Point Cloud Registration Based on Global and Local Feature Fusion

Wenping Ma, Mingyu Yue, Yongzhe Yuan, Yue Wu, Hao Zhu, Licheng Jiao

Point cloud registration based on global and local feature fusion

Wenping Ma[1], Mingyu Yue[1], Yongzhe Yuan[2], Yue Wu[2⋆], Hao zhu[1], and Licheng Jiao[1]

[1] School of Artificial Intelligence, Xidian University, Xi'an, China.
[2] School of Computer Science and Technology, Xidian University, Xi'an, China.
{wpma,lchjiao}@mail.xidian.edu.cn, myyue@stu.xidian.edu.cn, {ywu, haozhu}@xidian.edu.cn, magicyyz111@gmail.com

Abstract. Global feature extraction and rigid body transformation estimation are two key steps in correspondences-free point cloud registration methods. Previous approaches only utlize the global information while the local information is ignored. Moreover, global and local information may play different roles on multiple point clouds. In this paper, we verify the sensitivity of different types of point clouds to global and local information. We conducted extensive experiments on the ModelNet40 dataset by the SGLF-DQNet. Through the experimental results, we summarize the point cloud structure of the sensitivity to global and local features in the correspondence-free point cloud registration task.

Keywords: point cloud registration · featrue extraction · correspondence-free methods.

1 Introduction

As the most primitive three-dimensional data, the 3D point cloud[2] can accurately reflect the real size and shape structure of objects and gradually become a data form that visual perception[3, 4] depends on. Point cloud registration aims at finding a rigid transformation to align one point cloud to another. The most common algorithm for it is the Iterative Closest Point (ICP)[5].

Numerous deep learning registration methods[6–8] have been proposed to provide accurate alignments which improve the defects of traditional methods[9]. The baseline of correspondence-free methods is to extract the global features to regress the motion parameters of rigid transformation[10]. Early deep-learning methods usually utilize voxel-based[11] or pointcnn-based[12] feature extractors. PointNet[13] was firstly proposed for classification[14] and segmentation tasks, many correspondence-free registration architectures are utilize it for feature extraction Our previous work has proposed a network called GLF-DQNet[15], which fuse the global and local information and improve the defects of the PointNet.

⋆ indicates corresponding author

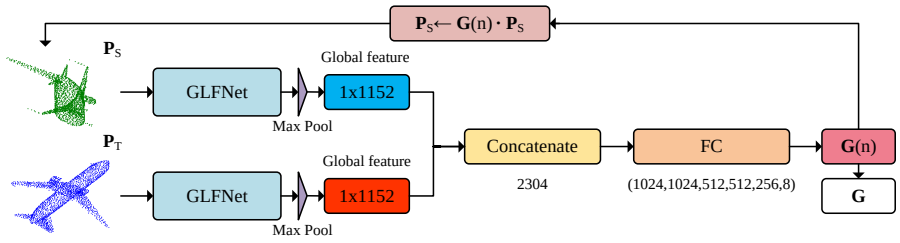


Fig. 1. GLF-DQNet Architecture

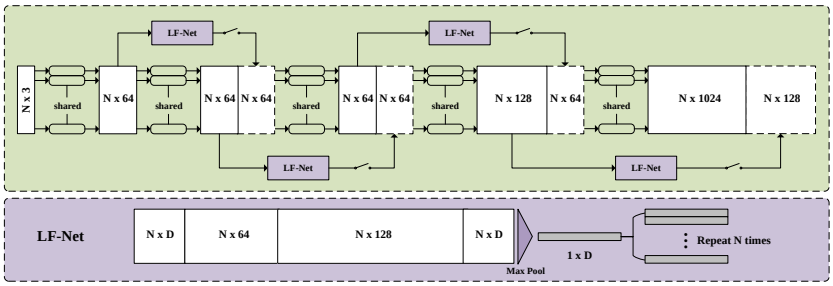


Fig. 2. SGLFNet Architecture

However, various samples may have different sensitivity to global and local features. We want to analyse the impact of global and local characteristics on correspondences-free registration tasks. To test this idea, based on GLF-DQNet, we add switches to the feature extractor to separate global and local features and do experiments on ModelNet40 dataset. Experiments show that for different types of samples, global features and local features play different leading roles.

Contributions: Our main contributions can be summarized as follows:

- We propose a new feature extraction network SGLFNet, which can make up for the deficiency of local features and separate local features and global features at the same time.
- We verify the sensitivity of different categories of point clouds to global and local features.
- We summarize the structure characteristics of global feature high sensitivity point cloud and local feature high sensitivity point cloud respectively.

2 Related Work

2.1 Traditional registration method

ICP is considered as the most standard algorithm to solve point cloud registration. ICP iteratively extracts the closest source point to each template point, and

refine the transformation by least-squares optimization. The varients of ICP improve the defects and enhances the registration accuracy. However, these methods are sensitive to the initial point cloud and hard to integrate these models into the end-to-end deep-learning pipelines.

2.2 Deep-learning registration method

PointNet is emerged as a milestone which is apply deep-learning to point cloud directly and solve the unordered arrangement problems. Amount of models utilize PointNet as the global feature extractor such as PointNetLK , PCRNet and CorsNet, etc. These learning-baesd methods all achieve higher accuracy and lower complexity than the traditional approaches. However, PointNet-based techniques loss the local information when extrat the feature.

2.3 GLF-DQNet

As shown in Figure 1, the model consists of a feature extraction network and the pose estimation using the dual quaternion. The source and template point clouds through the GLFNet and max-pooling to obtain the global feature with local information. The rigid transformation is calculated by using the dual quaternion[16]. In GLFNet, features obtained from sharing MLP are entered into LF unit. In each layer, the output of the MLP and LF unit are connected and used as input in the next shared MLP.

3 Method

3.1 Problem Statement

The input of network are two unsorted 3D point pairs $\mathbf{P_s}$ and $\mathbf{P_t}$. The aim of registration is to caculate the best linear transformation $\mathbf{G} \in SE(3)$ to align the source $\mathbf{P_s}$ to the template $\mathbf{P_t}$ as follows:

$$\mathbf{P}_{\mathrm{T}} = \mathbf{G} \cdot \mathbf{P}_{\mathrm{S}}. \tag{1}$$

The transformation $\mathbf{G}$ consisting of two parts: rotation matrix $R \in SO(3)$ and translation vector $t \in R^3$. In correspondences-free methods, the predicted rigid transformation $\mathbf{G}_{est}$ is usually generated as:

$$\mathbf{G}_{est} = \begin{pmatrix} \mathbf{R}_{est} & \mathbf{t}_{est} \\ \mathbf{0} & 1 \end{pmatrix} \tag{2}$$

$\mathbf{G}_{est}$ and $\mathbf{t}_{est}$ are the predicted rotation matrix and translation vector

3.2 SGLFNet Network

On the basis of GLFNet, we add switches in the network. It can be seen from Figure 2, there are switches behind each LF uint. When the switch is turned on, the network only extracts global features. But when the switch is turned off, the local information will be taken into account at the same time. In experiments, we can directly decide whether to extract local features by controling the switch.

Table 1. GLFNet verification experiments

Method	RMSE(R)	RMSE(R)	RMSE(t)	RMSE(t)
DirectNet	19.4791	5.0934(↓14.3857)	0.0122	0.0660(↑0.0538)
CorsNet	16.2356	12.7044 (↓3.5312)	0.0070	0.9608(↑0.9538)
PCRNet	3.8837	2.7966(↓0.2709)	0.0064	0.0029(↓0.0035)

3.3 Loss function

$\mathcal{L}_1$ is used to minimize the distance between the corresponding points in the source and the template point cloud. $\mathcal{L}_2$ is used to reduce the difference between estimated and ground truth transformation matrices. $\mathcal{L}_3$ combined two of them.

$$\mathcal{L}_1 = \frac{1}{N} \sum_{p_s \in \mathbf{P}_\mathrm{S}} \min_{p_t \in \mathbf{P}_\mathrm{T}} \|p_s - p_t\|_2 + \frac{1}{N} \sum_{p_t \in \mathbf{P}_\mathrm{T}} \min_{p_s \in \mathbf{P}_\mathrm{S}} \|p_s - p_t\|_2 \tag{3}$$

$$\mathcal{L}_2 = \|(\mathbf{G}_{est})^{-1}\mathbf{G}_{gt} - \mathbf{I}_4\|_F \tag{4}$$

$$\mathcal{L}_3 = \mathcal{L}_1 + 0.007\mathcal{L}_2. \tag{5}$$

4 Experiments

4.1 Experiments on GLFNet

In order to verify the effectiveness of local information for the correspondences-free registration task, we replace the feature extractors of DirectNet, CorsNet and PCRNet with GLFNet for experiments. From Table 1, when GLFNet is used as the feature extractor, the registration effect of these methods has been significantly improved after fusing local features.

4.2 Experiments on local registration

In order to verify the sensitivity of different categories to global and local features, we do experiments on 40 categories in ModelNet40 dataset respectively. By analyzing the point data of these categories, we find that point clouds which are more sensitive to global features usually include several special categories:

(a) Symmetry structure: As shown in Figure 3, *chair* and *toilet* are classical symmetrical structure. When extracting local information, it usually causes confusion of symmetrical part information, resulting in deviation.

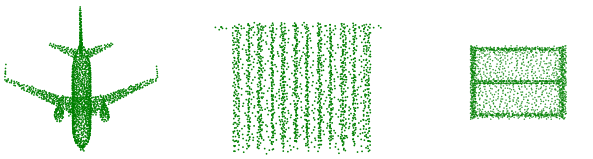
(b): Repetitive structure: It can be seen in Figure 3, *bookshelf* and *dresser* are been as the repetitive structure. For such samples, the existence of repeated structure will make the local information extracted from multiple points very similar, so it will cause confusion of information in different parts.

(c): Uniform distribution structure: Uniform distribution structure can be divided into compact structure and sparse structure. Compact structure like

(a) chair (b) toilet (c) bookshelf (d) dresser

Fig. 3. Visualization of symmetry structure and repetitive structure.

airplane and *curtain* in Figure 4 . When local information is incorporated, each point is equivalent to splicing the information of all points, resulting in the redundancy of individual point information. Sparse structure like *glass_ box* and *sink* in Figure 4. When extracting local information from the same number of points, the local features extracted from sparse point clouds often contain only one-sided information.



(a) ariplane (b) curtain (c) glass_box

Fig. 4. Visualization of uniform distribution structure.

4.3 Experiments on global registration

To verify the sensitivity of different categories to global and local features, we do experiments on 40 different categories respectively. The training set includes 40 categories, and the testing set only contains individual categories each time.

(a) flower_pot (b) piano (c) plant (d) stairs (e) vase

Fig. 5. Visualization of five special categories.

Global feature registration In this experiment, we turn on all the switches in the feature extractor. The GLFNet can only extract the global information.

The results are presented in Table 2, in ModelNet40 dataset, for 25 categories, the RMSE of rotation and translation can reach the optimal value at the same time when only extract the global features. For these categories, local features may become redundant information, thus reducing the registration results. There are

Table 2. Global feature experiments

Label	Global RMSE(R)	Global RMSE(t)
airplane	1.4680	0.0025
bathtub	1.1102	0.0021
bed	0.8038	0.0019
bench	1.8115	0.0015
bookshelf	1.2594	0.0033
car	1.2865	0.0018
chair	1.4376	0.0020
curtain	1.6539	0.0030
desk	1.6202	0.0025
dresser	0.9293	0.0020
door	0.9933	0.0018
flower_pot	4.2845	0.0027
glass_box	1.7052	0.0018
laptop	1.6977	0.0014
mantel	1.2350	0.0031
monitor	1.6495	0.0019
piano	2.8199	0.0023
plant	4.5942	0.0022
range_hood	1.7030	0.0026
sink	1.7412	0.0018
sofa	1.1124	0.0016
stairs	4.5999	0.0025
toilet	1.4200	0.0027
vase	4.5887	0.0025
wardrobe	1.1882	0.0018

five special types which are very poor when only global features are extracted, and will not be improved when local information is integrated. From the Figure 5, it is clearly to see the structure of them is very irregular and the distribution of them is very uneven.

Global feature with local feature registration In order to observe the influence of local features on different types of point clouds, we turn off all the switches to extract the global features and local features together. We found that there are 15 categories will improve the effect after incorporating the local features. Table 3 lists the results of only extracting global features and fusing local features and global features.

We believe that global information and local information have various effects on different kinds of point clouds. Therefore, we adjust the same weight of the

Table 3. The 1st and 2nd column means the result for global registration. The 3rd and 4th column means the decreasing value for global fusing local feature registration. The 5th and 6th column means the decreasing value after giving appropriate weights.

Label	RMSE(R)	RMSE(t)	RMSE(R) ↓	RMSE(t) ↓	RMSE(R) ↓	RMSE(t) ↓
bottle	4.9081	0.0031	0.3457	0.0015	0.4158	0.0007
bowl	4.9378	0.0033	0.2084	0.0010	1.9012	0.0011
cone	4.6873	0.0024	0.2200	0.0002	0.7047	0.0003
cup	4.7784	0.0023	0.2529	0.0001	0.6875	0.0001
guitar	3.3302	0.0034	1.1238	0.0013	1.1550	0.0015
keyboard	1.7053	0.0027	0.4463	0.0010	0.5886	0.0010
lamp	5.0006	0.0028	0.2366	0.0002	0.7944	0.0009
night_stand	2.4431	0.0021	0.1482	0.0006	1.0880	-0.0014
person	3.7108	0.0025	0.6800	0.0000	0.7742	0.0003
radio	2.9787	0.0033	0.3884	0.0007	1.0192	0.0015
stool	4.1324	0.0025	0.1730	0.0012	0.2453	0.0000
table	1.2904	0.0022	0.1755	0.0001	0.3507	-0.0005
tent	2.6244	0.0038	0.3316	0.0011	0.8698	0.0015
tv_stand	2.6446	0.0028	0.5681	0.0005	0.8877	0.0005
xbox	2.3177	0.0027	0.4837	0.0005	0.6106	0.0005

original global module and local module to the model adaptive weight. Through training, the weight that best matches the registration task is obtained, so as to obtain higher accuracy. The registration result is displayed in Table 3.

5 Conclusion

In this paper, we summarize the point cloud structure with better global feature effect and better local feature fusion effect. We carried out experiments by SGLF-DQNet, which can separate global information and local information. Through experiments, we found that near rotation or reflection symmetry structure, repetitive structure and uniform distribution structure are more sensitive to the global feature. Local uniform distribution is more sensitive to the global feature fusing the local information.

References

1. T. Jost and H. Hugli. A multi-resolution scheme icp algorithm for fast shape registration. In *International Symposium on 3d Data Processing Visualization & Transmission*, 2008.
2. Qian Wang and Min-Koo Kim. Applications of 3d point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced Engineering Informatics*, 39:306–319, 2019.
3. Y. Wu, J. W. Liu, C. Z. Zhu, Z. F. Bai, Q. G. Miao, W. P. Ma, and M. G. Gong. Computational intelligence in remote sensing image registration:a survey. *International Journal of Automation and Computing*, 18(1):17, 2021.

4. Wenping Ma, Na Li, Hao Zhu, Licheng Jiao, Xu Tang, Yuwei Guo, and Biao Hou. Feature split–merge–enhancement network for remote sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022.
5. Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In *Robotics: Science and Systems*, 2009.
6. Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 3212–3217, 2009.
7. Y. Wang and J. M. Solomon. Prnet: Self-supervised learning for partial-to-partial registration. 2019.
8. J. Y. Zi and G. H. Lee. 3dfeat-net: Weakly supervised local 3d features for point cloud registration. *Springer, Cham*, 2018.
9. Sofien Bouaziz, Andrea Tagliasacchi, and Mark Pauly. Sparse iterative closest point. In *Computer Graphics Forum*, pages 113–123, 2013.
10. Johannes Pöppelbaum and Andreas Schwung. Predicting rigid body dynamics using dual quaternion recurrent neural networks with quaternion attention. *arXiv preprint arXiv:2011.08734*, 2020.
11. G. N. Nikhil, M. Meraz, and M. Javed. *Automatic On-Road Object Detection in LiDAR-Point Cloud Data Using Modified VoxelNet Architecture*. Computer Vision and Image Processing, 2021.
12. Y. Li, R. Bu, M. Sun, and B. Chen. Pointcnn. 2018.
13. Vinit Sarode, Xueqian Li, Hunter Goforth, Yasuhiro Aoki, Animesh Dhagat, Rangaprasad Arun Srivatsan, Simon Lucey, and Howie Choset. One framework to register them all: pointnet encoding for point cloud alignment. *arXiv preprint arXiv:1912.05766*, 2019.
14. Wenping Ma, Na Li, Hao Zhu, Kenan Sun, Zhongle Ren, Xu Tang, Biao Hou, and Licheng Jiao. A collaborative correlation-matching network for multimodality remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022.
15. Yue Wu, Yongzhe Yuan, Mingyu Yue, Maoguo Gong, HaoLi, Mingyang Zhang, Wenping Ma, and Qiguang Miao. Feature mining method of multi-dimensional information fusion in point cloud registration. *Journal of Computer Research and Development*, (59), 2022.
16. Ben Kenwright. A beginners guide to dual-quaternions: what they are, how they work, and how to use them for 3d character hierarchies. In *Proceedings of International Conference on Computer Graphics, Visualization and Computer Vision*, pages 1–13, 2012.