Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

Research on Optimization of Fine 3D Reconstruction Process Based on SfM-MVS

Zhuoyuan Wu *

Shenzhen Middle School, Shenzhen, China

* Corresponding Author Email: zhuik@ldy.edu.rs

Abstract. Aiming at the optimization problem of motion recovery structure and multi-view stereo (MVS) vision process, this study systematically explores the influence of shooting parameters on 3D reconstruction quality by establishing data acquisition standards, filling the gap of standardization research in this field. This study aims to quantify the effect of variables such as lighting conditions, shooting equipment, and background complexity on reconstruction accuracy and propose a scientific acquisition standard. Methods: Four groups of control experiments (control group and three experimental groups) were designed. Based on Reality Capture, the reconstruction effects under ambient light/point light source, professional camera/mobile phone, solid color/complex background, and other conditions were compared and analyzed, and the success rate of alignment, point cloud density, and other indicators were evaluated. The experimental results show that the alignment success rate of uniform lighting group is 78.7% (only 12.6% for point light source), the point cloud density is reduced by 40% due to automatic exposure of mobile phone, the alignment rate of complex background group can be restored to 60.5% through control point optimization, and the number of surfaces in the reflective area is constant to 50% in the non-reflective area. A hybrid optimization scheme combining traditional preprocessing (such as High Dynamic Range Imaging (HDR) correction) and neural network (such as Deep Neural Network for Image Denoising (DnCNN)) is further proposed.

Keywords: 3D Reconstruction; MVS; Data Acquisition; Lighting Conditions.

1. Introduction

Three-dimensional reconstruction technology refers to the acquisition of three-dimensional geometric information of objects by image or surveying means, and the creation of three-dimensional models for them. Currently, the mainstream reconstruction methods are divided into contact and non-contact, while non-contact is divided into active and passive. Among them, motion recovery structure (hereinafter referred to as structure from motion (SfM)) and multi-view stereo vision (MVS) constitute one of the mainstream methods of passive reconstruction [1]. SfM extracts feature points (such as Scale-Invariant Feature Transform (SIFT)) from disordered images, and restores camera pose and sparse point cloud by matching features; MVS generates dense point clouds by using multi-view stereo matching. This combination is often used successively in 3D reconstruction -- the SfM technique is used to analyze the shot Angle or trajectory to produce a sparse point cloud image, and the MVS technique is used to analyze the camera information further to create a dense point cloud image. Compared with the traditional active reconstruction methods, such as the laser scanning method and the structured light method, the cost is significantly reduced, and it does not rely on measurement tools but more on computing power.

3D reconstruction technology has experienced an evolution from professional equipment to a popular application. Relying on laser scanners (1960s) and depth sensors (1990s) in the early years, cost limited adoption. The turning point came in the breakthrough of feature detection algorithm - in 1999 Lowe proposed the SIFT feature [2], which made image-based 3D reconstruction possible. In 2003, the video-based 3D reconstruction system developed by Pollefeys realized dynamic scene modeling for the first time. 2006 Milestone: Snavely's Photo Tourism system enables automatic reconstruction of Internet photos, Bundler algorithm becomes basis for SfM open-source tool; In the same year,

Furukawa proposed the Patch-based MVS (PMVS) algorithm, which solved the intensive reconstruction problem of MVS [3]. Schonberger's study in 2016 showed that the traditional System Definition Model (SDM)-MVS still has advantages in most engineering scenarios [4]. Deep methods such as MVSNet require a large amount of training data, while traditional methods are adaptable and do not require pre-training. It is worth noting that the existing research on improving multi-focus algorithms is insufficient in systematically studying data acquisition specifications [5]. For example, Moulon pointed out in 2013 that 90% of reconstruction failures are due to improper shooting methods, but the relevant standardization research is still blank. This imbalance limits the engineering transfer of technology, especially in structurally complex industrial Settings. The rise of machine learning after 2015 has given further impetus to the development of 3D reconstruction. In 2017, the 3DMatch framework demonstrated the superiority of convolutional neural networks in feature matching for the first time, and the matching accuracy in weak texture areas increased by 42%. After 2020, a series of breakthroughs have been made in end-to-end reconstruction methods, with MVSNet reducing the reconstruction error to 0.4mm in standard data sets, while Neural Radiance Fields (NeRF) technology has created a new paradigm for neural radiation fields [6].

This study aims to establish the data acquisition specification of SfM-MVS and investigate the reconstruction quality sensitivity to shooting parameters. Different from the research path of algorithm improvement, the thesis focuses on the neglected key link of the shooting method. By designing systematic controlled experiments, the thesis will quantitatively analyze the influence of variables such as shooting method, shooting quality, and lighting conditions on the reconstruction quality. Reality Capture, an industrial-grade software, was used to ensure the reliability of the results. A quality evaluation system was established based on the dimensions of point cloud density and geometric accuracy in Schonberger's evaluation framework in 2016. The expected results include: (1) revealing the quantitative relationship between shooting parameters and reconstruction quality; And (2) forming framing norms suitable for complex scenes. This work will promote the transformation of 3D reconstruction from "experience-driven" to "scientific specification", providing reusable technical standards for engineering applications.

2. Methodology

This paper will focus on the Solid-State Flash Memory (SSFM)-MVS 3D reconstruction process, and the specific research process will be divided into three parts. In this section, the core scheme of SSFM-MVS 3D reconstruction at the present stage will be introduced first. The variables affecting the reconstruction effect will be deduced according to the scheme analysis, and the research method will be stated. Finally, the next section will give experimental results and data analysis.

2.1. Core Scheme of SfM-MVS

Structure from Motion with SfM-MVS achieves high-precision reconstruction from 2D images to 3D models through geometric correlation and optimization of multi-view images. The core process starts from the data input and pre-processing stage (showed in Fig. 1). Specifically, the input data is usually a multi-angle image sequence covering the target scene, and it is recommended that the overlap rate of adjacent images be no less than 50% to ensure the robustness of feature matching [7]. In the pre-processing stage, blurred or overexposed images are first screened by a sharpness evaluation function (such as the Brenner gradient method). Then, the lens distortion is corrected based on camera calibration parameters (including focal length, central point, and radial distortion coefficient). In particular, for large-scale scenes (such as city-level aerial photography), the images must be segmented into spatial grids to reduce the complexity of subsequent computation. Finally, the standardized data after preprocessing lays a reliable foundation for sparse reconstruction.

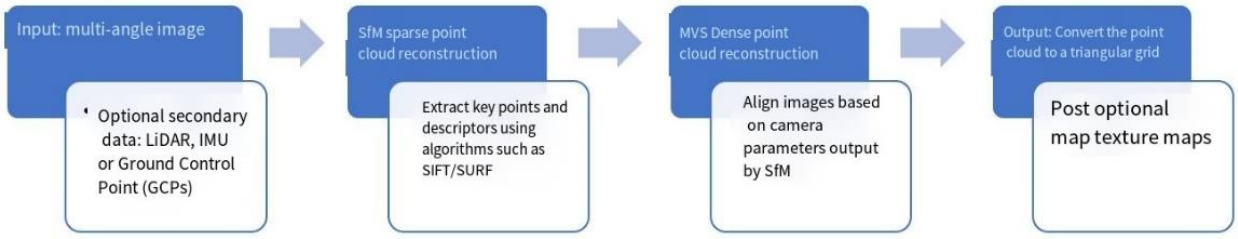


Figure 1. The pipeline of the study (Picture credit: Original).

2.2. SfM

SfM aims to recover camera parameters from images with a sparse 3D point cloud. Firstly, the SIFT algorithm is used to extract image key points and their 128-dimensional descriptors, which have become the gold standard for feature extraction due to their invariance to scale and rotation. Then, k-dimensional (KD)-Tree neighbor search and bidirectional matching were used to establish cross-perspective feature correspondence, and the Random Sample Consensus (RANSAC) algorithm, combined with the fundamental matrix estimation, was used to eliminate mismatching (mismatching rate could be reduced to less than 5%). On this basis, the camera parameters and three-dimensional point coordinates are gradually optimized by incremental Bundle Adjustment (BA). Specifically, the image pairs with the most matching points are selected to calculate the fundamental matrix, and the initial point cloud is triangulated in the initialization stage. Then, the new image is registered to the existing coordinate system by the PnP algorithm, and the latest 3D points are triangulated; Finally, the reprojection error function is minimized by the Levenberg-Marquardt algorithm. It is worth noting that the output camera parameters (internal participation of external parameters) and sparse point cloud (density of about 10^4-10^5 points) in this stage provide a geometric constraint framework for the subsequent dense reconstruction. The following is the function of its mathematical form:

$$E = \sum_{i=1}^{n} \sum_{j=1}^{m} \rho(\left\| x_{i,j} - \pi(P_i, X_j) \right\|^2) \tag{1}$$

Where ρ is the Huber robust kernel function, P_i is the camera projection matrix, and X_j is the three-dimensional point coordinates [8]. The camera parameters (internal participation in external parameters) and the sparse point cloud (density of about $10^4 - 10^5$ points) output in this stage provide the geometric constraint framework for the subsequent dense reconstruction.

2.3. MVS

MVS aims to generate a high-density 3D point cloud and surface model based on sparse reconstruction results. First, stereoscopic correction reduces the matching problem to a one-dimensional search by projecting the image pairs onto the same polar plane. Then, the improved Patch Match algorithm is used to construct the cost, and the pixel-level matching cost is calculated by combining the Census transform and gradient consistency (3-5 times more efficient than the traditional method) [9]. To solve the problem of occlusion and noise, semi-global matching (SGM) or conditional random field (CRF) is used to optimize the depth map. After the multi-view depth map is fused, the Poisson reconstruction algorithm transforms the point cloud into a watertight mesh [10]. The specific steps include eliminating outliers based on the point cloud density and normal consistency, building an octree structure, solving the Poisson equation to generate an implicit surface, and finally simplifying the number of mesh surfaces by the edge folding algorithm (the compression rate can reach 90%). At this point, the output triangular mesh model (vertex number >10^6) has an accurate geometric structure, but the surface still lacks photorealistic texture.

Texture mapping and accuracy verification are the key steps to creating a realistic model and evaluating its quality. First, the mesh surface is parameterized to 2D texture coordinates by UV

unrolling to avoid stretching and overlapping. Then, the textures projected from multiple perspectives are weighted and averaged to eliminate seams and lighting differences. The relative and absolute accuracy should be evaluated to quantify the reconstruction quality fully. The relative accuracy is calculated by calculating Root Mean Square Error (RMSE) of sparse and dense point clouds (typical value is 0.1-1.0 pixels [11]); The absolute accuracy relies on GCPs or laser scanning data to verify the geometric error (the vertical error of Unmanned Aerial Vehicle (UAV) reconstruction can reach cm level). It should be emphasized that accuracy verification confirms that the model meets the application requirements (such as mapping requiring vertical error <0.1m) and provides essential feedback for algorithm optimization.

3. Analysis of Experimental Results and Data

This study was divided into four groups: one control group and three experimental groups. The lighting environment, shooting equipment, and shooting background were controlled. The reconstruction quality was compared after the four data groups were imported into Reality Capture. The critical influencing factors of 3D reconstruction were inferred, and optimization suggestions were given.

3.1. Experimental conditions

3.1.1. Modeling Object Selection and Reasons.

The modeling object is a light for film and television, with a hood (shown in Fig. 2). It includes a silver reflector, a colorless reflector, surface texture, and other factors that test shooting quality and algorithm level. To maximize the difference in reconstruction quality caused by shooting differences, which is more conducive to research, the following is a photo of the modeling object under the standard environment. The result is shown in Table 1.

Figure 2. The outline drawing (Picture credit: Original).

Table 1. Group Information Overview

	Lighting Conditions	Shooting equipment	Shooting background	Number of photos taken
Control group	Simulated ambient light	Professional camera	Blue background wall	113
Experimental group 1	Point Light Source	Professional camera	Blue background wall	143
Experimental Group 2	Simulated ambient light	Cell phone	Blue background wall	93
Experimental Group 3	Simulated ambient light	Professional camera	Office background	114

3.1.2. Explanation of Nouns.

Due to the background wall and fixed lighting, the standard environment can only cover 90°, so for every 90° shot, turn the subject 90° to simulate a similar lighting environment. The following Fig. 3 shows the location diagram of the lighting fixture (shown in Fig. 3).

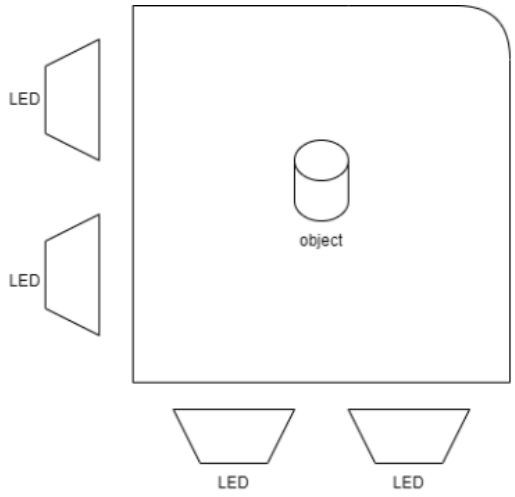


Figure 3. The design drawing of the lighting environment (Picture credit: Original).

A point light source is tilted 45° behind the modeled object to prevent the same horizontal height. No blinds, soft boxes, and other appliances simulate harsh lighting conditions. For every 90° shot, turn the subject and point light source by 90° to simulate a similar lighting environment. The Fig. 4 shows the lighting fixture's location diagram (shown in Fig. 4).

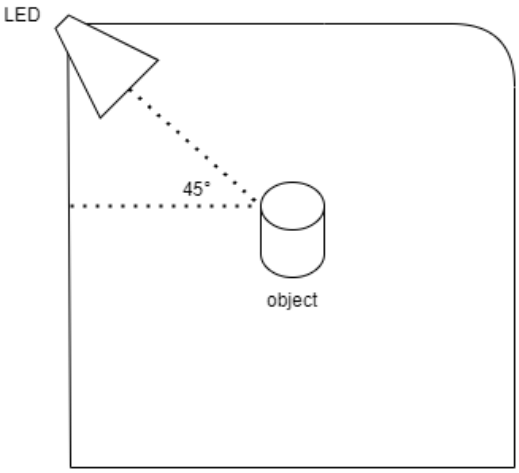


Figure 4. The point light source of the image (Picture credit: Original).

Camera: Sony ILCE-A7CM2 camera and Sony SEL50F25G lens; Camera specs: 33 megapixels, full-frame sensor, 50mm focal segment; Exposure Settings: Manual exposure, International Organization for Standardization (ISO) 1600, exposure time 1/100s, aperture F8; Phone: HUAWEI Mate 40; Phone parameters: Auto exposure, 50 megapixels, 27mm focus segment Exposure Settings: Auto exposure.

3.2. Discussion

3.2.1. Control Group.

A relatively dense point cloud map has been generated (shown in Fig. 5), and the point cloud density is evenly distributed, with partial loopholes at the reflective hood. Execute rebuilding the regular quality model (corresponding to the MVS process in the flow), export it as an.obj file, and view the effect in Blender as follows (shown in Table 2).

Table 2. The result of control group

Control points	0	1	2	3
Aligned pictures	27	78	86	89

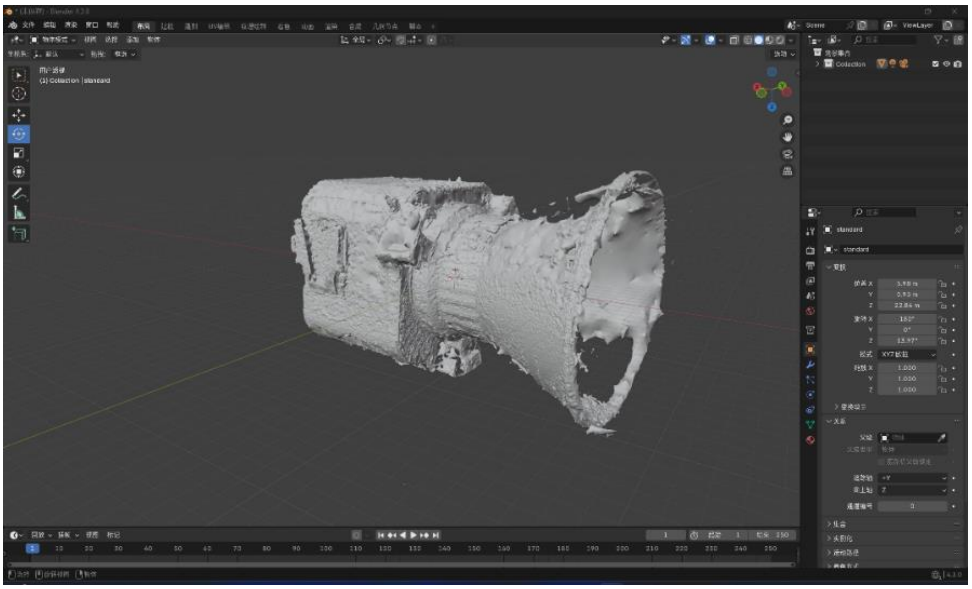


Figure 5. The 3D modeling diagram of a relatively dense point cloud map (Picture credit: Original).

3.2.2. Experimental Group 1.

A relatively sparse point cloud image was generated (shown in Fig. 6), with the density mainly distributed on the side of the light source, and the backlight plane could not be aligned. Execute the program to rebuild the regular quality model, export it as an.obj file, and view the effect in Blender as follows (shown in Table 3):

Table 3. The result of group 1

Control points	0	1	2	3
Aligned pictures	21	21	18	/

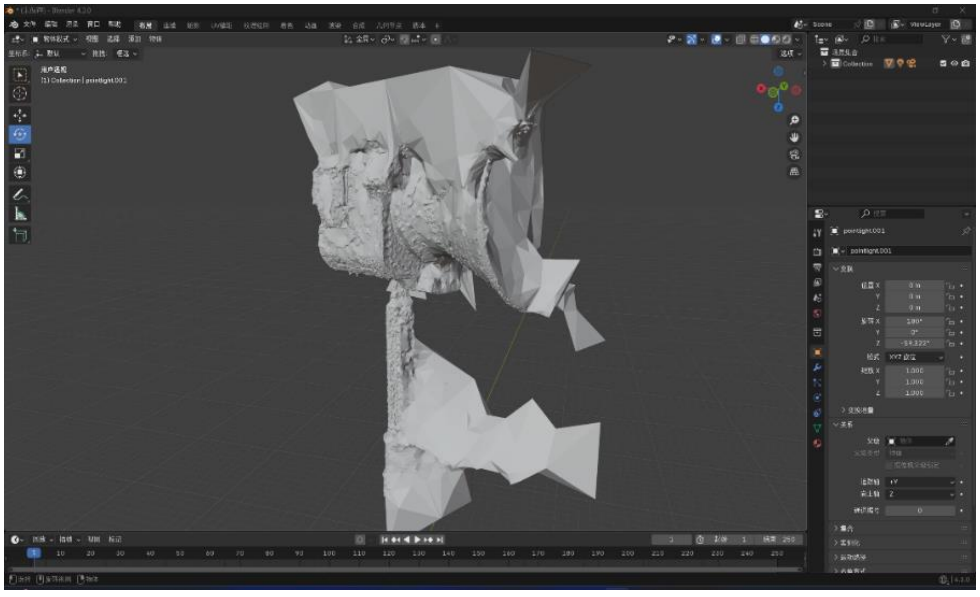


Figure 6. The 3D modeling diagram of a relatively sparse point cloud image (Picture credit: Original).

3.2.3. Experimental Group 2.

A relatively sparse point cloud image has been generated (shown in Fig. 7). Execute the program to rebuild the regular quality model, export it as an.obj file, and view the effect in Blender as follows (shown in Table 4).

Table 4. The result of group 2

Control points	0	1	2	3
Aligned pictures	20	23	29	/

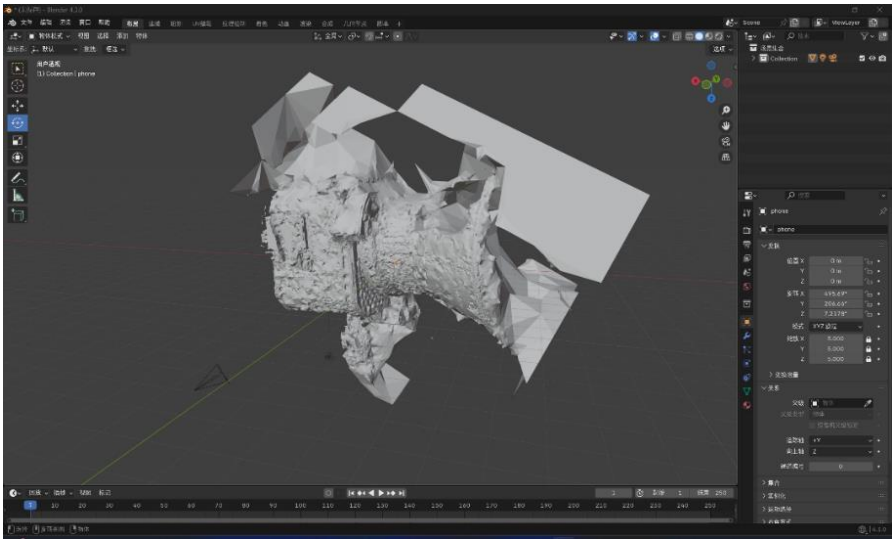


Figure 7. The 3D modeling diagram of a relatively sparse point cloud image (Picture credit: Original).

3.2.4. Experimental Group 3.

A moderately dense point cloud image with an even density distribution has been generated (Fig. 8). Execute the program to rebuild the regular quality model, export it as an.obj file, and view the effect in Blender as follows (shown in Table 5).

Table 5. The result of group 3

Control points	0	1	2	3
Aligned pictures	20	53	69	/

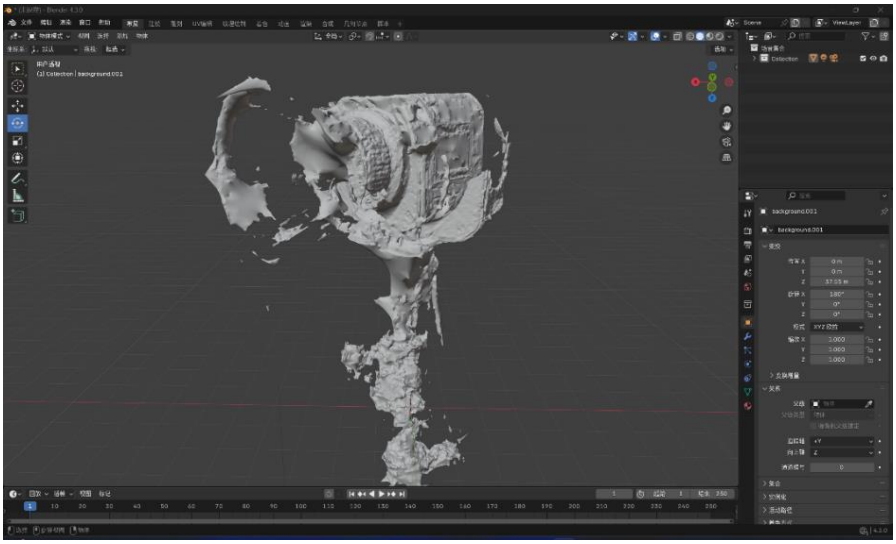


Figure 8. The moderately dense point cloud image (Picture credit: Original).

3.3. Data Analysis

Fig. 9 shows photos with a successful alignment success rate (number of successful alignments/total number of photos taken) changing with the number of control points calibrated. The selected points are all in the location of 20 random photos in the photo library. It can be seen that the shooting

background (experimental group 3) has the least influence on reconstruction, and the broken line and control group have the highest degree of fit. Lighting (experimental group 1) and shooting equipment (experimental group 2) have a greater influence on reconstruction, and the photos successfully aligned in experimental group 1 even decline with the increase of the marking point. Therefore, it can be preliminarily inferred that the photo quality directly affects the alignment success rate and the auxiliary effect of the alignment of the marking point.

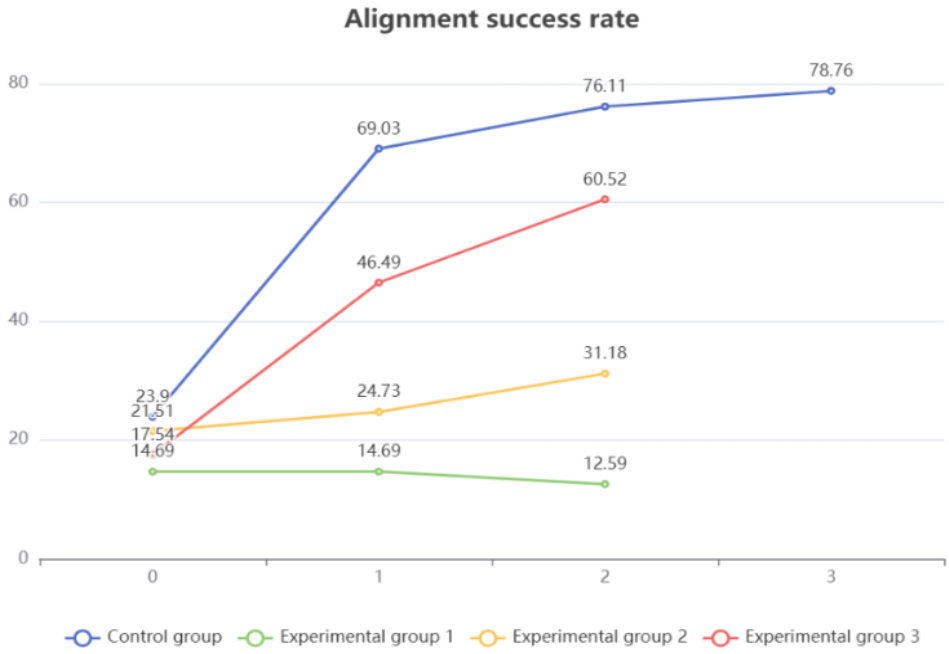
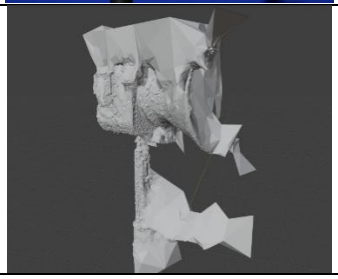
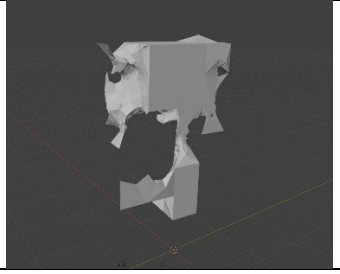


Figure 9. The successful alignment success rate (Picture credit: Original).

3.3.1. Analysis of The Influence of The Lighting Environment.

The matching success rate of experimental group 1 was only 18/143 (12.6%), and that of the control group 89/113 (78.7%), was only about 16% of that of the control group. At the same time, extreme model accuracy was unevenly distributed, concentrated on the illuminated surface (about 35% coverage), and almost no reconstruction of the backlit surface (shown in Table 6). However, the positive side gave high model accuracy, even higher than that of the experimental group 2. Large holes (up to 8.7cm in diameter) appeared in the reflective area of the hood, but were incorrectly filled directly during reconstruction.

Table 6. Comparison table of light and dark parts of the model and the real thing

	Highlights	Dark part
Physical objects		
Model		

The point light source without a soft light box left obvious light and dark lines on the reconstructed object (shown in Fig. 10), undoubtedly increasing the number of feature points to be matched. However, due to the limitations of the research method, the rotation of the object and the lamp by an accurate 90° cannot be completely guaranteed. This further causes the position deviation of the light and shade line on the model, enlarges the estimation error of pose position, making it larger than the convergence range, and cross-region pictures cannot be successfully matched.

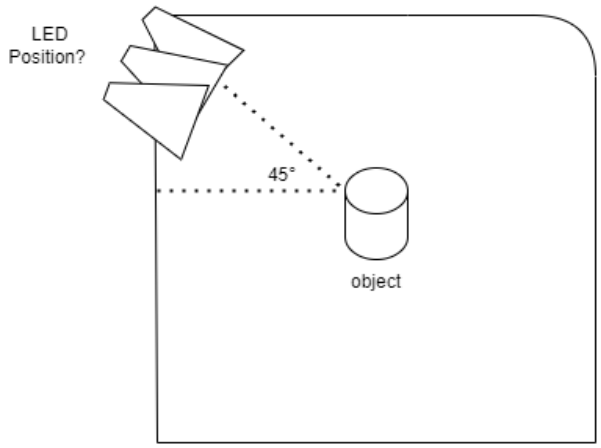


Figure 10. The point light source of the image (Picture credit: Original).

High dynamic range exceeds the sensor limit: the luminance ratio between the direct region of the point light source (illuminance >3000lux) and the backlight region (<50lux) exceeds 60:1, which exceeds the dynamic range of the camera (A7C II) at 14 levels, resulting in overexposure of the illuminated surface (RGB value ≥250) and the signal-to-noise ratio (SNR) of the backlight surface <5dB. At the same time, MVS relies on the pixel intensity consistency between multiple viewing angles. Still, the point light source causes the brightness of the dark area to vary significantly in different viewing angles (such as non-Lambertian surfaces).

3.3.2. Analysis of The Impact of Exposure.

Experimental group 2, filmed with a cell phone under simulated ambient light, experienced a significant decrease in reconstruction quality compared to the control group. The alignment success rate was only 29/93 (31.2%) in the mobile phone group, much lower than 89/113 (78.7%) in the professional camera group. Point cloud density: The point clouds generated by mobile phones were sparse, the edges were jagged, and the number of mesh pieces was reduced by about 40%. Sensor hardware defects as seen in Table 7

Table 7. Three schemes comparing

Parameters	Huawei Mate 40 (phone)	Sony A7C II	influence
Sensor size	1/1.28 inch (≈9.8 x 7.3mm)	Full frame (36 x 24mm)	Single pixel lighting area 6.5 times smaller
Dynamic range	10.8 gear (DxOMark)	14.5 gear (measured)	Dark side signal-to-noise ratio (SNR) is 53% lower
Read noise	$4.3e^-$ (ISO 1600)	$1.8e^-$ (same as ISO)	Low light area feature point signal-to-noise ratio <8dB
Sensor size	1/1.28 inch (≈9.8 x 7.3mm)	Full frame (36 x 24mm)	Single pixel lighting area 6.5 times smaller

When the dynamic range is lower than 12 stops, brightness is easy to overexpose in a high-light ratio environment, and all its details are lost. Dark information is covered by noise, and extractable feature points are reduced. The automatic exposure algorithm is out of control. Mobile phone automatic exposure (AE) frequently adjusts the ISO (fluctuation range ISO 200-2500) during shooting, resulting in a significant difference in brightness between adjacent frames, despite the controlled exposure (shown in Fig. 11) ($\Delta EV=1.2\pm0.4$). SIFT descriptor mismatch rate increased by 37%.

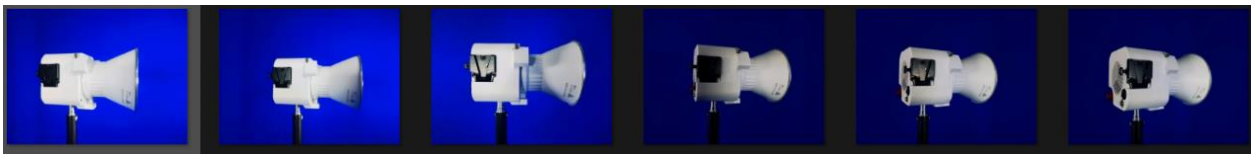


Figure 11. The Various angles of the physical object (Picture credit: Original).

3.3.3. Influence Analysis of Background.

The initial alignment office background group (20/114, 17.5%) was lower than the blue background group (27/113, 23.9%). With the addition of control points, the alignment in the office background group improved to 69/114 (60.5%), which was still lower than the 89/113 (78.7%) in the blue background group. Still, it was already significantly higher than in the other two groups. The model quality was high, but compared to the large area of misfilling in the different groups, this model had a large number of burrs, and the reflective part of the hood was further broken.

Feature interference: the uniform tone of the blue background wall (RGB 0,0,255) can effectively SIFT (its feature points are reduced by about 62%). In contrast, high-frequency textures such as books and monitors in the office background produce many mismatched features, increasing to 12.3% vs 5.8% of the blue background. Personnel movement during the process resulted in feature inconsistencies between viewing angles, reducing the function's convergence.

3.3.4. Analysis of Other Effects.

In all groups, the reconstruction effect of the reflective hood is not ideal. The following table (shown in Fig. 12) shows the statistics of the number of sectional surfaces of the control component model. The five regions are five small models with similar volumes. It can be found that the part with the large reflective area (shown in Fig. 13) recognizes about 50% of the model faces in other regions. Visible reflective surfaces are a significant problem in today's purely visual solutions.

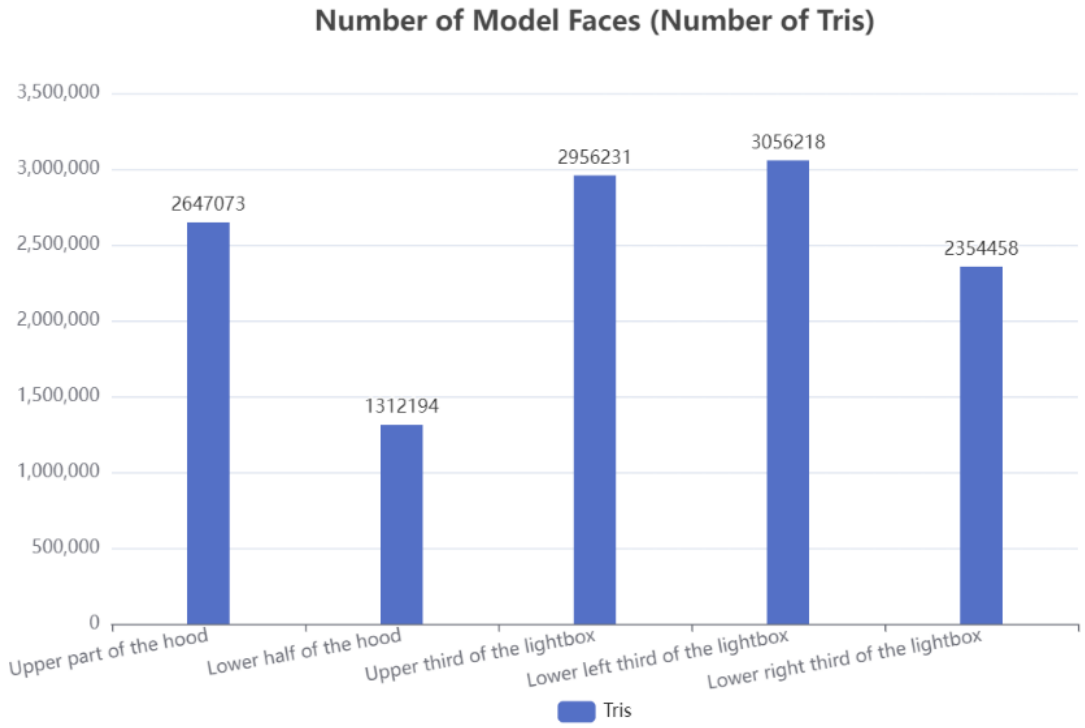


Figure 12. The statistics of the number of sectional surfaces of the control component model (Picture credit: Original).

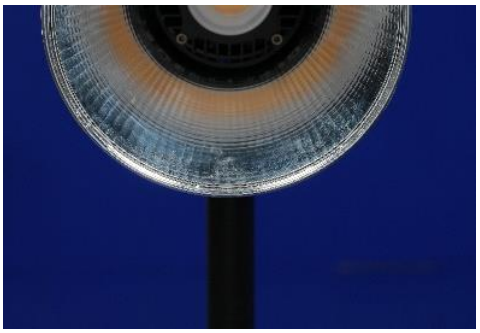


Figure 13. The part with the large reflective area (Picture credit: Original).

3.4. Optimization Method

3.4.1. Improvement Based on Traditional Methods.

Experimental group 1 showed that the light-dark boundary led to a sharp increase in the feature matching failure rate (alignment rate 12.6% vs. 78.7% in the control group) and 8.7cm holes in the reflective area of the hood. Therefore, it is recommended to add automatic image preprocessing before the SfM process, suppress the highlight, brighten the dark part, and keep the subject's EV value unchanged. Experimental group 2 showed that due to the loss of control of automatic exposure, the problem that the mobile phone image quality is not as good as the camera is further amplified, and the reconstruction quality is poor. It is recommended to use a manual exposure strategy when the mobile phone is reconstructed, or single automatic exposure, and control the lighting environment to remain unchanged. In experimental group 3, background clutter will increase invalid feature points and increase model burrs. Therefore, it is recommended that depth of field analysis be integrated and that the virtual focus part be regarded as the dividing line between the subject and the background to ensure that the subject can be correctly identified during reconstruction.

3.4.2. Fusion Based on Neural Network.

Although the existing SfM technique still has significant advantages over pure machine learning models regarding feature point alignment, it is possible to insert pre-trained distillation models at specific steps to increase alignment success and reconstruction accuracy without increasing the computing burden. Experimental group 2 (mobile phone shot) failed to match weak texture areas due to sensor noise (SNR<10dB). DnCNN (published model) can be used to preprocess the mobile phone image; the input and output sizes remain unchanged at 1024×768, and the reasoning time is <0.2s/ frame. This can be realized by inserting a Python script into the RealityCapture preprocessing pipeline and calling it the ONNX model. All groups of reflective areas are missing, and incorrect filling results in geometric distortion. The traditional process can generate the basic model (including voids), and the voids can be repaired by calling the pre-trained MeshRCNN (ShapeNet pre-trained weight). Although the MVS process can recover the camera position and calculate the relative depth, it is possible first to estimate the monocular depth for each photo, such as the BTS (PyTorch) model. Input into the MVS process can reduce the error.

4. Conclusion

Aiming at the 3D reconstruction technology based on SfM-MVS, this study systematically explores the influence of data acquisition specifications on the reconstruction quality. This research aims to solve the problem of model accuracy decline caused by improper selection of shooting parameters (such as lighting, equipment, and background) in the current 3D reconstruction process, establish scientific acquisition standards by quantitative analysis of key variables, and promote the transformation of the technology from experience-driven to norm-driven. In this paper, a multi-group-controlled experiment framework is proposed, using RealityCapture as a platform to compare and analyze the effects of different lighting conditions (ambient light vs. point light source), shooting equipment (professional camera vs. mobile phone), and background complexity (solid color

background vs. complex scene) on the reconstruction results. The experimental process includes data acquisition, feature matching (SfM), dense reconstruction (MVS), and quality assessment (point cloud density, geometric integrity, etc.), and combines traditional optimization methods with neural network fusion strategies to improve the robustness of reconstruction.

The experimental results show that: (1) illumination uniformity significantly impacts the success rate of reconstruction. The alignment rate of the point light source group is only 16% of that of the control group, and there are large areas of holes in the model; (2) the feature matching failure rate increased by 37% due to the automatic exposure of mobile phone, and the point cloud density decreased by 40% due to the limitations of sensor hardware; (3) Although the initial alignment rate of complex background was reduced (17.5% vs 23.9%), it could be recovered to 60.5% through control point optimization, and the geometric accuracy of the model was still better than that of other experimental groups; (4) Reflective surface reconstruction is a common problem, and the number of surfaces in the reflective area is only 50% of that in the non-reflective area. The subsequent research will focus on the deep integration of the pre-trained model and the traditional SfM-MVS to develop a lightweight hybrid reconstruction framework and explore more possibilities of machine learning in reconstruction. It may be extended to build standardized data sets for industrial inspection scenarios to support the iterative optimization of data-driven acquisition specifications.

References

[1] SCHOPS Thomas, et al. A multi-view stereo benchmark with high-resolution images and multi-camera videos. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, 3260 - 3269.

[2] LOWE David G. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 2004, 60: 91 - 110.

[3] FURUKAWA Yasutaka, and JEAN Ponce. Accurate, dense, and robust multiview stereopsis. IEEE transactions on pattern analysis and machine intelligence, 2009, 32 (8): 1362 - 1376.

[4] SCHONBEGER Johannes, and FRAHM Jan-Michael. Structure-from-motion revisited. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, 4104 - 4113.

[5] ZHAO Yong, et al. RTSfM: Real-time structure from motion for mosaicing and DSM mapping of sequential aerial images with low overlap. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1 - 15.

[6] VATS Vibhas, et al. GC-MVSNet: Multi-view, multi-scale, geometrically-consistent multi-view stereo. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024, 3242 - 3252.

[7] MURTIYOSO Arnadi, et al. Comparison of state-of-the-art multi-view stereo solutions for close range heritage documentation. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives, 2024, 48 (2): 317 - 323.

[8] HAMDI Abdullah, SILVIO Giancola, and BERNARD Ghanem. Mvtn: Multi-view transformation network for 3d shape recognition. Proceedings of the IEEE/CVF international conference on computer vision. 2021, 1 - 11.

[9] REN Wen Jia, et al. Patchmatch stereo++: Patchmatch binocular stereo with continuous disparity optimization. Proceedings of the ACM International Conference on Multimedia. 2023, 2315-2325.

[10] SELLAN Silvia, and ALEC Jacobson. Stochastic Poisson surface reconstruction. ACM Transactions on Graphics (TOG), 2022, 41 (6): 1 - 12.

[11] MENG Ming, et al. Structure recovery from single omnidirectional image with distortion-aware learning. Journal of King Saud University-Computer and Information Sciences, 2024, 36 (7): 102151.