Large-scale DSM registration via motion averaging

Ningli Xu^{1,3}, Rongjun Qin^{1,2,3,4,*}

^{1}Geospatial Data Analytics Lab, The Ohio State University, Columbus, USA
^{2}Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, USA
^{3}Department of Electrical and Computer Engineering, The Ohio State University, Columbus, USA
^{4}Translational Data Analytics Institute, The Ohio State University, Columbus, USA
Email: $\langle xu.3961\rangle\langle qin.324\rangle@osu.edu$

Keywords: Iterative Closest Point, 3D Data Registration, Motion Averaging, Digital Surface Model

Abstract

Generating wide-area digital surface models (DSMs) requires registering a large number of individual, and partially overlapped DSMs. This presents a challenging problem for a typical registration algorithm, since when a large number of observations from these multiple DSMs are considered, it may easily cause memory overflow. Sequential registration algorithms, although can significantly reduce the computation, are especially vulnerable for small overlapped pairs, leading to a large error accumulation. In this work, we propose a novel solution that builds the DSM registration task as a motion averaging problem: pair-wise DSMs are registered to build a scene graph, with edges representing relative poses between DSMs. Specifically, based on the grid structure of the large DSM, the pair-wise registration is performed using a novel nearest neighbor search method. We show that the scene graph can be optimized via an extremely fast motion average algorithm with O(N) complexity (N refers to the number of images). Evaluation of high-resolution satellite-derived DSM demonstrates significant improvement in computation and accuracy.

1. Introduction

Generating wide-area digital surface models (DSMs) is pivotal in establishing foundational geospatial datasets. This often involves registering a large number of individual DSMs generated either from the same or different sources, such as from satellite stereo-based reconstruction (e.g. WorldView 2 & 3[1], PlanetScope (Huang et al., 2022, Qin and Gui, 2023)), aerial sensors (Xu et al., 2024, Huang and Qin, 2023), lidar, Synthetic Aperture Radar (SAR) (e.g. SRTM (JPL, 2013)). These individual DSMs, however, may have varying degrees of absolute geometric accuracy, as well as partial and even very minimal overlap. Thus, registering DSMs under this context presents as a challenging problem both due to both the scale of the problem, as well as the sub-optimal quality of input. For example, registering these DSMs may require minimizing geometric errors between billions of points, easily causing memory overflow, at the same time, registering partial and low-overlap DSMs easily brings large errors leading to error accumulation when multiple DSMs are registered. To this end, We propose a framework based on motion averaging, which entails enumerating all overlapped DSM pairs to establish a scene graph. Our approach involves the utilization of an efficient pair-wise registration method to eliminate systematic errors at the pair-wise level and subsequently redistribute these errors across the graph. This process aims to reduce and evenly distribute systematic errors among multiple DSMs.

As mentioned above, the task of registering multiple DSMs raises several unique and difficult challenges. To be more specific, first of all, the huge volume of data adds to the difficulty of memory and computation. The standard method is the iterative closest point (ICP) (Besl and McKay, 1992), which iteratively minimizes the cloud-to-cloud distance between a pair of 3D data. The correspondences are formed by searching for the nearest neighbor (NN) point of reference data for each query point, which usually can be sped up by k-d tree (Besl and McKay, 1992, Greenspan and Yurick, 2003), or octree (Eggert and Dalyot, 2012), but the initialization of a k-d tree necessitates caching the entire reference data into memory, requiring $O(N)$ space, $O(NlogN)$ time. Such methods become impractical for city or terrain-level data. For instance, caching a single WorldView-2 DSM (usually around 20000 × 20000 equals 400 million points) consumes 22GB for the construction of k-d tree. Another challenge is accurately estimating global poses from pair-wise registration results. The prevailing method adopts a greedy approach, initiating with the initial pair and estimating subsequent DSMs based on maximum overlap. However, as the DSM count increases, this approach accumulates small registration errors, resulting in a significant deviation from the actual scene.

To address the above-mentioned challenges, we propose a motion averaging-based framework for removing the systematic errors among multiple unaligned DSMs, as shown in Figure 1. We first propose an ICP-based pair-wise registration method that can perform fast and exact NN searching by utilizing the grid structure of DSM (termed as DSM-ICP). Its space and computation complexity are independent of the data volume, allowing excellent scalability to large-scale datasets. Based on it, we perform pair-wise registration to all possible pairs of DSM, establishing a reliable scene graph, where each edge represents the relative poses and carries weights based on its overlap and registration error. Finally, a motion averaging approach over the weighted edge graph is performed to redistribute the registration error for the estimation of the global poses. In our experiments, we quantitatively and qualitatively evaluate our proposed method using a large number of individual satellite-stereo-based DSMs against the ground-truth lidar point cloud data. The rest of this paper is organized as follows: section 2

[1] https://resources.maxar.com/data-sheets/worldview-2

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-2024-275-2024 | © Author(s) 2024. CC BY 4.0 License.

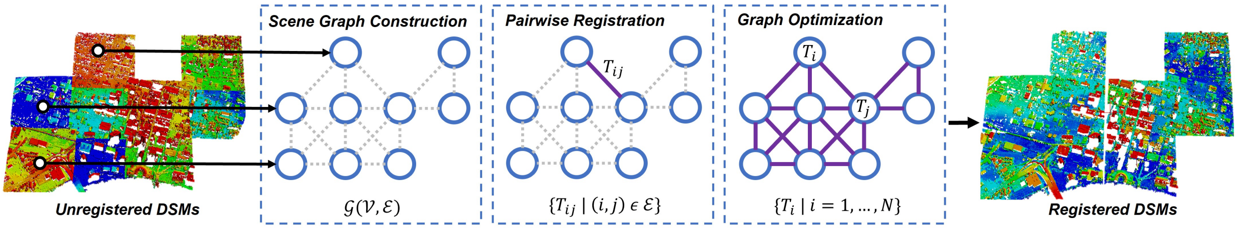


Figure 1. Overview of the proposed method. Given N unaligned DSMs, our target is to remove their systematic errors to produce seamless registered DSMs (the unaligned and registered DSMs are color-coded based on their heights). We first construct a scene graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with edges denoting all possible pairs of DSMs.Then, the proposed DSM-ICP (see Section 3.1) is performed for all edges determining their pair-wise transformation $\{T_{ij}|(i, j) \in \mathcal{E}\}$. Finally, a motion averaging approach is performed to estimate the global poses $\{T_i|i = 1, ..., N\}$ (see Section 3.2)

reviews the related works of 3D registration methods. section 3 introduces the proposed motion averaging framework in detail. section 4 presents the experimental results, comparative studies, and our analysis. section 5 concludes this work.

2. Related Work

Pair-wise 3D registration is the process of estimating the transformation (rigid-body transformation in our case) between a pair of 3D data sets to minimize the systematic error between two 3D data, which is widely used in 3D mapping and change detection (Xu et al., 2021) tasks. It can be categorized into coarse and fine registration according to the registration accuracy. Coarse registration assumes no initial alignment, involving the extraction of key primitives (e.g. key points (Rusu et al., 2009, Ao et al., 2021), lines (Chen and Yu, 2019), planes (Chen et al., 2019), 4-points congruent set (Mellado et al., 2014)), correspondence construction (Bai et al., 2021, Zhang et al., 2023), and transformation estimation (Barath and Matas, 2018, Yang et al., 2020). Such methods often yield suboptimal accuracy due to errors in primitive localization and matching, necessitating subsequent fine registration (Xu et al., 2023). The standard fine registration method is the ICP algorithm, which iteratively refines the transformation by searching for NN points and estimating the transformation. Variants of ICP aim to enhance its robustness (Sparse-ICP (Bouaziz et al., 2013), Robust ICP (Zhang et al., 2021)) and convergence speed (Point-to-Plane ICP (Arun et al., 1987), Symmetric ICP (Rusinkiewicz, 2019), AA-ICP (Pavlov et al., 2018)). The NN search typically employs k-d tree, requiring $O(N)$ memory for reference data caching and $O(logN)$ time for querying. Typically, CODEM[2] is also an ICP-based method for DSM registration while they transform DSM into point cloud format and perform standard ICP. In comparison, ours utilizes the grid structure of DSM and aims to optimize the memory and computation requirements of NN search to make it applicable to large-scale datasets.

Multi-view registration considers each DSM as a view, and simultaneously solves the registration of each DSM at the same time. This can be accomplished through growing-based and optimization-based approaches. Growing-based methods initiate registration with the pair possessing the largest overlap and progressively register new data to existing data. The order of data registration is determined using Kruskal's algorithm (Kruskal, 1956). This type of method is efficient while as the number of data sets increases, the accumulated registration error will lead to drift problems. Optimization-based methods involve constructing a scene graph, with each edge denoting relative pose and global pose estimation achieved by minimizing a predefined objective function — a concept known as motion averaging or pose graph optimization in the robotics community (Carlone et al., 2015). The closed-form solutions estimate the rotation using the singular value decomposition (SVD)-based methods (Arie-Nachimson et al., 2012, Gojcic et al., 2020, Wang et al., 2023) and then solve the translation using the least square methods (Gojcic et al., 2020). The iterative least-squares methods linearize the objective function and use Gauss-Newton or Levenberg-Marquardt methods to update the global poses. Our solution involves the SVD-based method to ensure computation and memory efficiency.

3. METHODOLOGY

Given a set of unaligned DSMs $\mathcal{P} = \{P_i|i = 1, ..., N\}$, our goal is to recover the global poses $\{T_i = (R_i, t_i) \in SE(3)|i = 1, ..., N\}$. In the following, we first introduce a computation and memory-efficient DSM registration method in Section 3.1. Based on it, we construct a scene graph with reliable weights for each edge and apply a motion-averaging approach to estimate the poses of every DSM in Section 3.2.

3.1 DSM-ICP

Aligning a pair of DSMs P and Q involves finding the rigid transformation $(R|t)$ such that applying the transformation to P causes the two DSMs to be as close as possible. The standard method is iterative closet point (ICP). It minimizes the distance between two DSMs by alternating between two steps:

- *Correspondence step*: search pairs of corresponding points (p_i, q_i^k), where q_i^k is the closest point to p_i given the current transformation $(R^k|t^k)$:

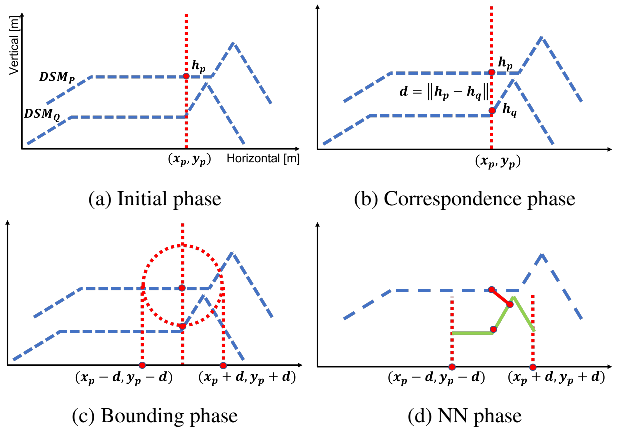
$$q_i^k = \arg\min_{q \in Q} \|R^k p_i + t^k - q\| \tag{1}$$

- *Estimation step:* estimate the new transformation (R^{k+1}, t^{k+1}) by optimizing the ℓ_2 distance given the current correspondences (p_i, q_i^k):

$$(R^{k+1}, t^{k+1}) = \arg\min_{R,t} \sum_i \|Rp_i + t - q_i^k\|^2 \tag{2}$$

Common methods solve the correspondence step by applying NN methods, such as k-d tree (Besl and McKay, 1992), approximate k-d tree (Greenspan and Yurick, 2003), or octree (Eggert

[2] https://github.com/NCALM-UH/CODEM



(a) Initial phase

(b) Correspondence phase

(c) Bounding phase

(d) NN phase

Figure 2. Illustration of the proposed NN search method. To simplify, DSMs are depicted as profiles in 2D space, with the x-axis representing the horizontal plane and the y-axis denoting height. The "Blue dashed" lines depict DSM data stored on disk, while the "Green solid" line represents cached data in RAM.

and Dalyot, 2012). These methods necessitate the initialization of the spatial structures to cache the whole reference data Q.Specifically, k-d tree needs $O(N)$ memory and $O(NlogN)$ time, and each NN searching requires $O(logN)$ time. In the estimation step, the closed-form solution involves separating the rotation and translation. Rotation estimation is achieved through SVD-based methods, while the translation can be readily determined.

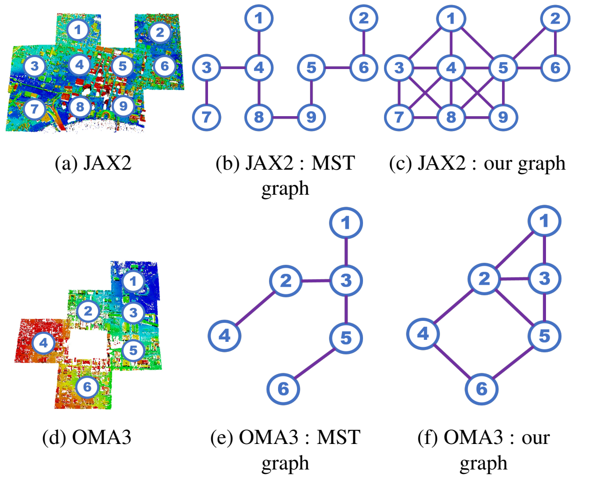
Efficient and exact NN searching. In a DSM, each pixel denotes a 3D point in world coordinates, where its uv coordinates $[u_i, v_i]$ indicate the geo-location in the horizontal plane and its value $h_i = dsm[u_i, v_i]$ represents the height in meters. The neighboring pixels within a range of x pixels can be queried in constant time using the operation $[u_i - x : u_i + x, v_i - x : v_i + x]$. Given a query pixel $[u_p, v_p]$ from the moving DSM P, we will introduce an efficient method to find the upper bound of its NN from the reference DSM Q in constant time.

Specifically, at the initial phase shown in Figure 2, we calculate its world coordinate (x_p, y_p) using Equation 3.1. Then, the pixel coordinate (u_q, v_q) from the other DSM Q covering the same world coordinates can be calculated using Equation 3.1. The height of that pixel can be looked up and termed as $h_q = DSM_q(u_q, v_q)$.

$$\begin{aligned}(x_p, y_p, h_p) &= uv2world(u_p, v_p, tfw_p)\\ (u_q, v_q) &= world2uv(x_p, y_p, h_p, tfw_q)\end{aligned} \tag{3}$$

Then, the pixel (u_q, v_q) serves as the initial correspondence to the query pixel (u_p, v_p), which is found in constant time. Moreover, their Euclidean distance also serves as the upper bound that the NN is guaranteed to be located within the 3D sphere ball centered at the query 3D point with a radius of the d, where $d = \|h_p - h_q\|$.

At the bounding phase, the 3D sphere ball is projected into the horizontal plane, formulating the 2D rectangle $[x_q - d : x_q + d, y_q - d : y_q + d]$ as the final bound. Finally, at the NN phase, the uv coordinates of the bounding rectangle are calculated and any pixels within the bounding rectangle are cached. By traversing all the cached pixels, we can find the NN.



(a) JAX2

(b) JAX2 : MST graph

(c) JAX2 : our graph

(d) OMA3

(e) OMA3 : MST graph

(f) OMA3 : our graph

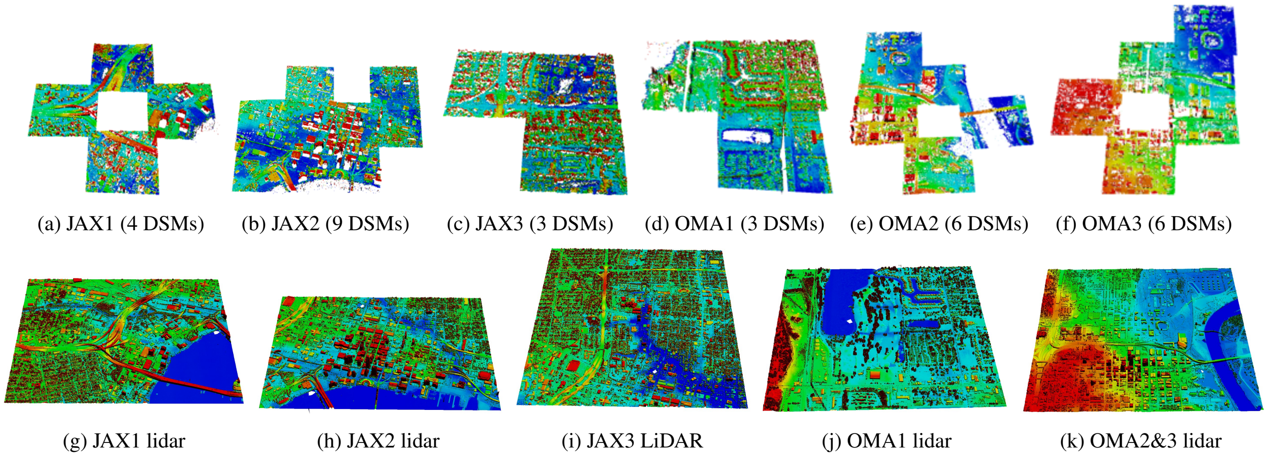
Figure 3. Comparison of two approaches for scene graph construction. "MST" denotes the minimal spanning tree, which identifies the minimum number of edges connecting all nodes, while our method establishes edges connecting all possible nodes. In this scenario, the path between any two nodes in our graph is shorter than or equal to those in the MST graph (e.g., node 4&6 in OMA3), significantly mitigating the accumulated error.

The primary impediment to the application of k-d tree to large data volumes lies in the necessity to initialize by caching the entire dataset, requiring $O(NlogN)$ time and $O(N)$ memory space. A frequently compromised solution involves building a k-d tree for downsampled datasets, albeit at the expense of accuracy reduction. In comparison, our NN search method eliminates the need for initialization. All data are stored on disk, and only local data around the query points are cached. For each NN search, our method demands $O(k)$ time and memory space, where the upper bound d is calculated in constant time and k is the number of pixels within the upper bound. In practice, $k \ll N$ and as the two DSMs are getting closer, k will decrease dramatically. Our approach achieves both memory and computational efficiency compared to the k-d tree method. This enables ICP to be applied to large-scale data without compromising accuracy.

3.2 Motion Averaging-based Framework

Scene graph construction. The scene graph serves as a potent tool for representing pair-wise information among multiple objects and finds extensive application in various engineering problems. In this context, our objective is to construct a scene graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where each vertex $v_i \in \mathcal{V}$ corresponds to a DSM P_i and each edge $(i, j) \in \mathcal{E}$ encodes the relative pose between DSM P_i and P_j.

Existing growing-based methods apply a greedy approach to search for the minimal number of edges connecting all vertex. As the number of vertex increases, the accumulated error between vertex that have a large path cannot be reduced, as shown in Figure 3. A reliable scene graph is intended to capture extensive relative information for providing redundant constraints to make sure any two vertex have a small path. Furthermore, the reliability of each edge must be taken into account to filter out potential outliers. Therefore, for each pair (P_i, P_j),



(a) JAX1 (4 DSMs) (b) JAX2 (9 DSMs) (c) JAX3 (3 DSMs) (d) OMA1 (3 DSMs) (e) OMA2 (6 DSMs) (f) OMA3 (6 DSMs)

(g) JAX1 lidar (h) JAX2 lidar (i) JAX3 LiDAR (j) OMA1 lidar (k) OMA2&3 lidar

Figure 4. Illustration of six test sets of satellite DSMs and the ground truth lidar data, where three test sets in Jacksonville area (JAX) and three test sets in Omaha area (OMA). The detailed description is in Section 4.1

we calculate the overlap score by

$$s_{ij} = \frac{\#overlap\,pixels}{\#valid\,pixels} \tag{4}$$

where $s_{ij} \in [0, 1]$; Assuming the provided DSMs are approximately geo-referenced, the identification of overlap pixels depends on the presence of corresponding reference pixels at the same geographical location. Once the s_{ij} is larger than a predefined threshold, the edge will be inserted into the scene graph. Their relative pose T_{ij} will further be estimated using our DSM-ICP outlined in Section 3.1.

Weight initialization. The edge reliability can be determined by two factors: the overlap ratio and the quality of pair-wise registration. We associate a weight on each edge to indicate the reliability of the estimated relative pose T_{ij}. It is calculated based on both the overlap score s_{ij} and the quality score of pair-wise registration r_{ij}:

$$\begin{aligned} w_{ij} &= s_{ij} * r_{ij} \\ r_{ij} &= \frac{e^{-err_{ij}}}{\sum_{(i,j)\in\mathcal{E}} e^{-err_{ij}}} \end{aligned} \tag{5}$$

where err_{ij} represents the pair-wise registration error representing the estimation error of registering the correspondences at the last iteration. It can be calculated based on Equation 2 with T_{ij}.

Motion averaging. Given the edge weights and the relative poses $\{w_{ij}, T_{ij} = (R_{ij}, t_{ij}) | (i, j) \in \mathcal{E}\}$, we estimate the global poses $\{T_i = (R_i, t_i)\}$. It is also known as the pose graph optimization problem in the robotics community (Carlone et al., 2015). Each relative pose $(R_{ij}|t_{ij})$ can be represented by global poses $(R_i|t_i)$ and $(R_j|t_j)$

$$t_{ij} = R_i^T(t_j - t_i) + t_{ij}^{\epsilon}, R_{ij} = R^T R_j R_{ij}^{\epsilon} \tag{6}$$

where the $t_{ij}^{\epsilon} \in \mathbb{R}^3, R_{ij}^{\epsilon} \in SO(3)$ denote the measurement noise. Therefore, we estimate the global poses $\{R_i, t_i\}$ by solving the optimization problem

$$\min_{\substack{\{R_i\}\in SO(3) \\ \{t_i\}\in\mathbb{R}^3}} \sum_{(i,j)\in\mathcal{E}} w_{ij}\|R_{ij} - R_i^T R_j\|_F^2 + w_{ij}\|R_i t_{ij} + t_i - t_j\|^2 \tag{7}$$

where $\|\cdot\|_F$ means the Frobenius norm of the matrix. The problem can be addressed by decoupling the optimization of rotation and translation. We first estimate the global rotation R_i using the closed-form solutions (Arie-Nachimson et al., 2012, Gojcic et al., 2020, Wang et al., 2023). Once the rotation is determined, the translation t_i can be computed using the standard least square method (Gojcic et al., 2020).

4. Experimental Results

In this section, two experiments were performed to evaluate the proposed method in terms of: 1) **pair-wise registration performance**: the memory and computation performance of the proposed DSM-ICP given varying volumes of datasets, introduced in Section 4.2, and 2) **multiple DSM registration performance**: the average relative error among all DSMs and the overall error with respect to the ground truth data, introduced in Section 4.3. The data and evaluation metrics used in the experiments are introduced in Section 4.1.

4.1 Datasets and Metrics

Datasets. We choose the public multi-view satellite datasets: DFC 2019 (tack 3) (Le Saux et al., 2019), encompassing the city of Jacksonville (short for JAX), Florida, USA, and Omaha (short for OMA), Nebraska, USA. These datasets were captured by WorldView-3[3] with ground-sampling distances of 35cm. In track 3, the satellite images were cropped into sub-images, totaling 118 sub-areas. Each subarea covers $700m \times 700m$ and is observed by 30 images. The sub-areas partially overlap, and we manually chose six sets of data as the test data, as illustrated in Figure 4. For each sub-area, we conducted dense matching to generate the DSM using satellite 3D reconstruction software[4]. The ground truth data for both regions are airborne lidar scanning data from the USGS 3DEP program[5], with a point density of 24.64 pt/m^2, 5.82 pt/m^2 in Omaha area.

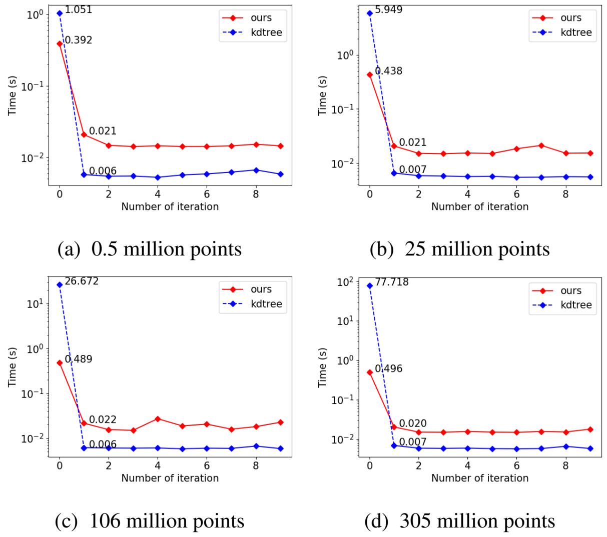
Metrics. The pair-wise registration performance can be reflected by the distance between two registered DSMs. In practical scenarios, the standard Root Mean Squared Error (RMSE) may not be effective due to the presence of outlier pixels

[3] https://resources.maxar.com/data-sheets/worldview-3
[4] https://u.osu.edu/qin.324/rsp/
[5] https://www.usgs.gov/3d-elevation-program

https://doi.org/10.5194/isprs-annals-X-1-2024-275-2024

Method	Scene Graph	$MEAN_{RMSE_\tau}$ [m]					
		JAX1	JAX2	JAX3	OMA1	OMA2	OMA3
-	-	2.213	1.715	2.632	2.231	1.877	2.126
Greedy	MST	**1.832**	1.514	**1.474**	1.490	1.760	1.903
Ours	Full	1.834	**1.424**	**1.474**	**1.486**	**1.727**	**1.856**

Table 1. Quantitative results of accumulated error reduction performance for two methods. ”$MEAN_{RMSE_\tau}$” represents the average $RMSE_\tau$ for all possible pairs. The first row represents the alignment of the raw dataset without any registration process. When the number of DSMs is small (e.g. JAX1, JAX3, and OMA1), the two methods achieve similar performance. When the number of DSMs is large (e.g. JAX2, OMA2, and OMA3), ours achieve better error reduction performance.



(a) 0.5 million points

(b) 25 million points

(c) 106 million points

(d) 305 million points

Figure 5. Computation efficiency of NN search for DSM-ICP and the standard ICP using k-d tree. The y-axis represents the time cost of the NN search at each iteration. In this experiment, 2065 points in total were performed NN search within the varying number of reference points from 0.5 million (a) to 305 million points (d). For ICP using k-d tree, the ”Blue dashed” line represents the initialization of k-d tree, followed by NN search.

within DSMs, which can significantly influence overall metrics. Therefore, we implement a straightforward outlier rejection mechanism in conjunction with RMSE:

$$rmse_\tau = \sqrt{\frac{1}{N}\sum_{i\in P}[\![\|p_i - q_i\| < \tau]\!](p_i - q_i)^2} \tag{8}$$

where the $[\![\cdot]\!]$ is the Iverson bracket, τ is a pre-defined inlier threshold (10m in our case). (p_i, q_i) is the pair of points corresponding to the same horizontal location.

4.2 Performance of Pair-wise Registration

The primary advantage of the proposed DSM-ICP lies in the elimination of the need for initializing spatial data structures such as k-d tree. In each NN search, its cached data is confined to the local area rather than encompassing the entire region. We assess its computational and memory efficiency across diverse volumes of reference data. Specifically, we measure the time cost for every NN search and the overall RAM consumption for both DSM-ICP and the standard ICP utilizing k-d tree.

Computation efficiency performance. We measure the time cost during the correspondence step (refer to Section 3.1) at each iteration for both the proposed DSM-ICP and the standard ICP employing k-d tree. Specifically, we fix the number of query points to 2065 while varying the volume of reference data from 0.5 million to 305 million points. The results are illustrated in Figure 5. It is evident that ICP using k-d tree initially consumed more time as it traversed all reference points on the local disk to construct the k-d tree in RAM. In contrast, our method bypassed this process. The time cost gradually decreased as the two DSMs moved apart initially, resulting in a larger searching range d (refer to details in Section 3.1). As the two DSMs approached each other, the searching range d decreased, leading to a smaller time cost for the NN search. After several iterations, the time cost for both methods stabilized. Our method required more time due to accessing data via disk, while k-d tree accessed data via RAM.

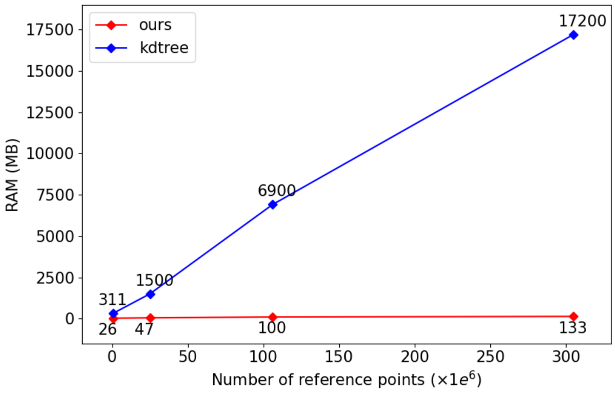


Figure 6. Memory consumption of the proposed DSM-ICP and the standard ICP using k-d tree with the varying number of reference points. The experiment setting is the same as Figure 5

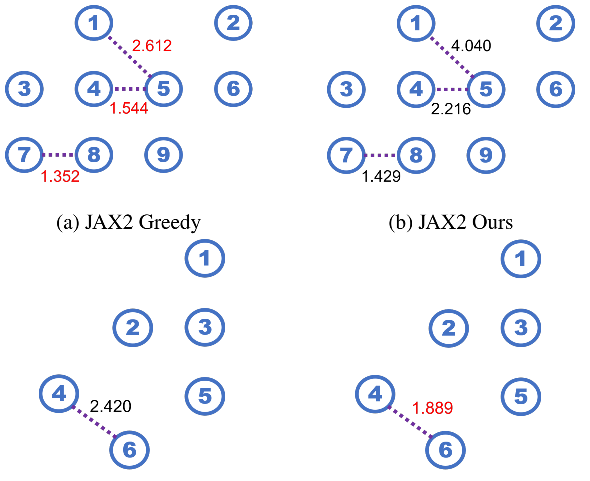
Memory efficiency performance. We measure the total RAM utilization throughout the entire iteration process with varying volumes of reference data. As depicted in Figure 6, the RAM consumption of the k-d tree exhibits linearity with the number of reference points, as the space complexity of constructing a k-d tree is $O(N)$, implying it loads all reference points into RAM. In contrast, our RAM consumption remains independent of the volume of reference data. When the number of reference points reached 305 million, our method required only 133 MB, while the k-d tree demanded 17200 MB—129 times larger than ours, revealing the superior memory efficiency of our approach.

4.3 Performance of Multiple DSM Registration

We assess the reduction in accumulated error and the reconstruction performance of the proposed graph-based method. The distinctive advantage of our approach lies in the elimination of accumulated registration errors through redundant pairwise constraints. The conventional approach for registering multiple DSMs employs a greedy method, which initiates registration with the pair exhibiting the largest overlap and sub-

Method	Scene Graph	$RMSE_\tau$ [m]					
		JAX1	JAX2	JAX3	OMA1	OMA2	OMA3
Greedy	MST	2.305	2.166	2.756	2.065	1.461	1.667
Ours	Full	**2.302**	**2.129**	2.756	2.065	**1.451**	**1.539**

Table 2. Quantitative assessment of the quality of the fused DSM, measured by $RMSE_\tau$ against the lidar ground truth. Our method demonstrates superior reconstruction quality in 4 out of 6 sets. As observed in Figure 9, particularly for scenarios with a large number of DSMs (e.g., JAX2, OMA2, OMA3), our approach outperforms the greedy method significantly.



(a) JAX2 Greedy

(b) JAX2 Ours

(c) OMA3 Greedy

(d) OMA3 Ours

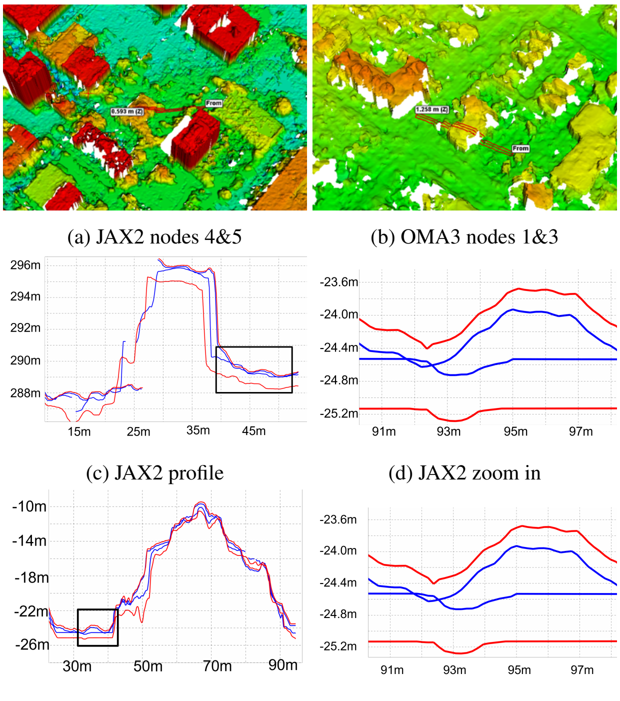
Figure 7. Registration errors of selected pairs in JAX2 and OMA3 area. The MST and our graphs are shown in Figure 3. We only display the pairs that are likely to accumulate the errors. Two nodes of these pairs are connected by a long path (more than 3 edges) in the MST graph.

sequently registers the next largest overlap for the remaining DSMs. This can be accomplished using Kruskal’s algorithm to find the minimal spanning tree (MST) of an undirected edge-weighted graph. Both the greedy method and our approach utilize the proposed DSM-ICP as the pair-wise registration method.

Accumulated error reduction performance. We apply the proposed method and the greedy method to six test sets (depicted in Figure 4) and gauge the average $RMSE_\tau$ for all possible pairs as the evaluation metric. As reported in Table 1, our method can reduce initial systematic errors by $0.2 - 1.2m$. It achieved the highest registration accuracy in 5 out of 6 sets. The greedy algorithm performs well when the graph size is small; however, as the graph enlarges, relative registration errors will accumulate along the path connecting two distant nodes. The advantage of our method lies in the provision of more pair-wise constraints, offering additional directions to distribute the accumulated error across the graph.

Specifically, we select two test sets and present their pair-wise $RMSE_\tau$ in Figure 7. Nodes 4 and 6 in OMA3 are four steps away from each other in the MST graph and their $RMSE_\tau$ is 2.420m. In comparison, ours reduced the registration error to 1.889m. The same holds for the nodes 1&5, 4&5, and 7&8 in JAX2.

Additionally, we showcase the qualitative results of nodes 4 & 5 of JAX2 and nodes 4 & 6 of OMA3 by illustrating the profiles of registered DSMs in Figure 8. In the JAX2 area, DSMs registered by the greedy method exhibit a rough 1m height difference, whereas ours nearly perfectly aligns the two DSMs, considering the quality of the DSM. Similarly, our method achieves



(a) JAX2 nodes 4&5

(b) OMA3 nodes 1&3

(c) JAX2 profile

(d) JAX2 zoom in

(e) OMA3 profile

(f) OMA3 zoom in

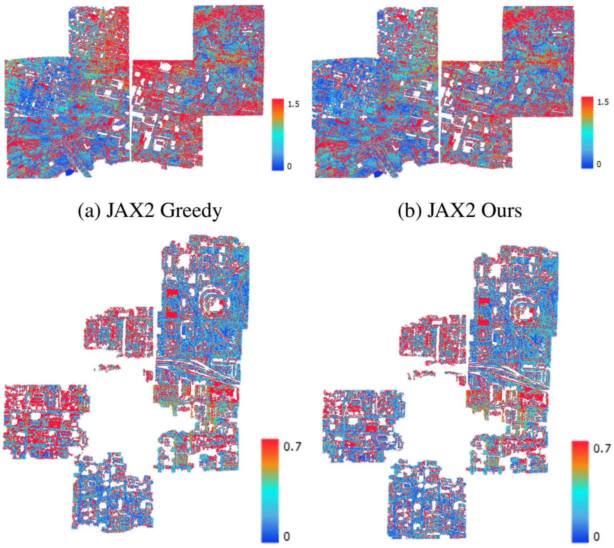
Figure 8. Profiles of selected buildings in JAX2 and OMA3 area. We show the profiles of registered DSMs by greedy method (red lines) and ours (blue lines).

nearly perfect registration of DSMs in OMA3, while the greedy method results in a rough 1m height difference.

Reconstruction performance. After eliminating systematic errors among the DSMs, we fuse all registered DSMs into a single DSM and assess its quality against lidar ground truth. Specifically, we align the fused DSM (achieved using RSP) with the ground truth using DSM-ICP and compute $RMSE_\tau$. The quantitative results are presented in Table 2, wherein our approach achieved superior reconstruction accuracy in 4 out of 6 areas. Moreover, we showcase qualitative results illustrating the quality of the fused DSM in two selected areas, JAX2 and OMA3, as depicted in Figure 9. Regions with significant errors are colorized in ”Red”. In the JAX2 area, large errors produced by the greedy method are mainly concentrated around the middle part, where accumulated errors between nodes 4&5 and 1&5 (see Figure 7) were not effectively reduced by the greedy method. Similar observations apply to the left part of the OMA3 area.

5. Conclusion

In this paper, we presented a motion averaging-based approach for the registration of multiple large-scale DSMs. Specifically, we proposed a fast and exact NN search method by utilizing the



(a) JAX2 Greedy (b) JAX2 Ours

(c) OMA3 Greedy (d) OMA3 Ours

Figure 9. Qualitative results of error map depicting the height difference between the fused DSM and the lidar ground truth. The greedy method produces more error data than ours, particularly around the middle-top part in JAX2 and the left part in OMA3.

grid structure of DSM. Its computation and memory complexity are independent of the volume of the input data, enabling our pair-wise registration method, DSM-ICP, to be scalable to large-scale data. Based on it, we perform pair-wise registration for all possible pairs of DSMs to construct a reliable scene graph, where the edge is weighted based on overlap and registration error. Then, the global poses for each DSM are estimated by redistributing the pair-wise registration errors over the whole scene graph. We evaluated the proposed method on public satellite datasets with the ground truth airborne lidar data. The experiment results demonstrate that our DSM-ICP achieves both superior computation and memory efficiency than k-d tree-based ICP. The proposed motion averaging-based method achieves better error reduction performance and reconstruction performance than the greedy method.

References

Ao, S., Hu, Q., Yang, B., Markham, A., Guo, Y., 2021. Spinnet: Learning a general surface descriptor for 3d point cloud registration. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11753–11762.

Arie-Nachimson, M., Kovalsky, S. Z., Kemelmacher-Shlizerman, I., Singer, A., Basri, R., 2012. Global motion estimation from point matches. *2012 Second international conference on 3D imaging, modeling, processing, visualization & transmission*, IEEE, 81–88.

Arun, K. S., Huang, T. S., Blostein, S. D., 1987. Least-squares fitting of two 3-D point sets. *IEEE Transactions on pattern analysis and machine intelligence*, 698–700.

Bai, X., Luo, Z., Zhou, L., Chen, H., Li, L., Hu, Z., Fu, H., Tai, C.-L., 2021. Pointdsc: Robust point cloud registration using deep spatial consistency. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15859–15869.

Barath, D., Matas, J., 2018. Graph-cut ransac. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6733–6741.

Besl, P. J., McKay, N. D., 1992. Method for registration of 3-d shapes. *Sensor fusion IV: control paradigms and data structures*, 1611, Spie, 586–606.

Bouaziz, S., Tagliasacchi, A., Pauly, M., 2013. Sparse iterative closest point. *Computer graphics forum*, 32number 5, Wiley Online Library, 113–123.

Carlone, L., Tron, R., Daniilidis, K., Dellaert, F., 2015. Initialization techniques for 3d slam: A survey on rotation estimation and its use in pose graph optimization. *2015 IEEE international conference on robotics and automation (ICRA)*, IEEE, 4597–4604.

Chen, S., Nan, L., Xia, R., Zhao, J., Wonka, P., 2019. PLADE: A plane-based descriptor for point cloud registration with small overlap. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4), 2530–2540.

Chen, X., Yu, K., 2019. Feature line generation and regularization from point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12), 9779–9790.

Eggert, D., Dalyot, S., 2012. Octree-based simd strategy for icp registration and alignment of 3d point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences; I-3*, 1(1), 105–110.

Gojcic, Z., Zhou, C., Wegner, J. D., Guibas, L. J., Birdal, T., 2020. Learning multiview 3d point cloud registration. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1759–1769.

Greenspan, M., Yurick, M., 2003. Approximate kd tree search for efficient icp. *Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.*, IEEE, 442–448.

Huang, D., Qin, R., 2023. a critical analysis of internal reliability for uncertainty quantification of dense image matching in multi-view stereo. *arXiv preprint arXiv:2309.09379*.

Huang, D., Tang, Y., Qin, R., 2022. An evaluation of PlanetScope images for 3D reconstruction and change detection–experimental validations with case studies. *GIScience & Remote Sensing*, 59(1), 744–761.

JPL, N., 2013. Nasa shuttle radar topography mission global 1 arc second [data set].

Kruskal, J. B., 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1), 48–50.

Le Saux, B., Yokoya, N., Hänsch, R., Brown, M., 2019. Data fusion contest 2019 (dfc2019).

Mellado, N., Aiger, D., Mitra, N. J., 2014. Super 4pcs fast global pointcloud registration via smart indexing. *Computer graphics forum*, 33number 5, Wiley Online Library, 205–215.

Pavlov, A. L., Ovchinnikov, G. W., Derbyshev, D. Y., Tsetserukou, D., Oseledets, I. V., 2018. Aa-icp: Iterative closest point with anderson acceleration. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 3407–3412.

Qin, R., Gui, S., 2023. Using PlanetScope-Derived Time-Series Elevation Models to Track Surging Glacier 3D Dynamics in Mid-Latitude Mountain Regions. *AGU23*.

Rusinkiewicz, S., 2019. A symmetric objective function for ICP. *ACM Transactions on Graphics (TOG)*, 38(4), 1–7.

Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. *2009 IEEE international conference on robotics and automation*, IEEE, 3212–3217.

Wang, H., Liu, Y., Dong, Z., Guo, Y., Liu, Y.-S., Wang, W., Yang, B., 2023. Robust multiview point cloud registration with reliable pose graph initialization and history reweighting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9506–9515.

Xu, N., Huang, D., Song, S., Ling, X., Strasbaugh, C., Yilmaz, A., Sezen, H., Qin, R., 2021. A volumetric change detection framework using UAV oblique photogrammetry–a case study of ultra-high-resolution monitoring of progressive building collapse. *International Journal of Digital Earth*, 14(11), 1705–1720.

Xu, N., Qin, R., Huang, D., Remondino, F., 2024. Multi-tiling neural radiance field (nerf) – geometric assessment on large-scale aerial datasets.

Xu, N., Qin, R., Song, S., 2023. Point cloud registration for LiDAR and photogrammetric data: A critical synthesis and performance analysis on classic and deep learning algorithms. *ISPRS open journal of photogrammetry and remote sensing*, 100032.

Yang, H., Shi, J., Carlone, L., 2020. Teaser: Fast and certifiable point cloud registration. *IEEE Transactions on Robotics*, 37(2), 314–333.

Zhang, J., Yao, Y., Deng, B., 2021. Fast and robust iterative closest point. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 3450–3466.

Zhang, X., Yang, J., Zhang, S., Zhang, Y., 2023. 3d registration with maximal cliques. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17745–17754.