SC^2-PCR++: Rethinking the Generation and Selection for Efficient and Robust Point Cloud Registration

Zhi Chen, Kun Sun, *Member, IEEE*, Fan Yang, Lin Guo, and Wenbing Tao, *Member, IEEE*

***Abstract*—Outlier removal is a critical part of feature-based point cloud registration. In this article, we revisit the model generation and selection of the classic RANSAC approach for fast and robust point cloud registration. For the model generation, we propose a second-order spatial compatibility (SC^2) measure to compute the similarity between correspondences. It takes into account global compatibility instead of local consistency, allowing for more distinctive clustering between inliers and outliers at an early stage. The proposed measure promises to find a certain number of outlier-free consensus sets using fewer samplings, making the model generation more efficient. For the model selection, we propose a new Feature and Spatial consistency constrained Truncated Chamfer Distance (FS-TCD) metric for evaluating the generated models. It considers the alignment quality, the feature matching properness, and the spatial consistency constraint simultaneously, enabling the correct model to be selected even when the inlier rate of the putative correspondence set is extremely low. Extensive experiments are carried out to investigate the performance of our method. In addition, we also experimentally prove that the proposed SC^2 measure and the FS-TCD metric are general and can be easily plugged into deep learning based frameworks.**

***Index Terms*—Point cloud registration, second-order spatial compatibility, constrained truncated chamfer distance, rigid transformation estimation.**

I. Introduction

THE alignment of two 3D scans of the same scene, known as Point Cloud Registration (PCR), plays an important role in areas such as Simultaneous Localization and Mapping (SLAM) [1], [2], [3], [4], augmented reality [5], [6] and robotics applications [7]. A canonical solution first establishes feature correspondences and then estimates the 3D rotation and translation that achieve optimal alignment of the common parts. However, due to challenges such as partial overlap or feature ambiguity, model estimation is prone to outliers in the correspondences, leading to inaccurate or wrong alignment.

RANdom SAmple Consensus (RANSAC) [8] pioneers the *generation-and-selection* strategy for model estimation. It generates a lot of hypothetical models through random sampling and selects the best hypothesis with the maximum consensus as the final result. The goal of the sampling process is to obtain an outlier-free set, so as to estimate a robust transformation while excluding the impact of outliers. However, it needs massive samplings or sometimes there is no guarantee of an accurate solution due to the low inlier rate. Spatial Compatibility (SC) [9], [10], [11], [12] is a widely used similarity measure for boosting the robustness and efficiency of the rigid transformation estimation. It assumes that two correspondences will have a higher score if the difference of spatial distance between them, e.g., $|\, d_{12} - d'_{12} \,|$ or $|\, d_{16} - d'_{16} \,|$ in Fig. 1(a), is small. Thus, sampling from compatible correspondences increases the probability of getting inliers. However, such kind of first-order metric still suffers from outliers due to locality and ambiguity. A toy example is shown in Fig. 1. There are five inliers {c1, c2, c3, c4, c5} and two outliers {c6, c7} in Fig. 1(a). As we can see from the yellow cells in Fig 1(b), c6 and c7 are outliers but they show high compatibility scores with some inliers by chance. As a result, the outliers would be inevitably involved in the model estimation process, leading to performance deterioration.

In this article, we propose a new global measure of the similarity between two correspondences. Specifically, we first binarize the spatial compatibility matrix into the hard form, as shown in Fig. 1(c). Then, for two compatible correspondences, we compute the number of correspondences that are simultaneously compatible with both of them as the new similarity between them. The globally common compatibility is set to 0 for any two incompatible correspondences. Therefore, the similarity between two inliers is at least the number of inliers excluding themselves from all the correspondences. However, the outliers do not have such good properties. To be specific, in Fig. 1(d), the similarities within the inliers {c1, c2, c3, c4, c5} are no less than 3, while the similarities related to the outliers {c6, c7} are no more than 1. Therefore, the global compatibility matrix in Fig. 1(d) can better distinguish inliers from outliers. Since the new measure can be expressed as the matrix product of the traditional first-order metric (See (8)), we name it the second-order spatial compatibility (SC^2) measure.

Manuscript received 28 October 2022; revised 19 April 2023; accepted 26 April 2023. Date of publication 3 May 2023; date of current version 5 September 2023. This work was supported by the National Natural Science Foundation of China under Grant 62176096. Recommended for acceptance by E. Kalogerakis. *(Corresponding author: Wenbing Tao.)*

Zhi Chen, Fan Yang, Lin Guo, and Wenbing Tao are with the National Key Laboratory of Science and Technology on Multi-spectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China (e-mail: z_chen@hust.edu.cn; fanyang@hust.edu.cn; linguo@hust.edu.cn; wenbingtao@hust.edu.cn).

Kun Sun is with the Hubei Key Laboratory of Intelligent Geo-Information Processing, School of Computer Science, China University of Geosciences, Wuhan, Hubei 430079, China (e-mail: kunsun@cug.edu.cn).

The code will be available at https://github.com/ZhiChen902/SC2-PCR-plusplus.

Digital Object Identifier 10.1109/TPAMI.2023.3272557

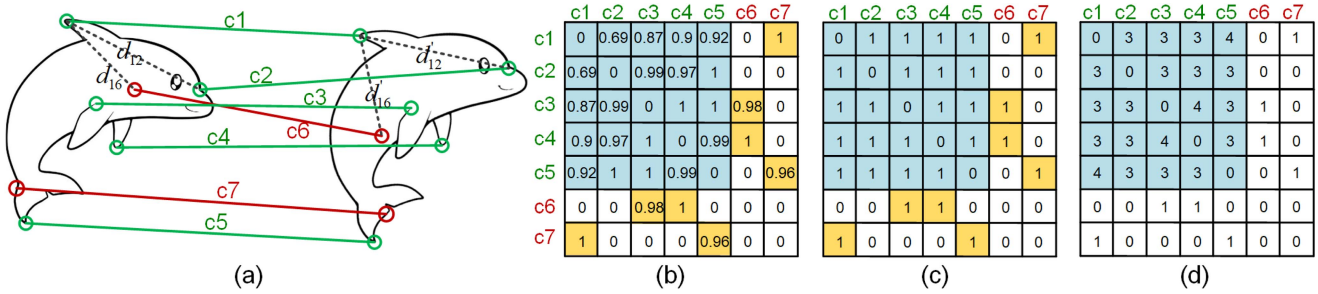


Fig. 1. **(a):** A toy example in which red and green line segments represent outliers and inliers, respectively. **(b):** The first order compatibility matrix of (a). As highlighted in yellow, the outliers have very high compatibility scores with some inliers. **(c):** A binarized compatibility matrix of (b) after thresholding. **(d):** The proposed second-order compatibility matrix of (a). By contrast, the values in the rows and columns of the outliers are small.

The proposed second-order spatial compatibility measure SC^2 has several advantages. 1) The inliers are much easier distinguished from the outliers. Suppose we have m inliers in n correspondences. The scores between any two inliers would be no less than m-2. However, it is difficult for an outlier to be simultaneously compatible with multiple correspondences and the score for it will be much smaller. 2) Based on the proposed SC^2 matrix, for each row vector corresponding to an inlier, we can easily find an outlier-free set by selecting the top k correspondences with the highest scores. In this way, the m valid samplings can be obtained by traversing all the n rows of the SC^2 matrix. Therefore, we can ensure m reliable model estimations by only n samplings, which makes the model estimation more efficient and robust. 3) We theoretically prove that the SC^2 matrix significantly reduces the probability of wrong sampling from a probabilistic view. We define an ambiguity event, in which the score between two inliers is smaller than that between an inlier and an outlier. By computing the probability distributions of this event for both the first-order metric and our second-order metric, the SC^2 matrix is much more robust to obtain reliable sampling (see Fig. 4).

Besides efficient sampling for model generation, another important issue is selecting the best model from the estimated hypotheses. RANSAC adopts the inlier count [8] as the hypothesis evaluation metric. It counts the number of correspondences whose alignment error is less than a pre-defined threshold after aligning by the estimated hypothesis and selects the model with the highest number as the final result. However, due to the limitation of feature descriptors or challenge brought by low overlap, sometimes the number of inliers in the putative correspondences is extremely small. In this case, even if an outlier-free set is sampled and a correct rigid transformation is estimated, the related inlier count value may be still small. This brings great challenges to selecting the correct hypothesis as the final result. To address this issue, we expect to introduce another widely used metric in point cloud processing, i.e., Chamfer distance (CD) [13], [14], [15], [16], to model selection. It calculates the distribution similarity of the two aligned point clouds. Taking the global alignment information into account, it gets rid of the dependence on the inlier rate of the putative correspondences. However, Chamfer distance totally ignores the appearance information, making it susceptible to noises and unstable in low-overlapped scenes. Meanwhile, computing Chamfer distance needs to search for the nearest neighbor in the global scope. Since it could generate lots of hypotheses during the model selection process, directly using it as the selection metric is time-consuming. Based on the above observations, we propose a new Feature and Spatial consistency constrained Truncated Chamfer Distance (FS-TCD) metric for hypothesis selection. Instead of searching for the nearest neighbor in the global set, it utilizes the feature information to narrow the search space, making it more efficient. Meanwhile, due to introducing the constraint of feature information, the anti-noise ability of our method is greatly enhanced. In order to eliminate the influence of noises, we further adopt the spatial consistency constraint [9], [10] into the FS-TCD metric. In Section V-B, we also show that FS-TCD is a more generalized inlier count, which greatly reduces the dependence on the inlier rate of the putative correspondences.

Based on the proposed SC^2 measure and the FS-TCD metric, we design an efficient and robust point cloud registration method, named SC^2-PCR++. Following [9], [17], [18], it first selects several seeds that are likely to be inliers. Then a two-stage sampling is carried out to construct a consensus set for each seed. Afterward, the weighted SVD is used to estimate a tentative model for each consensus set. Finally, the FS-TCD metric is adopted to select the best model as the final result. In a nutshell, this paper distinguishes itself from existing methods in the following aspects.

- For model generation, a second-order spatial compatibility (SC^2) measure is proposed. We prove that SC^2 significantly reduces the probability of an outlier being involved in the consensus set. Since the proposed method encodes richer information beyond the first-order metric, it enhances the robustness against outliers.
- For model selection, a new Feature and Spatial consistency constrained Truncated Chamfer Distance (FS-TCD) metric is proposed. It integrates the feature similarity, spatial consistency, and alignment quality, so that it can still find the best-estimated model even if the inlier ratio is extremely low.
- Compared with state-of-the-art learning methods such as [9], [11], [19], [20], our method is a light weighted solution that does not need training. It shows no bias

across different datasets and generalizes well on various scenarios, which is elaborated in the experiments.

- The proposed method is general. Although we implement it in a handcrafted fashion, it could be easily plugged into other deep learning frameworks such as PointDSC [9]. We show in the experiments that PointDSC produces better results when combined with the proposed SC^2 measure and FS-TCD metric.

This paper is an extension of our previous work SC^2-PCR (CVPR 2022) [21]. We have made several additions: 1) More exhaustive analyses and discussions about SC^2 measure are added. We present the detailed derivation of the ambiguity probability of SC^2 measure in Section IV-B, and add the experiments of parameters for SC^2 measure based sampling. 2) A new FS-TCD metric is employed to select the best hypothesis over the models produced by the SC^2 measure. Based on the FS-TCD metric, we develop a two-stage model selection strategy to accelerate the selection process. 3) A new pipeline named SC^2-PCR++ is designed. Different from SC^2-PCR, it adopts different matching methods for model generation and selection to overcome the wrong model selection in low-inlier-rate scenes. 4) More experiments are conducted to validate the performance of the proposed method. We use more datasets and competing methods in the experimental section.

II. Related Works

A. 3D Feature Matching

Traditional Feature Matching: An important step of feature-based point cloud registration is to establish correspondences by matching local descriptors. Some methods utilize the histograms of spatial distribution to generate the local descriptors. Spin Image (SI) [22], [23] uses the Principal Component Analysis (PCA) to compress the spin-images so that it is efficient enough for recognition from large model libraries. 3D Shape Context (3DSC) [24] directly extends the 2D shape context to three dimensions and shows that it is more robust to noisy scenes. Unique Shape Context (USC) [25] presents a comprehensive proposal that does not need to compute multiple descriptors for a detected key point. Some other works represent the local descriptor by geometric attribute histogram. Persistent Feature Histograms (PFH) [26] uses a 16D feature for characterizing the local geometry. It improves the robustness to the variations of position, orientation, or sampling density. Fast Point Feature Histogram (FPFH) [27] simplifies the PFH as SPFH by only considering the neighbors of the point, and generates the FPFH utilizing the weighted sum of SPFH. A more comprehensive study about hand-crafted feature matching can be found in [28].

Learning-Based Feature Matching: Recently, deep learning techniques are also introduced to learn 3D local descriptors [29], [30], [31], [32]. The pioneering 3DMatch [29] builds a Siamese Network for extracting local descriptors and provides the 3DMatch benchmark for evaluating the performance of point cloud registration. Following 3DMatch, some methods boost the performance by designing more suitable architectures for extracting local and global information. PPFNet [33] and the unsupervised PPFFoldNet [34] combine the point pair features with the global context generated by the PointNet [35]. 3DSmoothNet [36] and FCGF [30] build fully convolution networks by using voxelized smoothed density value (SDV) representation and Minkowski convolution [37] respectively. Keypoint detection modules [31], [38], [39] are also integrated into the descriptor learning networks. More recently, the successfully applied Transformer [40] architecture is also introduced to 3D feature matching area. Predator [41] designs an overlap-aware module by the self-cross-self attention operations. CoFINet [42] reformulates the Predator into a coarse-to-fine pipeline. Lepard [43] and GeoTransformer [44] add the new position encoding techniques into the attention operations.

Although these methods achieve remarkable performance improvements, they can hardly establish a totally outlier-free correspondence set. They depend on the model fitting method for robust rigid transformation estimation.

B. Model Fitting

Generation and Selection Framework: The Generation and Selection framework is the most significant pipeline for model-fitting, starting with the well-known RANSAC [8]. It can be applied to model estimation tasks in which the model is clearly defined and it is robust to outlier points. In past decades, many of its variants [45], [46], [47], [48] have been proposed. Instead of random sampling, PROSAC [47] computes the ordering by the similarity of local descriptors, and adopts the progressive sampling strategy to accelerate the RANSAC. EVSAC [49] further transforms the descriptor similarity to the confidence value for sampling by means of the extreme value theory. Lo-RANSAC [50] and accelerated FLo-RANSAC [51] perform local optimization to reduce the noises brought by the minimum subset. More recently, Graph-cut RANSAC [52], [53] proposes to use the Graph-cut technique for performing the local optimization step on the so-far-the-best model. Magsac [54] proposes a σ-consensus to build a threshold-free method for RANSAC. A more comprehensive study about RANSAC family can be found in [55].

Learning-Based Model Fitting: Recent works also adopt deep learning techniques, which were studied earlier in the 2D matching area, to model fitting tasks. The 2D correspondence selection network CN-Net [56] and its variants [57], [58], [59], [60], [61], [62], [63] formulate the model fitting as a combination of a correspondence classification module and a model estimation module. Recent attempts [9], [11], [19], [20] also introduce deep learning networks for 3D correspondence pruning. 3DRegNet [19] reformulates the CN-Net [56] into 3D form and designs a regression module to solve rigid transformation. DGR [20] introduces the full convolution to better capture global context for correspondence classification, and uses weight Procrustes for model estimation. PointDSC [9] develops a spatial consistency based non-local module and a Neural Spectral matching to accelerate the model generation and selection. DetarNet [64] presents decoupling solutions for translation and rotation. DHVR [11] exploits the deep Hough voting to identify the consensus from the Hough space, so as to predict the final transformation. COTReg [65] presents a coupled optimal transport based correspondence prediction module and integrates it into the network.

Spatial Compatibility: Spatial compatibility (SC), which is defined by the length consistency of inliers, is widely applied in point cloud registration. Spectral matching (SM) [66] directly predicts inliers and outliers based on how strongly a correspondence belongs to the main cluster in SC matrix. FGR [67] filters outliers by checking the compatibility between randomly sampled tuples as pre-processing. In CG-SAC [10], the SC is considered as the guidance for efficient sampling of the RANSAC pipeline. SAC-COT [12] further designs a compatibility triangle to benefit the sampling in the early iteration stage. In [68] and [69], a two-stage voting scheme is developed by ranking the geometric consistency in the global and local scope. Besides, the SC is also combined with deep learning techniques in more recent works. PointDSC [9] designs a non-local module taking advantage of the guidance of SC. DHVR [11] generates the hypotheses for deep Hough voting by the SC-validated tuples. TriVoC [70] reformulates the SC-based inlier voting into a deep learning framework. MI-PCR [71] uses the SC to build a learning pipeline for multi-instance point cloud registration. In this article, we try to solve the ambiguity problem when applying SC for efficient sampling.

C. Non-Feature-Matching-Based Methods

In addition to using feature matching, some methods adopt non-feature-matching pipelines to achieve end-to-end registration. The widely used Iterative Closest Point (ICP) algorithm [13] divides point cloud registration into two iterative sub-problems: establishing correspondences by searching the closest neighbor in the coordinate space and solving the rigid transformation by Singular Value Decomposition (SVD). Following the ICP pipeline, some methods try to speed up the searching and convergence process with different strategies, such as designing the point-to-plane error term [72] and adding a probabilistic model to the cost construction [73]. ICP and its variants have a simple form to be applied, but they are sensitive to the initial perturbation. To solve this drawback, some branch-and-bound (BnB) based methods are proposed, such as Go-ICP [74], Gogma [75], Gosma [76]. These methods alleviate the local minimum problem through global optimization, but they are time-consuming and not practical in some extreme scenarios.

Recently, some researchers also try to develop learning-based end-to-end registration frameworks. PointNetLK [77] modifies the Lucas & Kanade (LK) algorithm as a recurrent neural network and combines it with the PointNet [35] to achieve end-to-end registration. DCP [78] utilizes an attention-based module to learn the soft correspondences, and performs weighted SVD to get the rigid transformation. Following them, some methods integrate the optimal transport algorithm [79], graph matching framework [80] or Gaussian Mixture Model [81] to better learn the soft correspondence matrix. PR-Net [82] and OM-Net [83] try to solve the partial-to-partial problem by predicting key points and overlapping masks. RegTR [84] uses attention mechanisms to replace the role of explicit feature matching and RANSAC to directly predict the final set of correspondences.

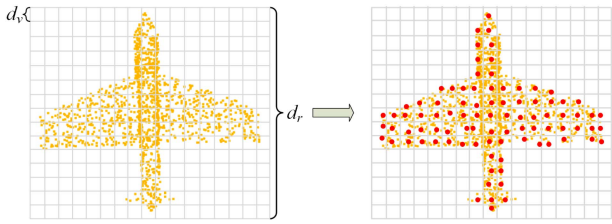


Fig. 2. The voxel grid based point cloud down-sampling. d_v is the length of a voxel grid, and d_r is the size of the voxelized scene. The red points in the right part are the retained points after down-sampling.

III. Background

A. Problem Definition

Given two point clouds to be aligned, we first build voxel grids to downsample the point clouds, as shown in Fig. 2, and the downsampled point clouds are denoted as source point cloud $\mathcal{X} \in \mathbb{R}^{N\times 3}$ and target point cloud $\mathcal{Y} \in \mathbb{R}^{M\times 3}$. Then, we use either hand-crafted or deep learning-based descriptors to extract feature descriptors for both of them, i.e. $\mathcal{F}_\mathcal{X} \in \mathbb{R}^{N\times D}$ and $\mathcal{F}_\mathcal{Y} \in \mathbb{R}^{M\times D}$. The downsampled point clouds with the extracted feature descriptors are taken as input to the proposed SC^2-PCR++. In our method, we first generate the correspondences by performing feature matching on the feature descriptors. Then, the proposed method finds the correct correspondences and estimates the rigid transformation between the two point clouds, i.e., the rotation matrices ($R \in \mathbb{R}^{3\times 3}$) and the translation vectors ($t \in \mathbb{R}^3$).

B. Overview

Our method can be considered as a variant of RANSAC, so we first briefly review the RANSAC and then clarify the difference between our method and RANSAC. RANSAC adopts the *generation-and-selection* pipeline for robust model fitting. In the generation step, it randomly samples the minimal set to generate massive potential rigid transformations for the two point clouds. In the selection step, it computes the inlier count (IC) metric for each estimation and returns the estimation with the highest IC as the final result. In the proposed SC^2-PCR++, the random sampling step is replaced by the second-order spatial compatibility (SC^2) measure guided sampling to improve the robustness and efficiency of the model generation, and the IC metric is replaced by the proposed Feature and Spatial consistency constrained Truncated Chamfer Distance metric (FS-TCD).

The remainder of this paper is organized as follows: In Section IV, we review the Spatial Compatibility (SC) measure and introduce the proposed SC^2 measure. Then, we compare the robustness of using them for model generation from a probabilistic view. In Section V, we describe the proposed FS-TCD. After that, we illustrate the whole pipeline of the SC^2-PCR++ in Section VI. Finally, we conduct extensive experiments to validate the performance in Section VII, followed by concluding remarks in Section VIII.

IV. Second-Order Spatial Compatibility

In this section, we first briefly review the commonly used Spatial Compatibility (SC) measure [9], [10], [11], [66]. Then, we introduce the new SC^2 approach proposed in this article. It is designed for measuring the similarity of correspondences to improve the quality of sampling. We consider that the robustness of a measure-based sampling method can be reflected in the probability of an ambiguity event. Here, the probability of ambiguity event is defined as:

$$\mathrm{P}_{am}(M) = \mathrm{P}(M_{in,out} > M_{in,in}), \tag{1}$$

where M is a specific metric for measuring correspondence-wise similarity. $\mathrm{P}(\mathrm{Z})$ is the probability of an event Z (For convenience we use this notation in the following part). $M_{in,out}$ is the similarity between an inlier and an outlier, while $M_{in,in}$ is the similarity between two inliers. When $M_{in,out} > M_{in,in}$, the outlier is a closer neighbor compared to the inlier, and the sampling in this case tends to fail. For example, as shown in Fig. 1, c1 is an inlier. When we use the SC measure as the guidance for sampling, we try to find the correspondences with higher similarity with c1 in Fig. 1(b). However, c7, which is an outlier, has higher similarity with c1 than other inliers, so it will also be selected to estimate the rigid transformation with c1. In this case, although c1 is an inlier, using it and its neighbors can not lead to a correct estimation. Instead, for the SC^2 matrix in Fig. 1(c), the similarity between c1 and other inliers is higher than outliers, which makes the near neighbors of c1 all inliers in SC^2 measure. So using c1 with its neighbors in the SC^2 matrix can result in a correct estimation. Thus, the lower the probability in (1) is, the method will be more robust to noises and the metric-based sampling will be more effective. Next, we will compare the ambiguity probability of the SC and SC^2 measure on 3DMatch dataset as an example.

A. Review of Spatial Compatibility

The Spatial Compatibility (SC) measure between correspondence i and j is defined as follows:

$$SC_{ij} = \phi(d_{ij}), d_{ij} = \mid d(x_i, x_j) - d(y_i, y_j) \mid, \tag{2}$$

in which (x_i, y_i) and (x_j, y_j) are the matched points of correspondences i and j. $\phi(\cdot)$ is a monotonically decreasing kernel function. $d(\cdot, \cdot)$ is the euclidean distance. As shown in Fig. 1, the distance difference between two inliers $d_{in,in}$ should be equal to 0 due to the length consistency of rigid transformation. However, because of the noises introduced by data acquisition and point cloud downsampling (Fig. 2), $d_{in,in}$ is not exactly equal to 0, but less than a threshold d_{thr}. Referring to [9], an approximate value of d_{thr} is twice of the voxel size (d_v in Fig. 2). For convenience, we assume that $d_{in,in}$ is uniformly distributed over d_{thr} and get the probability density function (PDF) of the distance difference between two inliers as follows:

$$\mathrm{PDF}_{in,in}(l) = 1/d_{thr}, 0 \leq l \leq d_{thr}. \tag{3}$$

Differently, there is no related constraint between two outliers or an inlier and an outlier due to the random distribution of outliers. We consider the distance difference between two unrelated points to be identically distributed and assume the probability density function (PDF) as $F(\cdot)$:

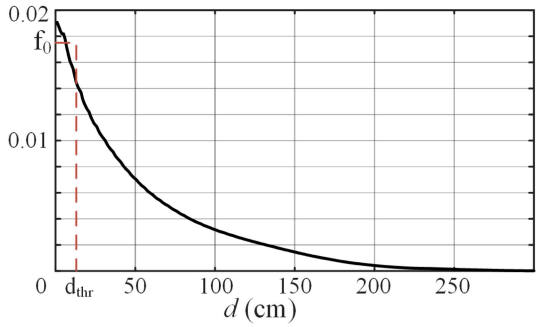


Fig. 3. The empirical probability density function (F) of the distance difference between two unrelated correspondences, i.e., $d_{in,out}$ and $d_{out,out}$.

$$\mathrm{PDF}_{in,out}(l) = F(l), \mathrm{PDF}_{out,out}(l) = F(l); 0 \leq l \leq d_r, \tag{4}$$

where d_r is the range of $d_{in,out}$ and $d_{out,out}$. In fact, it is hard to directly express F into a specific formula due to the randomness of the outliers. To make it clear, we plot the empirical F on 3DMatch dataset in Fig. 3. As we can see, the d_{thr} is much smaller than the size of voxelized scene d_r. For convenience, we assume that $F(l)$ is in uniform distribution within $(0, d_{thr})$ and the value of f_0 is the quotient of the integral in $(0, d_{thr})$ and d_{thr}. Then, we can get:

$$F(l) = f_0, 0 \leq l \leq d_{thr}. \tag{5}$$

Note that we can also model the F in $(0, d_{thr})$ as other forms, and it does not change the conclusion. The reason is that we just have to compute the integral in $(0, d_{thr})$, and f_0 is obtained by the integral in $(0, d_{thr})$. So modeling $F(l)$ as a constant does not change the integral result. Next, we compute the ambiguity probability of SC as (1), i.e., $\mathrm{P}(SC_{in,out} > SC_{in,in})$. According to (2), (3), (4) and (5), it can be computed as follows:

$$\begin{aligned}\mathrm{P}(SC_{in,out} > SC_{in,in}) &= \mathrm{P}(d_{in,out} < d_{in,in}) \\ &= \int_0^{d_{thr}} \int_0^{l} \mathrm{PDF}_{in,in}(l) \cdot \mathrm{PDF}_{in,out}(x) dx dl \\ &= \int_0^{d_{thr}} \int_0^{l} \frac{1}{d_{thr}} \cdot f_0 dx dl = \frac{d_{thr} \cdot f_0}{2}.\end{aligned} \tag{6}$$

B. Second-Order Spatial Compatibility

Next, we describe the proposed second-order spatial compatibility measure ($SC^2 \in \mathbb{R}^{N \times N}$). Specifically, we first build a hard compatibility matrix C ($C \in \mathbb{R}^{N \times N}$):

$$C_{ij} = \begin{cases} 1; d_{ij} \leq d_{thr}, \\ 0; d_{ij} > d_{thr}. \end{cases} \tag{7}$$

C considers that two correspondences satisfying length consistency are compatible ($C_{i,j}$ is set to 0 when $i = j$). Then, SC^2_{ij} counts the number of common compatibility correspondences

of i and j when they are compatible, as follows:

$$SC_{ij}^2 = C_{ij} \cdot \sum_{k=1}^{N} C_{ik} \cdot C_{kj}. \tag{8}$$

Similarly, we analyze the ambiguity probability of SC^2, i.e., $\mathrm{P}(SC_{in,out}^2 > SC_{in,in}^2)$. For convenience, suppose there are N pairs of correspondences and the inlier ratio is α.

Remark 1. The ambiguity probability of SC^2 measure, i.e., $\mathrm{P}(SC_{in,out}^2 > SC_{in,in}^2)$, can be written as follows:

$$\begin{aligned} &\mathrm{P}(SC_{in,out}^2 > SC_{in,in}^2) = p \cdot \mathrm{P}(X > (N \cdot \alpha - 2)), \\ &X \sim S((N\alpha - 1)p + (N(1-\alpha) - 1)p^2, N(1-\alpha)p^2), \\ &p = d_{thr} \cdot f_0, \end{aligned} \tag{9}$$

where $S(\cdot, \cdot)$ is the Skellam distribution [85], [86], [87].

Derivation of Remark 1: We first reformulate (8) as follows:

$$SC_{ij}^2 = C_{ij} \cdot M_{ij}, M_{ij} = \sum_{k=1}^{N} C_{ik} \cdot C_{kj}. \tag{10}$$

M_{ij} counts the quantity of the commonly compatible correspondences of i and j in the global set. According to (7) and (3), we can obtain that:

$$\mathrm{P}(C_{in,in} = 1) = 1. \tag{11}$$

According to (7), (4) and (5), we can get that

$$\mathrm{P}(C_{in,out} = 1) = \int_0^{d_{thr}} F(l)dl = d_{thr} \cdot f_0 = p. \tag{12}$$

$$\mathrm{P}(C_{out,out} = 1) = \int_0^{d_{thr}} F(l)dl = d_{thr} \cdot f_0 = p. \tag{13}$$

According to (10), to make $SC_{in,out}^2 > SC_{in,in}^2$ hold, two conditions need to be met: $C_{in,out} = 1$ and $M_{in,out} > M_{in,in}$. According to (12), we can obtain the following equation:

$$\begin{aligned} &\mathrm{P}(SC_{in,out}^2 > SC_{in,in}^2) \\ &= \mathrm{P}(C_{in,out} = 1) \cdot \mathrm{P}(M_{in,out} > M_{in,in}) \\ &= p \cdot \mathrm{P}(M_{in,out} > M_{in,in}). \end{aligned} \tag{14}$$

Next, we compute the distribution of $M_{in,out}$ and $M_{in,in}$. Since inliers have different distributions from outliers, we compute them separately and reformulate (10) as follows:

$$M_{ij} = \sum_{m \in \mathcal{I}} C_{im} \cdot C_{mj} + \sum_{n \in \mathcal{O}} C_{in} \cdot C_{nj}, \tag{15}$$

where $\mathcal{I}$ is the inlier set while $\mathcal{O}$ is the outlier set. (For convenience we use this notation in the following part).

We first discuss the value in M matrix between two inliers, i.e. $M_{in,in}$. According to (11), we can find that any two inliers are compatible. Thus, when correspondence i and j are inliers, the number of correspondences compatible with both of them in the inlier set is the number of inliers excluding themselves ($C_{ii} = 0, C_{jj} = 0$), i.e.:

$$\sum_{m \in \mathcal{I}} C_{im} \cdot C_{mj} = N \cdot \alpha - 2; i \in \mathcal{I}, j \in \mathcal{I}, \tag{16}$$

where α is the inlier rate. For outliers, according to (12), the probability that an outlier is compatible with an inlier is p. Then the probability that an outlier is compatible with both i and j is p^2. The number of outliers in the whole correspondence set is $N(1-\alpha)$. So the number of correspondences compatible with both of them in the outlier set is in a Bernoulli distribution [88] as follows:

$$\sum_{n \in \mathcal{O}} C_{in} \cdot C_{nj} \sim B(N(1-\alpha), p^2); i \in \mathcal{I}, j \in \mathcal{I}, \tag{17}$$

where $B(\cdot, \cdot)$ is the Bernoulli distribution. Thus, $M_{in,in}$ is in the following distribution:

$$M_{in,in} \sim N \cdot \alpha - 2 + B(N(1-\alpha), p^2). \tag{18}$$

After that, we discuss the distribution of the value in M matrix between an inlier and an outlier, i.e., $M_{in,out}$. For convenience, we assume correspondence i is an inlier while j is an outlier. For the inlier set except correspondence i ($C_{ii} = 0$), any of them is compatible with i (11), and the probability that one of them is compatible with j is p (12). So the number of correspondences compatible with both correspondence i and j in the inlier set is in the following distribution:

$$\sum_{m \in \mathcal{I}} C_{im} \cdot C_{mj} \sim B(N\alpha - 1, p); i \in \mathcal{I}, j \in \mathcal{O}. \tag{19}$$

Meanwhile, for each outlier except correspondence j ($C_{jj} = 0$), the probability that it is compatible with i or j are both p according to (12) and (13). So the probability that an outlier is both compatible with i and j is p^2. Thus, we can get the following distribution:

$$\sum_{n \in \mathcal{O}} C_{in} \cdot C_{nj} \sim B(N(1-\alpha) - 1, p^2); i \in \mathcal{I}, j \in \mathcal{O}. \tag{20}$$

So the distribution of $M_{in,out}$ is as follows:

$$M_{in,out} \sim B(N\alpha - 1, p) + B(N(1-\alpha) - 1, p^2). \tag{21}$$

Since p is a small value, the Binomial distribution in (18) and (21) can be approximately equivalent to the Poisson distribution [88], i.e.:

$$\begin{aligned} &M_{in,in} \sim N \cdot \alpha - 2 + \pi(N(1-\alpha)p^2), \\ &M_{in,out} \sim \pi((N\alpha - 1)p) + \pi((N(1-\alpha) - 1)p^2), \end{aligned} \tag{22}$$

where $\pi(\cdot)$ is the Poisson distribution. Furthermore, for two Poisson distribution: $X_1 \sim \pi(\lambda_1)$ and $X_2 \sim \pi(\lambda_2)$, their sum is also in the Poisson distribution [88] as follows:

$$X1 + X_2 \sim \pi(\lambda_1 + \lambda_2). \tag{23}$$

So we can convert $M_{in,out}$ in (22) into following form:

$$M_{in,out} \sim \pi((N\alpha - 1)p + (N(1-\alpha) - 1)p^2). \tag{24}$$

Meanwhile, we can convert $\mathrm{P}(M_{in,out} > M_{in,in})$ into following form:

$$\begin{aligned} &\mathrm{P}(M_{in,out} > M_{in,in}) \\ &= \mathrm{P}(M_{in,out} - M_{in,in} > 0) \\ &= \mathrm{P}(X > N \cdot \alpha - 2), \end{aligned} \tag{25}$$

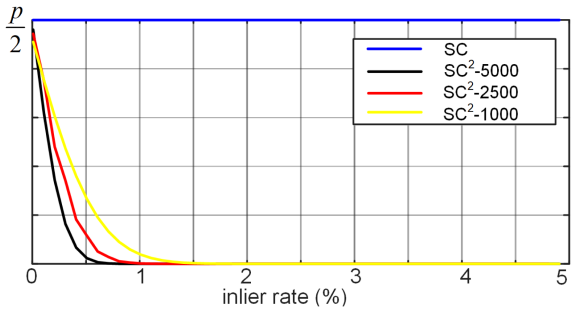


Fig. 4. The probability of ambiguity event. SC is spatial consistency measure. SC^2-N ($N = 5000, 2500, 1000$) is the second-order spatial consistency measure with N correspondences.

where X is in the following distribution:

$$\pi((N\alpha - 1)p + (N(1-\alpha) - 1)p^2) - \pi(N(1-\alpha)p^2). \tag{26}$$

For two Poisson distribution: $X_1 \sim \pi(\lambda_1)$ and $X_2 \sim \pi(\lambda_2)$, their difference is in the Skellam distribution [85], [86], [87], i.e.:

$$X1 - X_2 \sim S(\lambda_1, \lambda_2). \tag{27}$$

So the distribution of X in (26) can be converted as follows:

$$S((N\alpha - 1)p + (N(1-\alpha) - 1)p^2, N(1-\alpha)p^2). \tag{28}$$

Combining (14), (25) and (28), we compute the value of $\mathrm{P}(SC^2_{in,out} > SC^2_{in,in})$ as (9).

C. Ambiguity Probability Comparison

As derived above, the ambiguity probabilities of SC and SC^2 measures are shown in (6) and (9) respectively. The ambiguity probability of the SC measure is a constant independent of the inlier rate α. Take the 3DMatch [29] dataset as an example. Following [9], we set $d_{thr} = 2 \cdot d_v = 10$ cm, then the ambiguity probability of SC measure is about 0.1 according to (6). Considering the number of outliers might be large, the impact of wrong correspondences is not negligible even at this probability. Differently, according to (9), the ambiguity probability of SC^2 measure is related to the inlier rate α and the correspondence number N. According to the properties of Skellam distribution, the value of $\mathrm{P}(SC^2_{in,out} > SC^2_{in,in})$ is going to approach 0 very quickly as α increases.

In order to make a clearer comparison between the proposed SC^2 measure and the previous SC measure, we plot the curves of ambiguity probability with respect to the inlier rate α for both of them according to (6) and (9) on 3DMatch benchmark. Since the ambiguity probability of SC^2 measure is related to the correspondence number N, we plot its curves with different N. For the value of Skellam distribution, we use the Scipy library [89], which is a math kit, to compute it. As shown in Fig. 4, the ambiguity probability of the proposed SC^2 measure is significantly lower than the SC measure, even when the inlier rate is close to 0. It shows that using SC^2 measure as guidance for sampling is easier to obtain an outlier-free set. When the inliers rate reaches 1%, the ambiguity probability of SC^2 measure is close to 0, which ensures a robust sampling on the data with a low inlier rate.

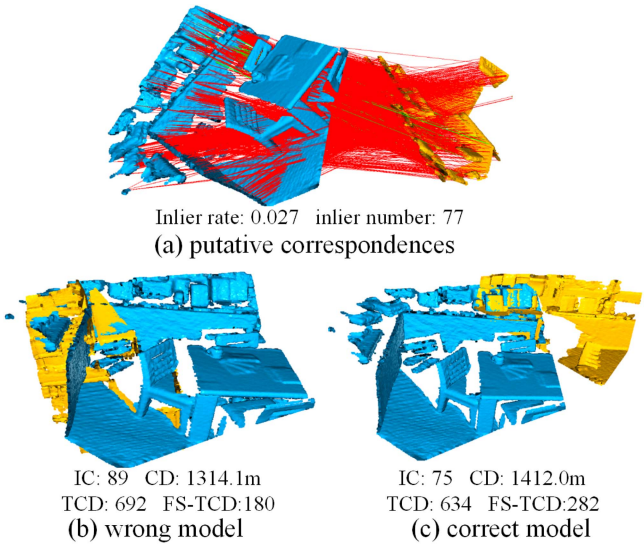


Fig. 5. (a) The putative correspondences of a pair of point clouds to be aligned, in which the inlier rate is low. (b) The wrong model selected by the IC metric. (c) The correctly estimated model. For IC, TCD and FS-TCD, higher is better. For CD, lower is better. The IC, CD, and TCD metrics can not instruct the correct model. Using the proposed FS-TCD metric can select the correct model.

V. Constrained Truncated Chamfer Distance

A. Challenges of Model Selection

In RANSAC, after generating some hypotheses, the inlier count (IC) is adopted as the metric for selecting the best estimation among them. The inlier count is computed based on a set of putative correspondences. More concretely, suppose there are two point clouds with pre-computed feature descriptors to be aligned: source point cloud $\mathcal{X} \in \mathbb{R}^{N\times 3}$ and target point cloud $\mathcal{Y} \in \mathbb{R}^{M\times 3}$. Formally, it first forms N pairs of correspondences by finding the nearest neighbor for each source point among target points. Then, the inlier count (IC) for evaluating k-th hypothesis (rotation R_k and translation t_k) is defined as follows:

$$\mathrm{IC}_k = \sum_{i=1}^{N} [\|R_k x_i + t_k - y_i\| < \tau], \tag{29}$$

where N is the number of putative correspondences. (x_i, y_i) is a pair of correspondence. $[\cdot]$ is the Iverson bracket, i.e., if the condition in [] is true, then it returns 1, and otherwise returns 0. Once R_k and t_k are correctly estimated, the IC_k should be close to the number of inliers. However, when there are only rare inliers in the putative correspondences, the IC value of the correct model is also small. An example of 3DLoMatch dataset is shown in Fig. 5. Fig. 5(a) shows a pair of point clouds to be aligned, with the putative correspondences. The inlier rate and inlier number in the putative correspondences are 0.027 and 77 respectively. Fig. 5(b) is the final model selected by the IC metric, and Fig. 5(c) is one of the correctly estimated models. Since the inlier number is only 77, the IC value of the correct model is

75, while IC value of the wrong model is 89 by chance. In this case, the best hypothesis can not be selected using the IC metric, leading to the failure estimation of point cloud registration.

The drawback of the IC metric is the dependence on putative correspondences and the lack of global alignment information. In the IC metric, the correspondences are counted independently without considering the spatial position relationship between them. In fact, if the rigid transformation between two point clouds is correctly estimated, then a continuous area should be aligned, which means it should be measured by a non-local metric. Therefore, we expect to introduce the Chamfer distance (CD) [14], [15], [16] metric to address the issue of IC metric. We first review the definition of CD metric. Generally, the one-way CD metric for R_k and t_k in point cloud registration is defined as follows:

$$\mathrm{CD}_k = \sum_{i=1}^{N} \min_{y_j \in \mathcal{Y}} \|R_k x_i + t_k - y_j\|. \tag{30}$$

For each source point, it finds the nearest neighbor in the target point set after aligning by R_k and t_k, and computes the distance between them. A smaller CD value measures a better alignment quality of the two point clouds after aligning. However, it is not suitable to directly use the CD metric for selecting the best hypothesis from the following three aspects: 1) For each source point, CD needs to search nearest neighbor in the whole set of target points. Considering that thousands of hypotheses could be generated and the selection is performed on them, computing CD for them is time-consuming. 2) CD metric assumes the two point clouds have a similar shape after aligning. However, when the two point clouds have low-overlapped areas, this assumption is not valid. 3) Since the CD metric does not consider the feature information, the nearest neighbor in the coordinate space could be a wrong alignment. This causes the CD to be sensitive to noises. As shown in Fig. 5, the wrong model in Fig. 5(b) achieves a lower CD metric than the correct model in Fig. 5(c). This also reveals the problem of using CD as the selection metric.

B. Constrained Truncated Chamfer Distance

Based on the above observations, we propose a Feature and Spatial consistency constrained Truncated Chamfer Distance metric, named FS-TCD, to address both the problems of CD and IC. Specifically, we first reformulate the CD metric as a truncated form (TCD):

$$\mathrm{TCD}_k = \sum_{i=1}^{N} [(\min_{y_j \in \mathcal{Y}} \|R_k x_i + t_k - y_j\|) < \eta], \tag{31}$$

where $[\cdot]$ is also an Iverson bracket. For each x_i, it finds the nearest neighbor among the target points after being aligned by the R_k and t_k, and computes the number of neighbor pairs whose alignment error is less than the threshold η. The reason for threshold truncation is that the shapes of the two point clouds are not exactly the same, so they cannot be perfectly aligned. Therefore, it is meaningless to calculate the alignment error between point clouds that are not in the overlapping region. Intuitively, TCD reflects the size of the overlap area between the two clouds after alignment, which measures the global quality of alignment. The greater the overlap between the two point clouds, the more likely they are to be aligned correctly.

In TCD, when R_k and t_k are incorrectly estimated, it is still possible that the alignment area is miscounted because some points are incorrectly aligned together by chance. In this case, it is likely to compute an abnormally big value of TCD for the wrong model, and select it as the final result. As shown in Fig. 5, the wrong model in Fig. 5(b) has a bigger TCD value than the correct model in Fig. 5(c). To suppress this situation, we introduce two constraints on the TCD. The first one is the feature matching constraint. Specifically, we first build a relaxed hard matching matrix $H \in \mathbb{R}^{N \times M}$ between the source point cloud and target point cloud by the feature descriptor information. We adopt a simple top-K strategy for building H matrix: if y_j is a top-K neighbor of x_i in feature space, then $H_{ij} = 1$. Otherwise, $H_{ij} = 0$. Note that each source point x_i can be matched with K points in target points, so H matrix represents a relaxed feature matching relationship. The H matrix is utilized as the feature constraint to build the F-TCD as follows:

$$\text{F-TCD}_k = \sum_{i=1}^{N} [(\min_{H_{ij}=1} \|R_k x_i + t_k - y_j\|) < \eta]. \tag{32}$$

In F-TCD, we search the nearest neighbor for each x_i based on the relaxed feature matching relationship represented by the H matrix. If we can find a y_j to ensure that the alignment error of (x_i, y_j) is less than η, and $H_{ij} = 1$ holds in the H matrix, then we consider x_i and y_j are successfully registered. We check all the source points and count the number of correctly registered pairs as F-TCD metric. F-TCD searches the neighbor by the H matrix instead of in the global scope, making it less expensive to calculate. Meanwhile, due to the introducing of the feature information, F-TCD is more robust to noises. Compared with the IC metric, F-TCD reduces the dependence on the inlier rate of putative matching. Feature information is not used to establish one-to-one matches in F-TCD, but to establish relaxed one-to-many matches. In fact, we can see that IC is a special case of F-TCD. When we set $K = 1$ and $\eta = \tau$, then F-TCD has the same formulation as IC.

We further integrate the spatial consistency constraint into F-TCD to obtain the proposed FS-TCD metric. As mentioned in Section IV-A, the inlier correspondences satisfy the spatial consistency. So we check the established match pairs in F-TCD, and find mutually compatible matches. The number of validated matches is viewed as the FS-TCD value. It further removes the potential mismatches which are possibly considered by the metric. As a result, the FS-TCD value of the model in Fig. 5(c) is greatly bigger than that of the model in Fig. 5(b), which helps to select the best result.

VI. Pipeline

The pipeline of the proposed method is shown in Fig. 6. It can be divided into two main components: model generation and model selection. Different from the classic RANSAC, we use different matching strategies in model generation and model selection. In the model generation process, for each point in the source point cloud, we find its nearest neighbor in the feature

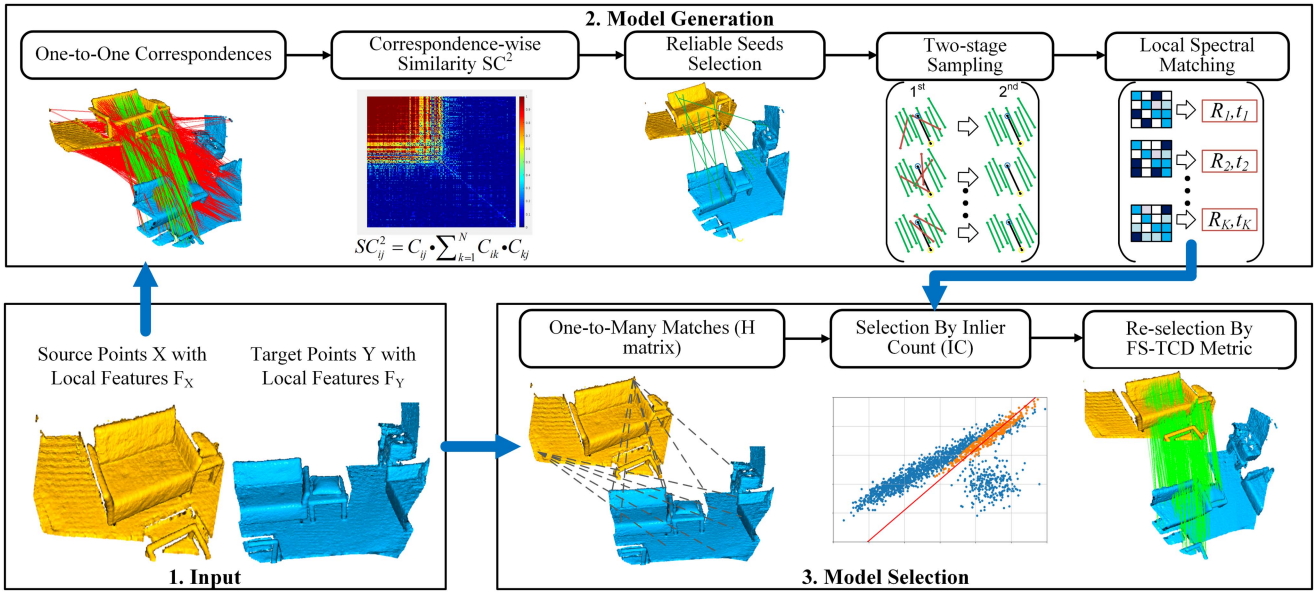


Fig. 6. Pipeline of our method. The input is the source and target points with extracted features. The proposed method rebuilds the model generation and model selection process of the classic RANSAC.

space among the target points to form N pairs of putative correspondences. Then, we use the proposed SC^2 measure to compute the correspondence-wise similarity matrix, and utilize it to guide the sampling for model generation. In the model selection process, we relax the matching condition, and allow one-to-many matching relationships to avoid failed selections due to the low inlier rate. Next, we will describe the model generation and model selection in detail.

A. Model Generation

First, we use the proposed SC^2 measure to compute the correspondence-wise similarity matrix. Then, we use a spectral matching technique with Non-Maximum Suppression (NMS) to select some reliable correspondences, termed seeds. Next, we propose a two-stage sampling strategy to build a consensus set for each seed. After that, we perform local spectral matching among the consensus set of each seed, and generate an estimation of rigid transformation (rotation R and translation t).

Reliable Seed Selection: As mentioned in Section IV, there are high similarities between inlier correspondences by the proposed SC^2 measure. Then, as long as we find an inlier correspondence, we can construct a consensus set by finding its k nearest neighbors in the metric space. Obviously, traversing all the correspondences must find an inlier, but it is not necessary. We only need to pick some reliable points called seed points to accelerate the registration process. We perform the spectral matching technique [66] to select seed points. Specifically, we first build the similarity matrix for all of the correspondences and normalize the value in the matrix to 0-1, following [66]. Then, following [9], [66], the association of each correspondence with the leading eigenvector is adopted as the confidence for this correspondence. The leading eigenvector is solved by the power iteration algorithm [90]. In order to ensure an even distribution of seed points, the correspondences with local maximum confidence score within its neighborhood of radius R are selected. The number of seed points (N_s) is determined by a proportion of the number of whole correspondences.

Two-Stage Consensus Set Sampling: As some seed points are selected, we extend each of them into a consensus set. We adopt a two-stage selection strategy to perform a coarse-to-fine sampling. In the first stage, we select K_1 correspondences for each seed by finding its top-K_1 neighbors in the SC^2 measure space. As mentioned before, the ambiguity probability $\mathrm{P}(SC^2_{in,out} > SC^2_{in,in})$ is very small. Thus, when a seed is an inlier correspondence, the consensus set also mainly contains inliers. Meanwhile, the similarity expressed by SC^2 measure focuses on global information instead of local consistency. Therefore, the neighbors selected in the SC^2 measure space are distributed more evenly rather than clustered together, which benefits the estimation of rigid transformation [9].

The second stage of the sampling operation is adopted to further filter potential outliers in the set obtained in the first stage. The SC^2 matrices are reconstructed within each set produced by the first stage instead of the whole set. We select top-K_2 ($K_2 < K_1$) correspondences of the seed by the newly constructed local SC^2 matrices. As shown in Fig. 4, since the higher inlier rate ensures a lower ambiguity probability, the potential outliers can also be further pruned. Note that we only discussed the case that the seed point is an inlier. In fact, when the seed point is an outlier, it can also form a local consistency, especially when there are aggregated false matches in the correspondence set. We encourage these sets to also generate hypotheses and filter them at the final hypothesis selection step (Section VI-B) rather than at the early stage. In this way, we can avoid some correct assumptions being filtered out early.

Local Spectral Matching: In this step, we perform the weighted SVD [91] on the consensus set to generate an estimation of rigid transformation for each seed. Although the previously proposed sampling strategy can obtain outlier-free correspondence set, we find that the weighted SVD achieves better performance than treating all correspondences equally. This may be because the inliers still have different degrees of noise. So correspondences with bigger noises should have smaller weights when estimating rigid transformation. Traditional spectral matching [66] method analyzes the SC matrix to assign a weight for each correspondence, which is affected by ambiguity problem [9]. Since the proposed SC^2 measure is more robust against ambiguity, we also replace the SC matrix with the SC^2 measure.

Specifically, for each consensus set, we build a local graph by considering each correspondence in the consensus set as a node, and the SC^2 value between the correspondences as edge. In order to facilitate matrix analysis, we convert the SC^2 measure into soft form ($\tilde{SC^2}$) as follows:

$$\begin{aligned} \tilde{SC^2} &= \tilde{C} \cdot (\tilde{C} \times \tilde{C}), \\ \tilde{C}_{ij} &= \mathrm{ReLU}(1 - d_{ij}^2/d_{thr}^2), (1 \le i \le K_2, 1 \le j \le K_2) \end{aligned} \tag{33}$$

where $\cdot$ is Hadamard product and $\times$ is matrix product. The $\tilde{SC^2}$ has a similar property with SC^2, i.e., there are higher similarity values between inliers and inliers than that between inliers and outliers. Thus, the inliers in the graph are clustered together. Then we conduct local spectral decomposition on the adjacent matrix of the local graph, i.e., $\tilde{SC^2}$, to obtain a weight w_i for correspondence i. According to [9], [66], the leading eigenvector of matrix $\tilde{SC^2}$ can be considered as the association of each correspondence with a main cluster. In the consensus set, when the seed is an inlier, this set mainly contains inliers, so the main cluster is the inlier set. Thus, the association value with the main cluster, i.e. the leading eigenvector can be explained as the inlier probability. Therefore, we use the leading eigenvector of the $\tilde{SC^2}$ as the weight for SVD. In our method, we use the power iteration algorithm [90] to efficiently compute the leading eigenvector. Finally, the rotation R_k and translation t_k of seed k are computed by performing weighted SVD [20] within its consensus set.

B. Model Selection

After model generation, we select the best estimation over the rigid transformations produced by all the consensus sets. We use the proposed Feature and Spatial consistency constrained Truncated Chamfer Distance (FS-TCD) metric to select the final estimation. As shown in Fig. 6, before obtaining the FS-TCD for each estimation, we first build the relaxed matching between source points and target points, represented as the H matrix, as described in Section V. Compared with Chamfer Distance (CD), FS-TCD is more efficient due to the reduction of the nearest neighbor search area. However, FS-TCD is still more time-consuming than Inlier Count (IC). In order to reduce unnecessary calculations of FS-TCD to accelerate the selection process, we use the IC to remove some spurious solutions in advance.

More concretely, as described in Section VI-A, we choose N_s seeds with consensus set, resulting in N_s hypotheses. Then, for the estimation of k-th seed R_k and t_k, we compute the IC metric, and retain N_s' estimations with the highest IC scores. After that, the FS-TCD metric is figured for each of the remaining hypotheses, and the R_k^* and t_k^* with the highest FS-TCD is selected as the final result.

VII. Experiment

A. Datasets and Experimental Setup

Indoor Scenes: We use the 3DMatch benchmark [29] for evaluating the performance on indoor scenes. It contains 1623 pairs of point clouds with ground-truth camera poses, which are obtained by 8 different RGBD sequences. For each pair of point clouds, we set the voxel size d_v as 5 cm to downsample the point cloud. Then we extract the local feature descriptors and match them to form the putative correspondences. In order to test the performance of each algorithm more comprehensively, we use FPFH [27] (handcrafted descriptor) and FCGF [30] (learning-based descriptors) as feature descriptors respectively.

Partial overlapping is challenging in point cloud registration. In order to further test the performance of our method, 3DLoMatch benchmark [41] is adopted to further verify the performance of the algorithm on low-overlapped point cloud registration. It contains 1781 pairs of point clouds with low overlapping. Following [9], [11], we use FCGF and Predator to generate putative correspondences.

Outdoor Scenes: The KITTI dataset [93] is composed of 11 outdoor driving scenarios of point clouds. Following [20], [30], we choose the 8 to 10 scenarios as test datasets. For all the LIDAR scans, we use the first scan that is taken at least 10 cm apart within each sequence to create a pair, which can obtain 555 pairs of point clouds for testing. Then we construct 30 cm voxel grids ($d_v = 30$ cm) to downsample the point cloud and form the putative correspondences by FPFH and FCGF respectively.

Multi-Way Registration Dataset: We use the Augmented ICL_NUIM [94], [95] dataset for testing the performance on the multi-way registration task. The dataset contains four scenes of indoor environments: two sequences of living rooms and two sequences of offices. For each pair of point clouds, we also use 5 cm voxel grids ($d_v = 5$ cm) to downsample the point cloud and extract FPFH descriptors.

Evaluation Criteria: We first report the registration recall (RR) under an error threshold (the unit of RR is % in the following experiments). For the indoor scenes, the threshold is set to (15 deg, 30 cm), while the threshold for outdoor scenes is (5 deg, 60 cm). For a pair of point clouds to be aligned, we calculate the errors of translation and rotation estimation separately. We compute the isotropic rotation error (RE) [96] and L2 translation error (TE) as follows:

$$\mathrm{RE} = \mathrm{acos}\left(\frac{\mathrm{trace}(\hat{\mathrm{R}}^{-1}\mathrm{R}) - 1}{2}\right), \mathrm{TE} = \|\mathrm{t} - \hat{\mathrm{t}}\|_2, \tag{34}$$

TABLE I
QUANTITATIVE RESULTS ON 3DMATCH DATASET. THE METRIC WITH ↑ MEANS THAT HIGHER IS BETTER, WHILE A ↓ MEANS THE OPPOSITE. METHODS WITH * ARE CORRESPONDENCE-FREE METHODS

	FPFH (traditional descriptor)						FCGF (learning-based descriptor)						
	RR(%) ↑	RE(deg) ↓	TE(cm) ↓	IP(%)↑	IR(%)↑	F1(%) ↑	RR(%) ↑	RE(deg) ↓	TE(cm) ↓	IP(%)↑	IR(%)↑	F1(%) ↑	Time (s)
DCP* [78]	-	-	-	-	-	-	3.22	8.42	21.40	-	-	-	0.07
PointNetLK* [77]	-	-	-	-	-	-	1.61	8.04	21.30	-	-	-	0.12
OM-Net* [83]	-	-	-	-	-	-	35.90	4.16	10.50	-	-	-	0.08
RegTR* [84]	-	-	-	-	-	-	92.00	1.57	4.90	-	-	-	0.18
3DRegNet [19]	26.31	3.75	9.60	28.21	8.90	11.63	77.76	2.74	8.13	67.34	56.28	58.33	0.05
DGR [20]	32.84	2.45	7.53	29.51	16.78	21.35	88.85	2.28	7.02	68.51	79.92	73.15	1.53
DHVR [11]	67.10	2.78	7.84	60.19	64.90	62.11	91.93	2.25	7.08	80.20	78.15	78.98	3.92
PointDSC [9]	77.57	2.03	6.38	68.45	71.56	69.75	92.85	2.08	6.51	78.91	86.23	82.12	0.10
SM [66]	55.88	2.94	8.15	47.96	70.69	50.70	86.57	2.29	7.07	81.44	38.36	48.21	0.03
ICP* [13]	5.79	7.93	17.59	-	-	-	5.79	7.93	17.59	-	-	-	0.25
FGR [67]	40.91	4.96	10.25	6.84	38.90	11.23	78.93	2.90	8.41	25.63	53.90	33.58	0.89
TEASER [92]	75.48	2.48	7.31	73.01	62.63	66.93	85.77	2.73	8.66	82.43	68.08	73.96	0.07
GC-RANSAC [52]	67.65	2.33	6.87	48.55	69.38	56.78	92.05	2.33	7.11	64.46	93.39	75.69	0.55
RANSAC-1M [8]	64.20	4.05	11.35	63.96	57.90	60.13	88.42	3.05	9.42	77.96	79.86	78.55	0.97
RANSAC-2M [8]	65.25	4.07	11.56	64.41	58.37	60.51	90.88	2.71	8.31	78.52	83.52	80.68	1.63
RANSAC-4M [8]	66.10	3.95	11.03	64.27	59.10	61.02	91.44	2.69	8.38	78.88	83.88	81.04	2.86
CG-SAC [10]	78.00	2.40	6.89	68.07	67.32	67.52	87.52	2.42	7.66	75.32	84.61	79.90	0.27
SC^2-PCR [21]	83.98	2.18	6.56	72.48	78.33	75.10	93.28	2.08	6.55	78.94	86.39	82.20	0.11
SC^2-PCR++	87.18	2.10	6.64	76.49	81.72	78.82	94.15	2.04	6.50	80.57	87.69	83.71	0.28

where R and t are ground-truth pose, while $\hat{R}$ and $\hat{t}$ are the estimated pose. The units of RE and TE are deg and cm respectively. Meanwhile, following [9], we also report the outlier removal results using the following three evaluation criteria: inlier precision (IP, %), inlier recall (IR, %) and F1-measure (F1, %). For the multi-way registration, following [20], we report the absolute trajectory error (ATE, cm) as the measurement.

Implementation Details: When computing the SC^2 matrix, the d_{thr} is set to twice as the voxel size for down-sampling (10 cm for indoor scenes and 60 cm for outdoor scenes). The number of seed (N_s in Section VI-A) is set to 0.2 * N, where N is the number of correspondences. When sampling the consensus set, we select 30 nearest neighbors ($K_1 = 30$) of the seed point at the first sampling stage, and remain 20 correspondences ($K_2 = 20$) to form the consensus set. When performing model selection, we use the Inlier Count (IC) to filter some wrong estimations, and remain 50 models to be further selected ($N'_s = 50$ in Section VI-B). All the experiments are conducted on a machine with an INTEL Xeon E5-2620 CPU and a single NVIDIA GTX1080Ti.

B. Evaluation on Indoor Scenes

We first report the results on 3DMatch dataset in Table I. We compare our method with 16 baselines: DCP [78], PointNetLK [77], OM-Net [83], RegTR [84], 3DRegNet [19], DGR [20], DHVR [11], PointDSC [9], SM [66], ICP [13], FGR [67], TEASER [92], GC-RANSAC [52], RANSAC [8], CG-SAC [10] and SC^2-PCR [21] (the previous version of our method in CVPR2022). The first 8 methods are based on deep learning, while the last 8 methods are geometric. For the deep learning methods, we use the provided pre-trained model of them for testing. DCP, PointNetLK, OM-Net, RegTR, and ICP are correspondence-free methods, so we do not report the correspondence-related metrics for them. For DCP, PointNetLK, OM-Net, and RegTR, they use neural networks for feature extraction, so we only compare them with learning-based descriptor for a fair comparison.

Combined With FPFH: We first use the FPFH descriptor to generate the correspondences, in which the mean inlier rate is 6.84%. As shown in Table I, the SC^2-PCR greatly outperforms all of the other methods. For the registration recall (RR), which is the most important criterion, SC^2-PCR improves it by about 6% over the closest competitors among the retested results (PointDSC and CG-SAC). Following [9], [20], since the part of failed registration can generate a large error of translation and rotation, we only compute the mean rotation (RE) and translation error (TE) of successfully registered point cloud pairs of each method to avoid unreliable metrics. This strategy of measurement makes methods with high registration recall more likely to have a large mean error, because they include more difficult data when calculating mean error. Nevertheless, SC^2-PCR still achieves competitive results on RE and TE. SC^2-PCR is slightly worse than PointDSC on TE and RE, and better than other methods. For the outlier rejection results, SC^2-PCR achieves the highest inlier recall (IR) and F1-measure. The F1 of SC^2-PCR outperforms the PointDSC by 5.35%.

Compared with the SC^2-PCR, SC^2-PCR++ further achieves a significant performance improvement. Since the proposed FS-TCD metric can better find the best hypothesis, SC^2-PCR++ can largely improve the registration recall (RR) from 83.98% to 87.18%. Meanwhile, the outlier rejection performance of SC^2-PCR++ is also better than SC^2-PCR, achieving 4.01%, 3.39%, and 3.72% improvement in terms of inlier precision (IP), inlier recall (IR) and F1-measure.

Combined With FCGF: To further verify the performance, we also adopt the recent FCGF descriptor to generate putative correspondences and report the registration results. The mean inlier rate of putative correspondences is 25.61%. As shown in Table I, since the inlier rate is higher than the correspondences obtained by FPFH descriptor, the performances of all of the feature-based methods are boosted. SC^2-PCR++ still achieves the best performance over all the methods, with 2.71%

TABLE II
QUANTITATIVE RESULTS ON 3DLOMATCH DATASET

	RR↑	RE↓	TE↓	IP↑	IR↑	F1↑	Time(s)
	FCGF						
DHVR [11]	54.41	4.14	12.56	41.96	38.60	39.22	3.55
DGR [20]	43.80	4.17	10.82	42.22	38.96	39.05	1.48
PointDSC [9]	56.09	3.87	10.39	44.51	52.38	47.57	0.10
FGR [67]	19.99	5.28	12.98	27.63	19.16	19.98	1.32
RANSAC [8]	46.38	5.00	13.11	40.70	44.61	42.02	2.86
CG-SAC [10]	52.31	3.84	10.55	42.16	47.02	44.61	0.25
SC²-PCR [21]	57.83	3.77	10.46	44.87	53.69	48.38	0.11
SC²-PCR++	61.15	3.72	10.56	47.12	56.52	50.85	0.26
	Predator						
	RR↑	RE↓	TE↓	IP↑	IR↑	F1↑	Time(s)
DHVR [11]	65.41	4.97	12.33	54.75	54.66	53.70	3.55
DGR [20]	59.46	3.19	10.01	51.38	54.24	51.62	1.48
PointDSC [9]	68.89	3.43	9.60	56.55	67.52	60.82	0.10
FGR [67]	35.99	4.77	11.64	47.18	38.76	39.10	1.32
RANSAC [8]	64.85	4.28	11.04	56.44	65.68	60.01	2.86
CG-SAC [10]	64.01	3.86	10.94	56.88	64.12	59.25	0.25
SC²-PCR [21]	69.46	3.46	9.58	56.98	67.47	61.08	0.11
SC²-PCR++	71.59	3.45	9.61	59.61	70.17	63.73	0.26
	GeoTransformer						
	RR↑	RE↓	TE↓	IP↑	IR↑	F1↑	Time(s)
DHVR [11]	73.83	4.49	10.21	61.06	71.85	64.21	2.71
PointDSC [9]	77.82	3.00	8.71	63.65	76.87	68.39	0.09
RANSAC [8]	77.48	3.37	9.69	64.91	73.98	68.68	2.03
CG-SAC [10]	76.92	3.34	9.81	62.10	75.27	67.05	0.22
LGR [44]	77.20	2.99	8.58	64.47	76.04	68.86	0.05
SC²-PCR [21]	78.33	3.04	8.81	64.63	76.67	69.19	0.08
SC²-PCR++	78.72	2.96	8.56	64.80	77.02	69.55	0.24

improvement over RANSAC on registration recall. Compared with SC²-PCR, SC²-PCR++ boosts the registration recall (RR) by 0.87%, and achieves 1.63%, 1.30% and 1.51% improvement in terms of inlier precision (IP), inlier recall (IR) and F1-measure.

Besides, the mean registration time for a pair of point clouds is also reported. Since SC²-PCR only needs to sample a few seed points with their consensus set rather than a large number of samples, it is competitive in terms of time-consuming. As shown in Table I, the mean registration time of SC²-PCR is 0.11 s. SC²-PCR++ adds some time overhead due to the more complex hypothesis selection strategy, requiring an average registration time of 0.28 s. Considering the performance improvement, the added time cost is well worth it. Nevertheless, SC²-PCR++ is still over 10× faster than RANSAC with 4 M iterations.

Robustness to Lower Overlap: Furthermore, we report the results on the low overlapped scenarios: 3DLoMatch [41]. Following PointDSC [9] and DHVR [11], we adopt the FCGF [30] and Predator [41] descriptors to generate correspondences. There are two versions of Predator. To avoid unnecessary misunderstanding, we specify that the version we used is the updated one. Similarly, the registration recall (RR), rotation error (RE), translation error (TE), inlier precision (IP), inlier recall (IR), and F1-measure (F1) are reported in Table II. As shown by the data, whether combined with FCGF or Predator descriptor, SC²-PCR++ achieves the highest registration recall. Compared with SC²-PCR, SC²-PCR++ improves the RR from 57.83% to 61.15% when combined with FCGF descriptor, and from 69.46% to 71.59% when combined with predator descriptor. The evaluation criteria related to the established correspondences, including IP, IR, and F1, have all been improved to a certain extent.

GeoTransformer [44], which uses a geometry-based transformer architecture to establish correspondences, achieves SOTA performance for the correspondence learning on 3DLoMatch dataset. In order to further validate the proposed method, we also use the GeoTransformer to generate correspondences and compare the registration results of our method with other methods. Since the Geotransformer produces correspondences in a coarse-to-fine method without descriptors, while DGR [20] and FGR [67] require descriptors as input, we do not report their results. LGR [44] is an outlier removal method proposed by GeoTransformer as the post-process. As shown in Table II, when combined with GeoTransformer, the proposed SC²-PCR++ still achieves the best performance.

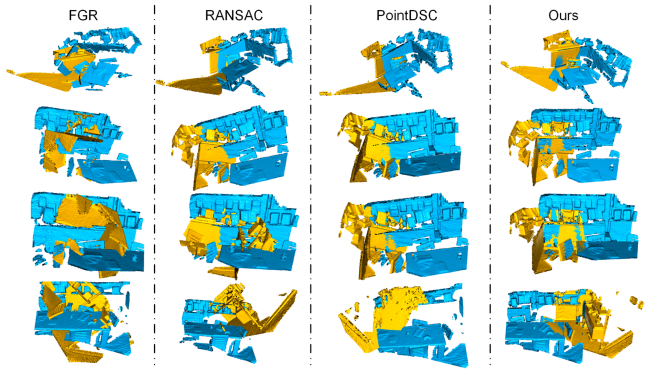


Fig. 7. Qualitative comparison on 3DMatch and 3DLoMatch dataset. From left to right are: FGR [67], RANSAC [8], PointDSC [9] and Ours.

Meanwhile, we also present some qualitative results on 3DLoMatch dataset. As shown in Fig. 7, we compare our method with RANSAC, FGR and PointDSC and report the visualized alignment results of some challenging scenes. Our method can successfully align two point clouds where the low overlap ratio is clearly visible.

C. Evaluation on Outdoor Scenes

In this experiment, we test on the outdoor KITTI [93] dataset. The results of DHVR [11], DGR [20], PointDSC [9], RANSAC [8], FGR [67], CG-SAC [10] are reported as comparison. DHVR, DGR, and PointDSC are deep learning based methods, while the remaining methods are non-learning. For DHVR, the authors have neither released the training code nor the pre-trained model on KITTI dataset, so we report the results provided by their paper. As shown in Table III, the SC²-PCR remarkably surpasses the non-learning methods, especially combined with the FPFH descriptor. The registration recall (RR) of our method is 25.23% higher than that of RANSAC when combined with the FPFH descriptor, and 0.54% higher when combined with the FCGF descriptor. The errors of translation and rotation are also lower than RANSAC. The SC²-PCR and SC²-PCR++ with FPFH descriptor obtain the results with the highest registration recall and lowest error of rotation and translation. This not only proves the flexibility of our method, but also proves the competitiveness of traditional hand-crafted descriptors in some scenarios. For the learning networks, our method can achieve close performance with them with high efficiency.

TABLE III
QUANTITATIVE RESULTS ON KITTI DATASET

	FPFH (traditional descriptor)						
	RR↑	RE↓	TE↓	IP↑	IR↑	F1↑	Time(s)
DHVR [11]	-	-	-	-	-	-	-
DGR [20]	77.12	1.64	33.10	78.39	54.12	62.15	2.29
PointDSC [9]	98.20	0.35	8.13	92.85	93.87	93.11	0.45
FGR [67]	5.23	0.86	43.84	4.93	0.05	0.10	3.88
RANSAC [8]	74.41	1.55	30.20	78.50	52.66	60.72	5.43
CG-SAC [10]	74.23	0.73	14.02	78.64	60.82	67.11	0.73
SC^2-PCR [21]	99.64	0.32	7.23	93.63	95.89	94.63	0.31
SC^2-PCR++	99.64	0.32	7.19	94.07	96.19	95.00	0.86
	FCGF (learning based descriptor)						
	RR↑	RE↓	TE↓	IP↑	IR↑	F1↑	Time(s)
DHVR [11]	99.10	0.29	19.80	-	-	-	0.83
DGR [20]	98.20	0.34	21.70	72.19	78.06	75.13	2.29
PointDSC [9]	98.02	0.33	21.03	82.00	90.84	85.83	0.45
FGR [67]	89.54	0.46	25.72	95.13	4.25	8.18	3.88
RANSAC [8]	98.02	0.39	23.17	81.89	90.36	85.52	5.43
CG-SAC [10]	97.84	0.37	22.91	81.85	90.84	85.74	0.73
SC^2-PCR [21]	98.20	0.33	20.95	82.01	91.03	85.90	0.31
SC^2-PCR++	98.56	0.32	20.61	82.17	91.23	86.09	0.86

TABLE IV
ABSOLUTE TRAJECTORY ERROR (ATE, CM) ON THE 4 SCENES OF AUGMENTED ICL-NUIM DATASET WITH SIMULATED DEPTH NOISES. THE AVERAGE ATE OVER ALL THE SCENES IS REPORTED IN THE LAST COLUMN. (LOWER IS BETTER.)

	Living1	Living2	Office1	Office2	AVG
ElasticFusion	66.61	24.33	13.04	35.02	34.75
InfiniTAM	46.07	73.64	113.8	105.2	85.68
BAD-SLAM	fail	40.41	18.53	26.34	-
Multiway + DGR	21.06	21.88	15.76	11.56	17.57
Multiway + PointDSC	20.25	15.58	13.56	11.30	15.18
Multiway + DHVR	22.91	16.37	12.58	10.90	15.69
Multiway+ FGR	78.97	24.91	14.96	21.05	34.98
Multiway + RANSAC	110.9	19.33	14.42	17.31	40.49
Multiway + SC^2-PCR	18.68	14.31	14.63	11.95	14.90
Multiway + SC^2-PCR++	17.56	14.37	13.24	9.49	13.67

D. Multi-Way Registration

In order to further validate the performance of the proposed method, we integrate it into a multi-way registration pipeline and test it on the ICL_NUIM [94], [95] dataset. Following [9], [20], we first extract the FPFH descriptor for each frame, and then use the proposed method to initialize the pose of each frame by pair-wise registration. After that, the poses are globally optimized by the graph optimization method (the g2o method [97] implemented in Open3d [98] library is utilized). We also combine other registration methods with the multi-way pipeline, including DGR, PointDSC, DHVR, FGR and RANSAC. Meanwhile, the results of the state-of-the-art online SLAM methods, including ElasticFusion [99], InfiniTAM [100] and BAD-SLAM [101], are also reported as a comparison. As shown in Table IV, we present the Absolute trajectory error (ATE) of each scene and the average result. Since BAD-SLAM can not lead to a successful result, we do not put the average result of it. As we can see, the non-learning methods FGR and RANSAC lead to worse results compared with deep learning based methods. The proposed SC^2-PCR and SC^2-PCR++ are also non-learning methods, but achieve great performance among all the methods. The SC^2-PCR++ achieves the best performance on the Living1 and Office2 scenes and the lowest average ATE over the four test scenes.

TABLE V
GENERALIZATION RESULTS. THE REGISTRATION RECALL (%) ON 3DMATCH, 3DLOMATCH AND KITTI DATASETS ARE REPORTED

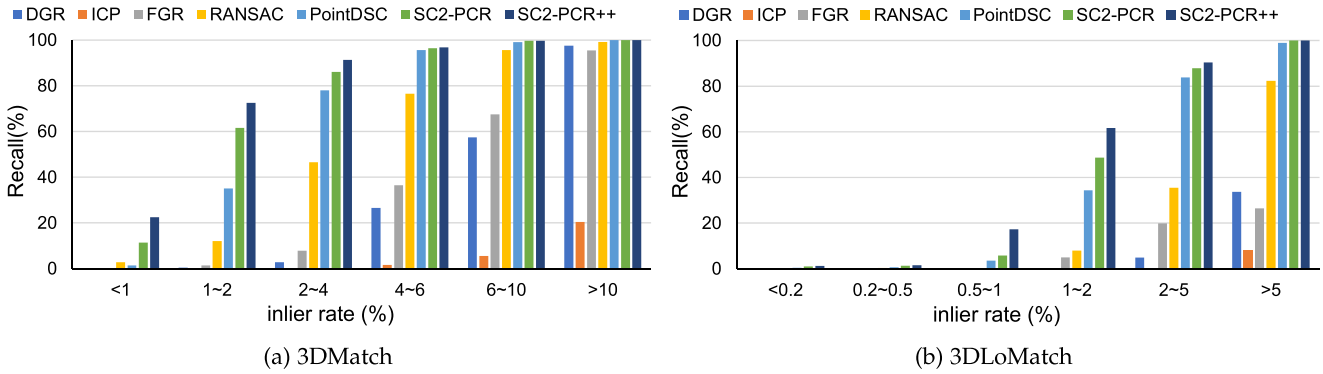
	3DMatch		3DLoMatch		KITTI	
	FPFH	FCGF	FCGF	Predator	FPFH	FCGF
DGR	49.48	81.89	23.75	45.03	73.69	86.12
PointDSC	68.12	87.74	40.65	53.79	90.27	92.97
SC^2-PCR	83.98	93.28	57.83	69.46	99.64	98.20
SC^2-PCR++	87.18	94.15	61.15	71.59	99.64	98.56

E. Generalization and Robustness

Generalization Experiments: As reported above, deep learning based methods also achieve competitive performance on the 3DMatch, 3DLoMatch and KITTI datasets. Compared with these methods based on deep learning, the other advantage of our method is that it has no bias cross different datasets, while deep learning based methods have performance degradation when generalized between different datasets. To demonstrate this, we perform the generalization experiments on both 3DMatch, 3DLoMatch and KITTI datasets. For the recent learning based methods, including DGR and PointDSC, we report the cross-dataset results. Specifically, we adopt their pre-trained model by KITTI to test on 3DMatch and 3DLoMatch and use 3DMatch's model to test on KITTI. As shown in Table V, both the previous version SC^2-PCR and the updated version SC^2-PCR++ show significant improvements in registration recall without the generalization problem. This further demonstrates the effectiveness of our method.

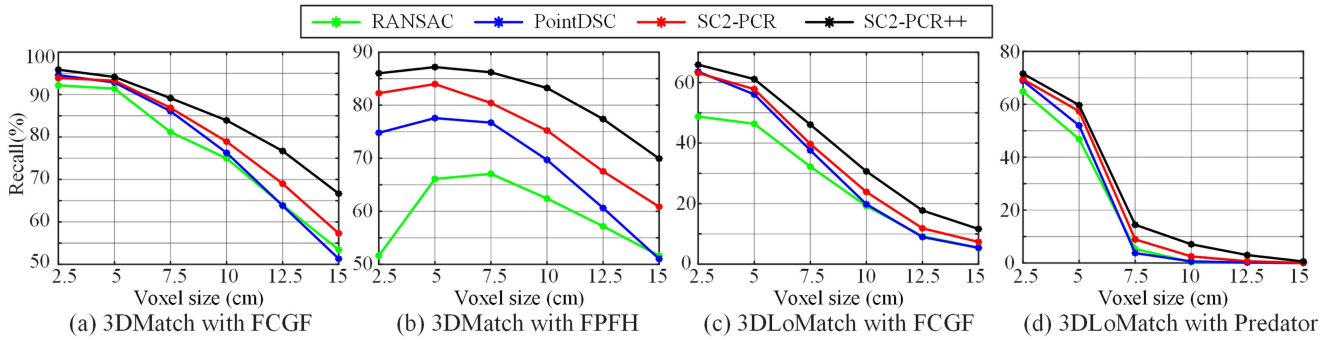
Robustness to Noises: An important factor to measure the model fitting method is the stability under low inlier rate. In order to further verify the performance of our method, we report the results under different inlier ratios in Fig. 8. Specifically, we first use FPFH to generate initial match pairs for the 3DMatch dataset. Then, according to the inlier ratio, all the point cloud pairs are divided into 6 groups: $< 1\%$, 1% - 2%, 2% - 4%, 4% - 6%, 6% - 10% and $> 10\%$. The number of point cloud pairs in each group is 141, 208, 346, 252, 323, and 353. As shown in Fig. 8(a), when the inlier rate is less than 2%, SC^2-PCR++ is significantly better than other baselines. Compared with SC^2-PCR, SC^2-PCR++ has a more robust performance in low inlier-rate scenes. Furthermore, we generate more challenging test pairs with lower inlier rates on 3DLoMatch dataset by FPFH descriptor, as shown in Fig. 8(b). The mean inlier rate of the putative correspondences on this dataset is 1.6%. Similarly, we also divided all the pairs into 6 groups according to the inlier rate: $< 0.2\%$, 0.2% - 0.5%, 0.5% - 1%, 1% - 2%, 2% - 5% and $> 5\%$, with 214, 322, 370, 410, 365 and 100 samples in each group. As we can see, when the inlier rate is lower than 0.5%, all methods fail to get the correct estimation. In this case, the inlier rate is too low, making the model fitting an ill problem. When the inlier rate is 0.5% - 1%, the success rate of most methods is extremely low, while SC^2-PCR++ significantly improves the success rate. The above experimental results demonstrate the robustness anti noises of our method.

Effect of Voxel Size: As mentioned in Section III-A, before extracting feature descriptors, we first use the voxel grid to downsample the two point clouds. Here we present the effect



(a) 3DMatch (b) 3DLoMatch

Fig. 8. The registration recall under the different inlier ratio of the putative correspondences.



(a) 3DMatch with FCGF (b) 3DMatch with FPFH (c) 3DLoMatch with FCGF (d) 3DLoMatch with Predator

Fig. 9. The registration recall with different voxel sizes for downsampling the point cloud.

of voxel size on registration results in Fig. 9. As we can see, as the voxel size increases, the number of points in the downsampled point clouds will decrease, and the registration results will also decrease, especially on 3DLoMatch dataset. The reason is that the two point clouds in 3DLoMatch dataset share a small overlapping area. When the voxel size increases, the overlapping points will be too insufficient, making it hard to establish correspondences between point clouds. In general, SC^2-PCR++ achieves the highest registration recall in different voxel sizes.

Efficiency With Different N_s': As described in Section VI-B, in SC^2-PCR++, we first use the IC metric to select N_s' candidate models, and then use the FS-TCD to select the best one. When N_s' is 1, SC^2-PCR++ will become SC^2-PCR. As N_s' increases, we can retain more possible models to be finely selected, but it will also add computation time. It not only means that SC^2-PCR++ is a more generalized version of SC^2-PCR, but also means that we can balance the efficiency and recall by the N_s' parameter. To better understand the SC^2-PCR++, we report the running time and registration recall with different N_s' in Table VI.

Robustness to Parameters: In order to decide the parameters for the method, we conduct a series of experiments in this part. The most important parameter for our method is the number of correspondences when sampling the consensus set. As described in Section VI-A, the SC^2-PCR adopts a two-stage sampling strategy. It finds K_1 instances for each seed in the first stage, and remains $K_2 (K_2 < K_1)$ samples for model estimation in the second stage. As shown in Fig. 10, we set (K_1, K_2) to be (10, 5), (20, 10), (30, 20), (40, 30), (50, 40), (60, 50), (70, 60) and (80, 70) respectively. The registration recall (RR) on 3DMatch and 3DLoMatch datasets with different descriptors are represented. As we can see from the curves, in general, RR increases first and then decreases as the number of correspondences in the consensus set increases. In fact, it only takes three correspondences to estimate the correct rigid transformation, but more correspondences may prevent the samples from clustering together, which can reduce the noise on the estimated transformation. On the whole, the results with (K_1, K_2) being (30, 20) are the best, so this combination of parameters is selected in the final version of our method. It is worth mentioning that changing parameters

TABLE VI
THE RUNNING TIME AND REGISTRATION RECALL (RR) WITH DIFFERENT N_s'

		3DMatch						
	N_s'	1	5	10	20	30	40	50
FPFH	Time (s)	0.11	0.14	0.16	0.19	0.22	0.25	0.28
	RR (%)	83.98	85.58	86.44	86.88	87.55	87.31	87.18
FCGF	Time (s)	0.11	0.14	0.16	0.19	0.22	0.25	0.28
	RR (%)	93.28	93.36	93.59	93.96	94.02	94.09	94.15
		3DLoMatch						
	N_s'	1	5	10	20	30	40	50
FCGF	Time (s)	0.11	0.13	0.15	0.18	0.20	0.23	0.26
	RR (%)	57.83	59.46	59.63	59.91	60.75	60.98	61.15
Predator	Time (s)	0.11	0.13	0.15	0.18	0.20	0.23	0.26
	RR (%)	69.46	70.19	70.30	70.69	71.25	71.36	71.59

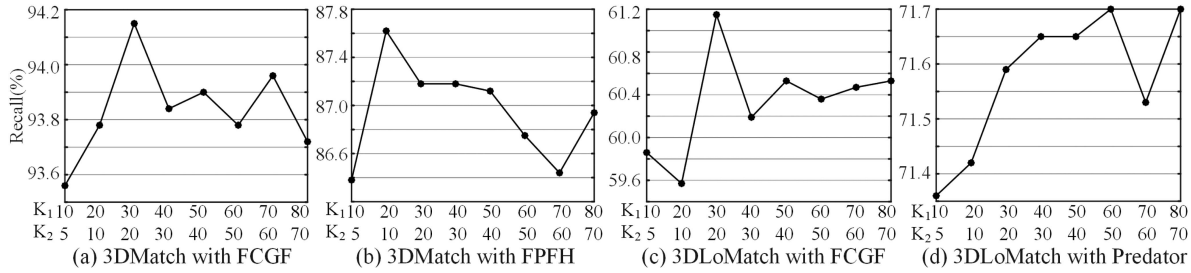


Fig. 10. Experiments results for the analysis of the sampling parameters.

TABLE VII
THE REGISTRATION RECALL OF COMBINING SC^2 MEASURE AND FS-TCD METRIC WITH LEARNING BASED NETWORK

	3DMatch		3DLoMatch		KITTI	
	FPFH	FCGF	FCGF	Predator	FPFH	FCGF
	Known					
PointDSC	77.57	92.85	56.09	68.89	98.20	98.02
+SC^2	83.24	93.10	57.05	69.04	99.10	98.02
+FS-TCD	81.52	93.60	57.72	70.30	98.38	98.20
+SC^2+FS-TCD	86.20	93.78	60.13	70.63	99.46	98.92
	Generalization					
PointDSC	68.12	87.74	40.65	53.79	90.27	92.97
+SC^2	74.12	89.59	44.81	56.71	97.48	98.02
+FS-TCD	73.14	89.16	45.37	57.68	95.14	97.84
+SC^2+FS-TCD	78.68	90.70	45.82	58.06	98.56	98.20

does not have a great impact on registration recall, which can also demonstrate the robustness of our method.

F. Combined With Learning Network

To verify the flexibility of our proposed approach, we combine our approach with a recent deep learning approach PointDSC [9]. It adopts the spatial consistency matrix to guide the non-local module. First, since the proposed SC^2 measure is more robust to the ambiguity, we replace the spatial consistency matrix in PointDSC with SC^2. Instead of retraining the network, we directly plug our metrics into it. The registration recall of their vanilla version and combined version are shown in Table VII. It can be seen that adding our metrics can significantly boost the performance of the network, especially for the generalization performance of the network. Next, we also integrate the FS-TCD metric into the PointDSC network. The original version of PointDSC generates some hypotheses and uses the Inlier Count (IC) metric to select the best estimation. We replace it with the proposed FS-TCD based selection strategy. It can be seen that the performance of all settings is improved. Finally, the SC^2 measure and FS-TCD metric are both added into the PointDSC, which achieves better results than only using one of them. The above results demonstrate that the proposed measure is flexible to combine with other methods.

G. Ablation Study

In this section, we perform ablation studies on 3DMatch dataset. We use the FPFH and FCGF descriptors to form correspondences respectively. The classic RANSAC is adopted as our baseline, as shown in Row 1 and Row 10 of Table VIII. We progressively add the proposed modules to the baseline and report the results.

Second-Order Spatial Compatibility: We first add the Second-Order Spatial Compatibility (SC^2) measure as the guidance for the sampling of RANSAC. Each correspondence is extended into a consensus set by searching the k-nearest neighbors in metric space. The Spatial Compatibility (SC) adopted by previous works [9], [10], [12] is also utilized as the sampling guidance, and the results are reported as a comparison. As shown in Row 1, 3, and 10, 12 of Table VIII, the registration recall obtained by using SC^2 measure as guidance is 14.79% higher than RANSAC when combined with FPFH, and 1.66% higher when combined with FCGF. Meanwhile, since SC^2 measure can narrow the sampling space, the mean registration time of SC^2 measure is much smaller than RANSAC. Besides, using SC^2 measure as guidance can achieve better performance than using SC measure by comparing Row 2, 3 and 11, 12. This is because SC is disturbed by the ambiguity problem, while SC^2 measure can eliminate the ambiguity.

Two-Stage Selection: We further adopt a two-stage selection strategy for generating the consensus set for each seed. When one seed is an inlier correspondence, it has almost removed most of the outliers in the consensus set formed in the first stage. Since SC^2 becomes more stable when the inlier rate increases, we construct a local SC^2 matrix to remove potential outliers. Comparing Row 3, 4, and 12, 13 in Table VIII, using two-stage selection achieves a recall improvement of 1.96% when combined with FPFH, and 0.12% improvement when combined with FCGF.

Local Spectral Matching: When a minimum set is sampled, RANSAC adopts the instance-equal SVD to generate an estimation of translation and rotation, which is sensitive to errors. We replace the instance-equal SVD [91] with the weighted SVD [19], [20], so that less reliable correspondences are assigned lower weights for robust registration. We construct a soft SC^2 matrix in each consensus set, and then use local spectral matching to compute the association between each correspondence with the main cluster. The association value is utilized as the weight for weighted SVD. Comparing Row 4, 5, and 13, 14 in Table VIII, using local spectral matching can boost the performance, especially for the mean rotation and translation error.

Seed Selection: So far, each correspondence is treated as a seed. However, it does not need to generate a consensus set for all correspondences and estimate a rigid transformation. We only need to select a few reliable points, and use the aggregation

TABLE VIII
ABLATION STUDY ON 3DMATCH DATASET. **SC**: SPATIAL COMPATIBILITY MEASURE. **SC^2**: SECOND-ORDER SPATIAL COMPATIBILITY MEASURE. **TS**: TWO-STAGE SELECTION FOR CONSENSUS SET SAMPLING. **LSM** LOCAL SPECTRAL MATCHING. **SEED**: USING SEED POINTS TO REDUCE THE NUMBER OF SAMPLING. **TCD**: TRUNCATED CHAMFER DISTANCE. **F-TCD**: FEATURE-CONSTRAINED TRUNCATED CHAMFER DISTANCE. **FS-TCD**: FEATURE AND SPATIAL CONSISTENCY CONSTRAINED TRUNCATED CHAMFER DISTANCE

		SC	SC^2	TS	LSM	Seed	TCD	F-TCD	FS-TCD	RR(%) ↑	RE(deg) ↓	TE(cm) ↓	IP(%)↑	IR(%)↑	F1(%)	Time(s)
FPFH	1)									66.10	3.95	11.03	64.27	59.10	61.02	2.86
	2)	✓								71.56	2.07	6.48	68.22	70.11	68.73	0.27
	3)		✓							80.89	2.34	6.92	71.56	77.14	73.27	0.31
	4)		✓	✓						82.85	2.32	6.69	72.68	78.01	74.99	0.33
	5)		✓	✓	✓					84.10	2.13	6.56	73.11	79.10	75.89	0.37
	6)		✓	✓	✓	✓				83.98	2.18	6.56	72.48	78.33	75.10	0.11
	7)		✓	✓	✓	✓	✓			73.76	1.70	5.93	86.95	22.90	35.14	1.57
	8)		✓	✓	✓	✓		✓		86.88	2.14	6.69	75.84	80.99	78.14	0.28
	9)		✓	✓	✓	✓			✓	87.18	2.10	6.64	76.49	81.72	78.82	0.28
FCGF	10)									91.44	2.69	8.38	78.88	83.88	81.04	2.86
	11)	✓								87.52	2.42	7.66	76.19	83.21	80.05	0.27
	12)		✓							93.10	2.16	6.76	77.81	85.53	81.21	0.31
	13)		✓	✓						93.22	2.10	6.88	78.80	86.47	82.16	0.33
	14)		✓	✓	✓					93.28	2.08	6.56	79.10	86.89	82.41	0.37
	15)		✓	✓	✓	✓				93.28	2.08	6.55	78.94	86.39	82.20	0.11
	16)		✓	✓	✓	✓	✓			86.96	1.78	6.33	95.11	45.51	60.03	1.57
	17)		✓	✓	✓	✓		✓		93.72	2.02	6.48	80.24	87.34	83.37	0.28
	18)		✓	✓	✓	✓			✓	94.15	2.04	6.50	80.57	87.69	83.71	0.28

among the inliers to collect the set without outliers, so as to further improve the efficiency of registration. We use the global spectral matching combined with Non-Maximum Suppression to find several correspondences as seeds instead of all of the set. Row 5, 6, and 14, 15 of Table VIII shows that Seed Selection can reduce registration time by more than half without much performance degradation.

Hypothesis Selection Strategies: Finally, we replace the Inlier Count (IC) with the proposed selection metric. As mentioned in Section VI-B, in order to accelerate the selection process, we use IC to filter some spurious models, and then use other metrics for finer chosen. First, the original Chamer distance (CD) is first utilized as a comparison. Since CD is prone to low-overlap and fails in most scenes, we use the Truncated form of Chamer distance, i.e. TCD as reformulated in (31). As shown in Row 7 and 16 of Table VIII, using TCD as re-selection metric leads to worse results than directly using IC for selection. The reason is that TCD finds the nearest neighbor for each source point in the whole target points, which considers too many wrong alignments. Next, we add the constraint into TCD and adopt the feature-constrained truncated Chamfer distance (F-TCD) to re-selection. Comparing Row 6, 8 and 15, 17, we can find that F-TCD achieves 2.90% and 0.44% improvement of RR when combined with FPFH and FCGF descriptors respectively. Meanwhile, the IP, IR, and F1 criteria are also improved, which means TCD also results in better correspondences. Finally, the spatial consistency constraint is also appended into the TCD as the proposed feature and spatial consistency constrained truncated Chamfer distance (FS-TCD). As shown in Row 9 and 18 of Table VIII, the registration performance can be further boosted.

VIII. CONCLUSION

In this article, we present a second-order spatial compatibility (SC^2) measure based point cloud registration method, called SC^2-PCR++. The core component of our method is to cluster inliers by the proposed SC^2 measure at an early stage while eliminating ambiguity. Specifically, some reliable correspondences are selected by a global spectral decomposition with Non-Maximum Suppression firstly, called seed points. Then a two-stage sampling strategy is adopted to extend the seed points into some consensus sets. After that, each consensus set produces a rigid transformation by local spectral matching. Finally, the best estimation is selected by the proposed Feature and Spatial consistency constrained Truncated Chamfer Distance (FS-TCD) metric as the final result. Extensive experiments demonstrate that our method achieves state-of-the-art performance and high efficiency. Meanwhile, we also demonstrate the proposed SC^2 and FS-TCD are flexible measures, which can be combined with learning networks to further boost their performance.

REFERENCES

[1] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robot. Automat. Mag.*, vol. 13, no. 3, pp. 108–117, Sep. 2006.

[2] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Automat. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[3] M. Montemerlo et al., "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *Proc. Conf. Assoc. Advance. Artif. Intell.*, 2002, pp. 593–598.

[4] K. Sun and W. Tao, "A center-driven image set partition algorithm for efficient structure from motion," *Inf. Sci.*, vol. 479, pp. 101–115, 2019.

[5] R. T. Azuma, "A survey of augmented reality," *Presence Teleoperators Virtual Environ.*, vol. 6, no. 4, pp. 355–385, 1997.

[6] M. Billinghurst, A. Clark, and G. Lee, "A survey of augmented reality," *Foundations Trends Human–Comput. Interact.*, vol. 8, no. 2-3, pp. 73–272, 2015.

[7] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robot. Auton. Syst.*, vol. 66, pp. 86–103, 2015.

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

[9] X. Bai et al., "PointDSC: Robust point cloud registration using deep spatial consistency," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15 859–15 869.

[10] S. Quan and J. Yang, "Compatibility-guided sampling consensus for 3-D point cloud registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7380–7392, Oct. 2020.

[11] J. Lee, S. Kim, M. Cho, and J. Park, "Deep hough voting for robust global registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15 994–16 003.

[12] J. Yang, Z. Huang, S. Quan, Z. Qi, and Y. Zhang, "SAC-COT: Sample consensus by sampling compatibility triangles in graphs for 3-D point cloud registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

[13] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures*, vol. 1611. Bellingham, WA, USA: SPIE, 1992, pp. 586–606.

[14] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "A papier-mâché approach to learning 3D surface generation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 216–224.

[15] Y. Yang, C. Feng, Y. Shen, and D. Tian, "FoldingNet: Point cloud auto-encoder via deep grid deformation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 206–215.

[16] X. Huang, G. Mei, and J. Zhang, "Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11 366–11 374.

[17] L. Cavalli, V. Larsson, M. R. Oswald, T. Sattler, and M. Pollefeys, "Handcrafted outlier detection revisited," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 770–787.

[18] K. Sun, W. Tao, and Y. Qian, "Guide to match: Multi-layer feature matching with a hybrid gaussian mixture model," *IEEE Trans. Multimedia*, vol. 22, no. 9, pp. 2246–2261, Sep. 2020.

[19] G. D. Pais, S. Ramalingam, V. M. Govindu, J. C. Nascimento, R. Chellappa, and P. Miraldo, "3DRegNet: A deep neural network for 3D point registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7193–7203.

[20] C. Choy, W. Dong, and V. Koltun, "Deep global registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2514–2523.

[21] Z. Chen, K. Sun, F. Yang, and W. Tao, "SC2-PCR: A second order spatial compatibility for efficient and robust point cloud registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 13 221–13 231.

[22] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

[23] A. E. Johnson and M. Hebert, "Surface matching for object recognition in complex three-dimensional scenes," *Image Vis. Comput.*, vol. 16, no. 9/10, pp. 635–651, 1998.

[24] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2004, pp. 224–237.

[25] F. Tombari, S. Salti, and L. Di Stefano, "Unique shape context for 3D data description," in *Proc. ACM Workshop 3D Object Retrieval*, 2010, pp. 57–62.

[26] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Aligning point cloud views using persistent feature histograms," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 3384–3391.

[27] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3212–3217.

[28] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, "A comprehensive performance evaluation of 3D local feature descriptors," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 66–89, 2016.

[29] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3DMatch: Learning local geometric descriptors from rgb-d reconstructions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1802–1811.

[30] C. Choy, J. Park, and V. Koltun, "Fully convolutional geometric features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 8958–8966.

[31] X. Bai, Z. Luo, L. Zhou, H. Fu, L. Quan, and C.-L. Tai, "D3Feat: Joint learning of dense detection and description of 3D local features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6359–6367.

[32] Y. Wang, C. Yan, Y. Feng, S. Du, Q. Dai, and Y. Gao, "STORM: Structure-based overlap matching for partial point cloud registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1135–1149, Jan. 2023.

[33] H. Deng, T. Birdal, and S. Ilic, "PPFNet: Global context aware local features for robust 3D point matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 195–205.

[34] H. Deng, T. Birdal, and S. Ilic, "PPF-FoldNet: Unsupervised learning of rotation invariant 3D local descriptors," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 602–618.

[35] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652–660.

[36] Z. Gojcic, C. Zhou, J. D. Wegner, and A. Wieser, "The perfect match: 3D point cloud matching with smoothed densities," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5545–5554.

[37] C. Choy, J. Gwak, and S. Savarese, "4D spatio-temporal convnets: Minkowski convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3075–3084.

[38] Z. J. Yew and G. H. Lee, "3DFeat-Net: Weakly supervised local 3D features for point cloud registration," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 607–623.

[39] A. Tonioni, S. Salti, F. Tombari, R. Spezialetti, and L. D. Stefano, "Learning to detect good 3D keypoints," *Int. J. Comput. Vis.*, vol. 126, no. 1, pp. 1–20, 2018.

[40] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[41] S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, and K. Schindler, "PREDATOR: Registration of 3D point clouds with low overlap," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 4267–4276.

[42] H. Yu, F. Li, M. Saleh, B. Busam, and S. Ilic, "CoFiNet: Reliable coarse-to-fine correspondences for robust pointcloud registration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 23872–23884.

[43] Y. Li and T. Harada, "Lepard: Learning partial point cloud matching in rigid and deformable scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5554–5564.

[44] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, and K. Xu, "Geometric transformer for fast and robust point cloud registration," 2022, *arXiv:2202.06688*.

[45] P. H. Torr, S. J. Nasuto, and J. M. Bishop, "NAPSAC: High noise, high dimensional robust estimation-it's in the bag," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 44.1–44.10.

[46] K. Ni, H. Jin, and F. Dellaert, "GroupSAC: Efficient consensus in the presence of groupings," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 2193–2200.

[47] O. Chum and J. Matas, "Matching with prosac-progressive sample consensus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 220–226.

[48] P. H. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[49] V. Fragoso, P. Sen, S. Rodriguez, and M. Turk, "EVSAC: Accelerating hypotheses generation by modeling matching scores with extreme value theory," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2472–2479.

[50] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Proc. Joint Pattern Recognit. Symp.*, Springer, 2003, pp. 236–243.

[51] K. Lebeda, J. Matas, and O. Chum, "Fixing the locally optimized RANSAC–full experimental evaluation," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 1–11.

[52] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6733–6741.

[53] D. Barath and J. Matas, "Graph-cut RANSAC: Local optimization on spatially coherent structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4961–4974, Sep. 2022.

[54] D. Barath, J. Matas, and J. Noskova, "MAGSAC: Marginalizing sample consensus," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 197–10 205.

[55] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, 2021.

[56] K. Moo Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2666–2674.

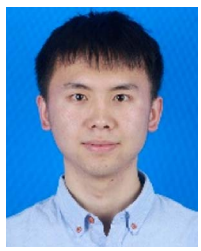
[57] E. Brachmann and C. Rother, "Neural-guided RANSAC: Learning where to sample model hypotheses," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4322–4331.

[58] C. Zhao, Z. Cao, C. Li, X. Li, and J. Yang, "Nm-net: Mining reliable neighbors for robust feature correspondences," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 215–224.

[59] Y. Liu, L. Liu, C. Lin, Z. Dong, and W. Wang, "Learnable motion coherence for correspondence pruning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3237–3246.

[60] J. Zhang et al., "Learning two-view correspondences and geometry using order-aware network," 2019, *arXiv: 1908.04964*.

[61] H. Chen et al., "Learning to match features with seeded graph matching network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 6301–6310.
[62] C. Zhao, Y. Ge, F. Zhu, R. Zhao, H. Li, and M. Salzmann, "Progressive correspondence pruning by consensus learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 6464–6473.
[63] Z. Chen, F. Yang, and W. Tao, "Cascade network with guided loss and hybrid attention for finding good correspondences," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 1123–1131.
[64] Z. Chen, F. Yang, and W. Tao, "DetarNet: Decoupling translation and rotation by siamese network for point cloud registration," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 401–409.
[65] G. Mei, X. Huang, L. Yu, J. Zhang, and M. Bennamoun, "COTReg: Coupled optimal transport based point cloud registration," 2021, *arXiv:2112.14381*.
[66] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. IEEE 10th Int. Conf. Comput. Vis.*, vol. 2, 2005, pp. 1482–1489.
[67] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 766–782.
[68] A. Glent Buch, Y. Yang, N. Kruger, and H. Gordon Petersen, "In search of inliers: 3D correspondence by local and global voting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2067–2074.
[69] H. Sahloul, S. Shirafuji, and J. Ota, "An accurate and efficient voting scheme for a maximally all-inlier 3D correspondence set," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2287–2298, Jul. 2021.
[70] L. Sun and L. Deng, "TriVoC: Efficient voting-based consensus maximization for robust point cloud registration with extreme outlier ratios," *IEEE Trans. Robot. Autom.*, vol. 7, no. 2, pp. 4654–4661, Apr. 2022.
[71] W. Tang and D. Zou, "Multi-instance point cloud registration by efficient correspondence clustering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 6667–6676.
[72] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image Vis. Comput.*, vol. 10, no. 3, pp. 145–155, 1992.
[73] A. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Proc. Conf. Robot. Sci. Syst.*, 2009, Art. no. 435.
[74] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016.
[75] D. Campbell and L. Petersson, "GOGMA: Globally-optimal Gaussian mixture alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5685–5694.
[76] D. Campbell, L. Petersson, L. Kneip, H. Li, and S. Gould, "The alignment of the spheres: Globally-optimal spherical mixture alignment for camera pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11 796–11 806.
[77] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "PointNetLK: Robust & efficient point cloud registration using PointNet," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7163–7172.
[78] Y. Wang and J. M. Solomon, "Deep closest point: Learning representations for point cloud registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3523–3532.
[79] Z. J. Yew and G. H. Lee, "RPM-Net: Robust point matching using learned features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11 824–11 833.
[80] K. Fu, S. Liu, X. Luo, and M. Wang, "Robust point cloud registration framework based on deep graph matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8893–8902.
[81] W. Yuan, B. Eckart, K. Kim, V. Jampani, D. Fox, and J. Kautz, "DeepGMR: Learning latent gaussian mixture models for registration," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, UK, Springer, Aug. 23–28, 2020, pp. 733–750.
[82] Y. Wang and J. M. Solomon, "PRNet: Self-supervised learning for partial-to-partial registration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8814–8826.
[83] H. Xu, S. Liu, G. Wang, G. Liu, and B. Zeng, "OMNet: Learning overlapping mask for partial-to-partial point cloud registration," 2021, *arXiv:2103.00937*.
[84] Z. J. Yew and G. H. Lee, "REGTR: End-to-end point cloud correspondences with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 6677–6686.
[85] J. O. Irwin, "The frequency distribution of the difference between two independent variates following the same poisson distribution," *J. Roy. Statist. Soc.*, vol. 100, no. 3, pp. 415–416, 1937.
[86] D. Karlis and I. Ntzoufras, "Analysis of sports data by using bivariate poisson models," *J. Roy. Statist. Soc. Ser. D*, vol. 52, no. 3, pp. 381–393, 2003.
[87] D. Karlis and I. Ntzoufras, "Bayesian analysis of the differences of count data," *Statist. Med.*, vol. 25, no. 11, pp. 1885–1905, 2006.
[88] G. E. Box et al., *Statistics for Experimenters*, vol. 664. Hoboken, NJ, USA: Wiley, 1978.
[89] P. Virtanen et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python" *Nat. Methods*, vol. 17, pp. 261–272, 2020, doi: 10.1038/s41592-019-0686-2.
[90] R. Mises and H. Pollaczek-Geiringer, "Praktische verfahren der gleichungsauflösung," *ZAMM-J. Appl. Math. Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 9, no. 1, pp. 58–77, 1929.
[91] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.
[92] H. Yang, J. Shi, and L. Carlone, "TEASER: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, Apr. 2021.
[93] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The kitti vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.
[94] S. Choi, Q.-Y. Zhou, and V. Koltun, "Robust reconstruction of indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5556–5565.
[95] A. Handa, T. Whelan, J. McDonald, and A. J. Davison, "A benchmark for RGB-D visual odometry, 3D reconstruction and slam," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2014, pp. 1524–1531.
[96] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*, vol. 26. Berlin, Germany: Springer, 2012.
[97] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G 2 O: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2011, pp. 3607–3613.
[98] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," 2018, *arXiv: 1801.09847*.
[99] T. Whelan, R. F. Salas-Moreno, B. Glocker, A. J. Davison, and S. Leutenegger, "ElasticFusion: Real-time dense SLAM and light source estimation," *Int. J. Robot. Res.*, vol. 35, no. 14, pp. 1697–1716, 2016.
[100] O. Kähler, V. A. Prisacariu, and D. W. Murray, "Real-time large-scale dense 3D reconstruction with loop closure," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 500–516.
[101] T. Schops, T. Sattler, and M. Pollefeys, "BAD SLAM: Bundle adjusted direct RGB-D SLAM," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 134–144.



Zhi Chen received the BS degree from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently working toward the PhD degree with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, with Prof. W. Tao. His research interests include image matching, point cloud registration and deep learning with geometry.

Kun Sun (Member, IEEE) received the PhD degree from the National Key Laboratory of Science and Technology on Multi-Spectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology in 2017. Since 2017, he has been an associate professor with the School of Computer Science, China University of Geoscience. His research interests include multi-view image matching, 3D reconstruction and point cloud processing.

Fan Yang received the BS degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2019. He is currently working with professor Wenbing Tao for the PhD degree with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology (HUST), Wuhan, China. His research interests include point cloud registration and 3D geometry.

Lin Guo received the B.S. degree from Zhengzhou University (ZZU), ZhengZhou, China, in 2021. He is currently working with Professor Wenbing Tao for the master's degree with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology (HUST), Wuhan, China. His research interests include point cloud registration and image matching.

Wenbing Tao (Member, IEEE) received the PhD degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2004. He was a research fellow with the Division of Mathematical Sciences, Nanyang Technology University, Singapore, from 2008 to 2009. He is currently a full professor with the School of Artificial Intelligence and Automation, HUST. He has authored numerous papers and conference papers in the area of computer vision and 3D reconstruction. His current research interests in 3D vision, including point cloud registration, multi-view stereo and surface reconstruction.