

## **Research Article**

# **3D Reconstruction and Measurement Analysis of a Dense Point Cloud Fused with a Depth Image**

## Yujing Qiao <sup>[b]</sup>,<sup>1</sup> Ning Lv,<sup>1</sup> and Siyuan Zhang<sup>2</sup>

<sup>1</sup>School of Mechanical Engineering, Yangzhou Polytechnic College, Yangzhou 225009, China <sup>2</sup>Intelligent Machine Institute, Harbin University of Science and Technology, Harbin 150080, China

Correspondence should be addressed to Yujing Qiao; qiaoyujing@sina.com

Received 24 March 2023; Revised 10 August 2023; Accepted 23 August 2023; Published 8 September 2023

Academic Editor: Haochong Huang

Copyright © 2023 Yujing Qiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To solve the problems of holes, noise, and texture information missing in the traditional incremental reconstruction of complex surface objects, a 3D reconstruction method of depth image fusion surface dense point clouds is proposed, and texture feature creation is combined to obtain a 3D reconstruction model that takes into account the main body and details of the reconstructed object. First, the mechanism of surface dense reconstruction based on the patch-based multiview stereo (PMVS) algorithm is analyzed. Combined with the principle of view angle selection of stereo images, surface point cloud density reconstruction is performed. Then, the depth value is optimized by the region growing method, and the optimization model is established. The depth image is fused into a dense surface, and the reconstructed part is supplemented by the depth information. Finally, the Markov random field (MRF) is introduced to describe the richness of image details, and combined with the calculating method of the area coordinate, the texture coordinates are accurately calculated to reproduce the texture details of the 3D reconstruction model. 3D reconstruction experiments are performed on multiple indoor and outdoor model surfaces, and the experimental results show that the proposed method can achieve complete and accurate reconstruction of complex surface objects. Our method provides technical support for complex surface topography detection and has industrial practical significance.

### 1. Introduction

3D reconstruction is always the most important step in accurate 3D topography measurement. With the development of computer technology and industrial technology, methods of capturing 3D information have become more feasible and have been applied in 3D topography measurement, such as the classical methods of capturing 3D information, laser 3D scanners, and depth cameras [1-3]. However, these active projection technologies have high requirements for illumination, the high price of 3D scanners, and the limited resolution of consumer-level depth cameras. Due to the depth, noise, or the complexity of the special surface, it is difficult for the scanner to completely restore the 3D structure of the surface during 3D reconstruction for the weak texture non-Lambert surface and complex surface, resulting in holes in the model, which cannot reproduce the surface details well. 3D scene reconstruction technology

based on multiview stereo (MVS) has better performance in terms of the reliability and adaptability of indoor and outdoor scene reconstruction [4–6]. Especially with the increasing resolution and decreasing cost of digital cameras, dense 3D reconstruction based on multiview images has become a hot topic in computer vision research. For example, many algorithms represented by KinectFusion can better restore the dense 3D model of the Lambert surface through RGB-D data collected by consumer cameras [7]. However, this method is not suitable for the 3D reconstruction of large outdoor scenes. Complex surface objects are blocked by surrounding objects, and there are still some difficult problems to be solved, such as the blocked part being difficult to convert from two-dimensional images to 3D information [8, 9].

In view of the key issues of 3D reconstruction, relevant research is mainly indicated by factors such as the intelligent control of the number of feature matching points, accurate camera calibration, and point cloud densification; the accuracy and speed of model reconstruction have also been improved over time. Jhony et al. used a new registration method of 2D-3D-free deformation transformation to improve the model coincidence degree, which has a better reconstruction effect on objects with a single texture or smooth surface [10]. Prakoonwit et al. studied a new method to quickly obtain the surface reconstruction of a 3D model from a small amount of data and predicted whether the fitted coordinate points contained features through the optimal distribution of landmark points [11]. This method has the best effect for simple small targets and has been widely used in 3D face reconstruction [12, 13]. Sohaib et al. propose a polarization-based photometric stereo vision shape recovery algorithm, which can effectively extract 3D surface information for some complex parts with concave surfaces [14]. Polarized 3D imaging technology has developed rapidly due to its high accuracy, long operating distance, and small impact from stray light. However, the problem of using the polarization characteristics of target reflected light to solve the normal vector accurately has not been truly solved, which has become the bottleneck restricting the development of this technique [15, 16]. The authors in [17] designed a confidence propagation algorithm based on gray level similarity probability to solve the problems of weak texture, shadow, and depth discontinuity. This method has an excellent effect on the 3D point cloud densification of objects in indoor environments. Qiao et al. proposes a 3D reconstruction hole repair algorithm based on structure from motion (SFM) multiview 3D point cloud organic fusion, which effectively solves the problem of hole repair on the surface of objects [18]. Currently, the widely developed 3D reconstruction method based on deep learning improves the robustness of stereo matching and the reconstruction accuracy, but it still does not solve the problem of holes in the process of 3D reconstruction [19, 20].

Through the abovementioned analysis, the existing image reconstruction methods have made great progress in improving the reconstruction accuracy and reducing the error. However, when reconstructing objects with weak textures and complex surfaces, the 3D point cloud reconstructed by the existing algorithms is relatively sparse, resulting in missing details of the reconstructed object model. At present, the solutions to the problem of 3D reconstruction of weakly textured objects can be basically divided into three types. The first is to improve the reconstruction quality by optimizing a certain step in 3D reconstruction or one of the processes in the whole set of algorithms. However, as the current 3D reconstruction methods are all a set of systematic theories, a single optimization of its local effect is less likely to improve the overall effect. The second is to combine some advanced intelligent algorithms in many traditional 3D reconstruction modes. Such schemes often achieve good results in academic experiments, but the actual industrial vision detection environment is complex and changeable, and the intelligent algorithm is not robust to such external factors. The third is some hardware products developed by some companies, which can indeed improve the reconstruction accuracy.

However, there are a wide variety of industrial products with different sizes, so whether they have perfect compatibility is the weakness of these products. Therefore, if the above factors can be taken into account in the 3D reconstruction of weak texture objects, it will essentially improve the reliability and practicability of reconstruction. The depth image obtained by the depth camera contains depth information, which can be used to realize 3D reconstruction by registration and splicing of multiview depth images in theory and practice. However, the accuracy of the registration results obtained by the existing consumer depth cameras cannot meet the measurement requirements. Other problems, such as low registration accuracy and blurred reconstruction results, are still inevitable in the realization of 3D reconstruction by the binocular stereo image stitching method.

In summary, this paper proposes a comprehensive 3D reconstruction method that takes into account both the main body and details of the reconstructed object. Especially for high-light weak texture surfaces and outdoor large structures, we proposed a 3D reconstruction method and multiview surface technology based on depth image fusion to obtain a 3D reconstruction model with high quality and high precision to further realize accurate 3D measurements.

The contributions of this work can be described as follows:

- (i) The regional growth expansion method is used to nonlinearly optimize the depth value, and the optimization model is established. The depth image was fused to the dense surface, and the missing part was reconstructed.
- (ii) Markov random field (MRF) is introduced to describe the texture details, and combined with the accurate calculation of texture coordinates, the texture details of the 3D reconstructed model can be reproduced. Finally, complete reconstruction of a 3D surface with rich texture information is realized.

## 2. Multiview Dense 3D Reconstruction Mechanism Based on Depth Image Fusion

The reconstruction points provided by feature matching are naturally not dense in the 3D reconstruction by the motion recovery structure of the camera. Therefore, to obtain dense point clouds, first, the patch-based multiview stereo (PMVS) is adopted to improve the density of point clouds [21]. In view of the limitations of this method in obtaining depth information, the depth image fusion reconstruction method is used to complete the unreconstructed part that is reconstructed based on the PMVS algorithm, and finally, a complete 3D model is obtained.

2.1. Dense 3D Reconstruction of a Multiview Point Cloud Based on PMVS. The patch-based multiview system (PMVS) algorithm first assumes some 3D rectangular patches in space and makes the patches cover the object surface by some regular expansion method. In principle, dense reconstruction using the PMVS algorithm can match almost every pixel in the photo and reconstruct the 3D coordinates of each pixel. As shown in Figure 1, the patch is a local tangent plane approximate to the object surface, including the center c(p), normal vector n(p), and reference image R(p). In addition to the corresponding reference image R(p), each patch p also corresponds to two image sets V(p)and V'(p), where the initialized V(p) represents the set with an angle less than 60° between the patch normal vector and the ray where the patch is located. V'(p) represents a set with a normalized correlation coefficient (NCC) greater than 0.4 between the patch projected onto the image and the reference image. The p in V(p) is actually visible, while the V'(p) in the patch cannot be recognized due to highlights, motion blur, or self-occlusion. R(p) is a subset of V(p).

The implementation process of PMVS includes three steps: feature matching initialization, patch diffusion, and patch filtering.

Step 1. Feature matching initialization. By allowing the polar line constraint with two-pixel errors to find the same type of feature points in other images to form a pair of matching points, the feature matching method is used to generate a series of 3D space points from these matching point pairs, arrange these points in the order of distance from small to large, and then try to generate patches one by one.

Step 2. Patch diffusion. Dense patches are obtained by diffusion of sparse seed points, and the goal is to have at least one patch in each mesh. The new patches are repeatedly generated through the aforementioned patches; specifically, given a patch p, a set of neighborhood image blocks C(p) satisfying certain conditions is obtained first, and then the patch generation process is carried out.

Step 3. Patch filtering. In the process of patch reconstruction, some patches with large errors may be generated, so filtering is required to ensure the accuracy of patches. The first filter implements filtering through visual constraints, and the second filter also considers stricter visual constraints. The third filter, for each patch p, in V(p), needs to map to the image block of patch p itself and all adjacent image blocks to collect such a set of patches. If the ratio between the number of patches in the eight neighborhoods of patch p and the number of patches collected is less than 0.25, then p is considered an abnormal value and is filtered out. Thus, the point cloud of patch densification based on the PMVS algorithm has been completed.

We take stainless-steel wire pair part with complex surfaces as the research object. Its multiview image sequence is shown in Figure 2(a). After the point cloud is densified using PMVS algorithm, iterative reconstruction is performed using the incremental structure from motion. The iterative reconstruction result is shown in Figure 2(b). There are obvious defects and holes in the reconstructed model.

The results indicate that although the PMVS-based method can achieve a relatively dense 3D model reconstruction, the passive method of acquiring depth



FIGURE 1: Schematic diagram of the patch and related parameters.

information by using multiview image acquisition has limitations; that is, the reconstructed model will have serious defects in the case of non-Lambertian surfaces or objects with occlusion.

2.2. Detail Reconstruction Based on Depth Image Fusion. A depth camera is an active method to acquire depth information of objects by actively projecting structured light. Although the point cloud image obtained by the depth camera contains depth information, due to its low resolution and low matching accuracy, it is difficult to achieve accurate three-dimensional reconstruction by using the multiview depth point cloud image method. However, a depth camera can be used to collect locally missing depth information.

2.2.1. The Vision Angle Selection of Stereoscopic Images. 3D point clouds can be reconstructed from multiple unordered images, but the selection of the spacing between images is related to the reconstruction quality, so it is necessary to make a more standard selection of the vision angle. The vision angle selection of the stereo image includes two parts: global vision angle and local vision angle.

There are two specific principles for selecting a global vision angle:

- (1) According to overlapping scenes and resolution between images, the image with the same scale should be selected as much as possible during matching. Otherwise, the patch will be too large or too small, and the calculation speed will be affected.
- (2) If there is a wide baseline between images, an appropriate length of baseline should be selected without affecting the reconstruction accuracy. The selection criteria are shown in the following formula:

$$g_R(V) = \sum_{f \in F_V \cap F_R} \omega_N(f) \omega_S(f), \tag{1}$$

where *R* represents the reference image, *f* represents the feature points reconstructed from 3D, and  $F_R$  and  $F_V$  represent the selected visual angle. The selected visual angle is evaluated by (1).  $\omega_s$  is the evaluation coefficient of the current image scale, and  $\omega_N$  is the evaluation coefficient of the angle between the lines of the adjacent images. The reasonable selection of the vision angle is determined by summing the



FIGURE 2: Dense 3D reconstruction process of the stainless-steel wire surface. (a) Image sequence. (b) The 3D reconstruction results.

coefficients of multiple vision angles. The higher  $g_R(V)$  is, the more appropriate the selected vision angle is.

The selection of the local viewing angle refers to taking the current global viewing angle as a candidate vision angle. The selection criteria mainly include the following two points:

- Determine candidate vision angles by NCC values: calculate the NCC values of every two candidate vision angles, and the viewing angle with a larger NCC value is determined.
- (2) The visual line shall be sufficiently scattered to ensure that it is at least not coplanar. Each vision angle corresponds to a visual line, which is calculated by the angle between the polar planes, as shown in (2), where  $\omega_e$  is the included angle of visual lines, so it is better to choose some scattered vision angles.

$$l_{R}(V) = g_{R}(V) \cdot \sum \omega_{e} \left( V, V' \right).$$
<sup>(2)</sup>

2.2.2. Depth Calculation and Optimized Fusion of Depth Image. The overall framework of the depth image fusion method mainly adopts the region growth method to expand [20], and the corresponding description is as follows:

Step 1. The priority queue is established according to the reconstructed confidence level

Step 2. Depth estimation from initial sparse feature points

Step 3. Nonlinear depth optimization was carried out for each seed point

Step 4. After each optimization, neighborhood pixels are added to the queue by judging the following two conditions: (1) there is no depth value in the neighborhood; (2) the confidence of the current pixel is higher than that of the neighboring pixels in a certain range.

First, nonlinear optimization of the depth value was carried out to establish an optimization model, as shown in Figure 3.  $O_R$  represents the center of the camera and follows

the visual line intersecting with the image at the center pixel. The pixel is used as the center to build the corresponding patch. The definition of the depth image patch is slightly different from that of the PMVS patch. The definition of a depth image patch is on the pixel of the reference image, which corresponds to a tiny plane in the space and represents the depth and normal vector of the pixel. The initial depth value of the pixel point (*s*, *t*) in the reference vision angle is h(s, t), and the unit vector of the ray in the corresponding three-dimensional space is  $r_R(s, t)$ . Then, the point in the 3D space corresponding to the pixel is shown in the following formula:

$$X(s,t) = O_R + h(s,t)r_R(s,t).$$
(3)

To optimize the depth at the pixel (s, t) and the normal vector in 3D, two variables  $h_s(s,t)$  and  $h_t(s,t)$  are introduced on the patch with the pixel as the center to represent the 3D coordinates of each pixel in the patch. Then, the depth corresponding to the pixel (s + i, t + j) is as shown in the following formula:

$$h(s+i,t+j) = h(s,t) + i * h_s(s,t) + j * h_t(s,t).$$
(4)

Assume that the ray direction at (s+i, t+j) is approximately  $r_R(s,t)$ ; then, the three-dimensional coordinates corresponding to the pixel (s+i, t+j) are shown in the following formula:

$$X(s+i,t+j) = O_R + r_R(s,t) (h(s,t) + i * h_s(s,t) + j * h_t(s,t)).$$
(5)

At this point, the calculation about the depth of pixels in the depth map is complete. The same as PMVS algorithm, the pair perspective depth image is merged into the neighborhood perspective by consistency constraint and visibility constraint, and the depth of all pixels is merged finally to complete the depth information recovery of the whole object.

2.3. Surface Texture Reproduction. Although the reconstruction effect of the depth image fusion dense point cloud method is relatively complete, it basically retains the



FIGURE 3: Nonlinear optimization of depth values.

three-dimensional morphological features of the object. However, the model surface is not smooth enough, and there are some redundant patches only through the reconstruction of the point cloud. Moreover, even after the optimization calculation of error compensation, such as camera calibration, the final reconstructed model, and the actual object cannot completely coexist. Therefore, the coordinates need to be recalculated with the reconstructed model, and the texture can be covered on the model as accurately as possible after one-to-one correspondence with the original object.

Before calculating texture coordinates, the Markov random field (MRF) [22, 23] is used to describe the richness of image details, and its expression is shown in the following formula:

$$E(l) = \sum_{F_i \in Faces} E_{data}(F_i, l_i) + \sum_{(F_i, F_j) \in Edges} E_{smooth}(F_i, F_j, l_i, l_j),$$
(6)

where  $E_{data}$  is the data item,  $E_{smooth}$  is the smooth item, and l is the optimization quantity tag. MRF is built on the patches of triangular mesh, and each patch corresponds to a vertex. Its model is shown in Figure 4, where  $F_i$  represents the patch and  $E_{data}$  represents the cost that label  $l_i$  assigned to patch  $F_i$ . The whole formula requires the minimum cost, so the minimum value of E(l) is needed.

The texture details to be considered in the data items are reflected in the average gradient and scale of the projected triangle, that is, the area of the projected triangle.  $E_{data}$  is described in the following formula:

$$E_{\text{data}} = -\int_{\Phi} (F_i, l_i) \left\| \nabla (I_{l_i}(p)) \right\|_2 \mathrm{d}p, \tag{7}$$

where  $l_i$  stands for the angle of view,  $\nabla (I_{l_i}(p))$  is the gradient at the projection point, and  $\Phi (F_i, l_i)$  is the projection area of a triangle. The richer the texture, the stronger the gradient response; in contrast, the gentler the gradient response. The next step is the calculation of smoothing terms, as shown in (8). If  $F_i$  and  $F_j$  are adjacent and labels l are inconsistent, a penalty constraint is added between them but not added if labels are consistent. Therefore, the purpose of the smoothing item is to keep the same angle of view between the adjacent patches, which can connect the textures in the mesh into pieces.

$$E_{\text{smooth}} = \left[ l_i \neq l_j \right]. \tag{8}$$

After describing the texture details, the texture coordinates are calculated. The texture coordinates are calculated by area coordinate calculation, and the relationship between the projection of the triangular surface in the visual angle is found from the texture image to be created and then used to determine the correspondence between space and area. As shown in Figure 5, the area coordinate is the ratio of three small triangles formed by these four points to the area of the whole large triangle. The three dimensions add up to 1, and each dimension is between 0 and 1. After converting to the point pixel coordinates, the values can be assigned to the triangular patch in the grid, and the texture is covered to obtain the model with texture.

#### 3. 3D Dense Reconstruction Process and Analysis of Depth Image Fusion

Taking the image sequence in Figure 2 as an example and using the vision angle selection principle, the results shown in Figure 6 are obtained through camera pose recovery and point cloud densification. It can be seen from Figure 6 that the spatial position of the camera is above the dense point cloud when 50 pictures are taken.

3.1. 3D Reconstruction by Depth Image Fusion. After the point cloud denseness experiment is implemented, the depth camera is used to shoot and reconstruct according to the abovementioned theory to solve the problem that the non-Lambert plane or object self-occlusion cannot be reconstructed in incremental reconstruction. The experimental equipment is an Astra prodepth camera, as shown in Figure 7(a). The camera depth range is 0.6 m to 8 m, the





FIGURE 6: Effect drawing of dense point cloud.

resolution of the color image is 1280 \* 720@30 FPS, the resolution of the depth image is 1280 \* 1024@7 FPS, and the accuracy is +-1-3 mm@1 m. To solve the problem of non-Lambertian surface mismatch, laser points are projected onto the captured object, as shown in Figure 7(b) [18]. The depth image obtained after light averaging and smoothing is shown in Figure 8. Here, the Figure 8(a) indicates the depth image detail after light averaging and smoothing processing.

The visualization effect of point cloud in the depth image is shown in the Figure 8(b).

After the depth image point cloud is obtained, the thread details that are not reconstructed in the incremental reconstruction are taken, and the MeshLab open-source software is used to integrate and add them to the point cloud of the main part, as shown in Figure 9. The approximate curved part is the dense point cloud part of the



FIGURE 7: Experimental equipment of the depth camera and laser projection. (a) Depth camera; (b) laser projection.



FIGURE 8: Depth image after light averaging and smoothing processing. (a) Depth image details after averaging and smoothing; (b) visualization of point cloud in depth map.



FIGURE 9: The effect after the fusion of the main dense point cloud and the detailed depth point cloud.



FIGURE 10: Dense reconstruction effect after point cloud fusion.

stainless-steel wire, and the point part is the supplementary detail part. It can be seen that the point cloud is well supplemented in the non-Lambert plane and self-occlusion area of the object. Figure 10 shows the dense reconstruction effect after point cloud fusion.

To further test the reconstruction effect of our method on objects with weak textures, the standard test image Tsukuba was used as the experimental image to illustrate the effectiveness of the method. In the standard test image experiment, a parallax map was used instead of a depth map as the control, and the control group was the graph cuts (GC) algorithm, as shown in Figure 11(a), the algorithm in studies of [24], as shown in Figure 11(b), and the algorithm in this paper, as shown in Figure 11(c). In the figure, (i) is the original image, (ii) is the disparity image, and (iii) is the reconstructed image. The results are shown in Figure 11 (for intuitive expression, the background was appropriately filtered during reconstruction).

It can be seen from the experimental results that compared with the control group, the method in this paper essentially retains most of the original object features and is more effective than the other two algorithms at compensating for hole defects.



FIGURE 11: Comparison of original images, disparity maps, and reconstruction maps of Tsukuba images. (a) GC algorithm, (b) the algorithm of literature [24], and (c) the algorithm of this paper. (i) The original image. (ii) Disparity map. (iii) Reconstructed image.

3.2. Reconstruction Experiment of Indoor and Outdoor Objects. To further analyze the effectiveness and robustness of the proposed reconstruction method, image sequences of multiple groups of objects were taken. In addition to stainless steel wire, three objects, as shown in Figure 12, were also selected for the experiment. Finally, several sets of reconstruction models of indoor objects are obtained, as shown in Figure 13. It can be seen that the complete reconstruction is realized, and the texture of the reconstructed model is well reproduced.

To reflect the robustness of the reconstruction method, surface reconstruction is performed again in the outdoor scene. However, because the light intensity of natural light in outdoor scenes is much greater than that of infrared speckle emitted by depth cameras, the shooting effect of depth cameras is extremely unsatisfactory. Since the disparity image has a linear relationship with depth, the depth image under each viewing angle takes two pictures with a binocular camera, and a composite disparity image is used to replace the depth image. The selected object for outdoor reconstruction is shown in Figure 14. Using a binocular camera, two images were taken after calibration and stereoscopic correction, and as an example, the Figure 14(c) shows the sequence image of a school motto stone. Taking statue in the Figure 14(a) as an example, the SGBM algorithm in the open-source computer vision library OpenCV is used to obtain the disparity image and convert it into the depth image, as shown in Figure 15. Due to the complex outdoor environment and excessive noise points, the effect of the parallax map is not very good, but the depth information of the detailed part is relatively complete after it is converted into the depth image.

After reconstruction according to the abovementioned method, the final three-dimensional model obtained is shown in Figure 16.

To accurately verify the reconstruction accuracy of the method in this paper, the vanishing point and the projective geometry are used to measure the size, surface area, and volume of the target object in the image [25]. Table 1 shows



FIGURE 12: Indoor reconstruction experiment object.

the comparison between the measured parameters after reconstruction and the actual parameters of the object itself. The statue and school motto stone are not compared in the table because of their huge size. The size error measured after reconstruction is controlled within  $1\sim 2$  mm.

To evaluate the effectiveness of our method more objectively, two performance indicators, the peak signal-tonoise ratio (PSNR) and structural similarity index (SSIM), are selected for evaluation, as shown in Figure 17. It can be seen from the figure that the PSNR value of the object reconstructed by our method is less than that of the traditional SFM method to varying degrees, indicating that the reconstruction of the proposed method has less noise, higher robustness, and better quality and can effectively describe the characteristics of the object, while the SSIM value is greater than that of the traditional SFM method, indicating that the model reconstructed by the proposed method is more complete and more similar to the original object. More information about object morphology is retained. In addition, it can be seen from the figure that the PSNR value of the T-joint in our method is the smallest, and the SSIM value is the largest, indicating that the size of the reconstructed object itself cannot effectively affect the reconstruction effect. Because the structure of the T-joint is relatively complex and there are many occurrences on the surface, the surface



FIGURE 13: Four kinds of object reconstruction in indoor environments: (a) reconstruction result of stainless a bolt joint; (b) reconstruction result of a pagoda-shape joint; (c) reconstruction result of a T-joint; and (d) reconstruction result of a flower pot.



FIGURE 14: Outdoor objects with weak texture to be reconstructed. (a) Statue; (b) school motto stone; and (c) the sequence image of the school motto stone.

topography complexity of the object is an important factor affecting the effect of the method proposed in this paper, while the traditional SFM method does not have such characteristics. Therefore, the method in this paper is more anisotropic, and it is more accurate for use in actual industrial vision measurements.



FIGURE 15: Disparity image and depth image of the human sculpture. (a) Disparity image and (b) depth image.



FIGURE 16: 3D reconstruction model of an outdoor object with weak texture. (a) 3D reconstruction model of the statue and (b) 3D reconstruction model of school motto stone.

TABLE 1: Measurement	parameters an	nd actual	parameters	of the	object	surface	after	reconstruction

Reconstructing objects	Dimension measurement (mm)	Dimension actual value (mm)	Area measurement (mm <sup>2</sup> )	Actual value of area (mm <sup>2</sup> )	Volume measurement (mm <sup>3</sup> )	Volume actual value (mm <sup>3</sup> )
Butt joint	31.046	30	3909.519	3972.584	31822.010	31655.506
Pagoda joint	39.656	40	52018.742	51922.747	24015.465	23181.982
T joint	55.968	55	5749.830	5675.966	26709.902	25988.520
Flower pot	136.301	137	90010.088	89929.651	2188359.630	2189257.363



FIGURE 17: Comparison of PSNR and SSIM of different reconstructed objects.

#### 4. Conclusions

In this paper, a dense 3D reconstruction method based on depth image fusion with multiple views is proposed, which takes into account both the main body and texture details of the reconstructed object and obtains a 3D reconstruction model with high quality and relatively high precision. The depth image is used to fuse the multiview dense point cloud obtained by the PMVS algorithm to obtain a complete 3D model. Combined with the reconstructed model, the texture coordinates are calculated, and a 3D surface with rich and accurate texture information is obtained. To verify the effectiveness and robustness of our method, a variety of indoor and outdoor objects were selected for reconstruction. The results show that the reconstruction accuracy can be controlled to within 1~2 mm, the structural similarity of the model is 90%, and the performance indices meet the measurement requirements. The method is also anisotropic, and its reconstruction effect is less affected by the size of the reconstructed object and more affected by the morphological complexity of the object itself. The complete 3D reconstruction method proposed in this paper has certain technical support for 3D measurement.

Because the method proposed in this paper requires image fusion and reconstruction from multiple cameras, it is difficult to unify the image parameters. How to quickly unify the parameters of pictures taken in the early stage is the primary problem to be solved in subsequent research. In addition, further studies can be performed to combine the proposed method with deep learning to achieve highprecision 3D reconstruction of large and complex structures.

#### **Data Availability**

The data from the 3D reconstruction section of this article used to support the findings of this study are available from the corresponding author upon request.

## **Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

#### Acknowledgments

This paper was supported by the Yangzhou City, 2021, leading talents of the "Green Yang Jinfeng Project" (Innovation in Colleges and Universities) (Grant no. YZLYJFJH2021CX044) and the Science and Technology Talent Support Project of Jiangsu Province, China (Grant no. FZ20221360).

#### References

- J. Wang and B. Li, "Analysis of aero-engine blade surface laser point cloud 3D reconstruction error," *Acta Photonica Sinica*, vol. 47, no. 05, pp. 187–195, 2018.
- [2] P. Yang, X. Gao, S. Lin, B. Ma, and C. Pan, "3D reconstruction of laser scanning at weld surface defects," *Transactions of the China Welding Institute*, vol. 41, no. 3, pp. 59–63, 2020.
- [3] G. Liu and J. Duan, "A deep image hole repairing method and a superpixel segmentation algorithm based on Kinect camera," *Computer Engineering & Science*, vol. 42, no. 5, pp. 851–858, 2020.
- [4] I. Kostrikov, E. Horbert, and B. Leibe, "Probabilistic labeling cost for high-accuracy multi-view reconstruction," in *Proceedings of the ieee conference on computer vision and pattern recognition*, pp. 1534–1541, Columbus, OH, USA, September 2014.
- [5] Y. Dai, Z. Zhu, Z. Rao, and B. Li, "Mvs2: deep unsupervised multi-view stereo with multi-view symmetry," in *Proceedings* of the 2019 International Conference on 3D Vision (3DV), pp. 1–8, Ieee, Quebec City, Canada, September 2019.
- [6] A. Kuhn, S. Lin, and O. Erdler, "Plane completion and filtering for multi-view stereo reconstruction," in *Proceedings of the Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019*, pp. 18–32, Dortmund, Germany, September 2019.
- [7] J. Li, W. Gao, Y. Wu, Y. Liu, and Y. Shen, "High-quality indoor scene 3D reconstruction with RGB-D cameras: a brief review," *Computational Visual Media*, vol. 8, no. 3, pp. 369–393, 2022.
- [8] X. Yunlong, L. Aihua, C. Zhigao, and F. Hao, "Moving target tracking algorithm based on improved Kernelized correlation filter," *Infrared and Laser Engineering*, vol. 45, no. S1, pp. 214–221, 2016.
- [9] L. Dong, Z.-W. Lu, and X.-Y. Liu, "Analysis and comparison of limit detection capabilities of three active synthetic aperture imaging techniques," *Chinese Optics*, vol. 12, no. 1, pp. 138– 147, 2019.
- [10] J. K. Pontes, C. Kong, A. Eriksson, C. Fookes, S. Sridharan, and S. Lucey, "Compact model representation for 3D reconstruction," 2017, https://arxiv.org/abs/1707.07360.
- [11] S. Prakoonwit, "3D reconstruction from few silhouettes using statistical models and landmark points," in *Proceedings of the*

2010 International Conference on Cyberworlds, pp. 120–125, IEEE, Singapore, October 2010.

- [12] E. Vezzetti, F. Marcolin, S. Tornincasa, L. Ulrich, and N. Dagnes, "3D geometry-based automatic landmark localization in presence of facial occlusions," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 14177–14205, 2018.
- [13] K. Khabarlak and L. Koriashkina, "Fast facial landmark detection and applications: a survey," 2021, https://arxiv.org/ abs/2101.10808.
- [14] A. Sohaib, A. R. Farooq, J. Ahmed, L. N. Smith, and M. L. Smith, "3D reconstruction of concave surfaces using polarisation imaging," *Journal of Modern Optics*, vol. 62, no. 11, pp. 927–932, 2015.
- [15] X. Li, F. Liu, and X. Shao, "Research progress on polarization 3D imaging technology," *Journal of Infrared and Millimeter Waves*, vol. 40, no. 2, pp. 248–262, 2021.
- [16] H. Luo, J. Cao, X. Gai, and Q. Ding, "Industrial vision based on polarization imaging and its key technologies," *Laser & Optoelectronics Progress*, vol. 59, no. 14, Article ID 1415003, 2022.
- [17] S. Gai, F. Da, L. Zeng, and Y. Huang, "Research on a hole filling algorithm of a point cloud based on structure from motion," *Journal of the Optical Society of America*, vol. 36, no. 2, pp. A39–A46, 2019.
- [18] Y.-J. Qiao, S.-Y. Zhang, and Y.-H. Zhao, "Surface robust reconstruction method for high light and weak textured objects," *Acta Photonica Sinica*, vol. 48, no. 12, Article ID 1212002, 2019.
- [19] G. Fahim, K. Amin, and S. Zarif, "Single-View 3D reconstruction: a Survey of deep learning methods," *Computers* & Graphics, vol. 94, pp. 164–190, 2021.
- [20] T. Samavati and M. Soryani, "Deep learning-based 3D reconstruction: a survey," *Artificial Intelligence Review*, vol. 56, no. 9, pp. 9175–9219, 2023.
- [21] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," *IEEE Transactions on Pattern Analysis* and Machine Intelligence, vol. 32, no. 8, pp. 1362–1376, 2010.
- [22] C. Wang, N. Komodakis, and N. Paragios, "Markov random field modeling, inference & learning in computer vision & image understanding: a survey," *Computer Vision and Image Understanding*, vol. 117, no. 11, pp. 1610–1627, 2013.
- [23] Z. Wu, D. Lin, and X. Tang, "Deep Markov random field for image modelling," in *Proceedings of the Computer Vision*-*ECCV 2016: 14th European Conference*, pp. 295–312, Amsterdam, Netherlands, October 2016.
- [24] Y.-K. Du and C. Liu, "Robust estimation method for dense disparity of binocular vision under textureless environment," *Optics and Precision Engineering*, vol. 25, no. 4, pp. 1086–1094, 2017.
- [25] A. Criminisi, "Single-view metrology: algorithms and applications," in *Proceedings of the Pattern Recognition, 24th DAGM Symposium*, Zurich, Switzerland, September 2002.