INDOOR SCENE REGISTRATION BASED ON KEY POINTS SAMPLING AND HIERARCHICAL FEATURE LEARNING

Mengchi Ai[1,2] ,Chun Liu[1,*], Hua Shen[2], Fanjin Cheng[1]

[1] College of Surveying and Geo-Informatics, Tongji Unversity, 200092 Shanghai, China - (aimengchi, liuchun, cfj) @tongji.edu.cn
[2] Information Sciences and Technology, The Pennsylvania State University, State College, USA - (mxa1097, hqs5468)@psu.edu

KEY WORDS: Indoor Scene, LiDAR Mapping, Point Cloud Registration, PointNet, Hierarchical Feature Descriptor

ABSTRACT:

PointNet has been widely considered as a popular representation for unstructured point clouds with the aim of classification and segmentation. To date, recent researches represent the limitation of the PointNet to pose estimation and alignment of real environment, due to the low performance in pattern learning to complex scenes. This paper presents an end-to-end deep learning method for point clouds registration of indoor environment. The proposed method involves three steps. Firstly, feature pre-processing extracts the key-points by adaptive Harris 3D algorithm and generate the local group by point grouping. Second, hierarchical feature learning network is trained to describe the local group as feature descriptors. Finally, loss function between feature descriptor is trained. The key contribution is that we innovatively use the key-points to generate multi-layer feature vector, which can provide the contextual local features of the indoor environment. The results shows that our method achieves comparable registration accuracy to the present state-of-art geometric methods in the indoor environment. We comprehensively validate the accuracy of our approach using S3DIS dataset. The high accuracy demonstrates that our method can be used in point clouds registration accurately.

1. INTRODUCTION

In recent years, three-dimensional (3D) mapping in indoor environment based on point clouds has received considerable critical attention. It has been an important data source for indoor 3D model and information visualization, which are an integral part of indoor service such as geo-hazard monitoring, urban asset management, and so on (Wang et al., 2019). Managing such indoor structures for timely maintenance, smooth operation, and safety can be quite challenging without up-to-date spatial information of the structural conditions and space use. Individual point clouds from single scanning position cannot provide the whole environment due to the limitation such as moving objects and long distance. It is necessary to register point clouds and obtain detailed 3D mapping from multi-position scanned point clouds.

Point cloud registration is a basic and important technology in spatial information. Considering that feature descriptor is one of the main factors in the process of registration, many researchers proposed different kinds of feature space for point clouds registration. Traditional methods to obtain the completed and detailed point clouds are mainly implemented by matching the geometric pairs of points and calculating the transformation. Geometric feature, such as ICP (Iterative Closest Point) and its variants, establish the point corresponding and performing a least squares optimization. However, these algorithms are sensitive to the quality of initialization and known to be susceptible to local minima since it is difficult to explicitly establish closest point correspondences due to the data noise. In addition, many of these descriptors do not work well in the real environment due to the noisy and the low density of the point clouds (Yew and Lee, 2018). Thus, developing the registration algorithm with higher feature space that can be used in the 3D mapping is necessary.

Inspired by the process of the deep learning for image-based tasks, such as moving object recognition and image understanding, several feature learning algorithms are proposed after the popularity of deep learning, which are proposed based on the properties of point clouds, such as PointNet, PointCNN, PointNet++ (Li et al., 2018, Qi et al., 2017a, Qi et al., 2017b). However, the unique aspects of point clouds limit the performance of registration and enhance the complexity of registration, including the lack of local feature description and efficient feature analysis. Thus, there is with great challenges to apply the training network for registration.

In this paper, we design a hierarchical learned feature to register point clouds. Firstly, to improve the accuracy of registration, the key points are extracted by an adaptive Harris3D algorithm. To generate the grouping for training network, neighbourhood points are selected around the key points with various size. Mini-PointNet is used as the training network to extract the feature vector. Multi-layer of the feature learning composes the feature descriptor. Transformation is trained by the LK algorithm to achieve the registration (Aoki et al., 2019). Our contributions mainly include two points: Firstly, for the input of mini-PointNet training network, key points sampling, which considered as the feature pre-processing is proposed, can be applied to the indoor point clouds with multi-objects. Secondly, the feature descriptor with hierarchical structure is constructed for registration, which improves the performance of feature learning.

2. RELATED WORK

2.1 Classical Handed-crafted Feature

Survey work from (Pomerleau et al., 2015) provides a comprehensive review of traditional registration algorithms. Classical handcrafted features are proposed before deep learning aiming to find the correspondence between target and source point clouds. The design of these features are mainly based on the geometric knowledge of the 3D point clouds (Besl and McKay, 1992, Magnusson et al., 2007, Yang et al., 2013). Some algorithms are designed by describing the geometry of each point locally. For example, in the (Zhong, 2009), feature points are selected by the principle direction or the unique curvatures and matched through descriptor. 3D-SIFT focuses on reducing the influence of scale

[*] Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-177-2020 | © Authors 2020. CC BY 4.0 License.

by using Difference-of-Gaussian (DoG) and representing the difference of the intensity values(Rusu and Cousins, 2011a). PFH, FPFH consider the feature histograms and use surface normal to describe the patch around each key point (Rusu et al., 2008, Rusu et al., 2009, Tombari et al., 2010). However, around the key points, these descriptors will fall into the spatial bin. Some handcrafted features are designed by describing the surface model. For example, Rotational Projection Statistics (RoPS) calculates the scatter matrix lying on the surface and obtaining the distribution of projected points on the 2D planes (Guo et al., 2013). However, it requires the surface data and can not be applied to the raw point clouds. More evaluation of the handcrafted features can be found in the review (Hänsch et al., 2014). Based on the evaluation, the handed-crafted feature works well for the high-quality surface and point clouds with low-noisy and high-density. However, features are unstable and sensitive to the the number of scans. As a result, point clouds in real world with noisy and lack of points number will not work well (Yew and Lee, 2018).

2.2 Learned Feature

With the development of deep learning, learned 3D feature descriptor are widely used to describe the unstructured point clouds. Some of them operate on the depth image, while others operate on the point clouds directly (Zeng et al., 2017, Kehl et al., 2016). For example, PointNet uses point clouds as input to realize classification and segmentation. The structure of training network follows the neural networks and realizes theoretical insight into raw point clouds (Qi et al., 2017a). As the extension of PointNet, PointNet++ is proposed to improve the performance of complex environment, by hierarchically extracting feature with local feature (Qi et al., 2017b). However, the inherent lack of structure presents difficulties in using point clouds registration directly in deep learning architectures.

Some researches have focused on developing deep leaning feature for the purpose of point clouds registration. For example, a deep learning feature descriptor extracted by a Siamese network is proposed to register mobile point clouds in the indoor environment. However, the descriptor use RANSAC to reduce the wrong matching points. The method also requires another refinement step using ICP to improve the accuracy (Zhang et al., 2019). PPFNet proposes the pairs of point clouds feature and global context to improve the descriptor (Deng et al., 2018). PointNetLK considered the PointNet as the "imaging function" and designed the modified Lucas Kanade (LK) algorithm as the loss function to minimize the distance between the candidate point clouds (Aoki et al., 2019). This work provides us a good intuition that classical imaging matching algorithm can be used as the loss function for the point clouds training. However, the performance of the various approaches does not consider different level of feature description (Groß et al., 2019). Despite the good performance and achievements of these works, none of the work consider both related feature with key points and comprehensive learning structure for point clouds registration. Thus, it is with great challenges and potential to apply the learning features for matching and registration.

3. METHOD

Our work can be considered as the design of deep learned feature with feature analysis and the application in the indoor environment. In section 3.1, we introduce the innovation and mathematics of the proposed algorithm. In section 3.2, the derivation of feature pre-processing is introduced. In section 3.3, we describe the hierarchical feature structure and training model used for the point cloud alignment.

3.1 Overview

Let ϕ denotes the feature function. ϕ presents the $R^{3\times N} \rightarrow R^{K}$. For an input point cloud P, $\phi(P)$ will obtain a K-dimensional feature vector descriptor. The Multi-Layer Perceptron (MLP) is operated to each point, then the output is the feature vector, with K-dimension. Following the PointNet ++, the ϕ is designed with multi-layer to extract both local and global feature (Qi et al., 2017b).

The registration process is formulated as follows. Let $\mathbf{P}_{\mathcal{S}}$, $\mathbf{P}_{\mathcal{T}}$ be two groups of input data, the source point clouds and target point clouds, respectively. The $\mathbf{T}$, $\mathbf{T} \in SE(3)$, which represents the rigid-transform, is the best aligns from source $\mathbf{P}_{\mathcal{S}}$ to target $\mathbf{P}_{\mathcal{T}}$. The alignment process can be described as finding $\mathbf{T}$ such that $\phi\left(\mathbf{P}_{\mathcal{T}}\right) = \phi\left(\mathbf{T} \cdot \mathbf{P}_{\mathcal{S}}\right)$.

In order to compute ΔT each time, an iterative optimization solution is designed as equation.

$$\phi\left(\mathbf{P}_{\mathcal{S}}\right) = \phi\left(\mathbf{P}_{\mathcal{T}}\right) + \frac{\partial}{\partial \xi}\left[\phi\left(\mathbf{T}^{-1} \cdot \mathbf{P}_{\mathcal{T}}\right)\right] \xi \tag{1}$$

Where Jacobian J will be denoted as $\mathbf{J} = \frac{\partial}{\partial \xi}\left[\phi\left(\mathbf{G}^{-1} \cdot \mathbf{P}_{\mathcal{T}}\right)\right]$, $\mathbf{J} \in SE(6)$. For each J, Jacobian can be approximated by a finite difference gradient, which calculated by the equation:

$$\mathbf{J}_i = \frac{\phi\left(\exp\left(-t_i \mathbf{T}_i\right) \cdot \mathbf{P}_{\mathcal{R}}\right) - \phi\left(\mathbf{P}_{\mathcal{T}}\right)}{t_i} \tag{2}$$

Where t_i is the infinitesimal perturbations of the parameters ξ. R is the generate of the exponential map with twist parameters. J^{+} is the Moore-Penrose inverse of J.

$$\mathbf{P}_{\mathcal{S}} \leftarrow \Delta\mathbf{T} \cdot \mathbf{P}_{\mathcal{S}} \quad \Delta\mathbf{T} = \exp\left(\sum_i \xi_i \mathbf{R}_i\right) \tag{3}$$

The transformation matrix will the re-computation with the looping function, using equation 3. Then a new source data will be updated by calculating with the new transformation matrix. The final pose estimation T is the composition of each iterative loop, as equation 4. The iterative computation is based on the threshold for ΔT.

$$\mathbf{T} = \Delta\mathbf{T}_n \cdot \ldots \cdot \Delta\mathbf{T}_1 \cdot \Delta\mathbf{T}_0 \tag{4}$$

3.2 Feature Pre-processing

In order to reduce the noisy in the point cloud data, a pre-process is designed including statistical filtering and voxel filter (Zhang et al., 2019, Rusu and Cousins, 2011a). The statistical filtering is used to remove the noise points and voxel filter is used to reduce the resolution. The process of the point cloud registration is to calculate the transformation for the coordinate alignment. Since the transformation matrix can be calculated from the several matching pairs of points between the source and target point clouds, it is more efficient to use the most informative points than all the points. Compared with the farthest point sampling (FPS) or random sampling, key points has better performance of the feature extraction given the same number of key points. Key points sampling is considered as the feature analysis in this paper.

In the review of key points sampling, several researches show that the Harris 3D method is robust to several transformations

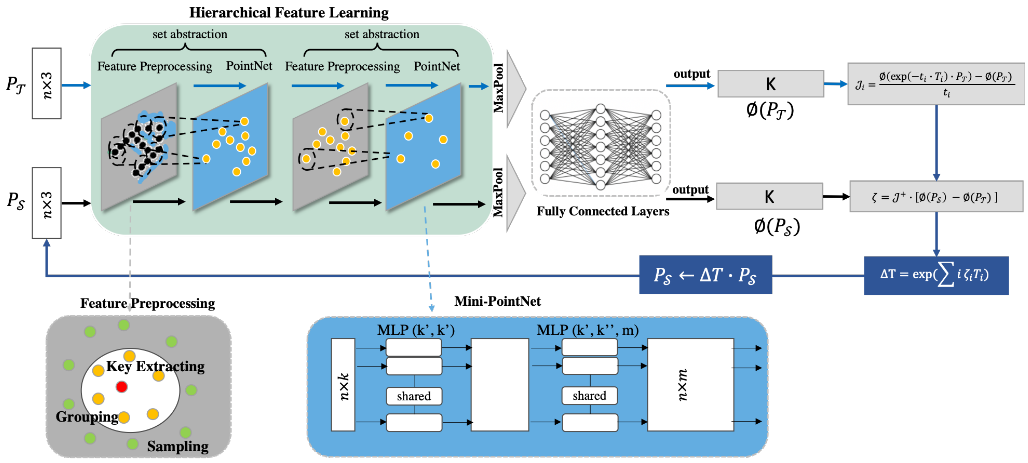
Hierarchical Feature Learning
set abstraction
set abstraction
Feature Preprocessing
PointNet
Feature Preprocessing
PointNet
$P_\mathcal{T}$
$n\times3$
$P_\mathcal{S}$
$n\times3$
MaxPool
MaxPool
Fully Connected Layers
output
output
K
K
$\emptyset(P_\mathcal{T})$
$\emptyset(P_\mathcal{S})$
$\mathcal{J}_i = \frac{\emptyset(\exp(-t_i \cdot T_i)\cdot P_\mathcal{T}) - \emptyset(P_\mathcal{T})}{t_i}$
$\zeta = \mathcal{J}^+ \cdot [\emptyset(P_\mathcal{S}) - \emptyset(P_\mathcal{T})]$
$\Delta T = \exp(\sum i\,\zeta_i T_i)$
$P_\mathcal{S} \leftarrow \Delta T \cdot P_\mathcal{S}$
Feature Preprocessing
Key Extracting
Grouping
Sampling
Mini-PointNet
MLP (k', k')
MLP (k', k'', m)
$n\times k$
shared
shared
$n\times m$

Figure 1. Point clouds input source $\mathbf{P}_\mathcal{S}$and target $\mathbf{P}_\mathcal{T}$ are passed through feature learning and a MLP to compute the feature vectors $\phi\left(\mathbf{P}_\mathcal{S}\right)$ and $\phi\left(\mathbf{P}_\mathcal{T}\right)$. The jacobian matrix $\mathbf{J}$ is computed by using the$\phi\left(\mathbf{P}_\mathcal{T}\right)$. Pose information Δ $\mathbf{T}$ will be updated incrementally if the value is higher than the thresh. During the training, one lose function is used, which is based on the difference between the estimated transform and the ground truth transform.

including noise, local scaling and presence of holes (Guo et al., 2014, Hänsch et al., 2014). In this paper, the adaptive Harris 3D method is used to extract the key points from source and target point clouds (Sipiran and Bustos, 2011). If one point is selected as the key points, the cluster of neighbourhood points will be considered as the local patch to represent points. The selection of the neighbourhood point is shown in the Algorithm 1.

Algorithm 1: Neighbourhood definition

```
Result: Size of neighbourhood points, $N$
initialization;
$N$ = 0; K = 4; $\delta = 0.025$
while $p \in P$ do
    Calculate the shortest path as:
    Calculate the Delaunay$Q$;
end
ring$_k(p) = \{q \in Q' || shortestpath(p, q)| = k\}$
while $k \leq K$ do
    if $d_{ring}\left(v, ring_k(p)\right) \geq \delta \& d_{ring}\left(v, ring_{k-1}(p)\right) \leq$
      $\delta$ then
        $N$ = $N$+1;
    else
        $N$ = $N$
    end
end
```

$$z = f(x, y) = \frac{p_1}{2}x^2 + p_2xy + \frac{p_3}{2}y^2 + p_4x + p_5y + p_6 \quad (5)$$

Based on the neighbourhood set for each point, given a point p, $p \in \mathbf{P}_\mathcal{S}$, $p \in \mathbf{P}_\mathcal{T}$, the neighbouring points are translated to fit a quadratic surface based on the equation 5. Then the derivatives of $f(x, y)$ is calculated in the point. A symmetric matrix E is defined using the derivatives of this function, as the equation 6. Then the highest Harris responses will be selected as the key points. The neighbourhood points will be selected as the local patches.

$$E = \begin{pmatrix} p_4^2 + 2p_1^2 + 2p_2^2 & p_4p_5 + 2p_1p_2 + 2p_2p_3 \\ p_4p_5 + 2p_1p_2 + 2p_2p_3 & p_5^2 + 2p_2^2 + 2p_3^2 \end{pmatrix} \quad (6)$$

3.3 Hierarchical Feature Learning

The feature descriptor is composed of a number of feature layers to achieve the hierarchical structure (Qi et al., 2017b). For each feature layer, an $N * (d + C)$ matrix is used as the input that represents N points with each point composes of d-dimensional coordinates and C-dimensional point feature. It outputs a $N' \times (d + C')$ matrix. The N' represents the sub-sampled points with d-dimensional coordinates and extracted C'-dimensional point feature. The layers of the descriptor will be introduced in the following paragraphs.

In the sampling layer, given input points$\{x_1, x_2, \ldots, x_n\}$, a subset of points $\{x_1, x_2, \ldots, x_m\}$ are extracted using the section 3.2, so that x_{xy} is the most informative point for each layer. Comparing with random sampling or farthest point sampling (FPS), this method provides more useful information to point cloud registration. The output of the sampling layer is N, which represents N selected points.

In the grouping layer, the input to this layer is $N * (d + C)$ with $N * d$ - dim coordinates and C - dim feature. It outputs several points set with a size of $N' \times K \times (d + C)$. Each point set corresponds to a local region for PointNet to convert the point into one fixed length feature vector. The K varies to adapt to points set. Neighbourhood points are selected in the step of feature pre-processing. Compared with the other methods, such as K nearest neighbor search or other threshold method, it has a better performance on requiring the local pattern (Jiang et al., 2018).

In the PointNet layer, points set with a size of $N' \times K \times (d + C)$ is the input. The output size is $N' \times (d + C')$, which is the basic building block for the local pattern learning from PointNet (Qi et al., 2017a). The function can be summarized as follows:

$$f\left(x_1, x_2, \ldots, x_n\right) = \gamma\left(MAX_{i=1,\ldots,n}\left\{h\left(x_i\right)\right\}\right) \quad (7)$$

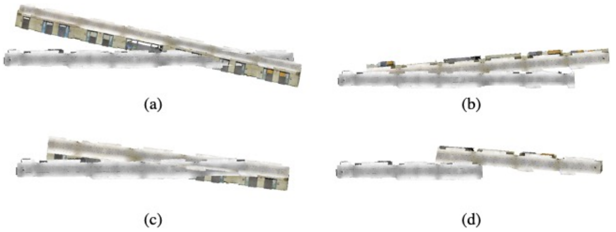

Figure 2. Different overlapping with the initial transformation. (a)ratio of overlapping is 95%, (b)ratio of overlapping is 75%, (c)ratio of overlapping is 50%, (d)ratio of overlapping is less than 30%. The point clouds in gray scale is target point clouds, while the point clouds in RGB is the source scale.

Where $\{x_1, x_2, \ldots, x_n\}$ represents the unordered point clouds, $x_i \in R^{(d)}$. The set function is designed to transform the points set to a feature vector. γ and h are the multi-layer perceptron (MLP) networks.f is the set function and invariant to the input source and target point clouds.

In the PointNet, maximum pooling function, average pooling function and weighted sum pooling function are used as the symmetric pooling function, following the MLP operation (Qi et al., 2017a), to realize the permutation invariance and unordered point clouds. In this paper, maximum pooling function will be used as the pooling function.

4. EXPERIMENTAL RESULTS

4.1 Experiment Design

We experiment with various type of objects with different real indoor scenarios. Stanford S3DIS indoor dataset is used to generate the source and target point clouds (Armeni et al., 2017). To evaluate the performance in the indoor environment, we demonstrate the use of proposed method to estimate the transformation in the Area 1 from S3DIS dataset. The dataset contains the corresponding semantic annotations and global XYZ images as well as surface normals.

To evaluate our method, we discuss the effect of different ratio of overlapping. To prepare the target point clouds and source point clouds, we implement the rotation and translation on the test data with different ratio of overlapping. Initial rotation and translation for test are in the range of [0, 5] meter and [0,90] degree. Figure 2 shows the different overlapping and initial transformation as the example. For evaluation purpose, three algorithm including ICP, NDT and Go-ICP are considered as the base line (Besl and McKay, 1992, Yang et al., 2015). The compared algorithms are implemented with the same point clouds without any additional process.

For each training dataset, we prepare the h5 file for the point clouds and property, respectively. For each input data, the size of training and testing data is $N * 4096$. Since we only consider the feature of geometry, only coordinate value is used and normalized from XYZ into $X'Y'Z'$. For parameters setting, the epochs, batch size, learning rate and the momentum are set to 200, 16, 0.01, 0.9. The experiments are implemented on a single GPU with Tensorflow 1.70.

4.2 Experiment Result

Point clouds are pre-processing by detecting key points and grouping. This step is implemented by PCL and Open3D (Rusu and Cousins, 2011b, Zhou et al., 2018). Figure 3 shows the points sampling (partial data in Conference_Room_1 from S3DIS), with the radius is 0.1. As shown in Figure 4, compared with other neighbourhood selection algorithm, the adaptive method will determine the size of neighbourhood points without a constant value.

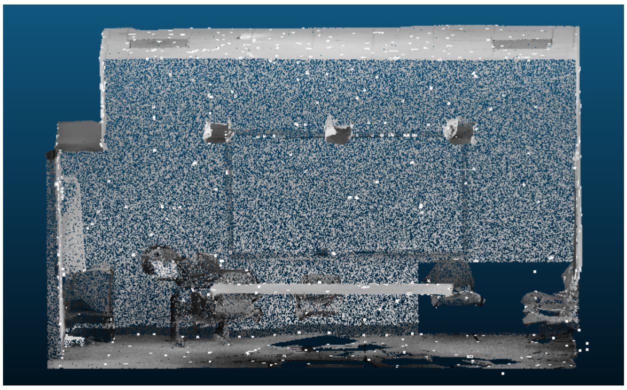

Figure 3. Feature pre-processing. Harris key points are selected and shown in the white point. The raw point clouds are shown in the gray scale.

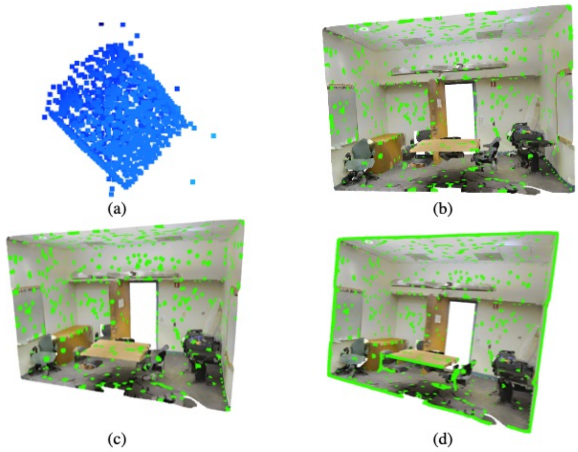

Figure 4. The group of neighbourhood points are selected from the point clouds around the key points. (a) Key points (b) Adaptive neighbourhood selection. (c) KNN neighbourhood selection. (d) Radius neighbourhood selection.

Different scenario	Our method	ICP	Go-ICP
Hallway	1.32	4.5	1.43
Conference Room	1.53	1.8	4.28

Table 1. Accuracy of different type of indoor environment (cm)

Two standard indoor environments are selected to evaluate our method. The ratio of overlapping between target and source point clouds is 100 %. Figure 5, 6 show the registration results conference and hallway, comparing to the ICP, NDT and Go-ICP. As shown in figure 5, 6 and Table 1, our method can get good performance in two standard indoor environments.

To evaluate the effect of different ratio of overlapping, the root mean square error (RMSE) is calculated, as shown in Table 2. For ICP, the RMSE is calculated based on corresponding points after registration. For our method, with the true transformation between the target and source point clouds, the RMSE is calculated as equation.

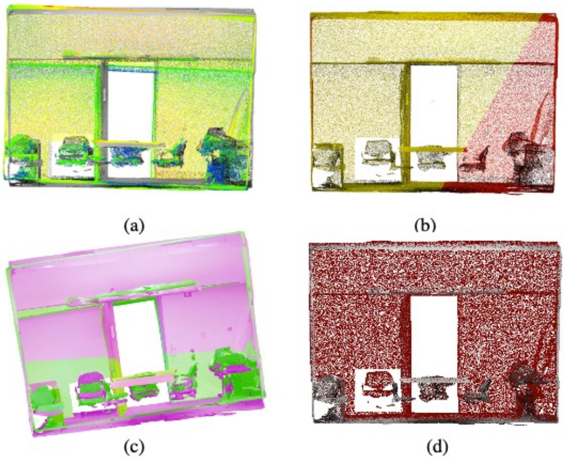

Figure 5. The conference environment. (a) is our method. (b)is the result of ICP. (c) is the result of NDT. (d) is the result of Go-ICP.

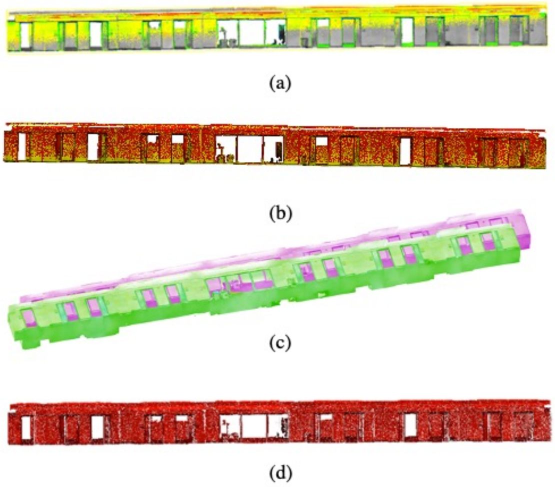

Figure 6. The hallway environment. (a) is our method. (b)is the result of ICP. (c) is the result of NDT. (d) is the result of Go-ICP.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|T_{st} \cdot S - T_{our} \cdot S\|^2}{N}} \tag{8}$$

Where N is the number of the corresponding points. T_{st} is the known transformation matrix between source and target point clouds. S represents the source point clouds.

Different overlapping	Our method	ICP
95%	1.3	4.5
75%	3.6	21.7
50%	5.4	21.9
Less than 30%	10.2	-

Table 2. Accuracy of different ratio of overlapping (cm)

5. CONCLUSION

A learned feature registration algorithm with point pre-processing and hierarchical training network is proposed in this paper. Adaptive Harris 3D algorithm is used to detect the key points and the hierarchical feature descriptor obtains the feature of the extracted points. The results show that our approach achieve good accuracy and computational efficiency with different ratio of overlapping in the indoor environment. In the future, we will evaluate on more dataset and scenes to improve the accuracy of the algorithm.

REFERENCES

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

Armeni, I., Sax, S., Zamir, A. R. and Savarese, S., 2017. Joint 2d-3d-semantic data for indoor scene understanding. arXiv preprint arXiv:1702.01105.

Besl, P. J. and McKay, N. D., 1992. Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures, Vol. 1611, International Society for Optics and Photonics, pp. 586–606.

Deng, H., Birdal, T. and Ilic, S., 2018. Ppfnet: Global context aware local features for robust 3d point matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 195–205.

Groß, J., Ošep, A. and Leibe, B., 2019. Alignnet-3d: Fast point cloud registration of partially observed objects. In: 2019 International Conference on 3D Vision (3DV), IEEE, pp. 623–632.

Guo, Y., Bennamoun, M., Sohel, F., Lu, M. and Wan, J., 2014. 3d object recognition in cluttered scenes with local surface features: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(11), pp. 2270–2287.

Guo, Y., Sohel, F., Bennamoun, M., Lu, M. and Wan, J., 2013. Rotational projection statistics for 3d local surface description and object recognition. International journal of computer vision 105(1), pp. 63–86.

Hänsch, R., Weber, T. and Hellwich, O., 2014. Comparison of 3d interest point detectors and descriptors for point cloud fusion. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2(3), pp. 57.

Jiang, M., Wu, Y., Zhao, T., Zhao, Z. and Lu, C., 2018. Pointsift: A sift-like network module for 3d point cloud semantic segmentation. arXiv preprint arXiv:1807.00652.

Kehl, W., Milletari, F., Tombari, F., Ilic, S. and Navab, N., 2016. Deep learning of local rgb-d patches for 3d object detection and 6d pose estimation. In: European conference on computer vision, Springer, pp. 205–220.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X. and Chen, B., 2018. Pointcnn: Convolution on x-transformed points. In: Advances in Neural Information Processing Systems, pp. 820–830.

Magnusson, M., Lilienthal, A. and Duckett, T., 2007. Scan registration for autonomous mining vehicles using 3d-ndt. Journal of Field Robotics 24(10), pp. 803–827.

Pomerleau, F., Colas, F., Siegwart, R. et al., 2015. A review of point cloud registration algorithms for mobile robotics. Foundations and Trends® in Robotics 4(1), pp. 1–104.

Qi, C. R., Su, H., Mo, K. and Guibas, L. J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660.

Qi, C. R., Yi, L., Su, H. and Guibas, L. J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in neural information processing systems, pp. 5099–5108.

Rusu, R. B. and Cousins, S., 2011a. 3d is here: Point cloud library (pcl). In: 2011 IEEE international conference on robotics and automation, IEEE, pp. 1–4.

Rusu, R. B. and Cousins, S., 2011b. 3D is here: Point Cloud Library (PCL). In: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China.

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

Rusu, R. B., Blodow, N., Marton, Z. C. and Beetz, M., 2008. Aligning point cloud views using persistent feature histograms. In: 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, pp. 3384–3391.

Sipiran, I. and Bustos, B., 2011. Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. The Visual Computer 27(11), pp. 963.

Tombari, F., Salti, S. and Di Stefano, L., 2010. Unique shape context for 3d data description. In: Proceedings of the ACM workshop on 3D object retrieval, pp. 57–62.

Wang, Y., Chen, Q., Zhu, Q., Liu, L., Li, C. and Zheng, D., 2019. A survey of mobile laser scanning applications and key techniques over urban areas. Remote Sensing 11(13), pp. 1540.

Yang, J., Li, H. and Jia, Y., 2013. Go-icp: Solving 3d registration efficiently and globally optimally. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1457–1464.

Yang, J., Li, H., Campbell, D. and Jia, Y., 2015. Go-icp: A globally optimal solution to 3d icp point-set registration. IEEE transactions on pattern analysis and machine intelligence 38(11), pp. 2241–2254.

Yew, Z. J. and Lee, G. H., 2018. 3dfeat-net: Weakly supervised local 3d features for point cloud registration. In: European Conference on Computer Vision, Springer, pp. 630–646.

Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J. and 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, Z., Wen, C., Chen, Y., Li, W., You, C., Wang, C. and Li, J., 2019. Indoor scene registration based on siamese network and pointnet. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences pp. 307–312.

Zhong, Y., 2009. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, IEEE, pp. 689–696.

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

ACKNOWLEDGEMENTS

This work was supported by the National Key RD Program of China (2016YFB0502102), National Key RD Program of China (2016YFB0502104), the China Scholarship Council (CSC) scholarship, and an International Exchange Program for Graduate Students, Tongji University.