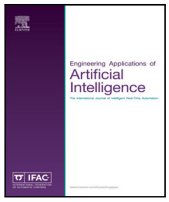
Engineering Applications of Artificial Intelligence 133 (2024) 107985

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Robust point cloud registration using Hough voting-based correspondence outlier rejection

Jihoon Han [a], Minwoo Shin [a], Joonki Paik [a,b,*]

[a] Department of Image, Chung-Ang University, Seoul, 06974, Republic of Korea
[b] Department of Artificial Intelligence, Chung-Ang University, Seoul, 06974, Republic of Korea

ARTICLE INFO

Keywords:
Point cloud registration
Outlier rejection
Hough voting

ABSTRACT

In this paper, we present a novel method for point cloud registration in large-scale 3D scenes. Our approach is accurate and robust, and does not rely on unrealistic assumptions. We address the challenges posed by scanning equipment like LiDAR, which often produce point clouds with dense properties. Additionally, our method is effective even in scenes with low overlap rates, specifically less than 30%. Our approach begins by computing overlap region-based correspondences. This involves extracting deep geometric features from point cloud pairs, which is especially beneficial in enhancing registration performance in cases with low overlap ratios. We then construct efficient triplets that vote in the 6D Hough space, representing the transformation parameters. This process involves creating a quartet from overlap region-based correspondences and then forming a final triplet following a sampling process.

To mitigate ambiguity during training, we use similarity values of the triplet as features of each vote when configuring votes for network input. Our framework incorporates the architecture of the Fully Convolutional Geometric Features (FCGF) network, augmented with a transformer's attention mechanism, to reduce noise in the voting process.

The final stage involves identifying the consensus of correspondence in the Hough space using a binning approach, which enables us to predict the final transformation parameters. Our method has demonstrated state-of-the-art performance on indoor datasets, including high overlap ratio data like 3DMatch and low overlap ratio data like 3DLoMatch. It has also shown comparable performance to leading methods on outdoor datasets like KITTI.

1. Introduction

Point cloud registration is a fundamental task to align a pair of three-dimensional (3D) scenes, and plays an important role in various 3D applications including: 3D reconstruction (Geiger et al., 2011; Chowdhury et al., 2021), simultaneous localization and mapping (SLAM) (Salas-Moreno et al., 2013; Zhang and Singh, 2015; Kim et al., 2018), and augmented reality (AR) (Azuma, 1997). Typical registration methods first assign coarse correspondences, and then align point clouds by estimating the rotation and translation in the 3D space (Choy et al., 2019b; Bai et al., 2020; Cao et al., 2021; Choy et al., 2020; Bai et al., 2021) (see Fig. 1).

Under challenging circumstances such as partially overlapped regions and feature ambiguity, a point cloud registration method cannot avoid outliers in the generated correspondences (Chen et al., 2022). To minimize distance between correspondences and to reject correspondence outliers, there are two approaches: (i) feature learning-based and (ii) end-to-end learning-based methods. The former extracts discriminative features using a learnable model, based on which robust correspondences are generated, and the transformation matrix can be predicted using various optimizer methods (*e.g.* RANSAC).

On the other hand, the latter approach generates correspondences using *a priori* learned features and predicts optimal transformation parameters. This is achieved by filtering out correspondence outliers through a learnable transformation estimation process. However, a limitation of these end-to-end learning-based methods is the assumption of a high overlap ratio between point cloud pairs, which may not always be realistic.

In this paper, we introduce a method that accurately estimates the transformation between point clouds, effectively rejecting correspondence outliers without relying on such assumptions. Our method predicts the most effective transformation by finding a consensus among valid correspondences in overlapped regions, utilizing Hough voting.

* Corresponding author at: Department of Image, Chung-Ang University, Seoul, 06974, Republic of Korea.
E-mail address: paikj@cau.ac.kr (J. Paik).

https://doi.org/10.1016/j.engappai.2024.107985
Received 13 July 2022; Received in revised form 21 December 2023; Accepted 24 January 2024
Available online 6 February 2024

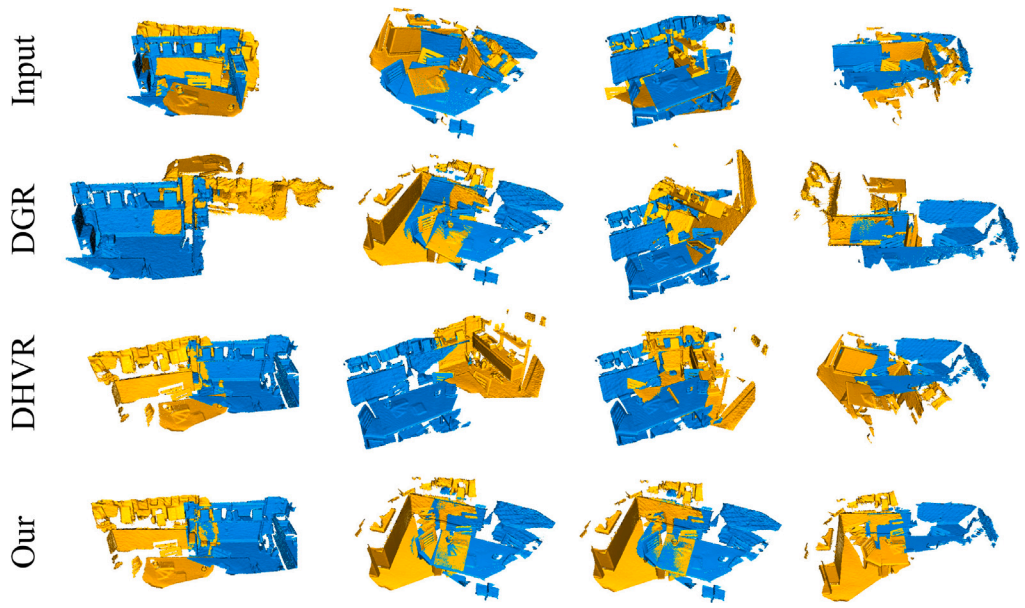


Fig. 1. Pairwise point cloud registration results of DGR, DHVR and ours on 3DLoMatch. Our method is robust in low overlap regions compared to the rest.

The approach for predicting valid correspondences in these regions draws inspiration from the works of (Yew and Lee, 2020). Additionally, our method for estimating the best consensus between correspondences through Hough voting is influenced by the Deep Hough Voting for Registration (DHVR) method (Lee et al., 2021).

However, as our proposed method is based on voting, it is crucial to clearly differentiate it from DHVR, the latest in voting-based approaches. Our method introduces unique improvements and variations, distinguishing it from existing techniques in both approach and efficacy.

1. In DHVR, the creation of triplets involves randomly selecting three points from the initial matches, which can include outliers. This leads to the possibility of having triplets that include correspondences with a large distance between them. To overcome this challenge, we put forward a solution that converts the distance between the correspondences into a probability, which is then used as the basis for the sampling process. With the proposed scheme, we can eliminate correspondences that have a significant distance from a randomly selected group of four correspondences.
2. To mitigate the impact of noise introduced by the acquisition of high-dimensional matrices, DHVR used the Simple Network of Fully Convolutional Geometric Features (FCGF) to effectively eliminate the noise. In contrast, the proposed solution combines self-attention modules from Transformers with FCGF's Simple Network to eliminate the noise from the votes in an end-to-end manner, without relying on Gaussian kernels.

The proposed algorithm consists of five steps: (i) filtering pairs of overlapped regions based on specific features (*e.g.*, spatial, textural) extracted from the backbone network, (ii) selecting triplet pairs from correspondence quartets based on a defined criterion (*e.g.*, geometric consistency), (iii) computing transformation parameters for these triplets and mapping these six-dimensional (6D) parameters into Hough space, (iv) applying a specified learnable model (*e.g.*, a neural network) in the Hough space to refine vote values, and (v) estimating the optimized transformation parameters by determining the maximum vote value through a specific method (*e.g.*, peak detection).

The proposed algorithm introduces two key innovations to address the challenges in point cloud registration. Firstly, it employs a novel sampling scheme that converts correspondence distances into probabilities, ensuring the selection of reliable triplets while eliminating outliers. This approach enhances the robustness of the method in challenging scenarios with partial overlaps and feature ambiguity.

Secondly, the algorithm combines self-attention modules from Transformers with FCGF's Simple Network to eliminate noise in the voting process in an end-to-end manner. By bypassing the reliance on Gaussian kernels, this integration enhances the accuracy of the transformation parameter estimation, making it more suitable for real-world applications with noisy data.

These innovations set the proposed method apart from the latest voting-based technique, DHVR, and make it a more robust and accurate solution for point cloud registration in large-scale 3D scenes.

2. Related work

Point cloud registration aims to estimate a three-dimensional (3D) rigid transformation between two sets of point clouds, which are referred to as $\mathcal{X}$ (the source) and $\mathcal{Y}$ (the target). The objective is to transform $\mathcal{X}$ in such a way that it aligns or overlaps optimally with $\mathcal{Y}$. This transformation involves a rotation matrix $\mathbf{R}$ that belongs to the special orthogonal group SO_3, and a translation vector $\mathbf{t}$ from the real number space $\mathbb{R}^3$.

The problem of point cloud registration seeks to determine the parameters of this rigid transformation. This is mathematically expressed as finding the minimum of the following equation:

$$\arg\min_{\mathbf{R},\mathbf{t}} \frac{1}{N} \sum_{\mathbf{x}\in\mathcal{X},\mathbf{y}\in\mathcal{Y}} |\mathbf{R}\mathbf{x}+\mathbf{t}-\mathbf{y}|^2, \tag{1}$$

where N denotes the number of points in each set of the point clouds. The goal is to minimize the average squared distance between the points in the transformed source cloud $\mathbf{R}\mathbf{x}+\mathbf{t}$ and the points in the target cloud $\mathbf{y}$.

2.1. Local registration

As the first seminal approach to local point cloud registration, the classic iterative closest point (ICP) aims to minimize the distance between corresponding points through iteratively updating the transformation (Besl and McKay, 1992). There are many variants of ICP to overcome the limitations of its original version. To make the ICP

more robust, Chetverikov et al. proposed Trimmed ICP (TrICP) that rejected outliers by applying continuous least trimmed squares (LTS) to all phases of the operation (Chetverikov et al., 2002; Besl and McKay, 1992). Picky ICP tried to reject all the points except minimum-distance correspondences (Zinßer et al., 2003), and Almhdie et al. used a comprehensive look-up matrix to find the best correspondence (Almhdie et al., 2007). Rusinkiewicz improved the registration performance in the case with noise and partial overlap using a symmetric loss function (Rusinkiewicz, 2019). However, the ICP and its variants have common drawbacks: (i) complexity increases with the number of points for the nearest correspondences estimation in their inner loops and (ii) dependence on the optimal initialization.

2.2. Global registration

RANdom SAmple Consensus (RANSAC) is the most popular approach to global registration by (i) iteratively sampling sets of candidate correspondences, (ii) performing pair-wise alignment, and (iii) evaluating the result of alignment (Fischler and Bolles, 1981). Yang et al. adopted the branch-and-bound strategy to obtain the optimal transformation. Yang's method provides a globally optimal solution to 3D ICP point-set registration (Go-ICP) and is faster than RANSAC (Yang et al., 2015). Zhou et al. sped up the global registration process using second-order gradients (Zhou et al., 2016). However, its performance depends on the quality of extracted features. Another Yang et al. formulated the registration problem using truncated least squares (TLS) and achieved high robustness against outliers using a general graph algorithm called TEASER (Yang et al., 2020). Chen et al. proposed the second-order spatial compatibility measure to compute similarity between correspondences (Chen et al., 2022). These registration approaches commonly increase computational cost to overcome various challenges of the point cloud registration problem such as noise variation, outliers, density variation, and partial overlap.

2.3. Feature-learning based methods

Traditional feature-based methods used hand-crafted feature extractions. Rusu et al. proposed fast point feature histograms (FPFH) (Rusu et al., 2009), and Salti et al. proposed SHOT (Salti et al., 2014). Both of them used RANSAC to perform registration. Early learning-based methods use deep neural networks (DNNs) to extract features, and then RANSAC-based correspondence generation and transformation estimation are separately performed afterward. After Zeng et al. proposed 3DMatch (Zeng et al., 2017), various learnable models extracted features using neural networks (Choy et al., 2019b; Wang and Solomon, 2019; Bai et al., 2020; Cao et al., 2021). Choy et al. used a 3D fully convolutional network to extract geometric features in the single path with a novel learning loss function (Choy et al., 2019a,b). To efficiently extract features from point clouds, the proposed solution adopted Minkowski Convolutional Neural Networks used in Choy et al. (2019a). These networks can perform high-dimensional matrix operations in the complex input space of point clouds, and thus provide an ideal fit for the proposed model. The network was used without any modifications. Bai et al. also proposed a new learning method based on a 3D fully convolutional network to predict detection score and description features (Bai et al., 2020).

On the other hand, Cao et al. computed final correspondences using confidence scores and then estimated the optimal transformation using singular value decomposition (SVD) without using RANSAC (Cao et al., 2021). Cao's method tried to find point pairs using a cross-attention matrix that combines low- and high-level contextual information. Wang et al. used learned point-wise features to generate correspondences and to predict the relative pose (Wang and Solomon, 2019). Since these two methods used a learning feature extraction network for point-to-point matching instead of the entire registration, their performance is degraded when the distribution of training data is not sufficiently similar to the target data.

2.4. End-to-end learning methods

The main idea of the end-to-end learning-based registration method is to obtain the transformation parameters between a pair of point clouds as input to the neural network to generate correspondences. To solve the registration problem, we combine registration-related optimization theories with deep neural networks. Aoki et al. proposed PointNetLK that uses Lucas-Kanade registration to find the best transformation matrix that aligns the point cloud in the feature space using PointNet (Aoki et al., 2019). However, this method proceeds repetitive learning to find the optimal twist parameters, resulting in increased computational complexity. Yuan et al. learns point correspondences using the Gaussian mixture model (GMM) (Yuan et al., 2020). They optimized the GMM to predict the optimal transformation. To remove outliers, Pairs et al. used a 3D regression network (3DRegNet) that finds the final set of candidates by classifying inlier and outlier pairs (Pais et al., 2020). Choy et al. used a six-dimensional convolutional network to predict the correspondence confidence and performed robust gradient-based $SE(3)$ optimization using differentiable weighted Procrustes algorithm-based pose estimation (Choy et al., 2020). However, 3DRegNet and DGR have a common problem of ignoring the properties of rigid transformation due to powerful additional information. Bai et al. proposed PointDSC that develops spatial consistency to prune outlier correspondences (Bai et al., 2021). Lee et al. applied Hough voting to decide a good consensus between candidate pairs (Lee et al., 2021). Chen et al. proposed consensus encoding unit (CEU) to generate distinguishable features from putative sets of correspondences, lightened the spatial attention model for better pairs and performed classification using instance-unique channel attention that combined spatial and channel attention (SCS) (Chen et al., 2021).

2.5. Hough voting

Hough transform (HT) converts pattern recognition problems in the image into a peak detection problem in the feature space (Hough, 1962). Ballard et al. generalized the HT to locate various objects in the image (Ballard, 1981).

Hough transform-based voting method to detect objects, called *Hough voting*, consists of learning and inference steps (Leibe et al., 2008): (i) Given an image set containing a labeled object bounding box, the learning step generates a mapping between image patches and the offsets from the corresponding object center in the form of a *codebook*. (ii) The inference step first selects points-of-interest, and extracts patches from the neighbor of the points. Next, the vote is calculated from the offset between the extracted patches and codebook patches. If the two patches are identical, a vote cluster is generated in the center of the corresponding object. Finally, the object bounding box is detected by tracking the center of objects with the maximum vote cluster.

Qi et al. applied the traditional Hough voting method to the point cloud object detection problem (Qi et al., 2019). Qi's method generated votes using a deep neural network instead of a codebook to reduce the feature ambiguity using larger receptive fields. It can also augment vote locations using feature vectors.

The Hough voting technique has been widely employed for detecting and recognizing geometric primitives in point cloud data. Limberger et al. utilized Hough voting to convert the problem of detecting planar regions in unorganized point clouds into a peak detection problem in the feature space. This approach enabled the robust identification of planar surfaces (Limberger and Oliveira, 2015). Building upon this framework, Birdal et al. extended the Hough voting approach to fit generic primitive shapes to point cloud data, allowing for the fitting of various types of quadrics. By leveraging the Hough voting method, accurate shape estimation and reconstruction from point cloud information were achieved (Birdal et al., 2019). Further advancements were made by Raffo et al., who introduced a localized voting procedure

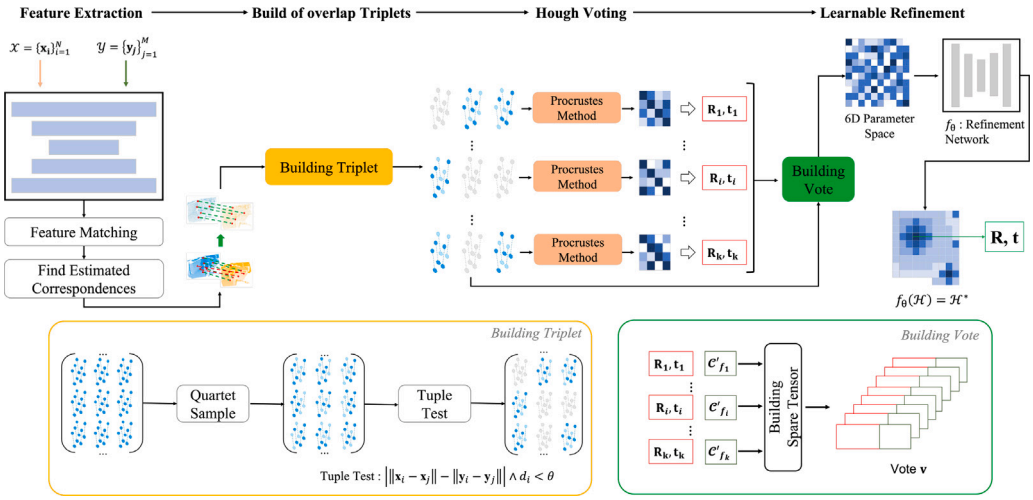


Fig. 2. The proposed method. We perform local feature-based matching between point cloud pairs using a set of triplets. Each triplet is cast with sparse Hough space divided by bin. We enhance the Hough space by integrating the attention mechanism from Transformers with FCGF's Simple Network using the proposed solution, effectively reducing noise from the Hough voting space. Finally, the bin with the maximum vote is used as the final conversion parameter.

based on the Hough voting approach. Their method focused on fitting and recognizing simple geometric primitives, such as lines and circles, in segmented 3D point clouds.

Another approach enhanced the detection and recognition of geometric primitives in point cloud data by leveraging localized voting. According to Raffo et al. (2022), this technique has been successfully applied to simpler geometric primitives, including planes, cylinders, cones, spheres, and tori. More recently, Romanengo et al. (2023) extended the application of Hough voting to more complex geometric primitives. Their study focused on the detection and recognition of generalized cylinders, generalized cones, surfaces of revolution, helical surfaces, and helical strips. The Hough voting technique, known for its ability to handle noise and partial observations, proved effective in reliably detecting and recognizing these complex shapes, as demonstrated in their work.

The proposed method introduces an innovative end-to-end learning-based framework for point cloud registration, utilizing Hough voting as a key component. Initially, we generate correspondences by extracting point features from a backbone network. Following this, we enhance the voting process by augmenting features related to the transformation parameters, using information derived from triplet correspondences. To improve the quality of our results, we filter out low-quality votes. This is achieved by combining an attention mechanism, inspired by Transformers, with the Simple Network from Fully Convolutional Geometric Features (FCGF). This integration is a crucial aspect of our proposed solution.

3. Methodology

The proposed point cloud registration algorithm consists of (i) feature extraction and matching, (ii) correspondences sampling for effective triplet, (iii) transformation of 6D parameters into the Hough space, and (iv) learning-based refinement of the Hough space as shown in Fig. 2.

3.1. Feature extraction

To generate the initial correspondences, we used the fully convolutional 3D feature descriptor (FCGF) (Choy et al., 2019b). The FCGF takes sparse tensors as input to produce a 32-dimensional descriptor for each point corresponding to the single-pass sparse point cloud. Given a pair of point cloud, $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^3 | i = 1, 2, \dots, N\}$ and $\mathcal{Y} = \{\mathbf{y}_j \in \mathbb{R}^3 | j = 1, 2, \dots, M\}$, local features are extracted from each point cloud. Using the extracted features, the top-1 nearest neighbor is selected to generate point pairs from $\mathcal{X}$ to $\mathcal{Y}$ and the corresponding counter pairs from $\mathcal{Y}$ to $\mathcal{X}$. Let $N_{\mathcal{X}\to\mathcal{Y}}$ and $N_{\mathcal{Y}\to\mathcal{X}}$ be the numbers of correspondence from $\mathcal{X}$ to $\mathcal{Y}$ and from $\mathcal{Y}$ to $\mathcal{X}$, respectively. Then we have a set of initial candidate correspondences, denoted as $\mathcal{C}$, having $N = N_{\mathcal{X}\to\mathcal{Y}} + N_{\mathcal{Y}\to\mathcal{X}}$ elements.

3.2. Correspondences sampling for effective triplet

Let $\mathbf{f}_{\mathbf{x}_i}$ and $\mathbf{f}_{\mathbf{y}_j}$ respectively represent descriptor vectors of $\mathbf{x}_i \in \mathcal{X}$ and $\mathbf{y}_j \in \mathcal{Y}$, then we define the set of correspondences from $\mathcal{X}$ to $\mathcal{Y}$ as

$$\mathcal{C}_{\mathcal{X}}^{\mathcal{Y}} = \{(\mathbf{x}_i, \mathbf{y}_j) | (\mathbf{x}_i, \mathbf{y}_j) \in \mathcal{X} \times \mathcal{Y}\}, \tag{2}$$

where $(\mathbf{x}_i, \mathbf{y}_j) = \mathrm{nn}(\mathbf{f}_{\mathbf{x}_i}, \mathbf{f}_{\mathbf{y}_j})$ and $\mathrm{nn}(\cdot,\cdot)$ is the nearest neighbor search operator (Johnson et al., 2019). The similarity calculation method proposed in Johnson et al. (2019) is the best fit for our paper, which utilizes the distance between correspondences as a feature. It has the advantage of being faster and more efficient in terms of GPU consumption compared to the widely used Euclidean method. Next, we adopted the fast global registration (FGR) (Zhou et al., 2016) and deep Hough voting for robust global registration (DHVR) (Lee et al., 2021) for the Hough voting step from three unique candidate correspondences. To remove triplets including spurious correspondences, we perform a simple tuple test. FGR proposes the use of triplets to calculate the optimal transformation matrix. It generates the triplet by randomly selecting three points from the initial correspondences, which can result in the inclusion of correspondences with a large distance between them. This method can impair the performance of point cloud registration. To address this issue, this paper proposes a two-step sampling process of the initial correspondences to generate triplets that have outliers removed. The triplet test proposed by FGR is considered the optimal solution for removing outliers from the triplets. Therefore, we adopted FGR to remove outliers from the triplets generated using the proposed solution. This task can be summarized as: (i) construction of overlap correspondence and (ii) triplet filtering. To construct the overlap correspondence $\mathcal{C}'$, we extract pairs in the overlap region of the initial correspondences $\mathcal{C}$. The distance between extracted pairs is defined as $D(\mathcal{C}_{\mathcal{X}}^{\mathcal{Y}}) = \{d_1, d_2, \dots, d_N\}$, and the probability of the ith pair distance is

$$p(d_i) = \frac{d_i^{-1}}{\sum_{i=1}^{N} d_i^{-1}}, \quad \sum_{i=1}^{N} p(d_i) = 1 \tag{3}$$

We construct $\mathcal{C}'$ using the high-probability pairs based on top-k. Since the original DHVR randomly extracts initial triplets from correspondences, the triplets may include pairs with a large distance. To avoid

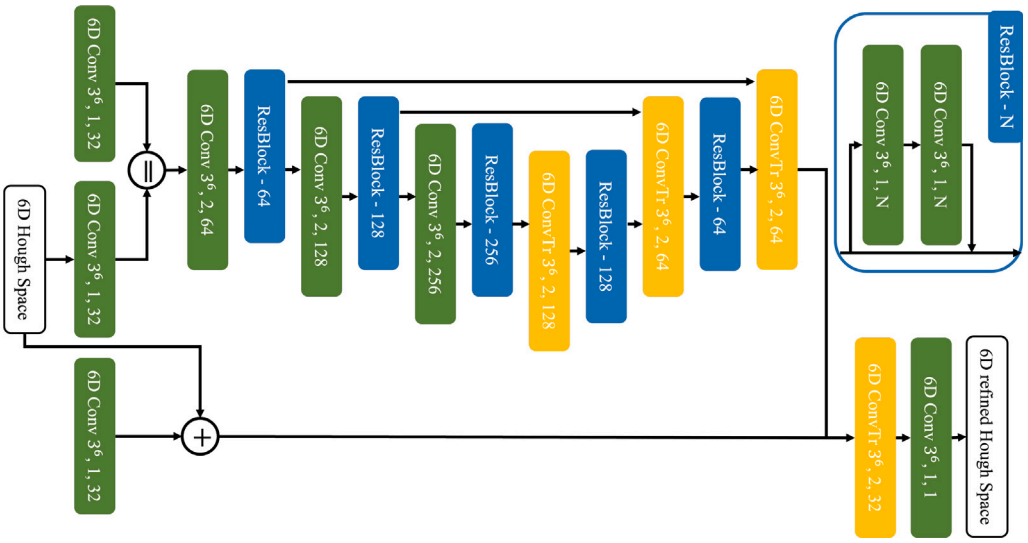


Fig. 3. Architecture of the refinement network. The numbers in each layer represent the kernel size, stride, and channel size, respectively. The network has 6D U-shaped sparse attention convolutional layers with skip connections.

this problem, we initially construct quartets and finally reject one with the largest distance. For triplet filtering, we adopt the simple tuple test used in DHVR, where the distance set function is defined as

$$D(\mathcal{C}^{\mathcal{Y}}_{\mathcal{X}}) = \mathcal{D}^{\mathcal{Y}}_{\mathcal{X}} = \{d_i \in \mathbb{R} | \forall i \neq j, d_i = |\|\mathbf{x}_i - \mathbf{x}_j\| - \|\mathbf{y}_i - \mathbf{y}_j\|| \wedge d_i < \theta\}, \quad (4)$$

where θ represents the distance threshold, and the upper bound of spatial hashing error which is approximately equal to 3 times the voxel size, such as $3 \times v$. The tuple test is efficient when rotation and translation are invariant to the point-to-point distance (Zhou et al., 2016).

3.3. Transformation of 6D parameters to the Hough space

We used Hough voting as an optimal consensus method to predict rotation and translation parameters between overlap correspondences C'. To represent parameters of the rotation and translation in the 3D space, we used the axis-angle representation with six degrees of freedom (Yang et al., 2015).

We construct a discretized 6D parameter space using rotation bin $b_{\mathbf{r}}$, translation bin $b_{\mathbf{t}}$, which is a discrete unit, to construct *Hough voting* consisting of the transformation parameters of each candidate overlap correspondences C'. The purpose of this process is to identify bins with transformation parameters that best represent *Hough voting* for final transformation parameter prediction.

Using Axis-angle presentation, we can represent the 3D vector $\mathbf{r} \in \mathbb{R}^3$ as a rotation as

$$\|\mathbf{r}\| = \arccos \frac{\mathrm{Tr}(\mathbf{R}) - 1}{2}, \quad \frac{\mathbf{r}}{\|\mathbf{r}\|} = \frac{1}{2\sin\theta}\begin{bmatrix}\mathbf{R}_{32} - \mathbf{R}_{23} \\ \mathbf{R}_{13} - \mathbf{R}_{31} \\ \mathbf{R}_{21} - \mathbf{R}_{12}\end{bmatrix}, \quad (5)$$

where $\mathbf{r}/\|\mathbf{r}\|$ represents the rotation axis, $\|\mathbf{r}\|$ the rotation angle, and TR(·) the trace of the matrix. In the axis-angle representation, the 3D rotation is located in a sphere with radius π, and the translation parameters are in a bounded cube $[-\xi, \xi]^3$ (Yang et al., 2015).

When voting in the 6D Hough space, two-parameter spaces are overlapped and the translation parameters may be unbounded (Lee et al., 2021). To solve that problem, DHVR used sparse 6D Hough space which can separate overlapped spaces using discrete bins and can bind the translation parameters. The sparse 6D Hough space, denoted as $\mathcal{H}$, can represent all transformation parameters between a point cloud pair and can aggregate votes in the form of bins in the set of triplets. We can express each triplet as

$$\hat{\mathbf{T}}_i = \min_{\mathbf{T} \in SE(3)} \sum_{(\mathbf{x},\mathbf{y}) \in C'_i} \|\mathbf{T}(\mathbf{x}) - \mathbf{y}\| \quad (6)$$

where C'_i represents the ith triplet, and $\mathbf{T} \in SE(3)$ represents the transformation parameters estimated from point cloud pairs. We used Procrustes method to obtain $\hat{\mathbf{T}}_i$ (Gower, 1975). $\hat{\mathbf{T}}_i$ consists of rotation $\hat{\mathbf{R}}_i$ and translation $\hat{\mathbf{t}}_i$, and can be obtained as

$$\hat{\mathbf{R}}_i = \mathbf{V}\mathrm{diag}(1, 1, |\mathbf{V}\mathbf{U}^T|)\mathbf{U}^T, \quad \text{and} \quad \hat{\mathbf{t}}_i = \bar{\mathbf{y}}_i - \hat{\mathbf{R}}_i\bar{\mathbf{x}}_i, \quad (7)$$

and the cross-covariance matrix $\boldsymbol{\Gamma}$ is expressed as

$$\boldsymbol{\Gamma} = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}}_i)(\mathbf{y}_i - \bar{\mathbf{y}}_i)^T, \quad (8)$$

where singular value decomposition $\boldsymbol{\Gamma} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ is used to decompose $\boldsymbol{\Gamma} \in \mathbb{R}^{3\times3}$, and $\bar{\mathbf{x}}_i, \bar{\mathbf{y}}_i \in \mathbb{R}^3$ represent the centroid of point clouds $\mathcal{X}_i, \mathcal{Y}_i$. If we transform the rotation matrix $\hat{\mathbf{R}}_i$ as (4), we can have an axis-angle representation of the rotation and translation for voting. $\mathbf{v}_i$ in the 6D Hough space can be expressed as

$$\mathbf{v}_i = \mathrm{concat}(\lfloor \frac{\hat{\mathbf{r}}_i}{b_{\mathbf{r}}} \rfloor, \lfloor \frac{\hat{\mathbf{t}}_i}{b_{\mathbf{t}}} \rfloor), \quad (9)$$

where $\lfloor \cdot \rfloor$ represents the floor operation, $b_{\mathbf{r}}, b_{\mathbf{t}}$ hyper-parameters representing the bin size of rotation and translation, and $\mathbf{v}_i$ consists of their concatenation. We used feature similarity values of correspondences in each triplet as a voting value (Li et al., 2020; Chen et al., 2022). We observed that accurate prediction was possible when the bin size was set to a sufficiently small value. After transforming all the votes, we can obtain sparse 6D Hough space of transformation parameters (see Fig. 3).

3.4. Refinement for Hough space

After casting all the votes, the sparse Hough space may become noisy since triplets have correspondence outliers. Furthermore, the transformation from the continuous 6D space to the discrete Hough space may generate additional quantization noise. To solve that problem, we refine the Hough space using a high-dimensional sparse convolution network, incorporating an attention module described as $f_\theta : \mathbb{R} \to \mathbb{R}$. We used a simple U-shaped 6D sparse convolution network, enhanced with an attention mechanism, to generate the refined space as

$$\mathcal{H}^* = f_\theta(\mathcal{H}). \quad (10)$$

We can train f_θ using binary cross entropy loss

$$\mathcal{L}_{bce}(\mathcal{H}^*, \mathbf{T}_{GT}) = \sum_{i=0}^{|\mathcal{H}^*|} (o_i \log \mathcal{H}^*_i + 1 - o_i \log \mathcal{H}^{*C}_i), \quad (11)$$

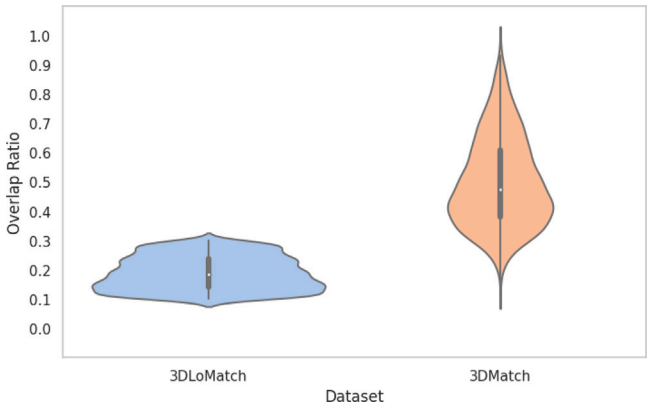


Fig. 4. Distribution of overlapping regions for registration pairs in the 3DMatch and 3DLoMatch datasets.

where $\mathbf{T}_{GT}$ represents the ground truth transformation matrix, and $\mathcal{H}_i^{*C} = 1 - \mathcal{H}_i^*$. If the ith bin of $\mathcal{H}^*$ is close to $\mathbf{T}_{GT}$, o_i becomes close to the unity. Otherwise, o_i becomes close to zero. We additionally used the transformation loss

$$\mathcal{L}_{trans}(\mathbf{R}, \mathbf{t}) = \|\mathbf{R}_{GT}^T \mathbf{R} - I\|^2 + \|\mathbf{t}_{GT} - \mathbf{t}\|^2. \tag{12}$$

The total loss is the weighted sum of (11) and (12)

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{bce}(\mathcal{H}^*, \mathbf{T}_{GT}) + \lambda \mathcal{L}_{trans}(\mathbf{R}, \mathbf{t}), \tag{13}$$

where α and λ are regularization hyper-parameters, and are experimentally chosen.

4. Experiments

In this section, we describe several ways to evaluate the performance of the proposed point cloud registration method. After summarizing the datasets and implementation details in Section 4.1, we explain various performance evaluation metrics including registration recall, relative translation error (RTE), and relative rotation error (RRE) in Section 4.2. Comparative experimental results are given in Section 4.3, and the ablation study is given in Section 4.4.

4.1. Implementation and dataset details

Indoor Benchmarks: 3DMatch and 3DLoMatch We used two standard datasets including 3DMatch (Zeng et al., 2017) and 3DLoMatch (Huang et al., 2021) to evaluate our algorithm. As used in the official data split, 3DMatch uses a total of 62 scenes, including 46 scenes for training, 8 scenes for validation, and 8 scenes for testing. We sampled down the point cloud to a 5 cm voxel grid. 3DLoMatch, which contains point cloud pairs with a low overlap ratio, was evaluated using the protocol proposed in Lee et al. (2021). 3DMatch consists of point cloud pairs with an overlap ratio of 30% or more, while 3DLoMatch consists of point cloud pairs with an overlap ratio of less than 30%. The distribution of overlapping regions for registration pairs in the 3DMatch and 3DLoMatch datasets is shown in Fig. 4.

We generate training data by applying less rigid transformations for augmentation.

Outdoor Benchmarks: KITTI odometry KITTI odometry conducts registration performance evaluation using a point cloud measured by precise LiDAR and a stereo sequence composed of ground truth positions provided by a GPS (Pathak et al., 2010). KITTI consists of 22 sequences including 11 sequences with ground-truth (00-10) and another 11 without ground-truth (11–21). We used sequences 00-05 for training, 06-07 for validation, and 08-10 for testing. We set the voxel size to 30 cm and sampled down the voxel-grid.

Table 1
Number of points in overlapping regions for 3DMatch and 3DLoMatch datasets.

	3DMatch	3DLoMatch
Average	2640.12	1938.86

Implementation Details We trained the refinement network using an Adam optimizer with a learning rate of 0.001 and weight decay of 0.0001 for 10 epochs (Kingma and Ba, 2014). We used bin sizes of $b_{\mathbf{r}} = 0.02$ rad, $b_{\mathbf{t}} = 0.02$ m for 3DMatch datsets, and $b_{\mathbf{r}} = 0.005$ rad, $b_{\mathbf{t}} = 0.05$ m for KITTI. The batch size was set to 6 for the 3DMatch dataset and 2 for the KITTI dataset. All the experiments were conducted on a personal computer with an Intel i9-10900KF CPU, 32 GB RAM, and NVIDIA RTX 2080Ti.

4.2. Evaluation metrics

We used two evaluation metrics to quantify the error between the transformation prediction of our algorithm and the ground-truth rigid transformation. The relative rotation error (RRE) evaluates the isotropic error (Yew and Lee, 2020), whereas the relative translation Error (RTE) (Yew and Lee, 2020) evaluates the ℓ_2 error and can be expressed as (14) and (15):

$$\text{RRE}(\hat{\mathbf{R}}) = \arccos \frac{\text{Tr}(\hat{\mathbf{R}}^\top \mathbf{R}^*) - 1}{2}, \tag{14}$$

and

$$\text{RTE}(\hat{\mathbf{t}}) = \|\hat{\mathbf{t}} - \mathbf{t}^*\|_2^2. \tag{15}$$

Registration recall refers to the percentage of successfully sorted point clouds. Pairwise registrations were considered successful only if RRE and RTE were below the thresholds.

4.3. Comparative experiment

Evaluation on indoor dataset For the evaluation of the indoor dataset, we used the 3DMatch benchmark dataset consisting of 8 scenes. We employed the official implementation of the FCGF descriptor and utilized pre-trained weights from the 3DMatch model with a consistent voxel size of 5 cm.

We compared the performance of our algorithm with traditional methods and deep learning-based methods that include state-of-the-art methods. The thresholds of RTE and RRE were respectively set to 30 cm and 15° for successful registration, and experiments were conducted using the set of parameters that gives the highest performance. The number and ratio of points within the overlapping regions, as well as experimental outcomes based on sampling methods (both proposed and baseline), are presented. Table 1 illustrates the count of points in the overlapping areas for each dataset used in the testing phase. Concurrently, Fig. 4 displays the proportion of overlap for each dataset.

Comparison with traditional and learning-based methods We compared 13 point cloud registration methods including RANSAC, SM, FGR, Super4PCS, Go-ICP, Point-to-Point and Point-to-plane based ICP, TEASER, SC^2-PCR, DCP, PointNetLK, DGR, DHVR, and PointDSC.

Eight of them are traditional methods, and five of them are deep learning-based methods. RANSAC and FGR used FPFH as a descriptor for generating corresponding points. We used a pre-trained model of learning-based methods for the best performance. We chose DGR and DHVR as the baseline since they focus on removing outliers among point cloud registration methods based on deep learning.

We used a learned FCGF descriptor to generate initial putative correspondences. The proposed method is different from other papers as it goes through a process of sampling correspondences that belong to the overlap region. This process can remove the outliers present in the initial correspondence. And, by using a different triplet generation method than DHVR, we can remove correspondences with long

Table 2
Comparison of state-of-the-art learning-based methods and our method using the 3DMatch dataset.

	Recall (% ↑)	RTE (cm ↓)	RRE (°↓)	Time (s)
SM (Leordeanu and Hebert, 2005)	86.57	7.07	2.29	**0.03**
FGR (Rusu et al., 2009)	42.7	10.6	4.08	0.31
TEASER (Yang et al., 2020)	85.77	8.66	2.73	0.07
SC^2-PCR (Chen et al., 2022)	93.28	6.55	**2.08**	0.11
RANSAC-2M (Fischler and Bolles, 1981)	66.1	8.85	3.00	1.39
RANSAC-4M	70.7	9.16	2.95	2.32
RANSAC-8M	74.9	8.96	2.92	4.55
Super4PCS (Mellado et al., 2014)	21.6	14.1	5.28	4.55
Go-ICP (Yang et al., 2015)	22.9	14.7	5.38	771.0
ICP(P2Point) (Zhou et al., 2018)	6.04	18.1	8.25	0.25
ICP(P2Plane) (Zhou et al., 2018)	6.59	15.2	6.61	0.27
DCP (Wang and Solomon, 2019)	3.22	21.4	8.42	0.07
PointNetLK (Aoki et al., 2019)	1.61	21.3	8.04	0.12
DGR (Choy et al., 2020)	85.2	7.73	2.58	0.70
DHVR (Lee et al., 2021)	91.4	6.61	**2.08**	0.46
PointDSC (Lee et al., 2021)	92.85	6.51	**2.08**	0.10
Our	**95.82**	**6.11**	2.14	0.53

Table 3
Quantitative results on 3DLoMatch dataset.

	FCGF (Choy et al., 2019b)			Predator (Huang et al., 2021)		
	Recall (% ↑)	RTE (cm ↓)	RRE (°↓)	Recall (% ↑)	RTE (cm ↓)	RRE (°↓)
FGR	19.99	12.98	5.28	35.99	11.64	4.77
RANSAC	46.38	13.11	5.00	64.85	11.04	4.28
PointDSC	56.09	**10.39**	3.87	68.89	9.6	3.43
DHVR	54.41	12.56	4.14	65.41	12.33	4.97
Our	**78.95**	12.04	**3.95**	**83.8**	**9.32**	**3.19**

Table 4
Scene-specific statistics for the 3DMatch and 3DLoMatch datasets.

	3DMatch			3DLoMatch		
	PointDSC	DHVR	Our	PointDSC	DHVR	Our
Kitchen	97.81	97.83	**99.60**	63.24	62.67	**77.71**
Home 1	96.44	**96.79**	**96.79**	48.45	46.02	**52.60**
Home 2	80.21	78.85	**80.29**	**57.96**	54.35	57.83
Hotel 1	97.54	**98.67**	98.23	61.78	62.39	**70.18**
Hotel 2	92.31	**95.19**	**95.19**	53.87	56.33	**72.78**
Hotel 3	92.59	90.74	**94.44**	54.67	51.02	**71.43**
Study	87.95	89.73	**92.81**	44.98	42.08	**66.67**
MIT lab	79.92	77.92	**87.01**	39.61	37.50	**61.11**
Average	92.45	92.61	**94.39**	53.07	54.18	**67.38**

distances between them which correspond to outliers in the triplets. Through this method, we can obtain triplets with outliers removed, allowing us to get the optimal transformation matrix.

As shown in Table 2, our method achieved a higher registration recall improvement than other methods. In registration recall, our method achieves more than 1.97% and 1.74% higher performance than PointDSC, a state-of-the-art method based on learning, and SC^2-PCR, a state-of-the-art method based on tradition, respectively. Compared to the baseline DGR of our paper, the recall was higher than 7.6% and higher performance than DHVR's recall was 3.4%.

Moreover, our algorithm was evaluated using the 3DLoMatch dataset because it aims to improve registration performance on point cloud datasets with low overlap rates without unrealistic assumptions of existing registration methods. Because our algorithm generated correspondences focused on the overlap, we achieved higher performance than other methods on the 3DLoMatch dataset. FCGF and Predator (Huang et al., 2021) were used as feature descriptors, and other methods used for comparison in this experiment were set to the conditions that respectively give the highest performance.

As shown in Table 3, in the case of FCGF descriptor, our method achieved approximately 20% improvement in registration recall, and

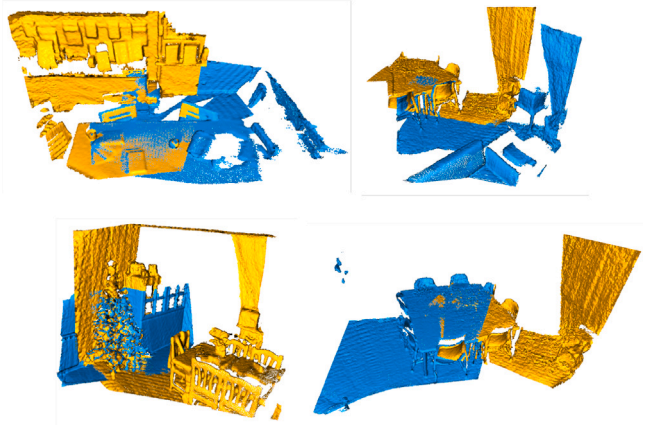


Fig. 5. Updated pairwise registration results on the 3DLoMatch dataset.

in the case of Predator descriptor, more than 15% improvement in registration recall.

Table 4 shows the proposed approach where to measure the resilience of the proposed model against others, we added noise points into the 3DMatch and 3DLoMatch datasets. This intervention subsequently influenced a scene-specific registration recall. As shown in Table 4, to test the robustness between our model and other models, we generated noise points on the 3DMatch and 3DDoMatch datasets, resulting in a scene-specific registration call.

Fig. 5 shows the alignment of a pair of point clouds with a low-overlap ratio using our method.

Evaluation on Outdoor Dataset We conducted experiments using KITTI, an outdoor dataset. We compared and analyzed traditional and learning-based methods including state-of-the-art methods. Traditional methods used FGR, SC^2-PCR, and RANSAC, while learning-based methods used DGR, DHVR, and PointDSC. Similar to other methods, the thresholds of RTE and RRE were set to 60 cm and 5° for successful registration, and experiments were conducted under the conditions set for the best performance.

As shown in Table 5, compared with the traditional method SC^2-PCR of the previous highest performance, our algorithm has lower errors in RTE and RRE, while they have the same registration recall. Our method has nearly twice as low errors in RTE than DGR and the state-of-the-art learning-based method PointDSC. However, compared to DHVR, our algorithm achieves 1% lower performance in the registration recall, but ours can have a higher recall through each threshold

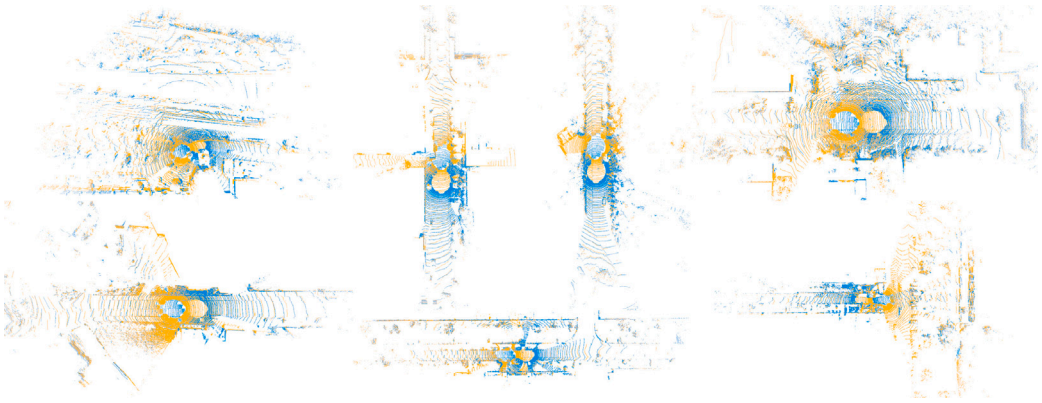


Fig. 6. Pairwise registration results on the KITTI odometry dataset.

Table 5
Registration recall for the KITTI dataset.

Method	Recall (% ↑)	RTE (cm ↓)	RRE (°↓)	Time (s)
FGR	0.2	40.7	1.02	1.42
SC^2-PCR	98.20	20.95	0.33	**0.31**
RANSAC	34.2	25.9	1.39	1.37
DGR	96.9	21.7	0.34	2.29
DHVR	**99.1**	19.8	0.29	0.83
PointDSC	98.02	21.03	0.33	0.45
Our	98.6	**11.43**	**0.25**	0.68

Table 6
Performance metrics on ETH dataset. Notably, all the techniques were trained using the 3DMatch dataset.

Method	Recall (% ↑)	RTE (cm ↓)	RRE (°↓)
SpinNet	72.07	5.55	1.295
DIP	61.31	14.61	1.923
GeoTransformer	4.87	21.06	0.91
FCGF + PointDSC	2.81	23.42	**0.57**
FCGF + DHVR	38.15	**14.02**	1.56
FCGF + Our	**77.42**	17.15	2.46

adjustment of both RTE and RRE. The qualitative evaluation of our algorithm in KITTI can be confirmed as shown in Fig. 6.

In the experiments conducted on the 3DMatch and the KITTI datasets, multiple bin sizes were tested, and the bin size that provided the best performance was finally selected. The selection was based on the results obtained by comparing the performance of different bin sizes. Evaluating various bin sizes through our experimental platform typically requires around three to four hours.

Evaluation on Unseen Dataset We performed the test using only the scenes from the ETH point cloud dataset that consists of 4 scenes mostly contained several partially overlapping scans of sparse and outdoor vegetation, such as trees and bushes (Pomerleau et al., 2012). All the methods are trained on 3DMatch datasets and tested on an unseen dataset. Following common practice we use only point clouds with an overlap of 30%. The thresholds of RTE and RRE were set to 30 cm and 5° for successful registration. As shown in Table 6, GeoTransformer and PointDSC achieve low recall which can be explained by the high RTE. DHVR also achieves a higher recall than GeoTransformer and PointDSC but achieves a lower recall than our method. The best recall is achieved by the proposed method.

4.4. Ablation study

We conducted an ablation study using 3DMatch to understand the important elements used in the experiment. We used the FCGF descriptor to generate initial correspondences. Table 7 shows the result

Table 7
A number of correspondences to construct triplets. Our method used the quartet.

Number of points	Recall (% ↑)	RTE (cm ↓)	RRE (°↓)	Time (s)
Triplet	92.4	6.61	**2.08**	**0.46**
Quintet	93.17	8.25	2.93	0.61
Hexad	91.42	9.23	3.44	0.64
Septet	85.44	10.44	4.29	0.58
Quartet (Ours)	**95.82**	**6.11**	2.14	0.53

Table 8
Ablation study on 3DMatch. **SQ**: Sampling using the quartet for triplet. **TL**: Transformation loss for transformation parameter training.

	SQ	TL	Recall (% ↑)	RTE (cm ↓)	RRE (°↓)
(1)			92.73	10.22	3.44
(2)		✓	93.2	9.88	3.32
(3)	✓		85.4	11.24	3.81
(4)	✓	✓	95.82	6.11	2.14

of applying a random extraction method according to the number of points in overlap correspondences for the triplet. As shown in Table 8, we can check the results according to the application of the module and the loss function.

Sampling for triplet Three points are required for each point cloud to obtain the rigid transformation parameter used for point cloud registration. To remove the value with a large distance between pairs before constructing the triplet, an experiment was conducted to construct the triplet by randomly extracting at least 4 to 7 points from correspondences $\mathcal{C}'$. For a fair comparison, the execution was carried out under the same implements. As shown in Table 7, we achieved higher performance than other methods in all results except the registration time of the quartet. When three points are randomly extracted from correspondences $\mathcal{C}'$, points with a large distance between pairs may be included. To solve that problem, we extracted four points and eliminated the case of large distances. When configuring the triplet, there is an improvement in registration performance. On the other hand, if five or more points are randomly extracted and configured as triplets, the probability of consisting only of pairs with a large distance to the extracted pairs increases. As a result, we confirmed that both RTE and RRE are increased, and the performance of registration recall is degraded.

Sampling module and loss functions. We construct the quartet in correspondences $\mathcal{C}^*$ to construct an efficient triplet by casting votes in the 6D Hough space representing the transformation parameter using a sampling module called SQ. In addition, we used transformation loss (TL) for accurate prediction of transformation parameters. As shown in Table 8, we conducted a comparative experiment based on the application of the quartet module and two losses. The experiments were conducted using the same hyperparameters. Experiment (1) means

DHVR, and experiment (4) means our experiment. Experiments (2) and (4) confirmed that the loss function TL is related to the prediction of the accurate transformation parameters.

5. Conclusion

We present a novel point cloud registration method for accurate and robust registration in large-scale 3D scenes, without using unrealistic assumptions. Our network uses correspondence points existing in overlap regions based on densely extracted features. It consists of Hough votes and uses a learnable refinement network. A core component of our method focuses on overlapping regions to generate correspondences and extract four points from correspondences to predict transformation parameters. Next, we construct a triplet to have the optimal consensus by removing pairs with a distance between corresponding points. We then obtain the transformation parameters of each element of the set of triplets and move them to a transformation parameter space constructed in a sparse manner to proceed with the vote to predict the final transformation parameters. Based on various experimental results, we achieved state-of-the-art performance without strong assumptions through matching with existent correspondences in overlapping regions.

CRediT authorship contribution statement

Jihoon Han: Conceptualization, Data curation, Methodology, Software, Validation, Writing – original draft. **Minwoo Shin:** Formal analysis, Investigation, Resources, Validation. **Joonki Paik:** Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Joonki Paik reports financial support was provided by Institute for Information and Communications Technology Planning and Evaluation (IITP) the National Research Foundation of Korea (NRF).

Data availability

Data will be made available on request.

Acknowledgments

This work was supported partly by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by Korea government(MSIT) [2021-0-01341, Artificial Intelligent Graduate School Program and 2021M3I1A1097911], and partly by National R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT [2020M3F6A1110350].

References

Almhdie, A., Léger, C., Deriche, M., Lédée, R., 2007. 3D registration using a new implementation of the ICP algorithm based on a comprehensive lookup matrix: Application to medical imaging. Pattern Recognit. Lett. 28 (12), 1523–1533.

Aoki, Y., Goforth, H., Srivatsan, R.A., Lucey, S., 2019. Pointnetlk: Robust & efficient point cloud registration using pointnet. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7163–7172.

Azuma, R.T., 1997. A survey of augmented reality. Presence: Teleoper. Virtual Environ. 6 (4), 355–385.

Bai, X., Luo, Z., Zhou, L., Chen, H., Li, L., Hu, Z., Fu, H., Tai, C.-L., 2021. Pointdsc: Robust point cloud registration using deep spatial consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15859–15869.

Bai, X., Luo, Z., Zhou, L., Fu, H., Quan, L., Tai, C.-L., 2020. 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. pp. 6359–6367.

Ballard, D.H., 1981. Generalizing the hough transform to detect arbitrary shapes. Pattern Recognit. 13 (2), 111–122.

Besl, P.J., McKay, N.D., 1992. Method for registration of 3-D shapes. In: Sensor Fusion IV: Control Paradigms and Data Structures, vol. 1611, SPIE, pp. 586–606.

Birdal, T., Busam, B., Navab, N., Ilic, S., Sturm, P., 2019. Generic primitive detection in point clouds using novel minimal quadric fits. IEEE Trans. Pattern Anal. Mach. Intell. 42 (6), 1333–1347.

Cao, A.-Q., Puy, G., Boulch, A., Marlet, R., 2021. PCAM: Product of cross-attention matrices for rigid registration of point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13229–13238.

Chen, Z., Sun, K., Yang, F., Tao, W., 2022. SC2-PCR: A second order spatial compatibility for efficient and robust point cloud registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13221–13231.

Chen, Z., Yang, F., Tao, W., 2021. DetarNet: Decoupling translation and rotation by siamese network for point cloud registration. arXiv preprint arXiv:2112.14059.

Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P., 2002. The trimmed iterative closest point algorithm. In: Object Recognition Supported By User Interaction for Service Robots, vol. 3, IEEE, pp. 545–548.

Chowdhury, S.A.H., Nguyen, C., Li, H., Hartley, R., 2021. Fixed-lens camera setup and calibrated image registration for multifocus multiview 3d reconstruction. Neural Comput. Appl. 33 (13), 7421–7440.

Choy, C., Dong, W., Koltun, V., 2020. Deep global registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2514–2523.

Choy, C., Gwak, J., Savarese, S., 2019a. 4D spatio-temporal ConvNets: Minkowski convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3075–3084.

Choy, C., Park, J., Koltun, V., 2019b. Fully convolutional geometric features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8958–8966.

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.

Geiger, A., Ziegler, J., Stiller, C., 2011. Stereoscan: Dense 3d reconstruction in real-time. In: 2011 IEEE Intelligent Vehicles Symposium. IV, IEEE, pp. 963–968.

Gower, J.C., 1975. Generalized procrustes analysis. Psychometrika 40 (1), 33–51.

Hough, P.V., 1962. Method and means for recognizing complex patterns.

Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., Schindler, K., 2021. Predator: Registration of 3d point clouds with low overlap. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4267–4276.

Johnson, J., Douze, M., Jégou, H., 2019. Billion-scale similarity search with GPUs. IEEE Trans. Big Data 7 (3), 535–547.

Kim, P., Chen, J., Cho, Y.K., 2018. SLAM-driven robotic mapping and registration of 3D point clouds. Autom. Constr. 89, 38–48.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lee, J., Kim, S., Cho, M., Park, J., 2021. Deep hough voting for robust global registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15994–16003.

Leibe, B., Leonardis, A., Schiele, B., 2008. Robust object detection with interleaved categorization and segmentation. Int. J. Comput. Vis. 77 (1), 259–289.

Leordeanu, M., Hebert, M., 2005. A spectral technique for correspondence problems using pairwise constraints.

Li, J., Zhang, C., Xu, Z., Zhou, H., Zhang, C., 2020. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration. In: European Conference on Computer Vision. Springer, pp. 378–394.

Limberger, F.A., Oliveira, M.M., 2015. Real-time detection of planar regions in unorganized point clouds. Pattern Recognit. 48 (6), 2043–2053.

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.

Pais, G.D., Ramalingam, S., Govindu, V.M., Nascimento, J.C., Chellappa, R., Miraldo, P., 2020. 3Dregnet: A deep neural network for 3d point registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7193–7203.

Pathak, K., Birk, A., Vaskevicius, N., Pfingsthorn, M., Schwertfeger, S., Poppinga, J., 2010. Online three-dimensional SLAM by registration of large planar surface segments and closed-form pose-graph relaxation. J. Field Robotics 27 (1), 52–84.

Pomerleau, F., Liu, M., Colas, F., Siegwart, R., 2012. Challenging data sets for point cloud registration algorithms. Int. J. Robot. Res. 31 (14), 1705–1711.

Qi, C.R., Litany, O., He, K., Guibas, L.J., 2019. Deep hough voting for 3d object detection in point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9277–9286.

Raffo, A., Romanengo, C., Falcidieno, B., Biasotti, S., 2022. Fitting and recognition of geometric primitives in segmented 3D point clouds using a localized voting procedure. Comput. Aided Geom. Design 97, 102123.

Romanengo, C., Raffo, A., Biasotti, S., Falcidieno, B., 2023. Recognizing geometric primitives in 3D point clouds of mechanical CAD objects. Comput. Aided Des. 157, 103479.

Rusinkiewicz, S., 2019. A symmetric objective function for ICP. ACM Trans. Graph. 38 (4), 1–7.

Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. In: 2009 IEEE International Conference on Robotics and Automation. IEEE, pp. 3212–3217.

Salas-Moreno, R.F., Newcombe, R.A., Strasdat, H., Kelly, P.H., Davison, A.J., 2013. Slam++: Simultaneous localisation and mapping at the level of objects. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1352–1359.

Salti, S., Tombari, F., Di Stefano, L., 2014. SHOT: Unique signatures of histograms for surface and texture description. Comput. Vis. Image Underst. 125, 251–264.

Wang, Y., Solomon, J.M., 2019. Deep closest point: Learning representations for point cloud registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3523–3532.

Yang, J., Li, H., Campbell, D., Jia, Y., 2015. Go-ICP: A globally optimal solution to 3D ICP point-set registration. IEEE Trans. Pattern Anal. Mach. Intell. 38 (11), 2241–2254.

Yang, H., Shi, J., Carlone, L., 2020. Teaser: Fast and certifiable point cloud registration. IEEE Trans. Robot. 37 (2), 314–333.

Yew, Z.J., Lee, G.H., 2020. Rpm-net: Robust point matching using learned features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11824–11833.

Yuan, W., Eckart, B., Kim, K., Jampani, V., Fox, D., Kautz, J., 2020. Deepgmr: Learning latent gaussian mixture models for registration. In: European Conference on Computer Vision. Springer, pp. 733–750.

Zeng, A., Song, S., Nieß ner, M., Fisher, M., Xiao, J., Funkhouser, T., 2017. 3Dmatch: Learning local geometric descriptors from rgb-d reconstructions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1802–1811.

Zhang, J., Singh, S., 2015. Visual-lidar odometry and mapping: Low-drift, robust, and fast. In: 2015 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 2174–2181.

Zhou, Q.-Y., Park, J., Koltun, V., 2016. Fast global registration. In: European Conference on Computer Vision. Springer, pp. 766–782.

Zhou, Q.-Y., Park, J., Koltun, V., 2018. Open3D: A modern library for 3D data processing. arXiv preprint arXiv:1801.09847.

Zinßer, T., Schmidt, J., Niemann, H., 2003. A refined ICP algorithm for robust 3-D correspondence estimation. In: Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429). Vol. 2, IEEE, pp. II–695.