

# Enhanced Shape-from-Silhouette reconstruction using voxel inverse projection and Marching Cubes Optimization

YuHan Chen<sup>1,2</sup>, Man Peng<sup>1</sup>, WenHui Wan<sup>1</sup>, WaiChung Liu<sup>1</sup>, and ZhaoQin Liu<sup>1\*</sup>

<sup>1</sup>State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China;

<sup>2</sup>School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China

## Key Points:

- A voxel–Marching Cubes fusion Shape-from-Silhouette (SfS) method is proposed, enhancing three-dimensional asteroid reconstruction accuracy and efficiency under low-resolution imaging.
- This method introduces morphologically optimized silhouette extraction and a peak signal-to-noise ratio–Hausdorff metric for robust threshold selection.
- The method achieves >90% storage reduction and ~50% time saving with improved visual-geometric fidelity in similar scenarios.

**Citation:** Chen, Y. H., Peng, M., Wan, W. H., Liu, W. C., and Liu, Z. Q. (2026). Enhanced Shape-from-Silhouette reconstruction using voxel inverse projection and Marching Cubes Optimization. *Earth Planet. Phys.*, 10(5), 812–822. <http://doi.org/10.26464/epp2026059>

**Abstract:** This study presents an improved Shape-from-Silhouette (SfS) framework for three-dimensional (3D) asteroid reconstruction during the spacecraft approach phases, where imaging resolution is often limited and illumination is uneven. The method integrates voxel inverse projection with the Marching Cubes (MC) algorithm to overcome the efficiency and accuracy limitations of the conventional voxel-based SfS. Morphological operations are used to extract robust contours under varying lighting. Subsequently, voxel inverse projection generates a cumulative value list, followed by MC-based isosurface extraction to produce a continuous 3D point cloud, effectively eliminating staircase artifacts and significantly reducing the computational load. A combined evaluation metric using the Hausdorff distance and peak signal-to-noise ratio is developed to guide threshold selection. Experiments using simulated datasets demonstrate a storage reduction of more than 90% and a runtime reduction of nearly 50% compared with traditional SfS methods while maintaining high geometric accuracy (up to 19.72 dB). These results highlight the potential of the proposed framework for efficient, high-fidelity 3D reconstruction in asteroid proximity operations, supporting navigation, sampling site evaluation, and mission planning for future deep-space exploration.

**Keywords:** asteroid exploration; three-dimensional reconstruction; Shape-from-Silhouette; Marching Cubes algorithm; accuracy evaluation

## 1. Introduction

Asteroids, often regarded as “fossils” of the early solar system, preserve crucial records of primitive material composition and dynamical environments during planetary formation. Their study has profound implications for understanding the origin of life, tracing the Earth’s material sources, and developing strategies for planetary defense (Shi YJ and Li L, 2025). With advances in deep-space exploration, asteroid investigation has become a central theme in international space science. Notable examples include the National Aeronautics and Space Administration’s (NASA’s) OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security-Regolith Explorer) mission, which successfully returned samples from asteroid Bennu (Lauretta et al., 2017),

JAXA’s (Japan Aerospace Exploration Agency’s) Hayabusa2 mission, which explored asteroid Ryugu (Tsuda et al., 2013), and China’s Tianwen-2 mission, currently targeting the near-Earth asteroid 2016 HO<sub>3</sub> (Li CL et al., 2024). These missions collectively mark a new era in asteroid exploration.

During the spacecraft approach phase, optical images become the primary data source for navigation, body parameter estimation, and sampling site evaluation. Three-dimensional (3D) reconstruction serves as a crucial bridge between raw image data and physical interpretation of asteroid morphology, supporting tasks such as establishing body-fixed coordinate systems, identifying candidate sampling zones, and guiding surface interaction strategies. In the case of Tianwen-2, 3D models of the target asteroid are required at multiple mission stages to refine orbital maneuvers and surface operation planning (Zhang RQ et al., 2025).

Among the available reconstruction methods, the Shape-from-Silhouette (SfS) technique offers particular advantages for long-range imaging conditions, where targets appear at low resolution

First author: Y. H. Chen, chenyuhan25@mails.ucas.ac.cn

Correspondence to: Z. Q. Liu, liuzq@aircas.ac.cn

Received 31 AUG 2025; Accepted 29 MAY 2026.

First Published online 26 JUN 2026.

©2026 by Earth and Planetary Physics.

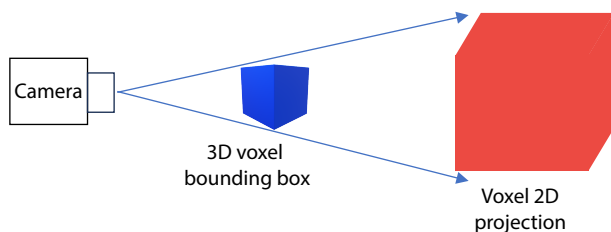
with limited surface texture. Shape-from-Silhouette reconstructs a 3D model from contours extracted from multi-view image sequences (Guo JR et al., 2024b). This approach requires only binary silhouettes, making it robust under conditions unfavorable for photometric or feature-based methods. Depending on the modeling strategy, SfS can be categorized into the voxel based, visual hull (Laurentini, 1994), and the conoxel based approaches (Casas and Salvador, 2006). Martin and Aggarwal (1983) were among the first to propose discretizing 3D space into voxels and eliminating those projected outside silhouette regions to approximate object shape.

Several improvements to SfS have been proposed for small-body applications. For instance, Li JX (2022) proposed an octree-based cropping algorithm combined with nonlinear least squares optimization of camera extrinsic parameters, reducing computational complexity. Zhang JL (2021) reformulated silhouette-based reconstruction as a pseudo-Boolean optimization problem, solved via local and hierarchical minimization. Most recently, Guo JR et al. (2024a) proposed an improved silhouette extraction algorithm to enhance reconstruction performance under variable illumination. Despite these developments, the voxel-based SfS method was selected for preliminary 3D reconstruction of asteroids during the approach phase.

Voxel-based SfS methods can be categorized into voxel set line-of-sight intersection and voxel inverse projection methods (Li JX, 2022). The voxel inverse projection method (Figure 1) projects rays from the camera center through all voxels onto the image plane, comparing the projections with binary silhouettes. Voxels inconsistent with the silhouettes are excluded, progressively refining the 3D volume.

Although conceptually straightforward, this approach has two main limitations. First, high-resolution reconstructions require evaluating an immense number of voxels, which severely reduces computational efficiency. Second, the cubic nature of voxels introduces staircase artifacts, degrading surface smoothness and geometric fidelity. These drawbacks hinder rapid and accurate 3D modeling during time-critical proximity operations.

To address these challenges, we propose an improved SfS framework incorporating the Marching Cubes (MC) algorithm. The main innovations include (1) introducing morphological preprocessing to reduce contour fragmentation caused by shadow-induced fragmentation and incomplete silhouettes under uneven lighting conditions; (2) generating voxel cumulative maps via inverse projection to retain high-confidence voxels; (3) reconstructing smooth 3D point cloud models through MC isosurface extraction,



**Figure 1.** The principle of voxel inverse projection.

thereby eliminating staircase effects and improving efficiency; and (4) developing a joint Hausdorff distance–peak signal-to-noise ratio (PSNR) metric for objective threshold selection. Experimental results show that this method outperforms traditional voxel-based SfS in both accuracy and efficiency, providing a reliable solution for asteroid 3D reconstruction during spacecraft approach.

## 2. Methods

The proposed framework includes several steps: sequential image acquisition, initial 3D voxel bounding box construction, silhouette extraction, voxel two-dimensional (2D) projection, 3D voxel reconstruction, and optimization with the MC algorithm. First, multi-angle sequential images covering one full asteroid rotation are acquired using the onboard camera. Second, a 3D voxel bounding box is defined based on the initial position of the asteroid, discretizing the surrounding space into uniform voxel units. Next, image contours are extracted and refined by using morphological operations to obtain binary silhouettes. Using the camera's intrinsic and extrinsic parameters, each voxel is projected onto the 2D image plane, and its cumulative values within the silhouettes are used to identify voxels that are consistent with the silhouettes. In the traditional SfS method, a threshold is applied to remove low-confidence voxels, producing a discrete 3D voxel model. In the improved approach, the MC algorithm is used to optimize surface reconstruction, thereby eliminating staircase artifacts and generating a smooth 3D point cloud model. The overall workflow is illustrated in Figure 2.

### 2.1 Silhouette Extraction Using Morphological Operations

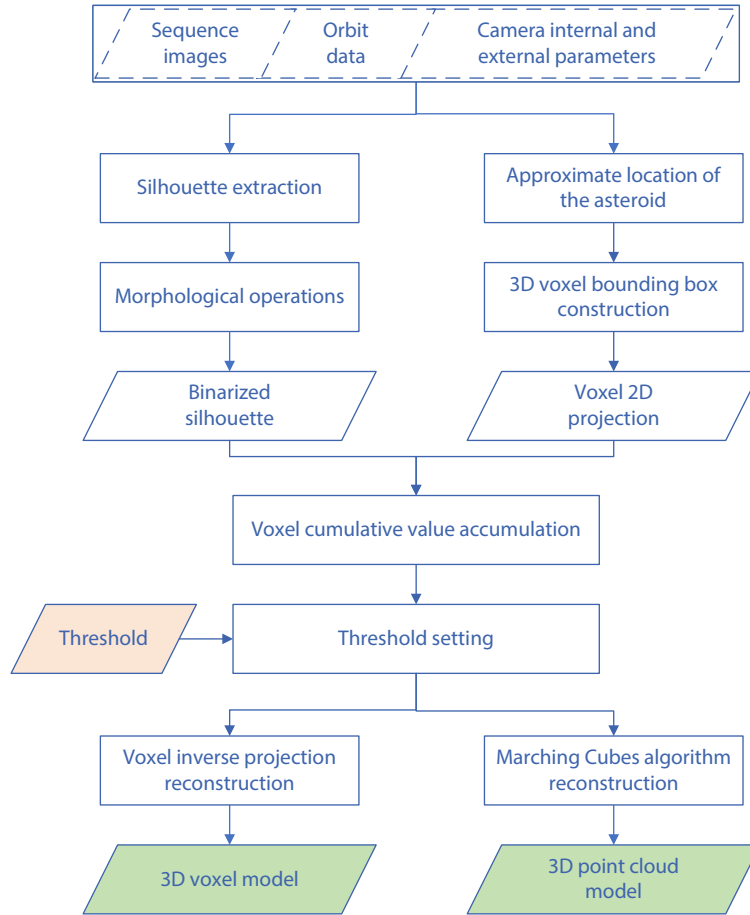
In asteroid imaging, rapid rotation and uneven illumination often produce shadowed regions whose grayscale values are close to the background, leading to incomplete or fragmented silhouettes when conventional thresholding is applied (Zheng TX et al., 2020). To address this problem, a morphological operation-based silhouette extraction method is introduced.

First, the grayscale images are binarized by using a fixed threshold to obtain initial silhouettes. To improve the robustness of silhouette extraction under uneven illumination and shadow interference, morphological processing is applied after the initial binarization stage. Specifically, morphological closing (dilation followed by erosion) is first used to fill small gaps and connect fragmented foreground regions caused by shadow-induced intensity variations. This operation helps preserve the structural integrity of the target silhouette by bridging narrow breaks along the contour.

Subsequently, a morphological gradient operation is used to enhance the object boundaries. The gradient operation highlights the transition between foreground and background regions, making the extracted contours more distinct and continuous. This operation further improves the reliability of the silhouette extraction in the presence of shadow artifacts.

Mathematically, the closing operation can be expressed as

$$I_{\text{close}} = (I \oplus S) \ominus S, \quad (1)$$



**Figure 2.** The process of the improved SFS method, from obtaining the parameters of the camera to 3D reconstruction. The threshold is one of the input elements of the reconstruction. The specific threshold selection needs to be evaluated through subsequent accuracy assessment.

where  $I$  denotes the binary image and  $S$  represents the structuring element in Equation (1). The morphological gradient, defined as the difference between dilation and erosion, preserves the object contour and is expressed as follows

$$I_{\text{grad}} = (I \oplus S) - (I \ominus S). \quad (2)$$

In Equation (2), erosion ( $\ominus$ ) and dilation ( $\oplus$ ) are fundamental morphological operations (Gonzalez and Woods, 2010). By adjusting the structuring element size and iteration count, this method adapts to different illumination conditions, resulting in silhouettes with improved completeness and robustness. Examples of silhouette extraction under variable illumination are shown in Figure 3.

A comparison of the preliminary binarization results with the images after applying morphological operations reveals that under favorable lighting conditions, the initial extraction yields satisfactory results although minor noise persists. However, under poor lighting conditions, shadowed areas prevent complete separation of foreground and background, resulting in incomplete silhouettes. Following the application of the proposed morphological operations, the extracted silhouettes exhibit enhanced integrity and clearer boundary definition. These results provide visual evidence supporting the role of morphological operations in mitigating incompleteness caused by shadows during silhouette extraction and enhancing the stability of silhouette extraction.

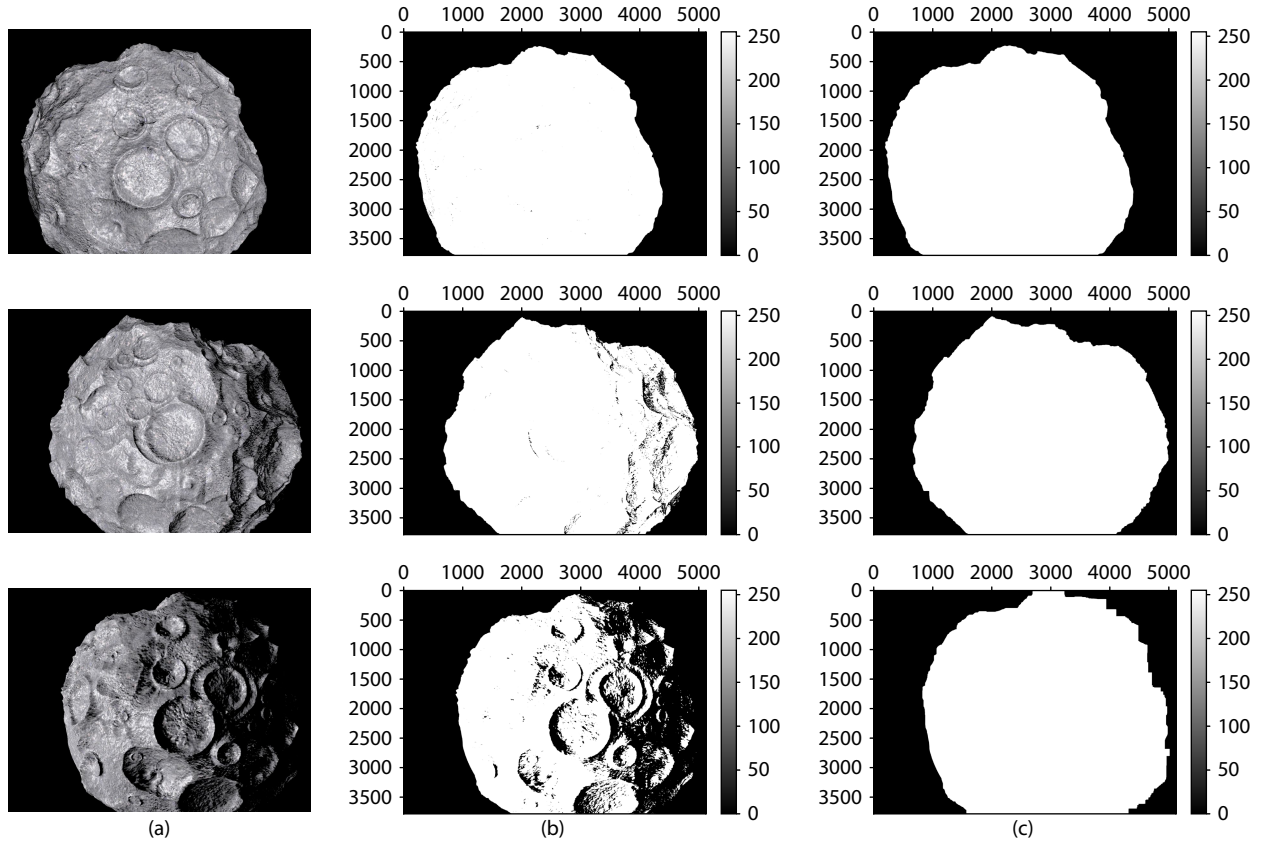
For both datasets, we first normalized grayscale images to  $[0, 1]$  and binarized them using a fixed threshold. We then applied morphological closing followed by a morphological gradient with a structuring element  $S$ .

## 2.2 3D Voxel Construction

The initial 3D voxel construction defines the spatial bounding box of the target object, providing a discretized computational basis for voxel inverse projection. First, the center of the bounding box is determined. Using onboard instruments such as inertial measurement units and sun sensors (Li JX, 2022), together with preliminary orbit determination and control data, we can derive the extrinsic parameters of the spacecraft camera relative to the asteroid for each captured image, along with the asteroid's initial position. Next, the side length of the bounding cube is set according to the estimated upper limit of the asteroid's size, ensuring the bounding box fully encloses the target. This initial definition can then be refined through corrections based on the camera's projection range.

The voxel resolution is determined by subdividing the bounding box into unit voxels. Let the cube side length be  $a$ , and the side length of each voxel be  $a_0$ . The relationship can be expressed as

$$a = n \cdot a_0, \quad (3)$$



**Figure 3.** Silhouette images extracted under different lighting conditions. The horizontal and vertical axes represent pixel coordinates (columns and rows), and the color bar shows the grayscale value (0–255). (a) Original asteroid simulation images. (b) Preliminarily binarized silhouette images without morphological processing. (c) Silhouette images after morphological processing.

where  $n$  represents the number of unit voxels along each dimension. Thus, the voxel resolution is  $n^3$ . It follows that, for a fixed bounding box, increasing the voxel resolution reduces the unit voxel size, thereby producing a more detailed 3D model of the asteroid.

### 2.3 Logic of Operation of the Voxel Inverse Projection

#### Method

Voxel inverse projection achieves 3D reconstruction by projecting voxels into the image plane and evaluating consistency with the silhouette. The mapping from world coordinates  $(X_w, Y_w, Z_w)$  to image plane points  $(u, v)$  is given by

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4)$$

$$= \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}.$$

Here,  $\mathbf{P}$  is the projection matrix,  $\mathbf{K}$  is the camera intrinsic parameter matrix,  $\mathbf{R}$  and  $\mathbf{t}$  denote the rotation matrix and translation vector from the asteroid-fixed frame to the camera frame, respectively;  $(u, v)$  are the pixel coordinates of the projected image point;  $f_x$  and  $f_y$  are the focal lengths in pixels in the  $x$  and  $y$  directions, respectively; and  $(u_0, v_0)$  are the principal point coordinates (offsets of

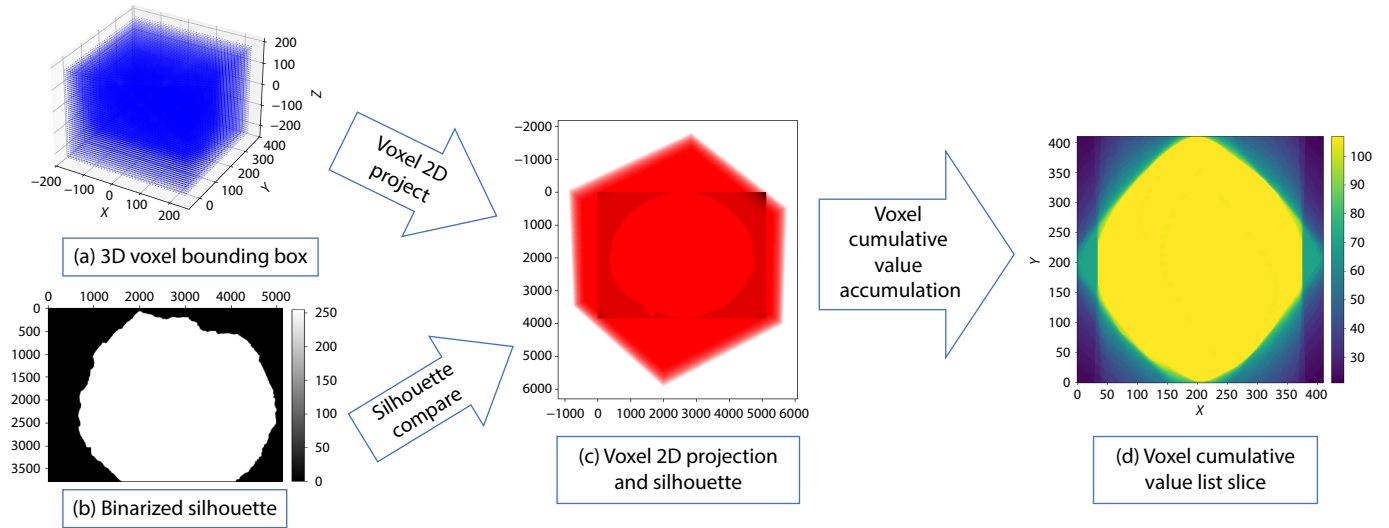
the image center from the origin) in the pixel coordinate system.

After obtaining the 2D projection of the voxels, they are compared with the silhouettes extracted from the corresponding viewpoint, and voxel cumulative values are computed. Specifically, for each voxel, the algorithm iterates over all viewpoints; if the projection of the voxel lies within the silhouette, its cumulative value is incremented by one. Once all viewpoints have been processed, a cumulative list for all voxels is generated, which serves as the basis for subsequent filtering. The outcome of the voxel inverse projection SfS method is therefore a cumulative list, which can be visualized by mapping the cumulative values into 3D slices. Finally, a threshold is applied to filter the voxels. Only those with cumulative values greater than or equal to the threshold are retained, yielding the reconstructed 3D model. The overall procedure is illustrated in Figure 4.

Equation (4) assumes that approximate camera extrinsic parameters, relative to the asteroid-fixed frame, are available through a combination of optical navigation, inertial measurements, and preliminary orbit determination. This study focuses on the silhouette-based reconstruction stage, rather than on solving the problem of joint navigation and shape estimation.

### 2.4 Reconstruction Using the MC Algorithm

In voxel inverse projection-based SfS reconstruction, the resulting model is composed of discretely stacked regular cubic units. This stacking leads to a noticeable staircase effect on the reconstructed



**Figure 4.** The workflow of the conventional voxel inverse projection SfS method. The subfigures are organized as follows. (a) 3D voxel bounding box, whose  $X$ ,  $Y$ , and  $Z$  axes denote the side lengths of the initial bounding volume. (b) Binarized silhouette image, with horizontal and vertical axes representing pixel coordinates (columns and rows), and the color bar showing grayscale intensity from 0 to 255. (c) Voxel 2D projection and silhouette, where the axes also refer to pixel coordinates with the origin at the top-left corner of the image. (d) Slice of the voxel cumulative value list, where the  $X$ - and  $Y$ -axes are the row and column indices of the voxel slice, and the color bar indicates the cumulative value (blue for low values, yellow for high values). The cumulative values are mapped onto the 3D grid, and the selected central slice is color-coded to illustrate the effect of the reconstruction threshold.

surfaces, making it difficult to accurately capture the continuity and fine-scale concavities and convexities of asteroid topography. Moreover, as voxel resolution increases, model complexity and storage overhead increase sharply, leading to a significant decline in computational efficiency.

To solve these problems, here we introduce the MC algorithm. With the cumulative list obtained from voxel inverse projection, the MC algorithm directly performs isosurface extraction and surface reconstruction, generating a continuous triangular mesh. This step produces a smooth 3D surface model that conforms more closely to the reference surface while mitigating the efficiency and storage bottlenecks of traditional voxel-based reconstructions.

The MC algorithm, proposed by Lorensen and Cline (1987), is an efficient method for extracting isosurfaces from 3D scalar fields. Each voxel is treated as a cube whose eight vertices are assigned scalar values (here, the voxel cumulative values). The relative position of the isosurface with respect to these vertices is then determined, and triangular facets are generated according to predefined topological configurations. Although 256 configurations are theoretically possible, symmetry reduces these to 15 canonical cases (Figure 5).

The scalar value  $f$  at any point  $(x, y, z)$  within a voxel can be expressed by using trilinear interpolation, as in Equation (5):

$$f(x, y, z) = a_0 + a_1x + a_2y + a_3z + a_4xy + a_5xz + a_6yz + a_7xyz. \quad (5)$$

Here,  $a_i (i = 0, 1, 2, \dots, 7)$  are constants determined by the 8 vertices of the unit voxel. If the threshold for constructing the 3D isosurface is  $T$ , representing the set of points with scalar value  $T$ , then the isosurface is given by Equation (6), which is a cubic algebraic

surface. When the voxel is sufficiently small, it can be represented by triangular facets (Xu XL et al., 2002):

$$\{(x, y, z) | f(x, y, z) = T\}. \quad (6)$$

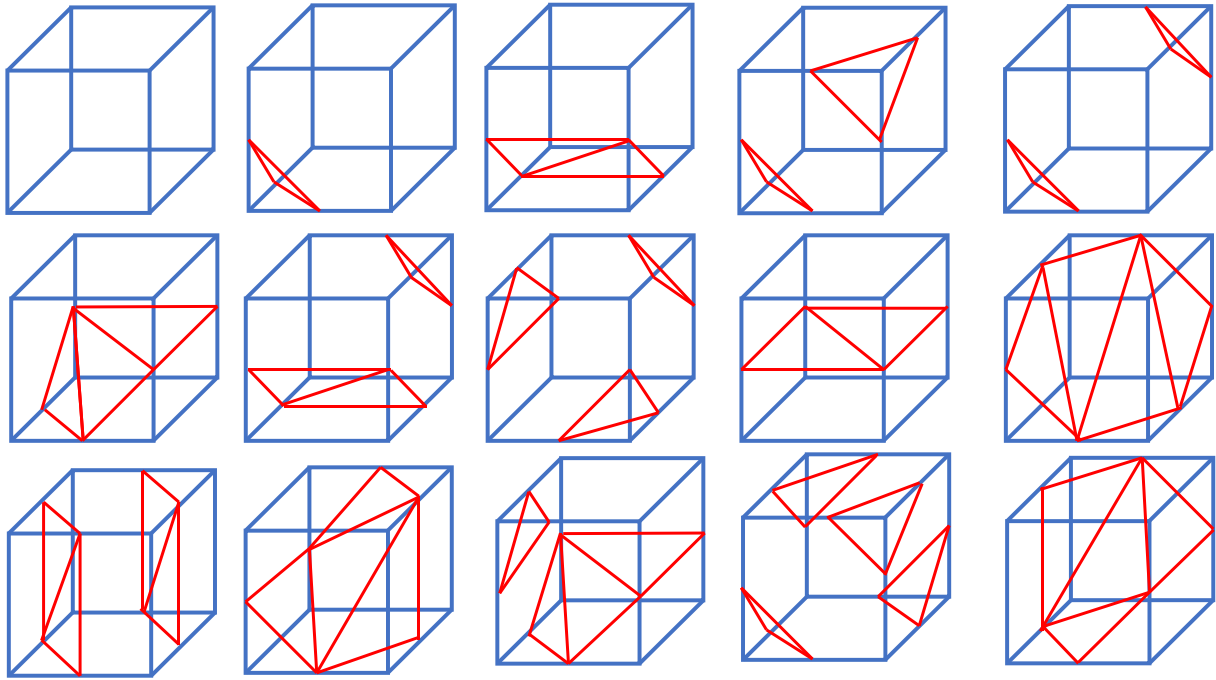
The coordinates and normal vectors of the intersection points are computed through linear interpolation based on the vertex values and a user-defined threshold  $T$ . The resulting triangular facets are then assembled to form a continuous mesh, yielding a smooth 3D point cloud model (Fan Q and Liu P, 2019).

In this study, the MC algorithm is integrated directly into the SfS reconstruction pipeline. Rather than explicitly constructing a 3D voxel model, isosurface extraction is performed directly from the voxel cumulative field by using a unified threshold.

As an algorithmic enhancement of traditional voxel inverse projection, the MC algorithm traverses each grid cube once. Its primary advantage lies in avoiding the substantial intermediate data storage required by the conventional method, instead directly generating a surface triangular mesh.

For a voxel grid with a resolution of  $n^3$ , the number of cubes is  $n^3$ . Each cube determines its corresponding index through the comparison of 8 vertex values and then performs a constant timetable lookup operation and up to 12 linear interpolation calculations. Therefore, the time complexity is  $O(n^3)$ .

Regarding memory, MC can operate in a streaming manner, given access to the scalar field; the dominant output cost is the extracted point cloud. The number of triangles generated is bounded by a constant per cube (at most 5 in the standard MC table), but in typical closed-surface objects, it scales approximately with the surface area, that is,  $O(n^2)$ . This explains why the MC-



**Figure 5.** Fifteen simplified configurations of cubic voxels.

based pipeline remains storage-efficient at high resolutions compared with storing a full occupied voxel volume,  $O(n^3)$ .

## 2.5 Threshold Selection via Hausdorff–PSNR Evaluation

In SfS-based 3D reconstruction, the choice of threshold critically influences the fidelity of the reconstructed edge details and surface structures. An excessively high threshold may eliminate valid voxels, resulting in the loss of fine terrain features, whereas a low threshold may retain noise voxels, introducing reconstruction artifacts. To determine the optimal threshold in a systematic manner, we introduce a joint evaluation framework that combines the Hausdorff distance and the PSNR. This framework quantitatively assesses the geometric differences between reconstructed models and a high-precision reference model, thereby providing an objective criterion for threshold selection.

The reference model is generated by applying Structure-from-Motion (SfM) reconstruction to the same sequence of images used in the experiments. Under close-range imaging conditions, SfM can produce high-resolution 3D models with rich surface detail and strong geometric consistency, owing to its ability to leverage multi-view feature matching and dense point cloud reconstruction. This method is both efficient and cost-effective (Fonstad et al., 2013), and it has been widely adopted in asteroid exploration. For example, Barnouin et al. (2019) used SfM in the OSIRIS-REx mission to reconstruct a centimeter-scale topographic model of asteroid Bennu. Similarly, Wang R et al. (2020) demonstrated that SfM applied to simulated asteroid images could achieve millimeter-level accuracy, with errors at control and checkpoints satisfying the requirements for detailed topographic mapping. Accordingly, SfM reconstructions serve as reliable high-accuracy references for evaluating the performance of SfS models in this study.

The proposed SfS method is not intended to outperform SfM

methods in close-range, high-resolution, and feature-rich imaging conditions. This study uses SfM as the reference model because the simulated image sequences permit stable multi-view feature matching, facilitating an objective evaluation. The practical value of the SfS technology lies in scenarios involving early detection, such as when objects occupy few pixels and the surface texture or disparity is insufficient for reliable key-point extraction and matching. Even under conditions in which SfM performance is diminished, the SfS method remains usable by leveraging the contour information provided by the images.

After the reconstructed and reference models are aligned, the Hausdorff distance is first computed to quantify geometric deviation. The Hausdorff distance effectively measures the maximum discrepancy between two point sets, defined as (Huttenlocher et al., 1993)

$$H(A, B) = \max[h(A, B), h(B, A)], \quad (7)$$

where  $A$  and  $B$  are non-empty point sets, and  $h(A, B)$  and  $h(B, A)$  denote the directed Hausdorff distances:

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|, \quad (8)$$

$$h(B, A) = \max_{b \in B} \min_{a \in A} \|b - a\|. \quad (9)$$

Here,  $\|\cdot\|$  denotes the Euclidean norm. Because the directed Hausdorff distance is asymmetric, the larger of the two directed values is taken as the final Hausdorff distance. This metric captures extreme deviations and is particularly useful for assessing maximum geometric errors between reconstructed and reference models.

However, the Hausdorff distance is scale dependent, which complicates comparisons across reconstructions with different

bounding box sizes. To address this issue, we convert the measure into the PSNR, which normalizes deviations relative to the model's spatial extent. Originally designed for image quality assessment, PSNR has been successfully adapted for evaluating 3D point cloud reconstructions (Nadipally, 2019; Xu YL et al., 2021). It is defined as

$$\text{PSNR}_{A,B} = 20 \cdot \log_{10} \left[ \frac{P}{H(A,B)} \right], \quad (10)$$

where  $P$  denotes the maximum side length of the bounding box enclosing the point cloud. A higher  $\text{PSNR}_{A,B}$  value indicates better geometric consistency between the reconstructed model and the reference.

By calculating  $\text{PSNR}_{A,B}$  values for models reconstructed under different thresholds, the influence of threshold selection on reconstruction quality can be objectively quantified. Moreover, comparing  $\text{PSNR}_{A,B}$  values before and after applying the MC algorithm allows us to directly evaluate the improvements in accuracy attributable to the proposed framework.

### 3. Experiments and Results

To evaluate the performance of the proposed voxel-MC SfS framework, we conducted experiments using two sets of simulated asteroid image sequences. The first dataset (hereafter referred to as Dataset-1) represents Bennu and was generated from the high-fidelity OSIRIS-REx shape model. It contains 107 sequential images with a resolution of  $5120 \times 3840$  pixels, captured over a complete asteroid rotation at a simulated altitude of 3 km (Dellagiustina et al., 2018). The second dataset (hereafter referred to as Dataset-2) corresponds to a simulated asteroid, generated using the orbital and physical parameters of the Tianwen-2 mission target (Li WZ et al., 2025), with 36 sequential images per illumination condition. Dataset-2 includes varying illumination scenarios, enabling a comprehensive assessment of reconstruction accuracy, efficiency, and robustness. To improve reproducibility, the main image and simulation parameters are summarized in Table 1.

As mentioned in Section 2.1, adjusting the morphological

processing parameters can enhance the effectiveness of silhouette extraction. In our experiments, these parameter choices were selected to reduce noise while preserving the overall contour. The parameter settings used in all experiments are summarized in Table 2, including the shape, size, and iteration count of the structuring element  $S$ .

Once sufficiently complete silhouettes have been obtained, voxel bounding boxes are generated based on the initial camera coordinates and the anticipated target size. A 3D voxel model is then obtained by using the traditional SfS method. As the SfS process traverses the silhouette images, the cumulative voxel values are acquired synchronously. At this stage, an MC-based framework is used for 3D model reconstruction, utilizing these cumulative values and selecting an appropriate threshold. The experiment retains the reconstructions under multiple voxel resolutions and thresholds for comparative analysis.

To comprehensively assess performance, quantitative experiments were carried out in terms of computational efficiency and geometric accuracy. For Dataset-1, Table 3 compares the average run time and storage requirements of the traditional voxel-based approach and the MC-based framework at voxel resolutions of  $41^3$ ,  $82^3$ ,  $205^3$ , and  $410^3$ . In parallel, Figure 6 presents visual comparisons of reconstructions, whereas Table 4 reports accuracy values expressed as  $\text{PSNR}_{A,B}$  at various thresholds for both methods at a resolution of  $205^3$ .

As shown in Table 3, the MC algorithm consistently outperforms the traditional method in both run time and storage efficiency. The reported run times represent end-to-end processing, from data input to model output. At all resolutions, the MC-based method required less time than did the voxel-only approach, with the advantage becoming increasingly pronounced as the resolution increased. For example, at  $205^3$ , the MC method required only 35 s, compared with 55 s for the voxel method—a 36% improvement. In terms of storage, reductions were even more significant. At the highest resolution ( $410^3$ ), the voxel-based

**Table 1.** Dataset parameters.

| Dataset   | Source                                     | Views                         | Image size         | Resolution  | Illumination condition |
|-----------|--------------------------------------------|-------------------------------|--------------------|-------------|------------------------|
| Dataset-1 | Rendered from OSIRIS-REx Bennu shape model | 107                           | $5120 \times 3840$ | 0.1 m/pixel | 100%                   |
| Dataset-2 | Simulated Tianwen-2-like asteroid          | 36 per illumination condition | $5120 \times 3784$ | 0.1 m/pixel | 100%, 75%, 50%         |

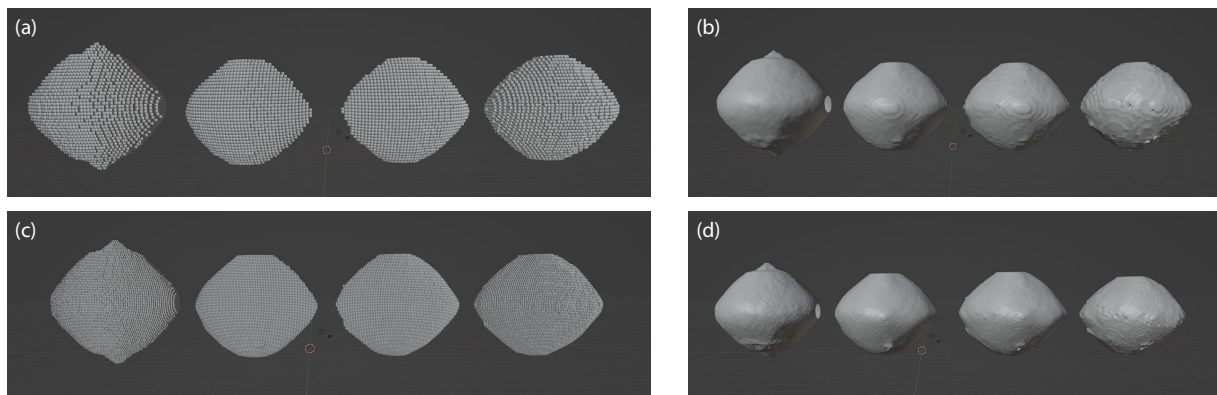
**Table 2.** Image morphological parameters.

| Dataset   | Illumination condition | Type     | Shape  | Size           | Iterations |
|-----------|------------------------|----------|--------|----------------|------------|
| Dataset-1 | 100%                   | Closing  | Square | $11 \times 11$ | 2          |
|           |                        | Gradient |        |                | 1          |
| Dataset-2 | 100%                   | Closing  | Square | $11 \times 11$ | 2          |
|           |                        | Gradient |        |                | 1          |
| Dataset-2 | 75%                    | Closing  | Square | $11 \times 11$ | 4          |
|           |                        | Gradient |        |                | 1          |
| Dataset-2 | 50%                    | Closing  | Square | $15 \times 15$ | 10         |
|           |                        | Gradient |        |                | 8          |

**Table 3.** Comparison of computational efficiency and storage requirements of the two methods under different voxel resolutions based on Dataset-1.

| Voxel resolution | Voxel inverse projection recon. |                      | MC algorithm recon. |                   |
|------------------|---------------------------------|----------------------|---------------------|-------------------|
|                  | Avg. time (s)                   | Avg. storage (kB)    | Avg. time (s)       | Avg. storage (kB) |
| 41 <sup>3</sup>  | 0.9                             | 8148                 | 0.7                 | 314               |
| 82 <sup>3</sup>  | 6                               | 80,882               | 3                   | 1458              |
| 205 <sup>3</sup> | 55                              | 1,217,824            | 35                  | 9873              |
| 410 <sup>3</sup> | Extreme <sup>a</sup>            | Extreme <sup>a</sup> | 339                 | 35,938            |

<sup>a</sup>The traditional method failed to complete reconstruction at 410<sup>3</sup> resolution because of insufficient memory. At 410<sup>3</sup>, the voxel-only pipeline requires  $\approx 9.29$  GB for voxel storage by cubic scaling from 205<sup>3</sup>, and the peak memory (including intermediate arrays) exceeded approximately 34 GB on a 64 GB RAM computer, leading to out-of-memory termination. All experiments were conducted on a computer equipped with an Intel Xeon w3-2435 processor (8 cores, 16 threads, base processing frequency of 3.10 GHz) and 64 G of RAM, running a Windows 10 operating system.

**Figure 6.** Comparison of reconstruction models with different thresholds at different resolutions based on Dataset-1. (a) 3D voxel model at 41<sup>3</sup> resolution. (b) 3D point cloud model at 41<sup>3</sup> resolution. (c) 3D voxel model at 82<sup>3</sup> resolution. (d) 3D point cloud model at 82<sup>3</sup> resolution. (a–d) The reconstruction thresholds of the four models from left to right are 70.5, 95.5, 101.5, and 106.5.

approach failed because of insufficient memory, whereas the MC algorithm successfully produced a lightweight mesh model requiring only 35.9 MB. These results confirm the scalability and engineering feasibility of the MC-based method for high-precision reconstruction tasks.

For the case in which the voxel grid resolution is  $n^3$ , the memory usage of the traditional voxel-based SfS is  $O(n^3)$ . In practice, this method requires the storage of (1) the cumulative voxel field, (2) the intermediate projection data, and (3) the explicit voxel coordinates and patches for output. Table 3 reports that the storage of the voxel model at 205<sup>3</sup> is 1,217,824 kB. Because 410<sup>3</sup> has 8 times more voxels, the voxel storage alone expands to approximately  $8 \times 1,217,824 \text{ kB} \approx 9,742,592 \text{ kB}$  ( $\approx 9.29$  GB), excluding other buffers and intermediate arrays. Adding the parallel processing overhead of the computer and system factors, the actual memory usage exceeds the available memory budget of our testing platform, causing the processing flow using the traditional voxel-based SfS pipeline to terminate at a grid resolution of 410<sup>3</sup> points, whereas the processing flow based on the MC-based pipeline remains feasible because it outputs a surface mesh whose storage scales closer to  $O(n^2)$  (surface area) instead of  $O(n^3)$ .

Table 4 further highlights the accuracy gains. Across all thresholds, reconstructions produced with the MC framework achieved higher  $\text{PSNR}_{A,B}$  values than those from the voxel-only method,

**Table 4.** Comparison of reconstruction accuracy ( $\text{PSNR}_{A,B}$ ) at different thresholds at a 205<sup>3</sup> resolution based on Dataset-1.

| Threshold (T) | Traditional SfS's $\text{PSNR}_{A,B}$ (dB) | MC algorithm's $\text{PSNR}_{A,B}$ (dB) |
|---------------|--------------------------------------------|-----------------------------------------|
| 70.5          | 16.32                                      | 17.16                                   |
| 95.5          | 18.05                                      | 19.02                                   |
| 98.5          | 18.16                                      | 19.16                                   |
| 101.5         | 18.39                                      | 19.58                                   |
| 103.5         | 18.33                                      | 19.72                                   |
| 104.5         | 18.14                                      | 19.48                                   |
| 106.5         | 8.34                                       | 9.94                                    |

demonstrating superior geometric fidelity in addition to smoother visual quality. The voxel method achieved its maximum accuracy at a threshold of 101.5, reaching 18.39 dB. The MC method also achieved optimal performance near this threshold, reaching 19.58 dB, thereby preserving details while suppressing noise. Although the absolute maximum  $\text{PSNR}_{A,B}$  occurred at 103.5, excessively high thresholds introduced minor detail loss resulting from facet interpolation. Thus, the threshold of 101.5 was selected as the optimal setting (95% rule).

In voxel inverse projection, each voxel  $v$  receives a cumulative value,

$$A(v) = \sum_{i=1}^N 1[\pi_i(v) \in S_i], \quad (11)$$

where  $N$  is the number of views,  $\pi_i(\cdot)$  is the projection under the  $i$ th camera pose, and  $S_i$  is the extracted binary silhouette. Therefore,  $A(v)$  is a view-consistency count: it measures in how many silhouettes voxel  $v$  remains inside the foreground region, reflecting its consistency across views.

In practical applications, if silhouette extraction is imperfect, a true voxel may be incorrectly excluded in a small fraction of views. Let  $\varepsilon \in (0,1)$  denote the tolerated inconsistency rate (i.e., up to  $\varepsilon N$  views may disagree because of segmentation noise or illumination artifacts). A theoretically motivated threshold can then be written as

$$T \approx (1 - \varepsilon)N, \quad (12)$$

which means that only the voxels that satisfy  $A(v) \geq T$  are retained, that is, the voxels that are consistent with at least  $(1 - \varepsilon)$  of all views. In Dataset-1, the optimal threshold selected by the Hausdorff-PSNR evaluation was close to the value obtained with  $\varepsilon = 0.05$ , yielding the empirical 95% rule. Importantly,  $\varepsilon$  should be regarded as a tunable parameter that depends on silhouette quality and view coverage; although  $\varepsilon = 0.05$  worked well in our experiments, different datasets may require adjustment.

In practical applications, a first-order perturbation analysis indicates that image projection coordinates are affected by position and attitude errors. These perturbations primarily affect voxels near contour boundaries, reducing the view-consistency count  $A(v)$ . Owing to the accumulation-and-threshold mechanism, moderate inconsistencies can be tolerated to the tolerance specified by  $\varepsilon$  in Equation (12). However, larger navigation errors degrade reconstruction integrity and geometric fidelity. Importantly, these results demonstrate that the MC algorithm achieves simultaneous optimization of both efficiency and accuracy, without trading one for the other.

To further validate the robustness of the method under different target and imaging conditions, experiments were conducted on Dataset-2, which contains  $N = 36$  sequential images per illumination condition. We observed silhouette incompleteness in 30 out of 36 images (approximately 83%), mainly near the projected polar (apical or basal) regions, including 22 appearing at the bottom and 8 at the top. Using the rendered ground-truth silhouettes available in simulation, we quantified the missing-foreground ratio for each view. The missing area averaged  $6\% \pm 3\%$  of the silhouette area, with a maximum of 9%. Furthermore, all missing pixels were located within the top or bottom 30% of the silhouette

bounding box height, supporting the view that the deficiency is concentrated in polar projections rather than uniformly distributed.

Because of these silhouette defects, absolute geometric accuracy  $\text{PSNR}_{A,B}$  was expected to be reduced. For this reason, the experiment primarily focused on computational efficiency and parameter portability, directly applying the threshold strategy derived from Dataset-1 (i.e., 95% rule corresponding to a threshold of 33.5). As shown in Table 5, the MC algorithm again demonstrated substantial efficiency advantages. Run time reductions ranged from 16.7% to 49.6%, with greater benefits at higher resolutions, whereas storage savings reached 90.9% at 410<sup>3</sup>. These results corroborate the conclusions drawn from Bennu, confirming the consistency of efficiency improvements of the MC method across different datasets.

Nevertheless, limitations were observed. Incomplete silhouettes in Dataset-2 produced localized geometric deficits, particularly in the apical and basal regions, leading to reduced overall  $\text{PSNR}_{A,B}$  values. As a result,  $\text{PSNR}_{A,B}$  values could not reliably serve as indicators of absolute geometric accuracy in this case. This outcome is consistent with the inherent limitations of the SfS method, which strongly depends on the completeness of multi-view contours. Importantly, however, these limitations did not compromise the evaluation of computational efficiency and storage requirements, which remained robust across both datasets.

To visually demonstrate the reconstruction quality, Figure 7 presents comparative models generated by the traditional SfS and MC methods alongside SfM reference reconstructions for the two datasets. The results highlight both the improved smoothness of MC-based reconstructions and the primary limitation of silhouette-based methods—namely, their inability to recover concave features such as pits on asteroid surfaces.

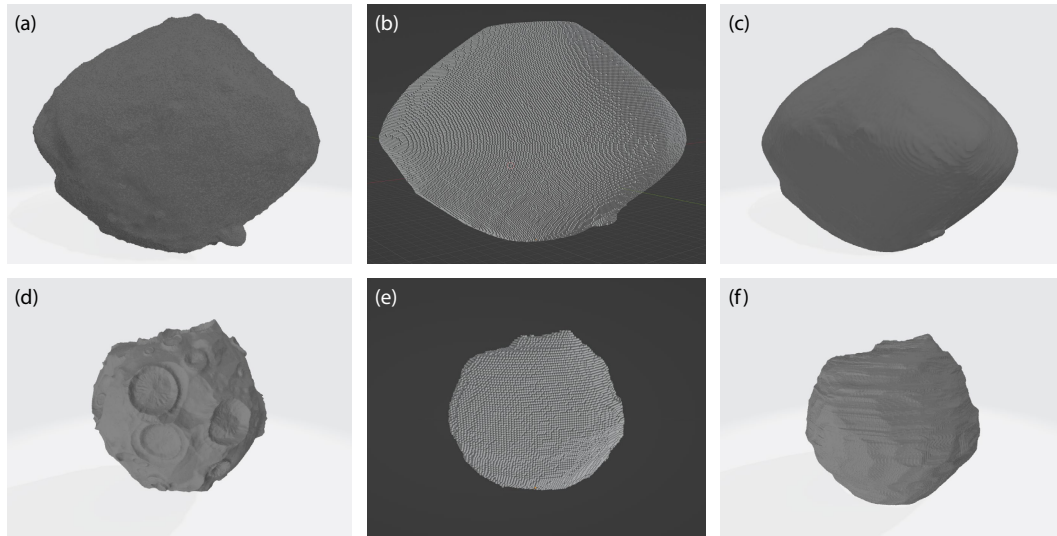
The experimental results show that when silhouette defects prevented the acquisition of precise reference values, the final threshold  $T$  for Dataset-2 was determined using the same  $\varepsilon$  value as Dataset-1, yielding visually plausible and computationally stable reconstructions. This result demonstrates the preliminary transferability of the threshold strategy across similar asteroid imagery. At the same time, the MC algorithm consistently shows stable advantages in reconstruction efficiency, confirming its robustness under varying imaging conditions.

#### 4. Conclusions

This study proposes an enhanced voxel-MC SfS framework for asteroid 3D reconstruction using sequential images. The main findings can be summarized as follows:

**Table 5.** Comparison of reconstruction computational efficiency and storage requirements based on Dataset-2 (threshold 33.5).

| Voxel resolution | 3D voxel recon. time (s) | 3D voxel model storage (kB) | MC recon. time (s) | 3D point cloud model storage (kB) |
|------------------|--------------------------|-----------------------------|--------------------|-----------------------------------|
| 41 <sup>3</sup>  | 0.6                      | 3668                        | 0.5                | 1876                              |
| 82 <sup>3</sup>  | 4                        | 62,999                      | 2                  | 7863                              |
| 205 <sup>3</sup> | 24                       | 213,613                     | 14                 | 34,061                            |
| 410 <sup>3</sup> | 125                      | 1,025,894                   | 63                 | 99,376                            |



**Figure 7.** Comparative visualization of 3D reconstruction results. (a) SfM reference model of Bennu. (b) Traditional SfS 3D voxel model of Bennu (threshold is 101.5 with a resolution of  $205^3$ ). (c) MC-based reconstructed 3D point model of Bennu (threshold is 101.5 with a resolution of  $205^3$ ). (d) SfM reference model of a simulated asteroid. (e) Traditional SfS 3D voxel model of a simulated asteroid (threshold is 33.5 with a resolution of  $205^3$ ). (f) MC-reconstructed model of a simulated asteroid (threshold is 33.5 with a resolution of  $205^3$ ).

(1) We proposed an improvement to the silhouette extraction method based on the voxel inverse projection SfS, effectively overcoming the challenges of silhouette extraction under conditions of low resolution and uneven illumination through morphological operations.

(2) The MC algorithm was integrated into the voxel-based SfS, achieving significant improvements in both storage (>90% reduction) and efficiency (~50% faster) while maintaining geometric accuracy.

(3) A joint Hausdorff–PSNR evaluation metric that enables objective and transferable threshold selection was validated across different asteroid datasets.

The proposed method assumes that a preliminary asteroid-fixed frame, or an equivalent body-relative frame, can be established through sequential optical navigation, light-curve analysis, or preliminary spin-state estimation. Inaccuracies in these parameters introduce systematic errors in the transformed camera poses, compromising projection consistency and reconstruction fidelity. Accordingly, this approach is best characterized as an approach-phase reconstruction technique applicable under approximate relative navigation conditions, rather than as a comprehensive solution for joint estimation of pose, spin state, and shape. The method is adaptable to various targets, demonstrating its feasibility for upcoming missions such as Tianwen-2. It shows promise for initializing asteroid shape reconstruction during the approach phase, provided that approximate relative pose and spin-state estimates are available. Future work should focus on validation using real mission imagery, integration of multi-source data, and the joint refinement of navigation parameters, rotational state, and 3D shape.

## Acknowledgments

This work was supported by the National Natural Science

Foundation of China (Grant No. 42530107).

## References

- Barnouin, O. S., Daly, M. G., Palmer, E. E., Gaskell, R. W., Weirich, J. R., Johnson, C. L., Al Asad, M. M., Roberts, J. H., Perry, M. E., ... The OSIRIS-REx Team. (2019). Shape of (101955) Bennu indicative of a rubble pile with internal stiffness. *Nat. Geosci.*, 12(4), 247–252. <https://doi.org/10.1038/s41561-019-0330-x>
- Casas, J. R., and Salvador, J. (2006). Image-based multi-view scene analysis using 'conexels'. In *Proceedings of the HCSNet Workshop on Use of Vision in Human–Computer Interaction – Volume 56* (pp. 19–28). Canberra, Australia: Australian Computer Society, Inc..
- Dellagiustina, D. N., Bennett, C. A., Becker, K., Golish, D. R., Le Corre, L., Cook, D. A., Edmundson, K. L., Chojnacki, M., Sutton, S. S., ... Lauretta, D. S. (2018). Overcoming the challenges associated with image-based mapping of small bodies in preparation for the OSIRIS-REx mission to (101955) Bennu. *Earth Space Sci.*, 5(12), 929–949. <https://doi.org/10.1029/2018EA000382>
- Fan, Q., and Liu, P. (2019). An improved Marching Cubes surface reconstruction algorithm based on median method. *Geomat. Spatial Inform. Technol. (in Chinese)*, 42(12), 11–14. <https://doi.org/10.3969/j.issn.1672-5867.2019.12.004>
- Fonstad, M. A., Dietrich, J. T., Courville, B. C., Jensen, J. L., and Carboneau, P. E. (2013). Topographic structure from motion: A new development in photogrammetric measurement. *Earth Surf. Proc. Land.*, 38(4), 421–430. <https://doi.org/10.1002/esp.3366>
- Gonzalez, R. C., and Woods, R. E. (2010). *Digital Image Processing* (3rd ed., Ruan, Q. Q., trans.) (in Chinese). Beijing: Publishing House of Electronics Industry.
- Guo, J. R., Chen, Y., Shao, W., Guo, W., and Liu, Y. J. (2024a). 3D reconstruction algorithm for small bodies from incomplete contour image. *Sci. Technol. Eng. (in Chinese)*, 24(10), 3974–3983. <https://doi.org/10.12404/j.issn.1671-1815.2305279>
- Guo, J. R., Liu, Q., Wen, K. K., Xuanyuan, M. Y., and Zhang, Y. X. (2024b). Review of small celestial body detection and 3D reconstruction methods. *Space Int. (in Chinese)*, (12), 18–23. <https://doi.org/10.3969/j.issn.1009-2366.2024.12.003>
- Huttenlocher, D. P., Klanderman, G. A., and Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(9), 850–863. <https://doi.org/10.1109/34.232073>
- Laurentini, A. (1994). The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(2), 150–162. <https://doi.org/10.1109/34.281861>

- doi.org/10.1109/34.273735
- Lauretta, D. S., Balram-Knutson, S. S., Beshore, E., Boynton, W. V., Drouet D' Aubigny, C., Dellagiustina, D. N., Enos, H. L., Golish, D. R., Hergenrother, C. W., ... Sandford, S. A. (2017). OSIRIS-REx: Sample return from asteroid (101955) Bennu. *Space Sci. Rev.*, 212(1-2), 925–984. <https://doi.org/10.1007/s11214-017-0405-1>
- Li, C. L., Liu, J. J., Ren, X., Yan, W., Zhang, Z. B., Li, H. Y., and Ouyang, Z. Y. (2024). Scientific objectives and payloads configuration of the Tianwen-2 mission. *J. Deep Space Explor. (in Chinese)*, 11(3), 304–310. <https://doi.org/10.15982/j.issn.2096-9287.2024.20230185>
- Li, J. X. (2022). Contour-based 3D modeling and application of small celestial bodies [Master's Thesis] (in Chinese). Harbin: Harbin Institute of Technology. <https://doi.org/10.27061/d.cnki.ghgdu.2022.001666>
- Li, W. Z., Wei, Y., Li, J. Y., and Yang, W. (2025). A promising view to the Moon: Anticipations for Tianwen-2 samples returned from Kamo'oalewa. *Earth Planet. Phys.*, 9(4), 782–788. <https://doi.org/10.26464/epp2025015>
- Lorensen, W. E., and Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 163–169). Anaheim, California, USA: ACM. <https://doi.org/10.1145/37401.37422>
- Martin, W. N., and Aggarwal, J. K. (1983). Volumetric descriptions of objects from multiple views. *IEEE Trans. Pattern Anal. Mach. Intell., PAMI-5*(2), 150–158. <https://doi.org/10.1109/TPAMI.1983.4767367>
- Nadipally, M. (2019). Optimization of methods for image-texture segmentation using ant colony optimization. In D. J. Hemanth, et al. (Eds.), *Intelligent Data Analysis for Biomedical Applications: Challenges and Solutions* (pp. 21–47). London: Academic Press. <https://doi.org/10.1016/B978-0-12-815553-0.00002-1>
- Shi, Y. J., and Li, L. (2025). Near-Earth asteroids: Characteristics, defense strategies and prospects for resource development. *Acta Mineral. Sin. (in Chinese)*, 45(2), 288–310. <https://doi.org/10.3724/j.1000-4734.2024.44.090>
- Tsuda, Y., Yoshikawa, M., Abe, M., Minamino, H., and Nakazawa, S. (2013). System design of the Hayabusa 2—Asteroid sample return mission to 1999 JU3. *Acta Astronaut.*, 91, 356–362. <https://doi.org/10.1016/j.actaastro.2013.06.028>
- Wang, R., Di, K., Wan, W., Liu, Z., Wang, Y., Liang, W., Wang, Y., Chen, X., and Zhi, S. (2020). Topographic mapping and analysis based on 3D reconstruction model of simulated asteroid. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B3-2020*, 1165–1170. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1165-2020>
- Xu, X. L., Li, X. M., Li, G. Q., Shi, R., and Li, H. (2002). A fast algorithm of marching cubes in volume reconstruction. *J. Syst. Simul. (in Chinese)*, 14(4), 509–513. <https://doi.org/10.3969/j.issn.1004-731X.2002.04.029>
- Xu, Y. L., Yang, Q., Yang, K. F., Hwang, J. N., and Zhu, J. F. (2021). Challenges and key technologies of point cloud quality assessment. *J. Commun. Univ. China (Sci. Technol.) (in Chinese)*, 28(5), 11–21. <https://doi.org/10.16196/j.cnki.issn.1673-4793.2021.05.002>
- Zhang, J. L. (2021). Research on 3D reconstruction of small body based on silhouette (in Chinese). Harbin: Harbin Institute of Technology. <https://doi.org/10.27061/d.cnki.ghgdu.2021.002849>
- Zhang, R. Q., Zhang, H., Liu, J. J., Geng, Y., Meng, L. Z., Li, H. T., Yu, T. Y., and Xue, C. B. (2025). Tianwen-2 small bodies exploration mission. *Sci. Sin.: Phys., Mech. Astron. (in Chinese)*, 55(7), 279501. <https://doi.org/10.1360/SSPMA-2025-0083>
- Zheng, T. X., Huang, S., Li, Y. F., and Feng, M. C. (2020). Key techniques for vision based 3D reconstruction: A review. *Acta Autom. Sin. (in Chinese)*, 46(4), 631–652. <https://doi.org/10.16383/j.aas.2017.c170502>