

RESEARCH

Open Access



SHC: soft-hard correspondences framework for simplifying point cloud registration

Zhaoxiang Chen¹, Feng Yu^{1,2*} , Shuqing Liu¹, Jiacheng Cao¹, Zhuohan Xiao¹ and Minghua Jiang^{1,2}

The source code is available at
<https://github.com/ChenPointCloud/SHC>.

*Correspondence:
yufeng@wtu.edu.cn

¹ School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan 430200, China

² Engineering Research Center of Hubei Province for Clothing Information, Wuhan 430200, China

Abstract

Point cloud registration is a multifaceted problem that involves a series of procedures. Many deep learning methods employ complex structured networks to achieve robust registration performance. However, these intricate structures can amplify the challenges of network learning and impede gradient propagation. To address this concern, the soft-hard correspondence (SHC) framework is introduced in the present paper to streamline the registration problem. The framework encompasses two modes: the hard correspondence mode, which transforms the registration problem into a correspondence pair search problem, and the soft correspondence mode, which addresses this new problem. The simplification of the problem provides two advantages. First, it eliminates the need for intermediate operations that lead to error fusion and counteraction, thereby improving gradient propagation. Second, a perfect solution is not necessary to solve the new problem, since accurate registration results can be achieved even in the presence of errors in the found pairs. The experimental results demonstrate that SHC successfully simplifies the registration problem. It achieves performance comparable to complex networks using a simple network and can achieve zero error on datasets with perfect correspondence pairs.

Keywords: Point cloud registration, 3D point matching, Computer vision

1 Introduction

Point cloud registration is a fundamental and critical technology in the field of computer vision. It can be extensively applied in various domains, including 3D pose estimation [1–3], 3D reconstruction [4–6], and simultaneous localization and mapping (SLAM) [7, 8]. In recent years, the remarkable achievements of deep learning have attracted numerous researchers to apply it to point cloud registration. Within the realm of registration problems, a significant amount of research has been dedicated to designing complex networks capable of accurately calculating robust rigid transformations. For instance, OMNet [9] incorporates a mask mechanism to counteract the detrimental effects of partial overlap in point clouds. RIENet [10] introduces an inlier evaluation module to identify reliable correspondences. Predator [11] introduces an overlap attention block to facilitate the exchange of information between two point clouds. These methods exemplify the utilization of complex network designs to address specific challenges, such as handling partial overlap and exploring reliable

correspondences in point cloud registration. Undoubtedly, those works have achieved remarkable success in point cloud registration. However, they primarily concentrate on designing complex networks to achieve robust performance, often overlooking the simplification of the registration problem itself. In contrast, our approach focuses on simplifying the registration task to a correspondence pair search task. This simplification enables us to address the problem using a simple network involving the stacking of a few convolution blocks [12] and a multiplication operation as the output layer. By simplifying the problem, we can achieve satisfactory results in point cloud registration without the need for intricate network architectures.

In this work, we introduce the soft-hard correspondence (SHC) framework to simplify point cloud registration. SHC offers two modes: soft correspondences and hard correspondences. The hard correspondence mode simplifies registration by converting it into a correspondence pair search problem, which is then addressed using the soft mode. In the soft correspondence mode, represented by the Siamese network in Fig. 1, we use a simple architecture with a feature extractor and multiplication to explore point cloud pair relationships. The hard correspondence mode, represented by the full SHC framework in Fig. 1, includes additional components and processes beyond the Siamese network, transforming the registration task into a correspondence pair search problem. By combining both modes, SHC provides a comprehensive and flexible solution for efficient point cloud registration.

To summarize, the proposed SHC trains a network in the soft correspondence mode to search for correspondence pairs between two point clouds. Subsequently, the hard correspondence mode, encompassing the complete SHC framework, is utilized to calculate the final rigid transformation for registration. The main contributions of our work can be summarized as follows:

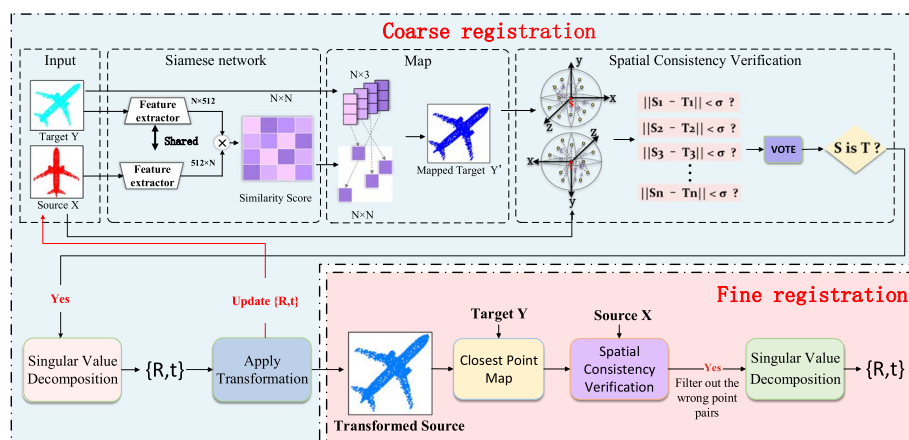


Fig. 1 The overview of the SHC framework. SHC extracts the feature descriptor for point clouds. Then, SHC calculates correspondence score M and uses it to gather correspondence pairs. The correspondence pairs are filtered by SCV and sent to SVD to calculate rigid transformation. The coarse registration iterates a few times to get coarse rigid transformation and then uses fine registration to refine the final result. During the iteration, the source point cloud undergoes rotation and translation based on the calculated R and t , resulting in a new source point cloud

1. Simplification of the point cloud registration problem: In this paper, SHC, which transforms the point cloud registration problem into a correspondence pair search problem, is introduced. We removed the optimization for R and t when calculating the loss because it would aggregate the errors from all points into the 3×3 rotation matrix and 3×1 translation vector.
2. Improves robustness: We have prior knowledge that the distance between two points on a rigid object does not change due to rotation. Based on this, we employ spatial consistency verification to let correspondence pairs mutually validate and assess their quality, removing pairs that do not support this prior knowledge. Our experiments demonstrate that accurate results can be achieved even if the found correspondence pairs contain fake correspondence pairs.
3. Reduction in the task complexity: SHC successfully simplifies the task, allowing for the utilization of a simple network architecture comprising only a few convolution blocks and an output layer to accomplish the new task. While complex networks may typically offer superior performance, our work demonstrates that by simplifying the task, a simple network can still achieve accurate results.

2 Related work

2.1 Correspondence-based registration method

Correspondence-based methods typically involve two steps: first, using a feature extraction network [12–14] to establish correspondences between points, and then using methods such as SVD [10, 15, 16] or RANSAC [11, 17] to solve for the rigid transformation. Many researchers have made significant improvements on this basis. Deng et al. [18] and Gojcic et al. [19] proposed to build better feature descriptors to search correspondence pairs. Deng et al. [18] implemented PPFNet, which uses point pair features to improve the quality of the initial input features. Shen et al. [10], Qin et al. [15] and Li et al. [20] used a weighted method to reduce the influence of incorrect predictions. However, weighted methods still retain some errors and limit the learning of feature embedding networks. RIENet [10] uses an inlier evaluation module to compute the weight of correspondence pairs. IDAM [20] uses hybrid features to compute the weights. GeoTransformer [15] builds multiple weight SVDs to compute multiple transformations and select the best transformation as the result. RGM [16] adopts the Hungarian algorithm to obtain hard correspondence pairs. However, using hard correspondence pairs without robust inlier evaluation modules will limit the registration performance. Wang and Solomon [21], Huang et al. [11] and Yew and Lee [22] used attention mechanisms to aggregate contextual information to extract discriminative feature extractors. DCP [21] utilizes attention mechanisms to extract more discriminative features. Predator [11] employs attention mechanisms to explore the overlapping regions. REGTR [22] uses attention mechanisms to predict whether a point belongs to the overlapping region and to find the corresponding points. Attention mechanisms are widely used in various networks, but their time and memory costs are not low. In addition, although embedding attention mechanisms in the network can improve the discriminative ability of the feature extractor, sometimes it provides no significant improvement. The other registration methods [23–25] also adopt correspondence pair searching methods.

2.2 Learning-based registration method

In the learning-based registration methods, a network is designed to compute the registration problem. PointLk [26, 27] integrates the Lucas and Kanada algorithm into deep learning for point cloud registration. FMR [28] aligns the global features of two point clouds to calculate the transformation parameter. OMNet [9] uses a mask mechanism to reduce the influence of outliers for registering partial overlapping point clouds. 3DSmoothNet [19] implements smoothed density value voxelization to process input data. 3DSmoothNet is robust for different sensor point clouds. FCGF [29] adopts a fully convolutional network to extract dense point cloud features. Regarding FMR, a new optimization goal for point cloud registration is proposed. Moreover, FMR computes transformations based on feature calculations and aligns the point clouds by matching their features. RPMNet [30] uses slack variables and weighted singular value decompositions (SVDs) to calculate transformation parameters and predicts annealing parameters to improve registration performance. In addition, Choy et al. [31], Gao et al. [32] and Zhu and Fang [33] also achieved remarkable performance in point cloud registration.

3 Methodology

In this section, we delineate the point cloud registration problem and elucidate how the SHC effectively addresses this challenge.

3.1 Problem description

The point cloud registration task involves determining a rigid transformation that aligns two point clouds. We will consider two point clouds, denoted as X and Y . Within these point clouds, there exist corresponding pairs (x_i, y_j) , where x_i belongs to Point Cloud X , and y_j belongs to Point Cloud Y . The relationship between the transformation parameters and the correspondence pairs can be described as follows:

$$R, t = \arg \min_{R, t} \sum ||Rx_i + t - y_j|| \quad (1)$$

where R and t denote the rotation matrix and translation vector, respectively. (x_i, y_j) denotes one of the correspondence pairs.

In practical scenarios, the input point clouds may not have a straightforward correspondence based on their indices alone. Therefore, we need a correspondence matrix M to establish a relationship between the two point clouds. The above equation can be described as follows:

$$R, t = \arg \min_{R, t} (RX + t - YM) \quad (2)$$

where X denotes the source point cloud, Y denotes the target point cloud, and M denotes the correspondence matrix, which describes the correspondence situation of two arbitrary points between X and Y . When M is known, the above equation has an analytical solution [34]. In contrast to fractal-wavelet analysis [35, 36], which decomposes the function into multiple functions, it merges multiple functions to ultimately obtain R and t . R and T are calculated from multiple correspondence pairs from the two point clouds,

and SVD combines the specific errors in the correspondence pairs, making it difficult for us to identify which specific pair is problematic. To address this issue, SHC employs a soft-hard correspondence framework, which abandons the direct search for the optimal R and t and instead focuses on finding matching point pairs.

Although deep learning methods have shown remarkable performance, it is challenging to achieve a 100% precision rate in identifying correspondence pairs. As a result, the utilization of the SHC method becomes crucial for achieving accurate registration performance, even in the presence of potential false correspondence pairs. While fractal-wavelet analysis [37–39] has many applications in the field of image processing, in registration, we have some prior knowledge that can be leveraged. For instance, the distance between two points will not change due to rotation, allowing SHC to employ simpler and more effective methods for outlier removal. We leverage the invariant property that the distance between two points does not change with rotation to discriminate the reliability of point pairs. However, it carries certain risks. When dealing with fractals, this property becomes ineffective due to the presence of self-similar geometric shapes within the fractal structure. Due to the lack of a fractal dataset, this paper did not delve further into the registration of fractal geometry. However, recent research on fractals [40] can offer new theoretical references for the registration of fractal geometry. Furthermore, the reconstruction of Shannon wavelet formula [41] also provides theoretical insights for point cloud decomposition.

3.2 SHC frame

The SHC is proposed to simplify the registration problem by translating it into a correspondence pair searching problem. After simplification, the difficulty of the task is reduced. On the one hand, for the registration problem, it is desirable for the results to be as close to the ground truth as possible, while for the correspondence pairs search problem, it is sufficient for the majority of the results to be accurate. On the other hand, the correspondence pairs search problem allows for the direct propagation of error gradients and feature extraction networks, avoiding the merging and offsetting of errors.

Regarding traditional deep learning methods [15, 42, 43], carefully designed networks are typically utilized to effectively address registration problems. In contrast, the network of SHC focuses on learning to differentiate correspondence pairs instead of learning the registration process itself. During the training step, the SHC method employs a soft correspondence mode, where a simple network is trained to discriminate between true and fake correspondence pairs. It is important to note that this mode alone is insufficient to solve the registration problem effectively. Therefore, during the evaluation step, the SHC method switches to a hard correspondence mode, utilizing the entire SHC framework to achieve precise and accurate registration results.

The complete SHC comprises two main components: coarse registration and fine registration. The objective of the coarse registration stage is to identify correspondence pairs on a global scale to estimate a coarse rigid transformation. The fine registration stage focuses on identifying correspondence pairs locally based on the estimated rigid transformation. Its purpose is to refine the registration outcome and reduce errors introduced during the coarse registration step.

3.2.1 Soft-hard correspondence

As previously mentioned, SHC simplifies the registration problem. We only need to train a simple network to solve the new problem. It is important to note that a simple network does not necessarily imply better performance. It is merely used to demonstrate that our method has indeed simplified the problem. Therefore, the structure of the network is very simple. The network consists of a few convolution blocks and a multiplication operation as the output layer. It can be described as follows:

$$S_{ij} = F_{x_i} * F_{y_j}^T \quad (3)$$

The convolution blocks extract the features of x_i in Point Cloud X and y_j in Point Cloud Y . Then, they are multiplied to obtain the similarity S_{ij} of two points (x_i and y_j). The convolutional blocks are responsible for extracting the features of x_i in Point Cloud X and y_j in Point Cloud Y . These features are then multiplied together to calculate their similarity, denoted as S_{ij} . During the training stage, the softmax function is utilized to convert the calculated similarity into similarity scores.

The trained network serves as a feature extraction network, enabling the generation of soft or hard correspondence pairs. During the training phase, to ensure differentiability, SHC applies the *softmax()* function to the similarity matrix S , producing the similarity scores matrix M . This matrix is then employed to compute the loss used for training the network. By multiplying the target point cloud with the matrix M , soft correspondence pairs are established to compute the rigid transformation. During the evaluation phase, since the trained model remains fixed, the differentiability of each operator no longer needs to be considered in the SHC framework. Consequently, the *argmax()* function is directly applied to each column of the similarity matrix S to identify the index of the maximum value in each column. These indices are then used to construct hard correspondence pairs. By utilizing these correspondence pairs, the SHC framework can accurately compute the rigid transformation.

We demonstrate the distinction between soft and hard correspondence modes by employing a base model with an appended SVD solver in the soft correspondence mode. The experimental results are depicted in Fig. 2. As observed in the figure, the performance of the trained model progressively improves throughout the training process. However, due to the robustness of SHC, it does not necessitate an impeccably precise feature embedding network. By the second epoch, the root mean square error of SHC is already below 1 degree, and by the third epoch, the performance of SHC exhibits no significant disparity from the best achievable performance. In essence, the enhancement of the base model's performance signifies that the network becomes more adept at searching for correspondence pairs during training. The flat curve of SHC illustrates that our method can yield accurate results even when the correspondence pair search problem is not perfectly solved.

3.2.2 Coarse registration

The coarse registration stage utilizes a network trained in soft correspondence mode to perform a global search for correspondence pairs. It is important to note that achieving zero error in deep learning networks is highly unlikely. Moreover, as the network's

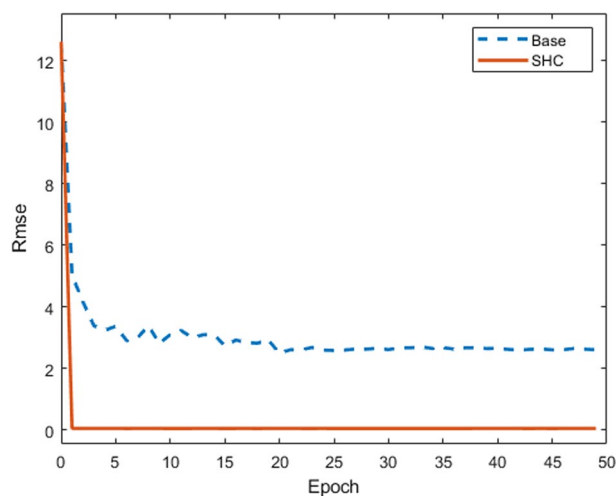


Fig. 2 The figure presented illustrates the performance of the trained model on the ICL-NUIM dataset throughout the training process. The horizontal axis depicts the number of epochs, while the vertical axis represents the root mean square error (RMSE) of rotation. The dotted line corresponds to the performance of the base model, which utilizes the soft correspondence mode concatenated with an SVD solver. The solid line represents the performance of the SHC framework

training objective is not explicitly focused on solving the registration problem, these errors can potentially impact the overall registration performance.

To address this issue, we introduce the spatial consistency verification (SCV) module, which ensures self-consistency among the correspondence pairs, resulting in reliable correspondences. The operation of the SCV module is illustrated in Fig. 1. We employ a voting-based approach to determine the authenticity of a point pair (S, T) as a true correspondence pair. Here, s_i represents the distance between Point S and the i -th point in the point cloud to which S belongs. Similarly, we obtain t_i through a similar operation. Assuming that most of the other point pairs are true correspondence pairs, it should hold that most of s_i is equal to t_i . By conducting a voting process and discarding correspondence pairs that contradict this assumption, we are able to obtain a set of self-consistent correspondence pairs. However, it is important to note that the effectiveness of the voting method depends on a sufficient number of reliable votes contributed by true correspondence pairs, as the votes provided by false correspondence pairs are considered unreliable.

To increase the number of accurate correspondence points, we adopt an iterative approach aimed at minimizing the initial pose differences between the two point clouds, thereby enhancing the reliability of the network. In contrast to the ICP algorithm [44], our approach limits the number of iterations to two or three. This is because the initial few iterations are notably effective in reducing the pose differences, while additional iterations do not yield substantial improvements in performance.

3.2.3 Fine registration

Since the coarse registration stage identifies correspondence pairs globally, there may be cases where points with similar feature descriptors are positioned far apart. Although the SCV module extracts a self-consistent subset of correspondence pairs, effectively

handling self-consistent yet similar local patches, such as two identical balls in the same scene, remains challenging. To address this issue, we incorporate fine registration to further refine the registration results. Among various deep learning approaches [21, 45], ICP is a widely used fine registration method. However, the iterative nature of ICP necessitates multiple iterations to approximate the optimal solution, leading to potential time consumption. In the fine registration stage of SHC, instead of relying on the iterative approximation approach of ICP, we leverage the SCV module to establish a self-consistent subset of correspondence pairs. This approach eliminates the need for multiple iterations and enables us to achieve optimized results through a single fine registration step.

Similar to ICP, we employ the closest point map (CPM) method to establish initial correspondence pairs. Subsequently, SHC utilizes the SCV module to obtain reliable correspondence pairs, which are essential for computing the final results. To align the point clouds based on the obtained correspondence pairs, we employ the SVD solver to solve for the rigid transformation. Another option for estimating the rigid transformation is RANSAC (random sample consensus), which is a robust algorithm capable of handling outliers and noisy data. RANSAC achieves this by iteratively selecting random subsets of correspondence pairs and estimating the transformation that best fits the pairs. However, RANSAC's reliance on random sampling makes it unreliable and necessitates hundreds or thousands of iterations, which contradicts our objective of simplifying the point cloud registration problem. Most importantly, the utilization of the SCV module to obtain reliable correspondence pairs enables us to employ the simpler SVD method.

3.3 Loss functions

Under ideal conditions, the problems of point cloud registration and correspondence pair search are mathematically equivalent. Knowledge of the correspondence pairs between two point clouds reveals their rigid transformation, and vice versa. However, in practice, the quality of the correspondence pairs does not always exhibit a strictly positive correlation with the performance of the point cloud registration. Figure 3 illustrates the relationship between correspondence loss and registration performance. Notably, the figure demonstrates that a lower correspondence loss does not necessarily lead to a lower registration error. In fact, in certain cases, a lower correspondence loss can even result in a higher registration error. Due to this inherent imperfection, many existing works [10, 21, 46] prioritize optimizing the final registration result. However, this approach can give rise to new problems. Achieving a lower registration error does not necessarily indicate that the trained network has learned a more effective feature descriptor. This is because the errors introduced during the feature extraction step are counteracted when calculating the final result. The feature error does not influence the registration errors in the same direction. For example, certain errors may lead to an overestimation of the computed angle compared to the ground truth, while other errors may result in an underestimation of the computed angle. If these errors reach an equilibrium state, it is possible to have correspondence pairs with errors yet still produce results that are close to the ground truth. Fortunately, SHC transforms the registration problem into a correspondence pair search problem, eliminating the reliance on error balancing to achieve accurate registration results. In SHC, our objective is to solve

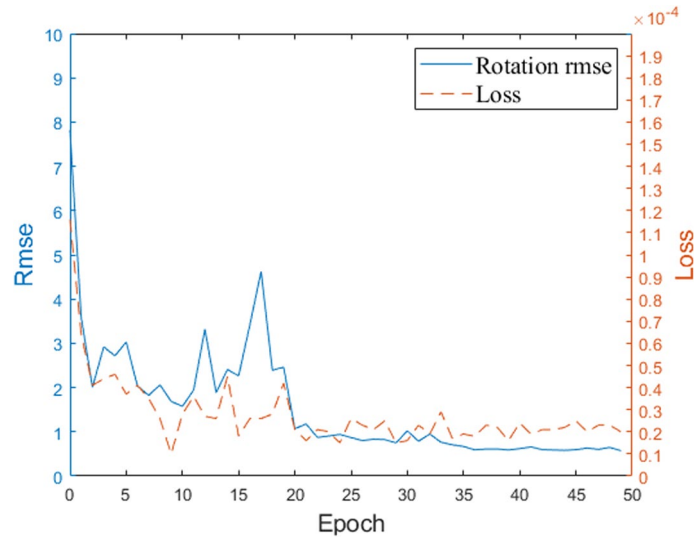


Fig. 3 This figure shows the loss and root mean square error of rotation during training. It is tested in dataset ICL-NUIM

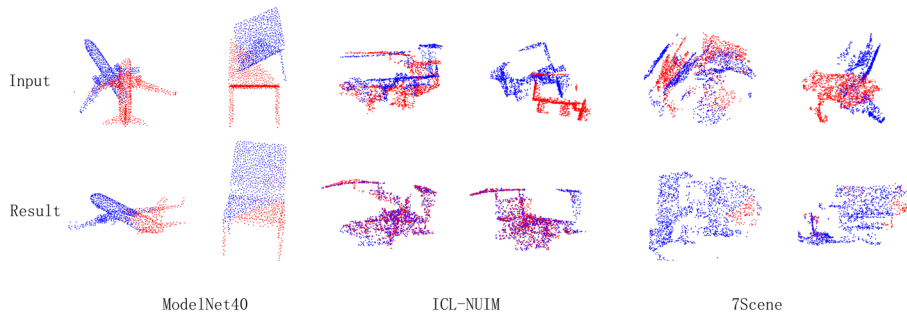


Fig. 4 Visualization results on ModelNet40, 7Scene and ICL-NUIM datasets

the correspondence pair search problem by minimizing the errors in correspondence. Hence, the loss function of SHC is formulated to minimize the correspondence loss. The loss function can be expressed as follows:

$$L = \sum_{i=1}^n \sum_{j=1}^n (1 - M_{ij}) \mathbb{I}(\|\hat{R}x_i + \hat{t} - y_j\| < \epsilon) \tag{4}$$

where M_{ij} denotes the similarity score of Point x_i and Point y_j , $\mathbb{I}(\cdot)$ is the indicator function that returns 1 if true, ϵ is the correspondence threshold, and \hat{R} and \hat{t} are the ground truth rotation matrix and translation vector, respectively.

4 Experiment

In this section, the SHC is implemented in multiple datasets to prove its performance (Fig. 4). Additionally, a few experiments are conducted to validate the rationality of its structural design. In these experiments, we compare classical traditional ICP [44] and Go-ICP [47], deep learning methods DCP [21], IDAM [20], CEMNet [48], and RIENet

[10]. The results of ICP, Go-ICP, IDAM and CEMNet are obtained from CEMNet [48], and DCP and RIENet are reproduced in the same environment as SHC. The experimental hardware environments include an NVIDIA GeForce RTX 2080 Ti and Intel(R) Core(TM) i7-9700k CPU.

4.1 Implement details

The SHC is implemented in PyTorch, and an ADAM optimizer is used for optimization. The initial learning rate is set to 0.001 and is multiplied by 0.1 at epochs of 20, 35, and 45. The total number of epochs is 50. During the training step, the batch sizes of training and testing are set to 16 and 4, respectively. In the evaluation step, the batch size for evaluation is set to 1. The correspondence threshold ϵ is set to 0.001.

4.2 Comparison evaluation on ModelNet40

To assess the performance of SHC, the experiment employs mean absolute error (MAE) and root mean square error (RMSE) metrics to quantify the rotation and translation errors. Both MAE and RMSE are utilized to evaluate the accuracy of the registration process, but RMSE is particularly sensitive to larger errors. Ideally, in a robust model, the values of RMSE and MAE should be relatively close. If there is a substantial difference between the RMSE and MAE values, it suggests that the registration performance varies significantly across different samples.

4.2.1 Dataset setting

The ModelNet40 dataset [49] is widely employed for point cloud registration tasks. It comprises 12,308 CAD models belonging to 40 different categories. In this experiment, a subset of 5,112 CAD models from the first 20 categories is used to train the SHC model. The experiment adopts the preprocessing approach introduced in RIENet [10], which involves removing 25% of the points from each point cloud to simulate partial overlapping. For the registration task, random rigid transformations are generated for each point cloud. The range of Euler angles is set to $[0, 45]$, and the translation range is set to $[-0.5, 0.5]$.

4.2.2 Clean point cloud

SHC first undergoes evaluation using clean point clouds consisting of 2468 CAD models. These clean point clouds do not contain Gaussian noise but exhibit partial overlapping. Moreover, the correspondence pairs in the clean point clouds are considered perfect correspondences, where the distance between these pairs becomes zero after applying the transformation to the source point cloud. This experiment demonstrates the upper limit performance of SHC under ideal conditions. It is important to note that deep learning methods inherently possess biases in the output of continuous values, making it nearly impossible to achieve a zero error. Conversely, traditional methods, while having the potential to attain zero error under ideal conditions, often struggle to achieve such ideal performance across the entire dataset. SHC addresses the registration problem by not directly utilizing the continuous values output by the deep network model for registration. It employs a soft correspondence mode to provide more ideal states and a hard correspondence mode to eliminate the bias in the network's output. As

Table 1 The registration performance on partial overlapping clean point cloud

Model	Clean			
	RMSE(R)	RMSE(t)	MAE(R)	MAE(t)
ICP	13.7952	0.0391	4.4489	0.0196
Go-ICP	14.7223	0.0328	3.5112	0.0127
DCP	9.3036	0.0936	6.3455	0.0615
IDAM	2.3384	0.0102	0.4711	0.0025
CEM	1.5018	0.0009	0.1385	0.0001
RIE	0.0062	0.0000	0.0018	< 0.0001
Our	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Bold indicates that it has the best performance

Table 2 The registration performance on partial overlapping unseen categories point cloud

Model	Unseen			
	RMSE(R)	RMSE(t)	MAE(R)	MAE(t)
ICP	14.7732	0.0351	3.5938	0.0132
Go-ICP	13.8322	0.0321	3.1579	0.0121
DCP	8.5182	0.0755	6.2885	0.0532
IDAM	2.1566	0.0151	0.6135	0.0037
CEM	1.1013	0.0020	0.0804	0.0002
RIE	0.0020	< 0.0001	0.0011	< 0.0001
Our	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Bold indicates that it has the best performance

shown in Table 1, SHC outperforms the runner-up approach significantly in all metrics. The near-zero error indicates that SHC possesses the capability to achieve zero error in most of the samples.

4.2.3 Unseen categories

To assess the generalization ability of SHC, we include the last 20 categories of CAD models that are distinct from the training set in this experiment. The trained SHC model is then tested on unseen class objects and compared with other methods. As shown in Table 2, the performance of SHC on unseen point clouds exhibits no significant difference compared to the clean point cloud scenario. These results demonstrate that SHC can achieve robust performance when dealing with point clouds from unseen classes.

4.2.4 Gaussian noise

In line with the settings of RIENet, we introduce Gaussian noise to the point cloud to simulate a noisy environment. As mentioned earlier, SHC simplifies the registration problem by transforming it into a correspondence pair search problem. This search problem does not necessitate a perfect resolution and allows for a certain tolerance toward errors. Consequently, SHC exhibits resilience toward additional errors introduced by noise. As depicted in Table 3, while the performance of some other methods

Table 3 The registration performance on partial overlapping point cloud with Gaussian noise

Model	Noise			
	RMSE(R)	RMSE(t)	MAE(R)	MAE(t)
ICP	12.5413	0.0398	4.2826	0.0184
Go-ICP	14.5225	0.0329	3.4252	0.0114
DCP	9.9544	0.0906	7.3685	0.0635
IDAM	3.5701	0.0206	1.0642	0.0066
CEM	2.2722	0.0014	0.3799	0.0008
RIE	0.0244	0.0001	0.0103	< 0.0001
Our	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Bold indicates that it has the best performance

significantly deteriorates, SHC continues to demonstrate robust performance in the presence of noise.

4.3 Comparison evaluation on 7Scene and ICL-NUIM

To further validate the performance of SHC, we employ the 7Scene dataset [50] and the ICL-NUIM dataset [51] for evaluation. The 7Scene dataset consists of seven indoor scenes, namely chess boards, fires, heads, offices, pumpkins, red kitchens, and stairs. Due to the powerful generalization capability of SHC, the model trained on the ModelNet40 dataset is able to successfully evaluate all 353 samples in the 7Scene dataset. The data processing for both the 7Scene and ICL-NUIM datasets follows the methodology outlined in RIENet [10], involving sampling 2048 points, removing 25% of the points, and applying random rigid transformations to the remaining points. Since the ICL-NUIM dataset does not provide perfect correspondence pairs, SHC cannot achieve zero registration error in this scenario. Nonetheless, SHC still achieves accurate results, and the small discrepancies between the RMSE and MAE indicate the stability of SHC's performance (Table 4).

4.4 The loss choice of SHC

In this experiment, three different loss functions are compared using the clean ModelNet40 dataset. For evaluating point correspondences, a point is considered an inlier if its closest distance to the corresponding point in another point cloud is less than 0.001. The correspondence loss is used to maximize the similarity score of inliers, the chamfer loss is used to minimize the nearest neighbor point distance between the transformed source point cloud and the target point cloud, and the transformation loss is used to minimize the difference between the calculated rigid transformation and the ground truth transformation. The other two losses primarily prioritize registration performance, and although they may yield fewer inliers, they still produce reliable rigid transformations. On the other hand, the correspondence loss emphasizes the search for inliers, leading to a substantial increase in the number of identified inliers, but it may not necessarily result in a significant improvement in the registration performance. Figure 5 illustrates that the rotation angle errors of these three loss functions are correlated with the number of identified inlier points. Indeed, the correspondence loss function is the only one that exhibits the ability to identify over 70% of the inlier points, whereas the other loss

Table 4 The registration performance on 7Scene and ICL-NUIM

Model	RMSE(R)	RMSE(t)	MAE(R)	MAE(t)
<i>ICL-NUIM</i>				
ICP	10.1247	0.3006	2.1484	0.0693
Go-ICP	1.5515	0.0601	0.6333	0.0241
DCP	6.7184	0.1516	4.7844	0.1096
IDAM	9.4539	0.3040	4.4153	0.1385
CEM	0.0821	0.0002	0.0211	0.0001
RIE	0.0654	0.0026	0.0440	0.0019
Our	0.0563	0.0021	0.0434	0.0016
<i>7Scene</i>				
ICP	19.9166	0.1127	7.5760	0.0310
Go-ICP	24.2743	0.0360	7.1068	0.0137
DCP	10.3565	0.0538	7.5818	0.0401
IDAM	10.5306	0.0539	5.6727	0.0303
CEM	0.1768	0.0012	0.0434	0.0002
RIE	0.0144	< 0.0001	0.0063	< 0.0001
Our	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Bold indicates that it has the best performance

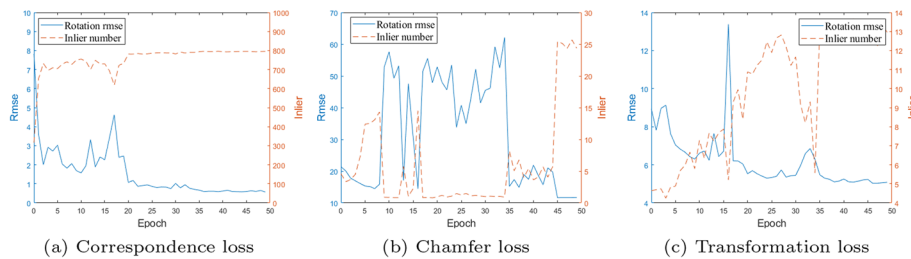


Fig. 5 This picture shows three loss functions performance during training. The full line is RMSE of rotation; the dotted line is the number of found inlier points

functions only manage to find less than 3% of the inlier points. While error counteraction can contribute to robust registration performance, it may also impede the learning of feature extraction capabilities.

4.5 Ablation study

In this section, ablation experiments are conducted to evaluate the effectiveness of the coarse and fine registration pipeline in SHC. The base model directly utilizes the built soft correspondences from the soft correspondence mode and applies an SVD solver to calculate the rigid transformation. As indicated in Table 5, incorporating the coarse registration pipeline leads to a significant improvement in performance. This outcome validates the efficacy of the SHC framework in transforming the point cloud registration problem into a correspondence pair search problem. Without leveraging the SHC framework, it would be challenging to achieve satisfactory results in point cloud registration. In detail, the trained model demonstrates the ability to robustly identify correspondence pairs, but achieving 100% accuracy can be challenging. The coarse registration process addresses the potential negative impact of these errors by translating

Table 5 The registration performance of different parts of SHC

Model	ICL-NUIM			
	RMSE(R)	RMSE(t)	MAE(R)	MAE(t)
Base	3.0375	0.1360	2.0708	0.0946
Base + coarse	0.0721	0.0029	0.0564	0.0022
Base + coarse + fine	0.0563	0.0021	0.0434	0.0016

the soft correspondence relationship into a hard correspondence relationship and building self-consistency correspondence pairs. Despite the coarse registration already producing a robust result, the fine registration process further enhances the registration performance. This finding suggests that the fine registration step can effectively refine the results.

5 Conclusion and future work

In this paper, we propose SHC, which simplifies the problem of point cloud registration by redefining it as a correspondence pair search problem. Through this simplification, the problem no longer requires perfect solutions and can be addressed using a simple network composed of a few convolution blocks and a multiplication operation. SHC offers two modes: soft correspondence mode and hard correspondence mode. In the hard correspondence mode, the registration problem is transformed into a correspondence pair search problem, while the soft correspondence mode trains a network to solve this new problem. The experimental results demonstrate that SHC successfully simplifies the registration problem. A simple network can achieve performance comparable to that of complex registration networks, and it has the ability to achieve zero error on ideal datasets. While SHC simplifies the registration problem and achieves comparable results with an extremely simple feature extractor network architecture, it is important to acknowledge that the feature extractor still plays a critical role, particularly for more complex point cloud registration tasks. In future work, we plan to design feature extractors that are more suitable for SHC to further enhance the registration performance.

Acknowledgements

We would like to express our sincere gratitude for the support received from the following funding sources: National natural science foundation of China (No. 62202346), Hubei key research and development program (No. 2021BAA042), China scholarship council (No. 202208420109), Wuhan applied basic frontier research project (No. 2022013988065212), MIIT's AI Industry Innovation Task unveils flagship projects (Key technologies, equipment, and systems for flexible customized and intelligent manufacturing in the clothing industry), and Hubei science and technology project of safe production special fund (No. SJZX20220908).

Author contributions

ZC designed the framework and implemented it in detail. FY provided research direction and professional guidance. SL contributed numerous suggestions regarding the feasibility of the proposed approach. JC and ZX conducted extensive literature review and provided significant assistance to the paper. MJ supervised the writing and revision of the manuscript.

Funding

This work was supported by national natural science foundation of China (No. 62202346), Hubei key research and development program (No.2021BAA042), China scholarship council (No.202208420109), Wuhan applied basic frontier research project (No. 2022013988065212), MIIT's AI Industry Innovation Task unveils flagship projects (Key technologies, equipment, and systems for flexible customized and intelligent manufacturing in the clothing industry), and Hubei science and technology project of safe production special fund (No.SJZX20220908).

Availability of data and materials

The dataset used in this paper is widely employed in point cloud registration, and relevant references have been cited in this study. The citations provide information on accessing the dataset used.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 14 June 2023 Accepted: 22 December 2023

Published online: 17 January 2024

References

1. A. Kamel, B. Liu, P. Li, B. Sheng, An investigation of 3D human pose estimation for learning Tai Chi: a human factor perspective. *Int. J. Hum. Comput. Interact.* **35**(4–5), 427–439 (2019)
2. A. Kamel, B. Sheng, P. Yang, P. Li, R. Shen, D.D. Feng, Deep convolutional neural networks for human action recognition using depth maps and postures. *IEEE Trans. Syst. Man Cybern. Syst.* **49**(9), 1806–1819 (2018)
3. P. Zhang, L. Zheng, Y. Jiang, L. Mao, Z. Li, B. Sheng, Tracking soccer players using spatio-temporal context learning under multiple views. *Multimed. Tools Appl.* **77**, 18935–18955 (2018)
4. S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S.M. Seitz, R. Szeliski, Building Rome in a day. *Commun. ACM* **54**(10), 105–112 (2011)
5. X. Ren, L. Lyu, X. He, W. Cao, Z. Yang, B. Sheng, Y. Zhang, E. Wu, Biorthogonal wavelet surface reconstruction using partial integrations, in *Computer Graphics Forum*, vol. 37, ed. by H. Hauser, P. Alliez (Wiley, Hoboken, 2018), pp.13–24
6. X. Li, B. Sheng, P. Li, J. Kim, D.D. Feng, Voxelize facial reconstruction using deep neural network, in *Proceedings of Computer Graphics International 2018* (2018), pp.1–4
7. J.-E. Deschaud, IMLS-SLAM: scan-to-model matching based on 3d data, in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2018), pp. 2480–2485
8. J. Zhang, S. Singh, Loam: Lidar odometry and mapping in real-time, in *Robotics: Science and Systems*, vol. 2 (2014), pp. 1–9
9. H. Xu, S. Liu, G. Wang, G. Liu, B. Zeng, Omnet: learning overlapping mask for partial-to-partial point cloud registration, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 3132–3141
10. Y. Shen, L. Hui, H. Jiang, J. Xie, J. Yang, Reliable inlier evaluation for unsupervised point cloud registration, in *AAAI* (2022), pp. 1–9
11. S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, K. Schindler, Predator: registration of 3D point clouds with low overlap, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 4267–4276
12. Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein, J.M. Solomon, Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph. (TOG)* **38**(5), 1–12 (2019)
13. C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: deep learning on point sets for 3D classification and segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 652–660
14. H. Thomas, C.R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, L.J. Guibas, Kpconv: flexible and deformable convolution for point clouds, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 6411–6420
15. Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, K. Xu, Geometric transformer for fast and robust point cloud registration, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 11143–11152
16. K. Fu, S. Liu, X. Luo, M. Wang, Robust point cloud registration framework based on deep graph matching, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 8893–8902
17. H. Wang, Y. Liu, Z. Dong, W. Wang, You only hypothesize once: point cloud registration with rotation-equivariant descriptors, in *Proceedings of the 30th ACM International Conference on Multimedia* (2022), pp. 1630–1641
18. H. Deng, T. Birdal, S. Ilic, Ppfnet: Global context aware local features for robust 3D point matching, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 195–205
19. Z. Gojcic, C. Zhou, J.D. Wegner, A. Wieser, The perfect match: 3D point cloud matching with smoothed densities, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 5545–5554
20. J. Li, C. Zhang, Z. Xu, H. Zhou, C. Zhang, Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration, in *European Conference on Computer Vision* (Springer, 2020), pp. 378–394
21. Y. Wang, J.M. Solomon, Deep closest point: learning representations for point cloud registration, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 3523–3532
22. Z.J. Yew, G.H. Lee, Regtr: end-to-end point cloud correspondences with transformers, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 6677–6686
23. H. Yang, L. Carlone, A polynomial-time solution for robust registration with extreme outlier rates (2019). arXiv preprint [arXiv:1903.08588](https://arxiv.org/abs/1903.08588)
24. H.M. Le, T.-T. Do, T. Hoang, N.-M. Cheung, Sdrsac: semidefinite-based randomized approach for robust point cloud registration without correspondences, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 124–133
25. T. Dong, Y. Zhao, Q. Zhang, B. Xue, J. Li, W. Li, Multi-scale point cloud registration based on topological structure. *Concurr. Comput. Pract. Exp.* (2022). <https://doi.org/10.1002/cpe.6873>
26. Y. Aoki, H. Goforth, R.A. Srivatsan, S. Lucey, Pointnetlk: robust and efficient point cloud registration using pointnet, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 7163–7172
27. X. Li, J.K. Pontes, S. Lucey, Pointnetlk revisited, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 12763–12772

28. X. Huang, G. Mei, J. Zhang, Feature-metric registration: a fast semi-supervised approach for robust point cloud registration without correspondences, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 11366–11374
29. C. Choy, J. Park, V. Koltun, Fully convolutional geometric features, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 8958–8966
30. Z.J. Yew, G.H. Lee, Rpm-net: robust point matching using learned features, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 11824–11833
31. C. Choy, W. Dong, V. Koltun, Deep global registration, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 2514–2523
32. J. Gao, Y. Zhang, Z. Liu, S. Li, Hdrnet: high-dimensional regression network for point cloud registration, in *Computer Graphics Forum*. ed. by H. Hauser, P. Alliez (Wiley, Hoboken, 2022)
33. J. Zhu, Y. Fang, Reference grid-assisted network for 3D point signature learning from point clouds, in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2020), pp. 211–220
34. K.S. Arun, T.S. Huang, S.D. Blostein, Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.* **5**, 698–700 (1987)
35. R.C. Guido, F. Pedroso, R.C. Contreras, L.C. Rodrigues, E. Guariglia, J.S. Neto, Introducing the discrete path transform (DPT) and its applications in signal analysis, artefact removal, and spoken word recognition. *Digit. Signal Process.* **117**, 103158 (2021)
36. M.V. Berry, Z. Lewis, J.F. Nye, On the Weierstrass–Mandelbrot fractal function. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **370**(1743), 459–484 (1980)
37. L. Yang, H. Su, C. Zhong, Z. Meng, H. Luo, X. Li, Y.Y. Tang, Y. Lu, Hyperspectral image classification using wavelet transform-based smooth ordering. *Int. J. Wavel. Multiresolut. Inf. Process.* **17**(06), 1950050 (2019)
38. X. Zheng, Y.Y. Tang, J. Zhou, A framework of adaptive multiscale wavelet decomposition for signals on undirected graphs. *IEEE Trans. Signal Process.* **67**(7), 1696–1711 (2019)
39. E. Guariglia, Primality, fractality, and image analysis. *Entropy* **21**(3), 304 (2019)
40. E. Guariglia, Harmonic Sierpinski gasket and applications. *Entropy* **20**(9), 714 (2018)
41. E. Guariglia, S. Silvestrov, Fractional-wavelet analysis of positive definite distributions and wavelets on $d'(c)$, in *Engineering Mathematics II: Algebraic, Stochastic and Analysis Structures for Networks, Data Classification and Optimization* (Springer, 2016), pp. 337–353
42. W. Chen, H. Li, Q. Nie, Y.-H. Liu, Deterministic point cloud registration via novel transformation decomposition, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 6348–6356
43. Z. Chen, F. Yang, W. Tao, Detarnet: decoupling translation and rotation by Siamese network for point cloud registration, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36 (2022), pp. 401–409
44. P.J. Besl, N.D. McKay, Method for registration of 3-D shapes, in *Sensor Fusion IV: Control Paradigms and Data Structures*, vol. 1611 (SPIE, 1992), pp. 586–606
45. P. Kadam, M. Zhang, S. Liu, C.-C.J. Kuo, R-pointhop: a green, accurate, and unsupervised point cloud registration method. *IEEE Trans. Image Process.* **31**, 2710–2725 (2022)
46. G.D. Pais, S. Ramalingam, V.M. Govindu, J.C. Nascimento, R. Chellappa, P. Miraldo, 3dregnet: a deep neural network for 3D point registration, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 7193–7203
47. J. Yang, H. Li, Y. Jia, Go-icp: Solving 3d registration efficiently and globally optimally, in *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 1457–1464
48. H. Jiang, Y. Shen, J. Xie, J. Li, J. Qian, J. Yang, Sampling network guided cross-entropy method for unsupervised point cloud registration, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 6128–6137
49. Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3d shapenets: a deep representation for volumetric shapes, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 1912–1920
50. J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, A. Fitzgibbon, Scene coordinate regression forests for camera relocalization in RGB-D images, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 2930–2937
51. A. Handa, T. Whelan, J. McDonald, A.J. Davison, A benchmark for RGB-D visual odometry, 3D reconstruction and slam, in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2014), pp. 1524–1531

Publisher's Note

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