K-means clustering for efficient and robust registration of multi-view point sets

Zutao Jiang\textsuperscript{a}, Jihua Zhu\textsuperscript{a,∗}, Georgios D. Evangelidis\textsuperscript{b}, Changqing Zhang\textsuperscript{c}, Shanmin Pang\textsuperscript{a}, Yaochen Li\textsuperscript{a}

\textsuperscript{a}School of Software Engineering, Xi’an Jiaotong University, P. R. China
\textsuperscript{b}DAQRI, Vienna, Austria
\textsuperscript{c}School of Computer Science and Technology, Tianjin University, P. R. China

Abstract

Generally, there are three main factors that determine the practical usability of registration, i.e., accuracy, robustness, and efficiency. In real-time applications, efficiency and robustness are more important. To promote these two abilities, we cast the multi-view registration into a clustering task. All the centroids are uniformly sampled from the initially aligned point sets involved in the multi-view registration, which makes it rather efficient and effective for the clustering. Then, each point is assigned to a single cluster and each cluster centroid is updated accordingly. Subsequently, the shape comprised by all cluster centroids is used to sequentially estimate the rigid transformation for each point set. For accuracy and stability, clustering and transformation estimation are alternately and iteratively applied to all point sets. We tested our proposed approach on several benchmark datasets and compared it with state-of-the-art approaches. Experimental results validate its efficiency and robustness for the registration of multi-view point sets.

∗Corresponding author: Jihua Zhu
Email address: zhujh@xjtu.edu.cn (Jihua Zhu)
1. Introduction

As a fundamental issue in many areas, the problem of point set registration has attracted immense attention in computer vision [1, 2, 3], computer graphics [4, 5], and robotics [6, 7]. The aim of point set registration is to estimate the optimal transformation between different point sets. According to the involved number of point sets, this problem can be roughly divided into the pair-wise registration and the multi-view registration, with the former being extensively addressed.

Commonly, the problem of pair-wise registration is solved by the iterative closest point (ICP) algorithm [8] or its variants [9, 10, 11]. To achieve registration, most of these approaches alternately build hard correspondences and estimate the transformation. Such approaches are efficient but may be not accurate enough. Furthermore, some registration approaches [12, 13, 14, 15] replace the hard assignment with the soft assignment so as to obtain more accurate results. Since the soft assignment should be built from each point to all points in the opposite point set, these approaches are time-consuming. Besides, most of ICP variants are locally convergent. To obtain the desired global minimum, particle filter [16, 17] and genetic algorithm [18, 19] can be combined to estimate the optimal transformation. 3D features can be also extracted from the point sets and they can be matched to provide initial transformation for the pair-wise registration [20, 21, 22].
The development of scanning equipment makes the 3D reconstruction of an object possible. Due to the occlusion, however, the object cannot be entirely scanned from a single viewpoint. Therefore, scanners should acquire point sets from different viewpoints so as to cover the entire object surface. These point sets can then be transformed into one common reference frame for the 3D model reconstruction[23].

Unlike pair-wise registration, multi-view registration is more difficult and has comparatively attracted less attention. Usually, the multi-view registration can be divided into two stages: the coarse registration and fine registration. For the coarse registration, some approaches apply the pair-wise registration method on all or some of point set pairs, then search the minimum set of pair-wise registration results to estimate the initial rigid transformations for the multi-view registration [24, 25, 26]. Since there are a lot of point set pairs involved in the multi-view registration, the coarse registration approaches are always time-consuming. To accelerate this stage, 3D features can be extracted and matched to provide good initial parameters for the pair-wise registration [27, 28, 29].

The fine registration has attracted considerably more attention. The common strategy is to sequentially align and merge two point sets until all point sets are merged into one model [30]. Although this approach may be efficient, it generally suffers from the error accumulation. To address this issue, Bergevin et al. [31] suggested to establish the correspondence between one point set and any other point set so as to estimate the rigid transformation of this point set. As the establishment of correspondence is time-consuming, the efficiency of this approach is very low. Therefore, Pulli [32] proposed to
align point sets with each other and utilize the pairwise alignments as constraints to evenly diffuse the pairwise registration errors in the multi-view registration step. Inspired by these approaches, the pair-wise registration algorithm can be sequentially utilized to align one point set to the coarse model constructed by all point sets [33]. The outcome can be used to update the coarse model, which can be further used to align each point set. Due to too many points in the reconstructed model, however, the complexity of these approaches is relatively high. The multi-view registration can be also viewed as graph optimization problem [34, 35, 36], where each node denotes one point set and each edge indicates the pair-wise registration of two connected nodes. Without the correspondence update, these approaches can diffuse the registration error over the graph of adjacent point sets. Despite their efficiency, these approaches only transfer registration errors among graph nodes and are unable to reduce the total registration errors.

Recently, Govindu and Pooja [37, 38] proposed the motion averaging algorithm for the multi-view registration. It can recover all rigid transformations simultaneously for multi-view registration from a set of relative motions, which can be estimated by the pair-wise registration. Given reliable and accurate relative motions [39], the motion averaging algorithm can achieve accurate multi-view registration. In practical applications, the relative motions obtained from the pair-wise registration should be confirmed [40, 41] so as to provide reliable and accurate relative motions to the motion averaging algorithm. Alternatively, the multi-view registration can also be cast into the problem of low-rank and sparse (LRS) matrix decomposition [42]. For some scan pairs with large overlap, the relative motions can be estimated and
concatenated into a large matrix, which has missed data corresponding to the scan pairs with low overlapping percentages. By the LRS decomposition, the matrix can be completed and the multi-view registration results can be recovered. Compared to other approaches, it is more likely to be affected by the sparsity of the uncompleted matrix.

More recently, Evangelidis and Horaud \cite{43} held the assumption that all the points involved in multi-view registration are drawn from a single Gaussian mixture, thus formulating registration as a soft clustering problem. An expectation maximization (EM) algorithm is utilized to estimate both the mixture parameters and the rigid transformations that optimally align the point sets. Although this approach is accurate, lots of parameters need to be estimated and the computational cost is quite high.

In this paper, we propose a new multi-view registration algorithm formulated as a joint clustering and alignment problem. We build on well known clustering paradigm, i.e., the K-means algorithm. We first design the new objective function and then present the proposed multi-view registration algorithm. Given initially posed point sets, this algorithm starts with estimating the initial centroids. Then it performs the clustering operation, which consists in the assignment of each point to a single cluster and the update of all cluster centroids. As the multi-view point sets are not well aligned, the shape comprised by all the updated centroids can be used as a reference point set to sequentially estimate the rigid transformation for each point set by the pair-wise registration. To obtain the desired results, clustering and transformation estimation are alternately and iteratively applied to all point sets. For the 3D model reconstruction, the K-means clustering was initially
introduced by Zhou et al. in [44], where the K-means clustering algorithm was utilized to detect and merge the corresponding points of the overlapping areas. It doesn’t involve the update of rigid transformations and only can be viewed as a post-processing step after multi-view registration step in 3D reconstruction.

The remaining of the paper is organized as follows. Section 2 gives a brief introduction of K-means clustering algorithm. Following that, Section 3 proposes the novel approach for registration of multi-view point sets. In Section 4, the proposed approach is tested on some benchmark data-sets. Section 5 concludes this paper.

2. K-means clustering

As an unsupervised learning method, K-means algorithm is effective and interpretable for the clustering. Given the number of clusters $K$ and the data set $X = \{\vec{x}_j\}_{j=1}^M$, the algorithm starts with initial estimation of the $K$ centroids $\{\vec{\mu}_k^0\}_{k=1}^K$, which can be either randomly selected or generated from the data sets. The clustering result is obtained from the iteration of the following two steps.

Step 1: data assignment

$$c^q(j) = \arg \min_{k \in \{1,2,\ldots,K\}} \left\| \vec{x}_j - \vec{\mu}_k^{q-1} \right\|_2^2$$  \hspace{1cm} (1)

Step 2: centroid update

$$\vec{\mu}_k^q = \frac{\sum_{j=1}^M \{c^q(j) = k\} \vec{x}_j}{\sum_{j=1}^M \{c^q(j) = k\}}$$  \hspace{1cm} (2)

Usually, good clustering results can be obtained by iteratively performing the above steps until some convergence criterion is achieved.
Figure 1: The proposed approach assumes that all points involved in multi-view registration are sampled from $K$ clusters. Given the multi-view point sets, the proposed approach starts with initial estimation for the $K$ cluster centroids. Then, clustering and transformation estimation are alternately and iteratively applied to all point sets until the stop criteria are met. The output contain both the multi-view registration and clustering results.

3. K-means based multi-view registration

Given an accurate 3D scene model, the multi-view registration problem can be divided into multiple subproblems, where each point set is pair-wise registered to the accurate model, respectively. However, when scanning an unknown scene, there is no accurate model beforehand. What is available is just the coarse model reconstructed from the initially posed point sets, which contains most likely unregistered points as well as too many redundant points due to the overlapping areas among different point sets.

To deal with these two issues, we here suppose that all the points, once registered, are drawn from $K$ clusters, whose centroids can make up an accurate model. Therefore, if all the points can be well clustered, the accurate
model can be constructed from the cluster centroids and the multi-view registration can be achieved by the pair-wise registration between the centroids and each point-set. Therefore, the multi-view registration can be viewed as the extended clustering problem, which incorporates a pair-wise registration solver. The flowchart of the proposed approach is displayed in Fig. [1]. Given the initial transformations that roughly align the point-sets, e.g. by any pair-wise registration scheme, the main steps of our approach are summarized as follows:

1. Initialize centroids from initially aligned point sets, e.g. random selection of $K$ points.
2. Assign each point to the nearest cluster and then update the centroids.
3. Align each point-set against the centroids
4. Iterate 2) and 3) until convergence

3.1. Problem formulation

We are given $N$ point sets to register. Let $\mathbf{P}_i = \{\vec{p}_{ij}\}_{j=1}^{M_i} (\vec{p}_{ij} \in \mathbb{R}^{3 \times 1})$ denote the $M_i$ 3D points that belong to the $i$th point set and let $M = \sum_{i=1}^{N} M_i$ denote the total number of points. As mentioned, we assume that all the $M$ points, once registered, are drawn from $K$ clusters. Given initial rigid transformations $\{\mathbf{R}_i^0, \mathbf{t}_i^0\}_{i=1}^{N}$ that roughly align the point-sets (e.g. by pairwise registration), the goal of multi-view registration is to simultaneously group the $M$ points into $K$ clusters and estimate accurate rigid transformations $\{\mathbf{R}_i, \mathbf{t}_i\}_{i=1}^{N}$ between each original point set and the centroids. This is an essential difference from [43] which registers Gaussian means with virtual points sets produced by the original ones.
Given all the point sets \( \{ \mathbf{P}_i \}_{i=1}^N \) and assuming that their points are samples drawn from \( K \) clusters, represented via their centroids \( \{ \bar{\mathbf{\mu}}_k \}_{k=1}^K \), the multi-view registration can be formulated as the following least-square (LS) problem:

\[
\begin{align*}
\text{arg min}_{\mathbf{R}_i, \bar{t}_i, c(ij) \in \{1,2,\ldots,K\}} & \sum_{i=1}^N \sum_{j=1}^{M_i} (w_{ij} \cdot \| \mathbf{R}_i \bar{p}_{ij} + \bar{t}_i - \bar{\mu}_{c(ij)} \|^2) \\
\text{s.t.} & & \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}_3, \det(\mathbf{R}_i) = 1
\end{align*}
\] (3)

where \( \bar{\mu}_{c(ij)} \) is the nearest cluster centroid of the transformed point \( (\mathbf{R}_i \bar{p}_{ij} + \bar{t}_i) \) and \( w_{ij} \) is a binary variable that quantifies the hard assignment. If the \( c(ij) \)th cluster only contains the point that belongs to the \( i \)th point set, we set \( w_{ij} = 0 \), otherwise \( w_{ij} = 1 \). Note that solving problem (3) implies solving both the clustering problem and the registration problem, that is, the cluster centroids \( \{ \bar{\mu}_k \}_{k=1}^K \) are also unknown.

3.2. Proposed multi-view registration algorithm

K-means algorithm is very effective for the clustering. However, to solve problem (3), it should be extended so as to cluster points and estimate the rigid transformations simultaneously.

Regardless of registration, initial centroids \( \{ \bar{p}_k \}_{k=1}^K \) are required for K-means. The initial centroids can be either randomly selected or generated from the data set. Given initial rigid transformations \( \{ \mathbf{R}_i^0, \bar{t}_i^0 \}_{i=1}^N \), the coarse model can be reconstructed by all point sets involved in the multi-view registration as follows:

\[
\mathbf{P} \triangleq \{ \mathbf{R}_i^0 \bar{p}_{ij} + \bar{t}_i^0 \}_{i=1,j=1}^N.
\] (4)
Then, the initial centroids of all clusters \( \{ \vec{\mu}_k \}_{k=1}^K \) can be uniformly sampled from this coarse model. The shape comprised by all centroids is similar to the coarse model. The only difference is the resolution of points.

Being provided with initial rigid transformations \( \{ R_i^0, \vec{t}_i^0 \}_{i=1}^N \), the K-means clustering algorithm can be used to achieve the registration of multi-view point sets by iterations. In each iteration, the following three steps are included:

Step 1: Assign each point to a single cluster:

\[
   c^q(ij) = \arg \min_{k \in \{1, 2, \ldots, K\}} \| R_i^q \vec{p}_{ij} + \vec{t}_i^q - \vec{\mu}_k^q \|_2^2. \tag{5}
\]

Step 2: Update the centroids for all clusters:

\[
   \vec{\mu}_k^q = \frac{\sum_{ij=1}^M \{ c^q(ij) = k \} (R_i^q \vec{p}_{ij} + \vec{t}_i^q)}{\sum_{ij=1}^M \{ c^q(ij) = k \}}. \tag{6}
\]

Step 3: Sequentially estimate each rigid transformation:

\[
   (R_i^q, \vec{t}_i^q) = \arg \min_{R_i, \vec{t}_i} \sum_{j=1}^{M_i} w_{ij}^q \| R_i \vec{p}_{ij} + \vec{t}_i - \vec{\mu}_{c^q(ij)} \|_2^2. \tag{7}
\]

As with the K-means algorithm, the above steps should be repeated until the iteration number \( q \) exceeds the maximum value \( Q \) or all the estimated rigid transformations have not drastically changed in successive iterations. Finally, it can simultaneously achieve the clustering and obtain the desired results for the registration of multi-view point sets.

In fact, Eq. (5) is the nearest neighbor (NN) search problem, which can be efficiently solved by the search method based on k-d tree. After data assignment, each cluster centroid can be directly calculated from all the assigned points. Besides, Eq. (7) can be solved by the singular value decomposition.
(SVD) method [45]. The main difficulty lies in the confirmation of each binary variable $w_{ij}^q$. In the multi-view registration, the $i$th point set covers some regions that other point sets cannot cover. If these regions are used to update the rigid transformation for itself in Eq. (7), the accuracy will go down. Therefore, these regions are invalid and should be eliminated. As these regions are only covered by one point set, the cluster lying in these regions would contain less points than other regions, which may be covered by more than two point sets. Hence, we can confirm the value of $w_{ij}^q$ according to the cardinality of the $c^q(ij)$th cluster. Here, if the cardinality of the $c^q(ij)$th cluster is less than four fifths of the mean value of all clusters, we set $w_{ij}^q = 0$ so as to eliminate the regions covered only by $i$th point set. Otherwise, we set $w_{ij}^q = 1$.

Based on the above description, our multi-view registration algorithm is summarized in Algorithm 1. As seen, to efficiently register multiple point sets, all points in the coarse reconstructed model should be clustered so as to get the reasonable model comprised by the cluster centroids.

The only parameter that affects the performance of the algorithm is the number of the clusters $K$, as discussed in Section 4.

3.3. Complexity

This section discusses the complexity of the proposed approach for registration of multi-view point sets. Since our approach is proposed for registering multiple point sets, the total number $M$ of points is the central quantity. For the multi-view registration, the maximum value of clustering iteration can be set as $Q$. At each iteration, four steps are executed to estimate each transformation:
Algorithm 1: Multi-view registration algorithm

Input: $\Theta^0 = \{ R_i^0, t_i^0 \}_{i=1}^N$, $K, \{ P_i \}_{i=1}^N$

- Get the initial model by Eq. (4);
- Sample the centroids $\{ \mu^0_k \}_{k=1}^K$ from the initial model;
- $q = 0$;

Repeat

- $q = q + 1$;
- Do data assignment according to Eq. (5);
- Update all cluster centroids according to Eq. (6);
- for $i = 2:N$
  - Calculate $(R_i^q, t_i^q)$ according to Eq. (7);
- end

Until ( $\Theta$’s change is negligible) or ($q > Q$ )

Output: $\Theta = \{ R_i, t_i \}_{i=1}^N$
Table 1: Complexity of the different operations in each iteration of the proposed registration approach

| Operation                  | Complexity     |
|----------------------------|----------------|
| Building k-d tree          | $O(K \log K)$ |
| Data assignment            | $O(M \log K)$ |
| Centroid update            | $O(M)$        |
| Transformation estimation  | $O(M)$        |

(1) Build the k-d tree. Before the estimation of all transformations, the k-d tree should be built once from all $K$ cluster centroids for the point assignment. This step results in a complexity of $O(K \log K)$.

(1) Assign each point to a single cluster. In this step, each point should be assigned to the nearest cluster centroid. To accelerate the assignment, the proposed approach utilizes the nearest neighbor search method based on the k-d tree, which leads to a complexity of $O(M \log K)$.

(2) Update the centroids for all clusters. By sequentially traversing each point, all the cluster centroids can be updated. Therefore, This step introduces a complexity of $O(M)$.

(3) Sequentially estimate each rigid transformation. As shown in Eq. [7], $M_i$ point pairs are used to estimate one rigid transformation, so the estimation of $N$ rigid transformations can introduce a complexity of $O(M)$.

Since $N$ transformations are required to be estimated, the complexity of our approach is shown in Table 1.
Table 2: 3D scans used for evaluation

|                | Bunny | Armadillo | Buddha | Dragon | Angel | Hand |
|----------------|-------|-----------|--------|--------|-------|------|
| Number of views \((N)\) | 10    | 12        | 15     | 15     | 36    | 36   |
| Total number of point \((M)\) | 1362272 | 307625 | 1099005 | 469193 | 2347854 | 1065575 |

4. Experiments

To evaluate the performance of the proposed approach, a set of experiments were conducted on six data-sets from the Stanford 3D Scanning Repository [46]. Table 2 displays some information of these data-sets, including the number of views and total point number of each data-set. These data-sets contain the multi-view point sets and the ground truth of transformations for their registration. To reduce the runtime of multi-view registration, all the raw point sets have been down-sampled by a factor \(S = 8\).

For the comparison, the error of rotation matrix and translation is here defined as  
\[
E_R = \frac{1}{N} \sum_{i=1}^{N} \| R_{i,m} - R_{i,g} \|_F \quad \text{and} \quad E_{\vec{t}} = \frac{1}{N} \sum_{i=1}^{N} \| \vec{t}_{i,m} - \vec{t}_{i,g} \|_2,
\]
where \((R_{i,g}, \vec{t}_{i,g})\) denotes the ground truth of the \(i\)th rigid transformation and \((R_{i,m}, \vec{t}_{i,m})\) indicates the one estimated by the multi-view registration approach.

During experiments, the maximum step for clustering was set as \(Q = 500\). All the baselines adopted the nearest-neighbor search method based on \(k\)-d tree to establish point correspondences and were implemented in MATLAB. To produce unregistered point sets, we randomly perturb the ground trough transformations. Experiments were performed on a four-core 3.6 GHz computer with 8 GB of memory.
Figure 2: Mean error and run time of the proposed approach tested on Armadillo with different values of $K$. (a) Rotation error. (b) Translation error. (c) Runtime.

4.1. Choice of $K$

The number of cluster centers $K$ seems to affect the performance of multi-view registration. The actual number of the clusters is unknown and may vary with the data set to be registered.

We conducted a group of experiments on Stanford Armadillo with different down-sampling factor $S$. To provide the initial registration parameters, a fixed perturbation was added to the ground-truth transformations of multi-view registration. Then, the proposed approach was tested with the number of cluster centers $K$ varying from 500 to 3500. To eliminate the randomness, 20 Monte Carlo (MC) trials were conducted with respect to each value of $K$. Fig. 2 shows comparison results of the proposed approach under different values of $K$.

As shown in Fig. 2, the minimum rotation error is obtained for the range $[1000, 1500]$ of $K$ for all down-sampling factors. The translation error decreases with the increase of the number of cluster centers $K$, while it may increase again after $K = 3500$. The run time obviously increases with $K$. 

15
While a large value of $K$ might be better for accuracy reasons, a small value of $K$ leads to a more efficient registration. By the consideration of these two factors, we set $K = 1500$ in the experiments below for the proposed registration algorithm.

4.2. Validation

As mentioned before, the reconstructed model is comprised by the cluster centroids, which are initially sampled from all the aligned point sets. Since each point set covers some regions that other point sets cannot cover, the proposed approach suggests that these regions should be eliminated to estimate the rigid transformation of this point set itself. To verify the effectiveness of this strategy, we compare the proposed approach with and without elimination of the invalid cluster centroids, which are abbreviated as KmeansReg and KmeansRegWOE (Kmeans Registration Without Elimination), respectively. They were tested on Stanford Bunny and Armadillo under different perturbation levels, where five uniform distributed noises were added to ground-truth transformations, respectively. To eliminate the randomness, 10 MC trials were conducted with respect to five perturbation levels. Fig. 3 illustrates the mean of rotation error, translation error and the mean run time of the two different strategies.

As shown in the Fig. 3, the proposed approach without elimination of the invalid cluster centroids can always obtain the less accurate registration results under varied noise levels for different data-sets. This is because the pair-wise registration of two identical point sets can only obtain the identity matrix. Since the $i$th rigid transformation should be estimated by aligning the $i$th point set to the model comprised by all the cluster centroids, these
Figure 3: Comparison of different strategies for dealing with the invalid cluster centroids
### Table 3: Comparison of accuracy and runtime for different approaches

| Dataset | Initial | CFTrICP [33] | MATrICP [38] | LRS [42] | JRMPC [43] | Ours |
|---------|---------|---------------|---------------|-----------|-------------|------|
| Bunny   | $E_R$   | 0.0360        | **0.0032**    | 0.0086    | 0.0082      | 0.0163| 0.0111|
|         | $E_T(\times 10^{-3})$ | 2.5211        | 0.4181        | **0.4176**| 0.6115      | 1.9902| 1.1439|
|         | T(min)      | –             | 1.0828        | 0.3380    | 0.9327      | 7.7482| 0.2207|
| Armadillo | $E_R$ | 0.0375        | 0.0077        | 0.1088    | 0.0650      | 0.0207| **0.0049**|
|         | $E_T(\times 10^{-3})$ | 4.5276        | **0.7717**    | 7.8606    | 7.4092      | 0.8416| 2.1517|
|         | T(min)      | –             | 0.7941        | 3.8178    | > 10        | 6.9728| **0.1426**|
| Buddha  | $E_R$   | 0.0334        | 0.0333        | **0.0107**| 0.0310      | 0.0213| 0.1320|
|         | $E_T(\times 10^{-3})$ | 2.7316        | 2.1410        | 1.0994    | 1.4452      | **0.7771**| 1.8696|
|         | T(min)      | –             | > 10          | 2.0492    | 4.8167      | > 10  | **0.7671**|
| Dragon  | $E_R$   | 0.0439        | 0.0429        | 0.2770    | 0.2539      | 0.0164| **0.0106**|
|         | $E_T(\times 10^{-3})$ | 5.6432        | 3.8421        | 39.6289   | 36.3570     | **1.2654**| 2.5963|
|         | T(min)      | –             | 7.6386        | 3.2058    | > 10        | > 10  | **0.3042**|
| Angel   | $E_R$   | 0.0382        | 0.0305        | **0.0022**| 0.0072      | 0.0105| 0.0045|
|         | $E_T(\times 10^{-3})$ | 184.7460      | 410.3960      | **51.8375**| 72.2118     | 66.2194| 94.2439|
|         | T(min)      | –             | > 10          | 7.7475    | > 10        | > 10  | **2.2303**|
| Hand    | $E_R$   | 0.0423        | 0.0518        | 0.0712    | 0.0569      | 0.0175| **0.0097**|
|         | $E_T(\times 10^{-3})$ | 227.1540      | 153.5512      | 261.3473  | 206.0000    | 96.6633| **85.8901**|
|         | T(min)      | –             | > 10          | > 10      | > 10        | > 10  | **1.3036**|
cluster centroids located in the regions only covered by the $i$th point set should be eliminated. Otherwise, the registration accuracy drops. Therefore, the proposed approach with elimination of the invalid cluster centroids can obtain more accurate registration results. Besides, the execution time of these two strategies are approximately equal. As a result, it is reasonable to discard the invalid cluster centroids so as to accurately estimate the transformations.

4.3. Comparison

We here compare the proposed algorithm against four state-of-the-art approaches: the coarse-to-fine TrICP approach [33], the motion averaging TrICP approach [38], the approach based on the low-rank and sparse decomposition [42] as well as the joint registration approach [43], which are abbreviated as CFTrICP, MATrICP, LRS, and JRMPC, respectively. Results are reported in run time, the error of rotation matrix and translation.

4.3.1. Accuracy and efficiency

As the locally convergent registration approaches, all these five approaches require the initial registration parameters. As mentioned, initial registration parameters were generated by perturbing the ground truth of rigid transformations with some noise. Subsequently, all the compared approaches start from the same initial parameters for the registration of multi-view point sets. For comparison, Table 3 shows the registration error and runtime of all compared approaches for different data sets, where the bold number denotes the best performance. To evaluate the registration accuracy in a more intuitive way, Fig. 4 displays the multi-view registration results in the form of cross-section.
| Dataset   | Initial | CFTrICP | MATrICP | LRS    | JRMPc  | Ours   |
|-----------|---------|---------|---------|--------|--------|--------|
| Bunny     | ![Bunny](image) | ![Bunny](image) | ![Bunny](image) | ![Bunny](image) | ![Bunny](image) | ![Bunny](image) |
| Armadillo | ![Armadillo](image) | ![Armadillo](image) | ![Armadillo](image) | ![Armadillo](image) | ![Armadillo](image) | ![Armadillo](image) |
| Buddha    | ![Buddha](image) | ![Buddha](image) | ![Buddha](image) | ![Buddha](image) | ![Buddha](image) | ![Buddha](image) |
| Dragon    | ![Dragon](image) | ![Dragon](image) | ![Dragon](image) | ![Dragon](image) | ![Dragon](image) | ![Dragon](image) |
| Angel     | ![Angel](image) | ![Angel](image) | ![Angel](image) | ![Angel](image) | ![Angel](image) | ![Angel](image) |
| Hand      | ![Hand](image) | ![Hand](image) | ![Hand](image) | ![Hand](image) | ![Hand](image) | ![Hand](image) |

Figure 4: Cross-section of multi-view registration results
As shown in Table 3, the proposed approach is the most efficient one among these competed approaches. Although JRMPC is also a clustering based registration approach, it is more time-consuming than the proposed approach owing to the soft assignment strategy and to the larger number of parameters that need estimating. Note that, the proposed approach only requires to establish correspondences between each point set and the model comprised by the cluster centroids. More specifically, it only requires to assign a single cluster centroid to one point in each point set. While, the CFTrICP should establish correspondences between each point set and the model constructed by other aligned point sets. Besides, both MA-TrICP and LRS should establish correspondences between point set pairs, which contains certain percentage of overlaps. Since the number of the cluster centroids are much less than the number of point involved in multi-view registration, the establishment of correspondences, which is the most time-consuming operation in the multi-view registration, is much faster in the proposed approach.

As shown in Table 3 and Fig. 4, the proposed approach is not very sensitive to noises and it can always obtain acceptable accuracy compared to other baselines. Since the proposed approach utilizes the model comprised by all cluster centroids for the estimation of each transformation, the resolution of this model is much lower than the model constructed by raw point sets. Therefore, the proposed approach may not always obtain the most accurate registration results. But other competed approaches may fail to achieve the multi-view registration without good registration parameters. Therefore, the proposed approach is comparable to other state-of-the-art approaches in
Table 4: Robustness comparison under varied noise levels

| Terms | CFTICP [33] | MATICP [35] | LRS [42] | JRMP [43] | Ours |
|-------|-------------|-------------|----------|-----------|------|
| $E_R$ [-0.01,0.01] $E_t(\times 10^{-3})$ | 0.0106 | 0.0745 | 0.0507 | 0.0093 | **0.0061** |
| $T(\text{min})$ | 1.0482 | 1.5948 | 4.6892 | 6.9014 | **0.1612** |
| $E_R$ [-0.02,0.02] $E_t(\times 10^{-3})$ | 0.0428 | 0.1157 | 0.0566 | 0.0156 | **0.0092** |
| $T(\text{min})$ | 1.3056 | 1.6660 | 7.9593 | 6.9335 | **0.1717** |
| $E_R$ [-0.03,0.03] $E_t(\times 10^{-3})$ | 0.0676 | 0.1450 | 0.0627 | 0.0259 | **0.0159** |
| $T(\text{min})$ | 1.7885 | 1.7847 | 8.6718 | 6.9945 | **0.2475** |
| $E_R$ [-0.04,0.04] $E_t(\times 10^{-3})$ | 0.0845 | 0.1523 | 0.0636 | 0.0664 | **0.0438** |
| $T(\text{min})$ | 1.7990 | 1.8480 | 10.1937 | 7.0830 | **0.3005** |

terms of accuracy.

4.3.2. Robustness

To further verify its robustness, the proposed approach was tested on Stanford Armadillo under different transformation noise levels, where values sampled from five uniform distributions were added to ground-truth transformations, respectively. To eliminate the randomness, 10 MC trials were conducted with respect to five noise distributions for all competed approaches. Table 4 shows the average of rotation error, the average of translation error, and the average run time for these approaches. To compare the robustness in a more intuitive way, Fig. 5 displays the mean error and the corresponding standard deviation for these approaches.
As shown in Table 4 and Fig. 5, the proposed approach can obtain the most robust registration results under varied initial registration. With the increase of noise level, the registration errors of the proposed approach increase gradually and its standard deviation is always small. Although the CFTrICP and JRMPC perform well under low noise levels, their registration error raises at higher noise levels. Besides, the standard deviation of the CFTrICP increases sharply with the noise levels. Meanwhile, both MATrICP and LRS approaches have difficulties to obtain accurate registration for the Stanford Armadillo under all noise levels. Therefore, the proposed approach seems to be the most robust one among all the competed approaches. Besides, its run time is far less than other three competed approaches.

5. Conclusion

This paper proposes a novel approach for registration of multi-view point sets. To the best of our knowledge, it is the first time applying the K-means clustering to the multi-view registration problem. The proposed approach
has been tested on the Stanford 3D Scanning Repository and the experimental results demonstrate that it can achieve the multi-view registration of initially posed point sets with good efficiency and robustness. Similar to most of the related approaches, the proposed approach requires initial parameters for the multi-view registration.

Our future work will investigate how to estimate the initial rigid transformations so as to automatically achieve the multi-view registration of unordered point sets without any prior information.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant nos. 61573273, 61573280 and 61503300. We would like to thank Federica Arrigoni for providing the MATLAB implementation of [42].

Reference

References

[1] T. Shiratori, J. Berclaz, M. Harville, C. Shah, T. Li, Y. Matsushita, S. Shiller, Efficient large-scale point cloud registration using loop closures, in: 3D Vision (3DV), 2015 International Conference on, IEEE, 2015, pp. 232–240.

[2] J. Yang, H. Li, D. Campbell, Y. Jia, Go-icp: a globally optimal solution to 3d icp point-set registration, IEEE transactions on pattern analysis and machine intelligence 38 (11) (2016) 2241–2254.
[3] W. Liu, Z. Lan, Y. Zhang, Z. Zhang, H. Zhao, F. Ye, X. Li, Global data registration technology based on dynamic coded points, IEEE Transactions on Instrumentation and Measurement 67 (2) (2018) 394–405.

[4] A. Dai, M. Nießner, M. Zollhöfer, S. Izadi, C. Theobalt, Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration, ACM Transactions on Graphics (TOG) 36 (3) (2017) 24.

[5] D. Aiger, N. J. Mitra, D. Cohen-Or, 4-points congruent sets for robust pairwise surface registration, Acm Transactions on Graphics 27 (3) (2008) 1–10.

[6] F. Yu, J. Xiao, T. Funkhouser, Semantic alignment of lidar data at city scale, in: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, IEEE, 2015, pp. 1722–1731.

[7] L. Ma, J. Zhu, L. Zhu, S. Du, J. Cui, Merging grid maps of different resolutions by scaling registration, Robotica 34 (11) (2016) 2516–2531.

[8] P. J. Besl, N. D. McKay, Method for registration of 3-d shapes, in: Sensor Fusion IV: Control Paradigms and Data Structures, Vol. 1611, International Society for Optics and Photonics, 1992, pp. 586–607.

[9] D. Chetverikov, D. Stepanov, P. Krsek, Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm, Image and Vision Computing 23 (3) (2005) 299–309.

[10] J. M. Phillips, R. Liu, C. Tomasi, Outlier robust icp for minimizing fractional rmsd, in: 3-D Digital Imaging and Modeling, 2007. 3DIM’07. Sixth International Conference on, IEEE, 2007, pp. 427–434.
[11] C. Langis, M. Greenspan, G. Godin, The parallel iterative closest point algorithm, in: 3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on, IEEE, 2001, pp. 195–202.

[12] S. Granger, X. Pennec, Multi-scale em-icp: A fast and robust approach for surface registration, in: European Conference on Computer Vision, Springer, 2002, pp. 418–432.

[13] B. Jian, B. C. Vemuri, Robust point set registration using gaussian mixture models, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8) (2011) 1633–1645.

[14] A. Myronenko, X. Song, Point set registration: Coherent point drift, IEEE transactions on pattern analysis and machine intelligence 32 (12) (2010) 2262–2275.

[15] Y. Tsin, T. Kanade, A correlation-based approach to robust point set registration, in: European conference on computer vision, Springer, 2004, pp. 558–569.

[16] R. Sandhu, S. Dambreville, A. Tannenbaum, Point set registration via particle filtering and stochastic dynamics, IEEE transactions on pattern analysis and machine intelligence 32 (8) (2010) 1459–1473.

[17] I. Kolesov, J. Lee, G. Sharp, P. Vela, A. Tannenbaum, A stochastic approach to diffeomorphic point set registration with landmark constraints, IEEE transactions on pattern analysis and machine intelligence 38 (2) (2016) 238–251.
[18] E. Lomonosov, D. Chetverikov, A. Ekárt, Pre-registration of arbitrarily oriented 3d surfaces using a genetic algorithm, Pattern Recognition Letters 27 (11) (2006) 1201–1208.

[19] J. Zhu, D. Meng, Z. Li, S. Du, Z. Yuan, Robust registration of partially overlapping point sets via genetic algorithm with growth operator, IET Image Processing 8 (10) (2014) 582–590.

[20] R. B. Rusu, N. Blodow, M. Beetz, Fast point feature histograms (fpfh) for 3d registration, in: Robotics and Automation, 2009. ICRA’09. IEEE International Conference on, IEEE, 2009, pp. 3212–3217.

[21] H. Lei, G. Jiang, L. Quan, Fast descriptors and correspondence propagation for robust global point cloud registration, IEEE Transactions on Image Processing 26 (8) (2017) 3614–3623.

[22] Z. Lou, T. Gevers, Image alignment by piecewise planar region matching, IEEE Transactions on Multimedia 16 (7) (2014) 2052–2061.

[23] M. J. Ren, L. J. Sun, M. Y. Liu, C. F. Cheung, Y. H. Yin, A reconstruction–registration integrated data fusion method for measurement of multiscaled complex surfaces, IEEE Transactions on Instrumentation and Measurement 66 (3) (2017) 414–423.

[24] D. F. Huber, M. Hebert, Fully automatic registration of multiple 3d data sets, Image and Vision Computing 21 (7) (2003) 637–650.

[25] A. S. Mian, M. Bennamoun, R. Owens, Three-dimensional model-based object recognition and segmentation in cluttered scenes, IEEE transac-
tions on pattern analysis and machine intelligence 28 (10) (2006) 1584–1601.

[26] J. Zhu, L. Zhu, Z. Li, C. Li, J. Cui, Automatic multi-view registration of unordered range scans without feature extraction, Neurocomputing 171 (2016) 1444–1453.

[27] S. Fantoni, U. Castellani, A. Fusiello, Accurate and automatic alignment of range surfaces, in: 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on, IEEE, 2012, pp. 73–80.

[28] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, M. Lu, An accurate and robust range image registration algorithm for 3d object modeling, ieee transactions on multimedia 16 (5) (2014) 1377–1390.

[29] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, An integrated framework for 3-d modeling, object detection, and pose estimation from point-clouds, IEEE Transactions on Instrumentation and Measurement 64 (3) (2015) 683–693.

[30] Y. Chen, G. Medioni, Object modelling by registration of multiple range images, Image and vision computing 10 (3) (1992) 145–155.

[31] R. Bergevin, M. Soucy, H. Gagnon, D. Laurendeau, Towards a general multi-view registration technique, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (5) (1996) 540–547.
[32] K. Pulli, Multiview registration for large data sets, in: 3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on, IEEE, 1999, pp. 160–168.

[33] J. Zhu, Z. Li, S. Du, L. Ma, T. Zhang, Surface reconstruction via efficient and accurate registration of multiview range scans, Optical Engineering 53 (10) (2014) 102104.

[34] G. C. Sharp, S. W. Lee, D. K. Wehe, Multiview registration of 3d scenes by minimizing error between coordinate frames, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (8) (2004) 1037–1050.

[35] S.-W. Shih, Y.-T. Chuang, T.-Y. Yu, An efficient and accurate method for the relaxation of multiview registration error, IEEE Transactions on Image Processing 17 (6) (2008) 968–981.

[36] A. Torsello, E. Rodola, A. Albarelli, Multiview registration via graph diffusion of dual quaternions, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 2441–2448.

[37] V. M. Govindu, Lie-algebraic averaging for globally consistent motion estimation, in: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 1, IEEE, 2004, pp. I–I.

[38] V. M. Govindu, A. Pooja, On averaging multiview relations for 3d scan registration, IEEE Transactions on Image Processing 23 (3) (2014) 1289–1302.
[39] Z. Li, J. Zhu, K. Lan, C. Li, C. Fang, Improved techniques for multi-
view registration with motion averaging, in: 3D Vision (3DV), 2014 2nd
International Conference on, Vol. 1, IEEE, 2014, pp. 713–719.

[40] V. M. Govindu, Robustness in motion averaging, in: Asian Conference
on Computer Vision, Springer, 2006, pp. 457–466.

[41] D. S. Pankaj, R. R. Nidamanuri, Robust algorithm for multiview regis-
tration, IET Computer Vision 11 (1) (2016) 96–103.

[42] F. Arrigoni, B. Rossi, A. Fusiello, Global registration of 3d point sets
via lrs decomposition, in: European Conference on Computer Vision,
Springer, 2016, pp. 489–504.

[43] G. D. Evangelidis, R. Horaud, Joint alignment of multiple point sets with
batch and incremental expectation-maximization, IEEE transactions on
pattern analysis and machine intelligence.

[44] H. Zhou, Y. Liu, Accurate integration of multi-view range images using
k-means clustering, Pattern Recognition 41 (1) (2008) 152–175.

[45] J. Zhu, D. Wang, X. Bai, H. Lu, C. Jin, Z. Li, Registration of point
clouds based on the ratio of bidirectional distances, in: 3D Vision (3DV),
2016 Fourth International Conference on, IEEE, 2016, pp. 102–107.

[46] M. Levoy, J. Gerth, B. Curless, K. Pull, The stanford 3d scanning repos-
itory, URL http://www-graphics.stanford.edu/data/3dscanrep.