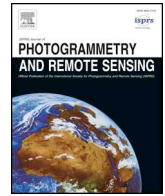




ELSEVIER

Contents lists available at ScienceDirect

## ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: [www.elsevier.com/locate/isprsjprs](http://www.elsevier.com/locate/isprsjprs)

# Inlier extraction for point cloud registration via supervoxel guidance and game theory optimization

Wei Li<sup>a,b</sup>, Cheng Wang<sup>a,b</sup>, Congren Lin<sup>a,b</sup>, Guobao Xiao<sup>c</sup>, Chenglu Wen<sup>a,b</sup>, Jonathan Li<sup>a,b,d</sup>

<sup>a</sup> Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, 422 Siming Road South, Xiamen, FJ 361005, PR China

<sup>b</sup> Digital Fujian Institute of Urban Traffic Big Data Research, Xiamen University, Xiamen, PR China

<sup>c</sup> College of Computer and Control Engineering, Minjiang University, PR China

<sup>d</sup> Departments of Geography and Environmental Management and Systems Design Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

## ARTICLE INFO

## Keywords:

Supervoxel segmentation  
Non-cooperative game  
Keypoint correspondences  
Point cloud registration

## 2018 MSC:

00-01  
99-00

## ABSTRACT

As a key step in Six-Degree-of-Freedom (6DoF) point cloud registration, 3D keypoint technique aims to extract matches or inliers from random correspondences between the two keypoint sets. The major challenge in 3D keypoint techniques is the high ratio of mismatched or outliers in random correspondences in real-world point cloud registration. In this paper, we present a novel inlier extraction method, which is based on Supervoxel Guidance and Game Theory optimization (SGGT), to extract reliable inliers and apply for point cloud registration. Specifically, to reduce the scale of keypoint correspondences, we first construct powerful groups of keypoint correspondences by introducing supervoxels, which involves 3D spatial homogeneity. Second, to select promising combined groups, we present a novel ‘fit-and-remove’ strategy by incorporating 3D local transformation constraints. Third, to extract purer inliers for point cloud registration, we propose a grouping non-cooperative game algorithm, which considers the relationship between the combined groups. The proposed SGGT, by eliminating the mismatched combined groups globally, avoids the false combined groups that lead to the failed estimation of rigid transformations. Experimental results show that when processing on large keypoint sets, the proposed SGGT is over 100 times more efficient compared to the state-of-the-art, while keeping the similar accuracy.

## 1. Introduction

As an important research direction in computer vision, 3D point cloud registration is a key point for 3D reconstruction (Haala and Kada, 2010), Simultaneous Localization and Mapping (SLAM) (Durrant-Whyte and Bailey, 2006; Engel et al., 2014), and autonomous driving (Levinson et al., 2011). The goal of 3D point cloud registration is to find a transformation to map a source point cloud  $\mathbf{P}$  to the corresponding target point cloud  $\mathbf{Q}$ . In this paper, we focus on a rigid transformation that involves only the six Degrees-of-Freedom (6DoF) parameters, i.e., rotation and translation parameters. As a popular technique, 3D keypoint techniques (Tam et al., 2013) for point cloud registration are to extract matches or inliers. Thus, if there are no mismatches or outliers, the 6DoF parameters usually can be fitted by the least squares sense (Arun, 1987). However, a high ratio of outliers inevitably leads to the generation of mismatches by most of 3D keypoint matching methods (Zeng et al., 2016; Huang et al., 2017; Yew and Lee, 2018; Gojcic et al., 2019). Therefore, an effective method for inlier extraction must be

constructed.

The technique of inlier extraction belongs to the category of geometric model fitting. As one of the standard approaches, Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981), and its variants (Chum et al., 2003; Tran et al., 2014), using least squares, randomly sample minimal subsets of size  $m$  ( $m = 2$  for rotation,  $m = 3$  for rigid transformation) to estimate the 6DoF parameters. However, because of sampling randomness, the consistency of their solutions cannot be guaranteed. Other than this, a high outlier ratio (approaching 99%) of keypoint correspondences, causes great resistance for RANSAC-like methods. Recently, Xiao et al. (2018), Xiao et al. (2016) proposed a Superpixel-based method to Deterministically Fit geometric model (SDF) in two-view images. However, first, the initial model hypothesis, which is generated based on the 2D superpixel segmentation method (Achanta et al., 2012), does not perform in 3D point clouds. Second, because of the high ratio of outliers, the model selection strategy, which is determined by matching scores in advance, is of low accuracy. Especially, a false model selection i.e., a mismatched group of

E-mail address: [cwang@xmu.edu.cn](mailto:cwang@xmu.edu.cn) (C. Wang).

<https://doi.org/10.1016/j.isprsjprs.2020.01.021>

Received 16 October 2019; Received in revised form 12 January 2020; Accepted 21 January 2020

Available online 31 March 2020

0924-2716/ © 2020 Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).

correspondences, that disturb the estimation of rigid transformations, usually leads to the failure of registration.

Due to the three-dimensional properties, inlier extraction for point cloud registration is different from the traditional geometric model fitting problem. The main differences are that, the task is more specific, that is, fitting 6DoF parameters. Besides, for the application, the inliers are applied to the estimation of 6DoF parameters that requires higher extraction accuracy. To serve 3D point cloud registration, we propose a novel inlier extraction method based on supervoxel guidance and game theory optimization. Especially, a supervoxel, which can be defined as a compact point cluster, ensures 3D spatial homogeneity (Papon et al., 2013; Lin et al., 2018). Thus, there is a high probability that keypoint correspondences are the inliers belong to the same supervoxel pair. Based on 3D supervoxel segmentation, we first group initial keypoint correspondences to generate powerful and significant groups, while reducing the size of keypoint correspondences. Then, to select more promising combined groups, incorporating 3D local spatial transformation constraints on the ‘fit-and-remove’ strategy (Xiao et al., 2018), we construct an improved ‘fit-and-remove’ strategy. Finally, to avoid the disturbance of falsely combined groups, we propose a grouping non-cooperative game algorithm that, considers the relationship between the combined groups, to reject the mismatched combined groups globally. The main contributions of this work are summarized as follows:

- (1) We present a supervoxel-guided method, incorporating 3D spatial characteristics, to extract promising groups of keypoint correspondences. This method achieves coarse extraction from initial keypoint correspondences with a high outlier ratio (approaching 99%).
- (2) We present a grouping non-cooperative game algorithm that further and globally removes the mismatched combined groups. This algorithm avoids the failures of 3D rigid transformation estimation due to some falsely combined groups.
- (3) We achieve inlier extraction with a nearly constant computational effort. When the size of keypoint correspondences reaches 10,000, the implementation of the proposed SGGT is more efficient and nearly 100 times faster than state-of-the-art methods, while keeping the similar accuracy.

## 2. Related work

The popular methods for point cloud registration can be mainly represented as ICP-like methods (Besl, 1992; Bae and Lichti, 2008; Yang et al., 2016; Campbell and Petersson, 2016), RANSAC-like methods (Aiger et al., 2008; Mellado et al., 2014), local-feature-based methods (Johnson and Hebert, 1999; Rusu et al., 2009; Zeng et al., 2016; Gojcic et al., 2019). ICP-like methods generally alternate between estimating the point correspondence and the transformation matrix. However, these methods rely on the assumption that all points have pairwise counterparts between two sets. Furthermore, they are sensitive to a given initialization. RANSAC-like methods, using the idea of planar congruent sets to compute optimal global rigid transformation, are a randomized alignment approach. However, because of their point-level operation, RANSAC-like methods are easy to be sub-optimal. Local-feature-based methods mainly contain two steps: local feature description and match or inlier extraction. The review of local-feature-based methods is given below.

Almost all the 3D keypoint matching methods were implemented based on local features. Therefore, we first briefly review some representative methods of point cloud registration based on local features. Many handcrafted methods were designed to describe geometric properties of local patches. Johnson and Hebert (1999) presented a data level shape descriptor that is invariant to rigid transformations and robust to clutter and occlusion. However, the spin image is sensitive to varying mesh resolutions and nonuniform sampling. Rusu et al. (2008)

proposed point feature histograms (PFH) to encode a local surface, and further constructed a Fast Point Feature Histogram (FPFH) (Rusu et al., 2009) that retains the majority discriminative power of the PFH with a reduced computational complexity. However, their methods are sensitive to outliers and noise. Tombari et al. (2010) proposed Signature of Histograms of Orientations (SHOT) that is very robust to noise, but sensitive to mesh resolution variation. With the advent of deep learning, representative works have significant superiority than the handcrafted methods. For examples, Zeng et al. (2016) proposed a 3DMatch that leverages millions of correspondence labels found in existing RGB-D reconstructions to learn local descriptors. However, 3DMatch ignore the nature of input: sparsity and unstructured-mess. Deng et al. (2018) presented a PPFNet that is highly aware of the global context on pure geometry. However, PPFNet is not fully rotation invariant. Gojcic et al. (2019) presented 3DSmoothNet with fully convolutional layers for 3D point cloud matching that outperforms the PPFNet by more than 20 percent points. Although Local-feature-based methods have been greatly improved, it is still challenging to find a unique and consistent keypoints (i.e., inlier). Therefore, the effective methods of inlier extraction need to be constructed.

The technique of inlier extraction belongs to the category of geometric model fitting. As a popular method, RANSAC (Fischler and Bolles, 1981) randomly samples a minimal subset of data points and attempt to estimate the parameters of model. However, RANSAC requires a huge amount of trial when the expected confidence of inlier extraction is high. In addition, RANSAC does not guarantee global optimality due to randomness. Many modifications of RANSAC developed random sampling and accelerating strategies to improve original RANSAC. Specially, (Kanazawa and Kawakami, 2004 and Chum et al., 2003) proposed to guide sampling minimal subsets. Despite better performance than RANSAC, they cannot get consistent and tractable fitting results. Chin et al. (2011) presented an accelerated sampling scheme by residual sorting information, which dramatically reduces the number of samples required. However, these methods still require the most mismatches to be pre-eliminated. More recently, Svärm et al. (2014) proposed another relevant method for camera localization from 2D to 3D correspondences. Before invoking a globally optimal algorithm in their approach, they also conducted a guaranteed outlier rejection scheme for 2D-3D point matches. Because our target problems (3DoF rotational and 6DoF rigid registration) differ from those of Svärm et al. (2014)’s, the core geometric motivations and operations of our work are vastly different from theirs. Barath and Matas (2018) presented a Graph-Cut RANSAC method in two-views images which is a locally optimized RANSAC alternating graph-cut and model re-fitting. The Graph-Cut RANSAC could run in real-time and is much simpler to implement than RANSAC-like methods. However, the high ratio of outliers (up to 99%) in real data would be the biggest obstacle to the algorithm.

Many global optimization algorithms have been proposed. Some studies (Enqvist and Kahl, 2008; Olsson et al., 2009; Hartley and Kahl, 2009; Parra et al., 2014) focused on parametric spatial search ( $SO(3)$  for rotation,  $SE(3)$  for rigid transformation) using Branch and Bound (BnB) (Salhi et al., 1994) to optimize their respective objective function. However, the runtime of BnB increases exponentially with the input size. By combining a spatial line process (Black and Rangarajan, 1996), Zhou et al. (2016) conducted a robust objective function to alleviate the effects of local optima. Hence, global optimality cannot be guaranteed. Briaies and Gonzalez-Jimenez (2017) presented a unified formulation for 3D registration problem that integrates common geometric registration modalities (point-to-point, point-to-line, and point-to-plane). However, the feasibility of this method mainly lies in their reasonable use of rotation constraints. The methods of Game Theory (GT) (Albarelli et al., 2009; Albarelli et al., 2010; Torsello et al., 2012; Albarelli et al., 2013), which consider the relationship between correspondences, have been proposed for point cloud registration. These methods first define a payoff matrix of the strategy, then attempt to find

an inlier subset of correspondences by maximizing average pairwise consistency. Thus, the solution is optimized according to an evolutionary stable state. However, the outliers are scattered in initial keypoint correspondences of real-world application, which makes it difficult to reject the mismatches directly. New types of methods, GORE (Bustos and Chin, 2017; Bustos and Chin, 2015), exploited the underlying geometry of the target model to reject mismatches. For surface symmetric models, these methods are prone to failure due to checks of 3D geometric consistency. In an extension of their work, Chin et al. (2016) investigated to perform GORE to a solver using Mixed Integer Linear Programming (MILP). At a scale below 2,000 pairs of correspondences, these algorithms are lightweight inlier extracting methods.

Recently, Xiao et al. (2018), Xiao et al. (2016) proposed an SDF method to fit and segment multiple-structure data. Here, Unfortunately, there are some drawbacks. First, a 3D point cloud consists of sparse discrete points and cannot be over-segmented based on 2D superpixels. Second, because of the high ratio of outliers, it is difficult to extract high-precision inliers that is only determined by matching scores. With the development of 3D point cloud processing, Papon et al. (2013), Lin et al. (2018) proposed the over-segmentation methods for 3D point clouds. Besides, many supervoxel-based methods (Aijazi et al., 2014; Fan et al., 2016; Li and Sun, 2018) have been proposed for the application such as object detection, segmentation and refinement of point clouds. Therefore, we extend SDF from 2D to 3D and applied to 3D point cloud registration in this paper.

### 3. Methodology

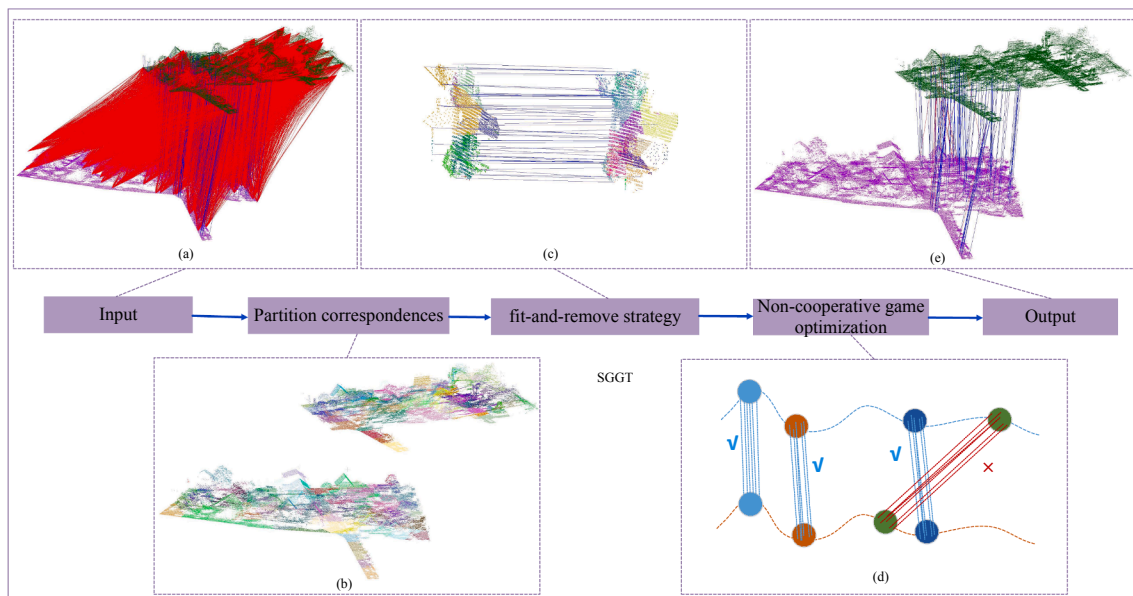
This section introduces the proposed SGGT for point cloud registration. As shown in Fig. 1, first, all pairwise keypoints are combined to generate keypoint correspondences between point cloud  $P$  and point cloud  $Q$ . To enhance the significance of keypoint correspondences, based on 3D supervoxels (Lin et al., 2018), keypoint correspondences are combined to generate more powerful groups. Then, an improved ‘fit-and-remove’ strategy, which considers feature appearance and three-dimensional characteristics, is presented to extract promising candidate groups. Finally, To extract more correct groups of correspondences, we construct a grouping non-cooperative game to reject

globally the mismatched combined groups. Here, we introduce the following two related concepts: matches and mismatches. **Matches:** Geometrically consistent correspondences, i.e., inliers; **Mismatches:** Geometrically inconsistent correspondences, i.e., outliers.

#### 3.1. Partition correspondences based on supervoxels

Based on supervoxel segmentation, each point on a point cloud is assigned to a unique supervoxel label. Therefore, a supervoxel contains a set of points. A supervoxel pair is combined by two point sets from the source and the target point clouds, respectively. Thus, any two keypoint correspondences with the same supervoxel label pair can be divided into a group. In fact, because of the 3D spatial homogeneity, keypoint correspondences have a high probability of belonging to the inliers in the same supervoxel pair. Especially, Lin et al. (2018) proposed novel supervoxel segmentation that well preserves the boundary for scene point clouds. Therefore, to acquire powerful correspondences in three-dimensional space, we combine supervoxel facets to group the keypoint correspondences. For each keypoint pair between two point clouds, the corresponding local features are described by the FPFH descriptor (Rusu et al., 2009).

Based on supervoxel segmentation, we group keypoint correspondences and generate a group set  $G = \{g_1, g_2, \dots, g_{K_0}\}$ , where  $K_0$  is the total number of supervoxel pairs, that is, the keypoint correspondences in the same pair of supervoxels are grouped into a set. Thus, the group set, formed by the combination of grouped keypoint correspondences, is more significant and greatly reduces the size of keypoint correspondences. It should be noted that, given the number of keypoint samples (approximately 1,000 in the evaluation experiment), the number of keypoint correspondences between two supervoxels (i.e., supervoxel correspondence) is affected by the supervoxel resolution and occlusion. Especially, 1) the supervoxel resolution  $r_s$  of source and target point cloud is small enough that there is no keypoint correspondence in a supervoxel pair, thereby the smaller the supervoxel resolution, the smaller the number of keypoints correspondences. 2) Because of occlusion in the scan, the difference of size between two supervoxels might vary greatly, which makes the number of keypoint correspondences unstable. Then, neighboring constraints are



**Fig. 1.** The framework of the proposed method. (a) The initial keypoint correspondences between two point clouds. (b) Supervoxel facets generation (each facet with the same color denotes a supervoxel) for grouping keypoint correspondences. (c) Selecting the most promising combined group based on matching score and checking if it is match or not. (d) Rejecting the false combined groups by grouping non-cooperative game optimization. (e) The final correspondences are used for point cloud registration. Note that, the blue lines stand for the inliers, the red lines represent outliers.  $\checkmark$  represents that a combined group is matching, and  $\times$  represents that a combined group is mismatching.

constructed by combining adjacent groups based on the first-order neighborhood of a supervoxel (i.e., a local patch). Here, the  $n$ -order neighborhood of supervoxel (Wang et al., 2017) can be defined as follows:

$$N_n(c) = \{c_t | D(c, c_t) \leq n, c_t \in C\} \quad (1)$$

where  $C$  is a vertex set composed of supervoxel centers and the edges exist only between directly neighboring supervoxels. The distance  $D(c, c_t)$  is defined as the minimum number of edges between two vertices  $c$  and  $c_t$ . For a supervoxel centered at  $c$ , all the adjacent supervoxels within the distance  $n$  constitute a local patch.

Instead of a 2D grid interval in superpixel-guided methods, a 3D spherical neighborhood metric of a supervoxel is computed and used to combine a group  $g_s \in G$  with each of its adjacent groups. A combined group  $\tilde{g}_s \in \tilde{G}$  is represented as follows:

$$\tilde{g}_s = g_s \cup N(g_s), \quad N(g_s) = \{g_l | g_l \in G, r(c_s, c_l) \leq r_s\} \quad (2)$$

where  $N(g_s)$  is the adjacent group set of  $g_s$  in three-dimensional space. The corresponding centers of first-order neighborhood of a supervoxel are represented as  $c_s$ , and similarly,  $c_l$ . The distance between two combined supervoxels in a point cloud is represented by  $r(\dots)$ . The expected supervoxel resolutions are represented by  $r_s$ . As seen in Fig. 2, we demonstrate the supervoxel segmentation of point cloud (bunny), and separately show the first-order neighborhood of a supervoxel. Based on first-order neighborhood of a supervoxel, the combining process of keypoint correspondences can be seen in Fig. 3.

Using the combining method of Eq. 2, any two adjacent groups, in which the distance of the centers between two neighboring supervoxels is smaller than  $r_s$ , are combined to improve the significance of the combined groups, and simultaneously reduce the scale of the combined groups. Then, to select promising we can achieve a coarse extraction by matching the similarities of two corresponding FPFH features. Especially, for a given combined group  $\tilde{g}_i$ , the corresponding matching similarities are represented by  $s_i = [s_i^1, s_i^2, \dots, s_i^{N_i}]$ . Here,  $s_i^l = \|f_{p_i^u} - f_{q_i^v}\|$ ,  $u \in \{1, \dots, U_i\}$ ,  $v \in \{1, \dots, V_i\}$ ,  $N_i = U_i \times V_i$  and  $l \in \{1, \dots, N_i\}$ ;  $U_i$  is the number of keypoints  $P_i \in \mathbf{P}$  in the  $i$ -th combined group,  $V_i$  is the number of keypoints  $Q_i \in \mathbf{Q}$  in the  $i$ -th combined group,  $\mathbf{P}$  and  $\mathbf{Q}$  represent source and target point clouds, respectively;  $f_{p_i^u}$  is the FPFH feature of keypoint  $p_i^u$ , where  $p_i^u \in P_i$  is the  $u$ -th keypoint in the  $i$ -th combined group;  $f_{q_i^v}$  is the FPFH feature of keypoint  $q_i^v$ , where  $q_i^v \in Q_i$  is the  $v$ -th keypoint in the  $i$ -th combined group. A combined group  $\tilde{g}_i$  is sorted according to their matching similarity and denoted as follows:

$$\tilde{g}_i = [x_i^1, x_i^2, \dots, x_i^{N_i}] \quad (3)$$

where  $\tilde{g}_i$  are the correspondences sorted in non-ascending order of the similarities. Thus,  $\tilde{g}_i$  is a sorted group to  $\tilde{g}_i$ , where  $x_i^l$  is a keypoint correspondence.  $N_i$  represents the number of keypoint correspondences in the  $i$ -th combined group. Thus, all the combined groups are denoted

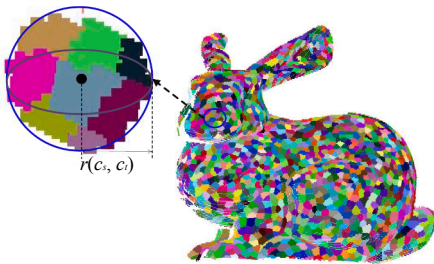


Fig. 2. First-order neighborhood of a supervoxel.

as  $\tilde{G} = \{\tilde{g}_i\}_{i=1}^{K_1}$ , where  $K_1$  is the size of the combined groups.

### 3.2. Fit-and-remove strategy with spatial constraints

Given the combined correspondence groups  $\tilde{G}$ , the focus in this subsection is mainly on selecting promising combined groups. The previous ‘fit-and-remove’ framework (Xiao et al., 2018) sequentially selected a combined group that has the largest number of inliers (i.e., promising group) and removed redundant combined groups. However, on one hand, because of the noise, outliers, varying occlusion, etc. in real-world point clouds, the feature description is insufficiently robust such that a selected promising combined group has a high probability of being a mismatched group. On the other hand, for each supervoxel, combined groups are constructed based on first-order supervoxels. Any neighboring combined group that partially overlap with the selected combined group, (See Fig. 4 (b)), are considered as a redundant combined group. To ensure the selected group is more reliable, we consider incorporating the 3D local spatial transformation constraints in the fit-and-remove strategy, which is as described in detail as follows.

The first step is to recognize whether or not a selected promising combined group is mismatched, if yes, remove it. As shown in Fig. 4 (a), based on the keypoint correspondences of selected promising combined groups, we estimate a transformation matrix  $T_1$  or  $T_2$  using Singular Value Decomposition (SVD). To recognize mismatched groups  $\tilde{g}_i^{mismatch}$ , the estimated transformation matrix is used to rotate and translate one of the keypoint sets to the other. Thus, any correspondence, whose spatial coordinate distance from the keypoint pair is less than a threshold  $\tau_0$  (in our experiments  $\tau_0$  is three times of resolution), is treated as a match. For 3D point cloud registration, at least four pairs of matches can be used for estimating the rigid transformation. Therefore, a selected group, which has at least four pairs of matches, is considered as a matched group, otherwise, it does not match. A selected promising combined group is denoted as  $\tilde{g}_i = \{(p_i^u, q_i^v) | u \in \{1, \dots, U_i\}, v \in \{1, \dots, V_i\}\}$ ,  $N_i$  represents the number of correspondences in the combined group. Thus, the mismatched groups are formulated as follows:

$$\tilde{g}_i^{mismatch} = \tilde{g}_i, \quad \text{if } \sum \mathbf{I}(\|R_i p_i^u + T_i - q_i^v\| < \tau_0) < p + 2, \quad (4)$$

$$u \in \{1, \dots, U_i\}, v \in \{1, \dots, V_i\}.$$

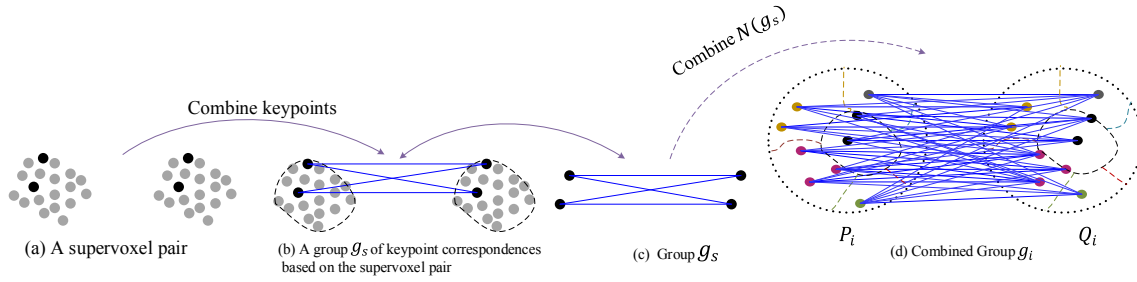
where  $\mathbf{I}$  represents an indicator function that  $\mathbf{I} = 1$  if  $\|R_i p_i^u + T_i - q_i^v\| < \tau_0$ , otherwise,  $\mathbf{I} = 0$ .  $R_i$  and  $T_i$  represent the rotation matrix and translation vector, respectively, of the corresponding combined group  $\tilde{g}_i$ , and are estimated by the correspondences of group  $\tilde{g}_i$ .

The second step is to remove redundant combined groups. As shown in Fig. 4 (b), the redundant combined group  $\tilde{g}_i^{neighbor}$ , i.e., neighboring combined group partially overlaps the selected promising group. Therefore, to determine whether a combined group is redundant, it is necessary to detect only whether a combined group  $\tilde{g}_j$  overlaps the selected combined group  $\tilde{g}_i$ . The redundant combined groups are formulated as follows:

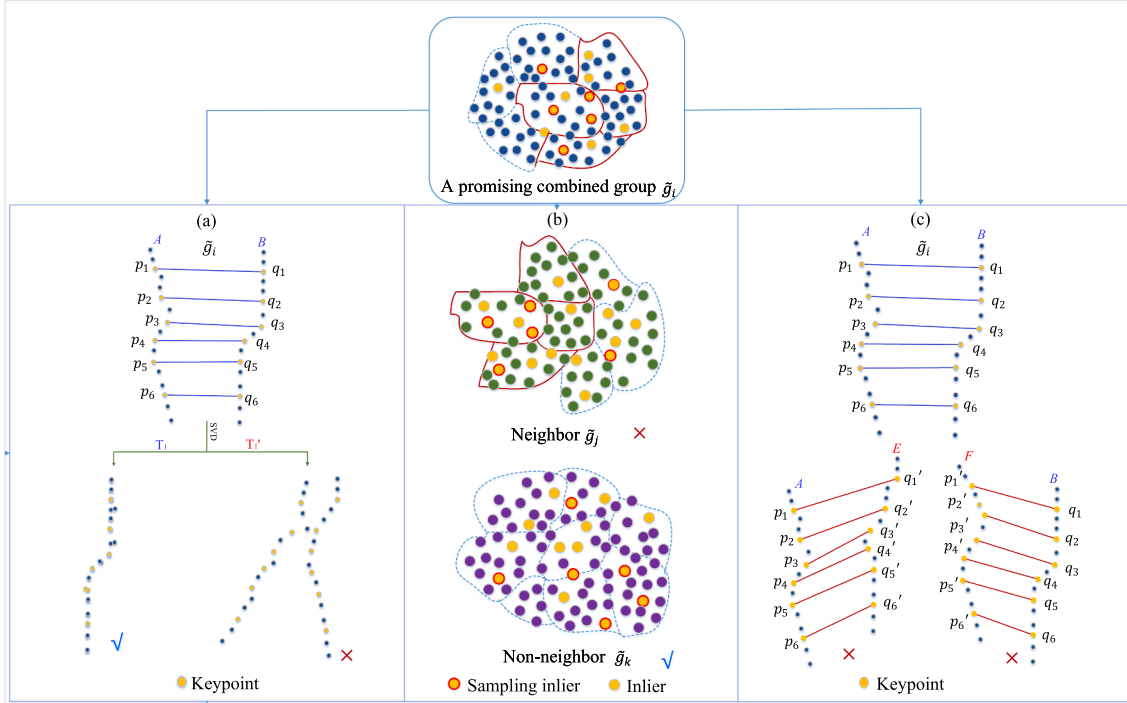
$$\tilde{g}_i^{neighbor} = \tilde{g}_j, \quad \text{if } \mathbf{SamS}(\tilde{g}_i) \cap \mathbf{InS}(\tilde{g}_j) \neq \emptyset \quad (5)$$

where  $\mathbf{SamS}(\tilde{g}_i) = \{x_i^u\}_{u=1}^{p+2}$  represents the top  $p + 2$  sorted inliers in the select combined group,  $u$  is the index of sorted inliers in the  $i$ -th combined group,  $p$  denotes the minimum size of matches. For 3D rigid transformation, four keypoint correspondences ( $p = 4$ ) can be used as the minimum size of the sampling subset to estimate a unique rotated and translated transformation matrix.  $\mathbf{InS}(\tilde{g}_j)$ , generated by a threshold ratio  $\lambda$ , is the inlier set of  $\tilde{g}_j$ .  $\mathbf{SamS}(\tilde{g}_i) \cap \mathbf{InS}(\tilde{g}_j)$  are used to decide if the sampled subset corresponding to the combined group  $\tilde{g}_i$  contains any inliers of  $\tilde{g}_j$ . Thus, the group  $\tilde{g}_i^{neighbor}$  is removed, and the non-neighboring combined groups are preserved for the next selection.





**Fig. 3.** The combining process of keypoint correspondences based on first-order neighborhood of a supervoxel. (a) Given a pair of supervoxels that each one has two keypoints; (b) A group of keypoint correspondences are combined in the supervoxel pair; (c) Group  $G_s$ ; (d) The group  $G_s$  combines with its adjacent groups  $N(G_s)$ .



**Fig. 4.** The improved ‘fit-and-remove’ strategy to select promising groups mainly includes three steps: (a) Removing the mismatched combined groups based on local transformation, (b) removing the redundant combined groups, (c) removing the non-promising combined groups. A promising combined group  $\tilde{g}_i$ , consists of two keypoint sets between the source and target point clouds. The blue line represents that the two keypoints are match, the red line represents that the two keypoints are mismatch.  $\checkmark$  indicates that the corresponding group will be selected.  $\times$  indicates that the corresponding group will be removed.

Finally, there is a high probability that the remaining non-promising combined groups  $\tilde{g}_i^{remain}$ , corresponding to the selected combined group, are mismatched. Thereby, to improve efficiency, we do not consider them as a matching combined group. As shown in Fig. 4 (c), any combined groups that intersect with one end of the selected promising combined group will be removed. For selected promising combined groups, we denote the corresponding keypoints as  $P_i = \{p_u^i\}_{u=1}^{N_i}$  and  $Q_k = \{q_w^k\}_{w=1}^{N_k}$ . The remaining non-promising combined groups are represented as follows:

$$\tilde{g}_i^{remain} = \tilde{g}_k, \quad \text{if } (P_i \cap P_k) \cup (Q_i \cap Q_k) \neq \emptyset \quad (6)$$

In summary, for a selected promising combined group, the corresponding groups that must be removed are formulated as follows:

$$\tilde{g}_i^r = \tilde{g}_i^{mismatch} \cup \tilde{g}_i^{neighbor} \cup \tilde{g}_i^{remain} \quad (7)$$

Consequently, the coarse extraction of inliers is achieved by Supervoxel Guidance (SG) method, and summarized as Algorithm 1.

**Algorithm 1.** The improved ‘fit-and-remove’ strategy for selecting promising combined groups

---

**Input:** Initial keypoints correspondences  $G$ , the inlier scale ratio  $\lambda$ .

- 1: Generate combined groups  $\tilde{G}$  by Eq. 2.
- 2: **for**  $i = 1$  to  $N$  **do**//  $N$  is the size of  $\tilde{G}$
- 3:    $m = \max_{i \in \{1, \dots, N-i\}} \{m_{\tilde{g}_i}\}$ ; //  $m_{\tilde{g}_i}$  represents the number of inlier in group  $\tilde{g}_i$
- 4:   Select the group  $\tilde{g}_i$  with  $m$  inliers;
- 5:   Generate the groups  $\tilde{g}_i^f$  by Eq. 7.
- 6:    $\tilde{G} \leftarrow \tilde{G} \setminus \{\tilde{g}_i^r\}$ ;
- 7: **end for**

**Output:** Promising combined groups  $G = \{\tilde{g}_k\}_{k=1}^K$ .

---

3.3. Grouping non-cooperative game optimization

Non-cooperative game methods (Bulò and Bomze, 2011; Torsello et al., 2012) have been proposed for rejecting false correspondences.

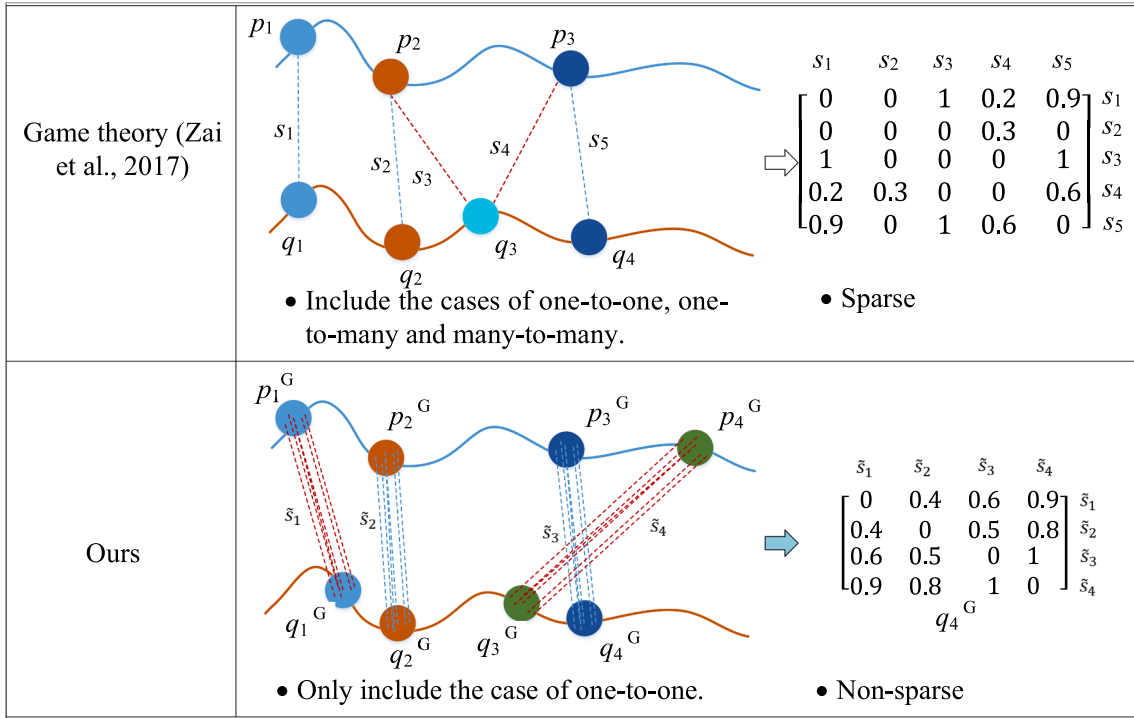


Fig. 5. The demonstration of constructing differences between Zai et al. (2017) and our proposed SGGT. The blue and red dotted lines denote correct and wrong combined groups, respectively.

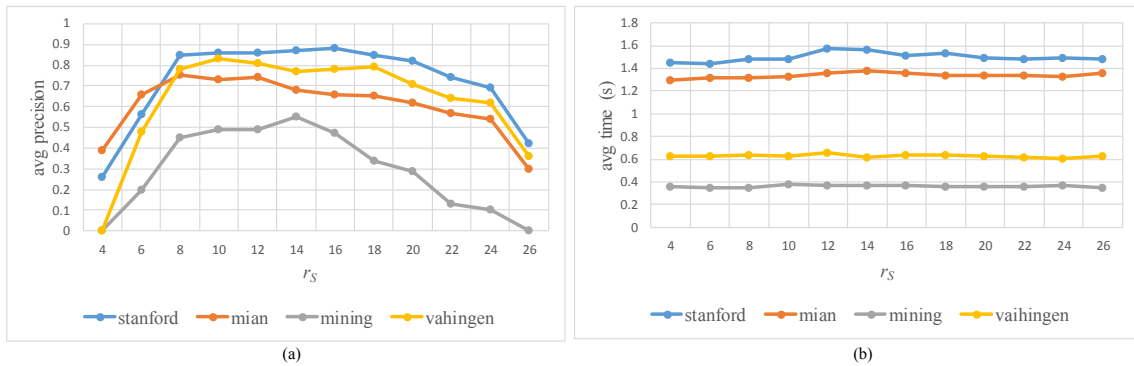


Fig. 6. Test of varying supervoxel resolutions on four kinds of models. (a) Demonstration of average precision; (b) Demonstration of average running time.

However, non-cooperative game methods are sensitive to the high proportion of outliers (See, Fig. 11), especially when the outlier rate is greater than 80%. Therefore, if the game theory is applied directly to the inlier extraction with high outliers, it will be difficult to satisfy the estimating requirements of three-dimensional rotation and translation transformation. Based on the above fit-and-remove selection strategy, we consider the relationship between combined groups to further improve the accuracy of the inlier extraction. Note that, the construction of combined groups avoids the one-to-many situation, making the non-cooperative game simpler. A grouping non-cooperative game method is proposed to further reject the mismatched combined groups.

Associated with each selected combined group, the isolation of mutually compatible colorredwith the outliers is measured by calculating payoffs. To extract the matched grouping subset which includes the largest number of correct combined groups, we first calculate the similarities (denoted as  $S^c$ ) between each of the two combined groups. Then, we extract the inlier grouping subset by imposing some constraints. Similar to the game construction method of Zai et al. (Zai et al., 2017), we summarize the game as a triplet  $U = \{\tilde{G}, S^c, A\}$ , where  $\tilde{G}$  is the player set (i.e., the selected combined groups set);  $S^c$  is the pure-

strategy set, and  $A$  is the combined payoff function. Instead of independent correspondences, the compatible combined groups are considered the groups that contain inliers. The non-cooperative game optimization is summarized as Algorithm 2.

### 3.3.1. Measure the similarity of combined groups

Based on the construction of a combined group, each combined group contains two groups of keypoints from source and target point clouds, respectively. However, by the Euclidean distance of FPFH features, the similarity of two keypoints in a keypoint correspondence is measured. To consider the relationship between combined groups, first, it is necessary to accurately measure the similarity of two groups of keypoints in each combined group by combining the corresponding correspondences. To enhance the difference in the similarity among the combined groups, using the size of inliers in a combined group  $\tilde{g}_i$ , we increase the difference in the similarities, the similarity in a combined group calculated as follows:

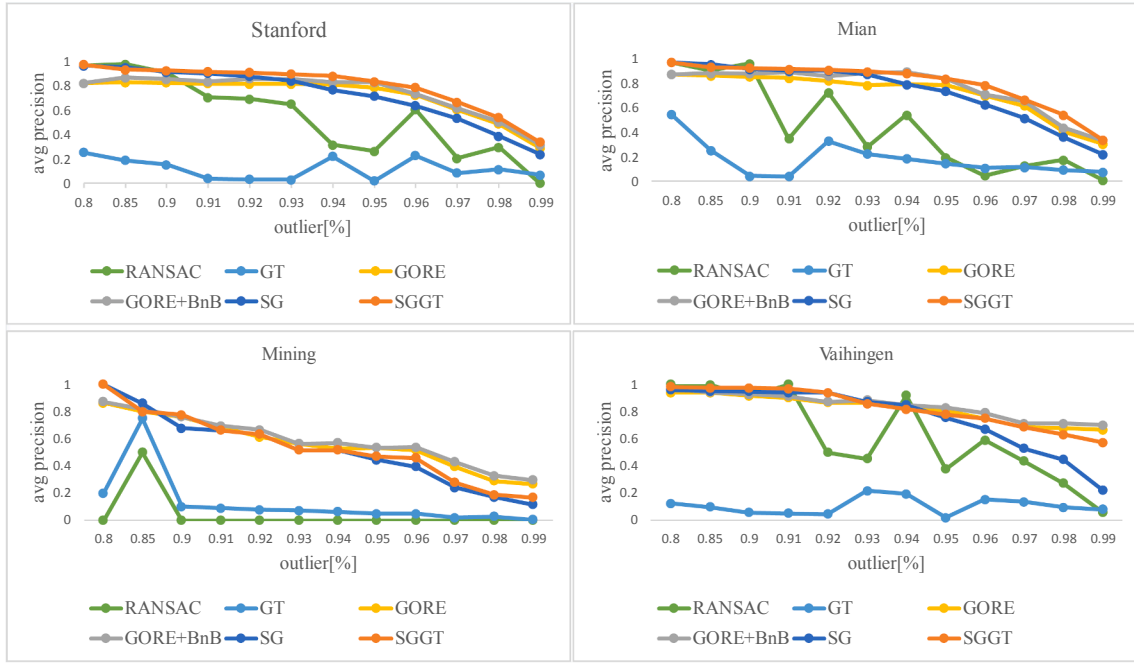


Fig. 7. Extraction precision of inliers.

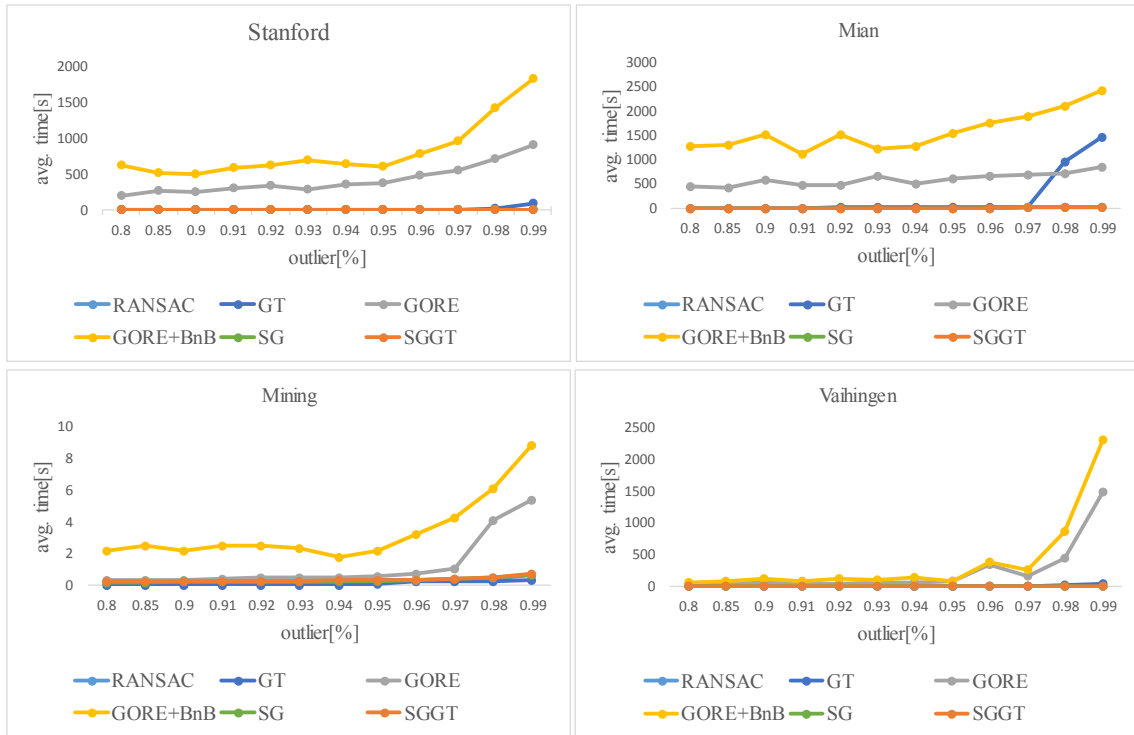


Fig. 8. CPU computational time of inlier extraction.

$$\tilde{s}_i = \frac{1}{N_i} \frac{\max_k \{N_k\} - N_i}{\left( \max_k \{N_k\} - \min_k \{N_k\} \right)} \bar{s}_i \quad (8)$$

where  $\tilde{s}_i \in \tilde{\mathcal{S}}$  represents the similarity metric of a combined group and  $N_i \in \{N_k | k = 1, 2, \dots, K_2\}$  represents the number of inliers in a combined group  $\tilde{g}_i$ . The average similarity  $\bar{s}_i$  is calculated as follows:

$$\bar{s}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} s_j \quad (9)$$

where  $s_j$  is a similarity metric of the FPFH features with respect to two keypoints.

### 3.3.2. Build a payoff matrix of combined groups

Considering that our problem conforms to a rigid transformation, to discard non-matching combined groups, it is natural to impose geometric constraints on those combined groups. Given two arbitrarily





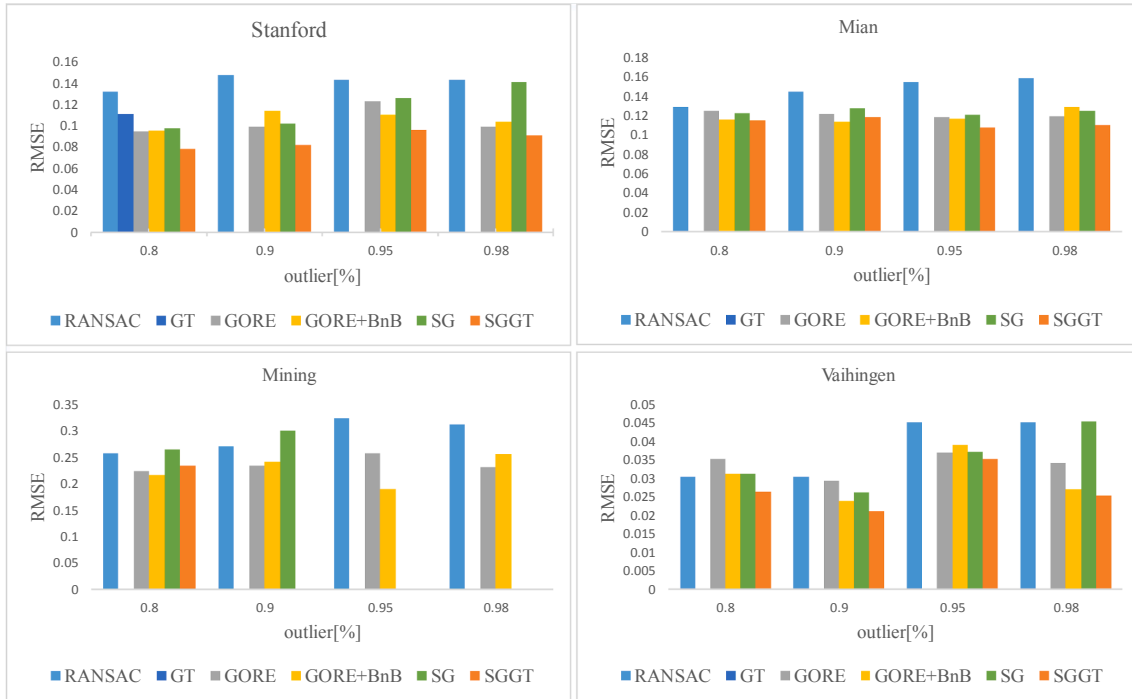


Fig. 9. RMSE is measured based on extracting inliers under the selected four different outlier ratios, i.e.  $\eta = \{0.8, 0.90, 0.95, 0.98\}$ .

follows:

$$\delta_F(\mathbf{x}^{(n)}) = \max_i(\mathbf{A}\mathbf{x})_i - \mathbf{x}\mathbf{A}\mathbf{x} \quad (14)$$

The Nash error is calculated according to the following:

$$e_n = \sum_{i=1}^N \left( \max_i(\mathbf{A}\mathbf{x})_i^{(n)} - (\mathbf{x}\mathbf{A}\mathbf{x})^{(n)} \right) \quad (15)$$

where  $n$  represents the  $n$ -th iteration.

#### Algorithm 2. Non-cooperative game optimization

**Input:** Combined groups  $\tilde{G}$ , initial probability  $\mathbf{x}^0$ , the threshold of nash error  $\varepsilon$ , The maximum number of iterations  $I$ .

- 1: Calculate the similarity of each combined group  $\tilde{G}$ .
- 2: Build payoff matrix  $A$  by Eq. 11.
- 3: **for**  $i = 1$  to  $I$  **do**
- 4: Play non-cooperative game by Eq. 12.
- 5: Calculate the Nash error  $e_n$  by Eq. 15.
- 6: **if** ( $e_n < \varepsilon$ ) **then**
- 7: break;
- 8: **end if**
- 9: **end for**

**Output:** the probability of updated strategy.

With the description of the proposed SGGT, we summarize the computational complexity of Algorithm 1 and Algorithm 2. Since the supervoxel (Lin et al., 2018) preserve well the boundaries of the point cloud with  $O(N_S \log(M/K_S))$ , where  $N_S$  is the total number of neighbor points in a supervoxel,  $M$  is the number of a point cloud,  $K_S$  is the number of supervoxels, and  $N < K_S$ . The computational complexity of Algorithm 1 is approximately proportional to  $O(K_S)$ . It cost the majority of computational time in the step of supervoxel segmentation. The evolution process of Algorithm 2 is characterized by a linear complexity (Torsello et al., 2012) and is running in the remaining candidate inliers. Therefore, the computational complexity of Algorithm 2 is  $O(I)$ , where  $I$  is the maximum number of iterations. Therefore, the total complexity of the proposed algorithm approximately amounts to  $O(K_S) + O(I)$ .

## 4. Experiments

To demonstrate the precision and efficiency of the proposed SGGT in point cloud registration, several experiments were conducted in C++ and evaluated in Matlab2018b, and on a PC with Windows 7, Intel Core(TM) i5-4460 3.2 GHz CPU and 16.0 GB RAM.

### 4.1. Experimental setup

In the evaluation experiments, we first sampled keypoints using a hash sampling method (Mellado et al., 2014). Hence, the keypoint correspondences were built at varying ratios of the outliers. The corresponding FPFH (Rusu et al., 2009) descriptors were implemented based on the Point Cloud Library (PCL)<sup>1</sup>. The scale and resolution of a point cloud affect the efficiency of supervoxel segmentation, and the proposed SGGT is based on supervoxel segmentation. The scale of a point cloud affects the efficiency of the algorithm; more specifically, the calculation consumes more time with the growth of the point cloud scale. Therefore, it is necessary to down-sample the point cloud first. For a large-scale point cloud, such as a Terrestrial Laser Scanning (TLS) point cloud, we down-sampled the source and target point clouds to 20% of their original resolution using a voxelized grid approach of the PCL. The input key point correspondences were obtained by combining every two key points from point clouds,  $\mathbf{P}$  and  $\mathbf{Q}$ .

We recorded the following measures for each approach:

- $\eta$ : The outliers ratio, i.e. the ratio of the number of mismatches to the number of all correspondences.
- $H$ : Initial correspondences, which are acquired by arbitrarily combining the keypoints between a source point cloud and a target point cloud.
- $H'$ : Final remaining correspondences (including matches and mismatches) which are acquired by the related methods, and  $H' \subset H$ .
- $|H'|$ : Size of the final remaining correspondences.
- $|J|$ : The final number of matches.

<sup>1</sup> <http://pointclouds.org/>.

**Table 2**  
Results of inlier extraction when  $\eta = \{0.95, 0.96, 0.97, 0.98, 0.99\}$ .

Outlier ratio	Model	input	Pipeline						
				RANSAC	GT	GORE	GORE + BnB	SG	SGGT
0.95	Stanford	10379	$ H $	–	1471	533	522	215	174
	$ P  = 40256$ $ Q  = 40097$	519	$ I $	15	7	337	330	180	141
	Mian	5999	$ H $	–	0	297	276	159	119
	$ P  = 69007$ $ Q  = 68681$	300	$ I $	2	0	199	198	113	102
	Mining	1379	$ H $	–	839	69	69	88	68
	$ P  = 29764$ $ Q  = 37761$	69	$ I $	2	42	36	36	39	39
	Vaihingen	6659	$ H $	–	1176	324	313	216	201
	$ P  = 138425$ $ Q  = 64255$	333	$ I $	9	35	230	229	174	169
	0.96	Stanford	12455	$ H $	–	617	409	523	224
$ P  = 40256$ $ Q  = 40098$		519	$ I $	7	225	327	331	176	120
Mian		7499	$ H $	–	0	296	276	166	98
$ P  = 69007$ $ Q  = 68682$		300	$ I $	4	0	198	199	119	84
Mining		1724	$ H $	–	1188	69	69	38	30
$ P  = 29764$ $ Q  = 37762$		69	$ I $	2	48	36	36	19	18
Vaihingen		8324	$ H $	–	1110	359	311	212	171
$ P  = 138425$ $ Q  = 64255$		333	$ I $	9	64	186	192	139	125
0.97		Stanford	17299	$ H $	–	652	494	203	259
	$ P  = 40256$ $ Q  = 40098$	519	$ I $	9	192	128	129	163	89
	Mian	9999	$ H $	–	0	291	276	169	46
	$ P  = 69007$ $ Q  = 68682$	300	$ I $	3	0	96	100	96	35
	Mining	2299	$ H $	–	1178	69	69	42	19
	$ P  = 29764$ $ Q  = 37762$	69	$ I $	2	21	16	16	8	7
	Vaihingen	11099	$ H $	–	714	326	360	228	143
	$ P  = 138425$ $ Q  = 64255$	333	$ I $	4	58	127	156	106	89
	0.98	Stanford	25949	$ H $	–	886	414	195	299
$ P  = 40256$ $ Q  = 40098$		519	$ I $	7	183	28	30	135	43
Mian		14999	$ H $	–	0	297	277	200	32
$ P  = 69007$ $ Q  = 68682$		300	$ I $	4	0	99	99	62	17
Mining		3449	$ H $	–	344	69	69	69	24
$ P  = 29764$ $ Q  = 37762$		69	$ I $	3	16	16	16	13	10
Vaihingen		16649	$ H $	–	908	327	307	235	73
$ P  = 138425$ $ Q  = 64255$		333	$ I $	9	18	132	137	103	50
0.99		Stanford	51899	$ H $	–	1311	577	203	435
	$ P  = 40256$ $ Q  = 40098$	519	$ I $	9	169	140	29	117	16
	Mian	29999	$ H $	–	0	297	278	299	36
	$ P  = 69007$ $ Q  = 68682$	300	$ I $	5	0	99	99	80	14
	Mining	6899	$ H $	–	669	72	70	154	16
	$ P  = 29764$ $ Q  = 37762$	69	$ I $	3	11	16	15	23	4
	Vaihingen	33299	$ H $	–	1082	334	315	244	56
	$ P  = 138425$ $ Q  = 64255$	333	$ I $	7	10	133	129	61	34

- **Precision:** Precision is defined as the ratio of the number  $|I|$  of matches to the size  $|H|$  of the remaining correspondences.
- $\bar{r}$ : Point cloud resolution is acquired by the mean of the distances between the points in the source point cloud and each of their closest points in the target point cloud.

4.2. Qualitative evaluation

Similar to the related work of Bustos and Chin (Bustos and Chin, 2017), we also validated the proposed SGGT for four different sources: (1) the Stanford 3D Scanning Repository (Curless and Levoy, 1996) (including bunny, armadillo, dragon and buddha), (2) Mian’s dataset<sup>2</sup> (including t-rex, parasaur, chef and chicken), (3) partially overlapping dataset, and (4) laser scans of underground mines (specifically mine-a, mine-b) (Bustos and Chin, 2017). Pairwise point clouds of one object per dataset are shown in Fig. 10.

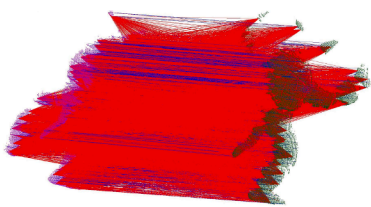
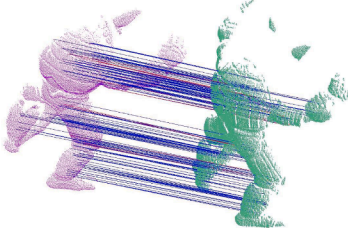
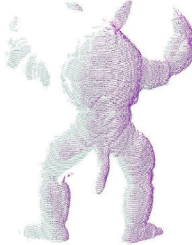
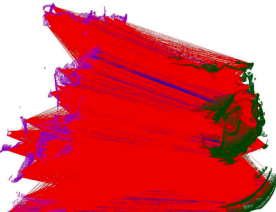
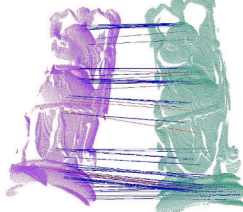
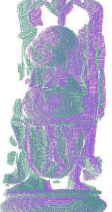
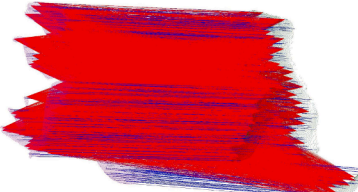
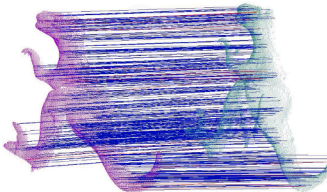

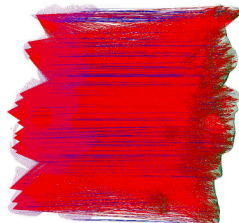
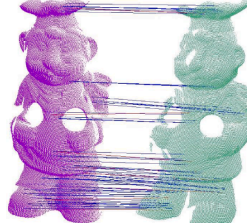

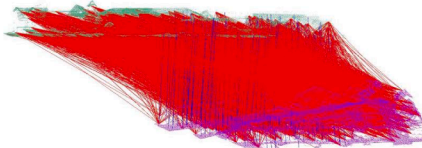
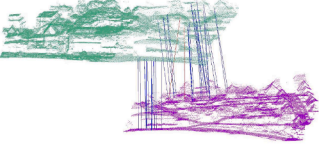
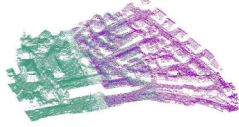
To evaluate the performance of the proposed SGGT, the testing experiments included two parts: (1) Analysis of varying supervoxel resolution, (2) analysis of varying outliers. Additionally, the methods

for comparison included the following five approaches: RANSAC (Fischler and Bolles, 1981), GT (Zai et al., 2017), GORE (Bustos and Chin, 2017), GORE + BnB (Bustos and Chin, 2017) and SGGT.

**Analysis of varying supervoxel resolutions.** When supervoxel segmentation is performed on a point cloud, the resolution of the supervoxels directly affects the number of supervoxel facets. The larger the supervoxel resolution, the smaller the number of supervoxels. However, the number of combined group pairs affects the extraction of matching pairs. Therefore, supervoxel resolution plays an important role in our proposed SGGT. It is necessary to test the influence of varying supervoxel resolution.

Fig. 6 shows the precision of extraction and computational time of the proposed SGGT with different supervoxel resolutions. We constructed a series of different supervoxel resolutions, denoted as  $\mathbb{R}^s = \{2i \times r_0\}_{i=2}^{13}$ , where  $r_0$  represents the resolution of raw point clouds. As shown in Fig. 6 (a), the average precision of each group of models remains relatively stable with the supervoxel resolution  $r_s$  in the range of  $8r_0$  to  $24r_0$ , where the outlier ratio  $\eta = 0.95$ . If the supervoxel resolution is outside that range, the size of the inliers in certain combined groups is smaller than  $p + 2$ . As a result, the corresponding group is removed. In such a case, sampling an all-inlier subset from each group is difficult. The quality of extracted inliers generated by the sampled

<sup>2</sup> <http://staffhome.ecm.uwa.edu.au/00053650/3Dmodeling.html>.

	Input ( $\eta : 0.95$ )	SGT ( $H'$ )	Registration using estimated T
Armadillo	 Match/Mismatch: 755/14344	 Match/Mismatch: 182/19 time: 0.79s	 RMSE: 0.1253
Buddha	 Match/Mismatch: 363/6896	 Match/Mismatch: 94/11 time: 1.38s	 RMSE: 0.1358
T-rex	 Match/Mismatch: 726/13793	 Match/Mismatch: 467/58 time: 1.43s	 RMSE: 0.1275
Chef	 Match/Mismatch: 300/5699	 Match/Mismatch: 70/11 time: 1.38s	 RMSE: 0.1278
Vaihingen-b	 Match/Mismatch: 577/10962	 Match/Mismatch: 40/2 time: 0.54s	 RMSE: 0.03778

**Fig. 10.** Qualitative results of SGGT for 6 DoF rigid registration with outlier ratio  $\eta = 0.95$ . Column 1: Input correspondences (true inliers are represented by blue lines, and true outliers by red lines). Column 2: Data remaining after SGGT. Column 3: Registration using approximate solution  $\tilde{T}$  produced by SGGT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

subset affects the extracting precision of SGGT. As shown in Fig. 6 (b), with the varying supervoxel resolutions  $r_s$ , we performed a histogram statistics analysis and calculated the average running time for each model. It is clearly shown that, the average running time for each model is close to a constant and less than 1.6 seconds. Therefore, although the proposed SGGT has a certain range where it adapts to the resolution of supervoxel segmentation, that range hardly affects computational efficiency.

**Analysis of varying outlier ratios.** For keypoint-based registration of point clouds, one of the most important techniques is the ability to accurately extract matches from a high percentage of outliers.

Therefore, five methods are designed to validate the performance of the proposed SGGT at varying outlier ratios, where we set  $\eta = \{0.8, 0.85, 0.90, \dots, 0.99\}$ .

Fig. 7 shows the variation curve for the extraction accuracy of the five compared methods at varying outlier ratios. The proposed SGGT although highly precise, decreases gradually with the increasing outlier ratios. The precision extracted by the GT method is low and unstable, demonstrating that it is difficult for GT to extract inliers from correspondences with a high proportion of outliers. The proposed SGGT achieves higher precision than SG. The precision of the proposed SGGT is approximately on a par with the methods of GORE and GORE + BnB.

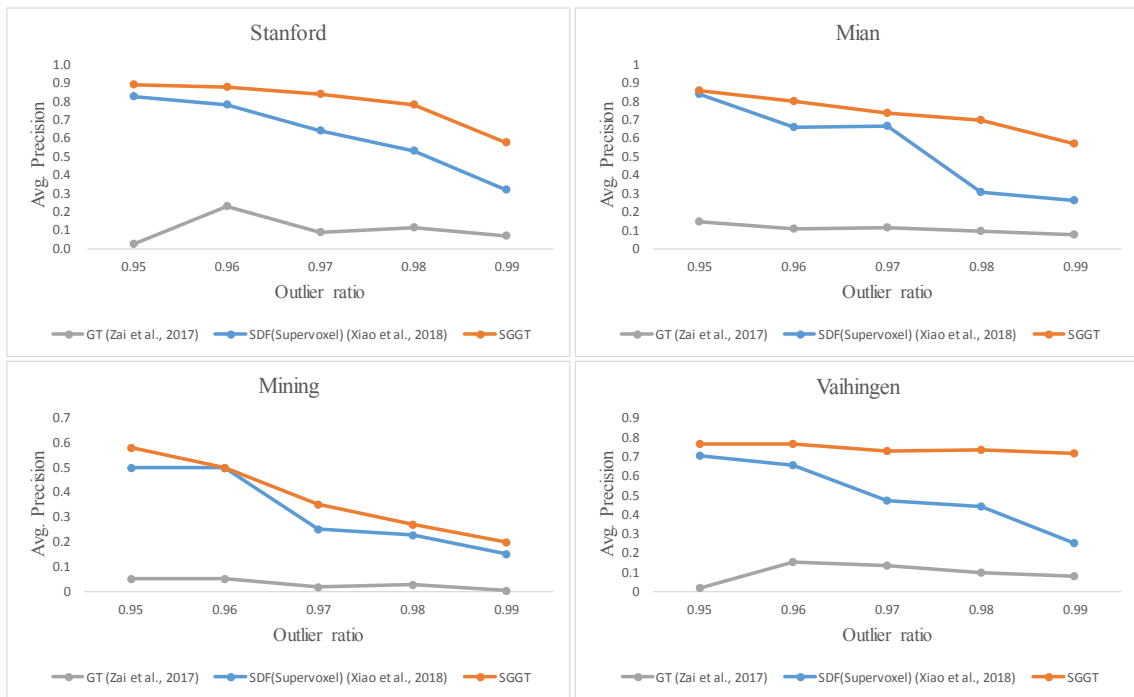


Fig. 11. Qualitative results from the methods of GT (Zai et al., 2017) and SDF (supervoxel) (Xiao et al., 2016).

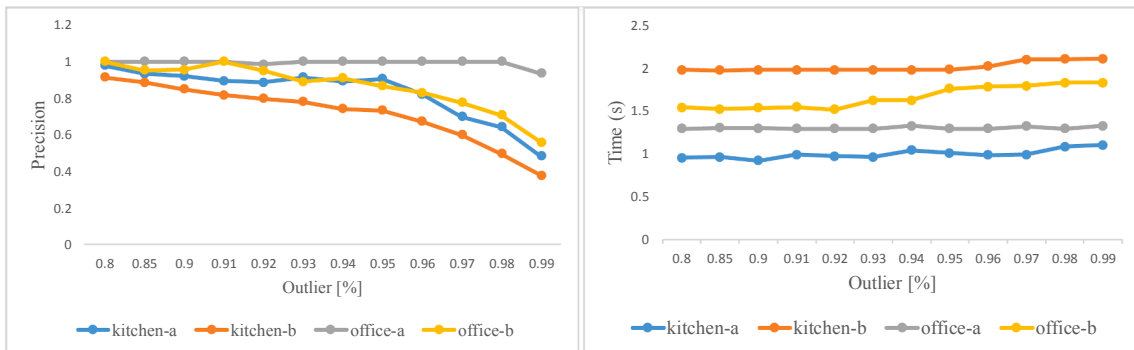


Fig. 12. The varying cures of precision and CPU computational time by the proposed SGGT used in depth maps.

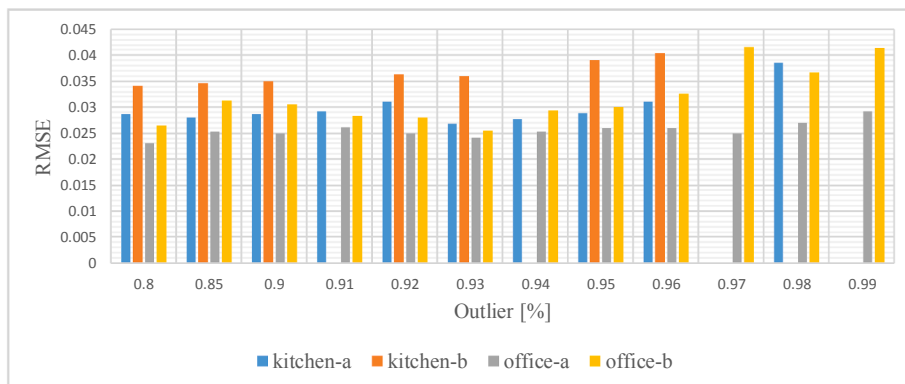
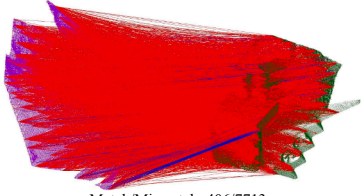
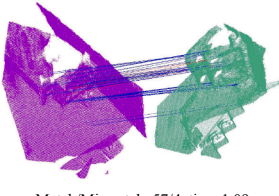
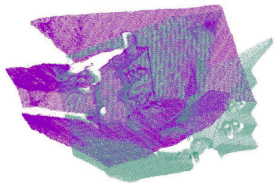
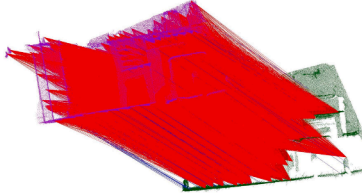
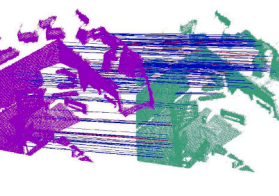
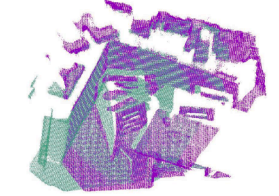
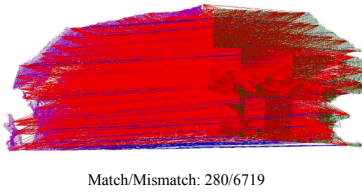
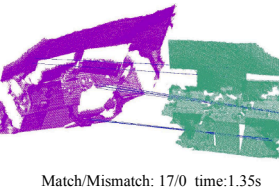
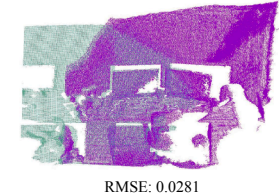
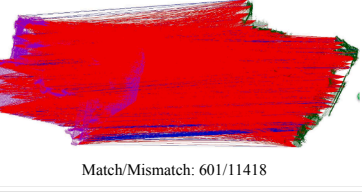
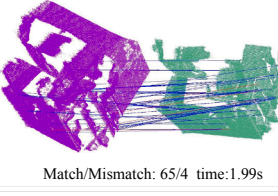
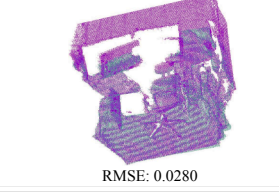


Fig. 13. The RMSE of depth maps registration.

Fig. 8 shows the CPU computational time corresponding to the experimental results in Fig. 7. The proposed SG and SGGT requires much less CPU time and does not significantly change with varied outlier ratios (For RANSAC, the number of iterations is limited to 5,000). The CPU computational time for GORE or GORE + BnB increases sharply

with the increasing outlier ratios. When the ratio of the outliers is less than 0.97, the CPU computational time by the GT method tends to zero seconds. However, when the outlier ratios are greater than or equal to 0.97, extraction times for the models such as Mian and Vaihingen increase greatly. The increase in the number of matching pairs leads to an



	Input	Remaining data $H'$	Registration results
Kitchen-a	 Match/Mismatch: 406/7713	 Match/Mismatch: 57/4 time:1.09s	 RMSE: 0.0258
Kitchen-b	 Match/Mismatch: 232/4407	 Match/Mismatch: 245/31 time:1.98s	 RMSE: 0.0312
Office-a	 Match/Mismatch: 280/6719	 Match/Mismatch: 17/0 time:1.35s	 RMSE: 0.0281
Office-b	 Match/Mismatch: 601/11418	 Match/Mismatch: 65/4 time:1.99s	 RMSE: 0.0280

**Fig. 14.** Qualitative results of SGGT for 6 DoF rigid registration with outlier ratio  $\eta = 0.95$ . Column 1: Input correspondences (true inliers are represented by blue lines, and true outliers by red lines). Column 2: Data remaining after SGGT. Column 3: Registration using approximate solution  $\tilde{T}$  produced by SGGT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

increase in the dimensions of the payoff matrix, thereby slowing down the solution of the payoff matrix. As the proportion of outliers increases, the CPU computational time for the proposed SGGT tends to zero, and there is no significant change. Thus, it is clear that the proposed SGGT is superior in efficiency and does not change as the outlier ratios increase, especially, when the size of the correspondences is larger than 5,000. For a more intuitive observation, we list the CPU computational time for all the compared methods in Table 1. It is clear that the proposed SGGT is much faster than the state-of-the-art methods.

Fig. 9 shows the performance of the point cloud registration, where the RMSE estimated by the result of inlier extraction is measured. Here, each RMSE is measured at a distance shorter than  $5\bar{r}$  between the point  $p_i$  in the source point cloud  $\mathbf{P}$  and its closest point  $q_j$  in the target point cloud  $\mathbf{Q}$ . RMSE is formulated as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \min_{j \in \{1, \dots, M\}} \left( dis(p_i, q_j) \right)}$$

$$min(dis(p_i, q_j)) < \tau \quad (16)$$

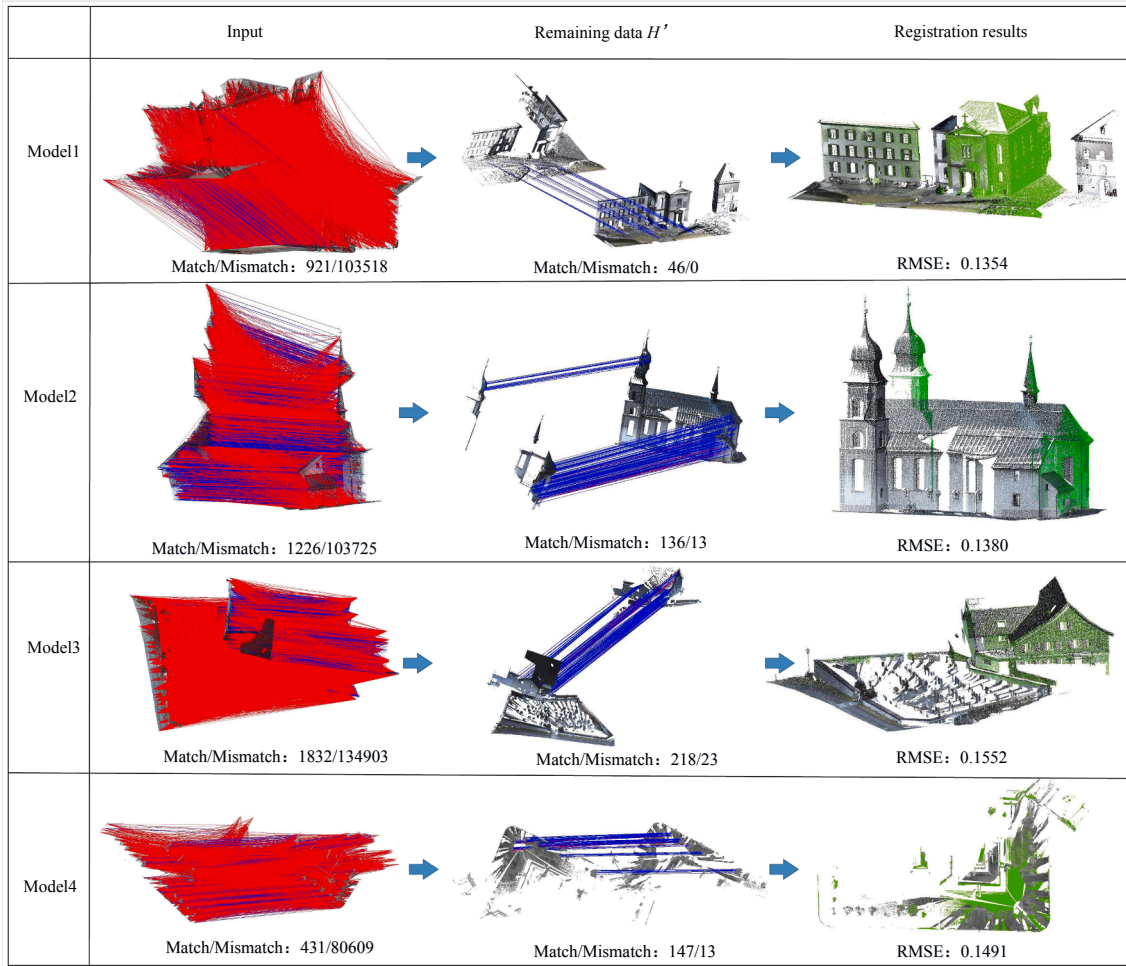
where  $dis(*, *)$  represents the Euclidean distance, and  $\tau$  is the threshold of distance,  $\tau = 5\bar{r}$ .

We mapped the RMSE histogram of the failed group with a blank. Hence, as observed that, except for the Stanford group, the registration results with the GT all failed under the outlier ratio  $\eta = 0.8$ . A plot of the RMSE histogram of selected groups, except “Mining” models from the mid row, shows that our method is effective. Because of varying density and many planar structures in the “Mining” models, corresponding feature appearances are insignificant. In summary, as seen

from the comprehensive performance of the precision; of the CPU computational time and RMSE evaluation of different methods, the proposed SGGT is effective and highly efficient.

In Table 2, the summary of median values for all eight groups of models demonstrates more clearly the superiority of the proposed SGGT with high outlier ratios  $\eta = \{0.95, 0.96, 0.97, 0.98, 0.99\}$ . The number  $|H'|$  of remaining correspondences of RANSAC obtains a rare pair of remaining points at a threshold ( $>2\bar{r}_0$ ) and is unstable mainly due to the randomness of the algorithm. The consensus sizes  $|I|$  with five different outlier ratios by GT are all too small relative to the size of the remaining correspondences,  $|H'|$ . Therefore, for a high outlier ratio, it is clear that GT is ineffective. The methods of GORE and GORE + BnB acquire more remaining correspondences. However, the ratio of inliers acquired by our SGGT is significantly higher than that acquired by the GORE and GORE + BnB methods. The inlier ratio is higher than for other methods. Besides, the sizes of initial candidate correspondences are all greater than 10,000 except for the “Minging” models, Combined with the experimental results shown in Fig. 8 and Table 1, it is seen that, when the size of keypoint correspondences reaches 10,000, our proposed SGGT is at least 100 times faster than GORE. Several selected registration results are shown in Fig. 10.

To examine the real impact of our proposed SGGT, we designed several comparative tests, including the following: GT, SDF (Supervoxel) and SGGT. GT is achieved in accordance with Zai et al. (Zai et al., 2017). SDF (supervoxel) from Xiao et al. (Xiao et al., 2016) replaces only superpixels with supervoxels. The extraction precision of the three methods on four point cloud pairs with different outlier ratios is shown in Fig. 11. GT achieves accuracy lower than SGGT and SDF



**Fig. 15.** Qualitative results of SGGT for 6 DoF rigid registration. Column 1: Input correspondences (true inliers are represented by blue lines, and true outliers by red lines). Column 2: Data remaining after SGGT. Column 3: Registration using approximate solution  $\tilde{T}$  produced by SGGT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

This table shows the information of TLS point cloud registration including input, intermediate sampling, inliers extraction time and RMSE after registration.

Model		Number of points	Resolution (m)	Number of points (Downsampled)	Overlap (%)	Number of sampling keypoints	RMSE (m)	Time(s)
Model1	P	824,509	0.0149	223,657	38	5,289	0.135363	22.861
	Q	13,986,927	0.0151	393,952		5,255		
Model2	P	2,204,475	0.0125	250,129	36	5,476	0.138014	14.998
	Q	925,352	0.0127	105,816		5,072		
Model3	P	5,923,555	0.0131	230,329	91	5,203	0.155219	17.595
	Q	112,007	0.0446	39,344		8,574		
Model4	P	5,800,968	0.0532	528,013	82	4,553	0.149116	39.929
	Q	3,885,373	0.0683	552,443		3,511		

(supervoxel). The proposed SGGT shows significant superiority over GT and SDF(supervoxel). Therefore, the supervoxel-guided method plays an important role in the proposed method. Thus, we evaluate (See next Section) only SGGT, which uses the supervoxel-guided method on the other testing dataset.

### 4.3. Test on the depth map

In this subsection, we evaluate the performance of the proposed SGGT on depth maps (Shotton et al., 2013) acquired from RGB-D cameras. As shown in Fig. 12, in the accuracy of extraction and CPU computational times vary as the outliers vary. As the outlier ratio increases, the precision by the proposed SGGT decreases. For CPU

computational times (all of which are less than two seconds), there are no significant changes. Therefore, the proposed SGGT for depth maps is extremely efficient. As shown in Fig. 13, some of the registering experiments fail when the outlier ratio is greater than 0.97.

Fig. 14 shows the qualitative results of our experiments for four groups of pairwise partially overlapping RGB-D data (kitchen-a, kitchen-b, Office-a, and Office-b) with an outlier ratio of 0.95. The number of keypoints extracted from each depth map, by the hash sampling method, is approximately 1,000. The input correspondences (matches and mismatches) are shown in the top row for kitchen-a. As seen in the middle row (kitchen-b and Office-a), the size of the input correspondences in each pairwise data is greater than or equal to 4,639. The accuracies of extraction by the proposed SGGT are all greater than

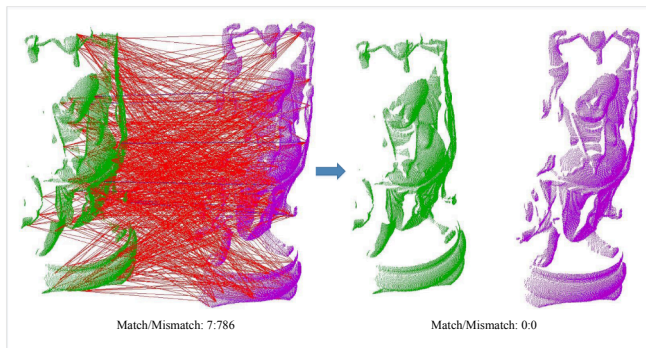


Fig. 16. Fail results by SGGT for 6 DoF rigid registration.

0.84, and even for 12,019 pairs of input correspondences, the computational CPU time is less than two seconds. The RMSE values of the four pairwise depth maps (Office-b), shown in the last row, are all less than 0.5. The experimental results verify the success of the corresponding depth maps registration. Therefore, it is clearly shown that the proposed SGGT is effective and efficient.

#### 4.4. Lidar point cloud registration

To further demonstrate the feasibility and effectiveness of the proposed SGGT, we tested the proposed SGGT on Terrestrial Laser Scanning (TLS) point clouds. Here, for each keypoint, we selected  $N$  best keypoint correspondences (sorted by the Euclidean distance of the FPFH descriptors), where  $N = 10$ . Data sets (Hackel et al., 2017) are recorded as Model1, Model2, and Model3 (see Fig. 15). The other data set (Model4) is from a square in Jiageng Chen Monument, Xiamen, China. These data sets consist of four point cloud pairs captured from different views. The corresponding information is described in Table 3. These data sets are all of high resolution ( $\bar{r} < 0.07m$ ) and large-scale. The keypoints were first extracted using a hash sample. The corresponding local features are described using FPFH. Then, candidate correspondences were obtained in ascending order of similarity. Hence, the true correspondences were obtained using the proposed SGGT. Finally, using the proposed SGGT, we aligned all the scans of tested models. The accuracy of the correspondences and the RMSE statistics are displayed in Table 3. The corresponding demos are shown in Fig. 15. The accuracies of extraction for the proposed SGGT are all greater than 0.84. The CPU computational time is less than 40 s, even for 19,985,583 pairs of input correspondences. The RMSE values of the four pairwise TLS point clouds are all less than 0.2 m. The experimental results verify the success of the corresponding TLS point clouds registration. Therefore, it is clearly shown that the proposed SGGT is effective and efficient.

## 5. Discussion

We introduce an inlier extraction method for point cloud registration. According to the algorithm introduction and experimental analysis, the proposed SGGT requires very few adjusted parameters (i.e., the resolution of supervoxels  $r_s$ , the resolution of point cloud  $\bar{r}$ , keypoint sampling number  $N_0$ ). We set the resolution of supervoxels to  $r_s = (5 \sim 10)\bar{r}$ , and set  $N_0 = (1\% \sim 5\%) * M$ , where  $M$ , for most pairwise point clouds, is the number of input point clouds.

Here we focus mainly on two points of discussion. First, supervoxel segmentation, involving 3D spatial homogeneity, provides powerful correspondence groups, resulting in a large reduction in the size of keypoint correspondences. Therefore, supervoxel segmentation helps achieve inlier extraction from 2D to 3D. Also, keypoint correspondences are grouped by supervoxel segmentation. The resolution of supervoxels directly determines the size of groups that play an important role in

grouping keypoint correspondences. Second, reject false grouping correspondences using non-cooperative match games. Because the above grouping and pruning strategy (see Fig. 5) avoids the cases of one-to-many and many-to-many, the proposed grouping game-theoretical method simplifies the massive combination. Via an infection and immunization dynamics equation, the population state evolves into an ESS and attains an optimal solution.

In summary, the superiority of our keypoint correspondences extraction framework can be attributed to at least two factors: (1) The proposed supervoxel-guided grouping keypoint correspondences are powerful and effective. (2) The grouping non-cooperative game-theoretic technique isolates mutually compatible correspondences from large outliers. Therefore, in terms of efficiency, the proposed SGGT outperforms the game-theoretic method by a large margin (See Fig. 11).

However, when the ratio of the inliers is too small, the number of matches in each combined group is fewer than  $p + 2$ . Therefore, SGGT fails to acquire the most promising combined group that does not extract the corresponding inliers. As shown in Fig. 16, the number of input matches/mismatches is 7/786. The numbers extracted by the proposed SGGT are 0/0. The proposed SGGT failed to extract matches, mainly because the matches are so few that the matches in each combined group are fewer than  $p + 2$  (at least 6). Thus, all the combined groups are removed using the ‘fit-and-remove’ selection strategy. Therefore, this failed case shows that the proposed SGGT has a certain condition that must be met for the number of matches in the candidate correspondences of the input.

## 6. Conclusion

We proposed an inlier extraction method (SGGT) for point cloud registration via supervoxel guidance and a grouping non-cooperative game optimization. Specially, we first presented a supervoxel-guided method, combining three-dimensional space attributes, to generate coarse promising groups of keypoint correspondences with greater compactness. Second, a grouping non-cooperative game was constructed to global evolve an optimal solution that generates purer matches. Tests in various data sets have shown that, first, our proposed SGGT outperforms the geometric model fitting of SDF in accuracy. Second, our proposed SGGT is more efficient, and, in accuracy, is on a par with state-of-the-art methods.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants U1605254, Grant 41471379. The authors would like to acknowledge the anonymous reviewers for their valuable comments.

## References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Machine Intell.* 34 (11), 2274–2282.
- Aiger, D., Mitra, N.J., Cohenor, D., 2008. 4-points congruent sets for robust pairwise surface registration. *ACM Trans. Graphics* 27 (3), 1–10.
- Aijazi, A.K., Checchin, P., Trassoudaine, L., 2014. Super-voxel based segmentation and classification of 3d urban landscapes with evaluation and comparison. *Springer Tracts Adv. Robot.* 92, 511–526.
- Albarelli, A., Bul, S.R., Torsello, A., Pelillo, M., 2009. Matching as a non-cooperative game. In: *IEEE International Conference on Computer Vision*, pp. 1319–1326.
- Albarelli, A., Rodola, E., Torsello, A., 2010. A game-theoretic approach to fine surface registration without initial motion estimation. In: *Computer Vision and Pattern*



- Recognition, pp. 430–437.
- Albarelli, A., Bergamasco, F., Torsello, A., 2013. A scale independent selection process for 3d object recognition in cluttered scenes. *Int. J. Comput. Vision* 102 (1–3), 129–145.
- Arun, K.S., 1987. Least squares fitting of two 3-d point sets. *Pattern Anal. Machine Intell.* 9 (5), 698–700.
- Bae, K.H., Lichti, D.D., 2008. A method for automated registration of unorganised point clouds. *ISPRS J. Photogramm. Remote Sens.* 63 (1), 36–54.
- Barath, D., Matas, J., 2018. Graph-cut ransac. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6733–6741.
- Besl, P.J., 1992. A method for registration 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14 (2), 193–200.
- Black, M.J., Rangarajan, A., 1996. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int. J. Comput. Vision* 19 (1), 57–91.
- Briales, J., Gonzalez-Jimenez, J., 2017. Convex global 3d registration with lagrangian duality. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4960–4969.
- Bulò, S.R., Bomze, I.M., 2011. Infection and immunization: A new class of evolutionary game dynamics. *Games Econ. Behav.* 71 (1), 193–211.
- Bustos, Á.P., Chin, T.-J., 2017. Guaranteed outlier removal for point cloud registration with correspondences. *IEEE Trans. Pattern Anal. Machine Intell.* 40 (12), 2868–2882.
- Campbell, D., Petersson, L., 2016. GOGMA: Globally-optimal gaussian mixture alignment. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5685–5694.
- Chin, T.-J., Yu, J., Suter, D., 2011. Accelerated hypothesis generation for multistructure data via preference analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (4), 625–638.
- Chin, T.-J., Heng Kee, Y., Eriksson, A., Neumann, F., 2016. Guaranteed outlier removal with mixed integer linear programs. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5858–5866.
- Chum, O., Matas, J., Kittler, J., 2003. Locally optimized ransac. *Lect. Notes Comput. Sci.* 2781, 236–243.
- Curless, B., Levoy, M., 1996. A volumetric method for building complex models from range images. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM, pp. 303–312.
- Deng, H., Birdal, T., Ilic, S., 2018. Ppfnnet: Global context aware local features for robust 3d point matching.
- DurrantWhyte, H.F., Bailey, T., 2006. Simultaneous localization and mapping. *IEEE Robot. Autom. Mag.* 13 (2), 99–110.
- Engel, J., Schops, T., Cremers, D., 2014. Lsd-slam: Large-scale direct monocular slam 8690, 834–849.
- Enqvist, O., Kahl, F., 2008. Robust optimal pose estimation. In: *European Conference on Computer Vision*, pp. 141–153.
- Fan, W., Wen, C., Guo, Y., Wang, J., Li, J., 2016. Rapid localization and extraction of street light poles in mobile lidar point clouds: A supervoxel-based approach. *IEEE Trans. Intell. Transp. Syst.* 18 (2), 292–305.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24 (6), 381–395.
- Gojcic, Z., Zhou, C., Wegner, J.D., Wieser, A., 2019. The perfect match: 3d point cloud matching with smoothed densities. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5545–5554.
- Gojcic, Z., Zhou, C., Wegner, J.D., Wieser, A., 2019. The perfect match: 3d point cloud matching with smoothed densities. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5545–5554.
- Haala, N., Kada, M., 2010. An update on automatic 3d building reconstruction. *Isprs J. Photogramm. Remote Sens.* 65 (6), 570–580.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. Semantic3d.net: A new large-scale point cloud classification benchmark.
- Hartley, R.I., Kahl, F., 2009. Global optimization through rotation space search. *Int. J. Comput. Vision* 82 (1), 64–79.
- Huang, H., Kalogerakis, E., Chaudhuri, S., Ceylan, D., Kim, V.G., Yumer, E., 2017. Learning local shape descriptors from part correspondences with multi-view convolutional networks. *Acm Trans. Graphics* 37 (1), 1–14.
- Johnson, A.E., Hebert, M., 1999. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (5), 433–449.
- Kanazawa, Y., Kawakami, H., 2004. Detection of planar regions with uncalibrated stereo using distributions of feature points. In: *BMVC*. Citeseer, pp. 1–10.
- Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., Kolter, J.Z., Langer, D., Pink, O., Pratt, V., 2011. Towards fully autonomous driving: Systems and algorithms. In: *Intelligent Vehicles Symposium*, pp. 163–168.
- Li, M., Sun, C., 2018. Refinement of lidar point clouds using a super voxel based approach. *ISPRS J. Photogramm. Remote Sens.* 143, 213–221.
- Lin, Y., Wang, C., Zhai, D., Li, W., Li, J., 2018. Toward better boundary preserved supervoxel segmentation for 3d point clouds. *ISPRS J. Photogramm. Remote Sens.* 143, 39–47. *ISPRS Journal of Photogrammetry and Remote Sensing Theme Issue Point Cloud Processing*.
- Mellado, N., Aiger, D., Mitra, N.J., 2014. Super 4pcs fast global pointcloud registration via smart indexing. In: *Computer Graphics Forum*. vol. 33. Wiley Online Library, pp. 205–215.
- Olsson, C., Kahl, F., Oskarsson, M., 2009. Branch-and-bound methods for euclidean registration problems. *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (5), 783–794.
- Papon, J., Abramov, A., Schoeler, M., Worgotter, F., 2013. Voxel cloud connectivity segmentation-supervoxels for point clouds. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2027–2034.
- Parra Bustos, A., Chin, T.-J., 2015. Guaranteed outlier removal for rotation search. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2165–2173.
- Parra, B.A., Chin, T.J., Eriksson, A., Li, H., Suter, D., 2014. Fast rotation search with stereographic projections for 3d registration. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3930–3937.
- Rusu, R.B., Blodow, N., Marton, Z.C., Beetz, M., 2008. Aligning point cloud views using persistent feature histograms. In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 3384–3391.
- Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. In: *IEEE International Conference on Robotics and Automation*, pp. 1848–1853.
- Salhi, A., Horst, R., Tuy, H., 1994. Global optimization: Deterministic approaches. *J. Oper. Res. Soc.* 45 (5), 595.
- Shotton, J., Glocker, B., Zach, C., Izadi, S., Criminisi, A., Fitzgibbon, A., 2013. Scene coordinate regression forests for camera relocalization in rgb-d images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2930–2937.
- Svärm, L., Enqvist, O., Oskarsson, M., Kahl, F., June 2014. Accurate localization and pose estimation for large 3d models. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 532–539.
- Tam, G.K.L., Cheng, Z.Q., Lai, Y.K., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X.F., Rosin, P.L., 2013. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *IEEE Trans. Visual Comput. Graphics* 19 (7), 1199–1217.
- Tombari, F., Salti, S., Di Stefano, L., 2010. Unique Signatures of Histograms for Local Surface Description. *Springer Berlin Heidelberg, Berlin, Heidelberg*, pp. 356–369.
- Torsello, A., Bergamasco, F., Albarelli, A., Bronstein, A.M., Rodola, E., 2012. A game-theoretic approach to deformable shape matching. *23(10)*, 182–89.
- Tran, Q.H., Chin, T.J., Chojnacki, W., Suter, D., 2014. Sampling minimal subsets with large spans for robust estimation. *Int. J. Comput. Vision* 106 (1), 93–112.
- Wang, H., Wang, C., Luo, H., Li, P., Chen, Y., Li, J., 2017. 3-d point cloud object detection based on supervoxel neighborhood with hough forest framework. *8(4)*, 1570–1581.
- Xiao, G., Wang, H., Yan, Y., Suter, D., 2016. Superpixel-based two-view deterministic fitting for multiple-structure data. In: *European Conference on Computer Vision*. Springer, pp. 517–533.
- Xiao, G., Wang, H., Yan, Y., Suter, D., 2018. Superpixel-guided two-view deterministic geometric model fitting. *Int. J. Comput. Vision* 1–17.
- Yang, J., Li, H., Campbell, D., Jia, Y., 2016. Go-ICP: A globally optimal solution to 3d icp point-set registration. *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (11), 2241–2254.
- Yew, Z.J., Lee, G.H., 2018. 3dfeat-net: Weakly supervised local 3d features for point cloud registration. In: *European Conference on Computer Vision*. Springer, pp. 630–646.
- Zai, D., Li, J., Guo, Y., Cheng, M., Huang, P., Cao, X., Wang, C., 2017. Pairwise registration of tps point clouds using covariance descriptors and a non-cooperative game. *ISPRS J. Photogramm. Remote Sens.* 134, 15–29.
- Zeng, A., Song, S., Niebner, M., Fisher, M., Xiao, J., Funkhouser, T., 2016. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions, 199–208.
- Zhou, Q.-Y., Park, J., Koltun, V., 2016. Fast global registration. In: *European Conference on Computer Vision*. Springer, pp. 766–782.