SC²-PCR: A Second Order Spatial Compatibility for Efficient and Robust Point Cloud Registration

Zhi Chen[1]　　　Kun Sun[2]　　　Fan Yang[1]　　　Wenbing Tao[1*]

[1]National Key Laboratory of Science and Technology on Multi-spectral Information Processing,
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China
[2]Hubei Key Laboratory of Intelligent Geo-Information Processing,
School of Computer Science, China University of Geosciences, China

`{z_chen, fanyang, wenbingtao}@hust.edu.cn; sunkun@cug.edu.cn`

Abstract

In this paper, we present a second order spatial compatibility (SC^2) measure based method for efficient and robust point cloud registration (PCR), called SC^2-PCR [1]. Firstly, we propose a second order spatial compatibility (SC^2) measure to compute the similarity between correspondences. It considers the global compatibility instead of local consistency, allowing for more distinctive clustering between inliers and outliers at early stage. Based on this measure, our registration pipeline employs a global spectral technique to find some reliable seeds from the initial correspondences. Then we design a two-stage strategy to expand each seed to a consensus set based on the SC^2 measure matrix. Finally, we feed each consensus set to a weighted SVD algorithm to generate a candidate rigid transformation and select the best model as the final result. Our method can guarantee to find a certain number of outlier-free consensus sets using fewer samplings, making the model estimation more efficient and robust. In addition, the proposed SC^2 measure is general and can be easily plugged into deep learning based frameworks. Extensive experiments are carried out to investigate the performance of our method.

1. 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) [7, 24, 40], augmented reality [4, 11] and robotics applications [34]. A canonical solution first establishes feature correspondences and then estimates the 3D rotation

---
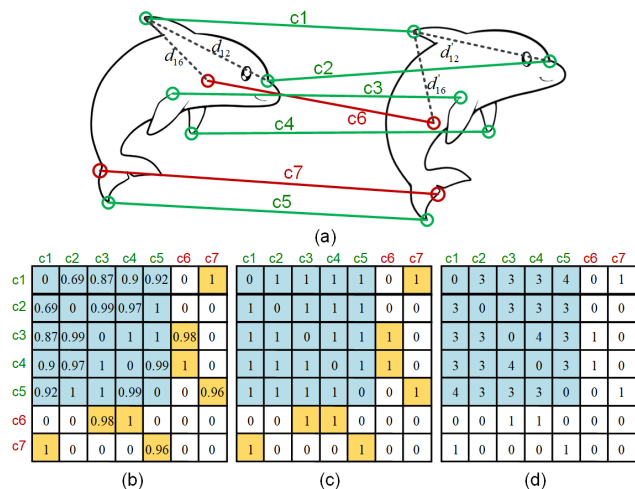
*Corresponding author.

[1]Code will be available at https://github.com/ZhiChen902/SC2-PCR.

Figure 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 by 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.

and translation that can best align the shared 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.

RANSAC [25] pioneers the iterative sampling strategy for model estimation. However, it needs more time to converge or sometimes there is no guarantee of an accurate solution due to massive outliers. Spatial Compatibility (SC) [5, 35, 45, 57] is a widely used similarity measure for boosting the robustness and efficiency of the rigid transformation estimation. It assumes that two correspondences will have 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 minor.

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. Fig. 1(b) is the spatial compatibility matrix of the correspondences in Fig. 1(a). As we can see from the yellow cells, c_6 and c_7 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 paper, 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 correspondences which are compatible, we compute the number of their commonly compatible correspondences in the global set. That is, 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 of two inliers is at least the number of inliers excluding themselves among all the correspondences. However, the outliers do not have such good properties. 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). From Fig. 1 (b) and Fig. 1 (c), we can see that the outliers c_6 and c_7 are compatible with some inliers, and the inliers are compatible with each other. But from Fig. 1 (d), we can see that the similarities between any two inliers are large while the similarities between the outliers with the other correspondences are small. 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 with 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 Eq. 8), we name it as the second order spatial compatibility (SC^2).

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, its difficult for an outlier to be simultaneously compatible with multiple correspondences and the score for it will be much smaller. 2) The traditional algorithms such as RANSAC and its variants [21, 25, 42, 53] need a large number of randomly samplings to find an outlier-free set for robust model estimation. However, 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 error 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 traditional first order metric and our second order metric, the SC^2 matrix is much more robust to obtain reliable sampling (see Fig. 3).

Based on the SC^2 measure, we design a full pipeline for point cloud registration, called SC^2-PCR. Following [5, 14, 50], it first selects several seeds that are likely to be inliers. Then we select a consensus set for each seed by finding those having the highest SC^2 scores with it. In order to further exclude outliers, a two-stage sampling strategy is carried out in a coarse-to-fine manner. Finally, we use the weighted SVD to estimate a tentative model for each seed and select the best one as the final output. In a nutshell, this paper distinguishes itself from existing methods in the following aspects.

- A second order spatial compatibility metric called SC^2 is proposed. We theoretically 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.

- Compared with state-of-the-art deep learning methods such as [5, 19, 35, 43], 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 also shown 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 [5]. We show in the experiment that PointDSC produces better results when combined with our method.

2. Related Work

3D Feature Matching. The widely used Iterative Closest Point [10] and its variants [12, 46, 49] establish correspondences by searching the nearest neighbor in coordinate space. Instead of using the distance in coordinate space, local feature descriptors aim to build 3D matches in feature space. Hand-crafted descriptors represent local feature by encoding spatial distribution histogram [26, 28, 31, 52], geometric histogram [15, 47, 48] or other attributes [61]. Recently, the deep learning techniques are also introduced to learn 3D local descriptors. The pioneering 3DMatch [62] builds a Siamese Network for extracting local descriptors. A number of recent networks try to boost the performance by designing rotation invariant modules [1, 22, 23, 58], fully

convolution modules [20], feature detection modules [6,59], coarse-to-fine modules [44, 60] and overlapping learning modules [29, 55]. Although these methods achieve remarkable performance improvement, it can hardly establish a totally outlier-free correspondence set.

Traditional Model Fitting. The model fitting methods estimate the geometric model from a noise correspondence set. The classic RANSAC [25] provides the most commonly adopted generation-and-verification pipeline for robust removing outliers. Many of its variants [21, 42, 53] introduce new sampling strategies and local optimization [21] to accelerate the estimation or boost the robustness. More recently, Graph-cut RANSAC [8] introduces the graphcut technique to better performing local optimization. Magsac [9] proposes a σ-consensus to build a threshold-free method for RANSAC. For specific 3D model-fitting, FGR [66] uses the Geman-McClure cost function and estimates the model through graduated non-convexity optimization. TEASER [56] reformulates the registration problem using a Truncated Least Squares (TLS) cost and solving it by a general graph-theoretic framework.

Learning based Model Fitting. Recent works also adopt deep learning techniques, which were first studied in 2D matching area, to model fitting tasks. The 2D correspondence selection network CN-Net [41] and its variants [13,16,17,37,51,63–65] formulate the model fitting as combination of a correspondence classification module and a model estimation module. Recent attempts [5,18,19,35,43] also introduce deep learning network for 3D correspondence pruning. 3DRegNet [43] reformulates the CN-Net [41] into 3D form and designs a regression module to solve rigid transformaton. DGR [19] introduces the fully convolution to better capture global context. PointDSC [5] develops a spatial consistency based non-local module and a Neural Spectral matching to accelerate the model generation and selection. DetarNet [18] presents decoupling solutions for translation and rotation. DHVR [35] exploits the deep hough voting to identify the consensus from the Hough space, so as to predict the final transformation.

3. Method

3.1. Problem Formulation

Given two point clouds to be aligned: source point cloud $X = \{x_i \in \mathbb{R}^3 \mid i = 1, ..., N_x\}$ and target point cloud $Y = \{y_j \in \mathbb{R}^3 \mid j = 1, ..., N_y\}$, we first extract local features for both of them. Then, for each point in the source point cloud, we find its nearest neighbor in the feature space among the target points to form N pairs of putative correspondences. The proposed method estimates the rigid transformation between the two point clouds, i.e., the rotation matrices ($R \in \mathbb{R}^{3\times3}$) and translation vectors ($t \in \mathbb{R}^3$). The pipeline of proposed method is shown in Fig. 2.

3.2. Second Order Spatial Compatibility

To analyze the effectiveness of a metric for sampling, we first define the probability of an ambiguity event as:

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

where M is a specific metric for measuring correspondence-wise similarity. $\mathrm{P}(Z)$ 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 that between two inliers. When $M_{in,out} > M_{in,in}$, outlier is the near neighbor of inlier, so the metric based sampling tends to fail. So the lower this probability is, the more robust the metric based sampling will be.

We first introduce the commonly used first order spatial compatibility (SC) measure [5,35,36,45]. The SC measure between correspondence i and j is defined as follows:

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

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 noise introduced by data acquisition and point cloud downsampling, $d_{in,in}$ is not exactly equal to 0, but less than a threshold d_{thr}. For convenience, we assume $d_{in,in}$ is uniformly distributed over d_{thr} and get the probability density function (PDF) of the distance difference between two inliers as follows:

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

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 $F(\cdot)$:

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

where d_r is the range of $d_{in,out}$ and $d_{out,out}$. An empirical F function on 3DMatch dataset is presented in Fig 3 (a). Obviously, d_r is much greater than d_{thr}. So we can approximate that $F(l)$ is a constant within $(0, d_{thr})$ as follows:

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

Next, we consider the ambiguity probability of SC as Eq. (1), i.e. , $\mathrm{P}(SC_{in,out} > SC_{in,in})$. According to Eq. (2), (3), (4) and (5), it can be computed as follows:

$$\begin{aligned} \mathrm{P}(SC_{in,out} > SC_{in,in}) &= \mathrm{P}(d_{in,out} < d_{in,in}) \\ &= \int_0^{d_{thr}} \int_0^l \mathrm{PDF}_{in,in}(l) \cdot \mathrm{PDF}_{in,out}(x) dx dl \\ &= \int_0^{d_{thr}} \int_0^l \frac{1}{d_{thr}} \cdot f_0 dx dl = \frac{d_{thr} \cdot f_0}{2} \end{aligned} \qquad (6)$$
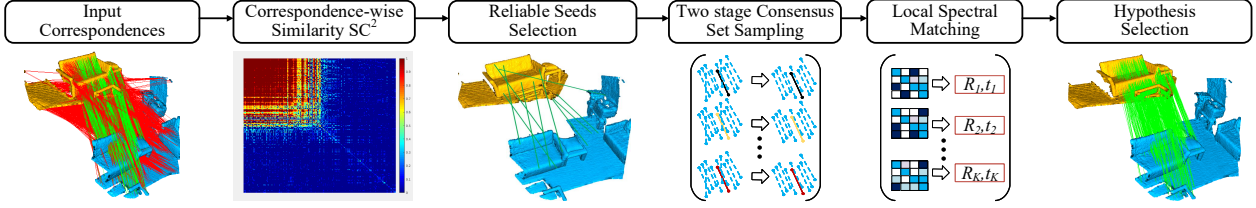
Figure 2. Pipeline of our method. **1.** Computing correspondence-wise second order spatial compatibility measure. **2.** Selecting some reliable correspondences as seeds. **3.** Performing the two-stage sampling around each seed. **4.** Performing local spectral matching to generate an estimation of R and t for each seed. **5.** Selecting the best estimation as final result.

Taking the 3DMatch [62] dataset as an example. Following [5], we set $d_{thr} = 10$cm, then the ambiguity probability of SC measure is about 0.1 as shown in Fig. 3 (a). Considering the amount of outliers might be large, the number of mistakes is not negligible even at this probability.

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:

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

Similarly, we analyze the ambiguity probability of this measure, i.e., $\mathrm{P}(SC^2_{in,out} > SC^2_{in,in})$. Suppose there are N pairs of correspondences and the inlier ratio is α. Then, we can prove that $\mathrm{P}(SC^2_{in,out} > SC^2_{in,in})$ can be expressed as (*The derivation is presented in the supplement materials*):

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

where $S(\cdot, \cdot)$ is the Skellam distribution [30, 32, 33]. 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 SC measure, we plot the curves of ambiguity probability for both of them according to Eq. (6) and (9). As shown in Fig. 3 (b), 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 anti the data with low inlier rate.
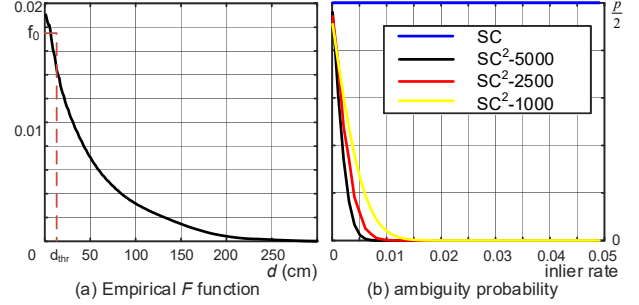

Figure 3. (a) The empirical probability density function (F) of $d_{in,out}$ and $d_{out,out}$. (b) The ambiguity probability. SC is spatial compatibility measure. SC^2-N (N = 5000, 2500, 1000) is the second order spatial compatibility measure with N correspondences.

3.3. Reliable Seed Selection

As mentioned above, 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 spectral matching technique [36] 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 [36]. Then, following [5, 36], the association of each correspondence with the leading eigen-vector is adopted as the confidence for this correspondence. The leading eigen-vetctor is solved by power iteration algorithm [39]. In order to ensure an even distribution of seed points, the correspondences with local maximum confidence score within its neighborhood with radius R are selected. The number of seed points (N_s) is determined by a proportion of the number of whole correspondences.

3.4. 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 $P(SC^2_{in,out} > SC^2_{in,in})$ is very small. Thus, when a seed is an inlier correspondences, 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 [5].

The second stage 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. 3 (b), since the higher inlier rate ensures a lower ambiguity probability, so that the potential outliers can also be further pruned. Note that we only discussed the case that the seed point is inlier. In fact, when the seed point is an outlier, it can also form a local consistency, especially when there are aggregated false matching in the correspondence set. We encourage these sets to also generate hypothesis and filter them at the final hypothesis selection step (Section 3.6) rather than at the early stage. In this way, we can avoid some correct assumptions being filtered out early.

3.5. Local Spectral Matching

In this step, we perform the weighted SVD [3] on the consensus set to generate an estimation of rigid transformation for each seed. Although the previous proposed sampling strategy can obtain outlier-free correspondence set, we find that the weighted SVD achieves better performance than treating all corrrespondences equally. This may be because that the inliers still have different degree of noises. So correspondences with bigger noise should have smaller weights when estimating rigid transformation. Traditional spectral matching [36] method analyzes the SC matrix to assign weight for each correspondence, which is effected by ambiguity problem [5]. Since the proposed SC^2 measure is more robust against ambiguity, we also replace the SC matrix with SC^2 measure. In order to facilitate matrix analysis, we convert the SC^2 measure into soft form ($\tilde{SC^2}$) as follows:

$$\tilde{SC^2} = \tilde{C} \cdot (\tilde{C} \times \tilde{C}),$$
$$\tilde{C}_{ij} = \text{ReLU}(1 - d^2_{ij}/d^2_{thr}), (1 \le i \le K_2, 1 \le j \le K_2) \tag{10}$$

where $\cdot$ is Hadamard product and $\times$ is matrix product. Then we conduct local spectral decomposition on $\tilde{SC^2}$ to obtain a weight w_i for correspondences i. Finally, the rotation R_k and translation t_k of seed k are computed by performing weighted SVD [19] within its consensus set.

3.6. Hypothesis Selection

In final step, we select the best estimation over the rigid transformations produced by all the consensus sets. We use the same criteria as RANSAC [25], i.e. inlier count, to select the final estimation. Specifically, for the estimation of k-th seed R_k and t_k, we count the number of correspondences that satisfy the constraints of R_k and t_k by a predefined error threshold (τ) as follows:

$$\text{count}_k = \sum_{i=1}^{N} [||R_k x_i + t_k|| < \tau], \tag{11}$$

where $[\cdot]$ is Iverson bracket. The R_k and t_k with the highest inlier count are selected as the final results.

4. Experiment

4.1. Datasets and Experimental Setup

Indoor scenes. We use the 3DMatch benchmark [62] for evaluating the performance on indoor scenes. For each pair of point clouds, we first use 5cm voxel grid to downsample the point cloud. Then we extract the local feature descriptors and match them to form the putative correspondences. Following [5], we use FPFH [47] (handcrafted descriptor) and FCGF [20] (learned descriptor) as feature descriptors respectively. The partial overlapping registration benchmark 3DLoMatch [29] is also adopted to further verify the performance of the method. Following [5,35], we use FCGF [20] and Predator [29] descriptors to generate putative correspondences.

Outdoor scenes. We use the KITTI dataset [27] for testing the effectiveness on outdoor scenes. Following [19,20], we choose the 8 to 10 scenes, obtaining 555 pairs of point clouds for testing. Then we construct 30cm voxel grid to down-sampling the point cloud and form the correspondences by FPFH and FCGF descritors respectively.

Evaluation Criteria. Following [5], we first report the registration recall (RR) under an error threshold. For the indoor scenes, the threshold is set to ($15°$, 30 cm), while the threshold of outdoor scenes is ($5°$, 60 cm). For a pair of point clouds, we calculate the errors of translation and rotation estimation separately. We compute the isotropic rotation error (RE) [38] and L2 translation error (TE) [19]. Following [5], we also report the outlier removal results using following three evaluation criteria: inlier precision (IP), inlier recall (IP) and F1-measure (F1).

Implementation Details. When computing the SC^2 matrix, the d_{thr} is set to the twice as the voxel size for downsampling (10cm for indoor scenes and 60cm for outdoor scenes). The number of seed (N_s in Section 3.3) is set to 0.2 * N, where N is the number of correspondences. When sampling consensus set, we select 30 nearest neighbors (K_1 = 30) of seed point at first sampling stage, and remain 20

Table 1:

		FPFH (traditional descriptor)						FCGF (learning based descriptor)						
		RR(%↑)	RE(∘↓)	TE(cm↓)	IP(%↑)	IR(%↑)	F1(%↑)	RR(%↑)	RE(∘↓)	TE(cm↓)	IP(%↑)	IR(%↑)	F1(%↑)	Time(s)
learning	DCP* [54]	3.22	8.42	21.40	-	-	-	3.22	8.42	21.40	-	-	-	0.07
	PointNetLK* [2]	1.61	8.04	21.30	-	-	-	1.61	8.04	21.30	-	-	-	0.12
	3DRegNet [43]	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 [19]	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 [35]	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
	DHVR-Origin [35]	80.22	2.06	6.87				91.40	2.08	6.61				0.46
	PointDSC [5]	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
traditional	SM [36]	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* [10]	5.79	7.93	17.59	-	-	-	5.79	7.93	17.59	-	-	-	0.25
	FGR [66]	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 [56]	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 [8]	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 [25]	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 [25]	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 [25]	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 [45]	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
	Ours	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

Table 1. Quantitative results on 3DMatch Dataset. Methods with * are correspondence-free methods.

	FCGF			Predator			
	RR(%)	RE(∘)	TE(cm)	RR(%)	RE(∘)	TE(cm)	Time(s)
DHVR [35]	54.41	4.14	12.56	65.41	4.97	12.33	3.55
DGR [19]	43.80	4.17	10.82	59.46	3.19	10.01	1.48
PointDSC [5]	56.09	3.87	10.39	68.89	3.43	9.60	0.10
FGR [66]	19.99	5.28	12.98	35.99	4.77	11.64	1.32
RANSAC [25]	46.38	5.00	13.11	64.85	4.28	11.04	2.86
CG-SAC [45]	52.31	3.84	10.55	64.01	3.86	10.94	0.25
Ours	57.83	3.77	10.46	69.46	3.46	9.58	0.11

Table 2. Quantitative results on 3DLoMatch Dataset.

correspondences ($K_2 = 20$) to form the consensus set. All the experiments are conducted on a machine with an INTEL Xeon E5-2620 CPU and a single NVIDIA GTX1080Ti.

4.2. Evaluation on Indoor Scenes

We first report the results on 3DMatch dataset in Tab. 1. We compare our method with 13 baselines: DCP [54], PointNetLK [2], 3DRegNet [43], DGR [19], DHVR [35], PointDSC [5], SM [36], ICP [10], FGR [66], TEASER [56], GC-RANSAC [8], RANSAC [25], CG-SAC [45]. The first 6 methods are based on deep learning, while the last 7 methods are traditional methods. For the deep learning methods, we use the provided pre-trained model of them for testing. The results of DHVR we tested have some difference with the original results, so we also report the results in their paper (DHVR-Origin in Tab. 1). DCP, PointNetLK and ICP are correspondence-free methods, so their results are not related with the descriptor.

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 Tab. 1, our method greatly outperforms all the methods. For the registration recall (RR), which is the most important criterion, our method improves it by about 6% over the closest competitors among the retested results (PointDSC and CG-SAC). Following [5, 19], 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
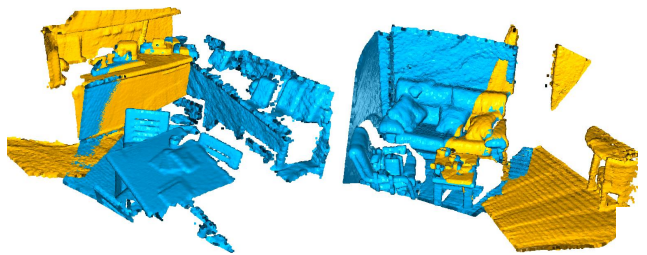

Figure 4. Qualitative registration results on 3DLoMatch dataset.

to avoid unreliable metrics. This strategy of measurement makes methods with high registration recall more likely to have large mean error, because they include more difficult data when calculating mean error. Nevertheless, our method still achieves competitive results on RE and TE. Our method is slightly worse than PointDSC on TE and RE, and better than other methods. For the outlier rejection results, our method achieves the highest inlier recall (IR) and F1-measure. The F1 of our method outperforms the PointDSC by 5.35%.

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 Tab. 1, since the inlier rate is higher than the correspondences obtained by FPFH descriptor, the performace of all of the feature based methods are boosted. Our method still achieves the best performance over all the methods, achieving 1.84% improvement over RANSAC on registration recall. Besides, the mean registration time for a pair of point clouds are also reported. Since the proposed method only need 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 Tab. 1, the mean registration time of our methods is 0.11s, which is over 20× faster than RANSAC with 4M iterations.

Robustness to lower overlap. Furthermore, we report

	FPFH			FCGF			Time(s)
	RR(%)	RE(∘)	TE(cm)	RR(%)	RE(∘)	TE(cm)	
DHVR [35]	-	-	-	99.10	0.29	19.80	0.83
DGR [19]	77.12	1.64	33.10	96.90	0.34	21.70	2.29
PointDSC [5]	98.20	0.35	8.13	98.02	0.33	21.03	0.45
FGR [66]	5.23	0.86	43.84	89.54	0.46	25.72	3.88
RANSAC [25]	74.41	1.55	30.20	80.36	0.73	26.79	5.43
CG-SAC [45]	74.23	0.73	14.02	83.24	0.56	22.96	0.73
Ours	99.64	0.32	7.23	98.20	0.33	20.95	0.31

Table 3. Quantitative results on KITTI Dataset. (For DHVR, the authors have neither released training code nor pretraining model on KITTI dataset, so we report the results in their paper.)

the results on the low-overlap scenarios: 3DLoMatch [29]. Following PointDSC [5] and DHVR [35], we adopt the FCGF [20] and Predator (There are two versions of Predator, and we use the updated one) [29] descriptors to generate correspondences. The registration recall (RR), rotation error (RE) and translation error (TE) are reported in Tab. 2. As shown by the data, whether combined with FCGF or Predator descriptor, our method achieves the highest registration recall. Meanwhile, we also present some qualitative results on 3DLoMatch dataset. As shown in Fig. 4, our method can successfully align two point clouds where the low overlap ratio is clearly visible.

4.3. Evaluation on Outdoor Scenes

In this experiments, we test on the outdoor KITTI [27] dataset. The results of DHVR [35], DGR [19], PointDSC [5], RANSAC [25], FGR [66], CG-SAC [45] are reported as comparison. DHVR, DGR and PointDSC are deep learning based methods, while the remaining methods are non-learning. As shown in Tab. 3, our method remarkably surpasses the non-learning methods. The registration recall (RR) of our method is 25.23% higher than that of RANSAC when combined with FPFH descriptor, and 17.84% higher when combined with FCGF descriptor. The errors of translation and rotation are also much lower than RANSAC. Our method with FPFH descriptor obtains the results of highest registration recall. For the learning networks, our method can achieve close performance with them.

4.4. 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 dataset. 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

	3DMatch		3DLoMatch		KITTI	
	FPFH	FCGF	FCGF	Predator	FPFH	FCGF
DGR-gen	49.48	81.89	23.75	45.03	73.69	86.12
PointDSC-gen	68.12	87.74	40.65	53.79	90.27	92.97
Ours	83.98	93.28	57.83	69.46	99.64	98.20

Table 4. Generalization results. The registration recall (%) on 3DMatch, 3DLoMatch and KITTI datasets are reported.
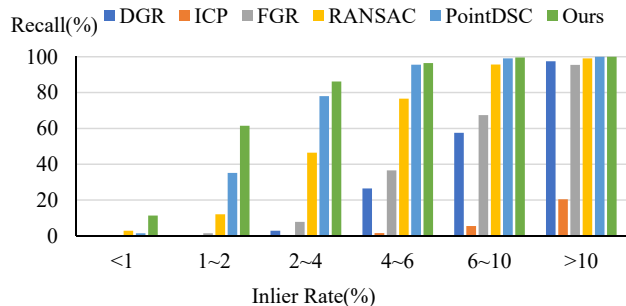

Figure 5. The registration recall under the different inlier ratio of the putative correspondences.

model to test on KITTI. As shown in Tab. 4, our method shows a significant improvement on registration recall without generalization problem. This further demonstrates the effectiveness of our method.

Robustness Anti Noise. An important factor to measure the performance of 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 ratio in Fig. 5. 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. 5, when the inlier rate is less than 2%, our method is significantly better than other methods. It demonstrates the robustness anti noise of our method.

4.5. Combined with learning network

To verify the flexibility of our proposed approach, we combine our approach with a recent deep learning approach PointDSC [5]. It adopts the spatial consistency matrix to guide the non-local module. 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 plugged our metrics into it. The registration recall of their vanilla version and combined version are shown in Tab. 5. It can be seen that adding our metrics can significantly boost the performance of network, especially for the generalization performance of network. It demonstrates that the proposed measure is flexible to combine with other methods.

	3DMatch		3DLoMatch		KITTI	
	FPFH	FCGF	FCGF	Predator	FPFH	FCGF
PointDSC	77.57	92.85	56.09	68.89	98.20	98.02
PointDSC+SC^2	83.24	93.10	57.05	69.04	99.10	98.02
PointDSC-gen	68.12	87.74	40.65	53.79	90.27	92.97
PointDSC-gen+SC^2	74.12	89.59	44.81	56.71	97.48	98.02

Table 5. The registration recall of combining SC^2 measure with learning based network.

4.6. Ablation Study

In this section, we perform ablation study 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 the Row 1 and Row 7 of Tab. 6. 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 seaching the k-nearest neighbors in metric space. The Spatial Compatibility (SC) adopted by previous works [5, 45, 57] is also utilized as the sampling guidance, and the results are reported as comparison. As shown in Row 1, 3 and 7, 9 of Tab. 6, 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 8, 9. 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 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 9, 10 in Tab. 6, 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 adopt the instance-equal SVD to generate an estimation of translation and rotation, which is sensitive to errors. We replace the instance-equal SVD [3] with the weighted SVD [19, 43], 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.

		SC	SC^2	TS	LSM	Seed	RR(%)	RE(○)	TE(cm)	Time(s)
FPFH	1)						66.10	3.95	11.03	2.86
	2)	✓					71.56	2.07	6.48	0.27
	3)		✓				80.89	2.34	6.92	0.31
	4)		✓	✓			82.85	2.32	6.69	0.33
	5)		✓	✓	✓		84.10	2.13	6.56	0.37
	6)		✓	✓	✓	✓	83.98	2.18	6.56	0.11
FCGF	7)						91.44	2.69	8.38	2.86
	8)	✓					87.52	2.42	7.66	0.27
	9)		✓				93.10	2.16	6.76	0.31
	10)		✓	✓			93.22	2.10	6.88	0.33
	11)		✓	✓	✓		93.28	2.08	6.56	0.37
	12)		✓	✓	✓	✓	93.28	2.08	6.55	0.11

Table 6. 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.

Comparing Row 4, 5 and 10, 11 in Tab. 6, 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 do 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 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 set. Row 5, 6 and 11, 12 of Tab. 6 shows that Seed Selection can reduce registration time by more than half without much performance degradation.

5. Conclusion

In this paper, 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 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 as the final result. Extensive experiments demonstrate that our method achieves the state-of-the-art performance and high efficiency. Meanwhile, we also demonstrate the proposed SC^2 is a flexible measure, which can be combined with learning networks to further boost their performance.

6. Acknowledgements

References

[1] Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11753–11762, 2021. 2

[2] Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7163–7172, 2019. 6

[3] K Somani Arun, Thomas S Huang, and Steven D Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5):698–700, 1987. 5, 8

[4] Ronald T Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997. 1

[5] Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15859–15869, 2021. 1, 2, 3, 4, 5, 6, 7, 8

[6] Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6359–6367, 2020. 3

[7] Tim Bailey and Hugh Durrant-Whyte. Simultaneous localization and mapping (slam): Part ii. *IEEE robotics & automation magazine*, 13(3):108–117, 2006. 1

[8] Daniel Barath and Jiří Matas. Graph-cut ransac. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6733–6741, 2018. 3, 6

[9] Daniel Barath, Jiri Matas, and Jana Noskova. Magsac: marginalizing sample consensus. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10197–10205, 2019. 3

[10] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992. 2, 6

[11] Mark Billinghurst, Adrian Clark, and Gun Lee. A survey of augmented reality. 2015. 1

[12] Sofien Bouaziz, Andrea Tagliasacchi, and Mark Pauly. Sparse iterative closest point. In *Computer graphics forum*, volume 32, pages 113–123. Wiley Online Library, 2013. 2

[13] Eric Brachmann and Carsten Rother. Neural-guided ransac: Learning where to sample model hypotheses. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4322–4331, 2019. 3

[14] Luca Cavalli, Viktor Larsson, Martin Ralf Oswald, Torsten Sattler, and Marc Pollefeys. Handcrafted outlier detection revisited. In *European Conference on Computer Vision*, pages 770–787. Springer, 2020. 2

[15] Hui Chen and Bir Bhanu. 3d free-form object recognition in range images using local surface patches. *Pattern Recognition Letters*, 28(10):1252–1262, 2007. 2

[16] Hongkai Chen, Zixin Luo, Jiahui Zhang, Lei Zhou, Xuyang Bai, Zeyu Hu, Chiew-Lan Tai, and Long Quan. Learning to match features with seeded graph matching network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6301–6310, 2021. 3

[17] Zhi Chen, Fan Yang, and Wenbing Tao. Cascade network with guided loss and hybrid attention for finding good correspondences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1123–1131, 2021. 3

[18] Zhi Chen, Fan Yang, and Wenbing Tao. Detarnet: Decoupling translation and rotation by siamese network for point cloud registration. *arXiv preprint arXiv:2112.14059*, 2021. 3

[19] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2514–2523, 2020. 2, 3, 5, 6, 7, 8

[20] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8958–8966, 2019. 3, 5, 7

[21] Ondrej Chum and Jiri Matas. Matching with prosac-progressive sample consensus. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 220–226. IEEE, 2005. 2, 3

[22] Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 602–618, 2018. 2

[23] Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppfnet: Global context aware local features for robust 3d point matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 195–205, 2018. 2

[24] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localization and mapping: part i. *IEEE robotics & automation magazine*, 13(2):99–110, 2006. 1

[25] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 1, 2, 3, 5, 6, 7

[26] Andrea Frome, Daniel Huber, Ravi Kolluri, Thomas Bülow, and Jitendra Malik. Recognizing objects in range data using regional point descriptors. In *European conference on computer vision*, pages 224–237. Springer, 2004. 2

[27] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012. 5, 7

[28] Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Jianwei Wan, and Min Lu. A novel local surface feature for 3d object recognition under clutter and occlusion. *Information Sciences*, 293:196–213, 2015. 2

[29] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4267–4276, 2021. 3, 5, 7

[30] Joseph Oscar Irwin. The frequency distribution of the difference between two independent variates following the same poisson distribution. *Journal of the Royal Statistical Society*, 100(3):415–416, 1937. 4

[31] Andrew E Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):433–449, 1999. 2

[32] Dimitris Karlis and Ioannis Ntzoufras. Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):381–393, 2003. 4

[33] Dimitris Karlis and Ioannis Ntzoufras. Bayesian analysis of the differences of count data. *Statistics in medicine*, 25(11):1885–1905, 2006. 4

[34] Ioannis Kostavelis and Antonios Gasteratos. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, 66:86–103, 2015. 1

[35] Junha Lee, Seungwook Kim, Minsu Cho, and Jaesik Park. Deep hough voting for robust global registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15994–16003, 2021. 1, 2, 3, 5, 6, 7

[36] Marius Leordeanu and Martial Hebert. A spectral technique for correspondence problems using pairwise constraints. 2005. 3, 4, 5, 6

[37] Yuan Liu, Lingjie Liu, Cheng Lin, Zhen Dong, and Wenping Wang. Learnable motion coherence for correspondence pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3237–3246, 2021. 3

[38] Yi Ma, Stefano Soatto, Jana Kosecka, and S Shankar Sastry. *An invitation to 3-d vision: from images to geometric models*, volume 26. Springer Science & Business Media, 2012. 5

[39] RV Mises and Hilda Pollaczek-Geiringer. Praktische verfahren der gleichungsauflösung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 9(1):58–77, 1929. 4

[40] Michael Montemerlo, Sebastian Thrun, Daphne Koller, Ben Wegbreit, et al. Fastslam: A factored solution to the simultaneous localization and mapping problem. *Aaai/iaai*, 593598, 2002. 1

[41] Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2666–2674, 2018. 3

[42] Kai Ni, Hailin Jin, and Frank Dellaert. Groupsac: Efficient consensus in the presence of groupings. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2193–2200. IEEE, 2009. 2, 3

[43] G Dias Pais, Srikumar Ramalingam, Venu Madhav Govindu, Jacinto C Nascimento, Rama Chellappa, and Pedro Miraldo.

3dregnet: A deep neural network for 3d point registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7193–7203, 2020. 2, 3, 6, 8

[44] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. *arXiv preprint arXiv:2202.06688*, 2022. 3

[45] Siwen Quan and Jiaqi Yang. Compatibility-guided sampling consensus for 3-d point cloud registration. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7380–7392, 2020. 1, 3, 6, 7, 8

[46] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *Proceedings third international conference on 3-D digital imaging and modeling*, pages 145–152. IEEE, 2001. 2

[47] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE international conference on robotics and automation*, pages 3212–3217. IEEE, 2009. 2, 5

[48] Radu Bogdan Rusu, Nico Blodow, Zoltan Csaba Marton, and Michael Beetz. Aligning point cloud views using persistent feature histograms. In *2008 IEEE/RSJ international conference on intelligent robots and systems*, pages 3384–3391. IEEE, 2008. 2

[49] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In *Robotics: science and systems*, volume 2, page 435. Seattle, WA, 2009. 2

[50] Kun Sun, Wenbing Tao, and Yuhua Qian. Guide to match: multi-layer feature matching with a hybrid gaussian mixture model. *IEEE Transactions on Multimedia*, 22(9):2246–2261, 2019. 2

[51] Weiwei Sun, Wei Jiang, Eduard Trulls, Andrea Tagliasacchi, and Kwang Moo Yi. Acne: Attentive context normalization for robust permutation-equivariant learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11286–11295, 2020. 3

[52] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique shape context for 3d data description. In *Proceedings of the ACM workshop on 3D object retrieval*, pages 57–62, 2010. 2

[53] Philip Hilaire Torr, Slawomir J Nasuto, and John Mark Bishop. Napsac: High noise, high dimensional robust estimation-its in the bag. In *British Machine Vision Conference (BMVC)*, 2002. 2, 3

[54] Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3523–3532, 2019. 6

[55] Hao Xu, Shuaicheng Liu, Guangfu Wang, Guanghui Liu, and Bing Zeng. Omnet: Learning overlapping mask for partial-to-partial point cloud registration. *arXiv preprint arXiv:2103.00937*, 2021. 3

[56] Heng Yang, Jingnan Shi, and Luca Carlone. Teaser: Fast and certifiable point cloud registration. *IEEE Transactions on Robotics*, 37(2):314–333, 2020. 3, 6

[57] Jiaqi Yang, Zhiqiang Huang, Siwen Quan, Zhaoshuai Qi, and Yanning Zhang. Sac-cot: Sample consensus by sampling

compatibility triangles in graphs for 3-d point cloud registration. *IEEE Transactions on Geoscience and Remote Sensing*, 2021. 1, 8

[58] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 206–215, 2018. 2

[59] Zi Jian Yew and Gim Hee Lee. 3dfeat-net: Weakly supervised local 3d features for point cloud registration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 607–623, 2018. 3

[60] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34, 2021. 3

[61] Andrei Zaharescu, Edmond Boyer, Kiran Varanasi, and Radu Horaud. Surface feature detection and description with applications to mesh matching. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 373–380. IEEE, 2009. 2

[62] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1802–1811, 2017. 2, 4, 5

[63] Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao. Learning two-view correspondences and geometry using order-aware network. *arXiv preprint arXiv:1908.04964*, 2019. 3

[64] Chen Zhao, Zhiguo Cao, Chi Li, Xin Li, and Jiaqi Yang. Nm-net: Mining reliable neighbors for robust feature correspondences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 215–224, 2019. 3

[65] Chen Zhao, Yixiao Ge, Feng Zhu, Rui Zhao, Hongsheng Li, and Mathieu Salzmann. Progressive correspondence pruning by consensus learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6464–6473, 2021. 3

[66] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In *European Conference on Computer Vision*, pages 766–782. Springer, 2016. 3, 6, 7