remote sensing

Article

An Efficient and Robust Hybrid SfM Method for Large-Scale Scenes

Zhendong Liu [1,2], Wenhu Qv [2], Haolin Cai [3], Hongliang Guan [1,*] and Shuaizhe Zhang [3]

[1] College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China
[2] Chinese Academy of Surveying & Mapping, Beijing 100036, China
[3] College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
[*] Correspondence: hlguan@cnu.edu.cn; Tel.: +86-010-6388-0555

Abstract: The structure from motion (SfM) method has achieved great success in 3D sparse reconstruction, but it still faces serious challenges in large-scale scenes. Existing hybrid SfM methods usually do not fully consider the compactness between images and the connectivity between subclusters, resulting in a loose spatial distribution of images within subclusters, unbalanced connectivity between subclusters, and poor robustness in the merging stage. In this paper, an efficient and robust hybrid SfM method is proposed. First, the multifactor joint scene partition measure and the preassignment balanced image expansion algorithm among subclusters are constructed, which effectively solves the loose spatial distribution of images in subclusters problem and improves the degree of connection among subclusters. Second, the global GlobalACSfM method is used to complete the local sparse reconstruction of the subclusters under the cluster parallel framework. Then, a decentralized dynamic merging rule considering the connectivity of subclusters is proposed to realize robust merging among subclusters. Finally, public datasets and oblique photography datasets are used for experimental verification. The results show that the method proposed in this paper is superior to the state-of-the-art methods in terms of accuracy and robustness and has good feasibility and advancement prospects.

Keywords: structure from motion; hybrid SfM methods; partition-merge strategy; compactness; connectivity; robustness

check for updates

Citation: Liu, Z.; Qv, W.; Cai, H.; Guan, H.; Zhang, S. An Efficient and Robust Hybrid SfM Method for Large-Scale Scenes. *Remote Sens.* **2023**, *15*, 769. https://doi.org/10.3390/rs15030769

Academic Editor: Benoit Vozel

Received: 25 December 2022
Revised: 20 January 2023
Accepted: 27 January 2023
Published: 29 January 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Structure from motion (SfM) is a process for estimating 3D scene structure and camera pose from a group of unstructured images. It is also called sparse reconstruction in the 3D reconstruction process [1,2]. SfM photogrammetry is widely used in many areas, such as geoscience [3], augmented reality [4,5], physical geography [6,7], archaeology [8,9], and autonomous navigation [10,11]. SfM includes three parts: feature extraction and matching, initial camera pose estimation, and bundle adjustment. According to the different initial camera pose estimation methods, SfM methods can be roughly divided into incremental [12–15], global [16–19], and hybrid [20–23] methods. The hybrid method combines incremental and global methods, that is, the partition-merge strategy [22], which has the characteristics of high efficiency and the ability to process large-scale scenes composed of tens of thousands or even hundreds of thousands of image data, and it effectively prevents error accumulation, drift problems, and memory bottlenecks [24,25].

According to the different partition-merge strategies, hybrid methods can be divided into hierarchical methods and flat methods. Hierarchical methods use hierarchical agglomerative clustering technology for scene partitioning and merging of large-scale datasets [22,26–30], but these methods need to initialize the leaf nodes of a tree with agglomerative clustering and use the bottom-up method for scene reconstruction and merging, so they lack a global perspective. The number of partitioned subgraphs is large, which easily leads to redundant calculations and merge failure. The flat method divides the dataset into several subclusters of the same size, and can use multicore technology or perform parallel reconstruction of small scenes on the distributed computing system, and the merging rules are relatively simple [20,31–34]. However, there are at least two problems in the existing flat methods: (1) Scene division stage: After subcluster clustering, the spatial distribution of images in the subcluster is relatively loose. When the subcluster is expanding, the efficiency of the expanded image is low, and the connectivity between the subclusters is not considered. (2) Subcluster merging stage: Due to the long merging path, the existing methods tend to cause error accumulation, which leads to the sparse reconstruction failure of the whole scene.

Therefore, this paper proposes an efficient and robust hybrid SfM method for large-scale scenes. The main contributions include the following three aspects:

(1) Image clustering with compact spatial distribution: To avoid the image clustering result of a weak connection degree in subclusters, a multifactor joint scene partition measure was constructed, including the number of image pairs with homologue points, the overlapping area of the image, and the number of common neighbors of the image;

(2) Image expansion considering connectivity: To improve the image expansion efficiency and ensure as much connectivity among subclusters as possible, a complete rate-guided preallocation equalization image expansion algorithm among subclusters is proposed;

(3) Robust subcluster merging: To avoid the large error accumulation problem caused by the long merging path of subclusters, a multilevel merging rule considering the subcluster connectivity is proposed. To ensure the accuracy of the merged subclusters as much as possible, considering the global perspective, the optimal internal camera parameter of the two subclusters to be merged was selected as the internal camera parameter of the merged cluster block.

The structure of this paper is as follows. The existing flat hybrid SfM methods and their shortcomings are introduced in Section 2. An efficient and robust hybrid SfM method for large scale is described in Section 3. The experiments and results analysis are presented in Section 4. The conclusions of this algorithm are introduced in Section 5.

2. Related Works

Because the hybrid SfM adopts the partition-merging processing strategy, this kind of method shows the incomparable advantages of the traditional SfM method in the large-scale scene sparse reconstruction problem. According to the different partition-merge strategies, the existing hybrid methods can be roughly divided into two categories: hierarchical methods and flat methods. The method in this paper belongs to the second category, so our discussion of earlier work focuses on the flat hybrid SfM method.

2.1. Existing Flat Hybrid SfM Method

To better adapt the initial camera pose estimation to large-scale images, scholars have continued to carry out research related to planar hybrid SfM methods, focusing on the two phases of scene partitioning and subcluster merging.

Bhowmick et al. [35] proposed a flat hybrid SfM, which directly divided the image dataset into N subclusters without establishing a tree structure, and merged the common camera pose and scene information of each subcluster after independent reconstruction, thus improving the efficiency and robustness of subcluster merging. Sweeney et al. [36] clustered the dataset through the distributed camera model and then merged each clustering result. On this basis, Zhu et al. [32] introduced the completeness and size constraint into scene partitioning to improve the connectivity between subclusters on the premise of ensuring the same size of all subclusters to ensure the robustness of subcluster merging. Lu et al. [33] took the image neighbor matrix obtained from the feature matching results as the input of the segmentation algorithm and reordered the images by repeatedly using the matrix bandwidth reduction algorithm to solve the subgraph discontinuity problem caused by the normalized cut.

Existing flat methods continuously improve and develop the two stages of scene partitioning and subcluster merging. On this basis, Chen et al. [20] proposed a graph-based method that regarded large-scale SfM as a graph problem. In the subcluster partition stage, the large-scale dataset is partitioned into several subclusters by a graph structure. In the merge stage, the minimum spanning tree in graph theory is used to find the optimal merging path to complete the merge task. The basic principle of this method is as follows:

Step 1: Subcluster partitioning, specifically subdivided into image clustering and image expansion. In the image clustering stage, each image is regarded as a node of the graph, the edge in the graph represents the connectivity between images, and the weight of the edge is the number of matching feature points between the geometrically filtered image pairs (the number of points with homologue points). In the image expansion process, the graph concept is extended to the subcluster level, and the subcluster is regarded as the node in the graph. The edge weight in the graph is the number of lost edges after adjacent subclusters are cut; to construct a maximum spanning tree, image edges are collected from the tree and sorted in descending order. Finally, the lost edges are added to the clusters that do not meet the completion readiness rate constraint.

Step 2: After the image dataset in the scene is divided into several subclusters, the subclusters can be reconstructed by the parallel local SfM method until all subclusters complete sparse reconstruction.

Step 3: Merge the subcluster reconstruction results. The similarity transformation is calculated by using the overlapping information between subclusters, and the disturbance of outliers is processed by RANSAC. The base subclusters and the optimal merging path are selected, and the leaf nodes are merged layer by layer from bottom to top through the precalculated transformation parameters. Finally, all the subclusters are transformed into a unified coordinate system.

2.2. Deficiencies of Existing Methods

The existing methods are very mature in the aspect of local sparse reconstruction of subclusters. However, in the subcluster division and merging stage, at least the spatial distribution of the images inside the subclusters is loose. When the subclusters expand, the efficiency of the expanded images is low, the connectivity between the subclusters is not considered, and the subcluster merging is not sufficiently robust.

(1) The spatial distribution of images in the subcluster is loose: the first step in the subcluster partition stage is image clustering. The clustered images have the problem of weak connectivity. To meet the reconstruction accuracy requirements, it is generally necessary to ensure a high degree of overlap between the captured images, so the angle between the camera shooting direction and the vertical direction of the ground may be very large (such as the oblique camera mounted on the UAV), as shown in Figure 1, camera C_b. Therefore, in 3D space, the images collected by the camera far away from the ground object (C_b) and the image collected by the camera close to the ground object (C_a) may still have some overlap and can form an image pair with the homologue point as the correlation factor, but this image pair generally has a weak association. As shown in Figure 1, cameras C_a and C_b of the two stations, with distance d, correspond to images I_a and I_b. There is a small overlap region between the two images, and the homologue points in the overlap region can satisfy the geometric filtering condition to form an image pair that can be used in the next stage of camera pose estimation. The existing methods use normalized cuts for image clustering, which may divide this type of image pair (I_a and I_b in Figure 1) into the same subcluster. The blue part in Figure 2 is a certain subcluster, $block_1$, whose image is divided into two parts of spatial fracture.

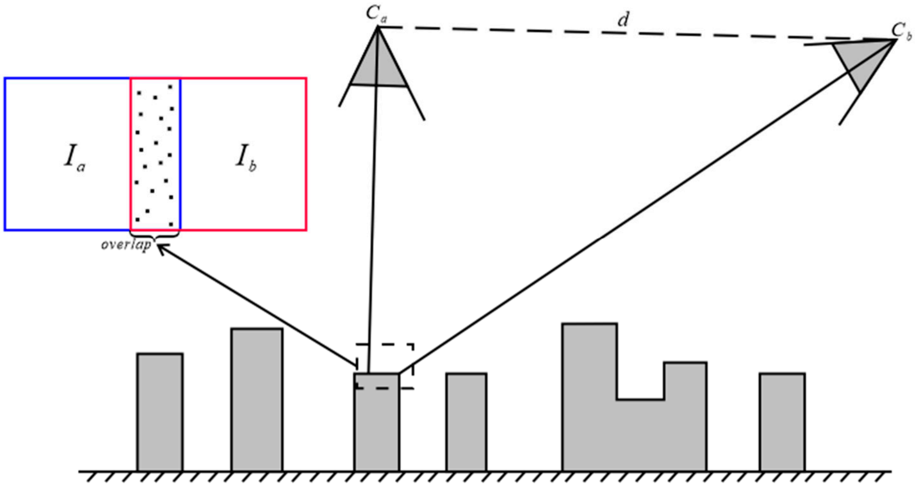


Figure 1. Camera station and image pair diagram.

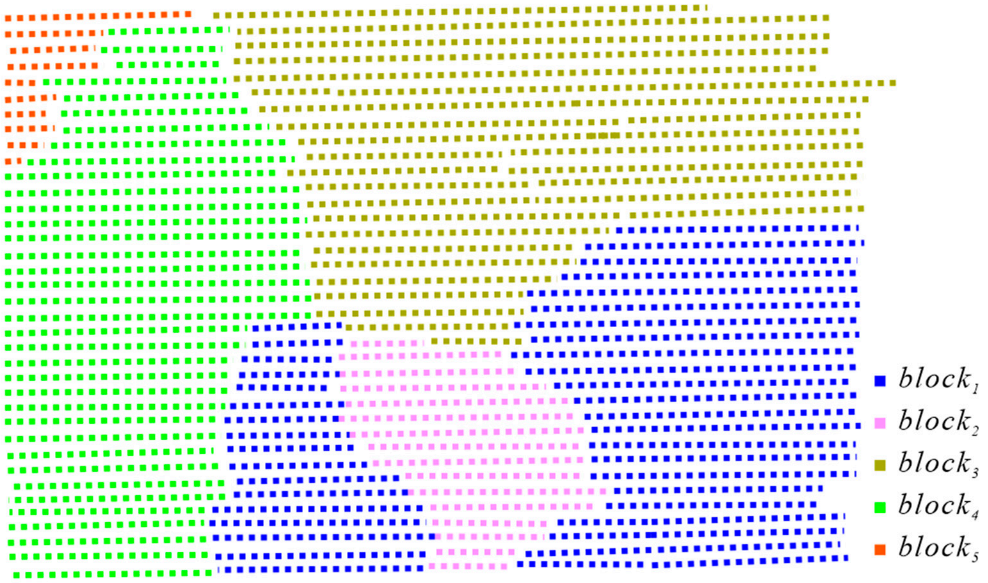


Figure 2. Diagram of the spatial discontinuity of image clustering.

(2) Low efficiency of subcluster expansion and lack of consideration for subcluster connectivity:

The image expansion of existing methods adds images associated with lost edges to subclusters that do not satisfy the completeness constraint. First, according to the completeness formula $\eta(i) = \frac{\sum_{j\neq i}|C_i \cap C_j|}{C_i}$ (where C_i and C_j are represented as subclusters i and j, respectively), we determine whether the current image to be added satisfies the preset completeness rate. The process is subcluster-by-subcluster and lost edge-by-edge cyclic, which consumes more time when faced with large-scale image expansion. The existing methods only add images associated with lost edges to the subclusters with the least number of images first under the constraint of completeness rate and do not consider the connectivity with the neighbor subclusters.

(3) The lack of subcluster merging robustness:

There are two main reasons for the lack of subcluster merging robustness: first, the subcluster connectivity is not strong; second, the benchmark cluster needs to be selected, and the merging path may be long, so the merging of subclusters easily causes error accumulation.

3. Methodology

In view of the shortcomings of the existing methods, an efficient and robust hybrid SfM method for a large scale is proposed in this paper. The core content includes three parts: (1) Scene partitioning. To avoid weakly associated image clustering results in subclusters, that is, an image pair with a loose, empty three-dimensional distribution, a multifactor joint image clustering measure is constructed with a multifactor union of the number of image pairs with homologue points, image overlap area, and the number of common image neighbors. To improve the image expansion efficiency while ensuring the connectivity between subclusters as much as possible, an intercluster balanced image expansion algorithm guided by the completeness constraint is proposed. (2) Local sparse reconstruction of subclusters. The robust and accurate global GlobalACSfM technique is used to complete the local sparse reconstruction of each subcluster, and this stage can be computed in parallel under the cluster architecture. (3) Multilevel subcluster merging considering subcluster connectivity. To avoid the large error accumulation problem caused by the long merging path of subclusters, a multilevel merging rule that considers subcluster connectivity is proposed. To ensure the accuracy of the merged subclusters as much as possible, considering the global perspective, the optimal camera internal parameter in the two subclusters is selected as the camera internal parameters of the cluster block. The flowchart of this method is shown in Figure 3.

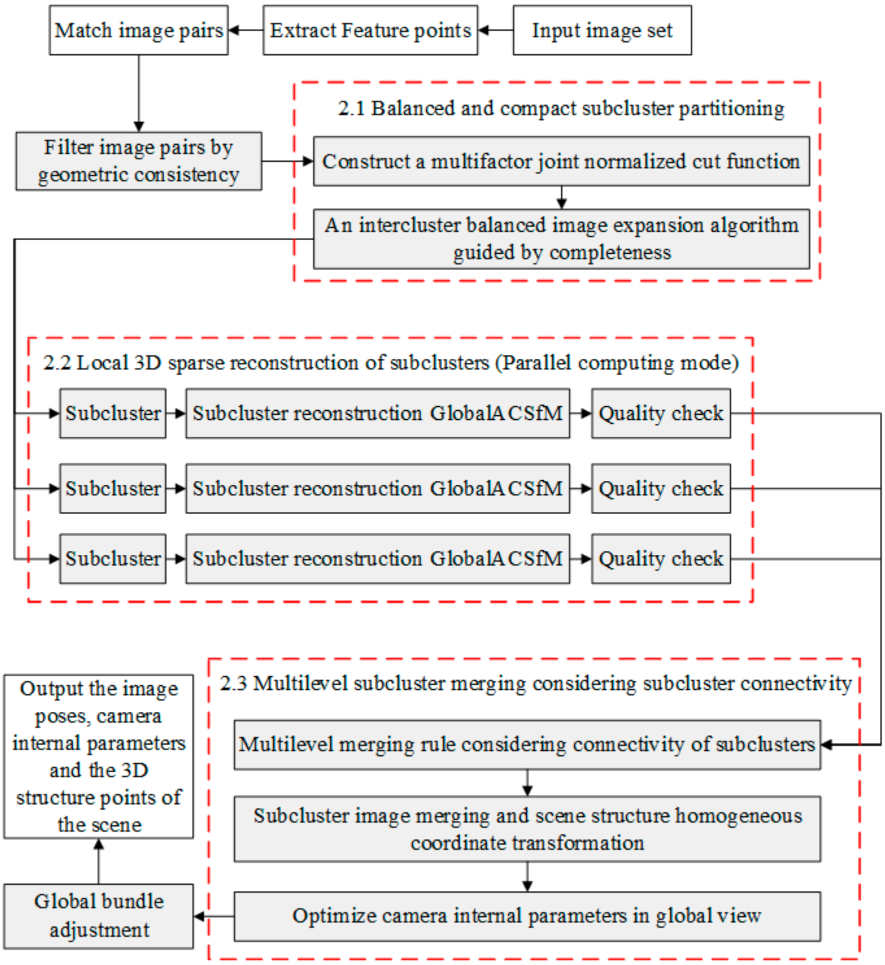

Match image pairs
Extract Feature points
Input image set
2.1 Balanced and compact subcluster partitioning
Filter image pairs by geometric consistency
Construct a multifactor joint normalized cut function
An intercluster balanced image expansion algorithm guided by completeness
2.2 Local 3D sparse reconstruction of subclusters (Parallel computing mode)
Subcluster
Subcluster reconstruction GlobalACSfM
Quality check
Subcluster
Subcluster reconstruction GlobalACSfM
Quality check
Subcluster
Subcluster reconstruction GlobalACSfM
Quality check
Output the image poses, camera internal parameters and the 3D structure points of the scene
2.3 Multilevel subcluster merging considering subcluster connectivity
Multilevel merging rule considering connectivity of subclusters
Subcluster image merging and scene structure homogeneous coordinate transformation
Global bundle adjustment
Optimize camera internal parameters in global view

Figure 3. Flowchart of this method.

3.1. Scene Partitioning

Scene partitioning is the first stage of the hybrid SfM method, whose main purpose is to divide the original scene into several related small scenes (hereinafter referred to as subclusters). Scene partitioning includes two steps: image clustering and subcluster expansion. Image clustering divides the original scene into N nonoverlapping subclusters by a clustering algorithm. Subcluster expansion adds repeated images to subclusters so that there are enough overlapping images between adjacent subclusters and enhances the connectivity between subclusters. For the problems of weak image connectivity and even spatial discontinuity in the image clustering results and low expansion efficiency and uneven connectivity in the image expansion in the existing methods, effective solutions are proposed in this paper.

3.1.1. Image Clustering Algorithm Based on Multifactor Joint Measure

Image clustering divides the scene into several nonoverlapping subclusters. Similar to the existing method [20], the image clustering algorithm in this paper completes image division by constructing the graph, calculating the edge weights in the graph, and using normalized cut [37]. The detailed steps are as follows:

Step 1: Build the graph. The clustering process in this paper is carried out on the graph. The image is taken as the node in the graph, and the connection between the images is taken as the edge in the graph to build the camera graph $G = \{V, E\}$, where the node $V_i \in V$ represents the camera $C_i \in C$, $E_{ij} \in E$ and the edge weight $w(E_{ij})$ represents the connection relation between the image pair C_i and C_j. Subcluster clustering divides all the cameras represented by graph $G = \{V, E\}$ into a group of N subclusters represented by graph $\{G_k | G_k = \{V_k, E_k\}\}$, and the subclusters do not overlap each other.

Step 2: Construction of edge weights. The edge weight $w(E_{ij})$ in graph $G = \{V, E\}$ is constructed to measure the connection relationship between two nodes. The existing method takes the number of points with homologue points of the image pair as the weight, and the weight function of the number of points with homologue points after normalization is:

$$w_{p_{ij}} = \frac{p_{num}}{\min(feat_i, feat_j)}, \tag{1}$$

where $feat_i$, $feat_j$ is the number of feature points of each image pair i, j and p_{num} is the number of points with the homologue points of the image pair. It is generally believed that the larger $w_{p_{ij}}$ is, the stronger the correlation between image pairs will be. If only the number of points with homologue points is used to calculate the edge weight, weakly connected image pairs with a long spatial distance may be divided into the same subcluster, while image pairs with a short spatial distance but a relatively small number of points with homologue points may be divided into different subclusters, resulting in the spatial distribution of images within the subcluster being loose or even fragmented (as shown in blue $block_1$ in Figure 2). Furthermore, it may lead to the rupture of the local reconstructed free network of the subcluster.

To solve the above problems, based on Formula (1), this paper introduces the overlapping area of image pairs, scene correlation degree information, and the number of points with homologue points to construct a new weight function $w(E_{ij})$. Among them, the overlapping area of the image pair is the representation of the same ground area captured in the two images in the image space. If the overlap area of an image pair is small, it is considered that the image pair is more likely to be weakly connected. The common neighbor number of images represents the number of images with the same ID in the neighbor set of two images in an image pair. The greater the number of common neighbors, the more likely the image pair is to be strongly connected. The number of points with homologue points in an image pair represents the number of matched feature points in two images, and the more points there are with homologue points, the stronger the correlation between the two images.

(1) The normalized overlap area of the image pair:

$$w_{a_{ij}} = \max\left(\frac{p_{area_i}}{area_i}, \frac{p_{area_j}}{area_j}\right), \tag{2}$$

where p_{area_i}, p_{area_j} represents the number of pixels occupied by the external polygons in the distribution range of the points with homologue points on their respective images. Due to the difference in the angle of view, the distribution range of the homologue points on the left and right visual images on the same image pair may be slightly different. For the convenience of calculation, the maximum value is adopted in this paper. $area_i$, $area_j$ represent the respective image area of images i, j. The larger the overlap area factor of the image pair, the more likely the image pair is to be strongly connected.

(2) Normalized scene correlation factor:

$$w_{m_{ij}} = \frac{com_{num}}{\max(match_{num})}, \tag{3}$$

where com_{num} is the number of common associated images of the image pair, that is, the number of common neighbor images contained in all image pairs composed of two images i, j, and $match_{num}$ is the maximum number of associated images shared by all image pairs in the scene. The more associated images the left and right visual images of the image pair have in the whole scene, the stronger the connection relation of the image pair is.

In this paper, the above three indicators are combined and regularized to construct a new weight function as follows:

$$w(E_{ij}) = \alpha \times w_{a_{ij}} + \beta \times w_{p_{ij}} + \delta \times w_{m_{ij}}, \tag{4}$$

where α, β, δ are the weights of the above three different factors (empirical values of 50, 30, and 20 are selected in this paper).

Step 3: Image clustering. First, the subcluster image scale is determined to determine the number of subclusters to be divided. The subcluster image scale should be moderate. Considering the limitations of computer memory and computing efficiency, the subcluster image scale should not be too large. Then, a normalized cut is used to divide the established graph. Finally, to facilitate subcluster expansion in the next stage, the image pairs that were disconnected in the process of subcluster clustering are collected and arranged in descending order by weight to construct the fracture relationship table E_{lost}:

$$E_{lost} = \left\{ \varepsilon_{k_1 k_2, i} \middle| k_1, k_2 \in [\mathrm{K}], i \in \left| \varepsilon_{k_1 k_2} \right| \right\}, \tag{5}$$

where K is the number of subclusters, k_1, k_2 represents two subclusters in the cluster pair, and $\varepsilon_{k_1 k_2}$ represents the matching pair that is broken during the partition.

3.1.2. Subcluster Expansion Algorithm Considering Partition Connectivity

After image clustering, image expansion of subclusters is performed to ensure the connectivity between neighboring subclusters and facilitate the merging of local sparse reconstruction results of subsequent subclusters. Based on the existing methods [20], an intercluster balanced image expansion algorithm guided by completeness is proposed to improve the image expansion efficiency and to consider the balanced connectivity of the expanded images of neighboring subclusters. The specific steps are as follows:

Step 1: Calculate the ratio to be expanded. Calculate the ratio of the target subcluster to the number of neighborhood clusters to be expanded:

$$r_{k_i k_j} = \frac{E_{lost|k_i k_j}}{sum\left(E_{lost|k_i}\right)}, \tag{6}$$

where $E_{lost|k_i}$ represents the total number of image pairs interrupted by subcluster i during the partition process, and $E_{lost|k_i k_j}$ represents the number of images interrupted by subcluster i and its neighborhood cluster j during the partition process.

Step 2: Calculate the number of images to be expanded. Combined with the completeness rate η_i and the allocation ratio calculated by Formula (6), the extended image number of the target subcluster i and its neighborhood subcluster j is calculated as follows:

$$N_{k_ik_j} = r_{k_ik_j} \times \left(\eta_i \times cluster_{k_i} - expand_{k_i}\right), \tag{7}$$

$$\eta_i = \frac{\sum_{j \neq i} |G_i \cap G_j|}{G_i}, \tag{8}$$

$$expand_{k_i} = \begin{cases} 0 & if\ i\ does\ not\ expand \\ \sum\limits_{i \neq j} N_{k_ik_j} & if\ i\ has\ done\ the\ expansion \end{cases}, \tag{9}$$

where $cluster_{k_i}$ represents the number of the current images of subcluster i, $expand_{k_i}$ represents the number of expanded images of current subcluster i, and G_i, G_j represents two different subclusters.

Step 3: Image addition. $N_{k_ik_j}$, the number to be expanded, is calculated by Formula (7), and the one with the smaller scale is selected among the cluster pairs to be expanded, restoring the first $N_{k_ik_j}$ matching pairs with the highest weight in E_{lost} to balance the size of the subclusters.

After the above two stages of image clustering and subcluster expansion, the images contained in the scene are divided into multiple subclusters of comparable size, with more compact image connectivity within the subclusters and reasonable and balanced connectivity established between the subclusters and neighboring subclusters by repeating images.

3.2. Local 3D Sparse Reconstruction of Subclusters

After subcluster partitioning, all subclusters need to be reconstructed individually using the local SfM approach. Unlike the hybrid hierarchical SfM method, the method in this paper belongs to the hybrid planar SfM method, which partitions large-scale scene data of tens of thousands or even hundreds of thousands of images into several or even tens of subscenes in a planar subcluster division so that either incremental SfM or global SfM methods can be chosen for the reconstruction of subscenes.

Unlike the previous hybrid SfM [20,31–33] that used incremental SfM for local sparse reconstruction, the local sparse reconstruction of subclusters is performed using global SfM. Compared with incremental SfM, the global SfM-based local sparse reconstruction strategy has the following advantages:

1. Global SfM does not require frequent global bundle adjustment, so it is more efficient in the local sparse reconstruction of subclusters.
2. There is no need to consider the risk of drift caused by incremental SfM local reconstruction, so the subcluster partitioning size in this paper can be as large as possible. The subcluster size can be set to hundreds or even thousands of images to reduce image number redundancy between subclusters and subcluster merging times.
3. The local SfM of subclusters adopts the global SfM, which has less possibility of drift and error accumulation. Therefore, only one global bundle adjustment is required during the subcluster merging process, which improves the efficiency of subcluster merging.

The GlobalACSfM method [35] used in this paper has been integrated into the open source library OpenMVG, which has been experimentally verified to be more efficient than incremental SfM is. The robustness of the method is better due to the use of an adaptive inverse estimation model for the initial pose estimation of the camera, as well as the use of adaptive thresholding to better reject noise and delineate interior points. The GlobalACSfM source code address is https://github.com/openMVG/openMVG, accessed on 24 December 2022.

Due to uncertainties such as image matching accuracy and flight sensor calibration characteristics, it is not guaranteed that all images in a subcluster will be successfully registered. If a subcluster has more image registration failures in local sparse reconstruction, it may cause the connectivity of that subcluster with other subclusters to decrease or it may even become an isolated subcluster, making the subcluster merge fail. Therefore, in this paper, a quality assessment of the sparse reconstruction results is performed after local sparse reconstruction. If the image registration of the subcluster sparse reconstruction result is less than 95% of the image size of the subcluster (the value is more reasonable after experimental verification), the local sparse reconstruction of the subcluster is judged to have failed. The set of camera internal parameters and aberration parameters with the smallest reprojection error in its neighboring subclusters is selected as the camera internal parameters and aberration parameters of the subcluster, and local sparse reconstruction is performed again to improve the robustness of this method.

3.3. Multilevel Subcluster Merging Considering Subcluster Connectivity

Subcluster merging is the final stage of hybrid SfM. In particular, when dealing with large image datasets, the existing methods are more prone to accumulate large merging errors or even merging failures due to the insufficient number of duplicate images or common structure points between neighboring subclusters or the long merging paths of these subclusters during the merging process. To solve this problem, the optimal merging path based on the minimum height tree (*MHT*) to select the benchmark cluster and other subclusters in the existing method [20] is no longer used, and the multilevel merging rule that considers the subcluster connectivity is proposed, which no longer selects the benchmark cluster but instead a decentralized multilevel merging process, and the merging process considers the global perspective. The specific steps are as follows:

Step 1: Build the set of subclusters to be merged. The local SfM reconstruction results of subclusters in Section 2.1 are judged by a simple quality check. If the number of registered images of a subcluster is small or the reprojection error is large, the local SfM reconstruction of this subcluster is considered to have failed and it cannot be included in the set of subclusters to be merged.

Step 2: Centerless method of selecting the cluster pairs to be merged. First, the subcluster with the largest number of neighboring clusters (or the best connectivity) is selected from the set of subclusters to be merged as the subcluster to be merged. Then, all subclusters paired with this subcluster are counted, the number of common images of the cluster pair is prioritized as the merging weight, and the neighboring subclusters with the highest priority are selected to form the cluster pair to be merged with the subcluster.

Step 3: Perform the merge operation. Since neighboring subclusters in the cluster pair to be merged may have been merged with other subclusters, different merging operations are applied depending on the merging of neighboring subclusters.

- The case where neither of the two subclusters has been merged. First, the feature points of the common image of the current cluster pair are collected, the 3D structure points corresponding to the feature points are used to perform alignment transformation (translation, rotation, scaling), and the camera internal parameters, image, and its external parameters and 3D structure points of the subcluster are merged into a new cluster block. Then, to solve the camera internal parameter consistency problem, the common camera internal parameter of two subclusters is combined with the 3D structure points of the cluster block for the rear intersection, the reprojection error is calculated, and the group with a smaller reprojection error is taken as the camera internal parameter of the cluster block.
- The case where the neighboring subclusters have been merged. First, the cluster blocks to be merged with subclusters are searched and determined. Then, the feature points of the common images of the current subcluster and the cluster block are collected, the corresponding 3D structure points of the feature points are used to perform alignment transformation (translation, rotation, and scaling), and the camera internal parameters, image, and its external parameters and 3D structure points of the subcluster are merged into the cluster block. Finally, to solve the camera internal

parameter consistency problem, the common camera internal parameter of subclusters and cluster blocks is combined with the 3D structure points of the cluster block for rear intersection, the reprojection error is calculated, and the group with smaller reprojection error is taken as the camera internal parameter of the cluster block.

Step 4: The subclusters that were merged in step 3 are removed from the set of subclusters to be merged, and steps 2–3 above are performed cyclically until all subclusters are merged. Then, the cluster blocks continue to merge until a complete scene is formed, and the merging process is the same as the above steps.

After the subcluster merging process, to ensure the accuracy of the sparse reconstruction results of the scene, the method in this paper adopts the idea of the literature [30] to perform another global bundle adjustment for the sparse reconstruction results of the merged scene, which is used to adjust the image poses, camera internal parameters, and the 3D structure points of the scene. In this paper, the work is conducted using Ceres Solvers [38], a library for nonlinear optimization with excellent performance, which is commonly used to optimize image poses (rotation matrix and translation matrix), camera internal parameters (including focal length, image principal point, radial aberration, tangential aberration), and 3D spatial point coordinates.

4. Experiment and Results

4.1. Experimental Data and Environment

The method proposed in this paper is embedded into the IMS software, which is a reality modelling software that was independently developed by the authors at the Chinese Academy of Surveying and Mapping. Multiple publicly available datasets of different sizes [39,40] and field-collected tilt-photography datasets were used for experimental verification. The detailed parameters of the experimental data are shown in Tables 1 and 2. The subcluster local SfM in the second stage of the proposed method can be computed in parallel using the cluster architecture of IMS software, so the experimental operating environment is a computing cluster consisting of one master node and five subnodes, where the master node is a workstation with Windows 10 64-bit operating system, an Intel Core i7-11700X CPU with a dominant frequency of 2.50 GHz, and 64 GB of memory; the subnodes are ordinary desktops with the Windows 10 64-bit operating system, an Intel Core Gold6132 CPU with a dominant frequency of 2.60 GHz, and 64 GB of memory. In addition, SIFT [41] provided by vlfeat is used to extract feature points in this paper.

Table 1. Public dataset information.

Dataset	Number of Images	Image Size
Graham-Hall	100	5616 × 3744
South-Building	128	3072 × 2304
Person-Hall	330	5616 × 3744
Echillais-Church	353	5616 × 3744
Gerrard-Hall	1273	5616 × 3744

4.2. Experimental Result of Subcluster Partitioning

To verify the effectiveness of subcluster partitioning in this paper, experimental validation and a description of experimental results are conducted from three aspects: image pair compactness of image clustering, image expansion connectivity, and image expansion efficiency.

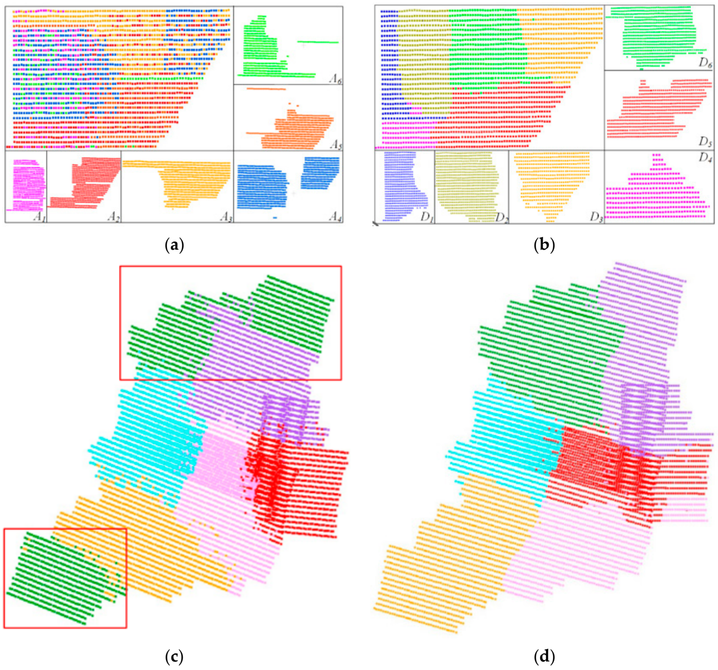
4.2.1. Image Clustering Compactness Verification

Three sets of oblique photography datasets Area 6 and Area 7 with rich and complex feature types (containing dense building areas, roads, vegetation, waters, etc.) were selected for comparison experiments between the proposed method and the traditional method [20] in the same environment.

Table 2. Aerial oblique photographic dataset information.

Dataset	Number of Images	GSD (cm)	Area (km^2)	Image Size	
				Ortho	Oblique
Area 1	162	4.08	0.075	5472 × 3648	5472 × 3648
Area 2	637	3.37	0.30	5472 × 3648	5472 × 3648
Area 3	2776	2.26	1.24	6000 × 4000	6000 × 4000
Area 4	5265	1.30	0.32	6000 × 4000	6000 × 4000
Area 5	8221	3.37	6.42	5472 × 3648	5472 × 3648
Area 6	11,795	6.50	40.30	11,674 × 7514	8900 × 6650
Area 7	34,364	3.00	10.62	7952 × 5304	7952 × 5304
Area 8	39,040	2.00	2.30	6000 × 4000	6000 × 4000
Area 9	48,335	1.20	2.93	6000 × 4000	6000 × 4000
Area 10	10,500	1.20	0.58	6000 × 4000	6000 × 4000

Area 6 data have the characteristics of higher acquisition height and larger ground resolution (GSD), and it is easier to have a large inclination angle between image pairs. With the same subcluster partition scale, the traditional method tends to partition the less connected images into a subcluster, and the partitioning result is incompact or even spatially discontinuous. As shown in the overall situation of subcluster partitioning (the upper-left corner of Figure 4a), there is spatial overlap in the positions of the images contained in adjacent subclusters; both subclusters A_1 and A_2 show the loose phenomenon of consecutive lost images in the airstrip; the images contained in three subclusters A_4, A_5, and A_6 show two to three small, aggregated groups, all of which show a spatial discontinuity. As seen in the overall situation of the subcluster partition in the upper-left corner and the image position distribution of $D_1 \sim D_6$ subclusters in Figure 4b, none of the results of the proposed method have image loosening or spatial discontinuity, and the subcluster partitioning results are more compact. This is because the weight calculation method of the multifactor union of the normalized cut function proposed in this paper takes the number of homologue points of the image pair, the overlap area of the image, and the number of common neighbors of the image pair as the factors to measure the connectivity of the image pair from different perspectives.



(a) (b)

(c) (d)

Figure 4. Comparison of image clustering results. The point set in the figure represents the position coordinates of the oblique image, and the same subcluster is represented by the same color. The left

side is the result of the traditional method, and the right side is the result of the proposed method. (**a**,**b**) The dataset is Area 6, and the subcluster partition size is 1800, where the top left is the overall partition results, and A_1~A_6 are the results of the six subclusters partitioned. (**c**,**d**) The dataset is Area 7, and the subcluster partition size is 5000.

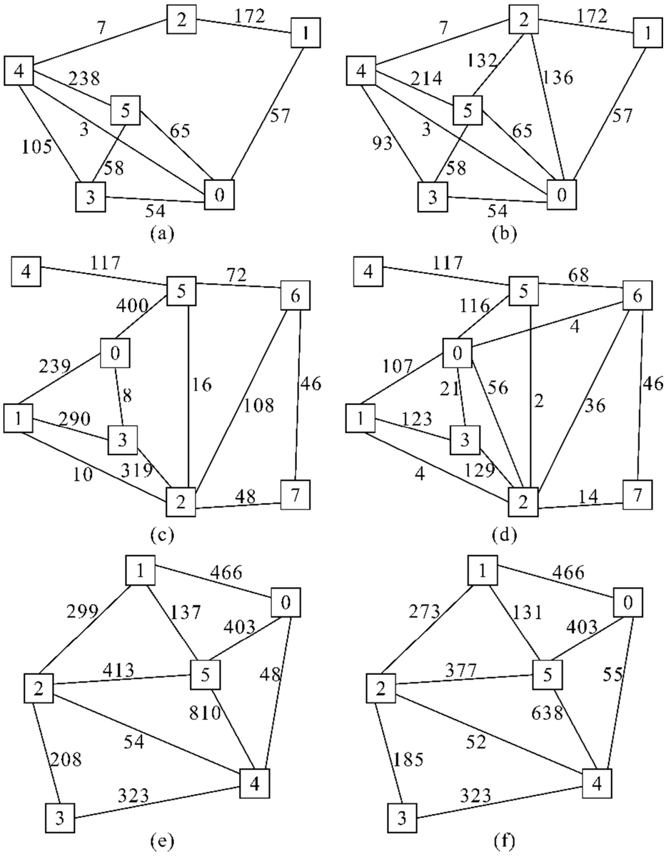
4.2.2. Subcluster Expansion Connectivity and Efficiency Verification

To verify the effectiveness of the proposed method in terms of subcluster expansion connectivity and efficiency, a comparison experiment was designed between the proposed method and the traditional method [20] in the same environment.

(1) Experimental result of subcluster expansion connectivity.

Using the experimental data Area 5, Area 6, and Area 7 and based on the scene partitioning results of the image clustering method in this paper, repeated images were added between neighboring subclusters using the subcluster expansion algorithm in the traditional method and the method proposed in this paper to ensure the connectivity between neighboring subclusters and facilitate the merging of subcluster reconstruction results. The subcluster completeness rate $\eta = 0.5$ was set in the experiment, and the number of neighboring subclusters with connectivity (presence of repeated images) with subclusters and the number of repeated images between them were counted.

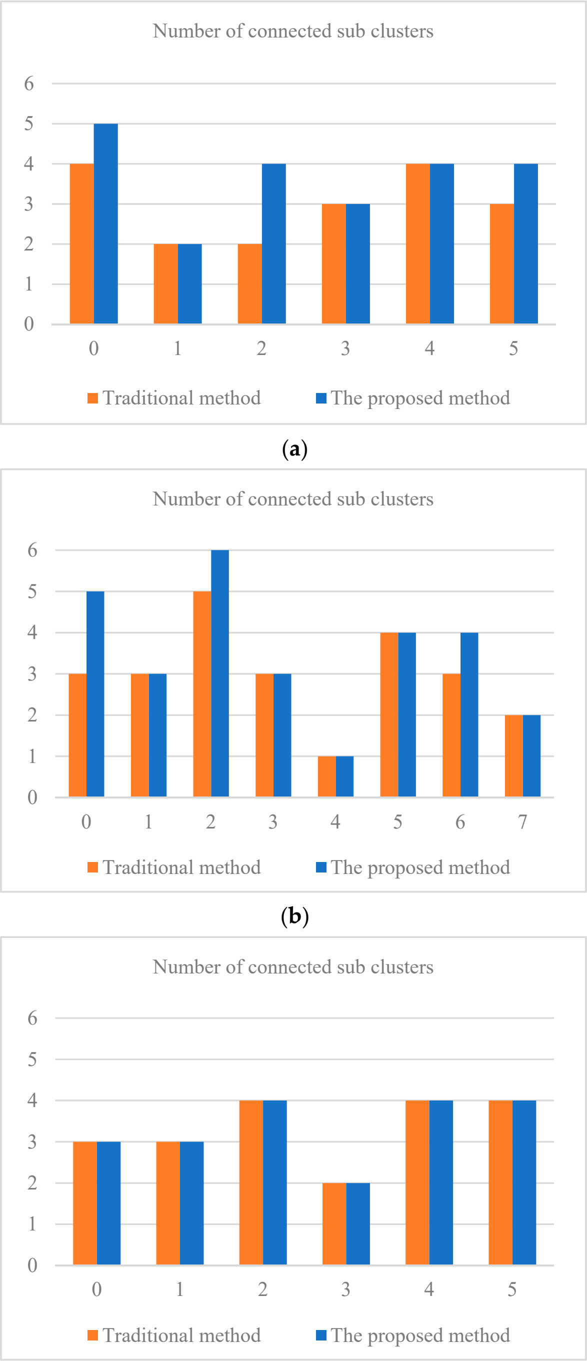
In Figure 5, the number of edges between nodes in (b) and (d) are both more than in (a) and (c), and the number of edges between nodes in (f) and (e) are equal. The histogram in Figure 6 is more intuitive, which indicates that the proposed method is better than the traditional method in terms of image expansion connectivity.



(a) (b) (c) (d) (e) (f)

Figure 5. Results of image expansion in subclusters (represented as undirected graphs). The left column shows the expansion results of the traditional method, and the right column shows the expansion results of the proposed method. The box represents a node in the undirected graph, and

the number in the box represents the subcluster ID; the edge represents the connectivity between nodes, and the number attached to the edge represents the number of repeated images added between nodes due to the expansion. (**a**,**b**) The dataset is Area 6, with a subcluster partition size of 1800; (**c**,**d**) the dataset is Area 5, with a subcluster division size of 1000; (**e**,**f**) the dataset is Area 7, with a subcluster division size of 5000.



(**a**)

(**b**)

(**c**)

Figure 6. Connectivity of subclusters after expansion, where the horizontal coordinate represents the subcluster ID and the vertical coordinate indicates the number of subclusters with connectivity to the subclusters. The orange color represents the results of the traditional method, and the blue color represents the results of the proposed method. (**a**) The dataset is Area 6, with a subcluster partition size of 1800; (**b**) the dataset is Area 5, with a subcluster division size of 1000; (**c**) the dataset is Area 7, with a subcluster division size of 5000.

(2) Experimental result of expansion efficiency

To verify the superiority of the efficiency of the proposed expansion algorithm, five sets of data of different sizes, Area 3, Area 6, Area 7, Area 8, and Area 9, were selected to measure the time consumed in the expansion stage using the proposed method and the traditional method [20], as shown in Table 3.

Table 3. Image expansion efficiency statistics.

Dataset	Number of Images	Subcluster Size	Traditional Method (ms)	The Proposed Method (ms)
Area 3	2776	500	2713	43
Area 6	11,795	2000	48,589	67
Area 7	34,364	2000	938,433	439
Area 8	39,040	2000	772,184	355
Area 9	48,335	2000	5,678,460	1340

As shown in Table 3, the efficiency of the proposed algorithm is faster than that of the traditional method on datasets of different sizes, and the advantage of the proposed algorithm becomes increasingly significant as the size of the dataset (the number of images) increases. For example, the number of images in Area 3 is 2776, and the time consumed by the proposed method is 1.58% that of the traditional method; the number of images in Area 9 is 48,335, and the time consumed by the proposed method is only 0.024% that of the traditional method.

For example, in Area 3 of Table 3, the number of images is 2776, and the time consumed by this method is 1.58% that of the traditional method; in Area 9, the number of images is 48,335, and the time consumed by this method is only 0.024% that of the traditional method. This is mainly because the image expansion in this paper adopts the preassignment strategy, which avoids the frequent adjustment of the subclusters to be added and the number of assigned images in the expansion process.

4.3. Experimental Result of Subcluster Merging Robustness

In terms of subcluster merging, the traditional method suffers from a low merging success rate. A multilevel subcluster merging rule that considers subcluster connectivity is proposed. To verify the robustness of the proposed method in subcluster merging, two public datasets, Echillais-Chuich and Granham-Hall, and a set of oblique photography data, Area 3, were used as experimental data. To ensure the rigor of the experiments, both the subcluster partitioning and subcluster local SfM steps before subcluster merging were performed using the proposed method. Subcluster merging was performed using the traditional GraphSfM method [20] and the merging rules proposed in this method. In addition, for a fuller comparative analysis, we added the well-known incremental SfM method Colmap [39], the global SfM method OpenMVG [35], and the hybrid SfM method 3Df [27] as control groups. The experimental results are shown in Figure 7 and Table 4.

Combining Figure 7 and Table 4, for the Echillais-Church data, all five methods could be reconstructed successfully, and the reprojection errors were all less than 0.5 pixels. However, for the Graham-Hall data, both the 3Df method and the GraphSfM method failed to reconstruct, the number of successfully registered images for the GraphSfM method was too small (Table 4), and the merging results were confusing (Figure 7). This is mainly due to the poor connectivity during subcluster merging, resulting in too few repeated images and common structure points among subclusters. Therefore, the exact chi-square coordinate transformation parameters could not be calculated.

4.4. Experimental Result of Accuracy and Time Comparative

To verify the overall performance of the proposed method, the reconstruction accuracy and efficiency of the proposed method, the most advanced incremental SfM method Colmap [39], the global SfM method OpenMVG [35], the hybrid SfM method 3Df [27], and

GraphSfM [20] were compared and analyzed using publicly available datasets and oblique photography datasets. The experiments were conducted on a single computing node, and the experimental results are shown in Tables 5–8.

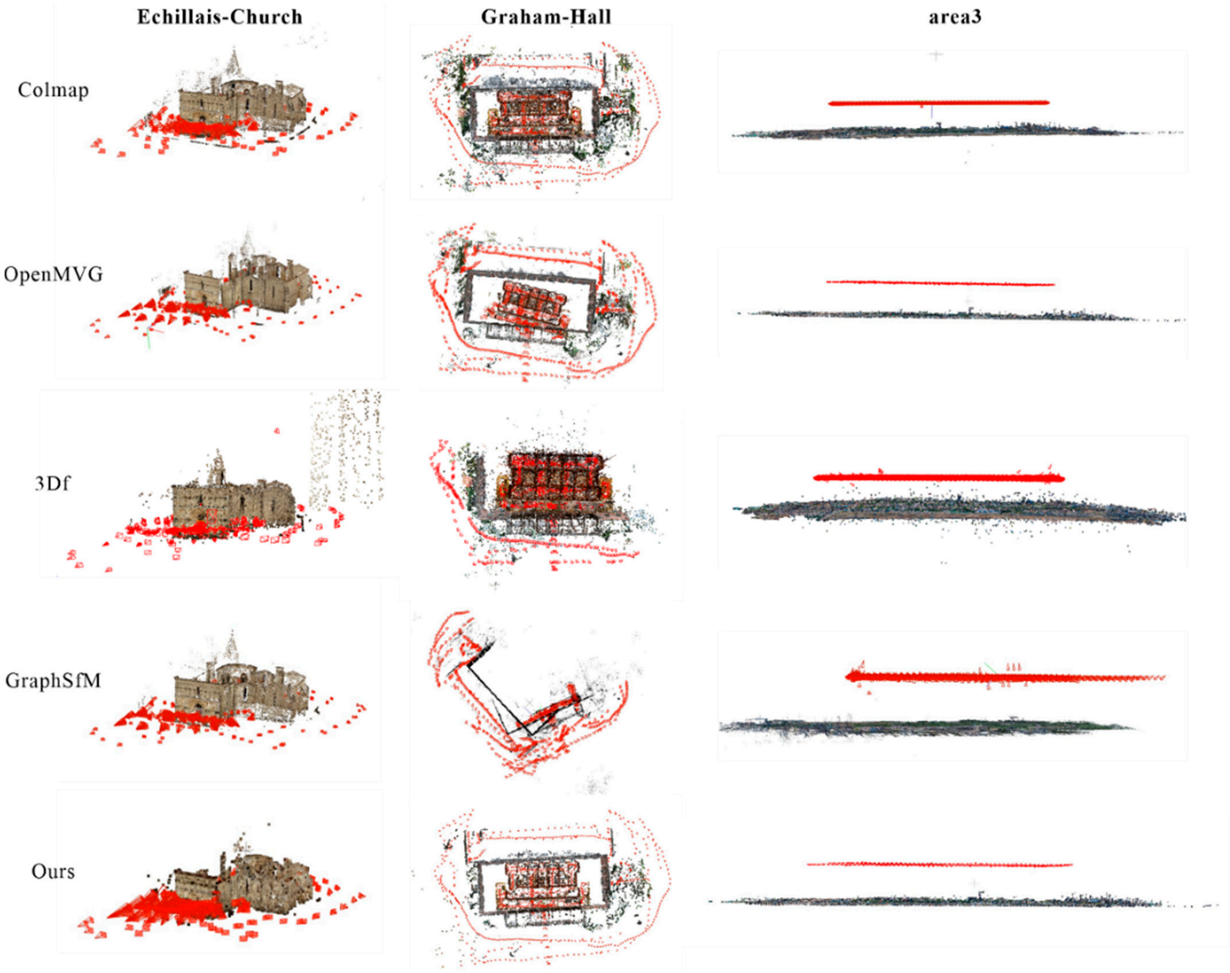


Figure 7. Subcluster merging results.

Table 4. Subcluster merge precision statistics. (N_c: number of registered images; N_p: number of structure points; *Err*: reprojection error (MSE), in pixels).

		Echillais-Church	**Graham-Hall**	**Area 3**
Colmap	N_c	288	1260	2688
	N_p	232,244	452,965	**1,588,456**
	Err	0.31	0.57	1.34
OpenMVG	N_c	342	1213	2694
	N_p	**634,240**	981,723	1,467,779
	Err	0.48	0.52	0.72
3Df	N_c	**352**	886	2713
	N_p	102,129	149,796	954,076
	Err	0.31	0.36	0.35
GraphSfM	N_c	**352**	855	1627
	N_p	181,424	**1,189,188**	1,308,133
	Err	0.39	0.36	0.45
Ours	N_c	**352**	**1265**	**2697**
	N_p	22,576	49,893	191,845
	Err	**0.29**	**0.31**	**0.33**

Table 5. Public dataset accuracy statistics. (N_c: number of registered images; N_p: number of structure points; Err: reprojection error (MSE), in pixels; — represents reconstruction failure).

		Gerrard-Hall	**South-Building**	**Person-Hall**	**Echillais-Church**	**Graham-Hall**
Colmap	N_c	**100**	**128**	**330**	288	**1260**
	N_p	**56,426**	86,972	179,168	232,244	**452,965**
	Err	0.48	0.41	0.48	0.31	0.57
OpenMVG	N_c	99	**128**	329	343	—
	N_p	33,280	86,465	**788,726**	**634,240**	—
	Err	0.47	0.30	0.45	0.48	—
3Df	N_c	**100**	**128**	—	—	—
	N_p	20,235	25,065	—	—	—
	Err	0.28	0.31	—	—	—
GraphSfM	N_c	**100**	**128**	328	352	—
	N_p	54,914	57,311	130,432	181,424	—
	Err	0.33	0.27	0.33	0.39	—
Ours	N_c	**100**	**128**	**330**	**353**	1259
	N_p	4811	6759	19,783	22,576	49,893
	Err	**0.24**	**0.27**	**0.28**	**0.29**	**0.31**

Table 6. Oblique photography dataset accuracy statistics. (N_c: number of registered images; N_p: number of structure points; Err: reprojection error (MSE), in pixels; — represents reconstruction failure; *** represents the reconstruction time is more than fifteen days).

		Area 1	**Area 2**	**Area 3**	**Area 4**	**Area 10**
Colmap	N_c	**162**	**637**	2688	***	***
	N_p	**72,655**	448,023	**1,588,456**	***	***
	Err	0.49	0.53	0.94	***	***
OpenMVG	N_c	**162**	**637**	2694	**5265**	**10,499**
	N_p	68,409	**448,206**	1,467,779	**2,270,331**	4,684,135
	Err	0.70	0.75	0.72	0.69	0.74
3Df	N_c	**162**	**637**	**2713**	5250	10,497
	N_p	54,801	220,078	954,076	2,219,910	2,885,045
	Err	0.26	0.35	0.35	0.34	0.73
GraphSfM	N_c	**162**	**637**	—	—	10,493
	N_p	47,634	302,216	—	—	**5,744,895**
	Err	0.50	0.54	—	—	0.45
Ours	N_c	**162**	**637**	**2713**	**5265**	**10,499**
	N_p	14,179	40,376	191,845	440,580	882,847
	Err	**0.22**	**0.30**	**0.33**	**0.33**	**0.34**

As shown in Tables 5 and 6, the proposed method could successfully perform sparse reconstruction on both the public dataset and the oblique photography dataset, and the main accuracy indexes, such as the number of registered images, the number of recovered structure points, and the reprojection error, all performed well and had high robustness. The GraphSfM method failed in sparse reconstruction on both types of datasets, such as the Graham-Hall data in Table 5 and Area 3 and Area 4 data in Table 6. In addition, on the public dataset and the oblique photographic dataset, the proposed method in this paper recovered only 3% to 20% of the structure points compared to the other methods.

The Colmap method performed better on the public dataset, but on the oblique photography dataset, the reconstruction time was exceedingly long when the number of images exceeded 5000, as in Table 6 for data Area 4 and Area 6, where the reconstruction time is indicated as ***. The OpenMVG method also failed when performing sparse reconstruction on the public dataset Graham-Hall. 3Df, a typical representative of open-source hybrid SfM methods, also performed poorly in terms of robustness.

Table 7. Public dataset efficiency statistics. (T_p: local reconstruction time; T_{Meg}: subcluster merging time; T_{BA}: global bundle adjustment time; $T_{\sum}$: total time. - represents no merging operation, such as with 3Df where scene partitioning is not performed with less than 1000 images, and time is no longer counted; — represents failure, time unit: seconds).

		Gerrard-Hall	**South-Building**	**Person-Hall**	**Echillais-Church**	**Graham-Hall**
Colmap	$T_{\sum}$	183	298	981	2257	19,384
OpenMVG	T_p	9	30	1366	241	—
	T_{BA}	25	97	3878	867	—
	$T_{\sum}$	34	127	5244	1108	—
3Df	T_p	12	19	—	—	—
	T_{Meg}	-	-	—	—	—
	T_{BA}	1	1	—	—	—
	$T_{\sum}$	**13**	**20**	—	—	—
GraphSfM	T_p	271	214	729	680	—
	T_{BA}	4	3	8	1	—
	$T_{\sum}$	275	217	**747**	**741**	—
Ours	T_p	101	181	3220	1235	4814
	T_{Meg}	1	1	48	59	34
	T_{BA}	2	14	19	40	630
	$T_{\sum}$	103	196	3287	1334	**5478**

Table 8. Oblique photography dataset efficiency statistics. (T_p: local reconstruction time; T_{Meg}: subcluster merging time; T_{BA}: global bundle adjustment time; $T_{\sum}$: total time. - represents no merging operation, such as with 3Df where scene partitioning is not performed with less than 1000 images, and time is no longer counted; — represents failure, time unit: seconds; *** represents the reconstruction time is more than fifteen days).

		Area 1	**Area 2**	**Area 3**	**Area 4**	**Area 10**
Colmap	$T_{\sum}$	341	5451	7,563,501	***	***
OpenMVG	T_p	20	343	1466	7639	41,367
	T_{BA}	30	896	2446	18,083	44,627
	$T_{\sum}$	50	1239	3912	25,722	85,994
3Df	T_p	23	402	1112	1809	22,143
	T_{Meg}	-	-	106	2136	2694
	T_{BA}	1	134	6	548	1912
	$T_{\sum}$	**24**	**536**	**1224**	4493	27,531
GraphSfM	T_p	179	2372	—	—	180,570
	T_{BA}	5	73	—	—	17,320
	$T_{\sum}$	184	2445	—	—	197,890
Ours	T_p	88	918	1408	2092	20,666
	T_{Meg}	1	3	90	34	494
	T_{BA}	4	23	227	1245	623
	$T_{\sum}$	93	944	1725	**3371**	**21,783**

In terms of efficiency, it can be seen in Tables 7 and 8 that (1) the proposed method in this paper did not perform the best in terms of the time required for sparse reconstruction (the 3Df method is the fastest) for both the open dataset and the skewed photographic dataset, but the time advantage of the proposed method in this paper became increasingly significant as the number of data increased. Compared with GraphSfM, which also uses a planar hybrid SfM framework, the efficiency of this paper's method was higher, with the proposed method consuming 90.3% to 12.0% of GraphSfM's time. (2) Overall, the time of the proposed method was not the shortest (3Df was the fastest) on the public dataset and the oblique photography dataset. However, as the data size increased, the gap between the time of the proposed method and the time of 3Df gradually decreased. For example, from Area 1 with 162 images and Area 4 with 5265 images, the time multiple of the two was reduced from 4 times to about 2 times. (3) When the number of images reached more than 10,000, the proposed method required less time than the 3Df method did, and the time complexity has obvious advantages. For example, for Area 10 with 10,500 images, the time consumption of the proposed method was only 80.6% of that of the 3Df method.

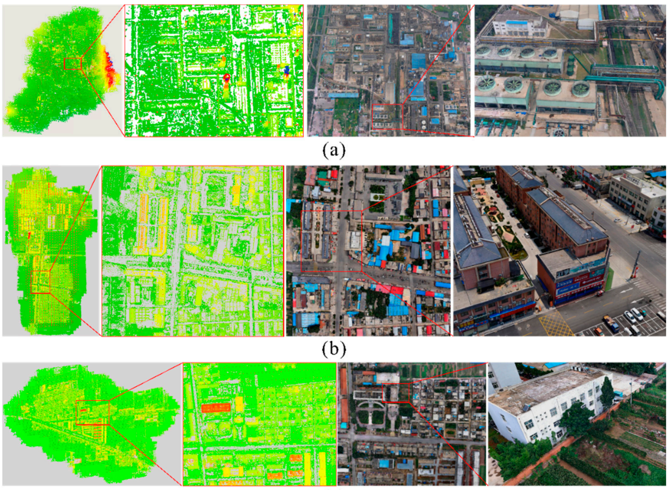
4.5. Large-Scale Aerial Image Dataset

To verify the performance of the proposed method on large-scale datasets, Area 6, Area 7, Area 8, and Area 9 were used as experimental data for statistical accuracy and efficiency of sparse reconstruction. Due to the large scale of the experimental data, the proposed method was integrated into the cluster architecture of IMS software for parallel computation (one master node and five subnodes; the detailed configuration is described in Section 3.1).

Combining Table 9 and Figure 8, under the abovementioned cluster mode, the proposed method could still perform sparse reconstruction with high efficiency while satisfying accuracy, which indicates that the proposed method is also more applicable to large-scale datasets. Taking the dataset Area 7 as an example, the total number of images was 34,435, the number of successfully registered images was 33,961, the reprojection error was only 0.35 pixels, and the reconstruction (the sum of subcluster partitioning, subcluster local reconstruction, and subcluster merging) time was 1038 min. The high quality of the 3D model (clear and realistic contours) can be seen in the local detail image on the rightmost of Figure 8a, which also reflects the high accuracy of the sparse reconstruction.

Table 9. Reconstruction results of the large-scale image dataset. (*N*: number of images; N_c: number of registered images; N_p: number of structure points; *Err*: reprojection error (MSE); *T*: reconstruction time, in minutes).

Data	*N*	N_c	N_p	*Err*	*T*
Area 6	11,795	11,326	890,934	0.33	178
Area 7	34,435	33,961	2,854,957	0.35	628
Area 8	39,040	38,819	3,354,213	0.34	1066
Area 9	48,335	48,146	3,957,959	0.35	1359


(a)

(b)

(c)

Figure 8. Reconstruction results for large-scale ariel datasets. The leftmost is the overall display of the sparse reconstruction results, the middle-left is the local detail display of the sparse reconstruction

results, the middle-right is the 3D model reconstruction results corresponding to the middle-left, and the rightmost is the detail display of the 3D model reconstruction results. (**a**) The result of oblique image dataset Area 7; (**b**) the result of oblique image dataset Area 8; (**c**) the result of oblique image dataset Area 9.

5. Discussion

5.1. Comparative Analysis of Subcluster Partitioning

To verify the effectiveness of subcluster partitioning in this study, experimental comparison and analysis were conducted from three aspects: image pair compactness of image clustering, image expansion connectivity, and image expansion efficiency.

5.1.1. Image Clustering Compactness Analysis

In the first stage of subcluster partitioning, i.e., image clustering, a multifactor joint normalized cut function weight calculation method is proposed to verify that the image clustering results of the proposed method do not contain image pairs with weak connectivity or spatial discontinuity when facing multiview image partitioning with high overlap and large inclination angles.

Figure 4e,f show the subcluster division results of another set of data. Compared with Area 6, the data of Area 7 are different in terms of acquisition height, ground resolution (GSD), route planning, and data scale. For example, the corresponding data of (a) and (b) were collected using tic-tac-toe route planning, whereas for (c) and (d), the corresponding data were collected using five-way flight path planning. As seen from the subcluster partitioning results in Figure 4a–d, the subcluster partitioning results of the traditional methods have different degrees of image loosening or spatial discontinuity, while the partitioning results of the proposed method are more compact and adaptable.

5.1.2. Subcluster Expansion Connectivity and Efficiency Analysis

In the first stage of subcluster division, i.e., image expansion, the traditional method [3] has problems such as poor connectivity between subclusters and low efficiency, and this study makes targeted improvements to address the above two aspects.

(1) Experimental comparison analysis of subcluster expansion connectivity.

In Figure 5, the number of edges between nodes in (b) and (d) are both more than in (a) and (c), and the number of edges between nodes in (f) and (e) are equal. The histogram in Figure 6 is more intuitive, which indicates that the proposed method is better than the traditional method in terms of image expansion connectivity. Comparing the numbers attached to the edges between nodes, as in Figure 5a,b, the proposed method is more balanced than the traditional method in terms of repeated image assignment. This is mainly because the proposed intercluster balanced image expansion algorithm guided by completeness solves the problem of low efficiency of extended image and the connectivity among subclusters in traditional methods [3].

(2) Experimental comparison analysis of expansion efficiency.

Since the repeated images are added cyclically subcluster-by-subcluster and lost edge-by-edge in the expansion of traditional methods, it consumes more time when faced with large-scale expansion. For this reason, the completeness-guided intercluster balanced expansion algorithm is proposed.

As seen in Table 3, the efficiency of the proposed algorithm is faster than that of the traditional method on datasets of different sizes, and the advantage of the proposed algorithm becomes increasingly significant as the size of the dataset (the number of images) increases. This is mainly because the proposed image expansion algorithm adopts a preassignment strategy, which avoids the frequent adjustment of the subclusters to be added and the number of assigned images in the expansion process.

5.2. Comparative Analysis of Subcluster Merging Robustness

In terms of subcluster merging, the traditional method suffers from a low merging success rate. A multilevel subcluster merging rule that considers subcluster connectivity is proposed.

Combining Figure 7 and Table 4, the three groups of data could be successfully reconstructed using the proposed method. Compared with other methods, the number of registered images was basically the same, the number of structure points recovered was less, and the reprojection error was less than 0.5 pixels.

However, for the Graham-Hall data, both the 3Df method and the GraphSfM method failed in reconstruction, the number of successfully registered images for the GraphSfM method was too small (Table 4), and the merging results were confusing (Figure 7). This was mainly due to the poor connectivity during subcluster merging, resulting in too few repeated images and common structure points among subclusters. Therefore, the exact chi-square coordinate transformation parameters could not be calculated.

For the Area 3 data, the GraphSfM method merged scene was also incomplete, and the number of successfully registered images using the GraphSfM method was less (Table 4), and the failure reason is the same as above. In contrast, the subclusters of the three datasets using the proposed method were successfully merged, and are comparable to the advanced incremental Colmap method and the global OpenMVG method in terms of the number of registered images and reprojection error of the scene reconstruction.

5.3. Accuracy and Time Comparative Analysis

To verify the overall performance of the proposed method, the reconstruction accuracy and efficiency are used as two main indicators for comparative analysis.

As shown in Tables 5 and 6, the GraphSfM method failed in sparse reconstruction on both types of datasets, such as the Graham-Hall data in Table 5 and the Area 3 and Area 4 data in Table 6. The main reasons include the following three aspects: (1) the image clustering results of the GraphSfM method may contain image pairs with weak connectivity or spatial discontinuity, resulting in a looser distribution of images in the subclusters; (2) the GraphSfM method does not consider the connectivity balance between subclusters and neighboring subclusters in image expansion; (3) the global optimal merging path selected based on the minimum height tree (MHT) in the GraphSfM method cannot guarantee the sufficient number of repeated images and common structure points required for merging. The proposed method proposes effective solutions to these three problems, so the proposed method can successfully perform sparse reconstruction with excellent accuracy.

In addition, on the public dataset and the oblique photographic dataset, the proposed method recovered fewer structure points than the other methods did, as shown in the ***Np*** values in Tables 5 and 6. The reasons mainly include two aspects: first, in the local 3D sparse subcluster reconstruction stage, GlobalACSfM was used in this study. This method uses an adaptive inverse estimation model to estimate the initial pose of the camera, and can use adaptive thresholds to correctly remove most noise and divide interior points. Therefore, the number of structure points of each subcluster is less than that of the other methods; second, in the subcluster merging stage, this paper proposes a multilevel subcluster merging considering subcluster connectivity. To ensure the accuracy of the 3D sparse reconstruction of the entire scene after merging, if two subclusters meet the merging conditions, the overlapping area of the two subclusters will be triangulated and adjusted when merging, and the 3D structure points of the area will be regenerated as the merged subcluster structure point.

In terms of efficiency, it can be seen from Tables 7 and 8 that, compared with GraphSfM, which also uses a planar hybrid SfM framework, the efficiency of the proposed method is higher. This is mainly because the proposed method adopts a global strategy in the subcluster sparse local reconstruction stage, which makes the number of parameters for a single solution decrease significantly, while GraphSfM adopts an incremental approach, which requires frequent global bundle adjustment. In addition, when the number of images reaches more than 10,000, the proposed method requires less time than the 3Df method does, and the time complexity has obvious advantages. This is mainly because the 3Df method sets a lower limit on the number of images using hybrid SfM; when the number of images is greater than 1000, the 3Df method will enable hybrid SfM, and its hierarchical division-merging strategy leads to a lower time complexity than that of the proposed method.

6. Conclusions

Existing hybrid SfM methods are mature in subcluster local sparse reconstruction, but at least in subcluster partitioning and subcluster merging, there are still problems such as noncompactness (weak connectivity) among images contained in subclusters, low efficiency and failure to consider subcluster connectivity in image expansion, and less robust subcluster merging. A large-scale hybrid SfM method with compact partitioning and a multilevel merging strategy is proposed. First, subcluster partitioning is performed using a multifactor joint normalized cut function and an inter-subcluster balanced image expansion algorithm guided by completeness. Second, the local sparse reconstruction of subclusters is performed using a global GlobalACSfM method in a cluster parallel framework. Finally, a multilevel merging rule is formulated to perform subcluster merging considering subcluster connectivity.

Experimental verification and analysis were conducted using public datasets and oblique photographic datasets, and the conclusions are as follows:

1. Image clustering in the subcluster partitioning stage. The multifactor joint normalized cut function weight calculation method proposed in this paper improves the compactness of subcluster images, especially when facing the partitioning of multiview images with high overlap and large inclination angles. The image clustering results of this method will not contain image pairs with weak connectivity or spatial discontinuity.
2. Image expansion in the subcluster partitioning stage. The completeness-guided image expansion algorithm proposed in this paper enhances the inter-subcluster connectivity as well as the balance, and the expansion efficiency is high, especially when facing large-scale datasets.
3. Subcluster merging stage. A multilevel subcluster merging rule that considers subcluster connectivity is proposed. Compared with the existing state-of-the-art methods, the proposed method can conduct subcluster merging on both public data and oblique photographic datasets and shows superiority.
4. Accuracy and time. The proposed method is compared with four advanced methods, Colmap, OpenMVG, 3Df, and GraphSfM, and the proposed method has the best performance in terms of reconstruction success rate, accuracy, and time.
5. Large-scale aerial image dataset. Under the cluster parallel computing framework, the proposed method performs well in terms of accuracy and time.

The multilevel subcluster merging rules proposed in the subcluster merging stage of this paper can only be adapted to the sequential execution of subcluster merging by a single compute node at present. In the next step, the merging rules can be improved to adapt to the cluster parallel architecture to further improve the efficiency of the proposed method. In addition, the proposed method cannot perform three-dimensional sparse reconstruction of ground images collected by mobile terminals such as mobile phones, because some ground images may have little or no overlap. In subsequent research, a hybrid SfM suitable for ground imagery data will be investigated.

Author Contributions: Conceptualization, Z.L. and H.G.; methodology, Z.L.; software, Z.L. and H.C.; experiment, Z.L., W.Q. and S.Z.; writing—original draft preparation, Z.L. and H.C.; supervision, H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Basal Research Fund of CASM (Grant No. AR2215).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Snavely, N.; Seitz, S.M.; Szeliski, R. Modeling the World from Internet Photo Collections. *Int. J. Comput. Vis.* **2008**, *80*, 189–210. [CrossRef]
2. Agarwal, S.; Furukawa, Y.; Snavely, N.; Simon, I.; Curless, B.; Seitz, S.M.; Szeliski, R. Building Rome in a Day. *Commun. ACM* **2011**, *54*, 105–112. [CrossRef]
3. Carrivick, J.L.; Smith, M.W.; Quincey, D.J. *Structure from Motion in the Geosciences*; John Wiley & Sons: Hoboken, NY, USA, 2016; ISBN 978-1-118-89583-2.
4. Carmigniani, J.; Furht, B.; Anisetti, M.; Ceravolo, P.; Damiani, E.; Ivkovic, M. Augmented Reality Technologies, Systems and Applications. *Multimed. Tools. Appl.* **2011**, *51*, 341–377. [CrossRef]
5. Yang, M.-D.; Chao, C.-F.; Huang, K.-S.; Lu, L.-Y.; Chen, Y.-P. Image-Based 3D Scene Reconstruction and Exploration in Augmented Reality. *Autom. Constr.* **2013**, *33*, 48–60. [CrossRef]
6. Anderson, K.; Westoby, M.J.; James, M.R. Low-Budget Topographic Surveying Comes of Age: Structure from Motion Photogrammetry in Geography and the Geosciences. *Prog. Phys. Geogr. Earth Environ.* **2019**, *43*, 163–173. [CrossRef]
7. Smith, M.W.; Carrivick, J.L.; Quincey, D.J. Structure from Motion Photogrammetry in Physical Geography. *Prog. Phys. Geogr. Earth Environ.* **2016**, *40*, 247–275. [CrossRef]
8. Green, S.; Bevan, A.; Shapland, M. A Comparative Assessment of Structure from Motion Methods for Archaeological Research. *J. Archaeol. Sci.* **2014**, *46*, 173–181. [CrossRef]
9. López, J.A.B.; Jiménez, G.A.; Romero, M.S.; García, E.A.; Martín, S.F.; Medina, A.L.; Guerrero, J.A.E. 3D Modelling in Archaeology: The Application of Structure from Motion Methods to the Study of the Megalithic Necropolis of Panoria (Granada, Spain). *J. Archaeol. Sci. Rep.* **2016**, *10*, 495–506. [CrossRef]
10. Ferrer, G.; Garrell, A.; Sanfeliu, A. Social-Aware Robot Navigation in Urban Environments. In Proceedings of the 2013 European Conference on Mobile Robots, Barcelona, Spain, 25–27 September 2013; pp. 331–336.
11. Huang, Y.-P.; Sithole, L.; Lee, T.-T. Structure From Motion Technique for Scene Detection Using Autonomous Drone Navigation. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 2559–2570. [CrossRef]
12. Havlena, M.; Torii, A.; Pajdla, T. Efficient Structure from Motion by Graph Optimization. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010; Springer: Berlin, Germany, 2010; pp. 100–113.
13. Kneip, L.; Scaramuzza, D.; Siegwart, R. A Novel Parametrization of the Perspective-Three-Point Problem for a Direct Computation of Absolute Camera Position and Orientation. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2969–2976.
14. Sweeney, C.; Hollerer, T.; Turk, M. Theia: A Fast and Scalable Structure-from-Motion Library. In Proceedings of the 23rd ACM International Conference on Multimedia, Ottawa, ON, Canada, 26–30 October 2015; pp. 693–696.
15. Cui, H.; Shen, S.; Gao, W.; Liu, H.; Wang, Z. Efficient and Robust Large-Scale Structure-from-Motion via Track Selection and Camera Prioritization. *ISPRS J. Photogramm. Remote Sens.* **2019**, *156*, 202–214. [CrossRef]
16. Govindu, V.M. Combining Two-View Constraints for Motion Estimation. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 8–14 December 2001; IEEE: Piscataway, NJ, USA, 2001; Volume 2, p. II.
17. Crandall, D.; Owens, A.; Snavely, N.; Huttenlocher, D. Discrete-Continuous Optimization for Large-Scale Structure from Motion. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 3001–3008.
18. Cui, Z.; Tan, P. Global Structure-from-Motion by Similarity Averaging. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 864–872.
19. Wang, X.; Rottensteiner, F.; Heipke, C. Structure from Motion for Ordered and Unordered Image Sets Based on Random K-d Forests and Global Pose Estimation. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 19–41. [CrossRef]
20. Chen, Y.; Shen, S.; Chen, Y.; Wang, G. Graph-Based Parallel Large Scale Structure from Motion. *Pattern Recognit.* **2020**, *107*, 107537. [CrossRef]
21. Zhu, S.; Zhang, R.; Zhou, L.; Shen, T.; Fang, T.; Tan, P.; Quan, L. Very Large-Scale Global Sfm by Distributed Motion Averaging. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4568–4577.
22. Farenzena, M.; Fusiello, A.; Gherardi, R. Structure-and-Motion Pipeline on a Hierarchical Cluster Tree. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, Japan, 27 September–4 October 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1489–1496.
23. Wang, X.; Xiao, T.; Kasten, Y. A Hybrid Global Image Orientation Method for Simultaneously Estimating Global Rotations and Global Translations. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *5*, 95–104. [CrossRef]
24. Cornelis, K.; Verbiest, F.; Van Gool, L. Drift Detection and Removal for Sequential Structure from Motion Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1249–1259. [CrossRef] [PubMed]
25. Cui, H.; Gao, X.; Shen, S.; Hu, Z. HSfM: Hybrid Structure-from-Motion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1212–1221.
26. Gherardi, R.; Farenzena, M.; Fusiello, A. Improving the Efficiency of Hierarchical Structure-and-Motion. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1594–1600.

27. Toldo, R.; Gherardi, R.; Farenzena, M.; Fusiello, A. Hierarchical Structure-and-Motion Recovery from Uncalibrated Images. *Comput. Vis. Image Underst.* **2015**, *140*, 127–143. [CrossRef]
28. Ni, K.; Dellaert, F. HyperSfM. In Proceedings of the Visualization & Transmission 2012 Second International Conference on 3D Imaging, Modeling, Processing, Zurich, Switzerland, 13–15 October 2012; pp. 144–151.
29. Zhao, L.; Huang, S.; Dissanayake, G. Linear SFM: A Hierarchical Approach to Solving Structure-from-Motion Problems by Decoupling the Linear and Nonlinear Components. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 275–289. [CrossRef]
30. Xu, B.; Zhang, L.; Liu, Y.; Ai, H.; Wang, B.; Sun, Y.; Fan, Z. Robust Hierarchical Structure from Motion for Large-Scale Unstructured Image Sets. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 367–384. [CrossRef]
31. Bhowmick, B.; Patra, S.; Chatterjee, A.; Govindu, V.M.; Banerjee, S. Divide and Conquer: Efficient Large-Scale Structure from Motion Using Graph Partitioning. In *Proceedings of the Computer Vision—ACCV 2014*; Cremers, D., Reid, I., Saito, H., Yang, M.-H., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 273–287.
32. Zhu, S.; Shen, T.; Zhou, L.; Zhang, R.; Wang, J.; Fang, T.; Quan, L. Parallel Structure from Motion from Local Increment to Global Averaging. *arXiv* **2017**, arXiv:1702.08601.
33. Lu, L.; Zhang, Y.; Liu, K. Block Partitioning and Merging for Processing Large-Scale Structure From Motion Problems in Distributed Manner. *IEEE Access* **2019**, *7*, 114400–114413. [CrossRef]
34. Jiang, S.; Li, Q.; Jiang, W.; Chen, W. Parallel Structure From Motion for UAV Images via Weighted Connected Dominating Set. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5413013. [CrossRef]
35. Moulon, P.; Monasse, P.; Marlet, R. Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion. In Proceedings of the IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; pp. 3248–3255.
36. Sweeney, C.; Fragoso, V.; Höllerer, T.; Turk, M. Large Scale SfM with the Distributed Camera Model. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 230–238.
37. Shi, J.; Malik, J. Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905. [CrossRef]
38. Ceres Solver. Available online: http://www.ceres-solver.org/ (accessed on 23 June 2022).
39. Schonberger, J.L.; Frahm, J.-M. Structure-From-Motion Revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
40. Moulon, P.; Monasse, P.; Perrot, R.; Marlet, R. OpenMVG: Open Multiple View Geometry. In *Proceedings of the Reproducible Research in Pattern Recognition*; Kerautret, B., Colom, M., Monasse, P., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 60–74.
41. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.