Article

Rapid Reconstruction of 3D Structural Model Based on Interactive Graph Cuts

Siyu Han, Linsheng Huo , Yize Wang, Jing Zhou and Hongnan Li

School of Civil Engineering, Dalian University of Technology, Dalian 116024, China; siyuu_han@163.com (S.H.); wanyizeshuai@sina.com (Y.W.); zjdlut01@163.com (J.Z.); hnli@dlut.edu.cn (H.L.)

* Correspondence: lshuo@dlut.edu.cn

Abstract: The image-based 3D reconstruction technique has been applied in many scenarios of civil engineering, such as earthquake prevention and disaster reduction, construction monitoring, and intelligent city construction. However, the traditional technique is time-consuming, and the modeling efficiency has become a bottleneck limiting its application in emergency scenarios. In this paper, a rapid reconstruction method is proposed which combines the traditional image-based 3D reconstruction technique and an interactive graph cuts algorithm. Firstly, a sequence of images is collected around the target structure. Then, the images are preprocessed using the interactive iterative graph cuts algorithm to extract the target from each image. Finally, the resulting sequence of images is used to perform the 3D reconstruction. During the preprocessing, only a few images require manual intervention while the rest can be processed automatically. To verify the modeling accuracy of the proposed method, a column that has been destroyed is selected as a target for 3D reconstruction. The results show that compared with the traditional method, the modeling efficiency of the fast reconstruction method is doubled. In addition, the modeling accuracy is 97.65%, which is comparable to the modeling accuracy of the traditional method (97.73%); as well, by comparing the point clouds, the alignment between the two models is tremendously close, with tiny difference. The proposed rapid reconstruction method can be applied in emergency scenarios, such as rapid assessment in post-disaster situations.

Keywords: 3D image reconstruction; structure from motion; interactive graph cuts; extract the target; rapid assessment

Citation: Han, S.; Huo, L.; Wang, Y.; Zhou, J.; Li, H. Rapid Reconstruction of 3D Structural Model Based on Interactive Graph Cuts. *Buildings* **2022**, *12*, 22. <https://doi.org/10.3390/buildings12010022>

Academic Editor: Tiago Miguel Ferreira

Received: 29 November 2021

Accepted: 27 December 2021

Published: 29 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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

With the continuous development of digitization in numerous industries, 3D modeling technology has arisen and has persisted as a hot issue of scientific research. For decades, this technology has been used in many applications of civil engineering, including assessment of earthquake-damaged structures [1,2], monitoring of construction [3,4], real-view 3D data, and the development of geographic information systems (GIS) technology [5,6]. This technology is also applied to protect cultural heritage, such as the restoration and virtual reconstruction of monuments, ancient murals, and other artifacts [7].

Many methods of 3D modeling are available and include range-based and image-based methods. Lidar and laser scanning are the dominant range-based methods. The basic principle of these methods is that the distance between the measured object and the background can be determined by the time difference between the transmitted signal and the received reflected signal. Lidar [8] has high sensitivity, wide scanning range, low energy consumption, and can obtain high-precision point cloud data of the measured object, which is suitable for engineering measurement projects that require high accuracy. Laser scanning [9] can directly provide a large point cloud data set and accurately reconstruct irregular 3D structure models. However, lidar and laser scanning rely on highly complex equipment and they are more expensive, limited in potential uses, and have poor real-time

performance. Image-based 3D modeling technology [10] recovers 3D information of an object from its 2D images. As for the image-based 3D modeling method, along with the continuous updating of electronic technology, anyone can use a mobile phone [11,12] or other photography equipment to image the structure. With the acquired photos, image processing and 3D modeling can be performed on a local device or the cloud. Because of its convenient operation and accurate 3D modeling data, image-based 3D reconstruction has been widely valued in the practical application of civil engineering.

In recent years, many researchers have focused on the important potential of image-based 3D reconstruction and have achieved excellent results in this area. Roberts et al. [13] proposed the possibility of using computer vision to recover the 3D structure from 2D images and set an important milestone for the field of 3D reconstruction. Varady et al. [14] classified the methods of acquiring the 3D image data into contact and non-contact methods. Non-contact methods use imaging and image processing to extract the surface information of the object. Compared to the contact method, the non-contact method is less likely to cause surface damage and has a much wider range of applications. Isgrò et al. [15] further divided the non-contact methods into active and passive methods. The laser scanning method and radar technology mentioned above are active methods, and the image-based 3D reconstruction is a passive method. In this paper, image-based 3D reconstruction is the focus of our research.

Structure from motion (SfM) is a traditional algorithm for image-based 3D reconstruction techniques. Its operation can generate the sparse point cloud, and the sparse point cloud will be further proceeded with by multi-view stereo (MVS) to generate the dense point cloud. SfM technology has become widespread and has permeated all aspects of scientific research. There are two main types of SfM, incremental SfM and global SfM. In the field of incremental SfM, Snavely et al. [16] originally proposed the use of bundle adjustment (BA) to optimize camera parameters and the 3D reconstruction. El Hazzat et al. [17] proposed a method for automating the internal and external parameters of the camera, introducing global-style constraints after the insertion of two initialized images and finally performing local constraints on the remaining images. Their method enabled full automation of 3D modeling. Compared with incremental SfM, global SfM is more likely to fail in the reconstruction process [18]. For such cases, Cui et al. [19] proposed a new framework that uses depth images to upgrade the essential matrices to similarity transformations with better camera parameter estimation for the purpose of global optimization and robust solutions. Moulon et al. [20] proposed a global calibration system for large-scale reconstruction with guaranteed robustness. The researchers also defined an efficient translation registration method to accurately locate the camera position and ensure the accuracy of the reconstruction. Zhou et al. [21] evaluated damaged residential buildings through the high-quality point cloud of image data obtained during Hurricane Sandy. Iheaturu et al. [22] used SfM for 3D terrain mapping, and experiments have proven the applicability of its method to produce highly accurate terrain data. Inzerillo et al. [23] used UAV-SfM and N-SfM (using cameras) to automatically measure the road surface and thus quickly identify road conditions and other key information.

The rapid development of deep learning has opened the door to previously unimagined possibilities [24,25]. Deep learning can handle huge amounts of data and is widely used in various applications such as image cutting, target detection, and localization in 3D reconstruction by combining it with traditional modeling methods. In recent years, many researchers have adapted and optimized deep learning techniques and neural network architectures for making important breakthroughs in image-based 3D reconstruction. Chaiyasarn et al. [26] proposed an image-based crack detection system using a combination of a deep convolutional neural network (CNN) as a feature extractor and a support vector machine (SVM) as a classifier. Their 3D structure reached 92.80% accuracy. Alidoost et al. [27] used the power of CNNs to extract inherent latent features from a single image, interpreted them as 3D information for building reconstruction, and used an optimized multi-scale convolution-deconvolution network (MSCDN) to predict building

heights. The WireNetV2 neural network model developed by Knyaz et al. [28] performed robust segmentation of the complex grid structure in the image, reduced the number of false feature matches, and improved the accuracy of 3D reconstruction. Zhao et al. [29] proposed a convolutional encoder-decoder network to identify pixel-level cracks from structural photos. Their system accurately measured the width and direction of cracks and has excellent potential for future research, particularly if trained with more images of cracks from three-dimensional structure databases and realizing real-time data processing.

Deep learning is powerful but requires large amounts of high-quality labeled training data. Compared with deep learning, especially for small structures, clustering techniques are more applicable. The clustering technique is a kind of unsupervised learning that is widely used in 3D reconstruction of digital images such as K-Means, density-based spatial clustering of applications with noise (DBSCAN), the Gaussian mixture model (GMM), and so on. The core idea of K-Means is to divide samples into different categories through an iterative process. It is simple in principle, easy to operate, and highly convergent. DBSCAN does not require prior knowledge of the number of cluster classes to be formed and is very useful for identifying and processing noisy data and outliers. GMM is an optimization of K-Means, which fits the data well and is highly flexible. In practical engineering, the combination of K-Means and GMM is a common method (i.e., K-Means is used for rough segmentation first, and then GMM is used for detailed iteration). This method has good performance in misclassification probability; it is time-consuming and has robustness in applications such as the clustering of remote sensing images, crack segmentation recognition of bridges, and corrosion monitoring of steel plates [30–32].

Although much research has been conducted in this area, there is still a gap in reducing the time consumed during the image processing to 3D reconstruction. Thus, this paper uses a traditional modeling approach in combination with an interactive graph cuts algorithm based on clustering techniques. In this method, 2D images are preprocessed so that target structures are segmented and the complex background and other irrelevant data are removed. The method greatly reduces the modeling time while maintaining high accuracy and can be widely used in practical engineering.

2. Methodology

In image-based 3D modeling, there is usually photographic information that has nothing to do with the object of interest. Therefore, removing the existence of irrelevant information and retaining the modeling object as much as possible can achieve the purpose of reducing the modeling time. In this paper, we improve the user-interactive graph cuts algorithm, Grabcut. The target of all images except for a few can be automatically segmented in the iterative process. The processed images are then used for 3D modeling. The flow chart of this method is shown in Figure 1 and the operational process is as follows.

1. Use mobile phone or other photography equipment to photograph the target from different angles in order to obtain adequate information about the surface of the target. Photos should be taken at the same plane or level surface and follow a circular path around the target. Lighting conditions, noise, internal and external parameters of the camera, and other factors should remain consistent to facilitate the post-image processing. The final result is an image set capturing all the features of the structure.
2. The user frames and marks eight azimuth pictures that cover the front, back, left, right, top, bottom, etc. of the target structure. The pictures should cover all the characteristic information of the structural members. The pixels outside the frame selection range are classified as the background, and the pixels within the frame selection range are treated as the target object.
3. Use K-Means to perform clustering of the target and background image data, respectively, by selecting the corresponding number of centroids, and obtain the parameters for the Gaussian mixture model (GMM).
4. Use the obtained Gaussian model mixture parameters to minimize and cut the pixels of the remaining pictures. This process is akin to iterative energy minimiza-

tion, and outputs a database with hundreds of segmented photos that have the background removed.

5. Finally, the pictures after image segmentation are used for 3D reconstruction.

Figure 1. The flow chart of interactive graph cuts and 3D reconstruction.

The following content will focus on the core algorithm Grabcut used in the interactive graph cuts and the steps of 3D reconstruction.

2.1. Grabcut

Grabcut is an interactive target extraction algorithm using iterative graph cuts and was proposed by Carsten et al. [33] in 2004. Different from traditional graph cuts, it applies the textural and color features of the images [34–37]. When Grabcut was used in this paper, most images could be segmented automatically while a few required manual intervention. The following sections first explain the segmentation principle of graph cuts [38] and how Grabcut is developed based on graph cuts.

2.1.1. The Segmentation Principle of Graph Cuts

The image is composed of an edge set E and a vertex set V . There are two types of vertices in V . The first are the vertices corresponding to the pixels of the image, and the other are the terminal vertices composed of t and b . An undirected graph with the two types of vertices is called a t - b graph, as shown in Figure 2. Since there are two types of vertices, there are also two types of edges. One is named as m -links, which connects neighboring vertices corresponding to the pixels in the image (i.e., the solid lines in Figure 2). The other is named as n -links, which connects the terminal vertices with the vertices corresponding to the pixels (i.e., dotted lines in the t - b graph).

Each edge in the image is assigned a non-negative weight, w_e . The cut edge C is a subset of the edge set E , and the weight of C is the sum of the weight values of all edges in the edge subset C , which can be represented by Equation (1).

$$|C| = \sum_{e \in C} w_e \quad (1)$$

Figure 2. The t - b diagram of graph cuts. The solid lines represent m -links, connecting the neighboring vertices of the image, and the dotted lines represent n -links, connecting the terminal vertices.

For a cut edge, if the summed weight of all its edges is the smallest, the minimum cut [39,40] will be obtained. The minimum cut divides the vertices of image into two disjoint subsets of T and B , and $t \in T$, $b \in B$, and $T \cup B = V$. T and B , respectively, correspond to the target and background in graph cuts. Therefore, the cut edge should be at the junction between the target and the background. Usually, we use pixel labeling to represent graph cuts. Define $L = \{l_1, l_2, l_3, \dots, l_i, \dots, l_p\}$ where p is the number of pixels in the image and $l_i \in \{0, 1\}$. Therefore, the pixel marked as 1 is regarded as the target object, and the pixel marked as 0 is regarded as the background. This process is obtained by minimizing the energy function. The energy formula of the image is shown in Equation (2):

$$E(L) = \gamma X(L) + Y(L) \quad (2)$$

where $X(L)$ indicates the probability that the pixel belongs to the background or target and $Y(L)$ is related to the degree of similarity between neighboring pixels. γ is the weighting factor, which determines the influence of $X(L)$ and $Y(L)$ on the energy. $X(L)$ can be expressed as follows:

$$X(L) = \sum_{p \in P} X_p(l_p) \quad (3)$$

$$X_p(1) = -\ln \Pr(I_p | t g t') \quad (4)$$

$$X_p(0) = -\ln \Pr(I_p | b k g') \quad (5)$$

$X_p(l_p)$ is the probability that pixel p belongs to label l_p and this probability is calculated based on the histogram of the image. It can be seen from the above formula that when $\Pr(I_p | t g t')$ is greater than $\Pr(I_p | b k g')$, $X_p(1)$ is less than $X_p(0)$. This means that when the probability of the pixel being the target object is greater than the probability of the background, the penalty assigned when the pixel is classified and grouped as the target is smaller. Therefore, when the pixels are correctly divided into two uncorrelated subsets of T and B , $X(L)$ is minimized and $E(L)$ can be reduced at the same time. $Y(L)$ can be obtained by Equations (6)–(8).

$$Y(L) = \sum_{\{p,q\} \in N} Y_{\langle p,q \rangle} \cdot \delta(l_p, l_q) \quad (6)$$

$$\delta(l_p, l_q) = \begin{cases} 0, & l_p = l_q \\ 1, & l_p \neq l_q \end{cases} \quad (7)$$

$$Y_{<p,q>} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \quad (8)$$

In the above equations, p and q are neighborhood pixels and σ is the camera noise. $Y_{<p,q>}$ is the discontinuity penalty between neighboring pixels, and it is lower when the difference in gray values between neighboring pixels is greater, and higher otherwise. Thus, $Y(L)$ is minimized at the boundary between the target and the background.

Using the above algorithms, $E(L)$ is minimized at the boundary between the target and the background, and can therefore be used to obtain the minimum cut.

2.1.2. Grabcut Based on Graph Cuts

Graph cuts use monochrome images, but Grabcut models the colors of the image using GMM. In the RGB color space, the target and background of the image are, respectively, modeled by a full covariance GMM with K Gaussian components.

Therefore, an additional vector $k = \{k_1, \dots, k_n, \dots, k_N\}$ ($k_n \in \{1, \dots, K\}$) is introduced to assign an exclusive Gaussian component to each pixel in the image, and k_n is used to denote the Gaussian component corresponding to the n th pixel. Each pixel is either assigned a corresponding Gaussian component in the target GMM or a corresponding Gaussian component in the background GMM. Grabcut is an iterative upgrade of graph cuts. In Grabcut, the energy function used for cutting is defined as follows:

$$E(\underline{\alpha}, k, \underline{\psi}, x) = V(\underline{\alpha}, k, \underline{\psi}, x) + W(\underline{\alpha}, x) \quad (9)$$

The image is an array $x = (x_1, \dots, x_n, \dots, x_N)$ with gray values. The segmentation of the image is represented as an array of “opacity” values $\underline{\alpha} = (\alpha_1, \dots, \alpha_N)$ on each pixel, where $\alpha_N \in \{0, 1\}$, 0 means background and 1 means target $\underline{\psi} = \{h(x; \alpha), \alpha = 0, 1\}$ is used to represent the distribution of the histogram of the gray values in the target and the background. V and W respectively correspond to the $X(L)$ and $Y(L)$ in graph cuts. The V function is expressed by the following equation:

$$V(\underline{\alpha}, k, \underline{\psi}, x) = \sum_n D(\alpha_n, k_n, \underline{\psi}, x_n) \quad (10)$$

where $D(\alpha_n, k_n, \underline{\psi}, x_n) = -\log \Pr(x_n | \alpha_n, k_n, \underline{\psi}) - \log \pi(x_n, k_n)$. $\Pr(\cdot)$ is the Gaussian probability distribution, and $\pi(\cdot)$ is the mixed weighting coefficient. Through performing a transformation, the following Equation (11) can be obtained.

$$\begin{aligned} D(\alpha_n, k_n, \underline{\psi}, x_n) = & -\log \pi(\alpha_n, k_n) + \frac{1}{2} \log \det \Sigma(\alpha_n, k_n) \\ & + \frac{1}{2} [x_n - \mu(\alpha_n, k_n)]^T \Sigma(\alpha_n, k_n)^{-1} [x_n - \mu(\alpha_n, k_n)] \end{aligned} \quad (11)$$

Therefore, the parameters of the model can be transformed into the following equation.

$$\underline{\psi} = \{\pi(\alpha, k), \mu(\alpha, k), \Sigma(\alpha, k), \alpha = 0, 1, k = 1 \dots K\} \quad (12)$$

From the above formula, the parameters of GMM are composed of the weight π , the mean vector μ and the covariance matrix Σ of each Gaussian component. Therefore, once the three parameters are determined, the target and background GMMs can be used to solve the probability of a pixel belonging to the target and background. The weight of the V function in the energy formula can also be solved after knowing the corresponding RGB values of the image.

The Euclidean distance is used to represent the W function as shown in Equation (13):

$$W(\underline{\alpha}, x) = \beta \sum_{(m,n) \in G} [\alpha_n \neq \alpha_m] \exp -\gamma \|x_m - x_n\|^2 \quad (13)$$

where G is the sum of adjacent pixel pairs, $\gamma = (2 \langle (x_m - x_n)^2 \rangle)^{-1}$, and $\beta = 50$.

2.1.3. Summary of the Implementation Process of Grabcut

Grabcut provides two enhancements to image cutting, namely iterative estimation and incomplete labeling. These two enhancements make it possible to achieve high-quality results for an image with only a small amount of user interaction. The iterative algorithm used in Grabcut replaces the one-time algorithm in graph cut. Each iteration process improves the parameters of both the target and background GMMs. Taking this paper's experiment as an example, the target only needs to be framed (see below) to obtain good segmentation (i.e., incomplete labeling).

In summary, the steps to implement Grabcut in this experiment are as follows.

- 1 Initialization: (i) The user obtains an initial trimap T by directly framing the target, that is, all the pixels outside the frame are used as background pixels T_B , and all the pixels surrounded by the frame are used as pixels T_T that may be part of the target. (ii) For the pixels belonging to the T_B , the initialized pixels are labeled $\alpha_n = 0$, and for each pixel belonging to the T_T , the initialized pixels n are labeled $\alpha_n = 1$. (iii) Through the above steps, the background and target GMMs are initialized from the sets $\alpha_n = 0$ and $\alpha_n = 1$, respectively. Lastly, (iv) the GMMs of the target and the background can be estimated through pixels. The K-Means algorithm is used to cluster the pixels from both groups into K categories so that K Gaussians are obtained. Each Gaussian corresponds to a certain pixel sample set. At this time, the mean vector and covariance of each GMM can be estimated from the RGB values of the pixels, and the weight of the Gaussian component can be determined by the ratio of the number of pixels in the Gaussian component to the total number of pixels.
- 2 Iterative minimization: Gaussian components are assigned to each pixel in the target image. The assigning process can be expressed by the following formula.

$$k_n := \underset{k_n}{\operatorname{argmin}} D_n(\alpha_n, k_n, \psi, x_n) \quad (14)$$

For a given data x , learn to optimize GMM parameters,

$$\underline{\psi} := \underset{\underline{\psi}}{\operatorname{argmin}} V(\underline{\alpha}, k, \underline{\psi}, x) \quad (15)$$

and use the minimum cut for segmentation estimation.

$$\min_{\{\alpha_n: n \in T_T\}} \min_k E(\underline{\alpha}, k, \underline{\psi}, x) \quad (16)$$

These steps (Equations (14)–(16)) are repeated until convergence.

2.2. Steps of Image-Based 3D Reconstruction

The problem of reconstructing 3D models from multiple 2D planar images has been extensively and deeply studied, and every step of 3D reconstruction is crucial [41]. This paper takes a structure as an example to demonstrate the proposed process of 3D reconstruction. The steps taken to reconstruct the 3D model are detailed below.

1. Image data collection: Before image processing, it is necessary to use a camera, a mobile phone, and other image capture devices to photograph the target from different angles and obtain the visual data of the target's surface. The photographic equipment should be kept on the same horizontal surface and follow a circular path around the

- target. When taking photos, careful attention should be paid to the influence of light conditions, noise interference, and internal and external parameters of the camera.
2. Reconstruction of sparse point cloud: The SfM algorithm is used to obtain a large amount of image data from multiple visual points. The scale-invariant feature transform (SIFT) [42–44] operator is then used to perform Gaussian scale-space transformation and detect the feature points of the imported image. The kd-tree model is subsequently used to perform recursive division and match feature points. The epipolar geometry is computed and matching pairs that do not satisfy the fundamental matrix are eliminated by random sample consensus (RANSAC). Finally, the bundle adjustment (BA) is used to further optimize the calibration of the camera position and the coordinates of the 3D point cloud. Through the above steps, the sparse point cloud of the 3D scene is reconstructed. The outline of the target structure is roughly distinguishable through the sparse point cloud reconstruction. The epipolar geometry is computed.
 3. Reconstruction of dense point cloud: Potentially, the sparse point cloud formed by the SfM algorithm contains fewer feature points than needed. A solution to this problem is the clustering multi-view stereo (CMVS) [45] algorithm, where the sparse cloud points generated by the SfM are clustered into different image clusters. Then, the patch-based multi-view stereo (PMVS) [46,47] algorithm is used to initialize feature matching to form a sparse patch set, diffuse the initial matching to adjacent pixels in order to obtain a denser patch, and finally filter and delete wrong matches through filtering to obtain a dense point cloud. Through the above steps, the details of the structure can be clearly revealed.
 4. Poisson surface reconstruction: The process of transition from point samples to reconstruction of the 3D structure surface is an important research topic. Surface reconstruction based on the Poisson [48,49] algorithm expresses the directed point set of the model surface through an indicator function. The surface integral is approximately solved by mapping between the indicator function and the vector field. The Poisson equation is constructed and solved to recover the indicator function and extract the isosurface.
 5. Texture mapping: In order to make the target more realistic and detailed, texture mapping is required. Through a process known as projection mapping, a two-dimensional texture is mapped onto the object's surface. A 2D pattern is defined at every point on the texture space as a color or grayscale value. Texture mapping puts the finishing touches on the reconstructed 3D model, and at this point, the model should closely resemble the target structure.

3. Full-Scale Validation

3.1. Overview of the Experimental Model

In this paper, a partially destroyed square column with a size of $500\text{ mm} \times 200\text{ mm} \times 200\text{ mm}$ was selected as the target object. In the experiment, a smartphone captured photos of the structure at a small angle to facilitate the matching of feature points between the pictures. A total of 116 pictures (resolution of 2976×3968) were collected.

First, according to the Grabcut interactive segmentation steps, eight photos, which collectively contained the overall information of the structure, were selected and framed. The results of frame selection are shown in Figure 3. The selection includes enough perspectives of the structure to allow better understanding of the basic characteristics of the structure.

Figure 3. Framework selection diagrams that roughly cover all aspects of the structure.

Afterwards, the Gaussian model was obtained through cluster analysis, and then the remaining hundreds of photos were iteratively segmented. A portion of the original images and the processed images are shown in Figure 4.

As can be seen from the above figure, the Grabcut interactive segmentation algorithm provides superior segmentation for the target object and the background. After the iterative segmentation, the three selected images lost minimal information, which did not affect the reconstruction result. Therefore, it can be considered that the segmentation effectiveness is close to 100%.

This paper used VisualSFM [50] to perform feature point detection and matching on experimentally contained images, camera parameter calibration, and sparse point cloud reconstruction. CMVS and PMVS were used to intensively process point clouds. Meshlab [51] was then utilized to perform Poisson surface reconstruction and texture mapping based on the generated dense point cloud. The above process is shown in Figure 5. Finally, the model reconstructed from the original images was compared with the model reconstructed from the segmented preprocessed image. The comparison result is shown in Figure 6. It can be seen that the reconstructed structure is largely similar to the original structure, and no important structural components are missing.

Figure 4. Comparison of the original images and the images after interactive segmentation. (a–c) original images, (d–f) the images after interactive segmentation.

Figure 5. Cont.

Figure 5. Images from each step of the 3D reconstruction process. (a) Sparse point cloud image; (b) dense point cloud image; (c) Poisson surface reconstruction image; (d) texture-mapped image.

Figure 6. Comparison of the 3D models of the original image and the segmented image: (a) 3D reconstruction model of the original picture; (b) 3D reconstruction model of the segmented picture.

3.2. Experimental Verification

3.2.1. Verification of Reconstruction Time

Before the image segmentation, many irrelevant elements are present around the structure. However, after segmentation, almost only the structure itself is left in the image, which improves the speed of the 3D reconstruction. To verify this, we separately counted the time for each stage and the total time for 3D reconstruction, and compared the times taken for the original structure versus the interactively segmented structure (Table 1).

Table 1. Comparison of consumed time in 3D reconstruction.

Reconstruction Steps	Modeling Time	
	Original Model (/s)	Grabcut Interactive Cutting Model (/s)
Sparse Point Cloud Reconstruction	504.000	258.000
Dense Point Cloud Reconstruction	1824.000	766.020
Poisson Reconstruction	170.610	109.670
Texture Mapping	222.010	92.810
Total Time	2720.620	1226.500

The above figures and data show that the images were effectively segmented after Grabcut target processing, which removed much of the irrelevant parts of the image, thereby reducing the computational time and resources needed. The truncation of irrelevant information nearly reduced the total time by half.

3.2.2. Verification of Reconstruction Accuracy

To further verify that the 3D model has high fidelity to the actual structure after interactive graph cuts, the accuracy of the reconstruction model was calculated. First, the 3D reconstructed model was resized according to the physical dimensions of the square column to find the degree of scaling. The calibration point of the square column is shown in Figure 7. The degree of scaling between the 3D reconstruction model and the actual structure is shown in Table 2. Table 3 lists additional calibration points used to verify the accuracy of the reconstruction model.

Figure 7. Location of calibration points on the square column. The purpose is to measure the distance of each calibration point as the criterion of accuracy.

Table 2. The size conversion relationship between the 3D reconstruction model and the actual structure.

Types of Images	Calibration Point Measured Dimension in Real Structure	A-B/mm 200.000	C-D/mm 500.000	E-F/mm 108.5000
Original Images	3D Reconstruction Model Scaling of the Model	6.256 31.969×	15.169 32.012×	3.385 32.056×
Grabcut Interactive Cutting Images	3D Reconstruction Model Scaling of the Model	9.274 21.566×	23.047 21.694×	5.032 21.561×

Table 3. Comparison of the size for actual structure and reconstruction model.

Types of Images	Calibration Point Measured Dimension in Real Structure/mm	O-P 98.000	P-Q 107.000	O-Q 200.000	L-M 88.000	M-N 99.000	L-N 185.000
Original Images	3D Reconstruction Model/mm	99.498	104.761	201.465	86.229	101.322	183.906
	Experimental Error	1.53%	2.09%	0.73%	2.01%	2.35%	0.59%
Grabcut Interactive Cutting Images	3D Reconstruction Model/mm	99.492	105.257	204.531	87.528	100.546	188.902
	Experimental Error	1.52%	1.63%	2.27%	0.54%	1.56%	2.11%

It can be seen from Table 3 that the maximum difference between the 3D model reconstructed from original images and the physical structure is 2.35%. That is, the accuracy is 97.65%. This error may be attributed primarily to human operation, and so it is within the allowable range for most practical projects. After the interactive image processing, the maximum error between the 3D model and the real structure is 2.27%. The low error demonstrates that the accuracy of the structural reconstruction has not been reduced due to the introduction of human error.

In addition, to further verify the accuracy of the model modeling, it is essential to compare the differences between the 3D models of the original image and the segmented one. Therefore, a comparison with the point clouds between the two models was implemented in CloudCompare. To distinguish the comparison, more redundant point clouds were kept in the original model. The models were first scaled to the same size, and then the 3D model of the original image was considered as a reference, and the distance of each point cloud of the model of the segmented image relative to the reference was iteratively calculated to obtain the minimum distance between the two models. Finally, its alignment result was achieved. As shown in Figure 8, the model ontology shows blue features (i.e., the more blue features, the better the structural alignment), which indicates that the two models are aligned with very successful results. The difference between the two models is tiny, which maximizes the accuracy of the modeling.

In summary, the image reconstruction model of segmented images meets the requirements of practicability and feasibility.

Figure 8. The results of the alignment between the 3D models of the original image and the segmented image. (a) The two models before alignment; (b) The two models after alignment.

4. Discussion

Compared to previous image-based 3D modeling methods that directly use raw images to reconstruct, the proposed method can reduce the modeling time by half by preprocessing images with the interactive graph cuts algorithm. This is of great significance for the rapid assessment of earthquake-damaged structures or other situations where a 3D structural model is needed in engineering. However, there are still some limitations to the practical applications of the proposed method. One challenge is the 3D reconstruction for large-scale and complex structures using the proposed method. Due to the limited resolution of a digital image, the reconstructed 3D model for a large-scale structure is low-precision and is hardly in use. We will develop a segmental merging algorithm in future work to solve this problem, where several sub-3D models of the target structure at different heights are reconstructed and then merged to generate the 3D model of the entire structure. Another limitation is that the lighting condition on images, such as the target structure being in the shadow of other objects, can greatly affect the modeling accuracy. The corresponding preprocessing methods to improve the image quality before 3D reconstruction will be studied in the future. In addition, this study only validates the proposed method with a simple structure in a laboratory environment, and its practicability will be further studied for large-scale and complex structures with the consideration of the above limitations.

5. Conclusions

Although the image-based 3D reconstruction technique has been applied in many scenarios of civil engineering, modeling efficiency has become the bottleneck against application in emergency scenarios. To tackle this issue, a rapid reconstruction method combining the traditional image-based 3D reconstruction technique and an interactive graph cuts algorithm is proposed and verified in this paper. Full-scale experimental validation has led to the following conclusions:

- (1) In the process of interactive graph cuts, only a few images require manual intervention, while the rest can be processed automatically. This achieves the purpose of removing the redundant background of the image to keep only the target objects as much as possible. Compared to the traditional image-based 3D reconstruction, the introduction of the algorithm can double the modeling efficiency without reducing the accuracy of the model. The modeling time can be reduced by more than one half.
- (2) The proposed rapid reconstruction method is robust to structural damage, and it can accurately reconstruct the damage of the structure. In the demonstration case, the modeling accuracy reached 97.65%, which is comparable to the modeling accuracy of the traditional method (97.73%). Moreover, the alignment between the models is excellent, and the difference between the two models is tiny.
- (3) The proposed method is simple, effective, and accurate, and it can be used in emergency scenarios, such as rapid assessment in post-disaster situations.

However, the experimental validation is still in the laboratory phase. Future work should focus more on applications in real-world scenarios and large-scale structures.

Author Contributions: Conceptualization, L.H.; methodology, Y.W.; software, S.H.; validation, Y.W.; formal analysis, S.H. and Y.W.; investigation, J.Z.; resources, L.H.; data curation, L.H.; writing—original draft preparation, S.H. and Y.W.; writing—review and editing, L.H. and J.Z.; visualization, J.Z.; supervision, L.H.; project administration, L.H. and H.L.; funding acquisition, L.H. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 51778111) and the Dalian High Level Talent Innovation Support Program (grant number 2019RD01).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

- Freeman, M.; Vernon, C.; Berrett, B.; Hastings, N.; Derricott, J.; Pace, J.; Home, B.; Hammond, J.; Janson, J.; Chiabrando, F.; et al. Sequential earthquake damage assessment incorporating optimized sUAV remote sensing at Pescara del Tronto. *Geosciences* **2019**, *9*, 332. [\[CrossRef\]](#)
- Barrington, L.; Ghosh, S.; Greene, M.; Har-Noy, S.; Berger, J.; Gill, S.; Lin, A.Y.M.; Huyck, C. Crowdsourcing earthquake damage assessment using remote sensing imagery. *Ann. Geophys.* **2011**, *54*, 680–687. [\[CrossRef\]](#)
- Zhao, S.; Kang, F.; Li, J.; Ma, C. Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction. *Autom. Constr.* **2021**, *130*, 103832. [\[CrossRef\]](#)
- Nguyen, C.H.P.; Choi, Y. Comparison of point cloud data and 3D CAD data for on-site dimensional inspection of industrial plant piping systems. *Autom. Constr.* **2018**, *91*, 44–52. [\[CrossRef\]](#)
- Yang, Y. Developing a mobile mapping system for 3D GIS and smart city planning. *Sustainability* **2019**, *11*, 3713. [\[CrossRef\]](#)
- Zhu, J.; Wu, P. Towards effective BIM/GIS data integration for smart city by integrating computer graphics technique. *Remote Sens.* **2021**, *13*, 1889. [\[CrossRef\]](#)
- Pietroni, E.; Ferdani, D. Virtual restoration and virtual reconstruction in cultural heritage: Terminology, methodologies, visual representation techniques and cognitive models. *Information* **2021**, *12*, 167. [\[CrossRef\]](#)
- Yang, S.; Fan, Y. 3D building scene reconstruction based on 3D LiDAR point cloud. In Proceedings of the IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taipei, Taiwan, 12–14 June 2017.
- Brenner, C. Building reconstruction from images and laser scanning. *Int. J. Appl. Earth Obs. Geoinf.* **2005**, *6*, 187–198. [\[CrossRef\]](#)
- Remondino, F.; El-Hakim, S. Image-based 3D modelling: A review. *Photogramm. Rec.* **2006**, *21*, 269–291. [\[CrossRef\]](#)
- Xie, B.; Li, J.; Zhao, X. Research on damage detection of a 3D steel frame model using smartphones. *Sensors* **2019**, *19*, 745. [\[CrossRef\]](#)
- Xie, B.; Li, J.; Zhao, X. Strain measurement based on speeded-up robust feature algorithm applied to microimages from a smartphone-based microscope. *Sensors* **2020**, *20*, 2805. [\[CrossRef\]](#) [\[PubMed\]](#)
- Roberts, L.G. Machine Perception of Three-Dimensional Solids. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, June 1963.
- Varady, T.; Martin, R.R.; Cox, J. Reverse engineering of geometric models-An introduction. *Comput.-Aided Des.* **1997**, *29*, 255–268. [\[CrossRef\]](#)
- Isgro, F.; Odone, F.; Verri, A. An open system for 3D data acquisition from multiple sensor. In Proceedings of the seventh International Workshop on Computer Architecture for Machine Perception, Palermo, Italy, 4–6 July 2005; pp. 52–57.
- Snavely, N.; Seitz, S.M.; Szeliski, R. Modeling the world from Internet photo collections. *Int. J. Comput. Vis.* **2008**, *80*, 189–210. [\[CrossRef\]](#)
- El Hazzat, S.; Merras, M.; El Akkad, N.; Saaidi, A.; Satori, K. 3D reconstruction system based on incremental structure from motion using a camera with varying parameters. *Visual Comput.* **2018**, *34*, 1443–1460. [\[CrossRef\]](#)
- Jiang, S.; Jiang, C.; Jiang, W. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. *ISPRS-J. Photogramm. Remote Sens.* **2020**, *167*, 230–251. [\[CrossRef\]](#)
- Cui, Z.; Tan, P. Global structure-from-motion by similarity averaging. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 864–872.
- 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 (ICCV), Sydney, Australia, 1–8 December 2013; pp. 3248–3255.
- Zhou, Z.; Gong, J.; Guo, M. Image-based 3D reconstruction for posthurricane residential building damage assessment. *J. Comput. Civil. Eng.* **2016**, *30*, 04015015. [\[CrossRef\]](#)
- Iheaturu, C.J.; Ayodele, E.G.; Okolie, C.J. An assessment of the accuracy of structure-from-motion (SfM) photogrammetry for 3D terrain mapping. *Geomat. Landmanage. Landsc.* **2020**, *2020*, 65–82. [\[CrossRef\]](#)
- Inzerillo, L.; Di Mino, G.; Roberts, R. Image-based 3D reconstruction using traditional and UAV datasets for analysis of road pavement distress. *Autom. Constr.* **2018**, *96*, 457–469. [\[CrossRef\]](#)
- Özdemir, E.; Remondino, F.; Golkar, A. Aerial point cloud classification with deep learning and machine learning algorithms. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *42*, 843–849. [\[CrossRef\]](#)
- Yuniarti, A.; Suciati, N. A review of deep learning techniques for 3D reconstruction of 2D images. In Proceedings of the 12th International Conference on Information and Communication Technology and System (ICTS), Surabaya, Indonesia, 18 July 2019; pp. 327–331.
- Chaiyasarn, K.; Buatik, A.; Likitlersuang, S. Concrete crack detection and 3D mapping by integrated convolutional neural networks architecture. *Adv. Struct. Eng.* **2021**, *24*, 1480–1494. [\[CrossRef\]](#)

27. Alidoost, F.; Arefi, H.; Tombari, F. 2D image-to-3D model: Knowledge-based 3D building reconstruction (3DBR) using single aerial images and convolutional neural networks (CNNs). *Remote Sens.* **2019**, *11*, 2219. [\[CrossRef\]](#)
28. Knyaz, V.A.; Kniaz, V.V.; Remondino, F.; Zheltov, S.Y.; Gruen, A. 3D reconstruction of a complex grid structure combining UAS images and deep Learning. *Remote Sens.* **2020**, *12*, 3128. [\[CrossRef\]](#)
29. Li, S.; Zhao, X. Automatic crack detection and measurement of concrete structure using convolutional encoder-decoder network. *IEEE Access* **2020**, *8*, 134602–134618. [\[CrossRef\]](#)
30. Neagoe, V.E.; Chirila-Berbentea, V. Improved Gaussian mixture model with expectation-maximization for clustering of remote sensing imagery. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 3063–3065.
31. Duan, M.; Lu, Y.; Li, Z.; Su, Y. An Improved Bridge Crack Image Segmentation Method. *J. Highway Transp. Res. Dev.* **2020**, *37*, 63–70.
32. Soo, K.B.; Jae-sung, K.; Choi, S.W.; Jung-pil, N.; Lee, K.; Jeong-hyeon, Y. Corrosion image monitoring of steel plate by using K-Means clustering. *J. Korean Inst. Surf. Eng.* **2021**, *54*, 278–284.
33. Rother, C.; Kolmogorov, V.; Blake, A. “GrabCut”—Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. [\[CrossRef\]](#)
34. Nayak, T.; Bhoi, N. Object detection and tracking using watershed segmentation and KLT tracker. *Glob. J. Comput. Sci. Technol.* **2020**.
35. Akay, B. A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding. *Appl. Soft. Comput.* **2013**, *13*, 3066–3091. [\[CrossRef\]](#)
36. Bhandari, A.K.; Kumar, A.; Singh, G.K. Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur’s, Otsu and Tsallis functions. *Expert Syst. Appl.* **2015**, *42*, 1573–1601. [\[CrossRef\]](#)
37. Kaur, R.; Juneja, M.; Mandal, A.K. A hybrid edge-based technique for segmentation of renal lesions in CT images. *Multimed. Tools Appl.* **2019**, *78*, 12917–12937. [\[CrossRef\]](#)
38. Yi, F.; Moon, I. Image segmentation: A survey of graph-cut methods. In Proceedings of the 2012 international conference on systems and informatics (ICSAI2012), Yantai, China, 19–20 May 2012; pp. 1936–1941.
39. Dantzig, G.; Fulkerson, D.R. On the max flow min cut theorem of networks. *Linear Inequal. Relat. Syst.* **2003**, *38*, 225–231.
40. Yuan, J.; Bae, E.; Tai, X. A Study on continuous max-flow and min-cut approaches. In Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 2217–2224.
41. El Hazzat, S.; El Akkad, N.; Merras, M.; Saaidi, A.; Satori, K. Fast 3D reconstruction and modeling method based on the good choice of image pairs for modified match propagation. *Multimed. Tools Appl.* **2020**, *79*, 7159–7173. [\[CrossRef\]](#)
42. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 1150–1157.
43. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
44. Hossein-Nejad, Z.; Agahi, H.; Mahmoodzadeh, A. Image matching based on the adaptive redundant keypoint elimination method in the SIFT algorithm. *Pattern Anal. Appl.* **2021**, *24*, 669–683. [\[CrossRef\]](#)
45. Bansal, M.; Sharma, D. A novel multi-view clustering approach via proximity-based factorization targeting structural maintenance and sparsity challenges for text and image categorization. *Inf. Process. Manag.* **2021**, *58*, 102546. [\[CrossRef\]](#)
46. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376. [\[CrossRef\]](#)
47. Li, B.; Venkatesh, Y.; Kassim, A.; Lu, Y. Improving PMVS algorithm for 3D scene reconstruction from sparse stereo pairs. In Proceedings of the Pacific-Rim Conference on Multimedia, Nanjing, China, 13–16 December 2013; pp. 221–232.
48. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson surface reconstruction. In Proceedings of the Fourth Eurographics Symposium on Geometry Processing, Sardinia, Italy, 26–28 June 2006.
49. Morel, J.; Bac, A.; Vega, C. Surface reconstruction of incomplete datasets: A novel Poisson surface approach based on CSRBF. *Comput. Graph.* **2018**, *74*, 44–55. [\[CrossRef\]](#)
50. Morgan, J.A.; Brogan, D.J. *How to VisualSFM*; Department of Civil & Environmental Engineering Colorado State University: Fort Collins, CO, USA, 2016.
51. Cignoni, P.; Callieri, M.; Corsini, M.; Dellepiane, M.; Ganovelli, F.; Ranzuglia, G. Meshlab: An open-source mesh processing tool. In Proceedings of the Eurographics Italian Chapter Conference, Salerno, Italy, 14–18 April 2008; pp. 129–136.