FAST REGISTRATION FOR CROSS-SOURCE POINT CLOUDS BY USING WEAK REGIONAL AFFINITY AND PIXEL-WISE REFINEMENT

Xiaoshui Huang\textsuperscript{1}, Lixin Fan\textsuperscript{2}, Qiang Wu\textsuperscript{1}, Jian Zhang\textsuperscript{1}, Chun Yuan\textsuperscript{3}

\textsuperscript{1}. GBDTC, University of Technology Sydney, Australia
\textsuperscript{2}. Nokia Technologies, Finland
\textsuperscript{3}. Graduate school at Shenzhen, Tsinghua University, China

ABSTRACT

Many types of 3D acquisition sensors are emerged in recent years and point cloud has been widely used in many areas. Accurate and fast registration of cross-source 3D point clouds from different sensors is an emerged research problem in computer vision. This problem is extremely challenging because cross-source point clouds contain mixture of various variances, such as density, partial overlap, large noise and outliers, viewpoint changing. In this paper, an algorithm is proposed to align cross-source point clouds with both high accuracy and high efficiency. There are two main contributions: firstly, two components, the weak region affinity and pixel-wise refinement, are proposed to maintain the global and local information of 3D point clouds. Then, these two components are integrated into an iterative tensor-based registration algorithm to solve the cross-source point cloud registration problem. We conduct experiments on synthetic cross-source benchmark dataset and real cross-source datasets. Comparison with six state-of-the-art methods, the proposed method obtains both higher efficiency and accuracy.

1. INTRODUCTION

With the rapid development of 3D data acquisition technologies, point clouds are becoming an effective way to express the surfaces of 3D objects and scenes. Because 3D point clouds have many advantages over 2D images, for example easily extracting spatial geometric, shape and pose information, point clouds become popular to be used in many fields and applications. Some current vision systems even contain more than one type of sensors, such as LiDAR and stereo camera in autonomous driving vision system. The existing applications show improved performance by using fused point clouds from different types of sensors. However, fusion of cross-source point clouds are very challenging because they contain mixture of various differences, such as different density, noise, outliers, viewpoint changing and missing data. The detail of the differences is explained in [1][2].

In this paper, a registration algorithm is proposed to deal with cross-source point cloud fusion problem by considering weak region affinity and pixel-wise refinement.

Definition 1 Same-source point clouds are homogeneous point clouds from same types of sensors. Cross-source point clouds are heterogeneous point clouds from different types of sensors.

Fig. 1: Two cross-source point clouds are captured about a same indoor scene. They contain mixture of variants of density, outliers, noise and missing data. The proposed algorithm can successfully align them with high accuracy and efficiency.

There are several attempts to solve the cross-source point cloud registration. CSGM [1] converts registration problem into graph matching problem. Then, ICP refinement (local) is done based on the graph matching results (global). [3] proposes a coarse-to-fine (global-to-local) strategy to solve the cross-source point cloud. [4] utilizes RANSAC and ICP to global align and local refine the cross-source point cloud registration. However, all the existing methods either cannot solve the cross-source point cloud registration accurately or cannot solve the problem efficiently. All the existing registration methods assume the strong structure consistency between two point-cloud sets being registered. However, it is not the case of cross source point cloud registration in which structure consistency is weak. The reliable correspondences which can be located between two sets are sparse. Larger inconsistency caused by different point cloud density, noise model and various outliers significantly degrade the estimation on rigid transformation.

According to the observation (One example is Figure 1), any two given cross-source point clouds still hold the intrinsic structure similarity although it is weak and degraded by
the various noise, inconsistent point cloud density and outliers. The proposed method is able to discover the salient but weak geometric structure affinity (not statistic feature affinity) by joining sufficient number of local matching i.e. weak regional affinity. In order to adjust the mismatching in weak regional affinity search process, a pixel-wise refinement process is proposed. It is similar to global and local strategy in previous existing methods. Different to previous separated streams, in the proposed method, these two processes i.e. co-affinity search and pixel-wise refinement is unified together. Thus, it can be sorted out in an uniform optimization process. Compared to separated streams, the advantages of uniform optimization process is that interaction between local and global components will be given to the optimization process by jointly considering both local and global information in one unified process.  

To mathematically formalize the above motivation, we need a tool that is a generalization of vectors and matrices. We select tensor as the mathematic tool to assemble this idea, because tensor provides an elegant and unified mathematic format to assemble the global weak regional affinity and local pixel-wise refinement. The weak regional affinity is triplet constraint that is stored in third-order tensors. The pixel-wise refinement is point-point or point-plane residual error that is stored in first-order tensor. Then, the correspondent finding problem can be optimized in a unified whole potential tensor space. Compared to previous separated global and local strategy, these components are interacted in the whole optimization process thanks to the tensor optimization. To solve the final registration problem, instead of doing tensor optimization once, we propose an iterative tensor optimization solution and a new energy function is formulated to obtain optimal geometric transformation. During the iterative optimization solution, in order to interact between transformation matrix estimation and correspondence estimation, the geometric transformation $T$ is integrated into tensor optimization and the tensor space will be updated when the $T$ is updated. When the energy function obtains convergence, both the correspondence and the geometric transformation are optimized.  

The contributions of this paper are three aspects:

- (1) two components, weak regional affinity and pixel-wise refinement, are proposed to keep global and local information in cross-source point clouds where structures are usually weak. We assemble them into tensors so that interaction between global and local information becomes possible.

- (2) An registration algorithm is proposed to integrate these two components to solve the cross-source point cloud registration. We iteratively updated tensor space so that interaction between correspondence estimation and transformation estimation is considered.

- (3) Comprehensive comparison experiments are conducted on cross-source point cloud registration problem.

2. PROPOSED ALGORITHM

Inspired by [5], we recognize a point-point match as a correct match only if the selected triplet points are matched. We regard these triplet points matched constraint as weak region affinity constraint because the triplet point selection is an elegant way to remain the salient structure in the rigid+scale geometric transformation problem. We assemble weak region affinity into three-order tensor. Also, the point-point match is regarded as pixel-wise refinement and can be assembled as first-order tensor. In our method, the three-order tensor and the first-order tensor are integrated into a tensor optimization framework. Then, an efficient power iteration solution is proposed to solve the tensor optimization. In order to be more robust to geometric deformations, we simultaneously solve the optimal correspondence $X$ as well as the optimal transformation $T$.  

Because the proposed method considers geometric constraint and tensor optimization, we define it geometric constraint tensor-based registration (GCTR).

2.1. Definition of pixel-wise refinement and weak region affinity

In the following, we will introduce how to formulate the above pixel-wise refinement and weak regional affinity into tensors. First-order tensor and third-order tensor are utilized to store pixel-wise refinement and weak regional affinity separately. In the following, we suppose Point cloud $C_1$ has $N_1$ points and point cloud $C_2$ has $N_2$ points. 

![Fig. 2](image)

Fig. 2: First-order tensor $H_{\alpha}$ to represent point-to-point similarity in two point clouds with $N_1$ and $N_2$ points. The first-order tensor is constructed by concatenating the columns of $H$. To index a value of the first-order tensor $H_{\alpha}$, for example the red box $H_{ii'}$, it is $H_{(i+i') \times N_1}$.

- **Pixel-wise Refinement**: Pixel-wise refinement is the potential point-to-point correspondence. In our algorithm, first-order tensor is used to store pixel-wise refinement represents correspondent similarity. For correspondent similarity, it is computed as the Euclidean distance between pixel-wise point pair:

$$
H_{ii'} = \exp(-\|f_i - f_{i'}\|^2)
$$

(1)

where $H$ is a $N_1 \times N_2$ correspondent point similarity matrix, $i \in [0, N_1]$ and $i' \in [0, N_2]$. Each element of $H$ stores the similarity of point $i$ of point cloud $C_1$ and point $i'$ of point cloud $C_2$. 
For the first-order tensor $\tilde{H}_a$, it is a $N_1N_2$ vector by concatenating the rows of a similarity matrix $H$. The index conversion between $H_a$ to $H_{i'i'}$ is $i = i + i' \times N_1$ (See Figure 2). $f_i$ is the feature vector of point $P_i$ in point cloud $C_1$ and $f_{i'}$ is the feature vector of point $P_{i'}$ in point cloud $C_2$. For feature vector, we use 3D point coordinate. The first-order tensor stores local information.

- **Weak Regional Affinity**: Weak regional affinity is the potential triplet-to-triplet correspondence (triplet points are a simple region). In our algorithm, third-order tensor is used to store weak regional affinity. In particular, triplet points are selected and are used to represent weak salient structure of cross-source point cloud (triplet points selection is detailed in Section 3). To estimate the correspondence between triplet points, we need to compute the similarity of these triplets. The similarity is computed by

$$H_{i'i'jj'kk'} = \exp(-\|f_{i'j} - f_{i'j'k'}\|^2)$$

where $H_{i'i'jj'kk'}$ is a 6D supersymmetric tensor such as invariant under permutations of indices in $(i, j, k)$ or $i', j', k'$. Each element of the 6D tensor stores the similarity of the two triplets. Points $(P_i, P_j, P_k)$ and $(P_{i'}, P_{j'}, P_{k'})$ are two triplet points with the correspondence of $P_i$ to point $P_j$ (correspondence matrix) and $P_{i'}$ to point $P_{j'}$. $H_{i,j,k}$ is a three-order tensor of size $(N_1N_2)^3$ which is rewritten from tensor $H_{i,i'jj'kk'}$. Because triplet points are selected as large triangles, the third-order tensor stores global information.

- **Triplet similarity**: For each triplet, we compute cosine value of three inner angles of the correspondence triangle combined by the triplet. Then, a descriptor $(f_{i,j,k}, s \in \{1, 2\})$ is formed to describe the triplet. The similarity between triplets are computed by using their descriptors $f_{i,j,k}$. Using this similarity computation strategy, all the elements of the above three-order tensor can be computed. In the three-order tensor, each dimension reflects the potential of point-point correspondence. Therefore, three dimensions are same and the node correspondent will be permutation of all points between two point clouds. Therefore, the tensor is a symmetric tensor.

### 2.2. Geometric Constraint Tensor-based registration method (GCTR)

The goal of our method is to find the optimal transformation matrix between cross-source point clouds. One key step is to finding point-point matching by considering triplet constraint. This is equal to finding the best matching solution in the whole tensor space. In this section, we integrate the above two components into an unified framework and solving the registration problem by maximizing the following objective function:

$$S = \sum_{i,j,k} H_{i'i'jj'kk'}(T)X_{i'i'}X_{j'j}X_{k'k} + H_{i'i'}(T)X_{i'i'}$$

where $X$ (correspondence matrix) and $T$ (transformation matrix) are two parameters need to estimate. $H_{i'i'jj'kk'}$ is a 6D supersymmetric tensor. Each node pair’s (i.e. $P_i$ and $P_i'$) similarity contributes a 2D dimension matrix to $H_{i'i'jj'kk'}$. i.e. $P_i$ and $P_i'$ can form a 2D matrix, and $P_j$ and $P_j'$ can form another 2D matrix. $H_{i'i'}$ is a 2D matrix to describe the pixel-wise similarity. $X$ is the $N_1 \times N_2$ assignment matrix where 1 means two points are matched and 0 otherwise. $\tilde{X} = vec(X)$ obtains $N_1N_2$ vector form of $X$ by concatenating the columns of $X$. $H_{i,i'} = vec(H_{i'i'jj'kk'})$ is a three-order tensor of size $(N_1N_2)^3$ where each element represents the similarity of two triplets. It is a rewritten tensor $H_{i'i'jj'kk'}$. $\hat{H}_i = vec(H_{i'i'})$ is vector form of $H_{i'i'}$ by concatenating the columns of $H_{i'i'}$, where each element represents the point-point similarity. With a geometric transformation $T(.)$ given, $H_{i'i'jj'kk'}(T)$ and $H_{i'i'}(T)$ are the two specific tensors and correspondence matrix $X$ can be estimated by tensor optimization. With an optimized $X$, transformation matrix $T$ can be estimated by singular-value decomposition (SVD).

For the scale computation, we compare triplet correspondent edges and compute the mean ratio as the final scale:

$$s = \frac{\sum_{i=1}^{n-1} (r_i^a/r_i^b)}{n - 1}$$

where $r_i^a$ is the length of point $A_i$ and point $A_{i+1}$ in point set $A$, $B_i$ is the correspondence of $A_i$, and $B_{i+1}$ is the correspondence of $A_{i+1}, r_i^b$ is the length of $B_i$ and point $B_{i+1}$. $n$ is
the number of correspondent pairs, $a, b$ identifies the length in point set A and point set B.

2.3. Power iteration solution

The power iteration solution aims to optimizing the objective function \( S \) by two iterative processes: the correspondence \( X \) is optimized by firming the geometric transformation \( T \) and the geometric transformation \( T \) is optimized by firming the new correspondence \( X \). Algorithm 1 shows the whole process.

2.3.1. Optimization for the correspondence

With specific geometric transformation \( T \) given, the optimization of formulation 3 is a tensor optimization problem. According to [6], two terms in objective function 3 are rank-1 tensors, so that the optimization can be formulated as \textit{Rank} – 1 tensor approximation (R1TA) problem. Inspired by [6] [7], we use tensor power iteration to solve the above R1TA problems. Line 6-10 in Algorithm 1 shows the correspondence optimization procedures.

2.3.2. Optimization for the geometric transformation

With the correspondences \( X \) are given, the estimation of the geometric transformation \( T \) is similar to ICP which can be solved in close-form solution. Suppose \( A \) and \( B \) is \( n \) matched pairs from correspondence matrix \( X \), \( A \) is the points from point set \( C_1 \) and \( B \) is the points from point set \( C_2 \). If \( U_A \) is the mean point of \( A \) and \( U_B \) is the mean point of \( B \), \( UDV = \text{svd}(AB^T) \), the geometrical transformation \( T \) is computed by \( R = UDV^T, t = U_A - s \ast R \ast U_B \). In this step, the key elements are scale estimation and tensor update. For scale estimation computation, we use formulation 4 defined before. For tensor update, we use formulation 4 defined before. For tensor update, we use formulation 4 defined before.

3. IMPLEMENTATION DETAILS

Initially, we introduce line 1 of Algorithm 1. Following [1], we use supervoxel segmentation method [8] to segment the point clouds and use the central points of these segments as the salient structures.

\textbf{Triplet point selections:} Inspired by [9], we randomly select triangles satisfying wide baseline strategy and use the three nodes of these triangles as triplet points.

Then, we introduce line 3-4 of Algorithm 1. Inspired by [9], we select triplets satisfying wide baseline strategy, so that they are more likely to be global aligned. In this algorithm, we use wide baseline strategy to randomly select \( N_1 N_2 \) large triangles in point cloud 1. We define the large triangles as three edges of the triangle are larger than 50% of the overlapping 3D containing voxel’s the diameter. That guarantees the

\begin{algorithm}
\caption{Power iteration algorithm for cross-source point cloud registration.}
\begin{algorithmic}
\State \textbf{Input:} Cross-source point clouds.
\State \textbf{Output:} Transformation matrices \( R, t, s \)
\State 1: Salient structures extraction
\State \textbf{repeat}
\State 2: Wide baseline triplets selection in \( C_1, C_2 \).
\State 3: Build different-order tensors.
\State \textbf{repeat}
\State 4: \[ \beta(m) = \sum_{i,j,k} H_3 \prod_{i,j,k} X_r^{(m)} \]
\State 5: \[ \gamma(m) = H_1 X(m) \]
\State 6: \[ S^{(m+1)} = \beta(m) + \gamma(m) \]
\State 7: \[ X^{(m+1)} = H_3 \otimes X^{(m)} \otimes X^{(m)} + H_1 \]
\State \textbf{until} convergence;
\State 8: \[ A_1, B_1 \leftarrow \|X^{(m+1)}\| \]
\State 9: \[ A, B \leftarrow \text{RANSAC}(A_1, B_1) \]
\State 10: \[ R = UDV^T, t = U_A - s \ast R \ast U_B \]
\State 11: \[ s = \frac{\sum_{i=1}^{n-1} (r_{ai}/r_{bi})}{n-1} \]
\State 12: Update point cloud \( C_2 = (sRC_2 + t) \)
\State \textbf{until} convergence;
\end{algorithmic}
\end{algorithm}

selected triangles are large triangles and make the final registration more prone to globally registered. For the overlapping ratio, if there is unknown, we automatically search as ratio is 0.25, 0.5, 0.75, 1.0.

4. EXPERIMENTAL RESULTS

In this section, we conduct thoroughly comparison experiments on synthetic and real datasets. Firstly, we compare on a synthetic cross-source benchmark datasets. Several state-of-the-art registration methods have been run on it and compared with the proposed method. Secondly, we compare the performance on real cross-source point clouds.

4.1. Experiments setup

The proposed algorithm is implemented by using standard C and Matlab. All the comparison experiments are executed on an i5 CPU, 8GB memory computer. We select ICP [10], GO-ICP [11], Super-4PCS [9], CPD [12], JR-MPC[13] and CSGM[14] as the comparison methods. Most of the existed state-of-the-art registration methods are focus on same-source data and designed to solve SE(3) transformation. They have not designed for scale variation. To compare fairly, we conduct an automatically scale normalization for all the other methods by following [13], which is assuming the size of the
point cloud’s 3D containing voxel is the same. Because JR-MPC becomes not practical when the point number increases significantly, we uniformly down-sample the point cloud to approximately 2000 points for JR-MPC.

For the evaluation, we compute the Frobenius Norm of the transformation matrix difference (TM) between ground-truth and estimation. The lower the value is, the higher accuracy the method achieves.

For the evaluation, we compute the Frobenius Norm of the transformation matrix difference (TM) between ground-truth and estimation. The lower the value is, the higher accuracy the method achieves.

**Fig. 4:** Overall performance (rotation, transformation and scale) on synthetic cross-source benchmark dataset.

### 4.2. Synthetic cross-source benchmark dataset

This section will demonstrate the ability of the proposed method on the cross-source benchmark dataset. For the construction of benchmark dataset, following [1], we start from Stanford 3D Scanning Models and simulate ten sets of cross-source benchmark dataset. Each set of cross-source dataset contains source A and source B which simulate cross-source problems (discussed in Section 1).

| Method   | R    | T    | S    | log(TM) | time(s) |
|----------|------|------|------|---------|---------|
| CPD      | 2.82 | 64.4 | 0.24 | 4.21    | 435     |
| Go-ICP   | 2.7  | 48.8 | 0.06 | 3.94    | 73.8    |
| ICP      | 3.38 | 50.1 | 0.06 | 3.98    | 103     |
| JR-MPC   | 1.84 | 63.3 | 0.06 | 4.18    | 255     |
| Super4PCS| 0.22 | 3.41 | 0.06 | 1.3     | 333     |
| CSGM     | 0.064 | 1.85 | 0.06 | 0.68    | 3024    |
| GCTR     | 0.069 | 1.89 | 0.03 | 0.69    | 61.6    |

**Table 1:** Comparison on Angle benchmark dataset.

We select Angle as examples for detailed quantitative evaluation on the cross-source benchmark datasets. We evaluate the Translation(T), Rotation(R), Scale (S) error separately, and compare the error of RMSE (log(TM)). Also, we compare the runtime on this dataset. Table 1 shows the evaluation results. It shows the GCTR obtains comparable accuracy to CSGM while shows much faster than CSGM. Compared to other methods, the proposed method obtains higher performance on accuracy and efficiency.

**Fig. 5:** Visual registration results of real cross-source point clouds.

### 4.3. Real cross-source point clouds

We capture the cross-source point cloud dataset by using KinectFusion and VSFM. KinectFusion outputs point cloud directly while VSFM reconstructs 3D point cloud from the images taking by mobile phone camera. We captured more than 30 datasets for different indoor scenes and the proposed methods obtain promising results in all the datasets.

**Fig. 5** shows the selected datasets and compares with other methods. The results illustrate our method obtain robust and visually correct registration results in real application. For all these datasets, the proposed methods align at a fast speed in approximately 120 seconds which is much faster than other methods. Although sometimes JR-MPC and
| Method       | CPD | Go-ICP | ICP  | JR-MPC | Super4PCS | CSGM  | GCTR |
|-------------|-----|--------|------|--------|-----------|-------|------|
| Runtime(s)  | 650 | 363    | 1491 | 324    | 1941      | 4648  | 287  |

Table 2: Average running time on the 10 pairs of cross-source benchmark datasets.

Go-ICP obtains registration results similar to our methods, their computation and memory complexity are very high. Our method can align the real cross-source point cloud accurately at a fast speed.

5. CONCLUSION

In this paper, we propose a fast registration method for cross-source point clouds registration. We have done two main works: firstly, weak regional affinity and pixel-wise refinement are proposed to keep global and local information in cross-source point clouds where structures are usually weak; secondly, an unified algorithm is proposed to integrate these two components to solve the cross-source point cloud registration. The experimental results show that the proposed method aligns the challenging cross-source point cloud fast and accurately.

6. REFERENCES

[1] X. Huang, J. Zhang, L. Fan, Q. Wu, and C. Yuan, “A systematic approach for cross-source point cloud registration by preserving macro and micro structures,” IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3261–3276, July 2017.

[2] X. Huang, J. Zhang, Q. Wu, L. Fan, and C. Yuan, “A coarse-to-fine algorithm for matching and registration in 3d cross-source point clouds,” in IEEE Transactions on Circuits and Systems for Video Technology. IEEE, 2018, vol. 28, pp. 2965–2977.

[3] Furon Peng, Qiang Wu, Lixin Fan, Jian Zhang, Yu You, Jianfeng Lu, and Jing-Yu Yang, “Street view cross-sourced point cloud matching and registration,” in Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 2014, pp. 2026–2030.

[4] N. Mellado, M. Dellepiane, and R. Scopigno, “Relative scale estimation and 3d registration of multi-modal geometry using growing least squares,” IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 9, pp. 2160–2173, Sept 2016.

[5] Haili Chui and Anand Rangarajan, “A new point matching algorithm for non-rigid registration,” Computer Vision and Image Understanding, vol. 89, no. 2, pp. 114–141, 2003.

[6] Xinchu Shi, Haibin Ling, Junling Xing, and Weiming Hu, “Multi-target tracking by rank-1 tensor approximation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2387–2394.

[7] Olivier Duchenne, Francis Bach, In-So Kweon, and Jean Ponce, “A tensor-based algorithm for high-order graph matching,” IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 12, pp. 2383–2395, 2011.

[8] Jeremie Papon, Alexey Abramov, Markus Schoeler, and Florentin Worgotter, “Voxel cloud connectivity segmentation-supervoxels for point clouds,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2027–2034.

[9] Nicolas Mellado, Dror Aiger, and Niloy J Mitra, “Super 4pcs fast global pointcloud registration via smart indexing,” in Computer Graphics Forum. Wiley Online Library, 2014, vol. 33, pp. 205–215.

[10] Paul J Best and Neil D McKay, “A method for registration of 3-d shapes,” IEEE Transactions on pattern analysis and machine intelligence, vol. 14, no. 2, pp. 239–256, 1992.

[11] J. Yang, H. Li, D. Campbell, and Y. Jia, “Go-icp: A globally optimal solution to 3d icp point-set registration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PP, no. 99, pp. 1–1, 2016.

[12] Andriy Myronenko and Xubo Song, “Point set registration: Coherent point drift,” IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 12, pp. 2262–2275, 2010.

[13] Georgios Dimitrios Evangelidis and Radu Horua, “Joint alignment of multiple point sets with batch and incremental expectation-maximization,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 6, pp. 1397–1410, 2018.

[14] Xiaoshui Huang, Jian Zhang, Qiang Wu, Lixin Fan, and Chun Yuan, “A coarse-to-fine algorithm for registration in 3d street-view cross-source point clouds,” in 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, 2016, pp. 1–6.