate), FNR (false negative rate), TNR (true negative rate), accuracy, F1-score, and time.

	TPR	FPR	FNR	TNR	Acc	F1 _{score}	time
TAnIPR	0.9613	0.0038	0.0387	0.9963	0.9757	0.9645	0.2901
TDoIPR	0.9944	0.0156	0.0056	0.9844	0.9843	0.9627	0.1681
TArIPR	0.9996	0.0058	0.0004	0.9942	0.9948	0.9845	0.5993

every algorithm into small-scale numbers (1000–5000) using voxel filtering to reduce the computational burden. Besides, we limited the number of putative corresponding pairs to $N = 1000$. All evaluations were performed on a laptop with a 2.6 GHz Intel Core i7-6700HQ and 16 GB of RAM. Note that in practical settings, ICP often fails to compute the correct transformation due to the fact that the initial guess is not in the basin of convergence of the optimal solution. Additionally, we did not compare our approach to deep learning-based pose estimation approaches, which typically require additional training for individual datasets.

A. Comparison Transformation Invariant Descriptors

This section evaluates Bunny, Buddha, and Dragon point clouds from the standard 3D scanning repository [17]. We scaled the point cloud into $[0, 1]^3$ to create the source point cloud, denoted as \mathcal{P} . The target point cloud, denoted as \mathcal{Q} is created by transforming the \mathcal{P} with a random transformation (\mathbf{R}, \mathbf{t}) , where $\mathbf{R} \in \text{SO}(3)$ is an arbitrary rotation matrix, and \mathbf{t} is bounded to $\|\mathbf{t}\| \leq 1$. Furthermore, three different noise level (0, 1/1000, 1/100) are added to each point. To generate a more realistic outlier correspondence, we replace a fraction of the q_i with other points on the same dataset but not included in \mathcal{Q} . It is a different approach than presented in [19], in which they uniformly sampled inside the sphere of radius. We evaluate the performance by increasing outlier rates from 10% to 90%. All statistics are computed over 40 times on a set of 40 random transformation matrices.. The table II summarizes the confusion matrix, accuracy, F-measure, and time by computing the mean of all three point Sets. The proposed TArIPR outperforms the other

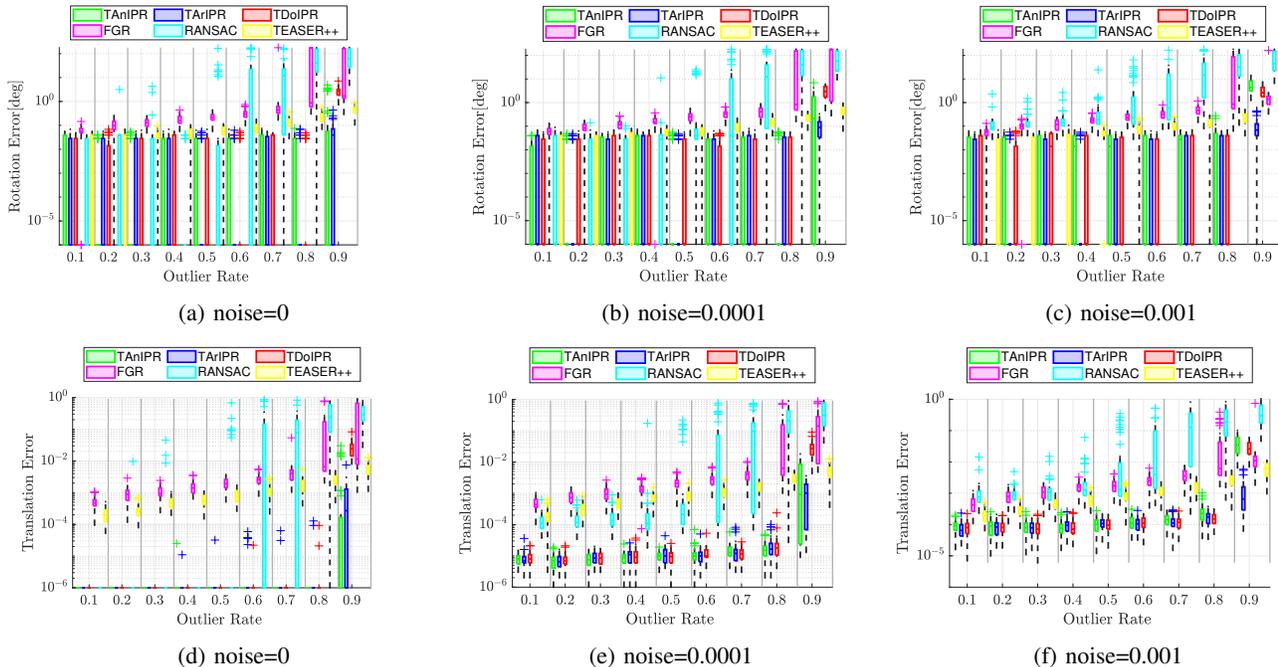


Fig. 3: Benchmark results were obtained for all algorithms on the Bunny dataset with three levels of Gaussian noise, while varying the outlier percentage from 10% to 90%. The corresponding point sets were generated randomly by introducing outliers.

TABLE III: Benchmark results are presented for all algorithms on three different point sets with three levels of Gaussian noise. The best result is highlighted in green color. The time is valuable only when the estimated transformation is correct.

		noise = 0.0000 Outlier: 10% – 90%			noise = 0.0001 Outlier: 10% – 90%			noise = 0.001 Outlier: 10% – 90%		
		$\epsilon_{\text{rotation}}$ [deg]	$\epsilon_{\text{translation}}$ [unit]	t [s]	$\epsilon_{\text{rotation}}$ [deg]	$\epsilon_{\text{translation}}$ [unit]	t [s]	$\epsilon_{\text{rotation}}$ [deg]	$\epsilon_{\text{translation}}$ [unit]	t [s]
Bunny [17]	RANSAC-1k	23.91890	0.1079	0.00714	23.4184	0.11172	0.0060	21.24770	0.1125	0.00633
	FGR	13.52710	0.0365	0.02405	17.1287	0.03566	0.0246	5.78010	0.0106	0.02518
	TEASER++	0.12122	0.0015	0.10733	0.1241	0.00154	0.1037	0.13055	0.0016	0.10543
	ours(TAnIPR)	0.07098	0.0003	0.29000	0.1223	0.00051	0.2925	0.63773	0.0050	0.29138
	ours(TDoIPR)	0.30984	0.0030	0.17034	0.3611	0.00341	0.1708	0.35333	0.0034	0.17120
ours(TArIPR)	0.01790	0.0001	0.58515	0.0217	0.00014	0.6044	0.02274	0.0002	0.60134	
Buddha [17]	RANSAC-1k	13.93100	0.0642	0.10656	13.7947	0.07142	0.0802	14.97420	0.0693	0.10033
	FGR	5.52130	0.0271	0.02463	9.9373	0.04819	0.0260	9.42950	0.0462	0.02443
	TEASER++	0.21389	0.0022	0.13121	0.2155	0.00212	0.1394	0.21020	0.0021	0.12843
	ours(TAnIPR)	0.49360	0.0045	0.28966	0.5152	0.00447	0.2889	1.14830	0.0082	0.28977
	ours(TDoIPR)	0.69047	0.0031	0.16699	0.6017	0.00316	0.1668	0.68674	0.0044	0.16678
ours(TArIPR)	0.32625	0.0023	0.60058	0.2109	0.00201	0.5994	0.33126	0.0025	0.60032	
Dragon [17]	RANSAC-1k	28.55990	0.1447	0.00394	22.2231	0.12044	0.0037	22.64380	0.1210	0.00423
	FGR	16.42770	0.0542	0.02400	13.4796	0.06122	0.0244	17.46470	0.0571	0.02436
	TEASER++	0.06505	0.0011	0.08062	0.0663	0.00109	0.0763	0.06849	0.0011	0.07912
	ours(TAnIPR)	0.03915	0.0001	0.28916	0.0668	0.00030	0.2896	0.43992	0.0050	0.29015
	ours(TDoIPR)	0.01817	0.0001	0.16641	0.0156	0.00007	0.1666	0.03920	0.0005	0.16670
ours(TArIPR)	0.01200	<0.0001	0.60219	0.0107	<0.00010	0.5990	0.01280	0.0001	0.60140	

two methods w.r.t TPR, Accuracy, and F1 score. However, the other two approaches are much more computationally efficient. In the context of point registration, the value of TPR should be higher since it reflects the true inliers numbers, which is returned as the Consensus set.

B. Benchmarking on Standard Datasets

We compared our approach with other state-of-the-art algorithms on three point sets with varying outlier rates

and three levels of noise. The evaluation metrics we used were rotation error, translation error, and computation time. The rotation error was defined as the geodesic distance between the rotation estimate \hat{R} and the ground-truth R° , while the translation error was defined as the 2-norm of the difference between the estimate \hat{t} and the ground-truth t° . The duration time was measured after searching for corresponding point sets. We used the same testing setup as described in Section V-A. Figure 3 shows the boxplots

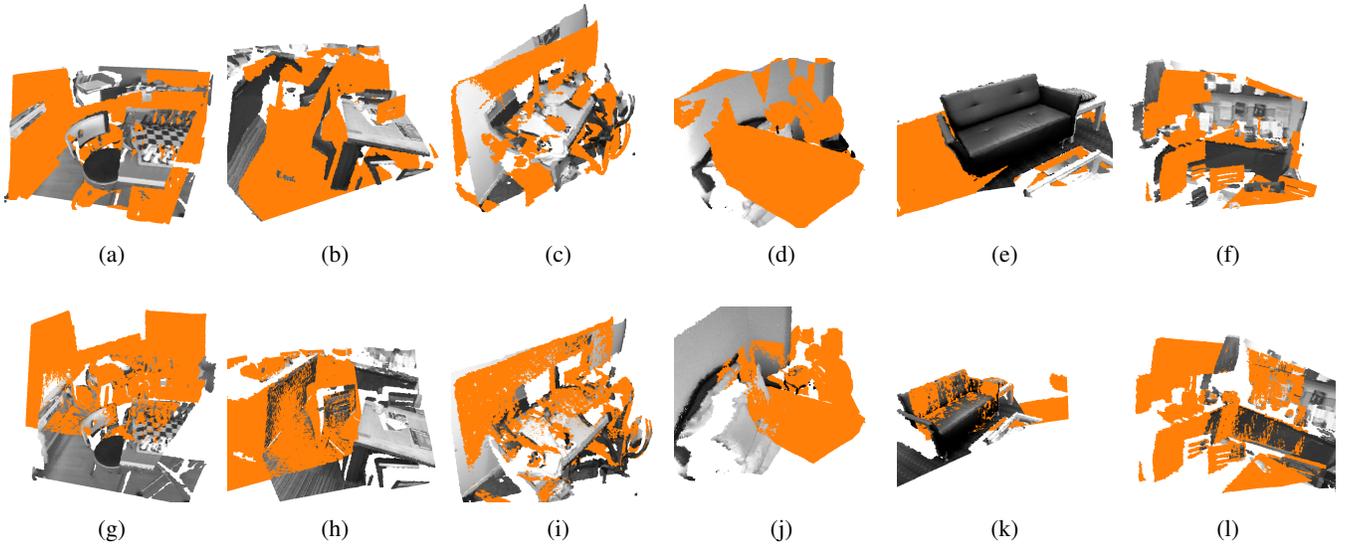


Fig. 4: Qualitative evaluation of the 3DMatch Dataset [18] was performed using TArIPR. The first row (a-f) shows the scans before alignment, while the second row (g-l) displays the results after alignment.

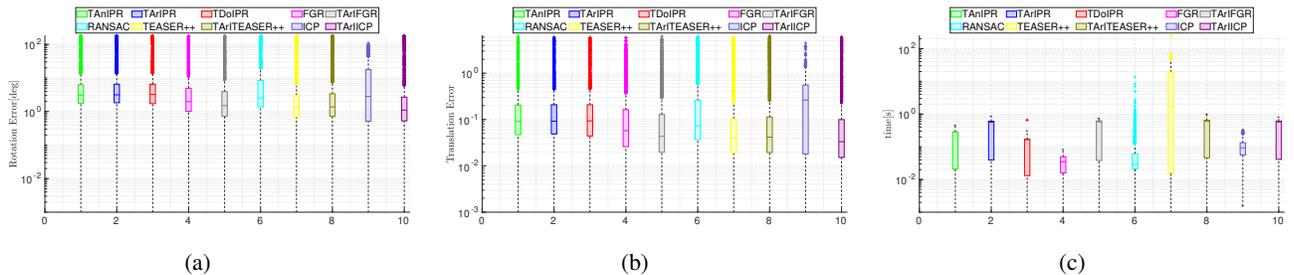


Fig. 5: Benchmark results for all algorithms on the 3DMatch dataset. TArIFGR, TArICP, and TArITEASER++ are the variant using the TArI descriptor as an outlier removal.

of the rotation errors and translation errors on the Bunny point set, and Table III provides a detailed comparison of the performance of each algorithm across all three datasets. Our results demonstrate that our approach outperforms the other methods in terms of robustness, accuracy, and efficiency.

1) **Outlier Robustness:** We examined the robustness of each algorithm against varying outlier rates, illustrated in Fig 3 and found that TAnIPR, TDoIPR, TArIPR, RANSAC, FGR, and TEASER++ can handle up to 70% outliers. However, our proposed algorithms (TAnIPR, TDoIPR, TArIPR) outperform the other approaches in terms of accuracy. Specifically, TArIPR can still produce a reasonable estimate up to 90%. In contrast, RANSAC, FGR, and TEASER++ start to fail after 70%.

2) **Noise Robustness:** Different levels of Gaussian noise can have a significant impact on the performance of point registration algorithms by changing the data distribution and significantly affecting the corresponding point set. To evaluate the algorithms under these conditions, we tested them with three different levels of Gaussian noise (0, 0.0001, and 0.001). As shown in Fig 3, TArIPR demonstrated the most robust performance on all three levels. The accuracy of FGR and TEASER++ gradually decreased as the noise level

increased. RANSAC had the worst performance among all the algorithms tested.

3) **FPFH for Corresponding Points Set:** In the previous section, we generated outliers by replacing true inliers randomly. To create a more realistic point registration scenario, we used FPFH to compute local descriptors for each point cloud and generated correspondences using a nearest-neighbor search. We utilized advanced matching strategies [23], [20] to find good correspondences and then evaluated all algorithms using the same putative correspondences. Fig 2 reports the rotation and translation errors at different noise levels. These results highlight the need for better keypoint detectors since even state-of-the-art local descriptors, such as FPFH, struggle to produce acceptable outlier rates in real problems. Our proposed algorithm outperforms the other approaches in terms of the metrics.

C. Evaluation on Scanning Dataset

Point cloud registration is widely used in robotics applications that require robust scan matching, such as 3D reconstruction and loop closure detection in SLAM. We evaluated our proposed algorithms using the 3DMatch dataset [18], which contains RGB-D scans from 62 real-world indoor

scenes. Without loss of generalization, we selected five indoor scenes and randomly chose two frames in each scenario for registration, resulting in a total of 2000 frame pairs. We evaluated all algorithms using the same testing setup described in section V-B.3, and we also added the proposed transformation invariant descriptor to the existing algorithm.

The results are reported in Fig. 5, which shows that FGR with TARl (TARlFGR) improves performance by reducing rotation and translation error. However, this comes at the cost of increased computational time. The variant of TARlTEASER++ shows similar performance, but with a significantly decreased computational time. Most notably, TARlICP shows similar performance to TEASER++, while ICP performs the worst in all experiments. This demonstrates that the proposed transformation invariant descriptor can improve robustness and is an efficient preprocessing step for robust registration.

In Fig. 4, we qualitatively demonstrate an excerpt of robust registration alignment using TARlPR on 3DMatch. These results highlight the potential of our proposed approach to improve the robustness and accuracy of point cloud registration in real-world scenarios.

VI. CONCLUSION

We have presented three different geometry-based, transformation-invariant descriptors for outlier removal in robust point registration, where outlier removal is formulated as a binary clustering problem. The proposed approach can be considered an efficient preprocessing step for robust point cloud registration by reducing the outlier ratio in the putative correspondences. Our approach is global, requiring no initialization, and extensive experiments have fully demonstrated its effectiveness. However, our approach has a limitation in that building a fully connected graph and computing the pair distance equation in (16) has a complexity of $\mathcal{O}(N^3)$, where N is the given number of putative corresponding pairs. We plan to address this issue in future work.

REFERENCES

- [1] J. Feldmar and N. Ayache, "Rigid and affine registration of smooth surfaces using differential properties," in *Proc. of the European Conf. on Computer Vision*, 2005, pp. 397–406.
- [2] F. Pomerleau, F. Colas, and R. Siegwart, "A review of point cloud registration algorithms for mobile robotics," *Foundations and Trends in Robotics*, vol. 4, no. 1, pp. 1–104, 2015.
- [3] S. Granger and X. Pennec, "Multi-scale EM-ICP: A fast and robust approach for surface registration," in *Proc. of the European Conf. on Computer Vision*, 2002, pp. 418–432.
- [4] A. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Proc. of Robotics: Science and Systems*, Seattle, USA, 2009.
- [5] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 32, no. 5, pp. 922–923, Sept. 1976.
- [6] —, "A discussion of the solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 34, no. 5, pp. 827–828, Sept. 1978.
- [7] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.
- [8] 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*, vol. PAMI-9, no. 5, pp. 698–700, Sept. 1987.
- [9] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 2009, pp. 3212–3217.
- [10] Z. Gojcic, C. Zhou, J. D. Wegner, and A. Wieser, "The perfect match: 3D point cloud matching with smoothed densities," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5545–5554.
- [11] 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.
- [12] C. Olsson, F. Kahl, and M. Oskarsson, "Branch-and-bound methods for Euclidean registration problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 783–794, 2008.
- [13] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2241–2254, 2016.
- [14] R. I. Hartley and F. Kahl, "Global optimization through rotation space search," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 64–79, 2009.
- [15] J. Lin, M. Rickert, and A. Knoll, "Deep hierarchical rotation invariance learning with exact geometry feature representation for point cloud classification," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2021, pp. 9529–9535.
- [16] A. Parra and T.-J. Chin, "Guaranteed outlier removal for point cloud registration with correspondences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, pp. 1–1, 11 2017.
- [17] G. Turk and B. Mullins, "Large geometric models archive," https://www.cc.gatech.edu/projects/large_models/, Mar. 2019.
- [18] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3DMatch: Learning local geometric descriptors from RGB-D reconstructions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [19] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Trans. Rob.*, vol. 37, no. 2, pp. 314–333, 2021.
- [20] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *Proceedings of the European Conference on Computer Vision*, 2016.
- [21] 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*, vol. PAMI-9, no. 5, pp. 698–700, 1987.
- [22] P. 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.
- [23] J. Lin, M. Rickert, and A. Knoll, "6D pose estimation for flexible production with small lot sizes based on CAD models using Gaussian process implicit surfaces," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 10 572–10 579.
- [24] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing ICP variants on real-world data sets," *Autonomous Robots*, vol. 34, no. 3, pp. 133–148, 2013.
- [25] F. Tombari, S. Salti, and L. Di Stefano, "Performance evaluation of 3D keypoint detectors," *International Journal of Computer Vision*, vol. 102, no. 1, pp. 198–220, 2013.
- [26] Á. P. Bustos, T. Chin, F. Neumann, T. Friedrich, and M. Katzmann, "A practical maximum clique algorithm for matching with pairwise constraints," *CoRR*, vol. abs/1902.01534, 2019. [Online]. Available: <http://arxiv.org/abs/1902.01534>
- [27] G. Wahba, "A least squares estimate of satellite attitude," *SIAM Review*, vol. 7, no. 3, pp. 409–409, 1965.
- [28] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.