A Systematic Approach for Cross-Source Point Cloud Registration by Preserving Macro and Micro Structures

Xiaoshui Huang, Jian Zhang, *Senior Member, IEEE*, Lixin Fan, Qiang Wu, and Chun Yuan

***Abstract*— We propose a systematic approach for registering cross-source point clouds that come from different kinds of sensors. This task is especially challenging due to the presence of significant missing data, large variations in point density, scale difference, large proportion of noise, and outliers. The robustness of the method is attributed to the extraction of macro and micro structures. Macro structure is the overall structure that maintains similar geometric layout in cross-source point clouds. Micro structure is the element (e.g., local segment) being used to build the macro structure. We use graph to organize these structures and convert the registration into graph matching. With a novel proposed descriptor, we conduct the graph matching in a discriminative feature space. The graph matching problem is solved by an improved graph matching solution, which considers global geometrical constraints. Robust cross source registration results are obtained by incorporating graph matching outcome with RANSAC and ICP refinements. Compared with eight state-of-the-art registration algorithms, the proposed method invariably outperforms on Pisa Cathedral and other challenging cases. In order to compare quantitatively, we propose two challenging cross-source data sets and conduct comparative experiments on more than 27 cases, and the results show we obtain much better performance than other methods. The proposed method also shows high accuracy in same-source data sets.**

***Index Terms*— Cross-source, point cloud, registration, graph matching, macro/micro.**

I. Introduction

THERE is currently a wide diversity of techniques for obtaining point clouds (e.g. Kinect, Lidar, range cameras, structure from motion (SFM) and simultaneous localization and mapping (SLAM)). Their registration is a long standing challenge in computer vision, computer graphics, robotics, and medical applications. Numerous methods have demonstrated convincing registration results for 3D data points which are captured by the same kind of sensors e.g. Kinect [1] or LiDAR [2]. When the point clouds come from different kinds of sensors, however, these successful techniques are invariably rendered useless because of the large variations presented in data points captured by different sensors.

Throughout this article, *Same-source* point clouds refer to 3D data from same kinds of sensor, while *Cross-source* refers to those from different sensors. In spite of the existence of same-source registration methods, their applicability is often limited. And in many real applications, multiple types of sensors have much greater ability than single sensors. For example, SLAM [3] constructs real-time complete depth and convert the depth to point cloud, SFM uses images captured by RGB cameras to create point clouds for urban scenes [4] and heritage objects [5]. Other devices, such as Kinect and Lidar, offer effective ways of producing standard point cloud datasets. With the development of new technology, there are increasing means of sensing 3D point clouds describing the same objects or scenes. Registering these datasets will have great value in cultural heritage protection, city development and technology. Unlike existing same-source methods, we propose in this paper a method which is able to tackle the cross-source point cloud registration problem.

There are a handful of attempts to deal with cross-source point cloud matching/registration. For instance, the registration of point clouds in [6] is executed using conventional iterative closest point (ICP) [7] and a number of ad hoc approaches are adopted, including removing sparse outliers and manually selecting the dense point regions; [2] uses RANSAC and downsampling strategy to conduct the registration. Although this simple downsampling strategy works well on highly dense cross-source point cloud, it faces challenge on missing data and different noise models. Super4PCS [8] use the largest common points (LCP) metric and coplanar 4-points sets to deal with the registration problem. It shows element of experiments that deal with cross-source problems. However, it faces slippage ambiguities and feature extraction ineffective in curved surface because of the coplanar 4-points mechanism. All the existed potential cross-source registration methods are originally designed for same-source registration. It was observed that adaptations applied to these approaches do not guarantee satisfactory performances for general cross-source registration problem (as demonstrated in Section V).

Manuscript received August 11, 2016; revised December 22, 2016 and April 10, 2017; accepted April 14, 2017. Date of publication April 19, 2017; date of current version May 9, 2017. This work was supported by the Nokia research funding under Grant MM12030846235. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jie Liang. *(Corresponding author: Xiaoshui Huang.)*

X. Huang, J. Zhang, and Q. Wu are with the Global Big Data Technologies Centre, School of Computing and Communication, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: xiaoshui.huang@student.uts.edu.au; jian.zhang@uts.edu.au; qiang.wu@uts.edu.au).

L. Fan is with Nokia Technologies, 33720 Tampere, Finland (e-mail: lixin.fan@nokia.com).

C. Yuan is with the Graduate School of Shenzhen, Tsinghua University, Beijing 518055, China (e-mail: yuanc@sz.tsing.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2695888

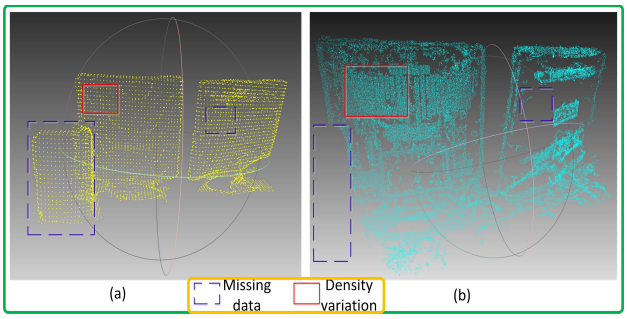


Fig. 1. Examples of cross-source point clouds coming from different sensors. 3D points are produced by two sensors. (a) Kinect point cloud, (b) Point cloud reconstructed from 2D images captured by an RGB camera.

Figure 1 shows two cross-source point clouds with two monitors and an audio equipment, which illustrates the difficulties confronting a robust registration method:

(1) *Varying densities*: large variations in the density of cross-source point clouds often lead to the failure of existing registration methods.
(2) *Missing data*: parts of point clouds are often missing when the same objects have different reflection or non-reflection in various types of sensors as a result of the imaging mechanisms of different sensing techniques. For instance, this problem is pronounced for point clouds created by SFM which is unable to generate points in uniform image regions.
(3) *Variation due to different viewpoints*: significant viewpoint differences often lead to *partial overlapping* of point clouds which is extremely challenging for many registration methods. Moreover, occlusions caused by viewpoint change may look remarkably different between cross-source point clouds.
(4) *Varying scale factors*: often there is a dramatic size change between data captured by different sensing techniques. However to estimate this scale factor reliably is a troublesome task due to serious missing data as well as large variation presented in cross-source point clouds.
(5) *Noise and outliers*: due to different kinds of sensors, large variations in noise model and outliers are common in cross-source 3D scanning data.

As demonstrated in our experiments, these combined challenges often give rise to adverse scenarios in which many existing registration methods fail miserably.

Despite the large variations in cross-source point clouds, our human vision system seems able to align them effortlessly with high accuracy. This is probably due to the fact that humans exploit the similarities between the *structures of two cross-source point clouds* instead of the detailed points. Motivated by this insight, we propose to extract the macro structure (e.g. the global outline of objects) and the micro structure (e.g. voxels and segments) of point clouds. These macro and micro structures extract from cross-source point clouds a set of local invariant components on the one hand, and preserve global geometrical structures within a graph matching framework on the other hand. They act like a net to robustly describe the invariant components of cross-source point clouds. Then, Ensemble of Shape Functions (ESF) and a novel scale-free edge feature can be used to convert the variant Euclidean graph similarity into a discriminative feature space, which overcomes missing data, scale, noise and outliers. The graph matching problem is solved by a improved graph matching method that considers global geometric constraint.

The contributions of this work are (1) a feasible structure-based framework to deal with the cross-source point cloud registration problem; (2) the macro and micro structures are robustly extracted and described into a robust feature space; and (3) an improved graph matching algorithm is proposed by considering the global geometrical constraint.

II. Related Work

Same-source point clouds are captured from the same kinds of sensors (e.g. all captured by Kinect), while cross-sources are captured from different kinds of sensors (e.g. one by Kinect, another by an RGB camera). In this section, the related methods are reviewed in terms of their ability to deal with the three challenges of cross-source point cloud registration. Existing same-source point cloud registration methods can be divided into two categories: direct methods and transformed methods. Direct methods usually minimize the distance between pairwised points or features [1], [7]–[14]. Transformed methods usually transform 3D points from Euclidean space to other models and convert the registration problem into a model correspondence problem [15]–[20].

A. Direct Methods

Direct methods aim to minimize the spatial distance between corresponding points. The associated correspondence problem and the minimization is often solved in an iterative manner. The most popular direct method is the ICP [7] algorithm, which alternates between estimating the point correspondence and estimating the transformation matrix for a given correspondence [1], [11], [12]. The vanilla ICP method [7] relies on the assumption that all points have pairwise counterparts between two sets and are very sensitive to a given initialization. Another direct methods are variants of the original ICP method which adopted either point-to-point [7], [13], or point-to-plane [21], [22] metrics to quantify the spatial distance between corresponding geometrical entities. References [22] and [23] show the point-to-plane distance metric is in general robust to the point-to-point distance metric, however, it can be less precise for large disturbances of the initial alignments and loses its advantages in unstructured environments.

The above direct methods are all heuristic methods, hence they cannot guarantee the global optimality of the solutions. Go-ICP [13] provides a globally optimal solution to ICP in 3D Euclidean registration, which combines ICP with a branch-and-bound (BnB) scheme. A roots-finding technique was used in [24] for affine invariant point cloud registration. The method is sensitive to outliers due to usage of moments. In order to deal with outliers, [25] proposes to adopt a robust L_2 Minimizing Estimate (L_2E) to assign lower probabilities to outliers in the optimization. This robust estimate is used to measure

distance between various feature descriptors for both 2D and 3D point clouds. Despite of good performance on same-source cases, like other ICP-based direct methods, it faces challenges on cross-source problems. Despite these improvements to the ICP method, the direct registration approaches above are intrinsically sensitive to missing data, large variations in point density, and scale differences, thus rendering them useless for cross-source point cloud registration (see the experimental results in Section V for examples).

RANSAC [26] has been adopted in our registration method to recover a 6DOF rigid transforms between two sets of point clouds. Despite of its popularity in computer vision applications such , RANSAC often suffers from the worst case $O(n^3)$ computational complexity for large scale 3D point cloud [14]. Various alternatives to RANSAC have been proposed to counter the cubic complexity, such as hierarchical representation in the normal space [27], super-symmetric tensors to represent the constraints between the tuples [28], stochastic non-linear optimization to reduce the distance between scan pairs [29], branch-and-bound using pairwise distance invariants [30], or evolutionary game theoretic matching [31], [32]. However, all these variants face $O(n^3)$ computation complexity and challenge in large scale cross-source point clouds.

Reference [2] utilizes RANSAC to solve cross-source point cloud registration when scale is estimated. It has several drawbacks: (1) the computation complexity is $O(n^3)$; (2) it obtains low accuracy when the geometry was scarcely representative; (3) it faces challenge in missing data. However, our method utilizes the macro and micro structures to organize the point cloud which shows better robustness and higher accuracy in the experiments.

Following the concept of RANSAC, another kind of method is 4PCS [14], which uses a randomized alignment approach and the idea of planar congruent sets to compute optimal global rigid transformation. The 4PCS method is widely used and has been extended to take into account uniform scale variations [33]. However, these methods have 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. It has great limitations when point numbers are large. To remove the quadratic complexity of the original 4PCS, [8] extends it to a fast algorithm with only linear computation time needed. This method reports the points or spheres in R^3 and uses a smart index to quickly find the matched plane in all candidate congruent 4-points planes. One cross-source point cloud registration experiment is reported in [8]. Despite of improved performance reported in [8] and [14], both methods suffer from challenges involved in cross-source registration (see Section V for thorough evaluation on a large scale dataset).

Although these direct methods show some ability in addressing elements of the cross-source problem, none of them can deal with the complete cross-source problem. To entirely deal with the cross-source problem, in this paper, a novel method is proposed. The method extracting and combining macro and micro structures is robust to large variations in density, noise and outliers. In addition, the enhanced graph matching globally registers two structures. Finally, a scale normalization step is used to eliminate most of the scale variation.

B. Transformed Methods

Transformed methods are a kind of methods transforming 3D points into other space or model, and the registration problems are converted into model fitting or distance minimizing in another space (e.g. feature space). One of the mathematical tools typically used for registration is Mutual Information (MI), which catches the non-linear correlations between the point clouds and the geometric properties of the target surface. The authors in [34] use ICP and mutual information (MI) to build one-to-one correspondence between an magnetic resonance (MR) surface and laser-scanned cortical surface. The work in [35] registers unstructured 3D point clouds by using K-means to form a set of codewords and using an estimator to optimize the MI value to obtain the final rigid relations. Cross correlation of the horizontal cross section images of the two point clouds is used in [36] to coarsely register the point clouds, and ICP is then used to refine the coarse results. These MI-based methods typically are insensitive to a variety of non-linear transformation [37]. Nevertheless they become unstable in the presence of severe missing data, which is one of the main challenges of cross-source registration.

Another type of transformed method is the feature-based method, which extracts features from 3D point clouds and transforms the point cloud registration Euclidean space into feature space. Typical 3D feature extraction methods[1] are FPFH [38], ESF [39], Spin image [40] and SHOT [41]. Deng *et al.* [18] transform each point in a point clouds into a shape representation, in order to cast the problem of point sets matching as a shape registration problem, which is the Schrodinger distance transform (SDT) representation. The problem is then transformed into solving a static Schrodinger equation in place of the consistent static Hamilton-Jacobi equation in the setting. These feature-based methods produce exciting results on same-source point clouds. However, it is very difficult to reliably extract similar features from cross-source point clouds, and performances of feature-based methods deteriorate significantly due to mismatching of distorted features.

Related to point cloud registration, another kind of methods is GMM-based methods. To deal with the noise and outliers existing in the point sets registration problem, Jian and Vemuri [16] proposed a method in which point clouds are represented as Gaussian Mixture Models (GMM) and go on to solve the registration problem by minimizing the statistical discrepancies between corresponding GMMs. This approach can be used for both rigid and non-rigid point cloud registration, and has demonstrated its ability to deal with noise and outliers to some extent. Myronenko and Song [15] introduced a motion drift idea into the GMM framework and achieved good results on rigid and non-rigid point set registration. GOGMA [42] combines Gaussian mixture model (GMM) with

[1]There is a tutorial about 3D features. http://robotica.unileon.es/index.php/PCL/OpenNI_tutorial_4:_3D_object_recognition_(descriptors)

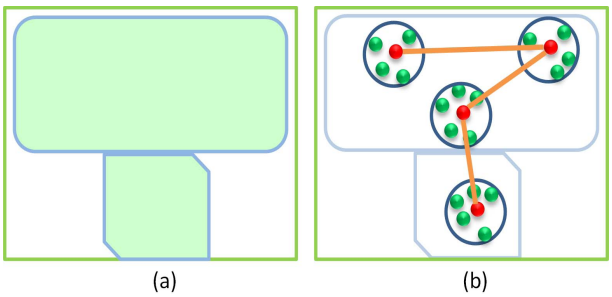


Fig. 2. Schematic diagram of macro and micro structures.

a BnB scheme. A solution to the GMM-based approach by recasting registration as a clustering problem was proposed in [43] and [44]. However, there are an increasing number of GMM models to robustly represent point clouds. When the point number increases to tens of thousands or millions, these methods are impractical in terms of both computational and memory cost. On the other hand, the GMMs depicting two point clouds are shown a lot of difference when there is missing data and large noise and outliers variations in cross-source point clouds, which makes the registration inaccurate or even fail. The experiments in Section V demonstrate these approaches do not lead to satisfactory results for cross-source point cloud registration.

The above transformed methods demonstrate ability in dealing with parts of noise and outliers or density variation, but none of them can successfully address the cross-source registration problem, which comprises issues of scale, density variation, noise and outliers and missing data. In this paper, we aim to address this tough cross-source problem. Motivated by our human registration process, a structure-based framework is proposed to robustly register two cross-source point clouds.

The remaining sections of this paper are organized as follows: Section III describes the proposed macro and micro structure representation; Section IV describes the proposed registration method based on our novel concept; Section V describes the experiments, and Section VI concludes the paper.

III. Macro and Micro Structure Representation

As mentioned in Section I, the significant challenges for 3D cross-source point cloud registration are the large variations in density, missing data, different perspective viewpoint and scale, noise and outliers between two point clouds. To address these variations, we propose to extract two geometrical structures (known as macro and micro structures) and use these structures, instead of original point clouds, for cross-source registration. Inspired by our human vision perception, these geometrical structures are designed to robustly describe the global and local invariance of cross-source point clouds in spite of large variations. When humans judge whether two objects are similar, they usually first consider the macro structure, and an overall alignment is obtained on this basis.

The macro structure is the overall outline or large-scale structure of an object or scene. It is important to note that it represents the global properties of the structure, such as the boundary outline, the contour and the shape, but not the global light, global color or global material. Figure 2(a) illustrates that the rectangle above the square (the blue outline) is the macro structure. The notion of micro structure, on the other hand, pertains to local properties such as stable cells or detailed parts of objects or scenes in question. In our work, micro structure consists of super voxels points with the same properties of 3D spatial geometry as illustrated in Figure 2(b). We use these micro and macro structures to iteratively obtain the corresponding relations between two point clouds.

IV. Cross-Source Point Cloud Registration

Definition: Cross-source point cloud registration refers to the robust registration process which is able to align 3D point clouds captured by different kinds of sensors. Concretely, the robust registration proposed in this paper aims to establish correspondence between point clouds of similar objects or scenes, in the presence of missing data, density variation, scale difference, and different noise model and outliers.

In this section, we describe the registration method based on the proposed macro and micro structure theory and describe the components that compose our system. Figure 3 provides an overview of our method in block form which comprises the following five components:

Step 1, Pre-processing: Two cross-source point clouds, which come from different sensors, are normalized to the similar scale. The details are given in Section IV-A.

Step 2, Macro/micro Structure Extraction: Cross-source point clouds are segmented into super voxels by using 3D geometric properties, and the statistical descriptor of each super voxel is to some extent invariant to scale, noise, outliers variations and missing data. These super voxels and the statistical descriptors are the micro structures constituting the building blocks of the macro structure, which is constructed in the following step (and elaborated in Section IV-B).

Step 3, Graph construction: In order to integrate invariant properties extracted at micro structure level, a graph is constructed whereas nodes are the extracted super voxels and edges represent adjacent relations between super voxels. The registration is then converted into a graph matching problem. Moreover, robust similarity measures are used to match graph nodes and edges as elaborated in Section IV-C.

Step 4, Optimization: Following factorized graph matching [47], which is an optimization algorithm with constant time complexity, we propose to adopt novel geometry constraints in the optimization. By doing so, relatively more robust registration results are achieved (see Section IV-D for details).

Step 5, Transformation estimation: This stage estimates the 6DOF rigid transformation between two point clouds. Given putative correspondences from the previous step, RANSAC [29] is first performed to remove outliers, followed by a final ICP refinement to recover the transformation (see Section IV-E for detailed account of this step).

A. Pre-Processing

The two point clouds come from different sensors and therefore have different scales. To deal with scale variation,

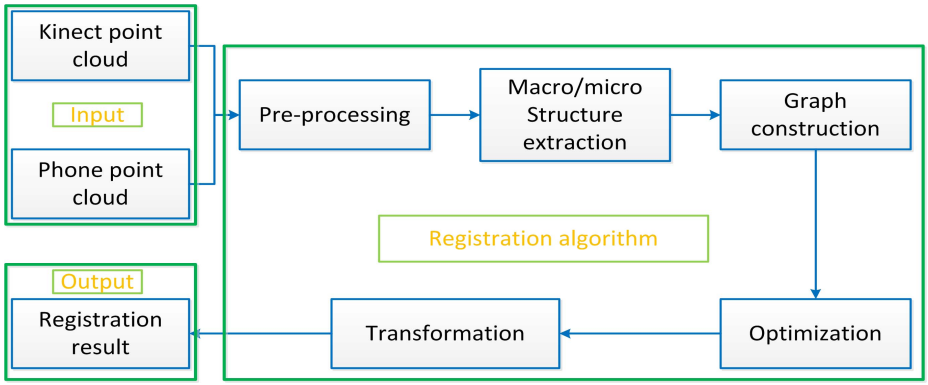


Fig. 3. Overall system workflow.

we conduct a pre-processing before the structure extraction step. For scale normalization, many automatic methods have been proposed (e.g. [2], [45]). Also, in some 3D detection or searching applications, we can estimate the scale by assuming the radius of 3D containing boxes are same and use the ratio to scale the point cloud. After this step, large scale changes are removed. Then, we introduce the new registration method to deal with small scale difference, missing data, density difference, perspective viewpoint difference, noise and outliers.

B. Macro/Micro Structure Extraction

Due to the large variations in cross-source point clouds, a method is needed to extract the invariable structures that our human can easily recognize. Figure 1 shows that even though the two cross-source point clouds have many variations, the structure can still be recognized. For these cross-source point clouds, therefore, the focus is on the structure information rather than the detailed information, since the latter is full of noise, outliers and different densities. Hence, with similar amount of segments extracted, the structure extraction can deal with different densities, noise and outliers.

We are motivated by the idea of cluster, where points with the same property are clustered together. As shown in Figure 1, humans have the ability to register these monitors at first glance. This is because the macro structure information remains in the cross-source data and when humans conduct the registration work, they are not concerned with detailed information (e.g. the location of a point). However, if we want to accurately register these two point clouds, macro structure information alone is insufficient, and micro structure information is also needed. Hence, to develop an intelligent registration algorithm, we need a method that will retain the common macro and micro structure information and ensure it is robust to varying densities and missing data.

To achieve this goal, we improve the recently developed segmentation method [46] to segment the two point clouds into many super voxels and extract the direct adjacency graph of these voxels. As the segmentation method adheres to object boundaries while remaining efficient by only using the 3D geometric property, it obtains robust results for two point clouds, regardless of different density, perspective viewpoint,

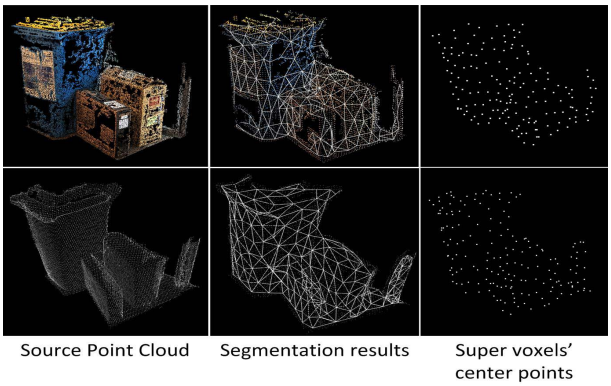


Fig. 4. Results of macro/micro structure extraction. The first column is the source point clouds of SFM (above) and KinectFusion (bottom); the second column is the segmentation results and the connection relationship; the third column is the segmented super voxels' central points.

noise and missing data (see the third column of Figure 4). Figure 4 shows that the center of the segmented super voxels deals with much of the noise, density and missing data problem. Unlike [46], we do not flow back at the extraction of each edge in the adjacency graph extraction step, which means that the direction information is considered in our new adjacency graph. This is because in the following optimization step (Section IV-D), direct graph matching achieves more robust results than indirect graph matching [47]. This revision is a key element to ensure that these extracted voxels are correctly and robustly registered. At the same time, the ESF descriptor [39] for each voxel is extracted to describe the statistical property as a local structure. Based on the definition of macro and micro structures, therefore, each segmented super voxel is a micro structure and the whole of the adjacency graph and voxel centers are macro structures. After these structures have been extracted, they are integrated as a graph in the graph construction stage.

C. Graph Construction and Description

The contribution in this section is that macro and micro structures are assembled into graph and robustly described without influence of small scale difference, noise and outliers. Therefore, the registration problem is transfered into a graph

matching problem. We select graph because it is a strong tool to maintain the properties(e.g. topology) of the structures.

Before introducing the new method, the graph matching notations are introduced. A graph with n nodes and m directed edges is defined as $\check{\mathrm{C}}= \{P, Q, G, H\}$. P and Q are the features for the nodes and edges of the graph, which are defined as $P = [p_1, \cdots, p_n] \in R^{d_p \times n}$and $Q = [q_1, \cdots, q_m] \in R^{d_q \times m}$ respectively. For example, p_i could be a SIFT descriptor or ESF descriptor extracted from the original data around the i^{th} node and q_i could be the length of the i^{th} edge. $G, H \in \{0, 1\}^{n \times m}$ is a node-edge incidence matrix which describes the topology of the graph. We define $g_{ic} = h_{jc} = 1$ if the c^{th} edge connects the i^{th} node and the j^{th} node, and zero otherwise. To perform graph matching, given a pair of graphs, we first need to define P and Q. Next, we compute two affinity matrices, $K_p \in R^{n_1 \times n_2}$ and $K_q \in R^{m_1 \times m_2}$ to measure the similarity of each node and edge pair, then $k^p_{i_1\ i_2} = \phi_p(p^1_{i_1}, p^2_{i_2})$ measures the similarity between the i_1^{th} node of $\check{\mathrm{C}}_1$ and the i_2^{th} node of $\check{\mathrm{C}}_2$, and $k^q_{c_1c_2} = \phi_q(q^1_{c_1}, q^2_{c_2})$ measures the similarity between the c_1^{th} edge of $\check{\mathrm{C}}_1$ and the c_2^{th} edge of $\check{\mathrm{C}}_2$. Only when we define these matrices correctly, can we use graph matching method.

In this section, a graph descriptor method is proposed to convert the similarity measurement from Euclidean space to a robust feature space. To robustly deal with the missing data, noise and outliers in cross-source point clouds, with exception of structure extraction, a structure-retaining similarity measurement method is needed. In other words, the graph should be robustly described. As previously discussed, humans can still register cross-source point clouds correctly by their structure. Similar to the human register's process, the graph is constructed as a expression of the relations between structures. This is another key element obtaining robust registration results. The graph has the ability of being robust to large variations in perspective viewpoint noise, outliers and missing data of cross-source point clouds. Here, we describe how to design the nodes and edges of these graphs, and their similarity measurement.

1) Graph Node and Similarity Measurement: To robustly represent the micro structures of point clouds, the method should be resilient to the large variations in density and missing data. We segment the super voxels of two point clouds and extract the centroid point of each super voxel. Hence, the density variation, most noise and outliers variation are solved by extracting similar amount of segments. The graph node E is then constituted by these centroid points. To correctly match these nodes, they need to be described discriminately. Due to the cross-source problems discussed above (i.e. varying density, missing data and variations in scale and perspective viewpoint), using only the coordinates of these centroid points cannot describe discriminately for nodes and the original matched node pairs are very rare. To robust deal with the cross-source problem, we select the ESF descriptor [39] instead of using conventional nodes' coordinate because the ESF descriptor is a global descriptor that adds up the properties of the distance, angles and area of the point clouds. Using the ESF descriptor, we transform the variable Euclidean space into feature space (ESF 640). If two points come from the corresponding segments, the ESF descriptors will mostly be the same and should be matched, even though the centroid point may not perfectly match in the Euclidean space. Due to we compute the ESF by scaling the 3D containing box in a unit sphere and use statistical property, the node descriptor is robust small scale difference, noise and outliers.

The node similarity matrix K_p is computed by comparing the distance between the nodes' ESF descriptors(see left hand of Figure 5). Here, the node similarity is not computed in Euclidean space but in feature space. Because ESF is a statistic and global descriptor, it has the ability to avoid the large local variations in Euclidean space and hence is more robust to the cross-source problem. The node similarity is

$$K_p = \overline{D}_p \tag{1}$$

where $\overline{D}_p$ is the normalized distance of two 3D points' ESF descriptors, $\overline{D}_p = D_p/||D_p||_\infty$. D_p is the square distance between i_{th} row of descriptor $p^1_{i_1}(esf)$ and j_{th} row of descriptor $p^2_{i_2}(esf)$ (i.e. $D_p(i, j) = \|p^1_{i_1(esf)}(:, i) - p^2_{i_{2(esf)}}(:, j)\|^2_2$). $p^1_{i_{1(esf)}}$, $p^2_{i_{2(esf)}}$ are the ESF descriptors of $p^1_{i_1}$ and $p^2_{i_2}$.

2) Graph Edge and Similarity Measurement: To robustly and discriminately describe the point cloud, it is necessary to build the edges accurately to reflect the relations between micro structures. We record the adjacent relations (extracted in Section IV-B) between super voxels and use these adjacent relations as edges Q. The adjacent relations correctly reflect the relations of the super voxels through the boundary property. The edges need to be described discriminatingly and meaningfully to ensure they are correctly matched. We need to reiterate that humans can register these two cross-source point clouds because their structures are almost the same. We therefore need to retain the structure property of these two graphs in describing edges. Edge direction is also an important factor for the structure of the graph, in spite of the edge distance.

In this paper, we use the spatial distance and geometric properties of these edges (see right hand of Figure 5). The Euclidean distance and Eular angles of two connected nodes are combined to construct a descriptor vector for describing the edges Θ_Q: (Θ, dr), where $\Theta = (\theta_x, \theta_y, \theta_z), dr = d/r, d = \|P_i - P_j\|_2$, r is the resolution of its belonging point cloud, $\theta_z = acos(z/dr), \theta_x = acos(x/(dr * sin(\theta_z)), \theta_y = acos(y/(dr * sin(\theta_z))$. We compare the similarity by comparing squared distance matrix of these descriptors and obtain D_q, where $D_q(i, j) = \|\theta_{Q}{}^1_i(:, i) - \theta_{Q}{}^2_j(:, j)\|^2_2$. To make a more robust comparison, we normalize the descriptor $\overline{D_q} = D_q/||D_q||_\infty$, and the edge similarity matrix K_q is computed by

$$K_q = \overline{D}_q \tag{2}$$

This is a simple mean of obtaining features in 3D point clouds (relative Euclidean distance ratio and Eular angles between two points). At the same, the edges geometry of cross-source point cloud is described by taking the spatial

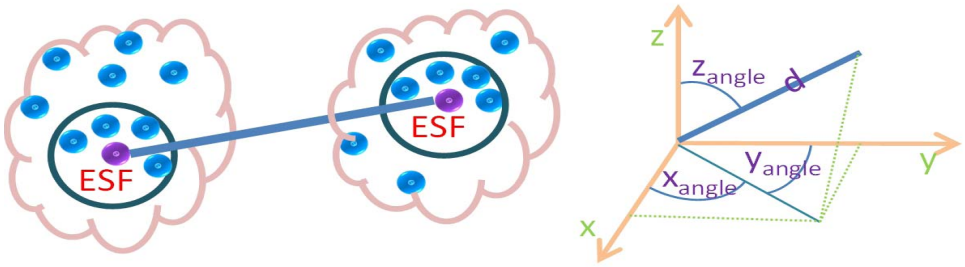


Fig. 5. Schematic diagram of graph nodes and edges. The node is described by ESF descriptors. The edge is described by 3D spatial relations in Euclidean space between neighbors including angles and relative distance.

relations and structures into consideration. The edge feature in Equation 2 is robust to scale difference, rotation and translation. The ability of registering the cross-source point clouds will be demonstrated in the experiment section.

D. Optimization

Due to the missing data, two graphs usually have different number of nodes. Moreover, the transformation of two cross-source point cloud often lead to poor initializations for graph matching. We select FGM [48] because it is initialization-free and can handle graphs with different nodes. To deal with the missing data, noise and outliers, we propose an enhanced factorized graph matching method which considers global geometry constraint to deal with the local minima problem in graph matching. Before introducing our method, we briefly review graph matching and FGM [48]. Suppose there is a pair of graphs, $\breve{\mathrm{C}}_1 = \{P_1, Q_1, G_1\}$ and $\breve{\mathrm{C}}_2 = \{P_2, Q_2, G_2\}$. The problem of graph matching consists of finding a correspondence between the nodes of $\breve{\mathrm{C}}_1$ and $\breve{\mathrm{C}}_2$ that maximizes the following score of global consistency:

$$J(X) = \Sigma_{i_1 i_2} x_{i_1 i_2} k^p_{i_1 i_2} + \Sigma_{\substack{i_1 \neq i_2, j_1 \neq j_2 \\ h^1_{i_1 c_1} g^1_{j_1 c_1} = 1 \\ h^2_{i_2 c_2} g^2_{j_2 c_2} = 1}} x_{i_1 i_2} x_{j_1 j_2} k^q_{c_1 c_2} \tag{3}$$

where $X \in \{0,1\}^{n_1 \times n_2}$ denotes the node correspondence, for example, if i^{th}_1 node of $\breve{\mathrm{C}}_1$ and the i^{th}_2 node of $\breve{\mathrm{C}}_2$ correspond, $x_{i_1 i_2} = 1$. $k^p_{i_1 i_2}$ is an element of K_p in i^{th}_1 row and i^{th}_2 col, $k^q_{c_1 c_2}$ is an element of K_q in c^{th}_1 row and c^{th}_2 col.

It is more convenient to write $J(X)$ in a quadratic form, $x^T K x$, where $x = vec(X) \in \{0,1\}^{n_1 n_2}$ is an indicator vector and $K \in R^{n_1 n_2 \times n_1 n_2}$ is computed as follows:

$$k^p_{i_1 i_2 j_1 j_2} = \begin{cases} k^p_{i_1 i_2} & \text{if } i_1 = j_1 \text{ and } i_2 = j_2 \\ k^q_{c_1 c_2} & \text{if } i_1 \neq j_1 \text{ and } i_2 \neq j_2 \text{ and} \\ & h^1_{i_1 c_1} g^1_{j_1 c_1} h^2_{i_2 c_2} g^2_{j_2 c_2} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

A factorized graph matching (FGM) method [48] is used to develop an initial-free optimization scheme with no accuracy loss to address the non-convex issue. This method divides matrix K into many smaller matrices. Using these smaller matrices, the graph matching optimization problem can be transformed to iteratively optimize the following non-linear problem:

$$\max_X J_\alpha(X) = (1-\alpha) J_{vex}(X) + \alpha J_{cav}(X) \tag{5}$$

where J_{vex} and J_{cav} are two relaxations in FGM [48].

Enhanced factorized graph matching. Although FGM iteratively uses a different α to apply the Frank-Wolfe (FW) algorithm to avoid local optima, which are nevertheless persistent in cross source registration according to our observations. To effectively deal with the local optima in FGM, we improve the algorithm by incorporating geometry constraints into the energy function 5 which is minimized by a novel iteration method:

$$\max_X J_\alpha(X) = (1-\alpha) J_{vex}(X) + \alpha J_{cav}(X) + J_{smooth}(X) \tag{6}$$

Geometrical constraints as such are imposed by enforcing neighboring points move consistently with two correspondence points in question. We use this property to avoid the local minima and obtain more accurate transformation relations. We design this regulation term by considering the projection difference of neighboring correspondence points. $J_{smooth}(X)$ is defined as

$$J_{smooth}(X) = -\sum_{i \in X} \sum_{j \in D} \frac{|\|p_i - p_j\| - \|p_{im} - p_{jm}\||}{(n_1 * n_2)} \tag{7}$$

where D represents connection points with point i, p_{im} is the matched point of p_i and p_{jm} is the matched point of p_j, n_1 and n_2 are node number of two graphs. We can easily obtain these points in D by searching matrix G in the graph.

To optimize this nonlinear problem, we use FW [49], which iteratively updates the solution of $X^* = X + \lambda Y$. Given an initial X_0, we update X through optimal direction Y and step size λ. As a smooth term needs a correspondence relation, we divide the computation of optimal direction Y into two steps: (1) compute initial Y_0 using J_{vex} and J_{cav}. We compute an initial Y_0 by solving the Hungarian algorithm which is linear programming similar to FGM [48]. (2) computes the final Y by using J_{vex}, J_{cav} and J_{smooth}. We compute the energy of the smooth terms using Y_0 and obtain the final Y using the new energy. As the computation of Y involves linear programming, adding one more computation step of Y is not computationally costly. Similar to the FGM strategy, we also use 100 times iteration to discard the inferior temporary solution and compute an alternative solution using another

Algorithm 1 Pseudo-code of the registration algorithm.

Input: Cross-source point clouds.
Output: Registration result and Transformation matrix

```
Scale normalization by Eq. (1).
Macro/micro Structure extraction.
Graph construction using Eq. (2) and Eq. (3).
Initialize X to be a doubly stochastic matrix;
for α = 0 : 0.01 : 1 do
    for nIt = 1 : 100 do
        Compute J_vex and J_cav from X_0
        Compute Y_0 using J_vex and J_cav
        Compute J_smooth using Y_0 as Eq. (8)
        Compute Y using J_vex, J_cav and J_smooth
        Compute the update direction Y = Y − X_0
        Compute update step λ
        Compute the updated X and set X_0 = X
    end for
end for
Transformation estimation.
```

FW step to optimize J(X). The final transformation matrix is computed in the next stage, following optimization.

E. Transformation Estimation

Given putative correspondences from the graph matching step, this stage aims to recover the 6DOF rigid transformation which aligns two point clouds. A 3D RANSAC [29], [52] is adopted to remove outliers in putative matching results, and remaining inlier points are used to compute the transformation directly [citation of applied method?].

Since the registration in our work is restricted to rigid transform, there exist inevitable non-zero distances between corresponding points. As a remedy to persistent re-projection error, a final Iterative Iterative Closest Point (ICP) refinement is adopted to further minimize the re-projection error.

After completing these steps, we register the two cross-source point clouds together. The pseudo code of the complete registration algorithm is shown in Algorithm 1.

V. Experiments

The proposed method provides a solution to the cross-source point cloud registration problem. In this section, we conduct comparative experiments with respect to state-of-the-art registration methods on both same-source datasets and challenging cross-source datasets.

A. Experimental Setup

For comparison purposes, we select the representative 3D registration algorithms ICP [23], Go-ICP [13], 4PCS [14], super-4PCS [8], TPS-RPM [9], GMMReg [16], CPD [15], SparseICP [22] and JP-MPC [43] as the comparison methods. Since the code of [2] was not made available, we reimplemented the registration part (RASANC algorithm) and used our scale normalization to replace its scale estimation part. In order to make a fair comparison, we follow [2] to use Constrained Poisson Disk approach [50] to downsample the original point cloud to approximately 10%. This is because experiments cannot be conducted on a large number of point cloud registrations using TPS-RPM, RANSAC and JR-MPC due to their large memory cost.

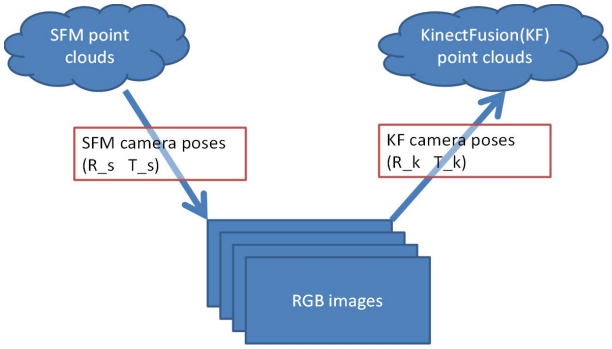


Fig. 6. A schematic diagram of cross-source point clouds construction using both VSFM and KinectFusion. See text below for the ground-truth registration between two cross-source clouds.

For the same-source database, we conduct a quantitative evaluation experiment with the 3D models "Bunny," "Lucy" and "Armadillo" from the Stanford 3D scanning repository.[2] We only consider points with positive z coordinates. For each view, following [43], the original models are rotated in the xz-plane and the points with negative z coordinates are rejected. In this way, only a part of the object is visual; the point sets do not fully overlap, and the extent of the overlap depends on the rotation angle, as in real scenarios.

There are three types of cross-source database that were used in all experiments reported in this article.

Database A: KinectFusion and Phones' RGB camera. We build a database with four sets of cross-source objects, which are typical examples of the different properties of cross-source point clouds. We use KinectFusion to build one source, and use VSFM to build another source for images which are captured by IPhone 6S Plus. As KinectFusion uses a physical device to capture 3D points, it can usually obtain dense and uniform point clouds on an object's surface. However, VSFM is a method by which 3D point clouds are built from 2D images. It uses keypoints to initially build highly accurate 3D points and uses CMVC [10] to build more dense 3D points. These two sources are typical examples of cross-source problems, as previously discussed.

Database B: KinectFusion and KinectFusion's RGB camera. We build the database in the following steps: Step 1, a modified KinectFusion method is used to output the image sequence and camera pose of each image when capturing KinectFusion point clouds. Step 2, another point cloud is computed using these images and VSFM. A set of camera poses is computed using VSFM. As these two cross-source point clouds come from the same set of image sequences, the camera poses of KinectFusion and VSFM should be the same. Figure 6 illustrates a general schematic diagram in which cross-source point clouds are reconstructed using VSFM and KinectFusion respectively. Ground truth registration between these point

[2] https://graphics.stanford.edu/data/3Dscanrep/3Dscanrep.html

clouds are established via following procedures. The VSFM point cloud is back-projected into the image coordinate system and then re-projected into the KinectFusion coordinate system. To avoid the inaccuracy of camera pose computation in VSFM and KinectFusion, we consider many poses whose reprojection error is less than σ (σ=0.5), and use these camera pose center points and the least-squares method to compute the final rigid transformation between these two camera center points. The rigid transformation matrix is built on critical prior information and can therefore be used as ground-truth. These benchmark data contain 13 datasets and can be used to perform quantitative evaluation for cross-source point cloud registration.

Database C: Synthetic cross-source point clouds. We build the synthetic datasets according to the cross-source properties. Simulating the cross-source problems discussed in Section I, we build the synthetic datasets in three steps. Step 1: Different density and different viewpoints. We up-sample the original point cloud by adding one point to the gravity center of each triangle of the original surface. We then remove all points whose z coordinate is less than 0 in the upsampling point cloud, and obtain view 1 as S1. The coordinate system is rotated 60° relative to the y axis and down-samples every 3 points. We obtain view 2 by removing all the $z \leq 0$ points. Step 2: Missed point cloud construction. Starting from view 2, we randomly delete ten parts in the plane to simulate a VSFM point cloud. Step 3: Rigid transformation. A random scale of 3 to 5, a random rotation matrix in the x, y, z axis of 30° to 60°, and a random translation in the z axis of 0 to 50% of the largest point-point distance are added to view 2. Step 4: Construction of noise and outliers. Each point is perturbed by a white Gaussian noise with predefined signal-to-noise ratio (SNR)[3] $SNR = 40dB$, and the noise is added to the original view 2 point cloud. The outliers are constructed by down-sampling the original view 2 to 30% and adding random offset[4] to the coordinate of the down-sampled point cloud. The noise and outliers are combined to form the final point cloud S2. Two point clouds (S1 and S2) constructed as such simulate large variations in density, scale, perspective viewpoint and missing data, which are typically presented in real cross source datasets. Ten cross-source datasets are synthesized using Stanford 3D objects.[5] Figure 7 shows one sample of the synthetic datasets.

B. Discussion About Parameters in Structure Extraction

Regarding to parameters, the main considered element is segmentation of structure extraction in Section IV-B. Figure 10 shows the measurement registration accuracy of different structure extraction parameters through the root-mean-square error (RMSE). It shows that low registration error when we set the radius of the super voxels as 1% of the point cloud radius for both the KinectFusion and SFM point clouds, and set seed time as 8 times. For the threshold of RANSAC and ICP, we set it as the resolution of the point cloud. For the proposed method, we first compute the transformation matrix on macro and micro structures and then use the transformation matrix to perform transformation on the original cross-source point cloud.

[3]Note that SNR is inversely proportional to the variance of added Gaussian noise. In this work we set SNR at a fixed value, thus, the variance of added Gaussian noise is actually adaptive to the variance of input point clouds.

[4]offtset ranges from 0 to 1% of the largest point-point distance

[5]http://graphics.stanford.edu/data/3Dscanrep/

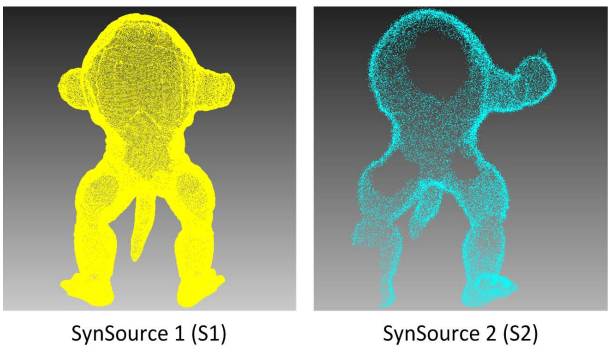


Fig. 7. Samples of synthetic cross-source datasets.

TABLE I
RMSE RESULTS OF THE JR-MPC, ICP AND CSGM

RMSE-D	JR-MPC	ICP	CSGM
Armadillo	1.456	1.725	0.508
Bunny	1.789	2.022	1.792

C. Experiments on Same-Source Point Cloud Datasets

We use the RMSE of the rotation parameters for the registration error since translation estimation is not challenging. We select "Armadillo" and "Bunny" with 30° and 45° respectively(SNR = 10dB and 20% outliers).

Extensive evaluation and comparison of registration methods has been conducted by JR-MPC on same-source databases. We only run JR-MPC, ICP and the proposed method(CSGM) on the same-source database. Table I shows the quantitative comparison results. Note that ICP is more sensitive to outliers as a result of the one-to-one correspondence and incurs a higher rate of error. JR-MPC demonstrates similar performance to the proposed method, because GMM models perform well when the overlapping areas do not have a significant amount of missing data or the scale problem. We can see from this experiment that the proposed method is robust to outliers, noise and angle variations on same-source point clouds. The quanlitative results are shown in Figure 9.

In addition, we test the robustness of the algorithms in terms of the rotation angle between two point clouds to capture the difference degree of the angles. We register the points under different angles from $-90°$ to 90° and use RMSE to test the performance. The results are shown in Figure 8 and it can be seen that the angles have a different effect on the final error. As the proposed method uses a macro and micro structure to describe the point clouds, it shows robustness in dealing with outliers, noise and missing data on same-source database. However, the error increases when the rotation angle increases, similar to other methods. With the increase in the

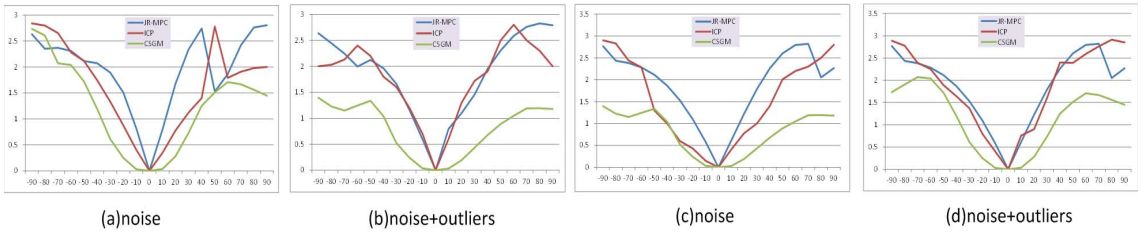


Fig. 8. RMSE as a function of the overlap (rotation angle) when two point sets are registered (SNR=20dB, 30% outliers) (a),(b) "Armadillo" (c), (d) "Lucy."

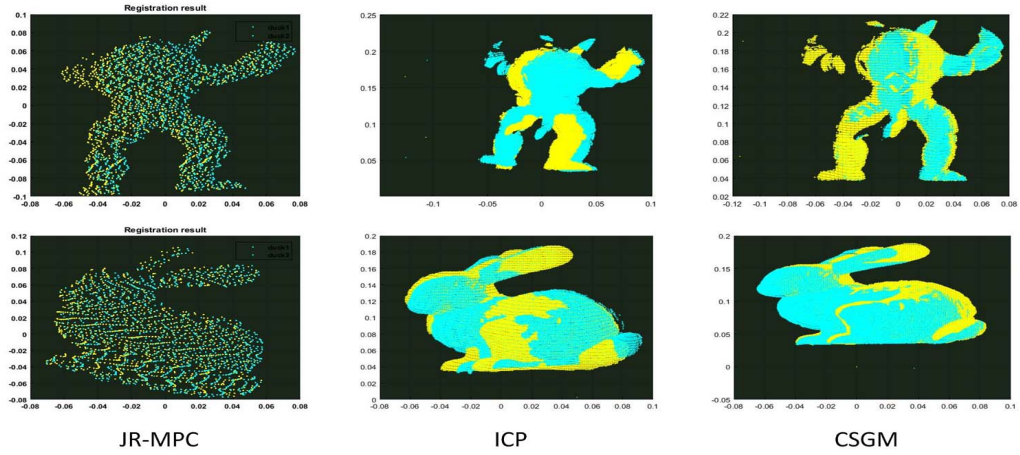


Fig. 9. Two point clouds registration results on same-source datasets.

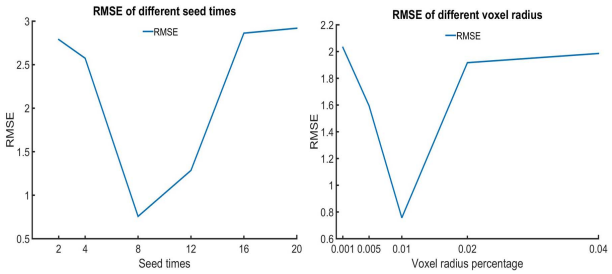


Fig. 10. The RMSE under different parameters of structure extraction.

rotation angle, the outliers and the mismatched parts become a larger proportion of each point cloud.

D. Evaluation on the Importance of Key Components

In this section, we conduct experiments to demonstrate the major contribution in dealing with cross-sourced point cloud registration. We done four experiments on four datasets: *Twobox, Threemonitor, Monitor, Chair*. (a) We remove the proposed node and edge descriptors, and other components are same to the proposed method. For graph description, like most other graph matching method, node is described by point's 3D coordinate and edge is described by Euclidean distance. (b) We only remove the edge descriptor and use Euclidean distance to describe the edge. (c) We remove the Macro/micro structures and use delaunay triangle method to build the edges on the downsample points. (d) We remove the RANSAC/ICP step. (e) The proposed method.

TABLE II

RMSE RESULTS OF COMPONENT EXPERIMENTS. (a) ONLY WITHOUT NODE AND EDGE DESCRIPTOR; (b) ONLY WITHOUT EDGE DESCRIPTORS; (c) ONLY WITHOUT MACRO/MICRO STRUCTURES; (d) ONLY WITHOUT RANSAC/ICP STEP; (ALL) ALL COMPONENTS COMBINED TOGETHER

RMSE	(a)	(b)	(c)	(d)	All
TwoBox	0.062	0.006	0.091	0.0045	**0.003**
Threemonitor	0.0021	0.0016	0.0036	0.0024	**0.0013**
Monitor	0.0017	0.0009	0.0015	0.0009	**0.0007**
Chair	0.0068	0.0017	0.0053	0.0018	**0.0016**

Table II illustrates the RMSE results of the four experiments. The results show without the structure extraction, the registration error increase dramatically. It shows the proper structure extraction plays the most important role in obtaining good registration results. Followed by the importance of node and edge description and then the importance of RANSAC/ICP. Figure 11 shows the quanlitative registration results. The results illustrate that the proposed macro/micro structures and the node/edge descriptor are the key elements in obtaining accurate registration result. Without these key elements, it cannot obtain satisfactory results.

E. Comparisons on Existing Datasets

According to Section 2, while several recent works are focus on the issue of point cloud registration, only a few of them can actually be applied on real cross-source point clouds.

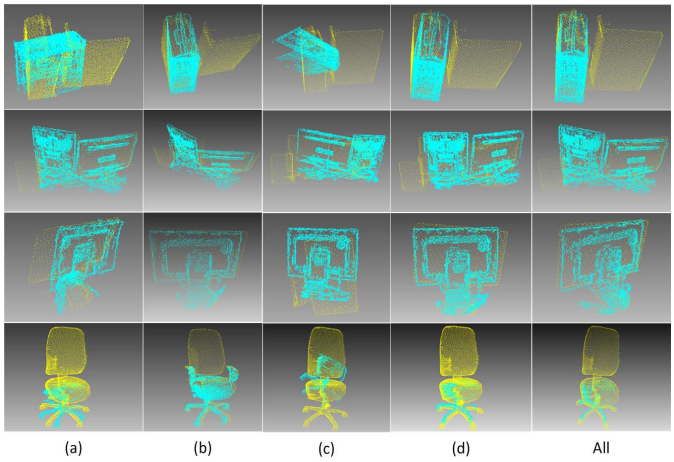


Fig. 11. Registration results of with and without components in our method. (a) without the proposed node and edge descriptors; (b) without the proposed edge descriptors; (c) without the proposed Macro/micro structures; (d) without the RANSAC/ICP step; (e) result of the proposed method.

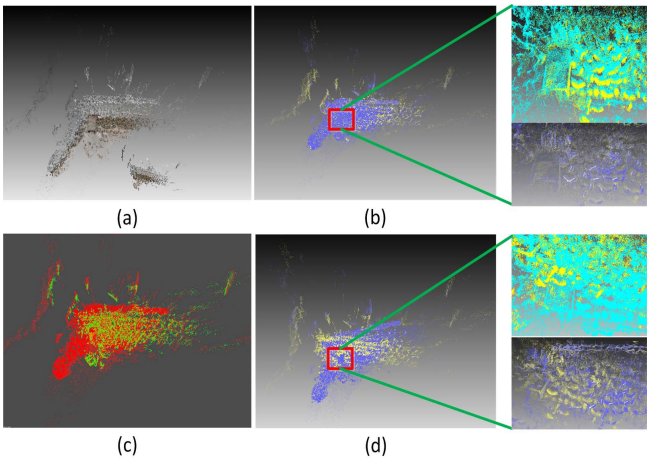


Fig. 12. The top row: (a) is original point clouds, (b) is our registration result. The bottom row: (c) is the manual registration result from [45], (d) is the RANSAC registration result of [2]. The last column is the registration details in different color.

This is mainly because the mixture of many challenges. Other techniques could not be reproduced [2], since the code was not made available to the community.

Reference [2] has two parts: scale estimation and RANSAC registration. We re-implement the registration part (RANSAC algorithm) of [2] and compare with it, which is to our best knowledge the only approach to handle cross-source registration problems. The registration results are shown in Figure 12. The results show the proposed method obtains accurate registration results which is similar to manually registration result in [45]. Our registration result is better than [2], see red box. The reason is that [2] obtains low accuracy when the geometry was scarcely representative. However, our method utilizes the macro and micro structures which has no such limitations.

Since our method aims at dealing with challenging real dataset, we also tested the proposed method on the models of Pisa Cathedral.[6] Figure 13 shows the proposed method is

[6]https://www.irit.fr/recherches/VORTEX/MelladoNicolas/category/datasets/

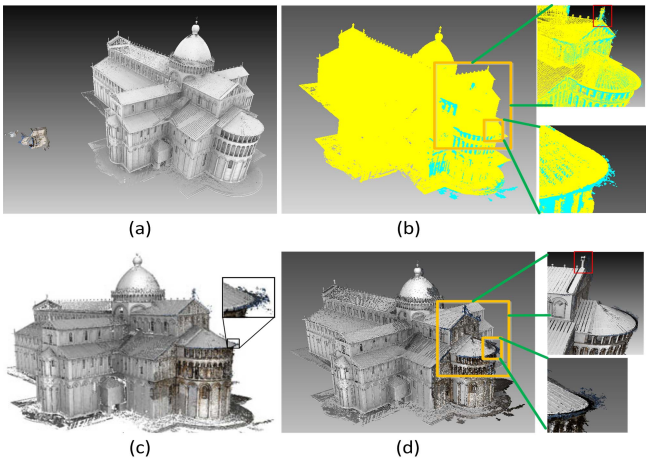


Fig. 13. The left is original point clouds of Pisa Cathedral taken by Lidar and PMVS; the right is registration result. (a) is the original point clouds in same coordinate system, (b) is our registration result, (c) is the registration result from [2]. (d) is the RANSAC registration result of [2]. Our registration result is better than [2], see red box.

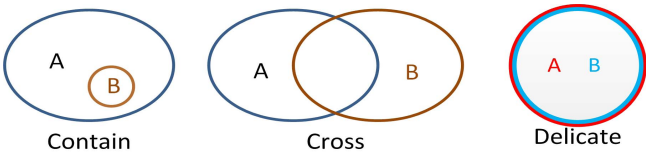


Fig. 14. The visual effect of three different kinds of overlapping.

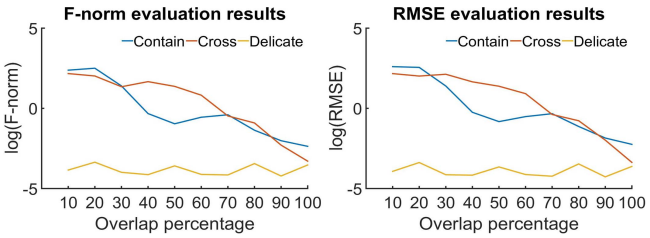


Fig. 15. The left is results under F-norm evaluation metric and the right is under RMSE evaluation metric. The results of different overlapping ratio in three cases: contain, cross and delicate. The overlap ratio ranges from 0.1 to 1.0.

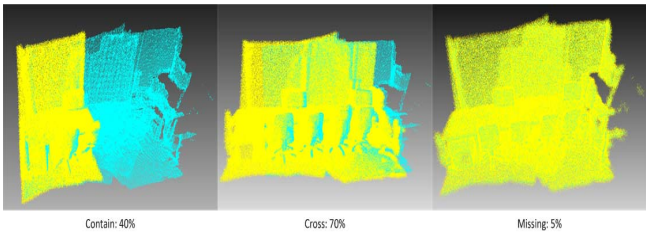


Fig. 16. Selected visual result for different overlapping ratio.

able to handle complex scenes when a coarse detection method is applied.

F. Evaluation on Different Overlap

To evaluate the ability of our method in dealing with different overlap variations (e.g. different coverage, missing data),

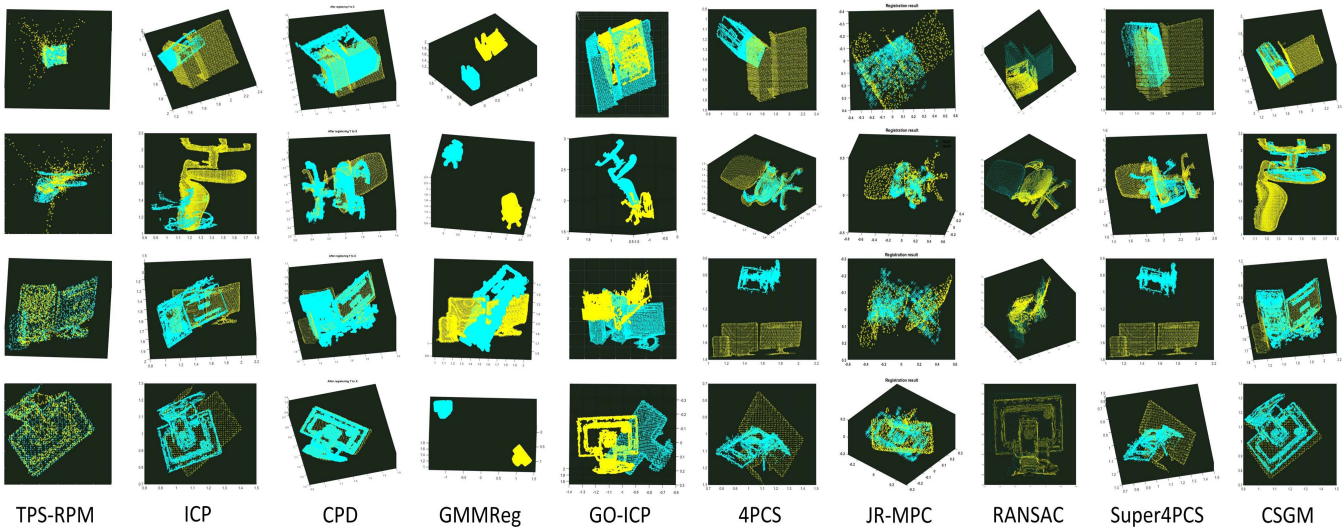


Fig. 17. Cross-source point cloud registration results on Database A.

we capture a scene by KinectFusion and simulate cross-source point clouds like Dataset C. We design three different kinds of overlapping which shows in Figure 14. 'Contain' means B is a subset of A, 'Cross' means there exists non-empty intersection between A and B, "Delicate" refers to two point clouds which look similar on large scales but have subtle differences e.g. in density and small missing parts.

We evaluate the proposed method on these overlapping cases. Figure 15 shows the evaluation results on different kinds of overlapping under F-norm and RMSE estimation metric. Figure 16 shows the registration results with 40% overlapping ratio in containing case, 70% overlapping ratio in containing case, 5% in missing data case. It shows the proposed method are very robust to containing and missing cases, and show part ability in the cross overlapping case. We observed that our approach tended to fail for data with very low overlap in the cross case (fails when the overlapping ratio less than 60%). However, other cross-source registration methods also face similar difficulty (e.g. [2]).

G. Qualitative Evaluation on Real Cross-Source Point Clouds

As discussed previously, cross-source point clouds have large variations in density, scale, perspective viewpoint and missing data which make the already difficult point cloud registration problem even more challenging. To test the ability of our method to register cross-source point clouds and compare with other related methods, we conduct qualitative analysis experiments on four real cross-source datasets: *Twobox, Chair, Threemonitor and Monitor*. To make a thorough comparison, TPS-RPM [9], ICP [23], CPD [15] , GMMReg [16], Go-ICP [13], 4PCS [14], JR-MPC [43], SparseICP [22], RANSAC of [2] and Super4PCS [8] are selected as our comparison methods. Since many of the selected methods are unable to handle the scale problem, we first normalize the scale difference for ICP, Go-ICP, 4PCS, Super4PCS, TPS-RPM, GMMReg, SparseICP, RANSAC and JR-MPC using the preprocessing step. In our proposed method, scale normalization is also an preprocessing step.

Figure 17 shows the final registration results which indicate that the proposed method gives successful registration results, whereas the other methods fail in almost all cases. This is because many of these methods cannot handle scale problem, density problem or missing data. Note that TPS-RPM obtains good result in *Threemonitor* and *Monitor*, but fails in *Twobox* and *Chair*. RANSAC [2] obtain similar registration results to us in *Chair* and *Monitor* while poor performance in Twobox and *Threemonitor*. The reason is that [2] obtains low performance when the geometry was scarcely representative (e.g. large flat surfaces). The proposed method obtains good results in cross-source datasets because it describes the micro and macro structure of point clouds robustly, and uses both the structures to optimize the registration relations. Also, the graph descriptors convert graph similarity from variant Euclidean space into a discriminative feature space which overcomes most of the Euclidean cross-source variations.

Note that we do not iteratively conduct enhanced graph matching and outliers detection (RANSAC). We find that when we use the outlier detection method to remove graph nodes, the graph structure in some cases is totally different. As a alternative solution, we use ICP to smoothly refine the graph matching result to obtain a final registration result.

H. Quantitative Evaluation on Real and Synthetic Cross-Source Point Clouds

To test the ability of the proposed method, we conduct quantitative evaluation on real and synthetic cross-source databases.

We first conduct quantitative evaluation on Databases B which contains real cross-source point clouds. We compare it in the quantitative evaluation experiments with methods that deal with rigid registration. Based on our knowledge, we compare our proposed method with ICP [23], GO-ICP [13], SparseICP [22], GMMReg [16], JP-MPC [43], CPD [15] and 4PCS [14], super-4PCS [8] and RANSAC of [2] on a cross-source database.

Many rigid methods are unable to handle the scale problem. To make a fair comparison, scale normalization is performed

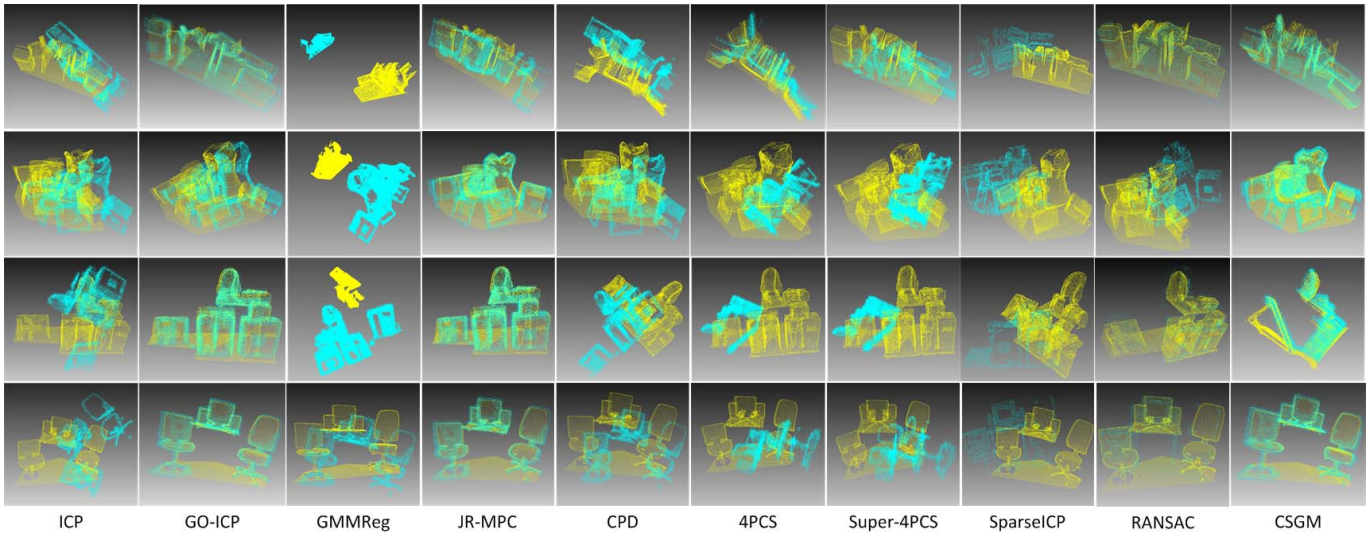


Fig. 18. Selected visual effect of cross source point clouds registration results on the Database B. Rows are datasets and columns are methods.

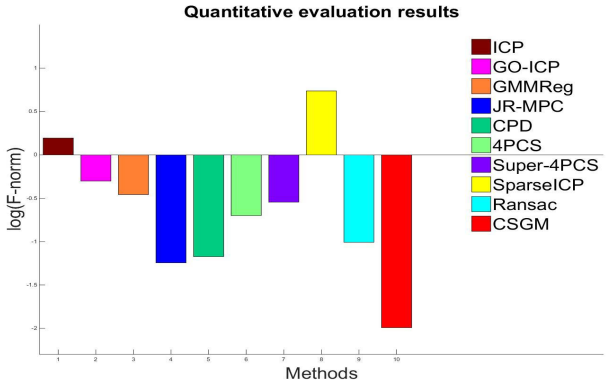


Fig. 19. Quantitative evaluation results of mean F-norm between transformation matrices on Database B.

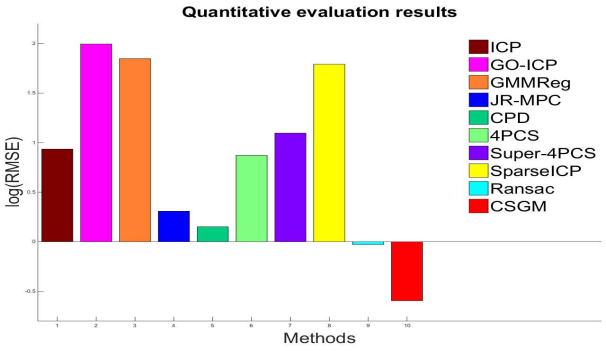


Fig. 20. Quantitative evaluation results of RMSE on Database C.

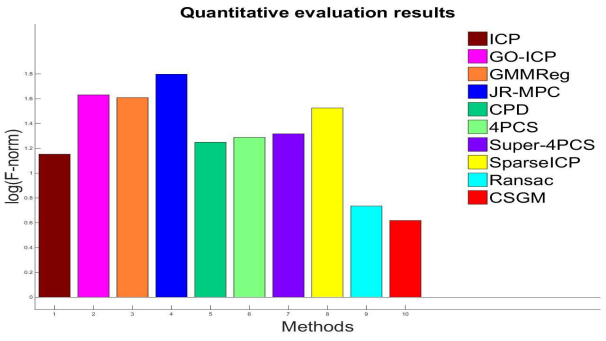


Fig. 21. Quantitative evaluation results of F-norm on Database C.

before running these methods except for CPD which estimates scale internally. The transformation matrix for each comparison method is then computed and these matrices are used for quantitative evaluation. In this experiment, the matrices are all transformed from VSFM point clouds to KinectFusion point clouds. The VSFM point clouds are initially performed by using new computed and ground truth transformation matrices. These transformed VSFM point clouds are then compared with the ground truth transformed point clouds. As in [43], we compare the Frobenius Norm (F-norm) between the newly computed matrices and the ground truth transformation matrices. To obtain better visual representations of comparison results, we use $log(RMSE)$ as the final performance value. The smaller the value, the better performance of the algorithm. We also compute the mean of the F-norm of all 13 datasets for each method and the results are shown in Figure 19.

Figure 19 shows the quantitative evaluation results. It illustrates the SparseICP obtains worst results, and ICP follows. It is because the point-point level strategy shows poor ability in cross-source problems. Also, SparseICP is designed to solve outliers while show overkilled to deal with noisy point clouds. The JR-MPC and CPD show more robust and higher accuracy than other comparison methods; to some extent, they demonstrate the advantage of using the statistical property. The proposed CSGM method obtains the highest accuracy on all dataset. This is because we use the macro structure to globally register two point clouds with paying little attention to the detail, and use the micro structure to accurately register the two point clouds. We also use RANSAC and ICP to further improve the accuracy and robustness.

Figure 18 shows several sample visual results of these methods. The results show that the proposed CSGM clearly achieves better results than the other methods. Go-ICP and

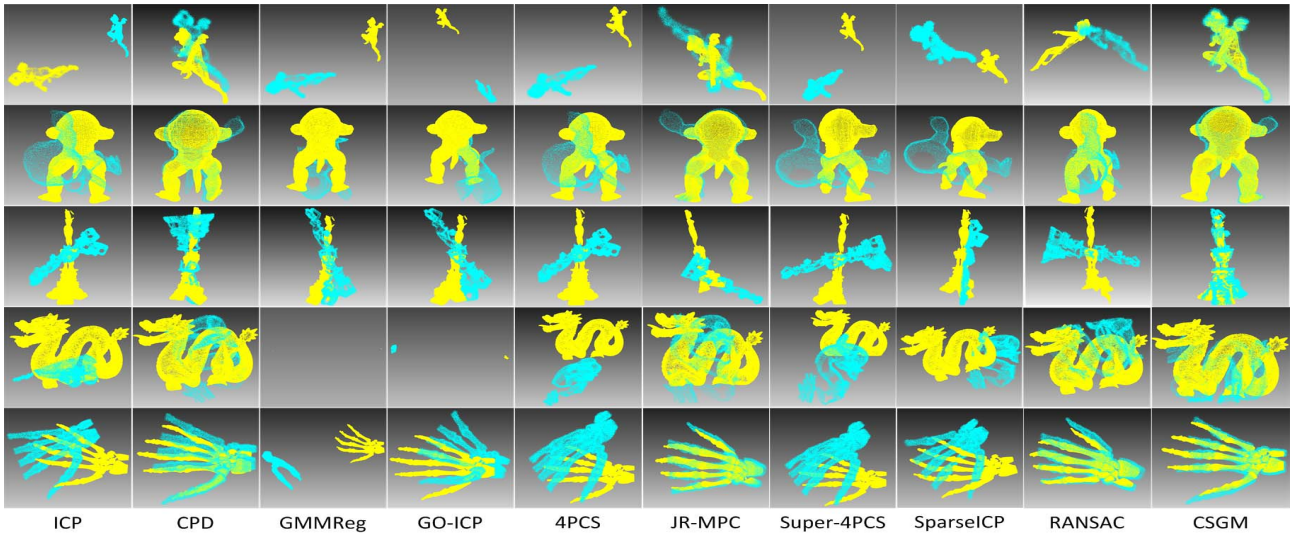


Fig. 22. Visual effect of registration results on Database C. Rows are different datasets and columns are methods.

TABLE III
THE MEAN RUNTIME COMPARISON ON DATASET C (THE UNIT IS SECOND (S))

Methods	ICP	GO-ICP	GMMReg	JR-MPC	CPD	4PCS	Super4PCS	SparseICP	Ransac	CSGM
Time (s)	1491	363	2480	324	650	3540	1641	387	652	1758

JR-MPC obtain similar results to the proposed CSGM in the fourth row dataset. Because of the BnB strategy in Go-ICP and the generative strategy in JR-MPC, good results are obtained if the scale is normalized very well and no large proportion of data are missing. If these conditions do not exist, these methods will completely fail. In the first two rows of Figure 18, for example, these methods show the results of such failure. However, the proposed CSGM achieves robust and accurate registration results in all cross-source datasets.

The proposed method is also compared on Database C which consists of synthetic cross-source point clouds. Transformation relation is estimated by the compared methods and the proposed method from view 2 to view 1 point cloud. The computed and ground truth transformation matrix are then utilized to transform the synthetic point cloud. The RMSE error is computed according to the statistical distance of these two transformed point clouds. Also, we compare the F-norm of the error of difference between transformation matrices.

Figure 20 shows the evaluation results of mean RMSE and Figure 21 shows the evaluation results of mean F-norm of the computed transformation matrix and the ground-truth transformation matrix on whole ten sets of Database C. The results show that our method achieves accurate registration results which are better than the other methods. Figure 22 illustrates the visual effects of the Synthetic evaluation. The results show that the proposed CSGM obtains robust and visually satisfactory registration results which are clearly better than those of the compared methods. Some of the comparison methods are even failed because the cross-source problem are really challenging to these methods.

We also compare the mean runtime on Dataset C and the results show in Table III. Our method is much faster than 4PCS and GMMReg, and it obtains similar speed with ICP and Super4PCS. The most time-consuming step of our method is the graph matching optimization. The reason that GO-ICP, JR-MPC, CPD, SparseICP and RANSAC run much faster is that they are all running with a downsample strategy. They cannot be directly used to deal with large scale cross-source point cloud registration. We ascribe the limited complexity of the proposed method to the fact that the coarse-to-fine computation of macro/micro structures mitigates excessive computational burdens.

VI. CONCLUSION

In this paper, we proposed a new registration pipeline to deal with the cross-source point cloud registration problem using three novelty components. Firstly, a registration pipeline was proposed which combines the initial correspondence from graph matching and refinement using RANSAC and ICP. Secondly, the macro and micro structures of cross-source point clouds are organized as graphs and described by a robust feature. Lastly, an improved graph matching method was proposed to solve the structure-based registration problem, which considers the global geometric constraints. The proposed method not only obtains accurate registration results in same-source point clouds, but also obtains state-of-the-art accuracy on cross-source point clouds registration.

ACKNOWLEDGMENT

The authors would like to thank the Nokia Corporation for their help and acknowledge the useful discussions with colleagues in GBDTC.

REFERENCES

[1] R. A. Newcombe *et al.*, "Kinectfusion: Real-time dense surface mapping and tracking," in *Proc. 10th IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2011, pp. 127–136.

[2] N. Mellado, M. Dellepiane, and R. Scopigno, "Relative scale estimation and 3D registration of multi-modal geometry using growing least squares," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 9, pp. 2160–2173, Sep. 2016.

[3] X. Huang, L. Fan, J. Zhang, Q. Wu, and C. Yuan, "Real time complete dense depth reconstruction for a monocular camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Apr. 2016, pp. 32–37.

[4] P. Musialski, P. Wonka, D. G. Aliaga, M. Wimmer, L. Gool, and W. Purgathofer, "A survey of urban reconstruction," in *Comput. Graph. Forum*, vol. 32, no. 6, pp. 146–177, 2013.

[5] A. M. Manferdini, "A methodology for the promotion of cultural heritage sites through the use of low-cost technologies and procedures," in *Proc. 17th Int. Conf. 3D Web Technol.*, Apr. 2012, p. 180.

[6] F. Peng *et al.*, "Street view cross-sourced point cloud matching and registration," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Jun. 2014, pp. 2026–2030.

[7] P. J. Besl and D. N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[8] N. Mellado, D. Aiger, and N. J. Mitra, "Super 4pcs fast global pointcloud registration via smart indexing," *Comput. Graph. Forum*, vol. 33, no. 5, pp. 205–215, 2014.

[9] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2000, pp. 44–51.

[10] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Towards Internet-scale multi-view stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 1434–1441.

[11] D. F. Huber and M. Hebert, "Fully automatic registration of multiple 3D data sets," *Image Vis. Comput.*, vol. 21, no. 7, pp. 637–650, 2003.

[12] A. Torsello, E. Rodola, and A. Albarelli, "Multiview registration via graph diffusion of dual quaternions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Apr. 2011, pp. 2441–2448.

[13] J. Yang, H. Li, and Y. Jia, "Go-ICP: Solving 3D registration efficiently and globally optimally," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2013, pp. 1457–1464.

[14] D. Aiger, N. J. Mitra, and D. Cohen-Or, "4-points congruent sets for robust pairwise surface registration," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–85, 2008.

[15] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.

[16] B. Jian and B. C. Vemuri, "Robust point set registration using Gaussian mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1633–1645, Aug. 2011.

[17] F. Wang, B. C. Vemuri, A. Rangarajan, and S. J. Eisenschenk, "Simultaneous nonrigid registration of multiple point sets and atlas construction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 2011–2022, Nov. 2008.

[18] Y. Deng, A. Rangarajan, S. Eisenschenk, and B. C. Vemuri, "A Riemannian framework for matching point clouds represented by the schrodinger distance transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2014, pp. 3756–3761.

[19] M. Torki and A. M. Elgammal, "Putting local features on a manifold," in *Proc. CVPR*, vol. 2. 2010, pp. 1–4.

[20] I. Cleju and D. Saupe, "Stochastic optimization of multiple texture registration using mutual information," in *Proc. Joint Pattern Recognit. Symp.*, 2007, pp. 517–526.

[21] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image Vis. Comput.*, vol. 10, no. 3, pp. 145–155, Apr. 1992.

[22] S. Bouaziz, A. Tagliasacchi, and M. Pauly, "Sparse iterative closest point," *Comput. Graph. Forum*, vol. 32, no. 5, pp. 113–123, 2013,

[23] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proc. IEEE 3rd Int. Conf. 3-D Digit. Imag. Modeling*, Sep. 2001, pp. 145–152.

[24] J. Ho, A. Peter, A. Rangarajan, and M.-H. Yang, "An algebraic approach to affine registration of point sets," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Oct. 2009, pp. 1335–1340.

[25] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2147–2154.

[26] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[27] Y. Diez, J. Martí, and J. Salvi, "Hierarchical normal space sampling to speed up point cloud coarse matching," *Pattern Recognit. Lett.*, vol. 33, no. 16, pp. 2127–2133, 2012.

[28] Z.-Q. Cheng, Y. Chen, R. R. Martin, Y.-K. Lai, and A. Wang, "SuperMatching: Feature matching using supersymmetric geometric constraints," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 11, pp. 1885–1894, Nov. 2013.

[29] C. Papazov and D. Burschka, "Stochastic global optimization for robust point set registration," *Comput. Vis. Image Understand.*, vol. 115, no. 12, pp. 1598–1609, 2011.

[30] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, "Robust global registration," in *Proc. Symp. Geometry Process.*, 2005, vol. 2. no. 3, pp. 1–5.

[31] A. Albarelli, E. Rodola, and A. Torsello, "Loosely distinctive features for robust surface alignment," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 519–532.

[32] E. Rodolä, A. Albarelli, F. Bergamasco, and A. Torsello, "A scale independent selection process for 3D object recognition in cluttered scenes," *Int. J. Comput. Vis.*, vol. 102, nos. 1–3, pp. 129–145, 2013.

[33] M. Corsini, M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello, and R. Scopigno, "Fully automatic registration of image sets on approximate geometry," *Int. J. Comput. Vis.*, vol. 102, nos. 1–3, pp. 91–111, 2013.

[34] T. K. Sinha, D. M. Cash, R. J. Weil, R. L. Galloway, and M. I. Miga, "Cortical surface registration using texture mapped point clouds and mutual information," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2002, pp. 533–540.

[35] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, "Toward mutual information based automatic registration of 3D point clouds," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Jan. 2012, pp. 2698–2704.

[36] A. Moussa and N. Elsheimy, "Automatic registration of approximately leveled point clouds of urban scenes," in *Proc. ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, 2015, pp. 145–150.

[37] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, Sep. 1997.

[38] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Apr. 2009, pp. 3212–3217.

[39] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3D object classification," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, May 2011, pp. 2987–2992.

[40] A. E. Johnson, "Spin-images: A representation for 3-D surface matching," Ph.D. dissertation, Citeseer, 1997.

[41] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 356–369.

[42] D. Campbell and L. Petersson. (2016). "Gogma: Globally-optimal Gaussian mixture alignment." [Online]. Available: https://arxiv.org/abs/1603.00150

[43] G. D. Evangelidis, D. Kounades-Bastian, R. Horaud, and E. Z. Psarakis, "A generative model for the joint registration of multiple point sets," in *Euro. Conf. Computer Vision (ECCV)*. New York, NY, USA: Springer, 2014, pp. 109–122.

[44] X. Huang, J. Zhang, Q. Wu, L. Fan, and C. Yuan, "A coarse-to-fine algorithm for registration in 3D street-view cross-source point clouds," in *Proc. Int. Conf. Digit. Image Comput. Techn. Appl. (DICTA)*, Nov. 2016, pp. 1–6.

[45] B. Lin, T. Tamaki, F. Zhao, B. Raytchev, K. Kaneda, and K. Ichii, "Scale alignment of 3D point clouds with different scales," *Mach. Vis. Appl.*, vol. 25, no. 8, pp. 1989–2002, 2014.

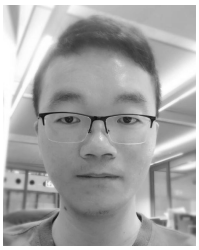
[46] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, "Voxel cloud connectivity segmentation—Supervoxels for point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, Jun. 2013, pp. 22–27.

[47] F. Zhou and F. De la Torre, "Deformable graph matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2013, pp. 2922–2929.

[48] F. Zhou and F. De la Torre, "Factorized graph matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Apr. 2012, pp. 127–134.

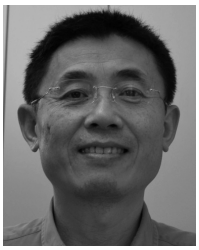
[49] M. Zaslavskiy, F. Bach, and J. P. Vert, "A path following algorithm for the graph matching problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2227–2242, Dec. 2009.

[50] M. Corsini, P. Cignoni, and R. Scopigno, "Efficient and flexible sampling with blue noise properties of triangular meshes," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 6, pp. 914–924, Jun. 2012.



Xiaoshui Huang is currently pursuing the Ph.D. degree with the Global Big Data Technology Center, University of Technology Sydney. His research interests include 3D point cloud registration and matching, 3D reconstruction, computer vision, and machine learning.

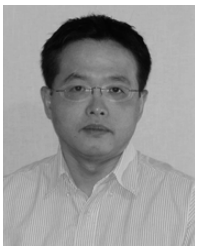
He received the B.Sc. degree in geographic information system from the Taiyuan University of Technology in 2010 and the M.Sc. degree in computer science from Shenyang Aerospace University in 2014. From 2014 to 2015, he was a Research Assistant with the Graduate School at Shenzhen, Tsinghua University.



Jian Zhang (SM'04) received the B.Sc. degree from East China Normal University, Shanghai, China, in 1982, the M.Sc. degree in computer science from Flinders University, Adelaide, Australia, in 1994, and the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, Australia, in 1999.

From 1997 to 2003, he was with the Visual Information Processing Laboratory, Motorola Labs, Sydney, as a Senior Research Engineer, and then became a Principal Research Engineer and a Foundation Manager with the Visual Communications Research Team. From 2004 to 2011, he was a Principal Researcher and a Project Leader with Data61, Sydney, Australia, and a Conjoint Associate Professor with the School of Computer Science and Engineering, UNSW. He is currently an Associate Professor with the Global Big Data Technologies Centre, Faculty of Engineering and Information Technology, School of Computing and Communication, University of Technology Sydney, Sydney. He has authored or co-authored over 120 paper publications, book chapters, and six issued patents filed in the U.S. and China. His current research interests include social multimedia signal processing, large-scale image and video content analytics, retrieval and mining, 3D-based computer vision, and intelligent video surveillance systems.

Dr. Zhang was the General Co-Chair of the International Conference on Multimedia and Expo in 2012 and the Technical Program Co-Chair of the IEEE Visual Communications and Image Processing 2014. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2006 to 2015. He has been an Associate Editor of the *EURASIP Journal on Image and Video Processing* since 2016.



Lixin Fan is currently a Principal Scientist with Nokia Technologies. He has (co-)authored over 50 international journal and conference publications. He (co-)invented dozens of granted and pending patents filed in the U.S., Europe, and China. His research areas of interests include 3D image and video processing, computer vision, machine learning, big data analysis, intelligent human–computer interface, augmented and virtual reality, mobile ubiquitous, pervasive computing, and Nokia/HERE map 3D data processing and rendering.

He received the M.Sc. and Ph.D. degrees in computer science from the National University of Singapore in 1998 and 2002, respectively. He was with the Xerox Research Center Europe, where he was involved in the well-recognized Bag of Keypoints method for image categorization.

Qiang Wu received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively, and the Ph.D. degree from the University of Technology Sydney, Australia, in 2004. He is currently an Associate Professor and a Core Member of the Global Big Data Technologies Centre, University of Technology Sydney. His research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. His research outcomes have been published in many premier international conferences, including ECCV, CVPR, ICIP, and ICPR, and the major international journals, such as the IEEE TIP, the IEEE TSMC-B, the IEEE TCSVT, the IEEE TIFS, PR, PRL, *Signal Processing*, and the IEEE SIGNAL PROCESSING LETTER. He also serves as a Reviewer for several journals and conferences, including TPAMI, PR, TIP, TCSVT, TSMC-B, CVIU, IVC, PRL, *Neurocomputing*, and the *EURASIP Journal on Image and Video Processing*.

Chun Yuan received the M.S. and Ph.D. degrees from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 1999 and 2002, respectively. He was with the INRIA-Rocquencourt, Paris, France, as a Post-Doctoral Research Fellow, from 2003 to 2004. In 2002, he was with the Microsoft Research Asia, Beijing, as an Intern. He is currently an Associate Professor with the Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University. His research interests include computer vision, machine learning, video coding and processing, cryptography, and digital rights management.