Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

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

Pairwise registration of TLS point clouds by deep multi-scale local features☆

Wei Li[a,b], Cheng Wang[a,b,*], Chenglu Wen[a,b], Zheng Zhang[a], Congren Lin[a], Jonathan Li[a,b,c]

[a] Fujian Key Laboratory of Sensing and Computing for Smart City, School of Informatics, Xiamen University, Xiamen, China
[b] Digital Fujian Institute of Urban Traffic Big Data Research, Xiamen University, Xiamen, China
[c] Department of Geography and Environmental Management, University of Waterloo, Waterloo, Canada

ARTICLE INFO

Article history:
Received 24 September 2018
Revised 29 October 2019
Accepted 18 December 2019
Available online xxx

Communicated by Dr. Zhu Jianke

2018 MSC:
00-01
99-00

Keywords:
MSSNet
Point cloud registration
Terrestrial laser scanning (TLS)
Data augmentation
Geometric constraints

ABSTRACT

Because of the mechanism of TLS system, noise, outliers, various occlusions, varying cloud densities, etc. inevitably exist in the collection of TLS point clouds. To achieve automatic TLS point cloud registration, many methods, based on the hand-crafted features of keypoints, have been proposed. Despite significant progress, the current methods still face great challenges in accomplishing TLS point cloud registration. In this paper, we propose a multi-scale neural network to learn local shape descriptors for establishing correspondences between pairwise TLS point clouds. To train our model, data augmentation, developed on pairwise semi-synthetic 3D local patches, is to extend our network to be robust to rotation transformation. Then, based on varying local neighborhoods, multi-scale subnetworks are constructed and fused to learn robust local features. Experimental results demonstrate that our proposed method successfully registers two TLS point clouds and outperforms state-of-the-art methods. Besides, our learned descriptors are invariant to translation and tolerant to changes in rotation.

1. Introduction

With the rapid development of laser scanning technology, 3D point clouds acquired from LiDAR, Kinect, Range Cameras, etc., are becoming more and more prevalent. As a technology for collecting point clouds, a Terrestrial Laser Scanning (TLS) system [1] provides an array of capabilities in terms of instrument range, scan speed, field of view, size and portability. At present, TLS is widely applied to many practical applications such as preserving cultural heritage, engineering surveying, and manufacturing. TLS point clouds registration is the foundation for 3D reconstruction, object retrieval, object pose estimation, and camera localization.

Classic methods, such as the Iteratively Closest Point (ICP) algorithm [2] and its variants [3–6], which rely on the initial position, cannot satisfy the registration of real-world 3D point clouds. Experimental results show the effectiveness of the 4-Points Congruent Sets-based (4PCS) methods [7,8]. However, as a result of operating at a point level, these methods are sub-optimal in the direction of slippage. Many hand-crafted 3D local descriptors, such as those by Rusu et al. [9], Zhang et al. [10], and Zou et al. [11], have been proposed for point cloud registration. These methods perform well in the 3D models having manifold completed surfaces, but are insufficiently robust to real-world 3D point clouds. Although carefully following the paradigm shift to deep neural networks [12] and [13], recent trends try to naively extend the networks from a 2D to a 3D domain. Due to sparse rendering data, lots of spatial details are lost which leads to sub-optimal. Therefore, TLS point cloud registration still presents a challenge because of low scene overlap severe occlusion and self-occlusion, and without prior positional information. The main difficulties are summarized as follows:

Data size: TLS point clouds can be acquired at a speed of 300,000 pts/s so that large-scale point clouds can be collected easily. Thus, many typical methods are invalid on such large amounts of data.

Noise and outliers: Noise is presented as a form of randomly fluctuating data, outliers are considered as those points far from the surface. Both noise and outliers are common and unavoidable.

Various occlusion: Because of different scanning views, data is often missing and incomplete when the objects are the same. Thus, the descriptions from corresponding keypoints are often inconsistent.

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author at: Fujian Key Laboratory of Sensing and Computing for Smart City, School of Informatics, Xiamen University, Xiamen, China.
E-mail address: cwang@xmu.edu.cn (C. Wang).

https://doi.org/10.1016/j.neucom.2019.12.074
0925-2312/

Varying densities: The scanning mechanism of the TLS system presents that, the closer the target is to the TLS system, the denser the acquired points are, which makes large variations in the density of TLS point clouds. The varying densities lead to failure of existing registration methods.

PointNet [14], pioneering efforts that directly processes 3D point sets, computes two types of features in a point cloud including a single global feature vector for the entire point cloud and a set of local features vectors for each point. However, this reliance on only either fully local or fully global feature vectors makes it difficult to estimate features of keypoints that depend on local neighborhood information. As an improved method of PointNet, Qi et al. [15] proposed PointNet++ that is able to learn local features with increasing contextual scales. Inspired by the method of PointNet++ [15], we design a novel multi-scale network to learn the description of local neighborhood. Specially, by varying neighbor sizes around a point, a novel multi-scale learning network is developed for robust estimation of local shape descriptors over a range of TLS point clouds. The key here is that local shape descriptors can be robustly estimated by suitably accounting for noise margin, occlusion, and varying densities.

The main contributions of this work are summarized as follows: (1) Based on varying local neighborhoods, we propose a Multi-Scale Siamese Network (MSSNet), which directly consumes unorder point clouds, to learn local shape descriptors. (2) To build a training set, we develop a novel data-augmented method by randomly removing points, adding noise, and rotating for 3D local patches. (3) Geometric constraints of matching are used to extract more correct corresponding pairs.

The remainder of this paper is as follows. Section 2 provides a brief review on the representative works of registration methods including ICP-like, RANSAC-like, and features of keypoints-based. Section 3 introduces the pipeline of our registration method, including learning local features by MSSNet and rejecting false correspondences with geometric constraints. Section 4 presents experimetal results and analyses. Section 5 contains some conclusions.

2. Related work

Using transforming parameters with six degrees of freedom, 3D point cloud registration is usually considered as rigid registration. Many related methods have been proposed for the related applications. In this Section, we briefly review the representative related work in rigid registration of point cloud.

ICP [2], one of most popular methods, alternates between estimating the point correspondence and the transformation matrix. Besides, many variations of this method [16–19] based on different applications have been proposed. However, such ICP-like methods first rely on the assumption that all points have pairwise counterparts between two sets. Furthermore, the methods are very sensitive to a given initialization. Yang et al. [20] proposed a globally optimal solution to ICP in 3D Euclidean registration, which combines ICP with a Branch-and-Bound (BnB) scheme. Similarly, Campbell and Petersson [21] replaced the objective function with a convolution of Gaussian Mixture Models (GMMs). However, these models cannot guarantee the global optimality of the solutions. To describe point cloud and surface normal densities, Straub et al. [22] proposed a method of Bayesian non-parametrics, which includes BnB optimization to estimate relative transformation. Because BnB is exponentially complex, these methods are computationally expensive, and are sensitive to missing data. Fan et al. [23] proposed a Convex Hull indexed Gaussian Mixture Model (CH-GMM) by incorporating proximity, area conservation and projection consistency. The CH-GMM algorithm works well on small-scale 3D point set, but it still faces challenges on large-scale TLS point clouds.

Following the method of RANSAC, Aiger et al. [7] proposed a randomized alignment approach and the idea of planar congruent sets to compute optimal global rigid transformation. However, their method has a complexity of $O(n*2+k)$, where n denotes the size of the point clouds, and k is the set of candidate congruent 4-points (called 4-PCS). Thus, 4-PCS has great limitations when sampling numbers are large. Mellado et al. [8] proposed Super-4PCS algorithm, which has a complexity of $(n+k_1+k_2)$, where k_1 is the number of pairs in source point cloud at a given distance, r and k_2 is the number of congruent sets. Theiler et al. [24] used DoG and Harris detectors to extract keypoints and then adapted the 4PCS algorithm for registration. However, due to their point-level operation, these RANSAC-like methods are easy to sub-optimal when computing their transformation relations. The problem of varying density in the TLS point clouds makes the performance of the 4PCS-based method even worse.

Many 3D local descriptors, based on hand-crafted methods, have been proposed. Johnson and Hebert [25] constructed a local shape descriptor, whose points are oriented with associated directions, and used it to match surfaces. Tombar et al. [26] pointed out the key issues of uniqueness and repeatability of the local reference frame, and encompassed a new unique, repeatable local reference frame as well as a new 3D descriptor. To optimize initial alignment, Rusu et al. [27] proposed a Sample-Consensus-based method that combines the local features called Fast Point Feature Histograms (FPFH), and Signatures of Histograms. However, their method still cannot guarantee global convergence. Yang and Zang [28] extracted crest lines as matching primitives and then proposed a deformation energy model to find correspondences. In their experiments, they obtained accuracy registration results. To reduce the required processing dramatically, Kechagias-Stamatis and Aouf [29] propose a 3D covariance descriptor, which overrides the necessity of a local reference frame or axis (LRF/A). Yang et al. [30] proposed a registration method based on semantic feature points extracted from large-scale urban scenes. Zai et al. [31] proposed an adaptive covariance (ACOV) descriptor that, in some TLS point clouds, is invariant to rigid transformation and robust to noise and varying resolutions. However, the robust ACOV descriptors relies on RGB color and intensity of the points. Yang et al. [32] proposed a local reference frame (LRF) together with a triple orthogonal local depth images (TOLDI) representation. Navarrete et al. [33] proposed a 3D compression and decompression method using GMMs for registration of 3D point clouds. Yang et al. [34] revealed that different spatial information encoding approaches would bring significant effect on a local shape descriptor. However, All of the manually designed descriptors are usually based on the property of the data itself and work well on small-scale models. Thus, it is difficult to adapt to large-scale TLS point clouds.

Recently, with the advent of deep learning, multi-scale information has been exploited both in 2D and 3D keypoint descriptions. To optimally combine feature maps from different scales for visual correspondence, Wang et al. [35] proposed a scale-attention network. However, this method is limited to 2D keypoint description. Fathy et al. [36] proposed a CNN-based hierarchical matching framework for 2D and 3D matching. However, to compare extracted feature vectors and establish correspondences, the training process of this method requires strong supervision in the form of per-pixel ground-truth labels. Besides, the tasks of 2D and 3D interest point matching and refinement are very different from our 3D point cloud registration. Learning local geometric representation on 3D point clouds is becoming more and more popular. Some works tried to naively extend the networks, usually in form of a binary-occupancy grid, from 2D to 3D domains. This idea was

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074

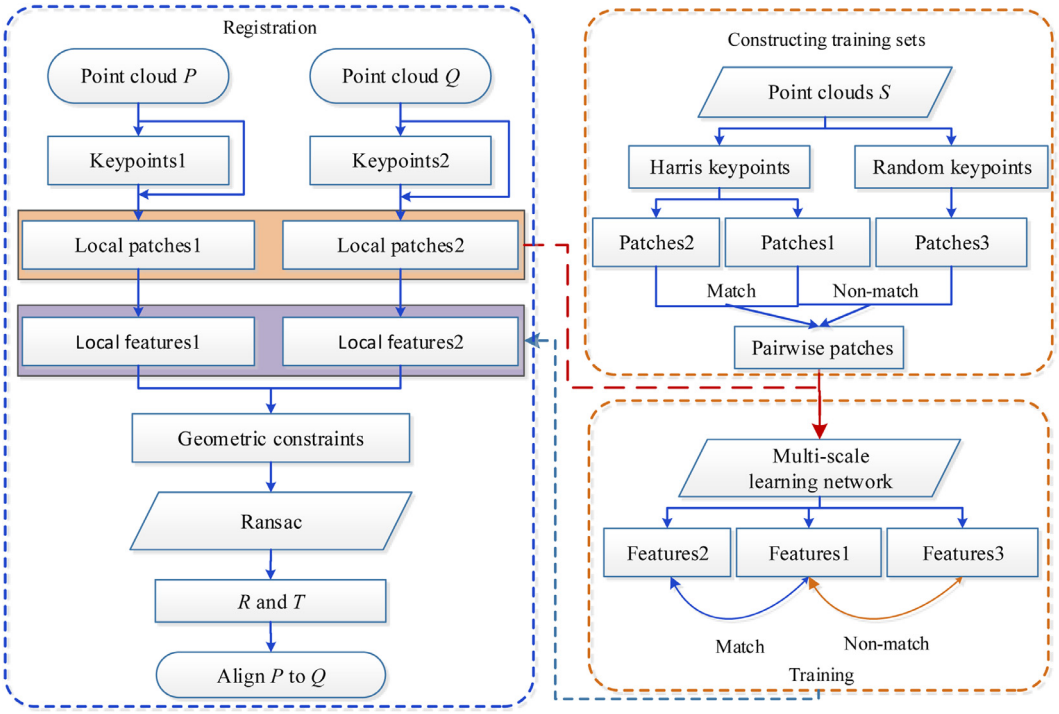

Fig. 1. Algorithm framework, consisting of testing (registration) and training. In the registration section, pairwise point clouds P and Q are used as the registration input. Then, the Harris algorithm is used to detect keypoints, and the trained MSSNet is used to describe the corresponding local patches. Finally, matches are extracted based on the similarity measure of local features and geometric constraints, and rigid transformation parameters are estimated. In the training section, first, a training dataset, including matching and non-matching patches, is constructed. Then, MSSNet is used to train the dataset. Finally, for TLS point cloud registration, the trained MSSNet, is used to transform the 3D local patches into robust local features.

quickly extended to more informative encoding such as TDF [37], TSDF [38] (Truncated Signed Distance Field). Because mainly used in the context of 3D retrieval, entire 3D objects were represented with small voxel grids 30^3 limited by the maximum size of the 3D convolution kernels. Zeng et al. [38] presented 3DMatch, a data-driven model that learns a local volumetric patch descriptor for establishing correspondences between partial 3D RGB-D data. Huang et al. [13] proposed a learning method of local shape descriptors using multi-view convolutional networks. Both of these methods ignore the raw nature of the input: sparsity and unstructured-ness. The methods use dense local grids and 3DCNNs to learn the descriptor. Thus, they fall short in performance of training/testing performance and recall. Elbaz et al. [39] proposed an auto-encoder deep learning network to learn local features for registration. However, such a method cannot adopt this approach for similar sized point clouds with partial overlap.

Compared with current methods of deep learning, in this paper, we propose a novel multi-scale network that directly consumes 3D raw point set, to learn robust local shape descriptors.

3. The proposed method

Inspired by the methods of PointNet [14] and Pointnet++ [15], which directly consume unordered 3D point sets, we designed a MSSNet network to learn robust local shape features. Then, combining corresponding learned local shape features, the keypoints were used to construct correspondences between two partial point clouds. Here, the pairwise matching of local descriptors is transformed into classifying as well as corresponding interest points (match) and non-corresponding interest points (no-match) by Euclidean distance. As shown in Fig. 1, the framework of our proposed algorithm can be summarized mainly as the following three sections: constructing training datasets, training local features and pairwise registering of point clouds. First, training datasets were built from a semantic3d dataset, which provides a large quantity of training datasets. Second, training datasets were constructed and fed into the MSSNet network. Third, superior local features, generated by the trained MSSNet network, were used for TLS point cloud registration.

3.1. Constructing training datasets

Because each point cloud is large-scale and contains millions of points in the semantic3d dataset,[1] it can provide a large quantity of 3D local patches to train of our MSSNet network. Constructed like 3DMatch [38], pairwise 3D local patches from many-view observations of a scene can provide more reliable training datasets. However, because there is only a small number of different observations of the same scene in a semantic3d dataset, it cannot provides enough training samples. Therefore, we developed a method of data augmentation to construct training samples. Especially, we first extracted keypoints from each point cloud by the 3D Harris method [40]. Second, one of the n_0 nearest neighbor points ($n_0 = 10$ in our experiments) of each keypoint is randomly selected as the matching keypoint. Any random point from another point cloud is selected as a non-matching keypoint. Third, based on the above keypoints (match and non-match), we randomly sampled 1024 points within a given neighborhood with a radius of 0.5 m.

To acquire robust 3D local descriptors, we constructed a semi-synthetic training dataset that includes processing as follows: (1) adding Gaussian random noise by randomly vibrating 5, 10 and 30 percent of the points; (2) randomly rotating the pairwise matching local patches (0°–360°) over the $Z-axis$; (3) simulating occlusions by removing 15, 25, 45 percent of the local patches. See Fig 2.

[1] http://www.semantic3d.net/.

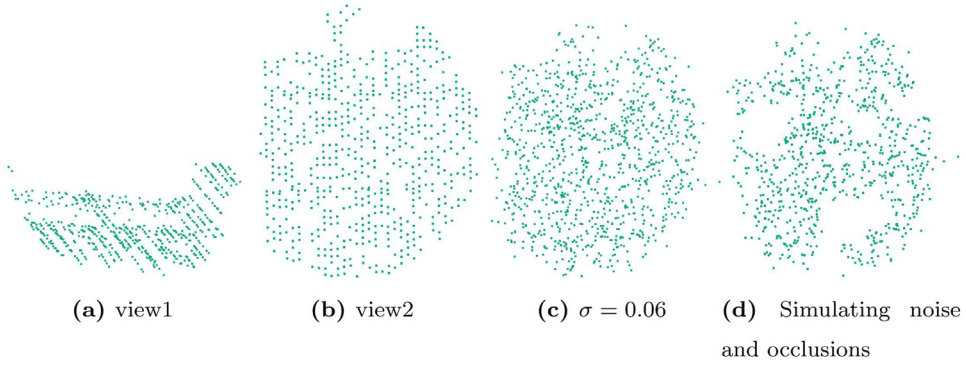

Fig. 2. simulating noise, density change, and occlusions. (a) View1 of a 3D local patch p_1. (b) view2 of the same local patch. (c) A new 3D local patch which is added Gaussian noise ($\sigma = 0.06$). (d) The 3D local patch simulating occlusions and noise is used as matching patch of p_1.

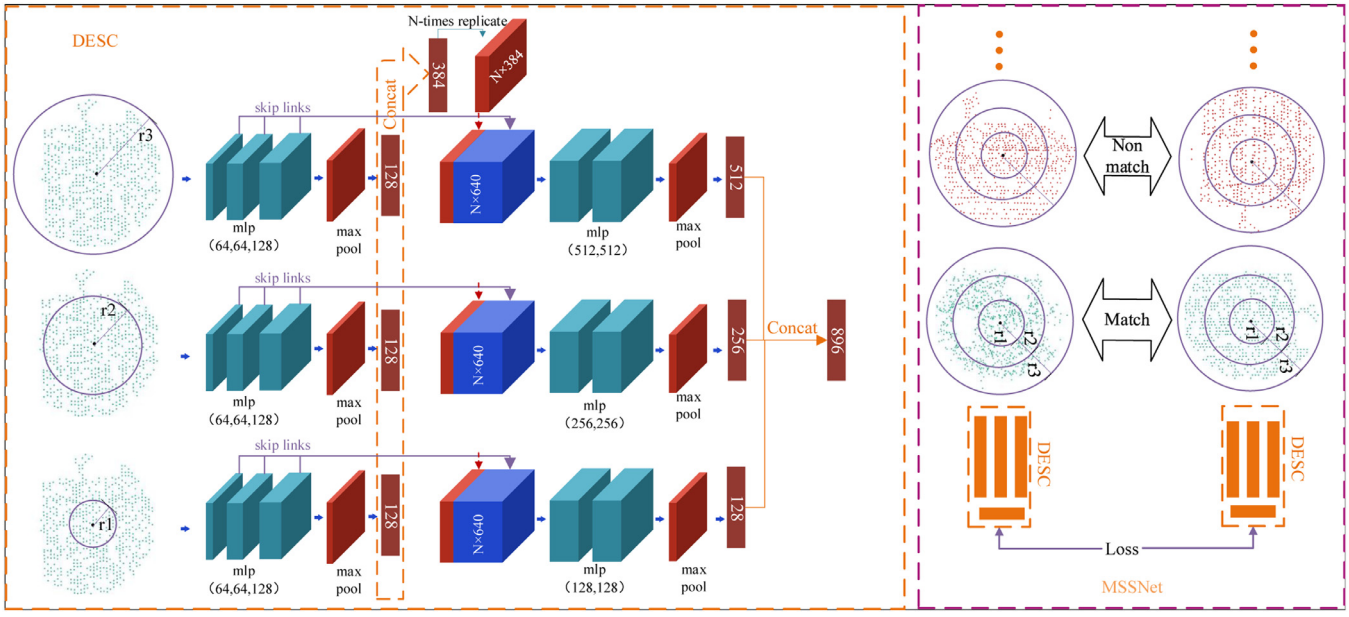

Fig. 3. Based on the Siamese architecture, we present a MSSNet, consisting of three subnetworks in each branch, to fuse the features of different scales. Specially, in the DESC section, based on varying neighborhood radii, we first construct three subnetworks to generate multi-scale features. Second, all the features were concatenated in each branch of the Siamese architecture.

These methods provide a large and diverse (match or non-match) number local patches for training.

Instead of the 3D volumetric representation of multi-view projected or voxel grid, each 3D local patch is the original representation (unordered point clouds). Given a 3D point cloud P, we extracted 3D Harris keypoints $\{c_i\}_{k=1}^{K}$ and obtained the corresponding 3D local patches, respectively. In the TLS coordinate system, each 3D local patch $p_{i=1}^{n}$, which is translated according to the corresponding keypoint, uses c_i as the origin, and is normalized by the distance of the furthest point from the origin. The translating formulate is as follows.

$$p'_i = p_i - c_k \tag{1}$$

where p'_i represents coordinates of a point in the local coordinate system.

3.2. Network architecture

Based on the PointNet-like encoder, MSSNet develops a multi-scale fusing approach to learn local features. It is designed to operate on three subnetworks with varying neighbors size, as summarized in Fig. 3.

The global features acquired by the PointNet-like approach signify that the larger the neighbor size, the more stable the global information. Given different neighbors of a keypoint, however, there exists great shape differences. Therefore, if a PointNet-like method is used, the larger neighbor size leads to considerably less local information. It is meaningful to fuse the features at different neighbor size to generate more significant and stable feature. As seen in the righthand segment of Fig. 3, local patch pairs (including match and non-match) are fed into our MSSNet at a ratio of 1:1.

To fuse the geometric characteristics of varying neighbors, as shown in the left of Fig. 3, for each branch, three local patches are sampled by varying neighborhood sizes r_1, r_2 and r_3. More specifically, a three-layer, point-wise Multi-Layer Perception (MLP) follows the input layer and subsequently a max-pooling is used to generate a local feature using skip-links. This results in a more powerful representation. To enhance the significance of local features, we concatenate the above three local features to a new global feature (384 dimension). Then, the global feature is used to concatenate the following two-layer MLP in each subnetwork, respectively. In the end, the local features are generated by Max pooling and concatenated to a codeword (896 dimension).

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074

In our experiments, a Siamese network based on PointNet (S-PointNet) was constructed from 1024 kernels with a neighborhood size (r1) of 1.0 m. A 3-PointNet-like Siamese network, which neighborhood sizes (r1, r2, and r3) of 0.3 m, 0.6 m and 1.0 m, was constructed from 128, 256, and 512 kernels, respectively. Additionally, we implement our MSSNet in tensorflow and use ADAM to train our network for 10,000 iterations using a base learning rate 10^{-3} and an initial momentum 0.9. We decay the learning rate at 10,000 steps and decaying rate is set to 0.7.

3.3. Loss function

We formulated matching issues as a binary classification including match and non-match. Our objective is to reduce the cost between two matched 3D patches, and, conversely, increase the dissimilarity between two non-matched 3D patches. To learn robust local features, we constructed a multi-scale network based on a Siamese fashion. To measure the similarity of 3D patch pairs, the L_2 norm as a metric, modeled by a contrastive loss function, was used. To avoid the phenomenon of over-fitting, we added the L_2 regularization term to the objective function. The loss function formula Eq. (2) is constructed as follows:

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)}d_i^2) + (1-y^{(i)})max(margin - d_i, 0)) + \frac{\lambda}{2n}\sum_{\omega}\omega^2 \tag{2}$$

where i represents the i-th pairwise local patches, n represents the number of training pairs of local patches; $y^{(i)}$ is a matching label that represents the matching of the pairwise points, where $y^{(i)} = 1$ if the pair-wise points match and $y^{(i)} = 0$ if the pair-wise points non-match. d_i is Euclidean distance, ω represents the learning weight parameters, θ represents the parameters when the object function is minimized.

3.4. Geometric constraints

Through the trained MSSNet, 3D local patches, with sampling based on keypoints, are transformed into robust features. To extract matches more effective, inspired by the reference [41], we introduce geometric constraints to reject mismatches. Here, we denote $\mathbf{F}(\mathbf{P}) = \{\mathbf{F}(p) : p \in \mathbf{P}\}$, where $\mathbf{F}(p)$, generated by trained MSSNet, is a local feature corresponding to the keypoint p. Similarly, $\mathbf{F}(\mathbf{Q}) = \{\mathbf{F}(q) : q \in \mathbf{Q}\}$. In general, the initial correspondence set κ is acquired by arbitrarily combining the keypoints between source point cloud $\mathbf{P}$ and target point cloud $\mathbf{Q}$. To extract good (matched) correspondences, we introduce the following geometric constraints.

First, for each keypoint $p \in \mathbf{P}$, we find the nearest neighbor of $\mathbf{F}(p)$ among $\mathbf{F}(Q)$, and at the same time, for each keypoint $q \in \mathbf{Q}$, we find the nearest neighbor of $\mathbf{F}(q)$ among $\mathbf{F}(P)$. Thus, the generated candidate correspondences, denoted by κ_I, have a very high ratio of outliers.

Second, for the pairwise test, a correspondence pair $(\mathbf{p}, \mathbf{q})$ is selected as match pair from κ_I if and only if $\mathbf{F}(p)$ is the nearest neighbor of $\mathbf{F}(q)$ among $\mathbf{F}(\mathbf{P})$ and $\mathbf{F}(\mathbf{Q})$. The resulting candidate correspondence set is denoted by κ_{II}.

Third, for the tuples test, three correspondence pairs $(\mathbf{p}_1, \mathbf{q}_1)$, $(\mathbf{p}_2, \mathbf{q}_2)$, $(\mathbf{p}_3, \mathbf{q}_3)$ are selected randomly from κ_{III}, and check if the tuples $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$ and $(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$ are compatible. In detail, we collect the candidate correspondence set κ_{III} if the following condition is satisfied:

$$\forall i \neq j, \lambda < \frac{\|\mathbf{p}_i - \mathbf{p}_j\|}{\|\mathbf{q}_i - \mathbf{q}_j\|} < 1/\lambda \tag{3}$$

where $\lambda = 0.9$, $\mathbf{p}_i, \mathbf{p}_j \in \mathbf{P}$ represents the keypoints from point cloud $\mathbf{P}$ and $\mathbf{q}_i, \mathbf{q}_j \in \mathbf{Q}$ represents the keypoints from point cloud $\mathbf{Q}$. $\|*\|$ represents the L_2 norm. This test verifies if the correspondences are compatible.

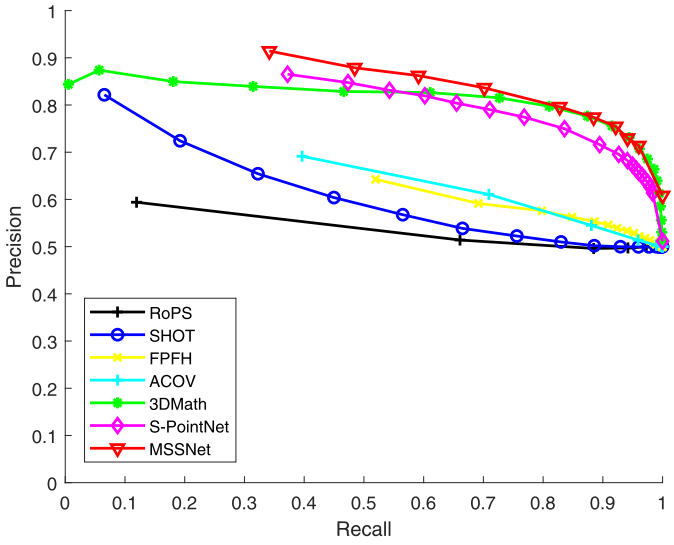

Fig. 4. Evaluation of our method compared with some representative methods on 3DMatch's benchmark, which contains 10,000 pairs of samples with match or not.

4. Experiment analysis

In this section, first, we evaluate the quality of our learned local shape features, which lies in the analysis of varying noise, varying neighborhood size, and varying rotation. Second, compared with the state-of-the-art methods, our learned 3D local features achieved superior registration results for 3D point clouds. Third, we show several pairwise registration results of point clouds. Our training methods of local patches were used with tensorflow 1.7 and cuda 9.0. Registration experiments were conducted in C++, and on a PC with ubuntu 16.04, Intel Core(TM) i5-4460 3.2 GHz CPU and 16.0 GB RAM.

4.1. Dataset

To validate our method, our evaluations for keypoints matching are against the diverse 3DMatch RGBD benchmark [38] in with 62 different real-world scenes retrieved from the pool of datasets. This collection is split into 2 subsets with 54 scenes for training and validation and 8 scenes for testing. The dataset typically includes indoor scenes like living rooms, offices, bedrooms, tabletops, and restrooms.

For TLS point cloud registration, our method was tested mainly in TLS point clouds. As a huge dataset of 3D local patches, the training dataset, labeled match and non-match from Semantic3d [42], was built to train our descriptors effectively. The dataset, containing fifteen outdoor TLS point clouds from the training set of semantic-8, and four fragments from the testing set of semantic-8, were used to construct local testing patches. Specially, we sampled keypoints from the TLS point clouds, and obtained 200,000 pairs of 3D local point clouds (i.e. local patches) at a given neighbor radius (0.5 m). For training, all the local patch pairs including match or non-match, were labeled at a 1:1 ratio for training, Similarly, 10,000 pairs of local patches were labeled for testing.

For registration test of TLS point clouds, eight fragments of TLS point clouds were selected from the Semantic3d, where two scenes including four fragments (See Fig. 10) were selected for evaluating of registration. The information of the evaluated fragments is summarized in Table 2, the average resolution $\bar{r}$ (the average of the distances of all two adjacent points in the point cloud) of each

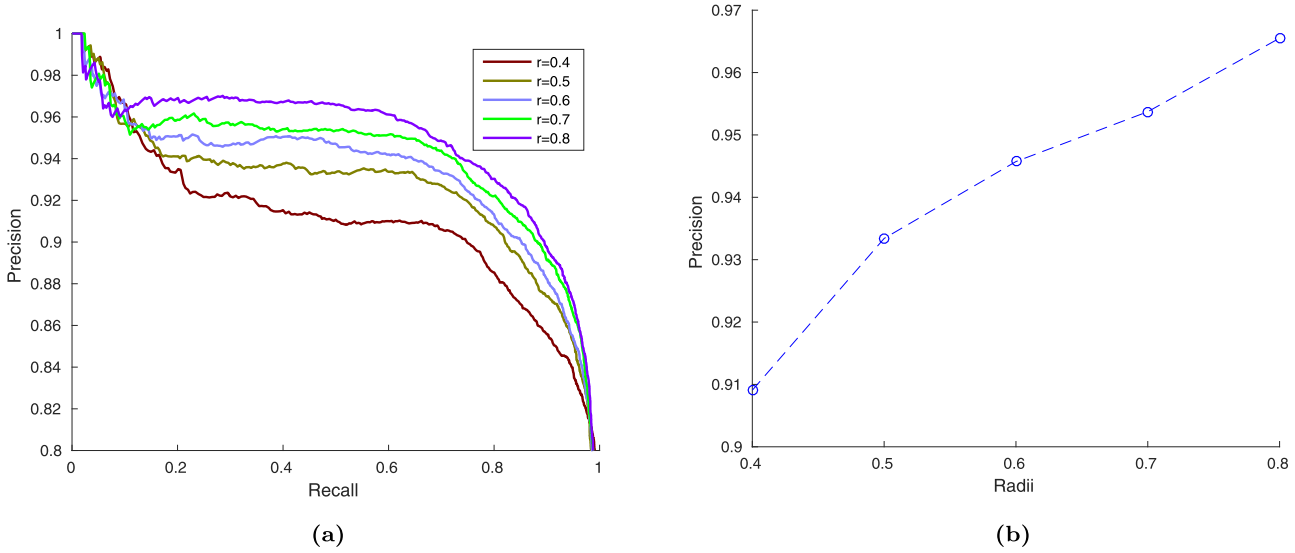

Fig. 5. (a) PR Curves achieved by varying neighbor size. (b) Precision Curve achieved by Recall ratio=0.5 of (a).

fragment. Additionally, another two fragments (See Fig. 11(g) and (h)) were selected for the registration test.

4.2. Keypoint matching

We used Precision–Recall (PR) curves to evaluate the distinctiveness of these descriptors. The Precision and Recall are computed as follows:

$$\begin{cases} Precision = TP/(TP+FP) \\ Recall = TP/(TP+FN) \end{cases} \tag{4}$$

where TP is the number of true positive matches, FP is the number of false positive matches, and FN is the number of false negative matches. The PR Curve is generated by tuning the parameter ratio.

To demonstrate the superiority of our learned descriptors, we focus on evaluating the different parameters of noise, varying neighborhood radii, and varying rotation of pairwise 3D local point clouds. Compared with several representative descriptors including RoPS [43], SHOT [26], FPFH [9], ACOV [31], and 3Dmatch [38], we tested the ability of our descriptors to distinguish between match and non-match. Implementation of ROPs was achieved based on the open-source MATLAB code.[2] SHOT and FPFH were implemented based on the open-source Point Cloud Library (PCL) [44]. The default parameters in PCL were selected in our implementation. We first tested our descriptors on the benchmark of Zeng et al. [38] that contains 10,000 pairs of local RGB-D patches and their labels of ground truth correspondence (binary "1" for match and "0" for non-match). As seen in Fig. 4, the training and testing datasets are from reference [38]. Our learned descriptors achieved the best performance.

To test the influence of neighborhood size, we evaluated the quality of the descriptors at different neighborhood sizes. First, keypoints were randomly extracted from training and testing point clouds. Then, the corresponding local patches with different neighborhood sizes were sampled by KNN. We designed five group experiments with different neighborhood sizes (0.4 m, 0.5 m, 0.6 m, 0.7 m, 0.8 m) to validate the performance of keypoint matching. As shown in Fig. 5, the larger the neighborhood size, the better the performance. However, a larger neighborhood size requires more computational time. Therefore, in the following experiments, the neighborhood radius for initial calculation is set to 0.5 m.

To test the robust performance, we designed six different levels of Gaussian noise ($\sigma = 0.00, 0.05, 0.10, 0.15, 0.20, 0.40$) for each 3D local patch. Fig. 6 shows the variation in a local point cloud with varying levels of Gaussian noise. As shown in Fig. 7 (a–f), the training and testing datasets are constructed like the methods of Fig. 6. It is observed that our learned descriptors achieved the best performance. Especially, our learned local descriptor has significantly better performance than the other descriptors.

To test the tolerance of varying rotations, we conducted twelve pairs of experiments on varying rotations. All the 3D local patches are first decentralized by the corresponding keypoints, thus, it is invariant to the shifting transformation of the 3D local patches. Specifically, we also built 10,000 pairs of 3D local patches (P_1 and P_2). P_2 was copied and rotated at every $\pi/12$ radian to build another 3D local patches ${P_2}^i$. Similarly, a series of 3D local patches $\{{P_2}^i\}_{i=1}^{12}$, were generated and tested by our learned detectors. We use precision 4 to evaluate the, performance. As Fig. 8a shows, the accuracy of our method oscillated between 0.8 and 0.9. Therefore, the experimental results indicate that data augmentation extends our network to be robust to rotation transformation. Additionally, as seen from Fig. 8b, the lowest precision is when the 3D local patches are rotated π.

4.3. Registration

To evaluate the registration performance, we first tested our proposed method on the public Geometric Registration Benchmark[3]. As seen in Table 1, our proposed method achieved the highest Precision and the second highest Recall on the public benchmark of 3DMatch. Therefore, it shows that our proposed method has a certain superiority in RGB-D data registration. For TLS point clouds, we designed several experiments involving point cloud registration. Especially, as seen in Fig. 9, two scenes, labeled "Scene1" and "Scene2", were selected for pairwise TLS point cloud registration. The "Scene1" contains two TLS point clouds denoted by P_1 and Q_1, and the "Scene2" also contains two point clouds denoted by P_2 and Q_2. The information pertaining to the tested point clouds is given in Table 2. Especially, the point clouds P_1 and Q_1 have 36.3% overlap, the point clouds P_2 and Q_2 have 35.3% overlap. Besides, to make the algorithm more efficient, we detected keypoints by 3D Harris detectors [40].

[2] http://yulanguo.me/img/RoPS.rar.

[3] http://3dmatch.cs.princeton.edu/geometric-registration-benchmark.

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074

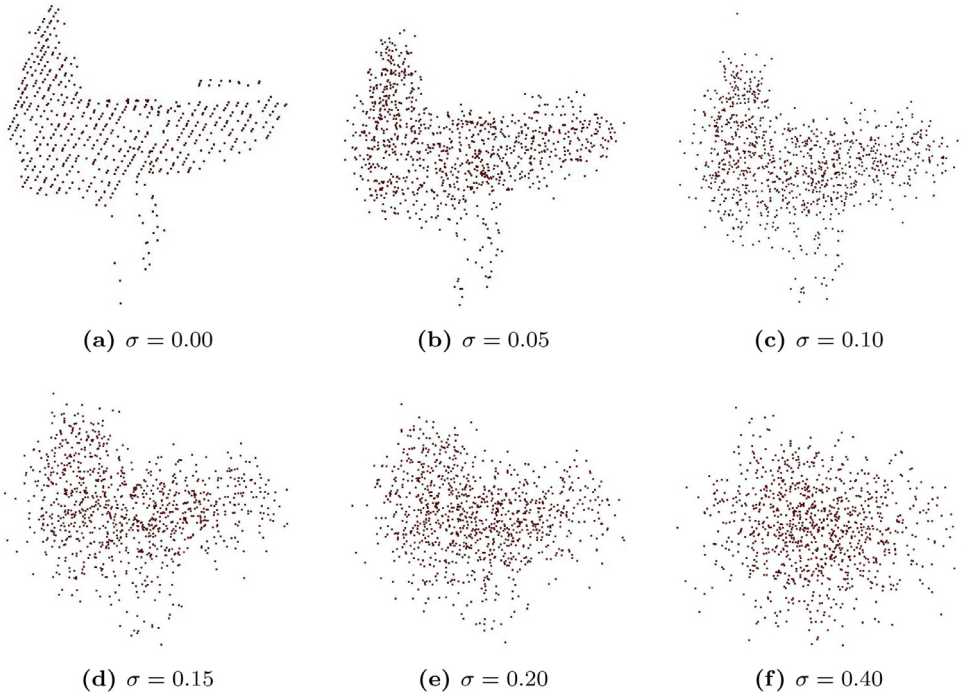

Fig. 6. The variations of a 3D raw local patch. (a) A 3D raw local patch. (b–f) Different levels of Gaussian noise ($\sigma = 0.05, 0.10, 0.15, 0.20, 0.40$).

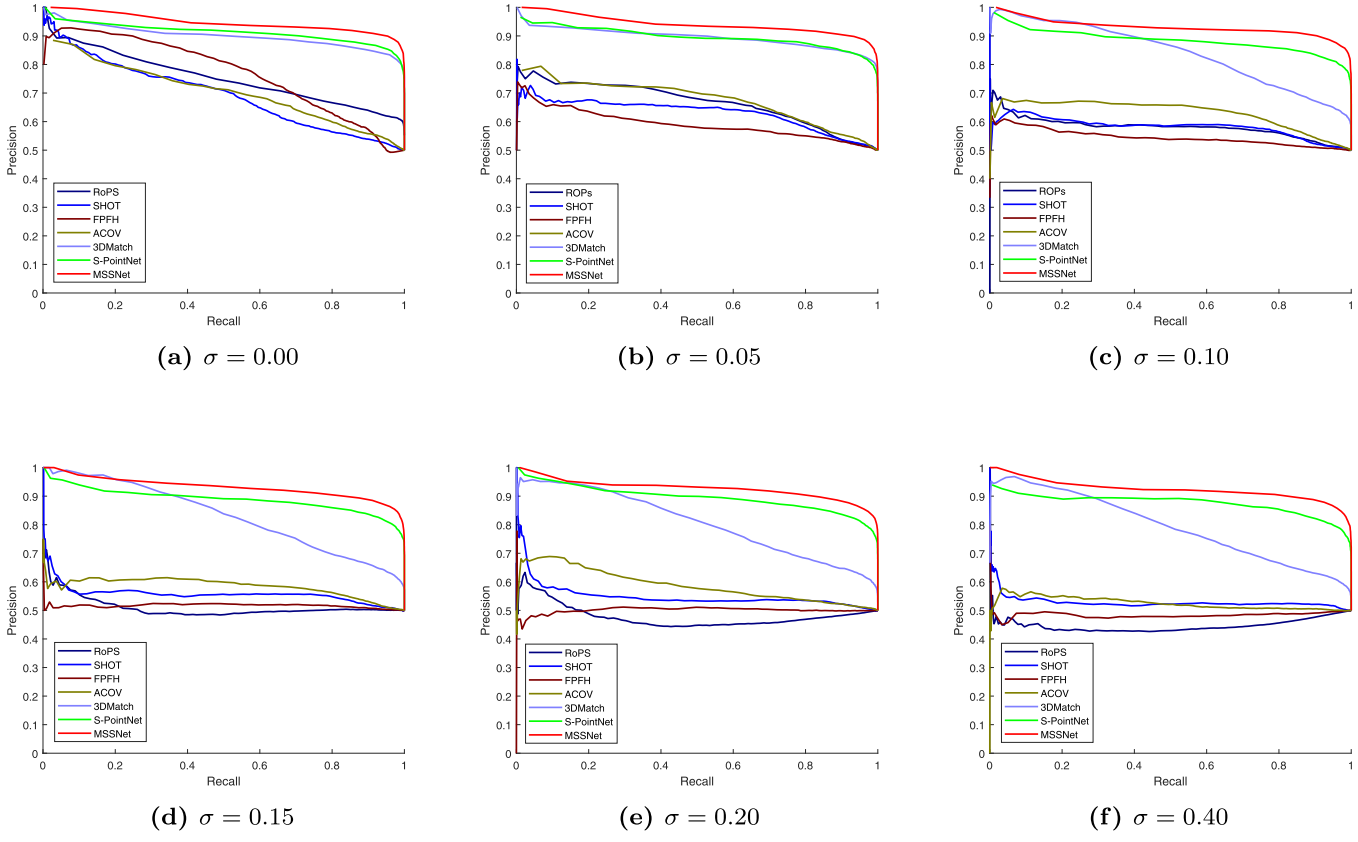

Fig. 7. PR Curves achieved by seven sets of comparative tests on our benchmark, which contains 10,000 pairs of samples with match or not. (b–f) show the corresponding PR Curves of different levels of Gaussian noise.

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074

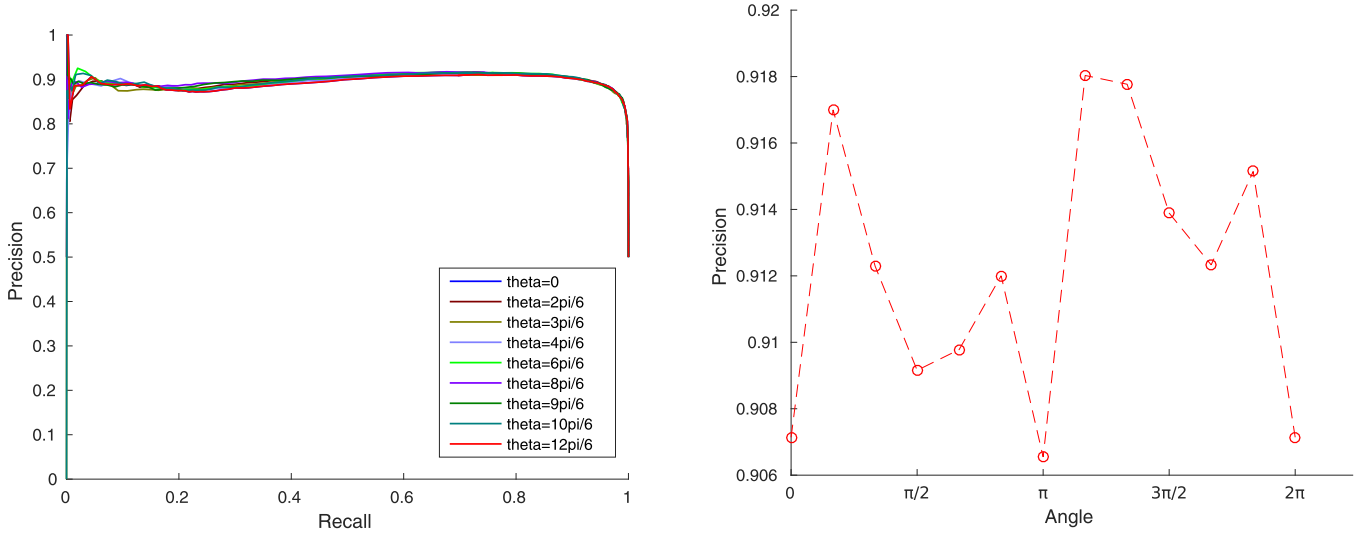

Fig. 8. (a) Rotation tolerance test. PR Curves achieved by varying rotation. (b) Evaluation of rotating variation achieved by Recall ratio=0.5 of Fig. (a).

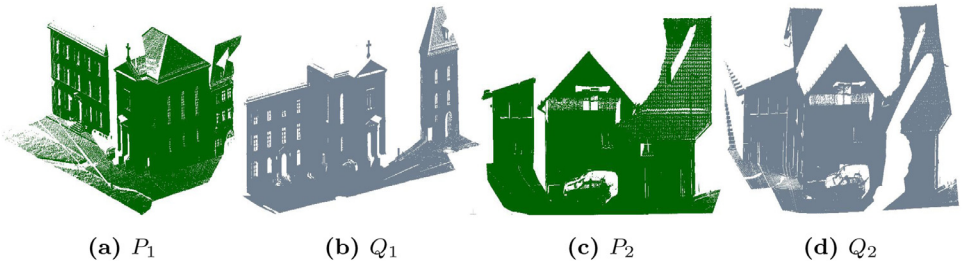

Fig. 9. Two scenes including two pairwise point clouds were selected for registration evaluation. (a,b) The first pairwise point clouds denoted by P_1 and Q_1 are acquired from "Scene1". (c,d) The second pairwise point clouds by P_2 and Q_2 are acquired from "Scene2".

Table 1
The performance of our proposed method is listed on the Benchmark Leaderboard.

Method	Recall (%)	Precision (%)
MSSNet	52.9%	**58.9%**
3DMatch	**66.8%**	40.1%
Spin-Images	51.8%	45.8%
FPFH	44.2%	30.7%

Table 2
Information of the tested point clouds.

Model	Number of points	Resolution $r(m)$	Overlap(%)
P_1	824,509	0.0411	
Q_1	13,986,927	0.0351	36.3
P_2	2,018,953	0.0201	
Q_2	2,271,520	0.0211	35.1

Table 3 shows the time cost of each step in our proposed registration framework. The keypoints extraction step requires the majority of time. As seen in Table 4, the comparative performance of some representative methods is listed in terms of average RMS distance and computational time. The proposed MSSNet was designed with a 3-scale subnetwork. As seen in Fig. 10, compared with features-based methods including FPFH, ACOV, 3DMatch and S-PointNet, (a-f) present the registration results of "Scene1", (g–l) present the registration results of "Scene2". RANSAC based on geometric registration was used to align the TLS point clouds. In addition, our method, combined with geometric constraints, is used to reject false correspondences. in terms of both registration error and computational time, our method outperforms the other methods. However, these methods, proposed by FPFH [9] and ACOV [31] (no RGB color and intensity), fail to register pairwise of point clouds. Especially, Fig. 10 (f) and (l) present the registration results of our method combined with ICP algorithm, which achieved the fine registration. Therefore, the experimental results demonstrate that our learned local descriptors are more robust than other methods.

To further demonstrate the feasibility and effectiveness, two pairs of TLS point clouds from Semantic3D and one pair of TLS point clouds acquired by RIEGL VZ-1000 system, were selected as testing data. The information of point clouds for the further test is

Table 3
Time cost of each step in the tested point clouds.

Step	Keypoints extraction	3D local patches generation	Descriptors calculation	Correspondence generation	Total time (s)
Scene1	154.0	12.0	26.5	5.2	**197.7**
Scene2	33.0	6.0	10.3	5.0	**54.3**

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074

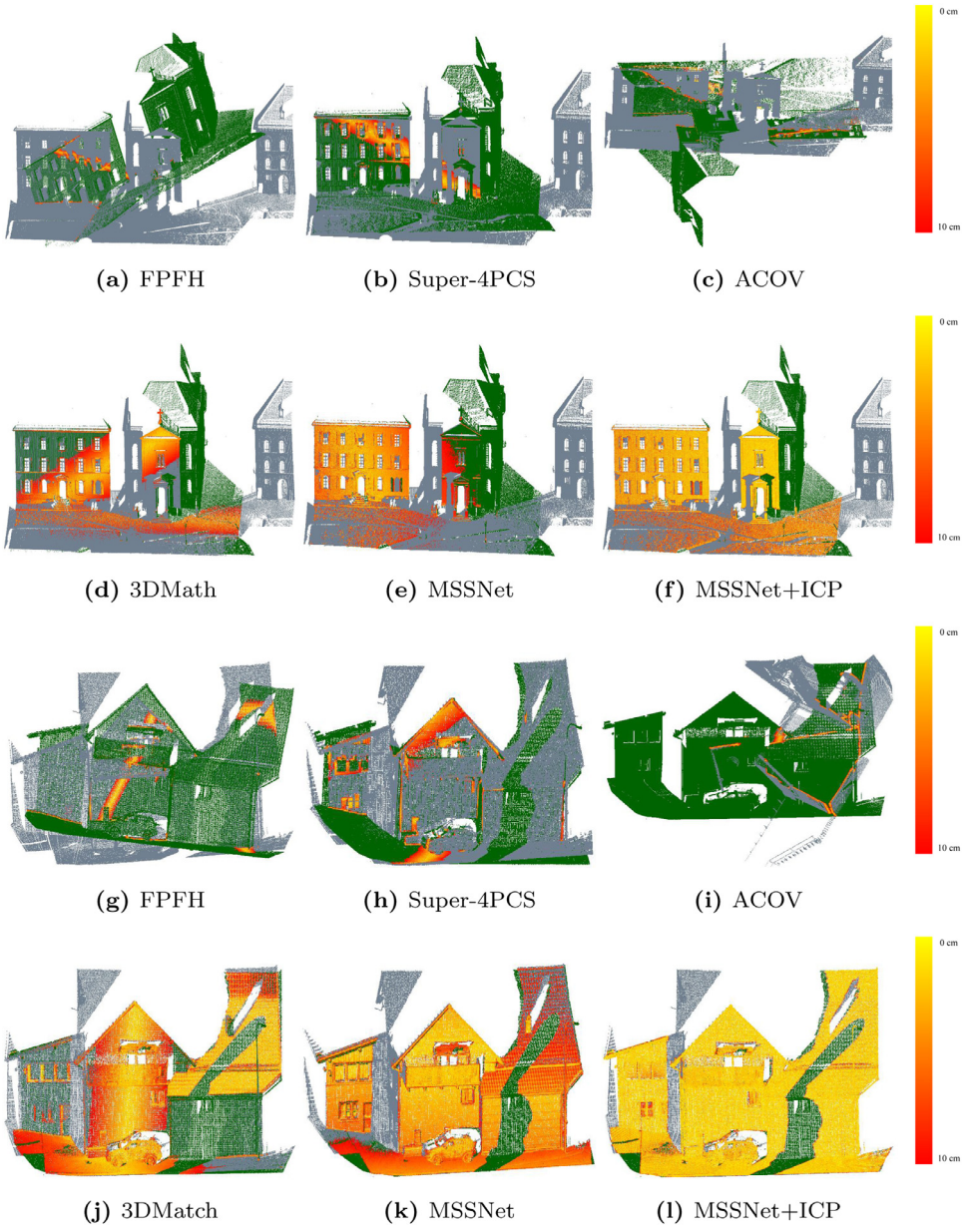

Fig. 10. Registration results on the point clouds of modelS Scene1 and Scene2. (a),(g) Achieved by FPFH. (b),(h) Achieved by Super4PCS. (c),(i) Achieved by ACOV. (d),(j) Achieved by 3DMatch. (e),(k) Achieved by MSSNet. (l) Achieved by MSSNet+ ICP.

Table 4
Compared with some typical methods.

	FPFH	Super-4PCS	ACOV	3Dmatch	Ours	Ours+ICP
Average RMS1 (cm)	–	7.81	–	5.88	**5.18**	2.90
Overlap1 (%)	–	28.3	–	32.5	**35.0**	36.0
Total time (s)	–	1500.3	–	327.8	**197.7**	218.8
Average RMS2(cm)	–	6.33	–	5.39	**4.51**	2.52
Overlap2(%)	–	33.4	–	31.3	**33.9**	35.0
Total time(s)	–	1008.1	–	356.8	**54.3**	77.8

shown in Table 5. The first pair of point clouds,P_3 and Q_3, covering a range of approximately 30 m by 25 m, are shown in Fig. 11(a) and (b). The true overlap rate, computed by means of manual registration and the ICP algorithm, is 40.5%. Fig. 11(c) shows the successful registration result with an overlap rate of 40.6%. The second pair of point clouds, P_4 and Q_4, covering a range of approximately 40 m by 15 m, are shown in Fig. 11(d) and (e). The true overlap rate is 20.9%. Fig. 11(f) shows the successful registration result with an overlap rate of 20.6%. The third pair of point clouds, P_5 and Q_5, covering a range of approximately 80 m by 20 m, are shown in Fig. 11(g) and (h). The true overlap rate is 44.5%. Fig. 11(i) shows the successful registration result with the overlap rate of 43.7%.

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074

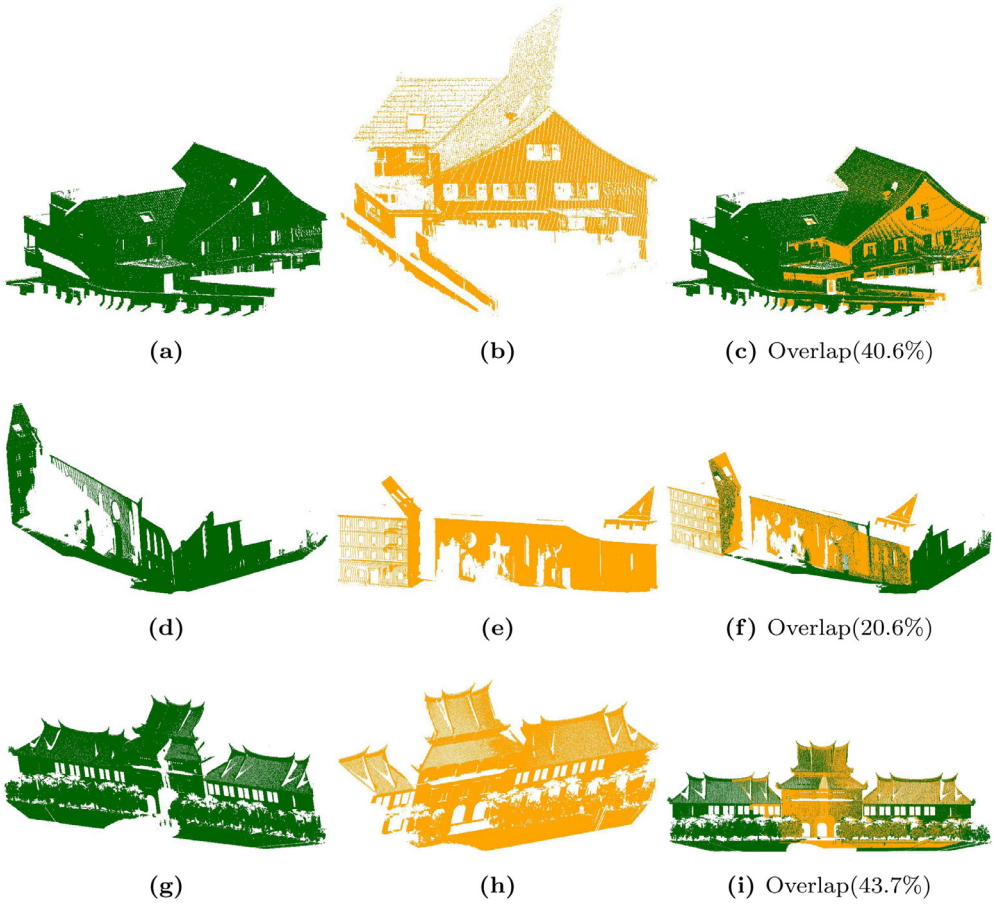

Fig. 11. Registration results on the point clouds of models. (a) and (b) are the first pairwise point clouds, (c) registration result of the first pairwise point clouds. (d) and (e) are the second pairwise point clouds, (f) registration result of the second pairwise point clouds. (g) and (h) are the third pairwise point clouds, (i) registration result of the third pairwise point clouds.

Table 5
Information of point clouds for further registration testing.

Model	Number of points	Resolution $r(m)$	Overlap(%)
P_3	618,521	0.0335	
Q_3	83,925	0.0551	40.7
P_4	13,789,069	0.0212	
Q_4	7,514,469	0.0121	21.0
P_5	8,787,836	0.0143	
Q_5	8,318,075	0.0166	45.1

Therefore, based on the local descriptors learned by MSSNet, we obtain successful registration results.

5. Conclusion

We proposed a novel method to register TLS point clouds. Based on inspired PointNet architecture, our method directly consume point clouds. Different from existing methods, 3D local shape descriptors, which learn by fusing multi-scale subnetworks, are more robust and reliable. Especially, compared with the state-of-the-art methods, first, the experiments involving public data sets show that the learned local shape descriptors are more accurate and robust to large noise; Second, the learned local shape descriptors are invariant to translation, and are tolerant to changes in rotation; Third, our method performs accurate and efficient registration even on very challenging scenes without any estimation of initial position. Overall, our proposed method outperforms existing local shape descriptors and registration methods by a significant margin.

In this paper, to extend the size of training data, data augmentation such as data rotation is used. The aim of rotation is to increase rotation invariance in the feature extraction. However, data augmentation, especially the rotation, decreases the capability to describe features. Therefore, our method would be improved by a future study of how to perform data augmentation without decreasing the capability to describe features.

Declaration of Competing Interest

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

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants U1605254, 41471379 and the Fujian Provincial Department of Education Science and Technology project JA14292. The authors would like to acknowledge the anonymous reviewers for their valuable comments.

References

[1] M. Lemmens, Terrestrial Laser Scanning, Springer Netherlands, Dordrecht, pp. 101–121.
[2] P.J. Besl, A method for registration 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 193–200.
[3] Y. Liu, Improving ICP with easy implementation for free-form surface matching, Pattern Recognit. 37 (2) (2004) 211–226.

[4] K.-H. Bae, D.D. Lichti, A method for automated registration of unorganised point clouds, ISPRS J. Photogramm. Remote Sens. 63 (1) (2008) 36–54.
[5] J. Yang, H. Li, Y. Jia, Go-ICP: solving 3d registration efficiently and globally optimally, in: Proceedings of IEEE International Conference on Computer Vision, 2013, pp. 1457–1464.
[6] Y. Liu, H. Liu, R.R. Martin, L.D. Dominicis, R. Song, Y. Zhao, Accurately estimating rigid transformations in registration using a boosting-inspired mechanism, Pattern Recognit. 60 (2016) 849–862.
[7] D. Aiger, N.J. Mitra, D. Cohenor, 4-Points congruent sets for robust pairwise surface registration, ACM Trans. Graph. 27 (3) (2008) 1–10.
[8] N. Mellado, D. Aiger, N.J. Mitra, Super 4pcs fast global pointcloud registration via smart indexing, in: Proceedings of Computer Graphics Forum, 33, Wiley Online Library, 2014, pp. 205–215.
[9] R.B. Rusu, N. Blodow, M. Beetz, Fast point feature histograms (FPFH) for 3D registration, in: Proceedings of IEEE International Conference on Robotics and Automation, 2009, pp. 3212–3217.
[10] Z. Zhang, S.H. Ong, X. Zhong, K.W.C. Foong, Efficient 3D dental identification via signed feature histogram and learning keypoint detection, Pattern Recognit. 60 (C) (2016) 189–204.
[11] Y. Zou, T. Zhang, X. Wang, Y. He, J. Song, BRoPH: A compact and efficient binary 3D feature descriptor, in: Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO), 2016, pp. 1093–1098.
[12] G. Elbaz, T. Avraham, A. Fischer, 3D point cloud registration for localization using a deep neural network auto-encoder, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2472–2481.
[13] H. Huang, E. Kalogerakis, S. Chaudhuri, D. Ceylan, V.G. Kim, E. Yumer, Learning local shape descriptors from part correspondences with multi-view convolutional networks, ACM Trans. Graph. 37 (1) (2017) 1–14.
[14] R.Q. Charles, H. Su, K. Mo, L.J. Guibas, PointNet: deep learning on point sets for 3D classification and segmentation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 77–85.
[15] C.R. Qi, L. Yi, H. Su, L.J. Guibas, PointNet++: deep hierarchical feature learning on point sets in a metric space, in: Proceedings of Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 5099–5108.
[16] P.J. Besl, N.D. Mckay, A method for registration of 3-d shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (2002) 239–256.
[17] K.H. Bae, D.D. Lichti, A method for automated registration of unorganised point clouds, ISPRS J. Photogramm. Remote Sens. 63 (1) (2008) 36–54.
[18] A. Gressin, C. Mallet, J. Demantk, N. David, Towards 3D lidar point cloud registration improvement using optimal neighborhood knowledge, ISPRS J. Photogramm. Remote Sens. 79 (I-3) (2013) 240–251.
[19] J. Stechschulte, C. Heckman, Hidden Markov Random Field Iterative Closest Point, CoRR abs/1711.05864 (2017).
[20] J. Yang, H. Li, D. Campbell, Y. Jia, Go-ICP: a globally optimal solution to 3d icp point-set registration, IEEE Trans. Pattern Anal. Mach. Intell. 38 (11) (2016) 2241–2254.
[21] D. Campbell, L. Petersson, GOGMA: globally-optimal gaussian mixture alignment, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5685–5694.
[22] J. Straub, T. Campbell, J.P. How, J.W. Fisher, Efficient global point cloud alignment using Bayesian nonparametric mixtures, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2403–2412.
[23] J. Fan, J. Yang, D. Ai, L. Xia, Y. Zhao, X. Gao, Y. Wang, Convex hull indexed Gaussian mixture model (CH-GMM) for 3D point set registration, Pattern Recognit. 59 (C) (2016) 126–141.
[24] P.W. Theiler, J.D. Wegner, K. Schindler, Keypoint-based 4-points congruent sets automated marker-less registration of laser scans, ISPRS J. Photogramm. Remote Sens. 96 (11) (2014) 149–163.
[25] A.E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3d scenes, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 433–449.
[26] F. Tombari, S. Salti, L. Di Stefano, Unique Signatures of Histograms for Local Surface Description, Springer, Berlin, Heidelberg, pp. 356–369.
[27] R.B. Rusu, N. Blodow, Z.C. Marton, M. Beetz, Aligning point cloud views using persistent feature histograms, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008, pp. 3384–3391.
[28] B. Yang, Y. Zang, Automated registration of dense terrestrial laser-scanning point clouds using curves, ISPRS J. Photogramm. Remote Sens. 95 (3) (2014) 109–121.
[29] O. Kechagias-Stamatis, N. Aouf, Histogram of distances for local surface description, in: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2016, pp. 2487–2493.
[30] B. Yang, Y. Zang, Z. Dong, R. Huang, An automated method to register airborne and terrestrial laser scanning point clouds, ISPRS J. Photogramm. Remote Sens. 109 (2015) 62–76.
[31] D. Zai, J. Li, Y. Guo, M. Cheng, P. Huang, X. Cao, C. Wang, Pairwise registration of TLS point clouds using covariance descriptors and a non-cooperative game, ISPRS J. Photogramm. Remote Sens. 134 (Supplement C) (2017) 15–29.
[32] J. Yang, Q. Zhang, Y. Xiao, Z. Cao, TOLDI: an effective and robust approach for 3D local shape description, Pattern Recognit. 65 (2017) 175–187.
[33] J. Navarrete, D. Viejo, M. Cazorla, Compression and registration of 3D point clouds using GMMs, Pattern Recognit. Lett. 110 (2018) 8–15.
[34] J. Yang, Q. Zhang, Z. Cao, The effect of spatial information characterization on 3D local feature descriptors: a quantitative evaluation, Pattern Recognit. 66 (2017) 375–391.
[35] S. Wang, L. Luo, N. Zhang, J. Li, Autoscaler: Scale-attention Networks for Visual Correspondence, arXiv preprint arXiv:1611.05837 [cs.CV] (2016) 1–10.
[36] M.E. Fathy, Q.-H. Tran, M. Zeeshan Zia, P. Vernaza, M. Chandraker, Hierarchical metric learning and matching for 2d and 3d geometric correspondences, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 803–819.
[37] S. Song, J. Xiao, Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images 139 (2) (2015) 808–816.
[38] A. Zeng, S. Song, M. Niebner, M. Fisher, J. Xiao, T. Funkhouser, 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions (2017) 199–208.
[39] G. Elbaz, T. Avraham, A. Fischer, 3D point cloud registration for localization using a deep neural network auto-encoder, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2472–2481.
[40] I. Sipiran, B. Bustos, Harris 3D: a robust extension of the harris operator for interest point detection on 3d meshes, Vis. Comput. 27 (11) (2011) 963.
[41] Q.-Y. Zhou, J. Park, V. Koltun, Fast global registration, in: Proceedings of European Conference on Computer Vision, Springer, 2016, pp. 766–782.
[42] T. Hackel, N. Savinov, L. Ladicky, J.D. Wegner, K. Schindler, M. Pollefeys, SEMANTIC3D.NET: a new large-scale point cloud classification benchmark, in: Proceedings of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-1-W1, 2017, pp. 91–98.
[43] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, J. Wan, Rotational projection statistics for 3d local surface description and object recognition, Int. J. Comput. Vis. 105 (1) (2013) 63–86.
[44] R.B. Rusu, S. Cousins, 3D is here: point cloud library (PCL), in: Proceedings of IEEE International Conference on Robotics and Automation, 2011, pp. 1–4.

Wei Li received the M.Sc. degree in college of Science from Jimei University, China, in 2013. He is currently working towards the Ph.D. degree in school of informatics with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University. His current research interests include computer vision, 3D point cloud processing, registration of 3D data, object detection and extraction, mobile laser scanning data processing.

Cheng Wang (M'11-SM'16) received the Ph.D. degree in information and communication engineering from National University of Defense Technology, Changsha, China, in 2002. He is a Professor with the Fujian Key Laboratory of Sensing and Computing for Smart Cities and an Associate Dean of the School of Informatics, Xiamen University, China. His current research interests include remote sensing image processing, mobile LiDAR data analysis, and multi-sensor fusion.

Dr. Wang has coauthored about 150 papers published in refereed journals such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and ISPRS Journal of Photogrammetry and Remote Sensing and conferences such as IGARSS and ISPRS. He is the Chair of the ISPRS Working Group I/6 on Multi-sensor Integration and Fusion (2016–2020). He is a Council Member of the Chinese Society of Image and Graphics.

Chenglu Wen (M'14-SM'17) received the Ph.D. degree in mechanical engineering from China Agricultural University, Beijing, China, in 2009. She is currently an Associate Professor with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen, China. She has coauthored more than 40 research papers published in refereed journals and proceedings. Her current research interests include 3D point cloud processing, machine learning, and robot vison. She is the Secretary of the ISPRS WG I/6 on Multi-Sensor Data Fusion (2016–2020), and Associate Editor of IEEE-GRSL.

Zheng Zhang received bachelor's degree in School of Fuzhou University, FuZhou, China, in 2016.

He is currently a master student in School of Information Science and Engineering, Xiamen University. His current research interests include point cloud registration and mobile scanning data processing.

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074

Congren Lin is a master student of Computer Technology at Xiamen University. He graduated from Minnan Normal University (China) in 2017. His major research interests focus on point cloud registration and distributed computing.

Jonathan Li received the Ph.D. degree in geomatics engineering from the University of Cape Town, South Africa. He is currently professor with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, China. He is also professor and head of the Mobile Sensing and Geodata Science Lab at the Faculty of Environment, University of Waterloo, Canada. His current research interests include information extraction from LiDAR point clouds and from earth observation images. He has co-authored more than 300 publications, over 130 of which were published in refereed journals including IEEE-TGRS, IEEE-TITS, IEEE-GRSL, IEEE-JSTARS, ISPRS-JPRS, IJRS, PERS and RSE. He is Chair of the ISPRS Working Group I/6 on LiDAR for Airborne and Spaceborne Sensing (2016–2020), Chair of the ICA Commission on Sensor-driven Mapping (2015–2019), and Associate Editor of IEEE-TITS and IEEE-JSTARS.

Please cite this article as: W. Li, C. Wang and C. Wen et al., Pairwise registration of TLS point clouds by deep multi-scale local features, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.12.074