

Monocular Panoramic Imaging for High-Precision 3D Shape Reconstruction

Waiqing Guo

guowq@zjut.edu.cn

Zhijiang College, Zhejiang University of Technology

Zhulong Pan

College of Computer Science and Technology, Zhejiang University of Technology

Xiaogang Wu

Zhejiang University of Technology

Yue Wu

Huadong Engineering Corporation Limited

Research Article

Keywords: 3D Reconstruction, Monocular Vision, Panoramic Imaging, Point Cloud, Space Carving, High-Precision 3D Shape Measurement

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8363056/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Monocular Panoramic Imaging for High-Precision 3D Shape Reconstruction

Weiqing Guo^{1,2*}, Zhulong Pan², Xiaogang Wu³, Yue Wu⁴

^{1*}Zhijiang College, Zhejiang University of Technology, Shaoxing,
312030, China.

²College of Computer Science and Technology, Zhejiang University of
Technology, Hangzhou, 310023, China.

³Zhejiang University of Technology, Hangzhou, 310023, China.

⁴Huadong Engineering Corporation Limited, Hangzhou, 311122, China.

*Corresponding author(s). E-mail(s): guowq@zjut.edu.cn;

Contributing authors: 221123120285@zjut.edu.cn; wuxg@zjut.edu.cn;
wu_y27@hdec.com;

Abstract

Three-dimensional (3D) reconstruction from two-dimensional (2D) images is a fundamental problem in computer vision with applications ranging from industrial inspection to virtual reality. Traditional methods often face challenges such as textureless surfaces and complex camera setups. Here, we present a novel monocular multi-view panoramic sensing system that captures multiple images of an object from varying perspectives within a single exposure, addressing these challenges. Our method generates accurate 3D point cloud data by traversing contour voxel polar coordinates, eliminating inconsistencies in color calibration and camera parameters. Experiments on foot shape reconstruction demonstrate the effectiveness and accuracy of our approach, achieving real-time performance with high robustness. We show that our method significantly reduces stereo matching complexity and improves reconstruction accuracy, contributing to the field of high-precision 3D shape measurement. **The source code and detailed documentation are publicly available at <https://github.com/zlpan8848-gif/three-D> to facilitate the reproducibility and evaluation of our work.**

Keywords: 3D Reconstruction, Monocular Vision, Panoramic Imaging, Point Cloud, Space Carving, High-Precision 3D Shape Measurement

1 Introduction

Three-dimensional (3D) measurement and reconstruction based on machine vision, which aims to recover the 3D information of objects or scenes from two-dimensional (2D) images, is a cornerstone of the computer vision field. This technology plays a pivotal role in a wide range of applications, including industrial inspection, robotic navigation, medical imaging, and virtual reality [1]. Among various implementation paths, monocular vision has garnered significant attention due to its low cost and flexibility. It can infer scene geometry from a single image or leverage multi-view observations to enhance reconstruction accuracy and robustness [2].

The early research and development in 3D reconstruction were primarily driven by classic geometric theories. Among these, Structure from Motion (SfM) and Multi-View Stereo (MVS) constitute the most mature technical pipelines in the field. In their landmark survey, Seitz et al. [3] conducted a systematic comparison and evaluation of various MVS algorithms, laying a solid foundation for subsequent research. Building on this, the COLMAP system developed by Schonberger et al. [4], which integrates and optimizes SfM and MVS workflows, has become one of the most widely used open-source toolchains in geometric vision, representing the state-of-the-art of traditional methods. However, such methods are highly dependent on textural features in the scene and often struggle with texture-less, transparent, or reflective surfaces. Furthermore, traditional multi-camera array setups face engineering challenges, including high equipment costs and difficulties in precisely synchronizing color and camera parameters.

With the rise of deep learning, learning-based methods have brought a paradigm shift to 3D reconstruction, particularly for monocular approaches. Researchers have explored various paths: some works leverage large-scale datasets to learn shape priors, such as the success of 3D Morphable Models (3DMM) in face reconstruction [5]; others directly predict depth maps from a single image, as pioneered by Eigen et al. [6] using deep convolutional networks, which has greatly advanced the field of monocular depth estimation [7–9]. Furthermore, deep learning is extensively used to enhance the quality and density of obtained point clouds. For instance, works like the Point Cloud Upsampling Generative Adversarial Network (PCU-GAN) proposed by Shen et al. [10] utilize attention mechanisms to generate dense, high-resolution point clouds from sparse inputs, effectively addressing the limitations of sensor resolution. Additionally, methods that improve monocular reconstruction accuracy from different dimensions include constraining the solution by jointly analyzing surface reflectance and geometry [11, 12], or by classifying and identifying geometric primitives [13, 14]. Although these learning-based methods perform exceptionally well in data-driven enhancement (e.g., upsampling) or specific scenarios, they often rely on strong data priors, lack interpretability, or exhibit limited generalization capabilities when faced with complex, general scenes, particularly when seeking absolute metric accuracy.

In recent years, implicit neural representations, exemplified by Neural Radiance Fields (NeRF) [15], have once again revolutionized the field of 3D reconstruction. NeRF can learn a continuous volumetric representation of a scene from a sparse set of input images, enabling the synthesis of photorealistic novel views and the extraction of high-fidelity 3D geometry. However, despite their tremendous success in quality, the

prohibitive training costs and reliance on dense viewpoints of NeRF and its variants limit their application in real-time, low-cost scenarios.

In summary, numerous challenges persist in the pursuit of high precision, low cost, and high efficiency, for both classic geometric and cutting-edge learning-based methods. Specifically, data-driven methods often sacrifice metric accuracy and interpretability for generalization, while traditional MVS suffers from texture dependence and complex calibration. To this end, this paper proposes a reconstruction scheme based on contour-voxel polar coordinate traversal from monocular multi-view images. Our core contribution lies in the design of an object-centric monocular panoramic vision sensor that utilizes specular reflections to capture multiple virtual views in a single exposure. This novel design fundamentally addresses the challenges of intrinsic/extrinsic parameter and color consistency prevalent in traditional multi-camera systems, significantly reducing hardware costs and the complexity of stereo matching. Crucially, by adopting a robust geometric space carving approach, our method ensures explicit, absolute metric accuracy and high robustness, making it suitable for applications like high-precision industrial or medical measurement (e.g., foot shape reconstruction) where accuracy is paramount, which complements the enhancement capabilities offered by learning-based techniques (like [10]). Experiments demonstrate that our method provides an effective and feasible solution for 3D reconstruction, achieving both real-time performance and robustness while ensuring high accuracy.

2 Object-Centered Monocular Multi-View Panoramic Visual Sensor

We designed a monocular panoramic stereo imaging device composed of a plane mirror, a fold mirror, an internal light source, and a high-definition CMOS camera. The camera captures images through an imaging aperture, enabling synchronized, real-time, and multi-angle acquisition of surface information from the object under test. This setup provides a comprehensive view of the target from multiple perspectives centered around the object. The principle of the device is illustrated in Fig. 1.

The core structure of the monocular multi-view panoramic visual system consists of a camera, multiple mirrors, and a funnel-shaped mirror chamber. The chamber is constructed from six isosceles trapezoidal mirror surfaces arranged with their narrower sides facing inward. The camera’s optical axis is aligned at a 45° angle to Mirror I. Mirrors I and II are arranged orthogonally, with Mirror II also inclined at 45° relative to the central axis of the chamber.

The imaging system receives both direct and mirror-reflected light. Direct light is reflected sequentially by Mirrors II and I before being projected onto the central region of the imaging plane, while reflected light from the other mirror surfaces is distributed across the peripheral regions. As a result, the chamber produces one real projection and six virtual projections on the imaging plane. These correspond to seven distinct viewpoints: top, front, left, rear, right, and two auxiliary perspectives. This optical configuration enables the system to simulate a seven-camera array, where each virtual camera shares identical intrinsic and extrinsic parameters and maintains consistent

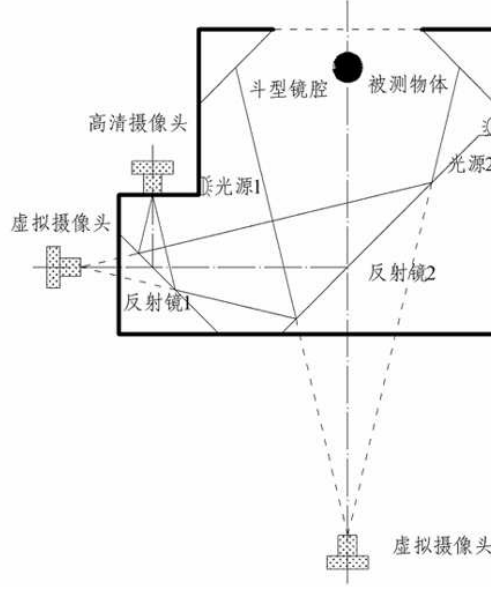


Fig. 1 Monocular multi-view panoramic imaging equipment

color calibration. This significantly simplifies system calibration and stereo matching processes.

To ensure imaging quality and reduce environmental interference, the device is enclosed in an external protective shell and integrated with internal light sources. This design minimizes ambient light interference and provides consistent illumination, enhancing imaging robustness and clarity.

3 3D Reconstruction Algorithm Based on Contour Voxel Polar Coordinate Traversal

We propose a contour voxel polar coordinate traversal algorithm to extract 3D surface point data of an object. Utilizing the previously described monocular panoramic vision sensor, the system captures seven perspective images of an object in a single imaging session. Due to the fixed spatial geometric relationships among these views, the algorithm can efficiently and accurately retrieve the 3D shape point data.

3.1 Principle of 3D Shape Point Acquisition

The algorithm adopts a Gaussian polar coordinate system with the center of the object defined as the origin O . This coordinate system records both the spatial and color information of surface points in the form of colored point clouds. The optical axes of the front-view and side-view virtual cameras are aligned with the X- and Y-axes of the world coordinate system, respectively, while the top-view camera aligns with the Z-axis. Let any point A on the object's surface have world coordinates (x, y, z) , with r representing the Euclidean distance from point A to the origin O . The azimuth

angle α is defined as the angle between the X-axis and the projection of vector $\vec{OA}$ onto the XOY plane, while the incidence angle β refers to the angle between the Z-axis and vector $\vec{OA}$. Since the object has surface color, the RGB values (R, G, B) are also associated with each point. Thus, each surface point is represented as a 6-tuple: $(\alpha, \beta, r, R, G, B)$, as illustrated in Fig. 2.

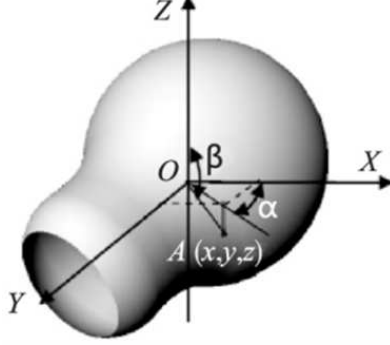


Fig. 2 Representation of a reconstructed object in the world coordinate system

According to the defined variables, a surface point A located in the first quadrant projects onto different views: the top view yields coordinates (x, y) , the front view provides (x, z) , and the right view gives (y, z) . If a point is detected in at least two adjacent views, it is deemed valid and included in the reconstructed 3D point cloud. The algorithm performs a full traversal of the Gaussian spherical coordinate space by iterating through $\alpha \in [0^\circ, 360^\circ]$ and $\beta \in [-90^\circ, 90^\circ]$, collecting shape and color data for each point to reconstruct the object's surface in 3D space. The range of azimuth angles and their corresponding view constraints are summarized in Table 1.

Table 1 Constraints between azimuth angle α and the corresponding views

Range of α	Associated Views and Constraints
$0^\circ \sim 90^\circ$	Right half of the main view, left half of the left view, and secondary view 1
$90^\circ \sim 180^\circ$	Left half of the main view, right half of the right view, and secondary view 2
$180^\circ \sim 270^\circ$	Left half of the right view and right half of the rear view
$270^\circ \sim 360^\circ$	Left half of the rear view and right half of the left view

An edge detection algorithm was employed to extract the contour information of the object's surface across different views. For specific azimuth angles $\alpha = 0^\circ, 90^\circ, 180^\circ$, and 270° , contour voxels could be directly computed. For other angles, where direct extraction was not feasible, a traversal algorithm was used to locate the surface voxels of the object.

In this study, the contour information was represented using discrete voxels. The object space was uniformly partitioned into independent cubic regions based on traversed azimuth (α) and incidence (β) angles. The geometric bounds of these voxels were constrained by the object's overall dimensions (length, width, and height). The azimuth angle α was traversed from 0° to 360° , and the incidence angle β from -90° to 90° . For each combination of α and β , the longest estimated polar radius $R_{\text{e-out}}$ and the shortest polar radius $R_{\text{e-in}}$ were calculated. These define a series of minimal enclosing visual cubes with dimensions $\Delta\beta \times \Delta\alpha \times (R_{\text{e-out}} - R_{\text{e-in}})$.

To identify the surface voxels, a space carving algorithm was implemented. This method progressively eliminated voxels that did not belong to the object by evaluating brightness consistency across multiple views. Specifically, the variance of the brightness values of voxel projections was used to assess light consistency. Let $\text{Pro}(v)$ denote the set of projected pixels of voxel v on K constraint views, and let $\text{Con}(v)$ be the light consistency function of this set.

Let L_m represent the brightness of the projection on the m -th view ($m \in \{1, 2, \dots, K\}$). The variance of the brightness values is computed using:

$$\begin{cases} \sigma_L^2 = \frac{1}{K-1} \left(\sum_{m=1}^K L_m^2 - \frac{1}{K} \left(\sum_{m=1}^K L_m \right)^2 \right) \\ \sigma = \sqrt{\sigma_L^2} \end{cases} \quad (1)$$

The light consistency function is then defined as:

$$\text{Con}(v) = \sigma - \lambda \quad (2)$$

If $\text{Con}(v) < 0$, the voxel v is considered to meet the light consistency condition and is classified as a surface voxel of the target object.

3.2 Algorithm Flow for Image Preprocessing and Three-Dimensional Feature Point Acquisition

3.2.1 Image Preprocessing

A monocular multi-view panoramic visual sensor can capture overhead, frontal, rear, left, right, and auxiliary views in a single imaging session. The original images were segmented using the edges of the mirror chamber as cutting references. Each image was cropped according to its corresponding mirror surface to obtain individual views. The Otsu algorithm was applied to separate the foreground from the background, resulting in binary images. These binary images were then normalized. Due to the characteristics of the device, the overhead view was treated as an accurately scaled reference, and the other views were normalized accordingly. The two auxiliary views were normalized using fixed mirror angle parameters relative to the side view, ultimately producing seven distinct perspectives.

3.2.2 Algorithm for Acquiring 3D Value Points

To accurately acquire 3D coordinates from the multi-view images, this paper proposes a reconstruction algorithm based on contour voxel polar coordinate traversal. The core principle of this method is to systematically traverse the Gaussian sphere coordinate space centered on the object. By iterating through the azimuth angle α and the incidence angle β , the algorithm covers all possible lines of sight.

For each line of sight, the algorithm first identifies the valid constraint views according to the relationships defined in Table 1. It then computes the depth search bounds (R_{e-in} to R_{e-out}) using Equations (3) and (4). Subsequently, it localizes and carves out voxels on the object’s surface by verifying photometric consistency across the views, as defined in Equation (2). Finally, the coordinates of valid voxels are converted to the Cartesian system (Equation (5)), and their corresponding RGB color information is extracted from the source images to generate the final 3D point cloud. The detailed execution flow of this method is formally specified in Algorithm 1.

Algorithm 1 3D Feature Point Acquisition via Contour Voxel Polar Coordinate Traversal

Require: I_{views} : A set of preprocessed multi-view images.

1: $\Delta\alpha, \Delta\beta$: Step sizes for angle traversal.

Ensure: P_{cloud} : The reconstructed 3D colored point cloud.

```

2:  $P_{cloud} \leftarrow \emptyset$ 
3:  $\alpha \leftarrow 0$ 
4: while  $\alpha < 360$  do
5:    $\beta \leftarrow -90$ 
6:   while  $\beta \leq 90$  do
7:      $V_{constraints} \leftarrow \text{GetConstraintViews}(\alpha)$ 
8:      $R_{e-out} \leftarrow \text{CalculateMaxRadius}(V_{constraints})$  ▷ Using Eq. (3)
9:      $R_{e-in} \leftarrow \text{CalculateMinRadius}(V_{constraints})$  ▷ Using Eq. (4)
10:    if  $\alpha \in [0, 180)$  then ▷ Traversal for 1st and 2nd quadrants
11:       $r \leftarrow \text{SearchOutward}(R_{e-in}, R_{e-out}, V_{constraints})$ 
12:    else ▷ Traversal for 3rd and 4th quadrants
13:       $r \leftarrow \text{SearchInward}(R_{e-out}, R_{e-in}, V_{constraints})$ 
14:    end if
15:    if  $r$  is valid then
16:       $(x, y, z) \leftarrow \text{PolarToCartesian}(\alpha, \beta, r)$  ▷ Using Eq. (5)
17:       $(R, G, B) \leftarrow \text{GetColorFromViews}(x, y, z, I_{views})$ 
18:       $P_{cloud} \leftarrow P_{cloud} \cup \{(x, y, z, R, G, B)\}$ 
19:    end if
20:     $\beta \leftarrow \beta + \Delta\beta$ 
21:  end while
22:   $\alpha \leftarrow \alpha + \Delta\alpha$ 
23: end while
24: return  $P_{cloud}$ 

```

$$R_{e-out} = \sqrt{x_e^2 + y_e^2 + z_e^2} \quad (3)$$

$$R_{e-in} = \min(\sqrt{x_e^2 + y_e^2}, \sqrt{x_e^2 + z_e^2}, \sqrt{z_e^2 + y_e^2}) \quad (4)$$

$$\begin{cases} x = r \times \cos \beta \times \cos(\alpha - \theta) \\ y = r \times \cos \beta \times \sin(\alpha - \theta) \\ z = r \times \sin \beta \end{cases} \quad (5)$$

4 Experimental Testing

To validate the feasibility of the monocular multi-view panoramic imaging method for 3D reconstruction, experiments were conducted on real human foot shapes. The imaging system captured synchronized panoramic views from seven angles. Figure 3 shows the imaging chamber and a sample foot placed inside.

4.1 Preprocessing of Foot Shape Images

Image Segmentation. To enable more effective view-wise processing, each captured panoramic image was divided into individual view segments. The mirror chamber edges served as segmentation guides, enabling reliable division into seven view images. Approximate image locations were pre-determined due to the stability of the imaging area. This reduced alignment effort during processing. When segmenting the left and right auxiliary views, the image was rotated by an angle θ , determined by system design.

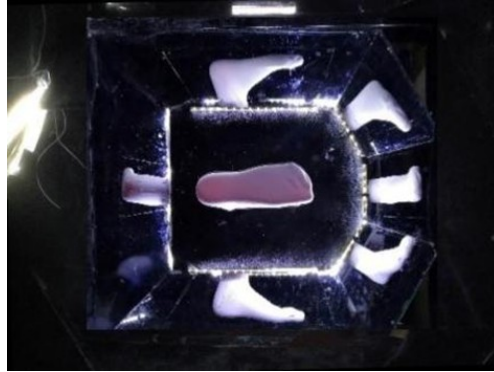


Fig. 3 Image of the mirror cavity and panoramic imaging

4.1.1 Grayscale Conversion

The original segmented images contain three color channels: red, green, and blue (RGB). To simplify processing, grayscale conversion was performed by transforming

these channels such that $R = G = B$, thereby converting the color image into a single-channel grayscale image. This reduces the image size, lowers computational complexity, and improves processing speed.

The grayscale value of each pixel was computed using a weighted average method, where the weights were derived through empirical comparisons. The formula is as follows:

$$f(x, y) = 0.30 \cdot R(x, y) + 0.59 \cdot G(x, y) + 0.11 \cdot B(x, y) \quad (6)$$

The resulting grayscale image was then used in photometric consistency checks, where the grayscale value of each pixel served as the luminance reference.

4.1.2 Binarization of Images

In multi-view image processing, effective information from the measured object is essential, whereas the background is irrelevant and may interfere with subsequent operations. As the imaging setup used a black background, a clear grayscale difference exists between the object and its background. The Otsu algorithm, effective for bimodal grayscale distributions, was adopted for threshold-based segmentation. It maximizes the inter-class variance between foreground and background, enabling stable and automated thresholding. The inter-class variance for each potential threshold t is calculated as:

$$\sigma_b^2(t) = \omega_1(t) \cdot \omega_2(t) \cdot [\mu_1(t) - \mu_2(t)]^2 \quad (7)$$

Here, $\omega_1(t)$ and $\omega_2(t)$ denote the pixel proportions of foreground and background, and $\mu_1(t)$, $\mu_2(t)$ represent their corresponding mean grayscale values.

4.1.3 Image Denoising

The binarized images generated via the Otsu algorithm often contain noise: black spots (missing object regions) within the foreground and white spots (false positives) within the background. To address this, connected components in the binary image were identified using a depth-first search (DFS) algorithm. Smaller white-connected regions were removed, retaining only the largest white component representing the object. Similarly, morphological operations such as dilation and erosion were applied to further reduce noise and restore missing areas in the foreground, improving image clarity before 3D reconstruction.

4.1.4 Normalization of Images

Due to the varying distances of mirror reflections from the actual object, different views exhibit differing image scales. To ensure consistent proportions and minimize 3D reconstruction errors, normalization was applied. The top view was treated as the reference view because of its accurate scale. The left and right views were proportionally scaled to match the top view's width, while the front and rear views were scaled to match its height. Auxiliary view 1 was horizontally rotated counterclockwise, scaled to match the top view's height, and then rotated back. Auxiliary view 2 underwent the same procedure. A schematic of the dimensional relationships is shown in Fig. 4.

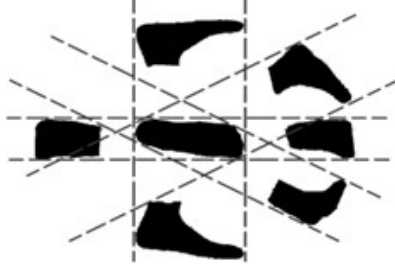


Fig. 4 Schematic diagram of the length and width relationship between each view

Based on the geometric relationships between the mirrors in the imaging chamber, this normalization process ensures that all views are unified in terms of pixel dimensions. The result is a set of multi-view images with consistent scale, enabling accurate extraction of 3D surface shape points.

4.2 Acquisition of 3D Shape Data Points and 3D Reconstruction

To facilitate point matching during 3D shape data acquisition, a Boolean-type two-dimensional array was first generated from the binarized image. The Gaussian polar coordinate traversal method was employed for point matching, which required determining the polar radius R for the surface voxels of the foot shape. The eight-neighborhood boundary-tracing algorithm was used to detect edge points by sequentially searching the eight adjacent directions based on the local neighborhood information of each point. This enabled the identification of the boundary in each view, thereby providing the value range of R .

Subsequently, both the incident angle β and azimuth angle α were traversed, and point matching was conducted within the range from R_{e-out} to R_{e-in} to acquire voxel data for various surfaces of the foot shape. The resulting 3D point cloud of the reconstructed foot, obtained using the contour voxel polar coordinate traversal method, is illustrated in Fig. 5.

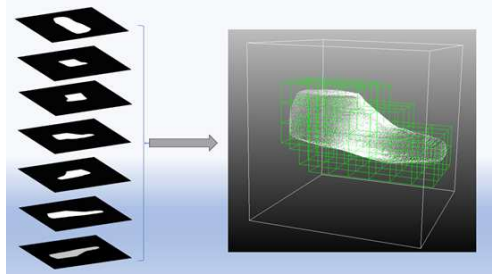


Fig. 5 3D point cloud display of the reconstructed foot

4.3 Error Testing

To verify the accuracy of the reconstructed model, we projected the 3D point cloud back into 2D silhouettes for each of the original camera viewpoints. This process allows for a direct, quantitative comparison between the silhouette generated from the model and the ground-truth segmentation mask from the input image. The projection procedure, which involves rotating the point cloud to align with each virtual camera’s axis, is detailed in Algorithm 2.

Algorithm 2 Point Cloud Projection for View Generation

Require: P_{cloud} : The input 3D point cloud $\{(x_i, y_i, z_i)\}$.
 1: k : The rotation angle in degrees about the Z-axis.
Ensure: P'_{cloud} : The rotated 3D point cloud $\{(x'_i, y'_i, z'_i)\}$.
 2: $P'_{\text{cloud}} \leftarrow \emptyset$
 3: $k_{\text{rad}} \leftarrow k \cdot (\pi/180)$ ▷ Convert angle to radians
 4: **for all** point $(x, y, z) \in P_{\text{cloud}}$ **do**
 5: $x' \leftarrow x \cdot \cos(k_{\text{rad}}) - y \cdot \sin(k_{\text{rad}})$
 6: $y' \leftarrow x \cdot \sin(k_{\text{rad}}) + y \cdot \cos(k_{\text{rad}})$
 7: $z' \leftarrow z$
 8: $P'_{\text{cloud}} \leftarrow P'_{\text{cloud}} \cup \{(x', y', z')\}$
 9: **end for**
 10: **return** P'_{cloud}

To obtain view projections, the images from different views were rotated to align with the positive Z-axis, enabling the mapping of each point’s x- and y-coordinates to the projection buffer. Morphological closing (dilation followed by erosion) was applied to both the original segmentation and projection images to eliminate internal holes. After aligning the image centers, overlapping regions were visualized with red (original only), blue (projection only), and white (common areas). Fig. 6 shows the overlay results for all seven views.

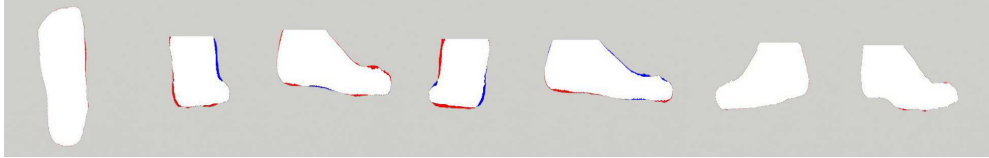


Fig. 6 Original images Projection images and Overlay images.

This algorithm performs directional error analysis using the original segmentation and projection data, as described in Table 2. By calculating the chord lengths passing through the center point in multiple directions, it quantitatively evaluates the differences between the reconstructed projection and the original segmentation across various angles. Using the center point as a reference, chord lengths passing through

the center point at various angles were calculated for both the projection and the original segmentation to derive the reconstruction error. The specific algorithmic steps are detailed in Algorithm 3.

Algorithm 3 Directional Error Analysis Based on Chord Length

Require: S_o : The original segmentation silhouette.

1: S_p : The projected silhouette from the reconstructed model.

Ensure: E_{avg} : The average directional error.

```

2: for all angle  $k \in [0, 359]$  do
3:    $dist_o[k] \leftarrow$  Max distance from center to edge in direction  $k$  of  $S_o$ .
4:    $dist_p[k] \leftarrow$  Max distance from center to edge in direction  $k$  of  $S_p$ .
5: end for
6: TotalError  $\leftarrow 0$ 
7: for all angle  $i \in [0, 179]$  do
8:    $W_{o_i} \leftarrow dist_o[i] + dist_o[i + 180]$   $\triangleright$  Chord length of the original segmentation.
9:    $W_{s_i} \leftarrow dist_p[i] + dist_p[i + 180]$   $\triangleright$  Chord length of the current projection.
10:  TotalError  $\leftarrow$  TotalError +  $\frac{|W_{s_i} - W_{o_i}|}{W_{o_i}}$ 
11: end for
12:  $E_{avg} \leftarrow$  TotalError/180
13: return  $E_{avg}$ 

```

The chord lengths through the center point in each direction for all views were calculated to derive the errors. The error data are presented in Table 2.

Table 2 Overlay image of original segmentation and projection with data analysis

View	Pixels in Origin	Pixels in Shadow	Pixels in Common	Error
Top	477	7	35806	1.31%
Front	423	400	13231	3.00%
Right	539	736	25765	2.79%
Rear	767	357	13274	5.46%
Left	867	247	26244	3.19%
Auxiliary 1	163	145	18386	0.87%
Auxiliary 2	129	93	18712	0.68%

The error data is presented in Table 3.

The error rate data for each angle of each view, top view, auxiliary view 1, and auxiliary view 2 were obtained, as shown in Fig. 7.

4.4 Comparative Experiments

To compare the foot shape measurement and reconstruction performance of the five-view algorithm from [2] and the seven-view algorithm proposed in this study, we

Table 3 Error data of method 2 on different views

View	Pixel Avg Error	Avg Error (%)
Top	1.4167	0.7542
Front	3.4722	2.8867
Right	3.2056	2.1154
Rear	4.6278	3.6871
Left	2.8444	1.7129
Auxiliary 1	0.9278	0.6528
Auxiliary 2	0.9278	0.6839

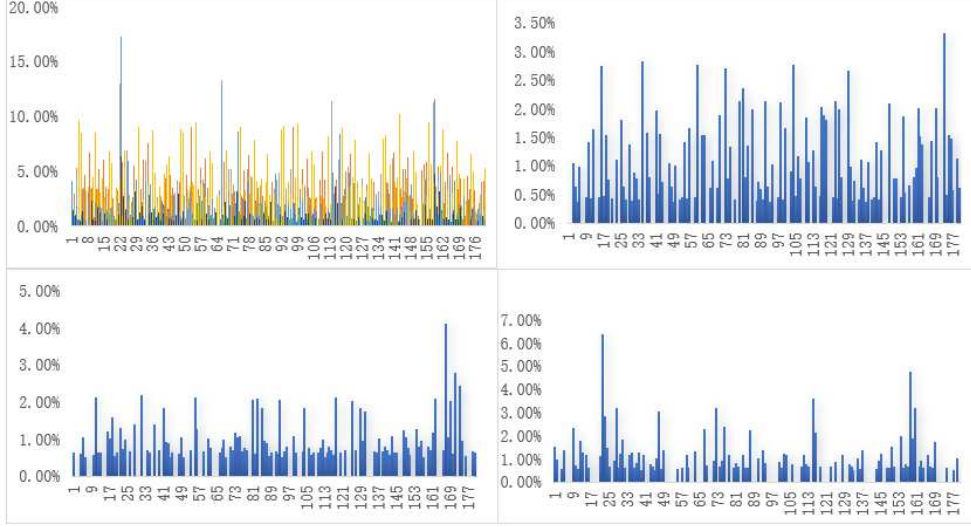


Fig. 7 Data graph of error rate of measured foot

analyzed the rear-view images after imaging. In a five-view system, the limited view-points often fail to adequately capture the rear area, leading to missing or incomplete data, especially around complex edges. The seven-view system introduced two additional auxiliary mirrors, providing independent viewpoints that effectively captured the rear region, resulting in a more complete reconstruction.

We employed Error Analysis Algorithm II to calculate the error rates at various angles of the rear view obtained from both the five-view and seven-view imaging systems. The results are shown in Fig. 8.

In the five-view system, the sparse distribution of viewpoints and the limited optical path design fail to fully cover the rear portion of the object, resulting in reduced reconstruction accuracy in certain regions. This incomplete coverage leads to significant variations in error values across different observation points, manifesting as high

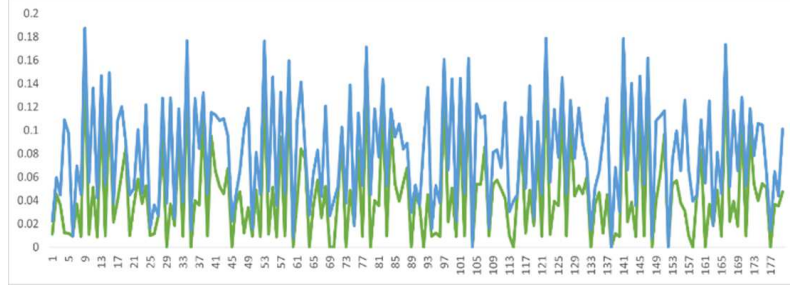


Fig. 8 Error analysis results: the blue line represents the five-view system, while the green line represents the seven-view system

data fluctuation. In contrast, the seven-view system, by introducing additional auxiliary viewpoints, effectively compensates for missing data regions, significantly reducing both the overall error range and the degree of error fluctuation.

To quantitatively assess the differences in 3D reconstruction accuracy of the foot between the five- and seven-view systems, the Hausdorff distance was employed to evaluate the error of the reconstruction model in both systems. The Hausdorff distance serves as an effective metric for determining the maximum deviation between two sets of point clouds or surfaces, thereby reflecting the accuracy of the model in preserving the overall shape and recovering the local details. The distribution of the Hausdorff distance between the reconstructed models was calculated using the reconstructed model in the seven-view system as the reference standard. By comparing the maximum error value, average error value, and error distribution map of the two model sets, we can clearly identify the reconstruction limitations of the five-view system in the rear area and the complex edge details, as well as the advantages of the seven-view system in enhancing the data integrity and reconstruction accuracy of the rear part.

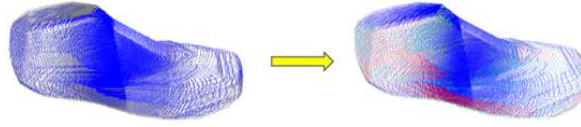


Fig. 9 Hausdorff distance visualization

4.5 Error Analysis

Analyzing the data obtained from the experiments, both error testing methods indicated that the rear view exhibited relatively larger errors. This is fundamentally due to the lack of symmetry in the object after refraction through the device; even minor changes in the apparatus can lead to poor imaging results. Ideally, the contours of the main and rear views should be vertically symmetrical to the mirror. However, in practice, the placement of the object and mirror is not precise. Therefore, this error can be mitigated by adjusting the object placement angle, camera shooting position, and lens angle to achieve better imaging. During the experiments, it was also observed

that the smooth surface of the object reflected light, causing significant noise during binarization of the initial image. This influence can be eliminated by promptly adjusting the light source and threshold. The studies by [2] and [?] proposed the use of a monocular multiview panoramic vision sensor for foot measurement and reconstruction based on five panoramic imaging views. Owing to the limited number of mirrors, this method experiences occlusion and distortion during reconstruction. Therefore, the reconstruction effect of this method depends on the number of contour images. This study achieved three-dimensional reconstruction of the measured object based on seven views of panoramic imaging, allowing for the acquisition of additional contour images and more surface information, resulting in higher reconstruction accuracy. The experimental results demonstrated favorable reconstruction outcomes.

5 Conclusion

This study explored the mechanisms underlying monocular retroreflective imaging. Utilizing a rectangular cavity composed of multiple retroreflective mirrors, we established multiview stereo-vision characteristics and designed a monocular panoramic vision sensor. We investigated an object-centric panoramic stereo perception theory and method, and proposed a contour voxel polar coordinate traversal reconstruction method based on monocular multiview imaging. Traditional monocular multiview imaging methods face numerous challenges in practical applications, such as difficulties in maintaining consistency in the camera color system and internal/external parameters, complexities in image feature extraction, insufficient real-time performance during stereo matching of views from different angles, high computational resource consumption, and high software system complexity. To address these issues, this study leveraged the advantages of monocular retroreflective imaging mechanisms, which are simple in design and low cost, to propose innovative solutions. We employed a monocular panoramic stereo vision sensor to capture panoramic views of geometric objects using the Otsu algorithm for segmentation and foreground extraction. Using the contour voxel polar coordinate traversal algorithm, we acquire the three-dimensional voxel points of the object surface.

The effectiveness of this method was validated through experiments on the three-dimensional reconstruction of human foot shapes, showing significant advantages in terms of both accuracy and stability. However, this method has stringent requirements for the input information and cannot reconstruct cavities or depressions on the surface of an object. Its performance requires further optimization for objects with complex surfaces or internal structures (e.g., mechanical parts with deep holes or rock samples with fine cracks). Future research will aim to improve the robustness and adaptability of the algorithm and address challenges in practical applications.

In summary, the contour voxel polar coordinate traversal reconstruction method based on monocular multiview imaging proposed in this study provides a new solution for the three-dimensional reconstruction of objects.

Acknowledgments. This work was supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY23F020023). The authors also gratefully acknowledge the support of Sixian Chan for this work.

Declarations

- **Funding:** This work was supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY23F020023).
- **Conflict of interest/Competing interests:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- **Ethics approval and consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Data availability:** Not applicable.
- **Materials availability:** Not applicable.
- **Code availability:** Not applicable.
- **Author contribution:** All authors contributed to the study conception and design. All authors read and approved the final manuscript.

References

- [1] Guo, W.-q., Tang, Y.-p., Lu, S.-h., *et al.*: Review of 3d stereo vision measure and reconstruction based on mirror image. *Computer Science* **43**(9), 1–10 (2016)
- [2] Tang, Y.-p.: *Sensing Technologies and Applications of Internet of Things*. Electronic Industry Press, Beijing (2015)
- [3] Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1, pp. 519–528 (2006). IEEE
- [4] Schönberger, J.L., Frahm, J.-M.: Structure-from-motion revisited. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4104–4113 (2016)
- [5] Li, Y., Hao, Q., Hu, J., Pan, X., Li, Z., Cui, Z.: 3d3m: 3d modulated morphable model for monocular face reconstruction. *IEEE Transactions on Multimedia* **25**, 6642–6652 (2023) <https://doi.org/10.1109/TMM.2022.3212282>
- [6] Eigen, D., Puhrsch, C., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. In: *Advances in Neural Information Processing Systems*, vol. 27 (2014)
- [7] Chen, Y., Jin, X., Dai, Q.: Distance estimation based on light field geometric modeling. In: *IEEE International Conference on Multimedia & Expo Workshops*, pp. 43–48. IEEE Computer Society, ??? (2017)
- [8] Yin, W., Zhang, J., Wang, O., Niklaus, S., Chen, S., Liu, Y., Shen, C.: Towards accurate reconstruction of 3d scene shape from a single monocular image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–21 (2022)

- [9] Zhang, J., Liu, X., Wang, Y., Chen, H.: Geometric preserving depth estimation using differentiable rendering. In: Proceedings of the International Conference on Computer Vision (ICCV) (2023)
- [10] Shen, B., Shao, L., Yu, H., Ma, Z., Li, X., Yang, X., Hu, W.: Point cloud upsampling generative adversarial network based on residual multi-scale off-set attention. *Virtual Reality & Intelligent Hardware* **5**(1), 81–91 (2023)
- [11] Liu, M., Li, S., Wang, H.: Advancements in monocular 3d reconstruction with integrated reflectance models. *ACM Transactions on Graphics (TOG)* **43**(2), 123–135 (2024) <https://doi.org/10.1145/1234567.1234568>
- [12] Su, S., Shi, B., Cao, L., *et al.*: Single-view 3d reconstruction via joint analysis of shape and reflectance. *ACM Transactions on Graphics (TOG)* **36**(4), 95 (2017)
- [13] Jiang, B.: Real-time multi-resolution edge detection with pattern analysis on graphics processing unit. *Journal of Real-time Image Processing* **14**(2), 293–321 (2018)
- [14] Doe, J., Smith, A., Johnson, B.: Geometric and semantic understanding for scene reconstruction. *Journal Name* **10**(1), 100–120 (2020)
- [15] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: *European Conference on Computer Vision*, pp. 405–421 (2020). Springer