Cheng et al. *Visual Intelligence* (2023) 1:20
https://doi.org/10.1007/s44267-023-00022-x

Visual Intelligence

RESEARCH Open Access

Sampling locally, hypothesis globally: accurate 3D point cloud registration with a RANSAC variant

Yuxin Cheng[1], Zhiqiang Huang[2], Siwen Quan[3], Xinyue Cao[1], Shikun Zhang[1] and Jiaqi Yang[1*]

Abstract

Correspondence-based six-degree-of-freedom (6-DoF) pose estimation remains a mainstream solution for 3D point cloud registration. However, the heavy outliers pose great challenges to this problem. In this paper, we propose a random sample consensus (RANSAC) variant based on sampling locally and hypothesis globally (SLHG) for 6-DoF pose estimation and 3D point cloud registration. The key novelties are efficient sampling by guiding the sampling process locally and accurate pose estimation by generating hypotheses with global information. SLHG first generates a correspondence subset via compatibility clustering on the initial set. Second, locally guided graph sampling is performed. Third, 6-DoF hypotheses are generated by incorporating global information with a voting scheme. The best hypothesis serves as the estimation result by repeating the second and third steps. Extensive experiments on four popular datasets and comparisons with state-of-the-art methods confirm that: SLHG manages to 1) achieve accurate registrations with a few iterations, and 2) yield better accuracy performance than most competitors.

Keywords: 3D point cloud, 3D registration, Random sample consensus, Pose estimation

1 Introduction

Registration of 3D point clouds is a fundamental problem in 3D computer vision, which contributes significantly to 3D reconstruction [1], 3D object recognition [2], remote sensing [3] and odometry [4]. The objective of the 3D registration method is to obtain the relative six-degree-of-freedom (6-DoF) pose between the source point cloud and the target point cloud to achieve alignment.

To perform accurate and robust point cloud registration, two different types of methods have been investigated, correspondence-free and correspondence-based methods. The 6-DoF estimation of correspondence-free methods does not rely on the local feature correspondences computed between point cloud pairs, and tends to first design an error function and then optimize it in an iterative manner. One of the most representative methods is the iterative closest point (ICP) method [5]. It performs registration by iteratively searching the closest points of the source point cloud in the target one and optimizing the point-to-point distance error. Many variants of ICP have been proposed with different error functions and closest point definitions [6–11]. Other typical correspondence-free methods include those based on branch-and-bound (BnB) [12, 13] and normal distribution transform (NDT) [14]. However, common drawbacks of correspondence-free methods include the high computational complexity and the risk of being trapped in local minima.

For correspondence-based methods, random sample consensus (RANSAC) [15] and its variants [16–20] are popular methods, because RANSAC-fashion methods are technically simple and robust to common nuisances. With abundant correspondences between point cloud pairs, RANSAC randomly samples three correspondences in each iteration, employs singular vector decomposition to

*Correspondence: jqyang@nwpu.edu.cn
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, 710129, China
Full list of author information is available at the end of the article

Springer

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

generate hypothesis and performs hypothesis evaluation algorithm to determine the optimal registration solution. In challenging registration cases with sparse correct correspondences, RANSAC usually requires a huge number of iterations and the accuracy still may not be guaranteed. Some methods attempt to resolve this issue by devising either more efficient sampling techniques [21–24] or more reliable performance evaluation metrics [16, 25, 26]. However, most of them still fail to achieve accurate 3D registration with a few iterations.

To this end, we propose a RANSAC variant based on sampling locally and hypothesis globally (SLHG) to achieve accurate 3D point cloud registration with a few iterations. The key insight of our algorithm is that we perform locally guided sampling in a compatibility graph to ensure the efficiency and hypothesis globally to improve the accuracy. SLHG first constructs several clusters among the raw correspondences based on the compatibility features, and the maximum cluster is modeled as a graph. Second, guided local sampling is performed in the graph based on a novel metric to evaluate the confidence of samples. Third, samples are mutually voted for each other to mine global information for hypothesis generation. The second and third steps are repeated to find the best hypothesis. Experiments are performed on four datasets, i.e., U3M [27], 3DMatch, 3DLoMatch [28], and RESSO [29], which incorporate different application contexts and data modalities, to validate the efficacy of our proposed method. In summary, this paper presents two main contributions:

1) A new metric is proposed for guided sampling in a compatibility graph, which allows the mining of good samples in an early stage.
2) A RANSAC variant called SLHG is proposed for 6-DoF pose estimation and 3D point cloud registration, which samples locally and hypothesizes globally. It is able to achieve accurate registrations with a few iterations.

The rest of the paper is structured as follows. Section 2 provides a review of correspondence-based and correspondence-free registration methods. Section 3 details our proposed SLHG method. Section 4 presents the experiments deployed to validate the effectiveness of our method with necessary discussions. Finally, Sect. 5 draws the conclusions and presents potential future research directions.

2 Related work

In this section, we present an overview of correspondence-free and correspondence-based 3D point cloud registration methods.

2.1 Correspondence-free registration

The current representative correspondence-free 3D registration methods are primarily based on the iterative closest point (ICP), branch and bound (BnB), and normal distributions transform (NDT).

ICP registration. ICP is one of the earliest 3D registration methods, proposed by Besl and McKay [5]. It employs a weighted Euclidean distance error function with an initial 6-DoF pose matrix, and iteratively solves the optimal 6-DoF pose. Some works focus on improving the closest point searching phase. Masuda et al. [30] proposed a method of randomly sampling points in each iteration of ICP. Weik [31] introduced the method of selecting points with high intensity gradient points in the convergence process of ICP. Turk [32] proposed removing boundary points in point selection to improve the accuracy. Some works improved ICP in the point matching stage. For instance, Godin et al. [33] proposed adding surface normal information based on Euclidean distance in closest match point searching. There are also several variants of ICP that define different error functions such as point-to-plane [6] and plane-to-plane [8] distance functions.

BnB registration. The BnB [34] method was applied to 3D point cloud registration by Olsson et al [12]. BnB registration depends on point-to-point, point-to-line, and point-to-plane correspondences to constrain the upper and lower bounds of the registration error function. One typical work is GO-ICP [13], which transfers the BnB method into ICP registration to enhance its efficiency. Liu et al. [35] proposed increasing the efficiency of BnB-based registration methods by decoupling translation and rotation separately, which performs better than GO-ICP. Brown et al. [36] introduced deterministic nested BnB (BnB-D) and probabilistic nested BnB (BnB-P), which use inner BnB to compute the upper and lower bounds to guide rotation searching to achieve faster convergence and the global optimal solution. However, BnB-based methods usually suffer from high computational complexity.

NDT registration. Magnusson et al. [37] proposed adapting NDT [38] to the field of 3D point cloud registration. For 3D registration in NDT, an objective function is set to measure the probability density of each target point in the source point cloud, and the Newton optimization method is applied iteratively to obtain the optimal solution. Since the NDT method is not limited by local feature correspondence computation, it is efficient for registration. Many subsequent approaches have been proposed, such as Fast-NDT [39], which uses the global description method to increase reliability and achieve real-time performance. Stoyanov et al. [40] presented the L_2 distance-based 3D-NDT method to formulate the objective function in NDT-based registration, which improves the registration accuracy and efficiency. Shi et al. [41] combined the ICP and NDT methods and proposed the NDT-ICP registration method, and then employed the ICP fine registration to achieve better accuracy.

Correspondence-free methods usually require a good initialization and hold limited efficiency performance because the solution search space $SE(3)$ is large without point-to-point correspondences.

2.2 Correspondence-based registration

Correspondence-based methods aim to minimize a registration error metric by correspondence searching and transformation estimation.

RANSAC registration. The RANSAC registration method remains the mainstream solution among various correspondence-based methods. Fischler et al. [15] proposed RANSAC to address model fitting problems, which shows significant effectiveness in the 3D registration task. However, in order to overcome the insufficiency of RANSAC in terms of efficiency and accuracy, a number of RANSAC variants [16–20] for 3D point cloud registration have been proposed. Rusu et al. [42] introduced the sample consensus-based initial alignment (SAC-IA) method, which employs a distance threshold to constrain the correspondence sample and uses the Huber penalty function for hypothesis evaluation. Yang et al. [43] developed a global distance metric in point cloud in assessing the accuracy of the estimated 6-DoF pose hypothesis. Chen et al. [44] designed a second-order spatial compatibility measurement (SC^2) to enhance the outlier rejection performance in correspondence sampling. To reduce RANSAC iterations, Yang et al. [22] proposed a guided correspondence sampling method that employs compatibility triangles as samples to speed up the sampling process in order to improve the registration efficiency and accuracy. Yang et al. [45] developed a mutual voting mechanism for high confidence correspondence ranking that involves graph construction among correspondences and has greatly improved inlier sampling performance. There are also two-point and one-point based sampling consensus methods. For instance, Yang et al. [46] introduced a two-point based sample consensus approach with global constraint (2SAC-GC). Quan et al. [21] presented the compatibility-guided sampling consensus (CG-SAC) approach, which involves point normal information for hypothesis generation. For one-point based RANSAC, Quan et al. [47] proposed a globally constrained one-point based sample consensus method that samples one correspondence with local reference frames per iteration. A comprehensive survey of RANSAC methods for 3D registration can be found in [48]. Nonetheless, we will experimentally show that most existing RANSAC methods still fail to achieve accurate registrations efficiently.

Learning-based registration. With the rapid development of deep learning, some trials have been completed toward performing registration with deep neural networks. Aoki et al. [26] proposed PointNetLK registration method, which employs PointNet [49] for local feature extraction. Bai et al. [24] designed PointDSC, which leverages a deep neural network based on spatial consistency for outlier rejection. Qin et al. [50] developed GeoTransformer, which can learn the geometric features of points and implement robust correspondence searching in low overlapping registration scenarios. Pais et al. [51] established a learning-based registration architecture called 3DRegNet, which includes an inlier-outlier classification block and a pose regression registration block for registration. Jiang et al. [52] formulated outlier rejection as a classification problem and proposed involving the variational Bayesian long-range dependency property in aggregating the inlier and outlier correspondence distinction. Lombardi et al. [53] proposed a variant of 3DRegNet named 3DReg-i-Net to achieve efficient registration with a lightweight model. Although learning-based registration methods have achieved promising results, the lack of sufficient training data may limit their efficacy in many practical applications.

Other methods. Aiger et al. [54] proposed the 4PCS method, which uses congruent co-planar four points between two point clouds to generate the pose hypothesis. Drost et al. [55] proposed a method to model the description globally based on all orientated point pair features and used local correspondences to obtain the 6-DoF pose hypothesis. Guo et al. [56] proposed a cluster-based method that estimates the 6-DoF pose by projecting the rotation and the translation spaces independently. Buch et al. [57] proposed using one correspondence to vote in a rotational sub-group of 6-DoF, and the final 6-DoF pose is obtained by combining the rotation sub-group with the highest kernel density obtained from the voting. Tombari et al. [58] proposed a voting-based method to generate the best hypothesis by obtaining voters from a 3D Hough space. These correspondence-based methods typically rely on point orientations such as normals and local reference frames, which are shown to be sensitive to noise and limited overlap [48].

Consequently, we focus our research on the geometric RANSAC-fashion method. We propose a novel RANSAC-based SLHG estimator. It constructs a graph based on a novel metric during correspondence sampling and incorporates mutual voting and global hypothesis generation in 6-DoF pose estimation, producing more accurate registrations with a few iterations.

3 Method

The pipeline of our method is represented in Fig. 1. It consists of four main steps: cluster construction, sampling locally, hypothesis globally, and hypothesis evaluation, where the last three steps are performed iteratively. These four stages play the following roles in our method.

1) *Cluster construction*. SLHG performs clustering for each correspondence based on the compatibility of each pair of correspondences, and the cluster with

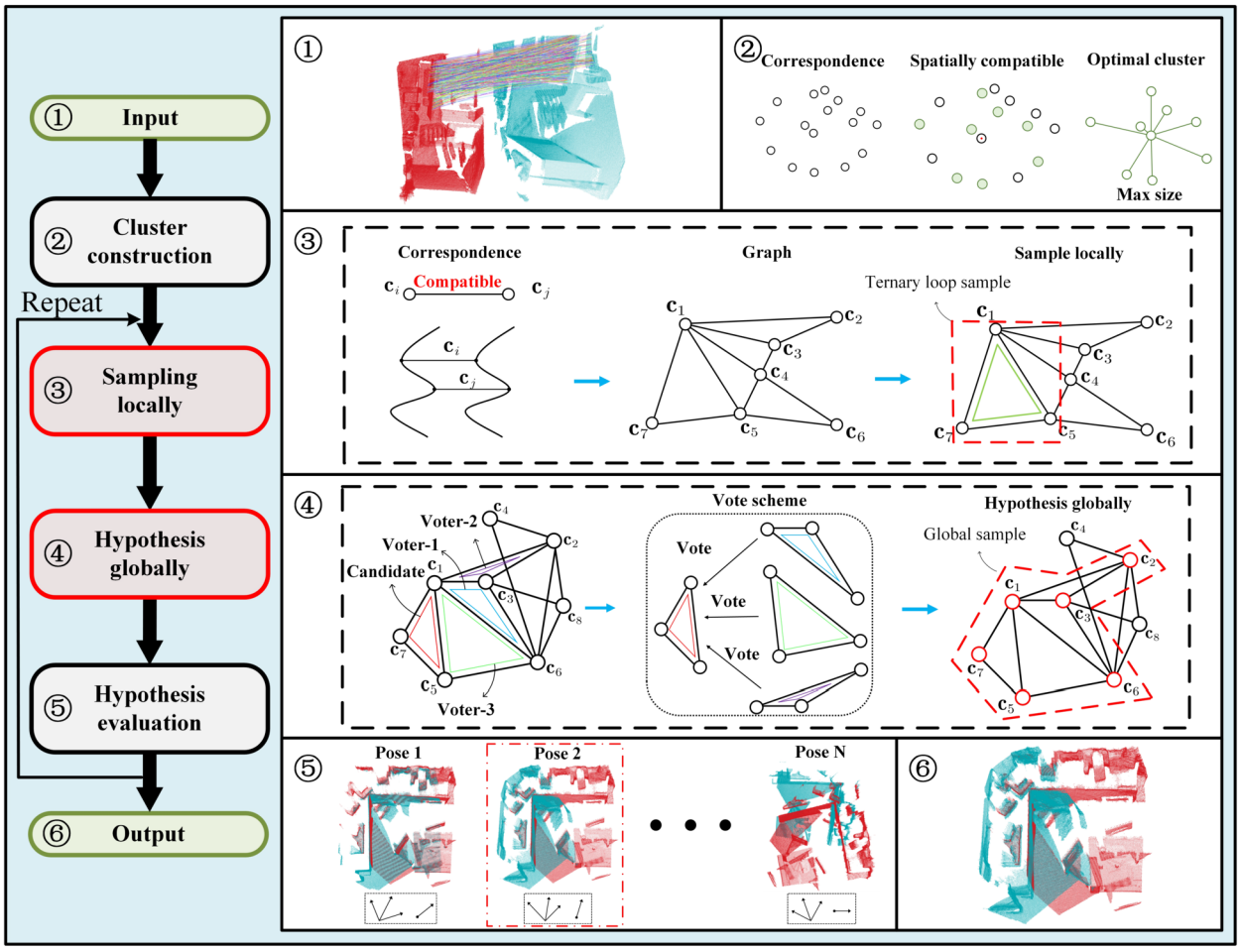


Figure 1 The pipeline of the proposed SLHG method. First, a cluster with the maximum number of correspondences that are compatible with each other is constructed. Second, a graph is constructed for the cluster and local guided sampling is performed in the graph. Third, for each local sample, we perform voting to mine globally consistent information and generate hypotheses from a global perspective. Fourth, each hypothesis is evaluated to determine the best one, which is the output of SLHG

the largest size is retained. Coping with this cluster is supposed to be more efficient than dealing with the initial correspondence set.

2) *Sampling locally*. The retained cluster will be modeled as a graph, in which local ternary loops serve as samples. In this step, we propose a new metric to rank ternary loops and guide the sampling process in order to reduce iterations.
3) *Hypothesis globally*. For each ternary loop sample, we perform voting in the graph to mine more stable global information for hypothesis generation. After voting, the singular value decomposition (SVD) algorithm is used to generate the 6-DoF pose.
4) *Hypothesis evaluation*. Finally, we use the mean absolute error (MAE) metric [59] to quantitatively evaluate the hypotheses and find the best one as the output.

To improve the readability, necessary notations are displayed in Table 1. The input of our method is the correspondence set $\mathbf{C}$ between the source point cloud $\mathbf{P}^s$ and the target point cloud $\mathbf{P}^t$, when the output is a 6-DoF pose matrix $\mathbf{M}$ that aligns $\mathbf{P}^s$ to $\mathbf{P}^t$.

3.1 Cluster construction

Given a set of initial feature correspondences $\mathbf{C}$, we address it in two steps to generate the maximum cluster, as shown in Fig. 2. First, we quantitatively analyze the compatibility between every two correspondences. Given two correspondences $\mathbf{c}_i$ and $\mathbf{c}_j$, the compatibility score between them is defined as:

$$s_{comp}(\mathbf{c}_i, \mathbf{c}_j) = \exp\left(-\frac{d(\mathbf{c}_i, \mathbf{c}_j)^2}{2d_{\text{thresh}}^2}\right), \tag{1}$$

Table 1 Notations

$\mathbf{P}^s$	$\mathbf{P}^s \in \mathbb{R}^3$: source point cloud
$\mathbf{P}^t$	$\mathbf{P}^t \in \mathbb{R}^3$: target point cloud
$\mathbf{p}^s$	$\mathbf{p}^s \in \mathbf{P}^s$: point in source point cloud
$\mathbf{p}^t$	$\mathbf{p}^t \in \mathbf{P}^t$: point in target point cloud
$\mathbf{c}$	$\mathbf{c}_i \in \mathbb{R}^6$: a pair correspondence of $\mathbf{p}^s$ and $\mathbf{p}^t$
$\mathbf{C}$	Correspondence set $\mathbf{C} = \{\mathbf{c}_i, \mathbf{c}_j, \ldots, \mathbf{c}_k\}$
$\mathbf{s}(\mathbf{c})$	$\mathbf{s}(\mathbf{c}) = \{\mathbf{c}_i, \mathbf{c}_j, \ldots, \mathbf{c}_k\}$: cluster for $\mathbf{c}$
$\mathbf{l}$	Ternary loop $\mathbf{l}_{ijk} = \{\mathbf{c}_i, \mathbf{c}_j, \mathbf{c}_k\}$
$\mathbf{L}$	Ternary loop set $\mathbf{L} = \{\mathbf{l}_{ijk}, \mathbf{l}_{mno}, \ldots, \mathbf{l}_{xyz}\}$
$\mathbf{M}$	$\mathbf{M} \in SE(3)$: 4 × 4 rigid transformation matrix
$\mathbf{R}$	$\mathbf{R} \in SO(3)$: 3 × 3 rotation matrix
$\mathbf{t}$	$\mathbf{t} \in R^3$: 3 × 1 translation vector

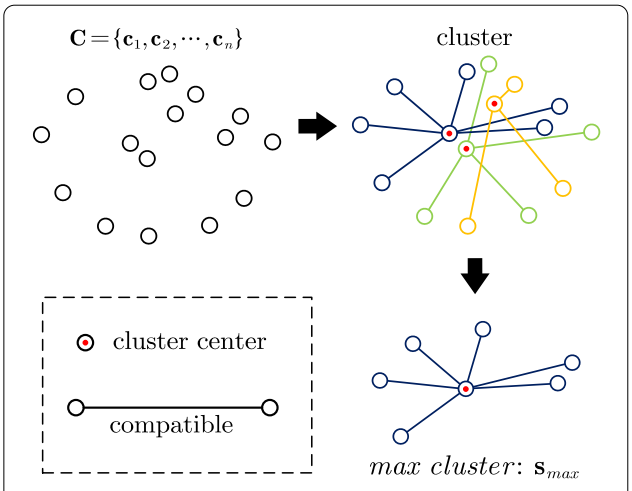


Figure 2 Illustration of the maximum cluster construction. First, each correspondence will find its compatible neighbors in the correspondence and form a cluster. Second, the cluster with the maximum number of correspondences is retained

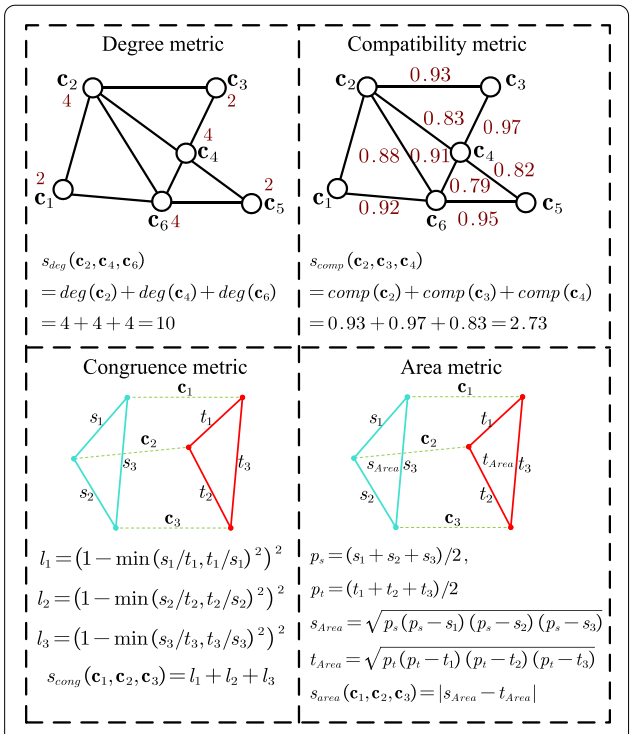


Figure 3 Illustration of four metrics for evaluating a ternary loop in a graph

where $d(\mathbf{c}_i, \mathbf{c}_j) = |\|\mathbf{p}_i^s - \mathbf{p}_j^s\| - \|\mathbf{p}_i^t - \mathbf{p}_j^t\||$ and d_{thresh} is a distance threshold. We empirically set d_{thresh} to 5 pr. Here, the unit 'pr' denotes the point cloud resolution, which is the average distance of each point in a point cloud to the closest neighbor to it.

Second, we make the assumption that inliers are only compatible with inliers, and inliers are supposed to form a cluster. Therefore, we establish a cluster $\mathbf{s}(\mathbf{c})$ for each correspondence $\mathbf{c}$ and ensure that elements in the cluster are compatible with $\mathbf{c}$. Then, we rank all clusters based on the cluster size and keep the maximum one $\mathbf{s}_{\max}(\mathbf{c})$. The scale of the cluster is supposed to be smaller than that of $\mathbf{C}$, while the inlier ratio is usually greater (verified in Sect. 4).

3.2 Sampling locally

In this step, we perform local sampling as follows. First, the maximum cluster $\mathbf{s}_{\max}(\mathbf{c})$ is modeled as an undirected graph $\mathcal{G}$. Specifically, in $\mathcal{G}$, nodes denote correspondences and edge link compatible nodes based on Eq. (1). Second, we serve local ternary loops in the graph as samples, because at least three correspondences are required for 6-DoF pose generation. Third, guided sampling is performed by ranking all ternary loops based on a new metric defined as follows.

In particular, we can mine the following four metrics to evaluate the priority of a ternary loop in $\mathcal{G}$, as illustrated in Fig. 3.

1) *Degree metric*: the sum of vertex degrees of a ternary loop.
2) *Compatibility metric*: the sum of the compatibility scores of all edges in a ternary loop.
3) *Congruence metric*: the congruent degree between the source and the target spatial triangles in two point clouds formed by a ternary loop.
4) *Area metric*: the area difference of the source and the target spatial triangles in two point clouds formed by a ternary loop.

Here, we propose a spatial consensus metric (SCM) by combining the congruence metric and compatibility metric as:

$$\text{SCM}(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3) = \exp\left(-s_{cong}(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3)\right) + s_{comp}(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3). \tag{2}$$

Intuitively, SCM combines constraints from the compatibility space (graph $\mathcal{G}$) and the spatial space (the source and target point clouds), and achieves a balance. Moreover, by

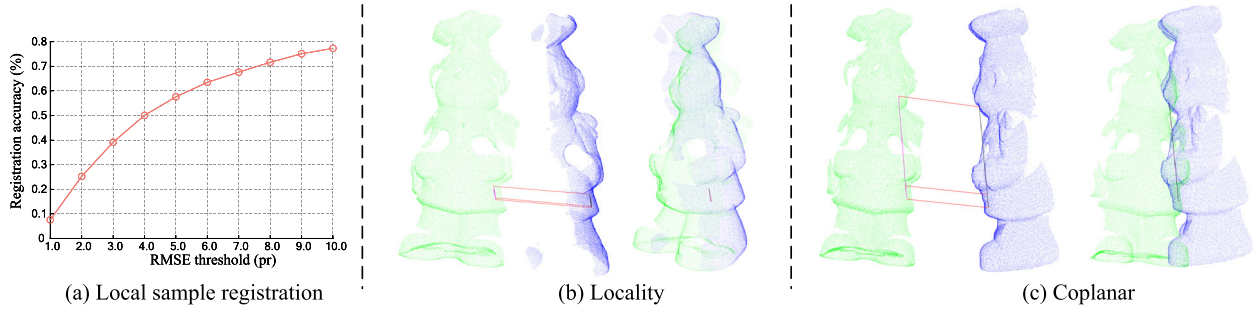


Figure 4 Illustration of the limitation of hypothesis locally. (a) The registration accuracy achieved by 6-DoF poses generated using correct local samples; (b) and (c) present two exemplar cases that local samples often result in inaccurate registrations

independently checking the performance of the four metrics, the compatibility metric and congruence metric are more stable. Then, we rank all ternary loops in $\mathcal{G}$ to achieve guided sampling.

3.3 Hypothesis globally

The ternary loop only contains local information. In some scenarios, we find that correct ternary loops (consisting of three inliers) do not always yield correct 6-DoF poses. In particular, we sample approximately 4000 correct ternary loops from the UWA 3D modeling dataset [27] and generate hypotheses based on these ternary loops to perform registration. As shown in Fig. 4(a), most of them are not able to achieve accurate registration, and more than 20% of the correct ternary loops fail to achieve successful registrations with a loose judging threshold, i.e., the threshold equals 10 pr. This can be explained by the fact that small scale and nearly co-planar ternary loops, which even consist of inliers, often fail to generate correct hypotheses. This motivates us to gather global information for more accurate hypothesis generation.

Specifically, global information is gathered via a voting process. Let the ternary loop $\mathbf{l}_{ijk} = (\mathbf{c}_i, \mathbf{c}_j, \mathbf{c}_k)$ be a candidate, voters for $\mathbf{l}_{ijk}$ are those ternary loops sharing at least one common node with $\mathbf{l}_{ijk}$. As such, each candidate will have a voter set $\mathbf{L}_{ijk}$ and the voters in $\mathbf{L}_{ijk}$ are ordered in a descending order based on the SCM score. Next, the correspondence set $\mathbf{C}_{ijk}$ incorporating global information will be generated for $\mathbf{l}_{ijk}$ as $\mathbf{C}_{ijk} = \{\mathbf{c} \mid \mathbf{c} \in \mathbf{l}, \mathbf{l} \in \mathbf{L}_{ijk}\}$. The confidence score for $\mathbf{C}_{ijk}$ is given as:

$$s(\mathbf{C}_{ijk}) = \mathrm{SCM}(\mathbf{l}_{ijk}) + \frac{1}{n}\sum\nolimits_{\mathbf{l}\in\mathbf{L}_{ijk}} \mathrm{SCM}(\mathbf{l}), \tag{3}$$

where $n = \max\{|\mathbf{L}_{ijk}|, V_{\mathrm{thresh}}\}$, and V_{thresh} is a threshold to limit the number of voters in case that bottom-ranked voters are involved. After voting, we rank all $\mathbf{C}_{ijk}$ by their confidence scores in a descending order and compute the hypotheses via the SVD algorithm.

Notably, in low inlier ratio cases, we find that even top-ranked voters may turn out unreliable. Therefore, we rely on a threshold T_{thresh} representing the number of ternary loops in the graph $\mathcal{G}$ to address this issue. More specifically, when the number of ternary loops in $\mathcal{G}$ is smaller than T_{thresh}, we judge it as an inlier ratio case and compute the confidence score of $\mathbf{C}_{ijk}$ as $s(\mathbf{C}_{ijk}) = \mathrm{SCM}(\mathbf{l}_{ijk})$.

3.4 Hypothesis evaluation

We employ the efficient mean absolute error (MAE) metric [59] to evaluate all hypotheses, which is defined as:

$$S_{\mathrm{mae}}(\mathbf{M}_i) = \sum_{j=1}^{|\mathbf{s}_{\max}(\mathbf{c})|} f(\mathbf{c}_j), \tag{4}$$

$$f(\mathbf{c}_j) = \begin{cases} \frac{|e(\mathbf{c}_j) - d_{\mathrm{thresh}}|}{d_{\mathrm{thresh}}}, & \text{if } e(\mathbf{c}_j) < d_{\mathrm{thresh}}, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where $e(\mathbf{c}_j) = \|\mathbf{R}_i\mathbf{p}_j^s + \mathbf{t}_i - \mathbf{p}_j^t\|$ represents the transformation error of $\mathbf{c}_j$. After N_{iter} iterations, the hypothesis yielding the highest MAE score is retained to perform registration.

In summary, our method is supposed to be efficient and accurate (verified in Sect. 4) for the following reasons: 1) a cluster is extracted from the initial set to reduce the input size while increasing the inlier ratio; 2) guided local sampling can reduce the iteration counts; and 3) global hypothesis generation can gather more convincing global information.

4 Experimental results

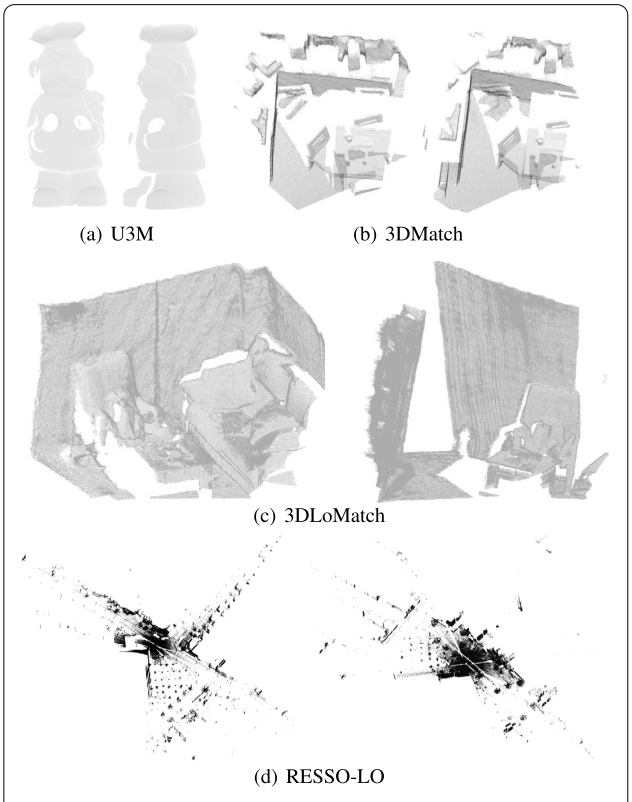
This section demonstrates the effectiveness and superiority of our proposed SLHG method in detail through various experiments.

4.1 Experimental setup

Datasets. To fully demonstrate the effectiveness of our proposed method under different scenarios, we consider four point cloud datasets, namely UWA 3D models (U3M) [27]

Table 2 Experimental datasets and their properties

Dataset	Scenario	Nuisances	Modality	#Matching pairs	Avg. inlier ratio
U3M [27]	Object Registration	Limited overlap, self-occlusion	LiDAR	496	0.1480
3DMatch [28]	Indoor Scene Registration	Limited overlap, self-occlusion, real noise	Kinect	1623	0.1143
3DLoMatch [60]	Indoor Scene Registration	Extreme Limited overlap, self-occlusion, real noise	Kinect	1781	0.0257
RESSO-LO [29]	Outdoor Scene Registration	Limited overlap, self-occlusion, real noise	LiDAR	187	0.1240



(a) U3M (b) 3DMatch (c) 3DLoMatch (d) RESSO-LO

Figure 5 Exemplar views of point cloud pairs in four experimental datasets

for object registration, 3DMatch [28] and 3DLoMatch [60] for indoor scene registration (3DLoMatch only includes point cloud pairs with less than 30% overlap), and the Real-world Scans with Small Overlap (RESSO-LO) [29] for large outdoor scene registration. The main properties of these four datasets are shown in Table 2 and exemplar views of point cloud pairs in these datasets are displayed in Fig. 5. One can see that the experimental datasets cover a variety of data modalities and application scenarios.

Metrics. We evaluate the performance of our method by computing the registration accuracy (RA). In our experiment, the root mean square error (RMSE) with point cloud resolution (pr) is used for evaluation on the U3M and RESSO-LO datasets as in [22], and the rotation error (RE) and translation error (TE) are used for evaluation on the 3DMatch and 3DLoMatch datasets according to [24]. For the RMSE error metric, given a pair of corresponding points $\mathbf{c}_i = (\mathbf{p}^s, \mathbf{p}^t)$, their registration error is defined as:

$$\epsilon\left(\mathbf{p}^s, \mathbf{p}^t\right) = \left\| \mathbf{R}_{\text{gt}}\mathbf{p}^s + \mathbf{t}_{\text{gt}} - \mathbf{p}^t \right\|, \tag{6}$$

and RMSE is calculated based on point-wise error $\epsilon(\mathbf{p}^s, \mathbf{p}^t)$ as:

$$\text{RMSE} = \frac{\sqrt{\sum_{(\mathbf{p}^s,\mathbf{p}^t)\in\mathbf{C}_{\text{gt}}} \epsilon(\mathbf{p}^s, \mathbf{p}^t)^2}}{|\mathbf{C}_{\text{gt}}|}, \tag{7}$$

where $\mathbf{R}_{\text{gt}}$ and $\mathbf{t}_{\text{gt}}$ represent the ground-truth rotation matrix and translation vector, respectively; $\mathbf{C}_{\text{gt}}$ represents the ground-truth set of corresponding points between two point clouds. Under this error metric, we define a registration result as correct when its RMSE error is lower than a judging threshold t_{rmse}.

For the rotation error and the translation error, we compute the them according to $\mathbf{R}_{\text{gt}}$ and $\mathbf{t}_{\text{gt}}$ as:

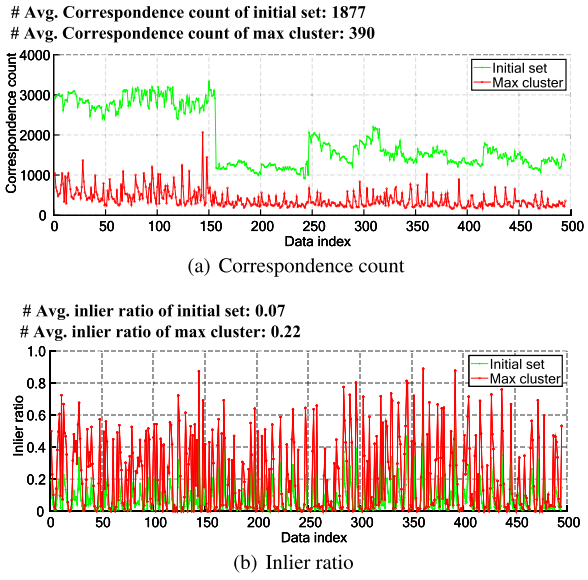
$$\begin{cases} \epsilon(\mathbf{R}_{\text{est}}) = \arccos\frac{\text{Tr}(\mathbf{R}_{\text{est}}^T\mathbf{R}_{\text{gt}})-1}{2}, \\ \epsilon(\mathbf{t}_{\text{est}}) = \|\mathbf{t}_{\text{est}} - \mathbf{t}_{\text{gt}}\|_2, \end{cases} \tag{8}$$

where $\mathbf{R}_{\text{est}}$ and $\mathbf{t}_{\text{est}}$ represent estimated rotation matrix and translation vector, respectively; $\epsilon(\mathbf{R}_{\text{est}})$ and $\epsilon(\mathbf{t}_{\text{est}})$ represent the rotation error and the translation error, respectively.

RA is then given as the ratio of successfully registered point clouds in a dataset.

Compared methods. In our experiments, both geometry-only and learned methods are involved for comparison with SLHG. For geometry methods, Go-ICP [13], RANSAC [15], SAC-COT [22], SAC-IA [42], CGSAC [21], GC1SAC [47], FGR [61], 1PRANSAC [47], and PPF [55] are employed; for learning-based methods, DGR [62] and PointDSC [24] are considered. These methods will be compared with SLHG based on different bechmark protocols.

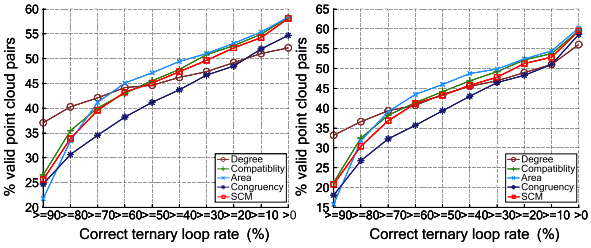
Implementation details. Our method is implemented based on the point cloud library (PCL) using C++ programming language. For U3M and RESSO-LO, we use the Harris3D keypoint (H3D) detector and the signatures of histograms of orientations (SHOT) descriptor for initial correspondence generation as in [22]. For the 3DMatch and 3DLoMatch datasets, we use the fast point feature



(a) Correspondence count

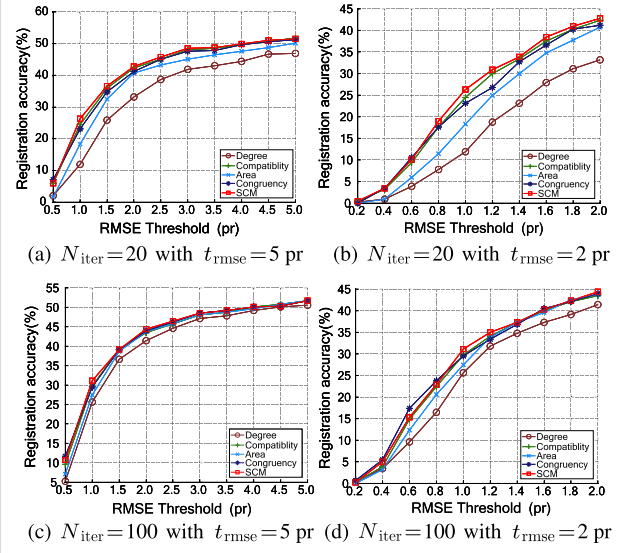
(b) Inlier ratio

Figure 6 Comparison of the initial correspondence set and the max cluster in terms of correspondence count and inlier ratio



(a) $N_{\text{iter}} = 20$ (b) $N_{\text{iter}} = 100$

Figure 7 The percentages of valid point cloud pairs with different correct ternary loop rates resulting from five ternary loop scoring metrics



(a) $N_{\text{iter}} = 20$ with $t_{\text{rmse}} = 5$ pr (b) $N_{\text{iter}} = 20$ with $t_{\text{rmse}} = 2$ pr

(c) $N_{\text{iter}} = 100$ with $t_{\text{rmse}} = 5$ pr (d) $N_{\text{iter}} = 100$ with $t_{\text{rmse}} = 2$ pr

Figure 8 The registration accuracy results achieved by five ternary loop scoring metrics

histograms (FPFH) descriptor to generate the initial correspondence set. All the experiments are conducted with a computer equipped with a 2.3 GHz CPU and 16GB RAM (no parallel computing is involved).

4.2 Method analysis

We conduct a series of experiments to verify the rationality of each stage of SLHG and analyze key parameters. These experiments are deployed on the U3M dataset.

4.2.1 The benefit of cluster construction

To validate the necessity of cluster construction in the first stage of SLHG, we statistically calculate the correspondence counts and inlier ratios for each point cloud pair with and without cluster construction on the U3M dataset in Fig. 6.

As seen from Fig. 6, the size of the correspondence set decreases significantly on each pair of point clouds, while the inlier ratio increases dramatically on most point cloud pairs. Overall, the average number of correspondences decreases from 1877 to 390 and the average inlier ratio rises from 0.07 to 0.22. This indicates that cluster construction can reduce the input size and increase the inlier ratios simultaneously, which is beneficial for efficient and accurate 3D registration.

4.2.2 Ternary loop scoring metric analysis

Guided local sampling is based on the ternary loop scoring metric. We first check the performance of the four metrics introduced in Sect. 3.2, i.e., "compatibility", "degree", "congruence" and "area", and then compare SCM with them to demonstrate its rationality. The performance is checked concerning two terms: 1) the correct rate of top-N_{iter} ternary loops ranked by examined metrics; 2) the registration accuracy with top-N_{iter} ternary loops ranked by examined metrics. Here, a ternary loop is defined as correct if it consists of three correct correspondences.

For correct ternary loop sampling results as shown in Fig. 7, it can be observed that SCM is not the best performer. However, regarding the registration performance as shown in Fig. 8, SCM achieves the best overall performance under different iterations and RMSE thresholds. The results indicate two points. First, correct ternary loops are not guaranteed to achieve correct registrations, as explained previously in Fig. 4. Second, SCM effectively combines constraints from the compatibility space and the spatial space, resulting in reliable guided sampling results.

Table 3 Comparison of the numbers of successfully registered point clouds with and without hypotheses globally on the U3M dataset

	0.2 pr	0.4 pr	0.6 pr	0.8 pr	1.0 pr	1.2 pr	1.4 pr	1.6 pr	1.8 pr	2.0 pr
(a) Hypothesis locally										
SLHG (local)	2	16	36	79	117	145	154	170	182	191
(b) Hypothesis globally										
$T_{\text{thresh}} = 10{,}000$, $V_{\text{thresh}} = 3$	3	18	48	88	130	153	159	174	183	191
$T_{\text{thresh}} = 10{,}000$, $V_{\text{thresh}} = 5$	3	24	52	88	132	155	162	177	186	192
$T_{\text{thresh}} = 10{,}000$, $V_{\text{thresh}} = 10$	3	28	59	92	130	152	162	179	188	194
$T_{\text{thresh}} = 10{,}000$, $V_{\text{thresh}} = 20$	**3**	30	60	99	**132**	157	166	**181**	**189**	**196**
$T_{\text{thresh}} = 1000$, $V_{\text{thresh}} = 3$	3	19	51	92	130	152	162	170	180	190
$T_{\text{thresh}} = 1000$, $V_{\text{thresh}} = 5$	3	24	56	89	129	148	158	172	180	184
$T_{\text{thresh}} = 1000$, $V_{\text{thresh}} = 10$	3	28	62	94	129	150	163	174	183	187
$T_{\text{thresh}} = 1000$, $V_{\text{thresh}} = 20$	3	**31**	**64**	**101**	129	**159**	**168**	177	184	189

Table 4 The numbers of successfully registered point cloud pairs by SLHG with different iterations on the U3M dataset

	0.5 pr	1.0 pr	1.5 pr	2.0 pr	2.5 pr	3.0 pr	3.5 pr	4.0 pr	4.5 pr	5.0 pr
5 iters.	11	59	97	140	164	181	194	200	206	212
10 iters.	27	79	124	160	184	199	211	215	221	225
20 iters.	49	126	160	191	207	215	222	224	229	234
100 iters.	47	128	178	193	210	217	224	231	236	240
200 iters.	51	131	182	199	208	219	228	233	237	241

4.2.3 The rationality of hypothesis globally

To validate the rationality of the hypothesis globally, we conducted a comparative experiment for hypothesis locally and hypothesis globally. Here, hypothesis locally indicates computing the hypothesis with a single ternary loop. The two parameters in the hypothesis globally stage are also discussed. The registration results are presented in Table 3.

The results suggest that hypothesis globally clearly can align more point cloud pairs, especially under strict RMSE thresholds. This demonstrates that hypothesis globally can achieve more accurate registrations. In addition, T_{thresh} = 1000 and V_{thresh} = 20 achieve the best overall performance and we keep this as the default parameter setting.

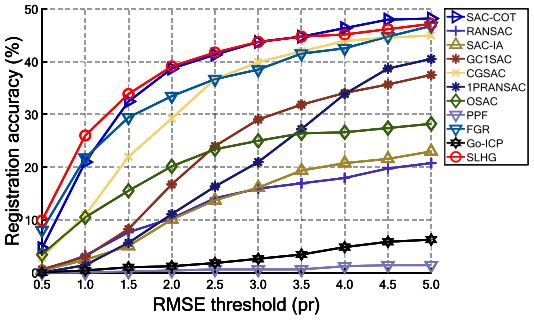


Figure 9 The registration accuracy achieved by eight methods is compared in 20 iterations with a 5 pr RMSE error

4.2.4 Varying iteration count

To demonstrate that SLHG can achieve good registration performance with a few iterations, we vary the number of iterations N_{iter} to check the performance variation. The results are presented in Table 4.

According to the results, it can be found that our method can achieve a remarkable registration accuracy with only 5 iterations. In addition, by observing the registration results under different iterations varying from 20 to 200, it can be seen that our proposed SLHG method can achieve pleasurable performance with 20 iterations and the performance gain becomes marginal as N_{iter} further increases. Thus, we set N_{iter} to 20 for SLHG by default.

4.3 Registration accuracy results

To further verify the effectiveness and accuracy of SLHG, we test it on four experimental datasets addressing different application scenarios. Only the 3DMatch and 3DLoMatch datasets have benchmark records of deep learning methods and we will compare SLHG with learning-based methods on these two datasets.

4.3.1 U3M

The registration accuracy results under different RMSE thresholds on the object-scale U3M dataset are shown in Fig. 9.

It can be seen from Fig. 9 that SLHG and SAC-COT are the two top-ranked methods on this dataset. The two methods outperform most competitors by a clear margin. In particular, when the RMSE threshold is smaller than 2 pr, our SLHG method still outperforms SAC-COT, indicating that SLHG can achieve very accurate registrations.

4.3.2 3DMatch & 3DLoMatch

On the 3DMatch and 3DLoMatch datasets, we use FPFH descriptors to generate initial correspondences with ap-

proximately 1k scale. In some scenarios, outlier rejection is employed before 6-DoF pose estimation to perform registration. As such, we test two cases for SLHG: 1) the initial correspondences are directly fed to SLHG; 2) outlier rejection is performed prior to SLHG (we employ PoinDSC for outlier rejection). In addition, we test methods under two thresholding conditions: 1) loose threshold, i.e., the rotation error is within 15° and translation error is within 30 cm; 2) tight threshold, i.e., the rotation error is within 10° and translation error is within 10 cm. The results of the two datasets are shown in Table 5, Table 6, Table 7 and Table 8.

On the 3DMatch dataset, as can be seen from Table 5 and Table 6, SLHG + PointDSC achieves the best registration performance with both loose-threshold and tight-threshold methods. Impressively, even without the outlier rejection module, SLHG outperforms a state-of-the-art learning method, i.e., PointDSC, under a tight threshold. Moreover, we find that our method with only 20 iterations surpasses RANSAC with 10k iterations by a significant margin.

On the 3DLoMatch dataset, as suggested by Table 7 and Table 8, the performance of all methods drops significantly. However, PointDSC + SLHG still achieves the best performance. It also demonstrates that our method can effectively boost the performance of existing correspondence-based methods.

To demonstrate that our method is capable of achieving very accurate results, we present more detailed results in Table 9 and Table 10. The results suggest that our method can achieve smaller RE and TE than the compared methods. This can be explained that hypothesis globally incorporating more stable global information can bring more accurate pose hypotheses.

Table 5 Registration accuracy performance (%) with loose thresholds on the 3DMatch dataset (the best and the second best results are shown in bold and underlined, respectively)

	Kitchen	Home 1	Home 2	Hotel 1	Hotel 2	Hotel 3	Study room	MIT Lab	Avg.
(i) Traditional									
CG-SAC [21]	48.71	21.29	18.36	17.33	30.09	28.30	11.68	23.68	27.97
GC1SAC [47]	2.18	2.58	2.42	1.78	0.97	1.89	0.69	0.00	1.73
RANSAC [15]	17.62	34.19	24.64	18.22	20.39	26.42	9.96	21.05	19.29
RANSAC (1k iters.) [15]	60.87	71.15	53.37	67.26	58.65	70.37	44.52	51.95	58.60
RANSAC (5k iters.) [15]	71.49	76.13	58.45	73.33	63.11	75.47	48.45	46.79	64.94
RANSAC (10k iters.) [15]	72.48	76.13	58.93	78.22	65.05	69.81	47.42	57.89	65.80
SAC-COT [22]	73.66	76.13	58.94	80.04	69.90	73.58	54.30	51.32	67.83
FGR [61]	44.07	51.28	37.98	46.90	38.46	48.15	26.71	46.75	41.16
(ii) Deep learned									
DGR [62]	66.80	69.23	54.81	69.47	57.69	70.37	49.32	51.95	61.55
PointDSC [24]	78.06	80.77	**67.79**	**86.28**	**72.12**	83.33	63.36	62.34	74.55
Ours	76.48	78.85	61.06	84.07	68.27	75.93	59.93	**64.94**	71.72
Ours (100 iters.)	77.08	78.50	62.02	84.07	70.19	77.78	59.93	**64.94**	72.27
PointDSC + Ours	**78.46**	**81.41**	**67.79**	85.40	71.15	**85.19**	**64.38**	62.34	**74.74**

Table 6 Registration accuracy performance (%) with tight thresholds on the 3DMatch dataset

	Kitchen	Home 1	Home 2	Hotel 1	Hotel 2	Hotel 3	Study room	MIT Lab	Avg.
(i) Traditional									
CG-SAC [21]	24.70	2.56	5.77	2.65	10.58	11.11	4.11	5.19	11.09
GC1SAC [47]	0.00	0.64	0.48	0.00	0.00	0.00	0.00	0.00	0.12
RANSAC [15]	6.91	9.62	11.54	6.64	5.77	1.85	2.74	6.49	6.72
RANSAC (1k iters.) [15]	36.56	44.23	30.29	33.63	26.92	38.89	16.44	24.68	31.36
RANSAC (5k iters.) [15]	43.68	50.00	35.09	42.48	38.46	50.00	18.83	35.06	38.01
RANSAC (10k iters.) [15]	49.01	49.36	36.54	46.46	39.42	51.85	20.55	32.47	40.66
SAC-COT [22]	51.58	50.00	32.21	45.58	35.58	48.15	25.68	31.17	41.34
FGR [61]	29.64	33.97	24.52	27.43	17.31	22.22	10.27	28.57	24.52
(ii) Deep learned									
DGR [62]	41.30	40.38	26.92	39.38	38.46	50.00	22.60	29.81	35.30
PointDSC [24]	59.88	50.00	37.98	48.23	40.38	**62.96**	30.82	38.96	47.13
Ours	58.89	52.56	38.94	53.10	**48.08**	53.70	**32.88**	36.36	48.31
Ours (100 iters.)	59.49	55.13	39.90	53.54	**48.08**	57.41	**32.88**	36.36	49.04
PointDSC + Ours	**60.28**	**57.05**	**40.87**	**54.87**	42.31	**62.96**	32.53	**42.86**	**49.85**

Table 7 Registration accuracy performance (%) with loose thresholds on the 3DLoMatch dataset

	Kitchen	Home 1	Home 2	Hotel 1	Hotel 2	Hotel 3	Study room	MIT Lab	Avg.
(i) Traditional									
CG-SAC [21]	12.15	3.65	2.99	4.81	6.90	15.56	0.92	3.03	6.48
GC1SAC [47]	1.62	0.00	0.00	0.48	0.69	0.00	0.00	0.00	0.61
RANSAC [15]	1.33	1.38	0.43	0.46	1.90	6.12	0.42	0.00	1.12
RANSAC (1k iters.) [15]	9.90	8.99	16.52	14.68	12.02	22.45	1.67	2.78	10.33
RANSAC (5k iters.) [15]	15.24	13.49	23.04	21.56	17.72	32.65	3.75	4.16	15.44
RANSAC (10k iters.) [15]	16.95	14.53	22.17	25.23	17.72	28.57	4.17	6.94	16.51
SAC-COT [22]	21.52	13.39	27.15	33.92	20.11	27.17	5.61	9.98	19.59
FGR [61]	0.95	3.11	3.91	1.38	1.90	6.12	0.42	2.78	1.97
(ii) Deep learned									
DGR [62]	20.57	13.84	25.22	29.82	17.71	34.69	5.42	9.72	18.87
PointDSC [24]	26.10	21.11	30.43	41.74	23.43	38.78	9.17	**18.06**	25.27
Ours	22.86	17.30	27.83	36.70	19.62	**40.82**	8.33	9.72	22.01
Ours (100 iters.)	24.57	19.03	28.26	38.53	20.25	**40.82**	8.33	12.5	23.25
PointDSC + Ours	**26.29**	**21.80**	**31.30**	**42.20**	**24.68**	38.78	**12.92**	**18.06**	**26.22**

Table 8 Registration accuracy performance (%) with tight thresholds on the 3DLoMatch dataset

	Kitchen	Home 1	Home 2	Hotel 1	Hotel 2	Hotel 3	Study room	MIT Lab	Avg.
(i) Traditional									
CG-SAC [21]	2.22	0.73	0.99	0	1.37	6.67	0.46	0.00	1.27
GC1SAC [47]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RANSAC [15]	0.40	0.37	0.50	0	0.69	2.22	0.00	0.00	0.36
RANSAC (1k iters.) [15]	1.90	3.81	6.52	2.75	3.16	14.28	0.83	0.00	3.14
RANSAC (5k iters.) [15]	4.00	5.54	9.56	5.04	5.79	20.40	0.00	0.00	4.99
RANSAC (10k iters.) [15]	4.95	6.23	9.13	7.34	5.06	18.36	0.83	2.78	5.72
SACCOT [22]	6.28	4.81	8.89	11.76	7.14	16.72	2.15	2.85	6.62
FGR [61]	0.00	0.35	1.74	0.92	1.27	0.00	0.00	0.00	0.51
(ii) Deep learned									
DGR [62]	3.62	5.19	5.22	11.47	5.06	20.41	0.83	2.78	5.22
PointDSC [24]	10.10	9.00	14.35	15.14	13.92	18.37	1.25	2.78	10.16
Ours	8.76	8.00	13.04	13.30	13.29	18.37	1.67	4.17	9.15
Ours (100 iters.)	9.52	9.00	13.91	14.22	13.29	18.37	1.67	4.17	9.88
PointDSC + Ours	**12.38**	**9.34**	**14.78**	**15.60**	**15.19**	**22.45**	**2.50**	**5.56**	**11.51**

Table 9 Registration performance with loose thresholds on the 3DMatch dataset

	RA	RE (°)	TE (cm)
(i) Traditional			
CG-SAC	27.97	5.22	12.11
GC1SAC	1.73	7.31	13.73
RANSAC	19.29	6.17	14.84
RANSAC (1k iters.)	58.60	4.14	11.24
RANSAC (5k iters.)	64.94	3.67	10.23
RANSAC (10k iters.)	65.80	3.46	9.51
SACCOT	67.78	3.55	9.97
FGR	41.16	4.07	10.22
(ii) Deep learned			
DGR	61.55	3.38	9.91
PointDSC	74.55	3.07	9.43
Ours	71.72	3.22	9.61
Ours (100 iters)	72.27	3.14	9.31
PointDSC + Ours	**74.74**	**3.04**	**9.18**

Table 10 Registration performance with loose thresholds on the 3DLoMatch dataset

	RA	RE (°)	TE (cm)
(i) Traditional			
CG-SAC	6.48	7.63	17.69
GC1SAC	0.61	8.68	19.11
RANSAC	1.21	8.14	17.21
RANSAC (1k iters.)	10.33	6.67	15.10
RANSAC (5k iters.)	15.44	5.65	13.12
RANSAC (10k iters.)	16.51	5.8	13.72
SACCOT	19.59	5.56	13.88
FGR	1.97	7.34	14.73
(ii) Deep learned			
DGR	18.87	5.62	14.13
PointDSC	25.27	5.11	12.37
Ours	22.01	5.43	13.07
Ours (100 iters)	23.25	5.27	12.41
PointDSC + Ours	**26.22**	**5.03**	**12.21**

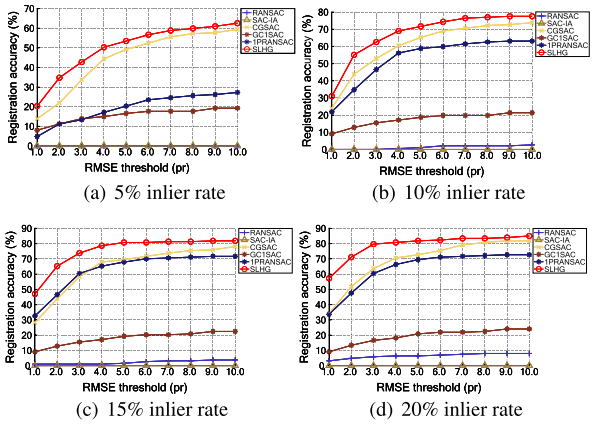


Figure 10 Results on the RESSO-LO dataset with different inlier rates

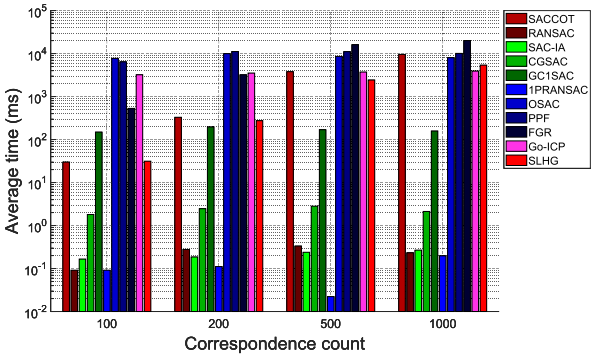


Figure 11 Visual registration results. From left to right: the source point cloud, the target point cloud, the registration results of three tested methods, and the point-wise errors visualized in pseudo-color results from the three tested methods

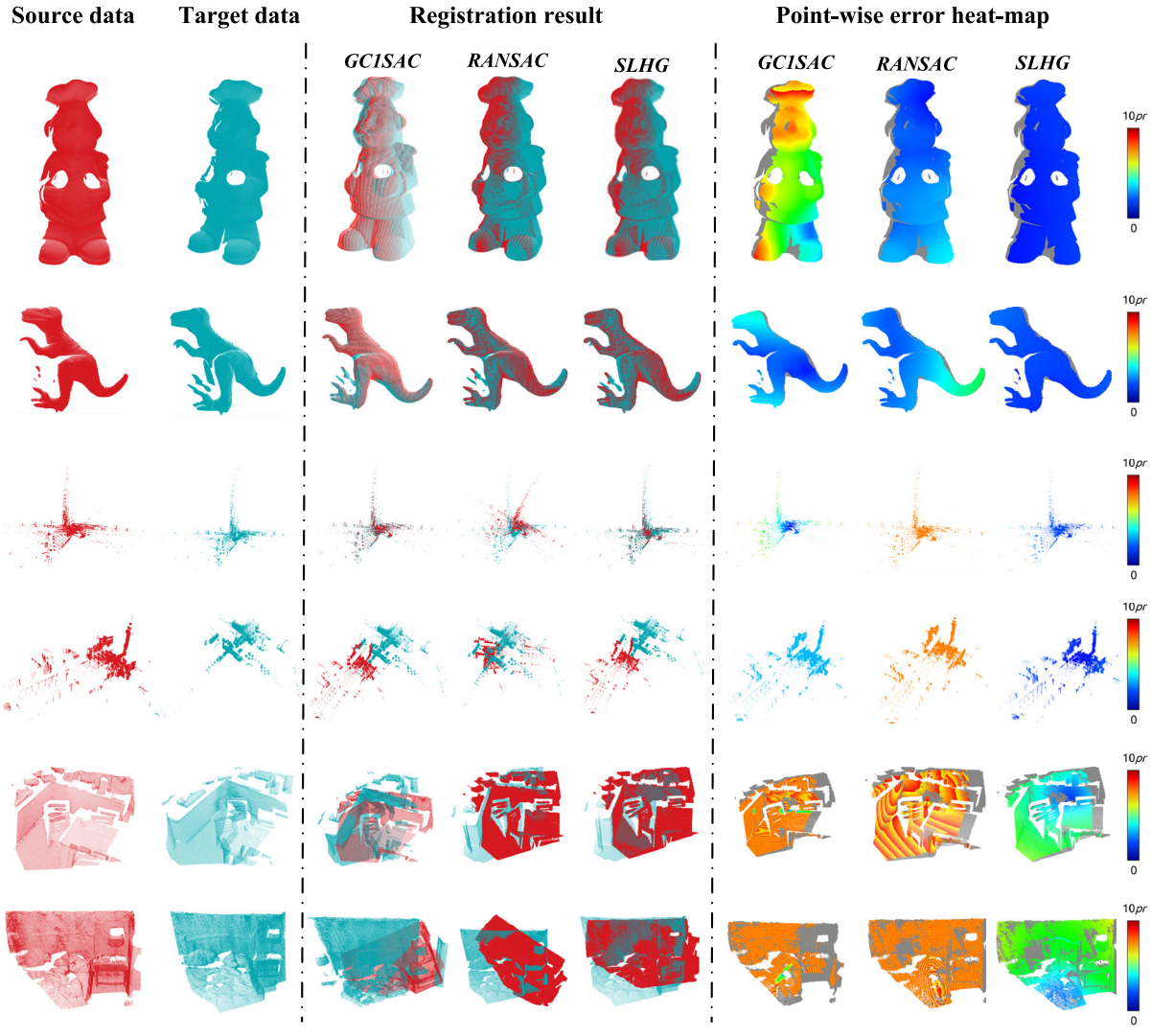


Figure 12 Average time costs of each estimator for registering one point cloud pair with different scales of inputs

4.3.3 RESSO-LO

Finally, on the large-scale outdoor RESSO-LO dataset, we follow [22] and generate correspondences with different inlier rates as inputs for the tested methods. Note that some methods are ultra-consuming on this dataset, such as OSAC, Go-ICP and PPF, and they are not included for comparison here. The results are shown in Fig. 10.

It can be seen from Fig. 10 that SLHG is the best performer on all tested conditions. The superiority of SLHG still remains clear even with only 5% inlier rate.

4.3.4 Visualization

In Fig. 11, we visualize the registration results of the proposed SLHG method and two competitors.

Fig. 11 suggests that the registration results achieved by SLHG are clearly more accurate than those achieved by GC1SAC and RANSAC. The results are consistent across different datasets, indicating the good generalization ability of SLHG.

4.4 Efficiency

In Fig. 12, we report the efficiency results by different methods with different scales of input correspondences. Learning-based methods are not compared here because our method as well as compared methods here are implemented with CPU only without parallel computing.

First, one can see that SLHG is less efficient than RANSAC, SAC-IA, CG-SAC and 1PRANSAC. This is because SLHG additionally performs cluster construction and voting during the hypothesis generation stage. However, SLHG achieves significantly better registration results. Second, SL-HG still can finish 6-DoF pose estimation with around 10 ms given 100 correspondences, which can already satisfy the requirement of real-time applications. Overall, SLHG strikes a good balance in terms of efficiency, accuracy, and robustness.

5 Conclusions

In this paper, we present a RANSAC variant called SLHG for 6-DoF pose estimation and point cloud registration. The key idea is performing guided local sampling to reduce iterations while hypothesizing globally to increase registration accuracy. Experiments on the object, indoor, and outdoor datasets as well as comparison with both geometry-based and learning-based methods verify the overall superiority of our method.

In the future, we plan to explore a more stable voting scheme to further reduce the impacts of heavy outliers so as to improve the performance in cases with extreme low inlier ratios.

Acknowledgements
The authors would like to acknowledge the publishers of the employed datasets for making them publicly available.

Funding
This work was supported by the National Natural Science Foundation of China (Nos. 62002295, 62006025 and U19B2037), Natural Science Basic Research Program of Shaanxi (No. 2021JCW-03), Postdoctoral Research Foundation of China (No. 2020M673319), and Shaanxi Provincial Key R&D Program (No. 2021KWZ-03).

Abbreviations
Avg., Average; comp, Compatibility; cong, Congruence; est, Estimation; gt, Ground truth; iter/iters, Iteration/iterations; mae, Mean absolute error; pr, Point cloud resolution; RA, Registration accuracy; RE, Rotation error; rmse, Root mean square error; TE, Translation error; thresh, Threshold.

Availability of data and materials
The datasets used in this study were obtained from multiple sources, including UWA 3D models, 3DMatch, 3DLoMatch, and the Real-world Scans with Small Overla. The details of the datasets used in this study are as follows: UWA 3D models: the dataset used in this study was obtained from [https://ieee-dataport.org/documents/three-dimensional-object-recognition-and-segmentation-cluttered-scenes]. 3DMatch dataset: the dataset used in this study was obtained from [https://3dmatch.cs.princeton.edu/]. 3DLoMatch dataset: the dataset used in this study was obtained from [https://overlappredator.github.io/]. The Real-world Scans with Small Overlap dataset: the dataset used in this study was obtained from [https://3d.bk.tudelft.nl/liangliang/publications/2019/plade/resso.html].

Declarations

Competing interests
The authors declare no competing interests.

Author contributions
This study was conceptualized and designed by YC, ZH and JY. Material preparation, data collection and analysis were performed by YC, ZH and SQ. Experiments of evaluation methods were performed by YC, XC and SZ. The first draft of the manuscript was written by YC and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Author details
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, 710129, China. [2]Da-Jiang Innovations, ShenZhen, 511464, China. [3]School of Electronic and Control Engineering, Chang'An University, Xi'an, 710064, China.

Received: 9 April 2023 Revised: 1 August 2023 Accepted: 4 August 2023
Published online: 02 September 2023

References

1. Lv, C., Lin, W., & Zhao, B. (2022). Voxel structure-based mesh reconstruction from a 3D point cloud. *IEEE Transactions on Multimedia*, *24*, 1815–1829.
2. Arvanitis, G., Zacharaki, E. I., Váša, L., & Moustakas, K. (2022). Broad-to-narrow registration and identification of 3D objects in partially scanned and cluttered point clouds. *IEEE Transactions on Multimedia*, *24*, 2230–2245.
3. Lalos, A. S., Nikolas, I., Vlachos, E., & Moustakas, K. (2017). Compressed sensing for efficient encoding of dense 3D meshes using model-based Bayesian learning. *IEEE Transactions on Multimedia*, *19*(1), 41–53.
4. Wang, Z., Zhang, L., Shen, Y., & Zhou, Y. (2022). D-LIOM: tightly-coupled direct liDAR-inertial odometry and mapping. *IEEE Transactions on Multimedia*. Advance online publication. https://doi.org/10.1109/TMM.2022.3168423.
5. Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256.
6. Chen, Y., & Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing*, *10*(3), 145–155.
7. Censi, A. (2008). An ICP variant using a point-to-line metric. In *2008 IEEE international conference on robotics and automation* (pp. 19–25). Los Alamitos: IEEE.
8. Segal, A., Haehnel, D., & Thrun, S. (2009). Generalized-ICP. In *Proceedings of robotics: science and systems* (pp. 1–8). Cambridge: MIT Press.

9. Blanco, J.-L., González-Jiménez, J., & Fernández-Madrigal, J.-A. (2012). An alternative to the Mahalanobis distance for determining optimal correspondences in data association. *IEEE Transactions on Robotics*, *28*(4), 980–986.
10. Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. In *The 3rd international conference on 3-D digital imaging and modeling* (pp. 145–152). Los Alamitos: IEEE.
11. Guo, Y., Sohel, F., Bennamoun, M., Wan, J., & Lu, M. (2014). An accurate and robust range image registration algorithm for 3D object modeling. *IEEE Transactions on Multimedia*, *16*(5), 1377–1390.
12. Olsson, C., Kahl, F., & Oskarsson, M. (2008). Branch-and-bound methods for Euclidean registration problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(5), 783–794.
13. Yang, J., Li, H., Campbell, D., & Jia, Y. (2015). Go-ICP: a globally optimal solution to 3D ICP point-set registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(11), 2241–2254.
14. Magnusson, M. (2009). *The three-dimensional normal-distributions transform: an efficient representation for registration, surface analysis, and loop detection*. Ph.D. dissertation, Örebro Universitet.
15. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.
16. Li, J., Hu, Q., & Ai, M. (2021). Point cloud registration based on one-point RANSAC and scale-annealing biweight estimation. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(11), 9716–9729.
17. Raguram, R., Chum, O., Pollefeys, M., Matas, J., & Frahm, J.-M. (2012). USAC: a universal framework for random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 2022–2038.
18. Barath, D., Noskova, J., Ivashechkin, M., & Matas, J. (2020). Magsac++, a fast, reliable and accurate robust estimator. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1304–1312). Los Alamitos: IEEE.
19. Brachmann, E., Krull, A., Nowozin, S., Shotton, J., Michel, F., Gumhold, S., & Rother, C. (2017). DSAC-differentiable RANSAC for camera localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6684–6692). Los Alamitos: IEEE.
20. Brachmann, E., & Rother, C. (2019). Neural-guided RANSAC: learning where to sample model hypotheses. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4322–4331). Los Alamitos: IEEE.
21. Quan, S., & Yang, J. (2020). Compatibility-guided sampling consensus for 3-D point cloud registration. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(10), 7380–7392.
22. Yang, J., Huang, Z., Quan, S., Qi, Z., & Zhang, Y. (2021). SAC-COT: sample consensus by sampling compatibility triangles in graphs for 3-D point cloud registration. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–15.
23. Jiang, H., Shen, Y., Xie, J., Li, J., Qian, J., & Yang, J. (2021). Sampling network guided cross-entropy method for unsupervised point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6128–6137). Los Alamitos: IEEE.
24. Bai, X., Luo, Z., Zhou, L., Chen, H., Li, L., Hu, Z., et al. (2021). Pointdsc: robust point cloud registration using deep spatial consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 15859–15869). Los Alamitos: IEEE.
25. Pankaj, D. S., & Nidamanuri, R. R. (2016). A robust estimation technique for 3D point cloud registration. *Image Analysis & Stereology*, *35*(1), 15–28.
26. Aoki, Y., Goforth, H., Srivatsan, R. A., & Lucey, S. (2019). Pointnetlk: robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7163–7172). Los Alamitos: IEEE.
27. Mian, A. S., Bennamoun, M., & Owens, R. A. (2006). A novel representation and feature matching algorithm for automatic pairwise registration of range images. *International Journal of Computer Vision*, *66*(1), 19–40.
28. Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., & Funkhouser, T. (2017). 3DMatch: learning local geometric descriptors from RGB-D reconstructions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1802–1811). Los Alamitos: IEEE.
29. Chen, S., Nan, L., Xia, R., Zhao, J., & Wonka, P. (2019). Plade: a plane-based descriptor for point cloud registration with small overlap. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(4), 2530–2540.
30. Masuda, T., Sakaue, K., & Yokoya, N. (1996). Registration and integration of multiple range images for 3-D model construction. In *The 13th international conference on pattern recognition* (pp. 879–883). Los Alamitos: IEEE.
31. Weik, S. (1997). Registration of 3-D partial surface models using luminance and depth information. In *Proceedings of the IEEE international conference on recent advances in 3-D digital imaging and modeling* (pp. 93–100). Los Alamitos: IEEE.
32. Turk, G., & Levoy, M. (1994). Zippered polygon meshes from range images. In *Proceedings of the 21st annual conference on computer graphics and interactive techniques* (pp. 311–318). New York: ACM.
33. Godin, G., Rioux, M., & Baribeau, R. (1994). Three-dimensional registration using range and intensity information. In *Proceedings of Videometrics III* (Vol. 2350, pp. 279–290). Bellingham: SPIE.
34. Land, A. H., & Doig, A. G. (2010). An automatic method for solving discrete programming problems. In M. Jünger, T. M. Liebling, D. Naddef, et al. (Eds.), *50 years of integer programming 1958–2008* (pp. 105–132). Berlin: Springer.
35. Liu, Y., Wang, C., Song, Z., & Wang, M. (2018). Efficient global point cloud registration by matching rotation invariant features through translation search. In *Proceedings of the 15th European conference on computer vision* (pp. 448–463). Berlin: Springer.
36. Brown, M., Windridge, D., & Guillemaut, J.-Y. (2019). A family of globally optimal branch-and-bound algorithms for 2D–3D correspondence-free registration. *Pattern Recognition*, *93*, 36–54.
37. Magnusson, M., Lilienthal, A., & Duckett, T. (2007). Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics*, *24*(10), 803–827.
38. Biber, P., & Straßer, W. (2003). The normal distributions transform: a new approach to laser scan matching. In *Proceedings of the IEEE international conference on intelligent robots and systems* (Vol. 3, pp. 2743–2748). Los Alamitos: IEEE.
39. Stoyanov, T., Magnusson, M., Andreasson, H., & Lilienthal, A. J. (2012). Fast and accurate scan registration through minimization of the distance between compact 3D NDT representations. *The International Journal of Robotics Research*, *31*(12), 1377–1393.
40. Stoyanov, T., Magnusson, M., & Lilienthal, A. J. (2012). Point set registration through minimization of the L2 distance between 3D-NDT models. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 5196–5201). Los Alamitos: IEEE.
41. Shi, X., Peng, J., Li, J., Yan, P., & Gong, H. (2019). The iterative closest point registration algorithm based on the normal distribution transformation. *Procedia Computer Science*, *147*, 181–190.
42. Rusu, R. B., Blodow, N., & Beetz, M. (2009). Fast point feature histograms (FPFH) for 3D registration. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 3212–3217). Los Alamitos: IEEE.
43. Yang, J., Cao, Z., & Zhang, Q. (2016). A fast and robust local descriptor for 3D point cloud registration. *Information Sciences*, *346*, 163–179.
44. Chen, Z., Sun, K., Yang, F., & Tao, W. (2022). SC2-PCR: a second order spatial compatibility for efficient and robust point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13221–13231). Los Alamitos: IEEE.
45. Yang, J., Zhang, X., Fan, S., Ren, C., & Zhang, Y. (2023). Mutual voting for ranking 3D correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Advance online publication. https://doi.org/10.1109/TPAMI.2023.3268297.
46. Yang, J., Zhang, Q., & Cao, Z. (2017). Multi-attribute statistics histograms for accurate and robust pairwise registration of range images. *Neurocomputing*, *251*, 54–67.
47. Quan, S., Ma, J., Hu, F., Fang, B., & Ma, T. (2018). Local voxelized structure for 3D binary feature representation and robust registration of point clouds from low-cost sensors. *Information Sciences*, *444*, 153–171.
48. Yang, J., Huang, Z., Quan, S., Cao, Z., & Zhang, Y. (2022). RANSACs for 3D rigid registration: a comparative evaluation. *IEEE/CAA Journal of Automatica Sinica*, *9*(10), 1861–1878.
49. Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2019). Pointnet: deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7163–7172). Los Alamitos: IEEE.
50. Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., & Xu, K. (2022). *Geometric transformer for fast and robust point cloud registration*. arXiv preprint. arXiv:2202.06688.
51. Pais, G. D., Ramalingam, S., Govindu, V. M., Nascimento, J. C., Chellappa, R., & Miraldo, P. (2020). 3DRegNet: a deep neural network for 3D point registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7193–7203). Los Alamitos: IEEE.

52. Jiang, H., Dang, Z., Wei, Z., Xie, J., Yang, J., & Salzmann, M. (2023). Robust outlier rejection for 3D registration with variational Bayes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1148–1157). Los Alamitos: IEEE.
53. Lombardi, M., Riccardi, A., Savardi, M., & Signoroni, A. (2019). 3DReg-i-Net: improving deep learning based 3D registration for a robust real-time alignment of small-scale scans. In M. Agus, M. Corsini, & R. Pintus (Eds.), *Italian Chapter conference 2019-smart tools and apps in computer graphics* (pp. 51–60). Eindhoven: Eurographics Association.
54. Aiger, D., Mitra, N. J., & Cohen-Or, D. (2008). 4-points congruent sets for robust pairwise surface registration. *ACM Transactions on Graphics*, *27*(3), 1–10.
55. Drost, B., Ulrich, M., Navab, N., & Ilic, S. (2010). Model globally, match locally: efficient and robust 3D object recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 998–1005). Los Alamitos: IEEE.
56. Guo, Y., Sohel, F., Bennamoun, M., Lu, M., & Wan, J. (2013). Rotational projection statistics for 3D local surface description and object recognition. *International Journal of Computer Vision*, *105*(1), 63–86.
57. Buch, A. G., Kiforenko, L., & Kraft, D. (2017). Rotational subgroup voting and pose clustering for robust 3D object recognition. In *Proceedings of the IEEE international conference on computer vision* (pp. 4137–4145). Los Alamitos: IEEE.
58. Tombari, F., & Di Stefano, L. (2010). Object recognition in 3D scenes with occlusions and clutter by hough voting. In *Proceedings of the fourth Pacific-Rim symposium on image and video technology* (pp. 349–355). Los Alamitos: IEEE.
59. Yang, J., Huang, Z., Quan, S., Zhang, Q., Zhang, Y., & Cao, Z. (2020). *On efficient and robust metrics for ransac hypotheses and 3D rigid registration*. arXiv preprint. arXiv:2011.04862.
60. Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., & Schindler, K. (2021). Predator: registration of 3D point clouds with low overlap. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4267–4276). Los Alamitos: IEEE.
61. Zhou, Q.-Y., Park, J., & Koltun, V. (2016). Fast global registration. In B. Leibe, J. Matas, N. Sebe, et al. (Eds.), *Proceedings of the 14th European conference on computer vision* (pp. 766–782). Berlin: Springer.
62. Choy, C., Dong, W., & Koltun, V. (2020). Deep global registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2514–2523). Los Alamitos: IEEE.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com