Research Article

Automated 3D Scenes Reconstruction Using Multiple Stereo Pairs from Portable Four-Camera Photographic Measurement System

Qi Peng,[1] Lifen Tu,[1] Kaibing Zhang,[1] and Sidong Zhong[2]

[1]*School of Physics and Electronic Information Engineering, Hubei Engineering University, Xiaogan 432000, China*
[2]*School of Electronic Information, Wuhan University, Wuhan 430072, China*

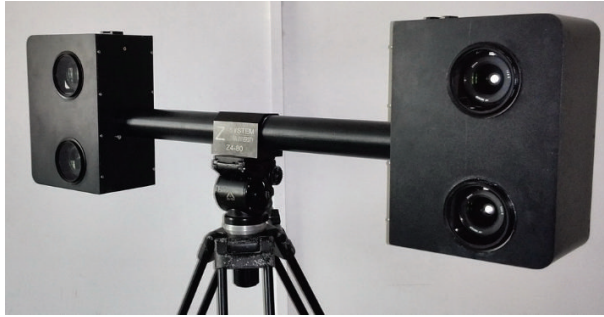Correspondence should be addressed to Lifen Tu; tulifen_0301@163.com

An effective automatic 3D reconstruction method using a portable four-camera photographic measurement system (PFCPMS) is proposed. By taking advantage of the complementary stereo information from four cameras, a fast and highly accurate feature point matching algorithm is developed for 3D reconstruction. Specifically, we first utilize a projection method to obtain a large number of dense feature points. And then a reduction and clustering treatment is applied to simplify the Delaunay triangulation process and reconstruct a 3D model for each scene. In addition, a 3D model stitching approach is proposed to further improve the performance of the limited field-of-view for image-based method. The experimental results tested on the 172 cave in Mogao Grottoes indicate that the proposed method is effective to reconstruct a 3D scene with a low-cost four-camera photographic measurement system.
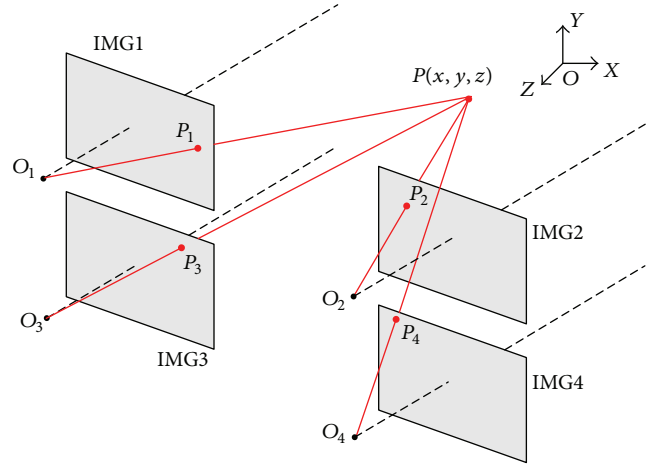
1. Introduction

The 3D models have been widely used in various applications such as virtual cultural heritage conservation [1], architecture design [2], and cartography [3]. The virtual cultural heritage conservation refers to 3D digitization and replication of cultural heritage, which has become an interesting and challenging research field [4]. The cultural heritage is the witness of the historical development and an important basis for historical research. Reconstructing the heritage in 3D model allows users to view the heritage online, regardless of their locations [5]. Besides, the 3D digital technology will not damage the objects when the observers get their surface information, which is beneficial to preserve the 3D digital cultural heritage for a long time. For these reasons, it is of importance to develop a fast, automated, and yet cost-effective 3D model reconstruction method.

Real object models can be reconstructed automatically using active and passive methods. Object range scanning by laser [6] and structured light [7] are typical examples of the active methods. The advantage of laser scanners is

their high accuracy in geometry measurement. However, these methods often demand expensive equipment and a separate step for object texture acquisition and registration [8]. Image-based methods are typical passive methods which require a simpler and less expensive setup to reconstruct a 3D geometry [9]. Nevertheless, their weakness is the capture of a partial observation of the surrounding environment if a normal camera with a limited field-of-view is used. As a result, a large number of views located at different positions are required to reconstruct a complete model for the 3D scene, which leads to the problem of calibrating multiple cameras and the pose of each camera. Especially for the traditional image-based methods based on single camera, which need complex and time-consuming postprocessing in order to obtain 3D information for the scene [10]. Other image-based methods are based on stereo image pairs, in which a 3D model for each pair of images is reconstructed by stereo matching [11]. The advantage of this group is that they are flexible in operating and 3D information can be obtained easily. Despite their effectiveness, there remain two open problems: one is high accuracy stereo matching to acquire

(a)

(b)

FIGURE 1: The main hardware system: (a) the PFCPMS and (b) the structure.

reliable point cloud, and the other is 3D model stitching for adjacent scenes captured in different locations.

A 3D model reconstruction system using images acquired from multiple stereo pairs by PFCPMS [12] has been proposed. First, stereo images and the checkerboard projection are assisted to get reliable point cloud by PFCPMS. Second, the point reduction and classification algorithm is applied to generate Delaunay triangulation [13] for texture mapping [14]. Last, a complete 3D model is generated by integrating partial scene models. The major contributions are the following: (1) The PFCPMS is made of four identical cameras, which constitute six groups of stereo vision system. This special structure can be used to effectively obtain high accuracy stereo matching points. (2) The checkerboard is projected on the surface of the models for obtaining dense and evenly distributed point cloud, especially for the regions lack of textures. (3) The point reduction algorithm can reduce both the time complexity and space complexity dramatically, and the classification algorithm can boost up the correctness of the Delaunay triangulation effectively. (4) The point cloud can be used as a space constraint in feature point matching, and thus the distance between two matched points in two different 3D models can suppress false matches when calculating the rotation and translation parameters between two 3D models.

2. Data Collection

The data collection process is divided into two parts. The first is the hardware system. The second is the system's matching method and how to ensure the correctness of the matches.

2.1. The Main Hardware System. The PFCPMS is shown in Figure 1(a), where the four cameras are in a rectangular distribution. As shown, the baseline between the left two cameras is short, so the difference of images captured by them can be minimized, and the accuracy of feature matching between the left two cameras can be improved. The right two cameras constitute another stereo pair which is similar to

the left two cameras. We can see that the baseline between the upper left (UL) and upper right (UR) cameras is long, similar to the down left (DL) and down right (DR) cameras. The long baseline helps to improve the accuracy of the spatial locations for points. For the PFCPMS, the long baseline and short baseline camera have different characteristics, which are complementary to each other and propitious to achieve efficient and accurate stereo matching.

The optical axes of four cameras are parallel to each other, and their parameters should be consistent as far as possible in order to ensure the precision. The four image planes and their optical centers are on an identical plane. Let the four image planes be IMG1, IMG2, IMG3, and IMG4, respectively, and the four optical centers be O_1, O_2, O_3, and O_4, as shown in Figure 1(b). Given any visible point P in the world coordinate, it can be imaged in the four cameras denoted as P_1, P_2, P_3, and P_4, respectively.

2.2. Reliable Point Cloud. The PFCPMS can perform high accuracy stereo matching quickly. As shown in Figure 2(a), $l_{i,j}$ $(i, j = 1, 2, 3, 4)$ are the epipolars, and they can be calculated easily when all the camera parameters are known for this system [15]. The matching flow can be divided into six stages: (1) Detect the Harris [16] feature points for the UL. (2) For one feature point P_1 in UL, search the matched point P_3 around l_{13} in DL using the common correlation method [17]. Because the baseline between UL and DL is very short, their images are similar to each other. It narrows the search within a small area, so the matching process is fast and accurate. (3) Calculate the matched point P_2 in UR by epipolar constraint, which is the intersection between l_{12} and l_{23}. (4) Search the matched point P_4 around l_{24} in DR for P_2 by correlation method, similar to Step 2. (5) Calculate the auxiliary matched point P_{44} in DR by l_{24} and l_{34}, similar to Step 3. In an ideal condition, P_4 and P_{44} overlapped, or the match is false. In practice, a small area around P_{44} is permitted because of the effect of the point spread function. (6) Obtain the auxiliary matched point P_{11} in UL by l_{13} and l_{14}, and determine whether
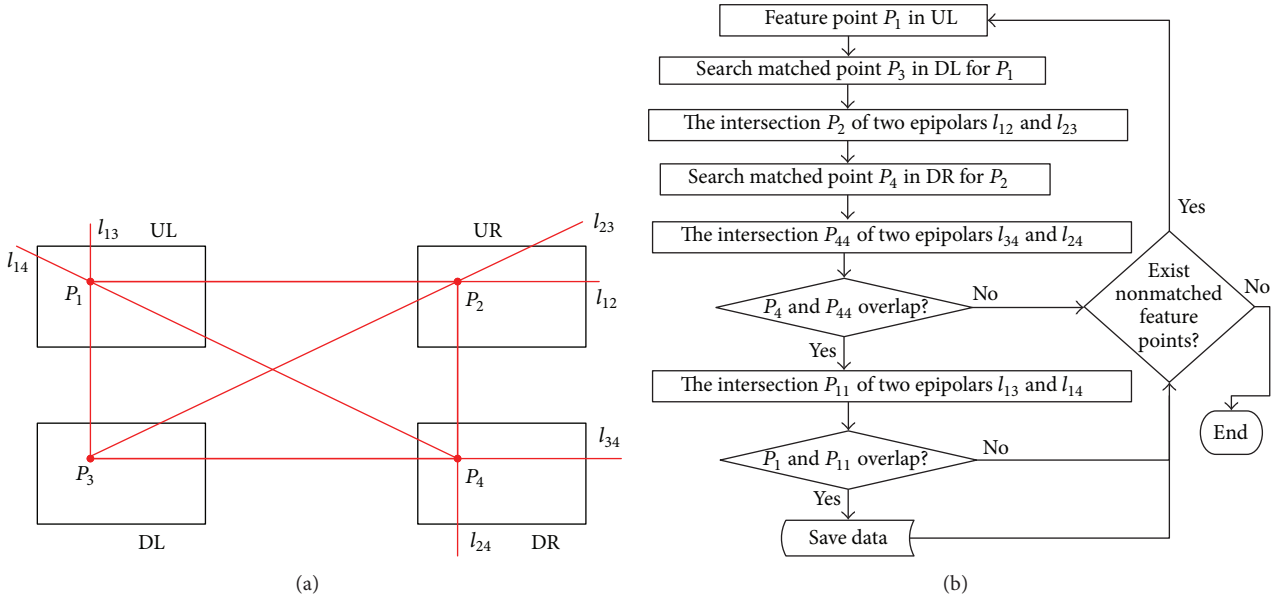
FIGURE 2: The feature matching method for the PFCPMS: (a) the epipolar characteristic and (b) the flow chart for feature matching.

P_1 and P_{11} are overlapped, which is similar to Step 5. When P_4 and P_{44} and P_1 and P_{11} overlapped each other at the same time, an effect matched pair is found. The other feature points in UL should be detected the same. A simple flow chart for feature matching is illustrated in Figure 2(b).

The last two steps of the matching flow are very important and necessary, though the short baseline can narrow the search area and reduce the mismatches markedly; the errors will appear when the surface has repetitive texture, especially the dense and small one. Usually, more than one matched P_3 and P_4 can be detected around their epipolars. Fortunately, the overlap ratio between the matched point and the auxiliary matched point added another constraint which can adapt to repetitive texture surface effectively.

The PFCPMS improved the stereo matching effectively, but a problem should be taken into consideration in order to get enough reliable point cloud. It is the poor texture, as shown in Figure 3(a). The surface of the Buddha is very smooth, no feature point exists in large areas. As a result, a few initial feature points will be detected and insufficient matches will be obtained in the first step of the matching flow in data collection process, so the smooth surface is difficult to reconstruct by traditional image-based methods.

In order to get enough reliable point cloud, a projection method is applied, as shown in Figure 3(b). A simple program is designed to generate black and white checkerboard, as shown in Figure 3(c). Then it projects onto the object to add texture information by a projector, as shown in Figure 3(d). The square's size should big enough to ensure the matching correctness. In the meantime, the checkerboard should be dense enough to get enough point cloud, and then the square's size should be small. To meet these requirements, a larger square is used; then it subdivided by small translation of the projection; at last all the feature points in different position are fused to form a complete point cloud. Usually, the square

size and subdividing number are not very strict; they can be adjusted according to precision requirement. Define the focal length of the projector as f, the distance between projector and object as D, and the square size as w, and it was subdivided nine times; then the actual distance between two connected corner points h satisfy

$$h = \frac{w}{9} \times \left(\frac{D}{f} - 1 \right). \tag{1}$$

h is the minimum measurable distance between two points, and this can be defined as the measurable precision.

3. Data Analysis

After obtained 3D point cloud, Delaunay triangulation [13] and texture mapping [14] can be applied to reconstruct 3D model for every scene, and the neighboring 3D models can be jointed together to gain a complete 3D model.

3.1. 3D Reconstruction. The original point cloud obtained from Section 2 is shown in Figure 4(a). We can see that more than two hundred thousand points have been obtained for this model, and they are dense and uniformly distributed. The point cloud is so large that the computational complexity is very intensive, especially for the Delaunay triangulation process. Usually, the overdense data is not necessary, so some point cloud reduction algorithms can be applied to reduce the computational cost [18]. In this paper, an improved bounding box method is used. The traditional bounding box method [19] is divided into three steps: First, search a minimum box which can surround all the points. Next the box is divided into many connected small boxes. Finally, one point nearest to the center of gravity for every small box is reserved. The reduction result is shown in Figure 4(b), where less than

(a)

(b)

(c)

(d)

FIGURE 3: Intensive feature point generation method: (a) poor texture, (b) square projection, (c) the checkerboard, and (d) the projector.
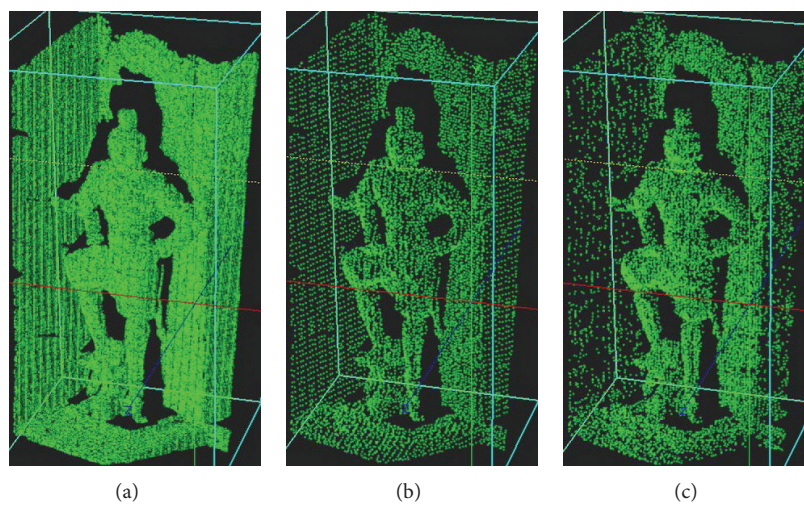

(a)

(b)

(c)

FIGURE 4: Point cloud data processing: (a) original data, (b) uniform sampling, and (c) nonuniform sampling.
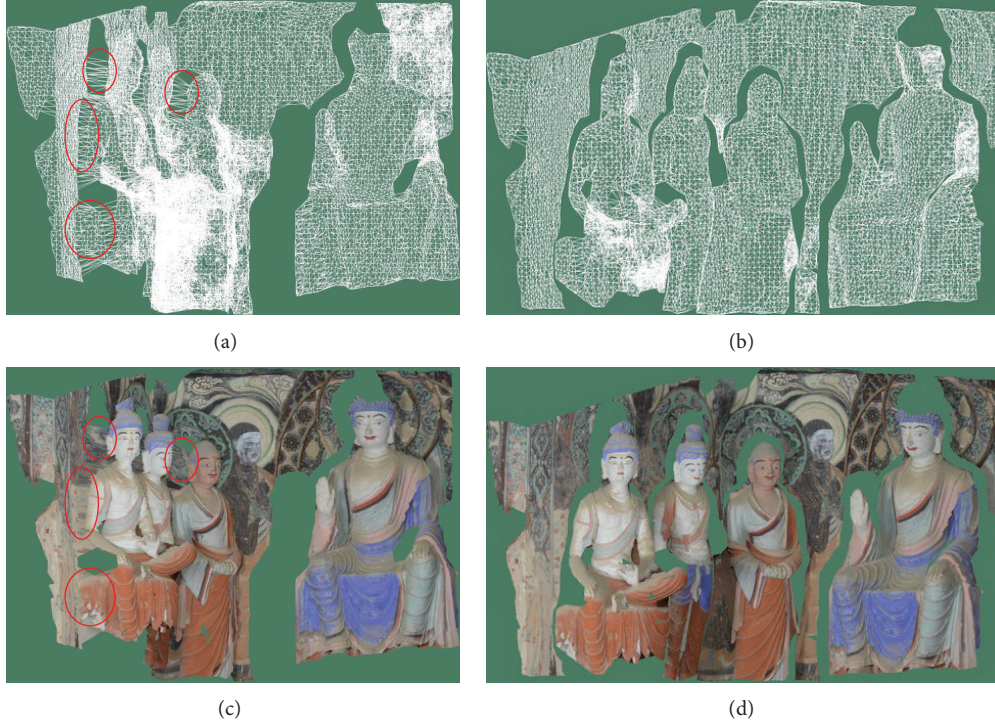
Figure 5: The point cloud classification optimization: (a) original triangular mesh, (b) triangular mesh after clustering, (c) original 3D model, and (d) 3D model after clustering.

10% points are retained. Obviously, the traditional bounding box method has not considered the characteristic of the scene. It is a uniform sampling for the original data. But in fact a few numbers of points are needed to describe a flat area, for example, wall, floor, and ceiling. On the contrary, if the curvature of surface is changeful, for example, face, hair, and clothes, more points are needed such that the 3D reconstruction precision can be guaranteed. As a result, an improved point cloud reduction algorithm is proposed. In the method, two complementary thresholds are used: the higher one is T_H, and the lower one is T_L, which can be selected flexibly according to different precision requirements. The former two steps are the same as the traditional one. After getting the small boxes, a plane is fitted to all of them. Then, the average distance D from all the points to the plane is calculated in each box. The corresponding operations are as follows:

   (i) If $D < T_L$, give up all the points in this box.

   (ii) If $T_L < D < T_H$, reserve one point nearest to the center of gravity.

   (iii) If $D > T_H$, the box will be divided into eight smaller boxes, whose length of side is half of the original one, and, then, reserve one point nearest to the center of gravity for all of the eight smaller boxes.

The result is shown in Figure 4(c). In this way, a few points are preserved in a relatively flat area to further simplify the calculation, while more points are preserved to obtain high precise models.

After point cloud reduction, the Delaunay triangulation process is applied to build a large number of small grids, as shown in Figure 5(a), and then mapping texture on every grid to reconstruct 3D model, as shown in Figure 5(c). Though long triangle sides are suppressed in order to distinguish different objects, errors still exist between some near models; see the red ellipse parts in Figures 5(a) and 5(c). In order to overcome this problem, different models should be clustered first by point density distribution. The interval between two models is used to classify different parts because no matched points can be detected in these areas and the point density is zero. The process is as follows.

Step 1. Define the point cloud as M, where every point named as $m_i \in M$ ($i = 1, 2, 3, \ldots$), and the corresponding space coordinate is ($m_{i\_}x, m_{i\_}y, m_{i\_}z$).

Step 2. Select one point m_j randomly and count the number N in its r-neighborhood points. Set $N = 0$, and then, for every point m_i ($i \neq j$) in M, if m_j satisfies (2), N plus one is as follows:

$$\left| m_{i\_}x - m_{j\_}x \right| < r \cap \left| m_{i\_}y - m_{j\_}y \right| \\ < r \cap \left| m_{i\_}z - m_{j\_}z \right| < r. \quad (2)$$

The r-neighborhood points constitute another set Q.

Step 3. If N in Q is less than threshold T, the point m_j selected in Step 2 is inappropriate; then return to Step 2 and choose a new m_j. Otherwise, the point m_j can be chosen as a core

(a)   (b)

FIGURE 6: 3D model stitching effect: (a) the initial model and (b) after stitching.

point, and perform the following Step 4. The threshold T is an experience according to the spatial distribution of point cloud, which is various for different cases, but the same is for different models from one case.

Step 4. When a core is found, then find r-neighborhood points for all other elements in Q. If one point in M meets (2), add this point to Q. Repeat this process until no new point can be added.

Step 5. The points in Q construct a new cluster. Delete all the points that belong to Q for M and repeat all the above processes to find other new clusters.

After clustering, the errors have been solved. The newly generated triangular mesh and 3D model are shown in Figures 5(b) and 5(d), respectively.

3.2. 3D Model Stitching. The photographic measurement system can reconstruct 3D model effectively, but a key problem remains because of limited field-of-view for normal camera. This can be reflected from Figure 6(a), where two 3D models captured in different positions are fused together primitively. Although each model can be reconstructed correctly, their coordinates are inconsistent, so how to reconstruct a large and integrated scene is a very important factor to be considered. For a rigid object, the attitude can be determined by three noncollinear points. As a result, finding more than three spatial matched points between different 3D models can achieve 3D model stitching. The process is described as follows.

Step 1. Get the Harris corners and stereo matching in order to calculate the point's 3D coordinates. The method is similar to Section 2.2.

Step 2. Feature point matching between different 3D models. First, the common correlation method is applied. Then, calculate the distance DL_{ij} $(i \neq j)$ for any two points in one

3D model and the distance DR_{ij} $(i \neq j)$ for their matched points in another 3D model. The error is

$$D_{ij} = \left| DL_{ij} - DR_{ij} \right|. \tag{3}$$

When D_{ij} has a minimum value, these two matched points are chosen as referential points. Last, check the other matched points. Calculate the distances between one matched point and referential points; the distance in one model is d_L and in the other model is d_R, if the distances between two referential points satisfy

$$\left| d_L - d_R \right| < T_d. \tag{4}$$

Reserve the matched point, or delete it. The T_d is a threshold that can control the match precision.

Step 3. Calculate the rotation parameters **R** and the displacement parameters **T**. Suppose the space coordinates of matched points in 3D models are $\mathbf{P}_{Li} = (X_{Li} \ Y_{Li} \ Z_{Li})'$ and $\mathbf{P}_{Ri} = (X_{Ri} \ Y_{Ri} \ Z_{Ri})'$; then the parameters can be computed as follows:

$$\mathbf{P}_{Ri} = \mathbf{R}\mathbf{P}_{Li} + \mathbf{T}. \tag{5}$$

Usually, more than three matched points can be detected, so the least square method can be applied to improve the computational accuracy. The stitching result is shown in Figure 6(b).

4. Experimental Results

To validate the effectiveness of our proposed 3D model reconstruction system, the experiment is conducted on the 172 cave in Mogao Grottoes, which is located 25 kilometers southwest of Dunhuang city, Gansu Province in China. In the meantime, another group has reconstructed the 159 cave by traditional 3D laser scanning technique. The new and traditional methods have been tested on different caves in the
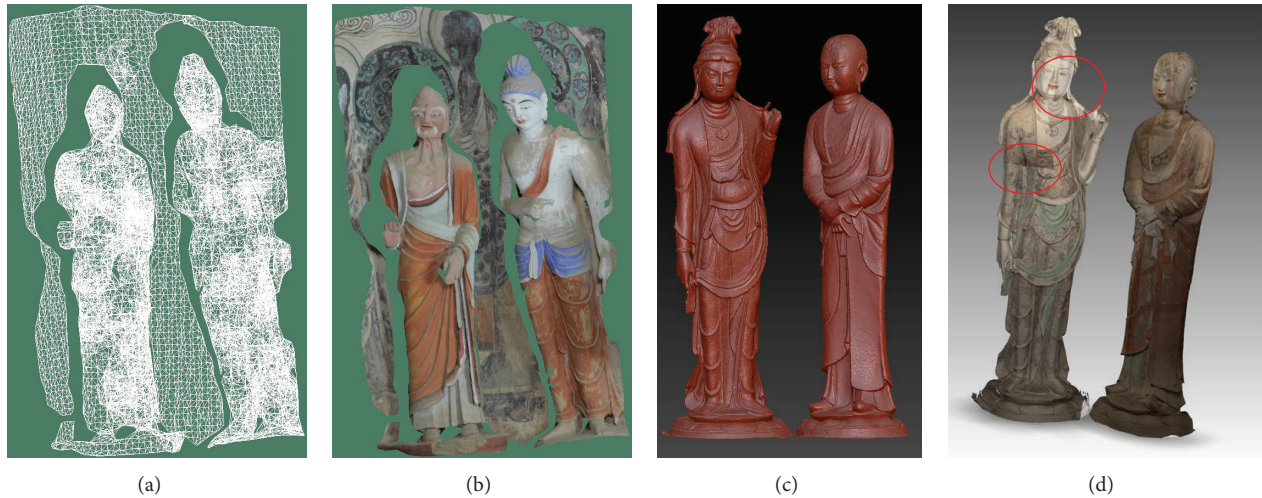
FIGURE 7: The 3D reconstruction results by our method and laser: (a) the triangular mesh by PFCPMS, (b) the 3D model reconstructed by PFCPMS, (c) the 3D model scanning by laser, and (d) the laser 3D model after texture registration.
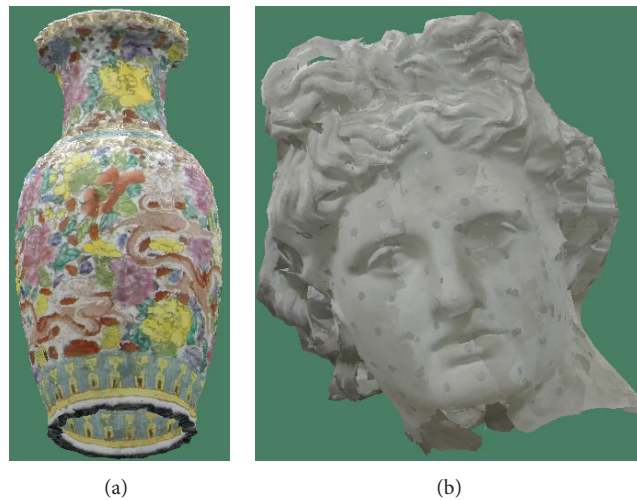

FIGURE 8: The real and touchable 3D models: (a) ornamental vase and (b) plaster head.

meantime because of the restricted condition of data sampling. Although the scenes are not exactly the same for these two methods, the characteristic, difficulty, and workload are similar. As a result, the contrast experiment between them is significant. A typical comparison result can be seen in Figure 7. Figures 7(a) and 7(b) are the triangular mesh and 3D model reconstructed by the proposed PFCPMS, respectively. Figures 7(c) and 7(d) are the 3D model scanning by laser before and after texture registration, respectively. While the laser scanning method performs well in reconstructing a high-precision 3D model by huge amounts of point cloud data, it is clumsy at rendering rich textures on the surface. The texture is discontinuous because of various illuminations on different surfaces, as shown in the red ellipse parts in Figure 7(d). By contrast, our method is superior to laser scanning in that lots of detailed textures can be effectively generated, leading to pleasing visual effect, which is very important in the virtual cultural heritage conservation field.

To quantitatively illustrate the effectiveness of this method, two real and touchable objects are reconstructed by a smaller PFCPMS. The 3D models can be seen in Figure 8. Figure 8(a) is an ornamental vase and (b) is a plaster head.

Since all points of the space coordinates are a relative value, the value of a single coordinate is not significant. So the reconstruction accuracy is evaluated by the space distance between two points. For each of the two models in Figure 8, five groups of random point distances are calculated by the space coordinates from the 3D models. The true value between two points is measured by two steel rules and an angle measuring instrument relying on triangulation principle. The measurement results are shown in Table 1.

All of the errors from Table 1 are less than 1 cm, and the relative error is less than 10%.

In Section 3.2, we use a 3D model stitching method to reconstruct a large and integrated scene using several 3D models from different perspective views. With different

TABLE 1: Quantitative analysis of reconstruction results (unit: meters).

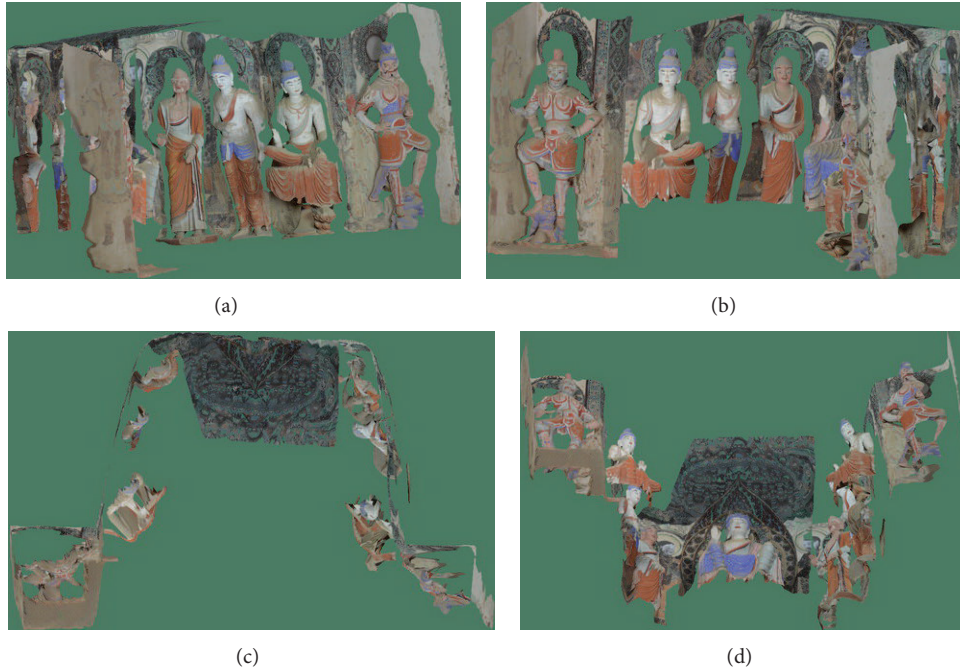|	Ornamental vase					Plaster head				
Measured value	0.1342	0.1282	0.3438	0.1388	0.2538	0.1610	0.1233	0.0835	0.1303	0.1095
True value	0.1281	0.1352	0.3340	0.1290	0.2480	0.1700	0.1180	0.0765	0.1230	0.1185
Error	0.0061	−0.007	0.0098	0.0098	0.0058	−0.009	0.0053	0.0070	0.0073	−0.009

(a)

(b)

(c)

(d)

FIGURE 9: The integrated 3D model: (a) left view, (b) right view, (c) top view, and (d) bottom view.

rotation and displacement parameters, we can stitch all the 3D models together for providing a large range of 3D view. The stitched result is shown in Figure 9. We can observe that the 3D scene can be displayed from any perspective. In Figure 9, four representative views are demonstrated. Because all the points in this model have 3D spatial information, the model not only can be preserved, but also enables to provide precise geometry information for reproduction.

During the 3D reconstruction process, the most time consuming procedure is Delaunay triangulation. The original point cloud has 1257305 points; the computation speed is too slow that is impossible to calculate. After point reduction and clustering, only 41649 points are preserved, and the 3D reconstruct process spends about half an hour.

5. Conclusion

An effective automatic 3D reconstruction method using PFCPMS has been proposed. The method has been tested on the 172 cave in Mogao Grottoes and compared with the 3D model scanning by laser. It indicated that the proposed method is effective to reconstruct a 3D scene with a low-cost four-camera photographic measurement system.

The limitation of the system is that no good quantitative evaluation standard has been developed. A rough idea is

evaluated performance by LIDAR scans. But the cost is so high that is against our design philosophy. Future research is to investigate an appropriate method.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] A. Koutsoudis, B. Vidmar, G. Ioannakis, F. Arnaoutoglou, G. Pavlidis, and C. Chamzas, "Multi-image 3D reconstruction data

evaluation," *Journal of Cultural Heritage*, vol. 15, no. 1, pp. 73–79, 2014.

[2] Y. Jiang and J. Tang, "Application of 3D stereo visualization system in virtual architectural design," *Applied Mechanics and Materials*, vol. 174–177, pp. 1706–1709, 2012.

[3] A. K. Aijazi, P. Checchin, and L. Trassoudaine, "Handling occlusions for accurate 3D urban cartography: a new approach based on characterization and multiple passages," in *Proceedings of the 2nd International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT '12)*, pp. 112–119, IEEE, Zürich, Switzerland, October 2012.

[4] G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, and C. Chamzas, "Methods for 3D digitization of cultural heritage," *Journal of Cultural Heritage*, vol. 8, no. 1, pp. 93–98, 2007.

[5] N. Ibrahim, K. A. Azmi, F. H. M. Salleh, and S. Yussof, "Cultural heritage preservation: 3D modeling of traditional Malay House using hidden surface removal approach," in *Proceedings of the International Conference on Software Engineering and Computer Systems*, Academic, 2009.

[6] H. Baltzakis, A. Argyros, and P. Trahanias, "Fusion of laser and visual data for robot motion planning and collision avoidance," *Machine Vision and Applications*, vol. 15, no. 2, pp. 92–100, 2003.

[7] C. Je, S. W. Lee, and R. H. Park, "High-contrast color-stripe pattern for rapid structured-light range imaging," in *Proceedings of the 8th European Conference on Computer Vision*, vol. 3021 of *Lecture Notes in Computer Science*, pp. 95–107, Springer, 2004.

[8] F. Blais, "Review of 20 years of range sensor development," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 231–243, 2004.

[9] G.-H. Liu, X.-Y. Liu, and Q.-Y. Feng, "High-accuracy three-dimensional shape acquisition of a large-scale object from multiple uncalibrated camera views," *Applied Optics*, vol. 50, no. 20, pp. 3691–3702, 2011.

[10] S. Agarwal, Y. Furukawa, N. Snavely et al., "Building rome in a day," *Communications of the ACM*, vol. 54, no. 10, pp. 105–112, 2011.

[11] Z. Ren, J. Liao, and L. Cai, "Three-dimensional measurement of small mechanical parts under a complicated background based on stereo vision," *Applied Optics*, vol. 49, no. 10, pp. 1789–1801, 2010.

[12] S. D. Zhong and Y. Liu, "Portable four-camera three-dimensional photographic measurement system and method," 2013, http://worldwide.espacenet.com/publicationDetails/biblio?CC=CN&NR=102679961B&KC=B&FT=D.

[13] V. J. D. Tsai, "Delaunay triangulations in TIN creation: an overview and a linear-time algorithm," *International Journal of Geographical Information Systems*, vol. 7, no. 6, pp. 501–524, 1993.

[14] Y. Weng, D. Li, and Y. Tong, "Texture mapping subdivision surfaces with hard constraints," *Visual Computer*, vol. 29, no. 11, pp. 1231–1241, 2013.

[15] S. Xu, D. Xu, and H. Fang, "Stereo matching algorithm based on detecting feature points," in *Materials Science and Information Technology*, pp. 6190–6194, Academic, 2012.

[16] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the Alvey Vision Conference*, pp. 189–192, The Plessey Company, 1988.

[17] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 87–119, 1995.

[18] W. Song, S. Cai, B. Yang, W. Cui, and Y. Wang, "A reduction method of three-dimensional point cloud," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09)*, pp. 1–4, IEEE, October 2009.

[19] H. P. Kriegel, M. Petri, M. Schubert, M. Shekelyan, and M. Stockerl, "Efficient similarity search on 3D bounding box annotations," in *Medical Imaging 2012: Advanced PACS-based Imaging Informatics and Therapeutic Applications*, vol. 8425 of *Proceedings of SPIE*, 2012.