Supplementary information for Vision-based sampling implementation in the Chang'e-6 lunar farside sample return mission

Congjia Chen[1], Xiaoyu Jia[2], Gao Zhang[2], Yanhong Zheng[2], Xiangjin Deng[2], Yufu Qu[1]

[1]School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China

[2]Beijing Institute of Spacecraft System Engineering, Beijing 100094, China

E-mail: Yufu Qu (qyf@buaa.edu.cn) or Xiangjin Deng (dengxiangjin@sina.com)

In supplementary information, we provide more introduction about the relevant parts of our method. In the first section, we introduce the detail of the 3D reconstruction, and analyze the accuracy of the reconstructed point cloud. In the second section, we introduce the calculation of the mechanical arm's end localization.
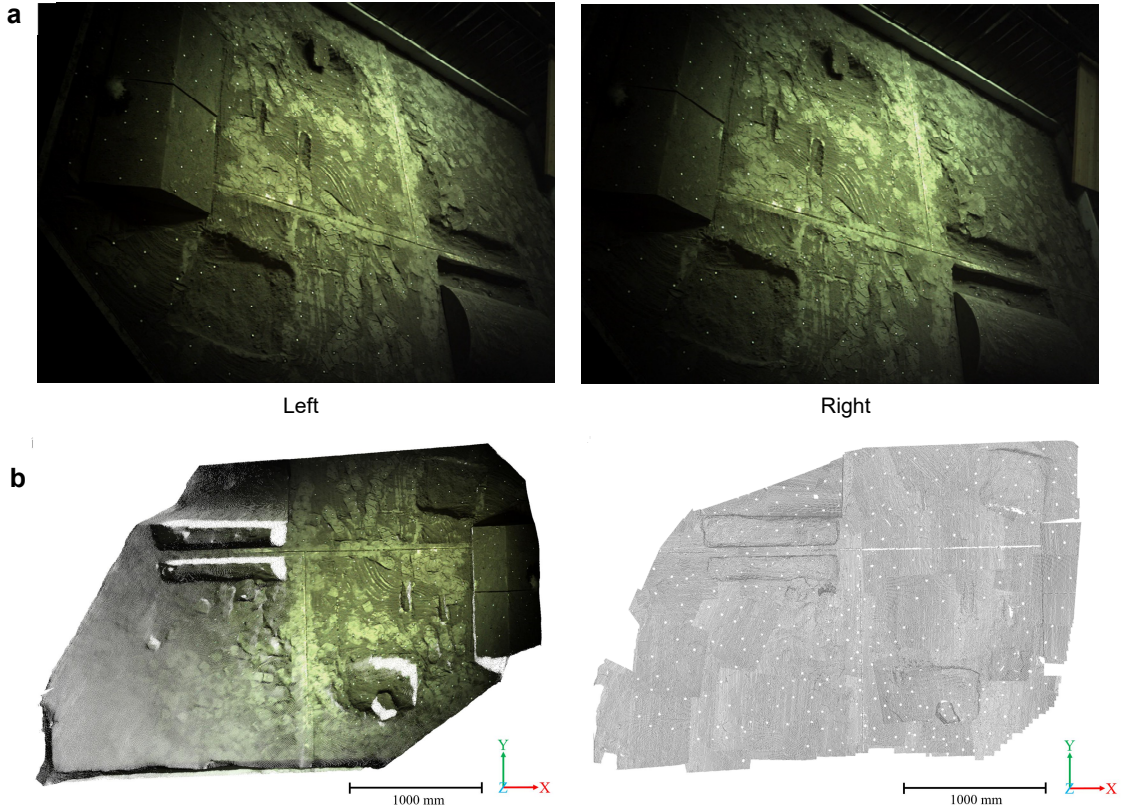
Detail of 3D reconstruction

For 3D reconstruction, the intrinsics and extrinsics of the binocular camera, and the transformation between camera coordinate system and mechanical arm coordinate system are required. The calibration was performed on earth, using three theodolites and the targets placed on the testing ground to measure corresponding points in both camera coordinate system and mechanical arm coordinate system, and determine the required parameters[1]. As shown in Fig. S1, to reconstruct 3D point cloud, the original image pair was first rectified and then input to the stereo matching networks to determine the disparity of each pixel in the left image. Then, the 3D coordinates of each pixel were reconstructed as follows:

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = \mathbf{Q} \begin{bmatrix} u \\ v \\ d \\ 1 \end{bmatrix} \tag{1}$$

where d is the disparity of image pixel (u, v), $\mathbf{Q}$ is the 4×4 disparity-to-depth mapping matrix obtained by stereo rectification. The pixels were then projected into 3D space using the reconstructed 3D coordinates to obtain point cloud, and the point cloud was transformed into the mechanical arm coordinate system for subsequent process. Notably, the distant portion of the reconstructed point cloud exceeded the range of motion of the mechanical arm. Therefore, only the part within the reachable region of the mechanical arm was chosen for subsequent process.
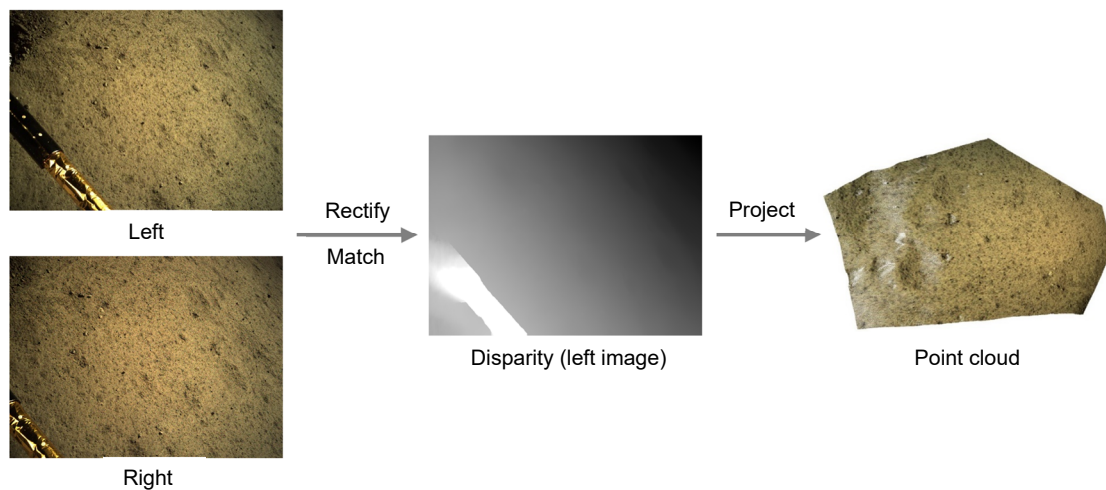
Fig. S1 Pipeline of 3D reconstruction. The disparity was normalized into [0, 255] for image visualization, the invalid points of the reconstructed point cloud were removed.

Fig. S2 Reconstruction of testing ground. a The image pair captured by the binocular camera. **b** The reconstructed point cloud (left) and the ground truth point cloud (right).

For evaluating the accuracy of 3D reconstruction, a structured light scanner was used to obtain ground truth point cloud of the testing ground. The captured images of the testing ground are shown in Fig. S2a, the reconstructed point cloud and the ground truth point cloud are shown in Fig. S2b, the resolution of the ground truth point cloud is about 1 mm.

The mean absolute distance is used as the evaluation metric of 3D reconstruction[2], which measures the distance from points in one point cloud to the closest point in the other point cloud:

$$\mathbf{P}^t = \{\mathbf{P} \in \mathbf{P}^r \mid \lVert \mathbf{P} - \underset{\mathbf{P}' \in \mathbf{P}^g}{\operatorname{argmin}} \lVert \mathbf{P} - \mathbf{P}' \rVert_2 \rVert_2 \leq \lambda\} \tag{2}$$

$$d_a = \frac{1}{|\mathbf{P}^t|} \sum_{i=1}^{|\mathbf{P}^t|} \lVert \mathbf{P}_i^t - \underset{\mathbf{P}' \in \mathbf{P}^g}{\operatorname{argmin}} \lVert \mathbf{P}_i^t - \mathbf{P}' \rVert_2 \rVert_2 \tag{3}$$

where $\mathbf{P}^r$ is the reconstructed point cloud and $\mathbf{P}^g$ is the ground truth point cloud, λ is the distance threshold used to eliminate the influence of non-overlapped points. After alignment, the mean absolute distance d_a between reconstructed point cloud and ground truth point cloud was 2.84 mm, with a threshold λ set to 30 mm. A point-wise absolute distance analysis is also provided, the results are shown in Figs. S3a and S3b. The results show that the points with high reconstruction errors are only located in further distance. Most of the points within the sampling region have an absolute distance below 2 mm, which demonstrates the remarkable accuracy of 3D reconstruction. Since only the points within the sampling region (the reachable region of the mechanical arm) are used for subsequent process, the reconstructed point cloud accurately describes the local terrain property, and provides reliable precision for the pose measurement.
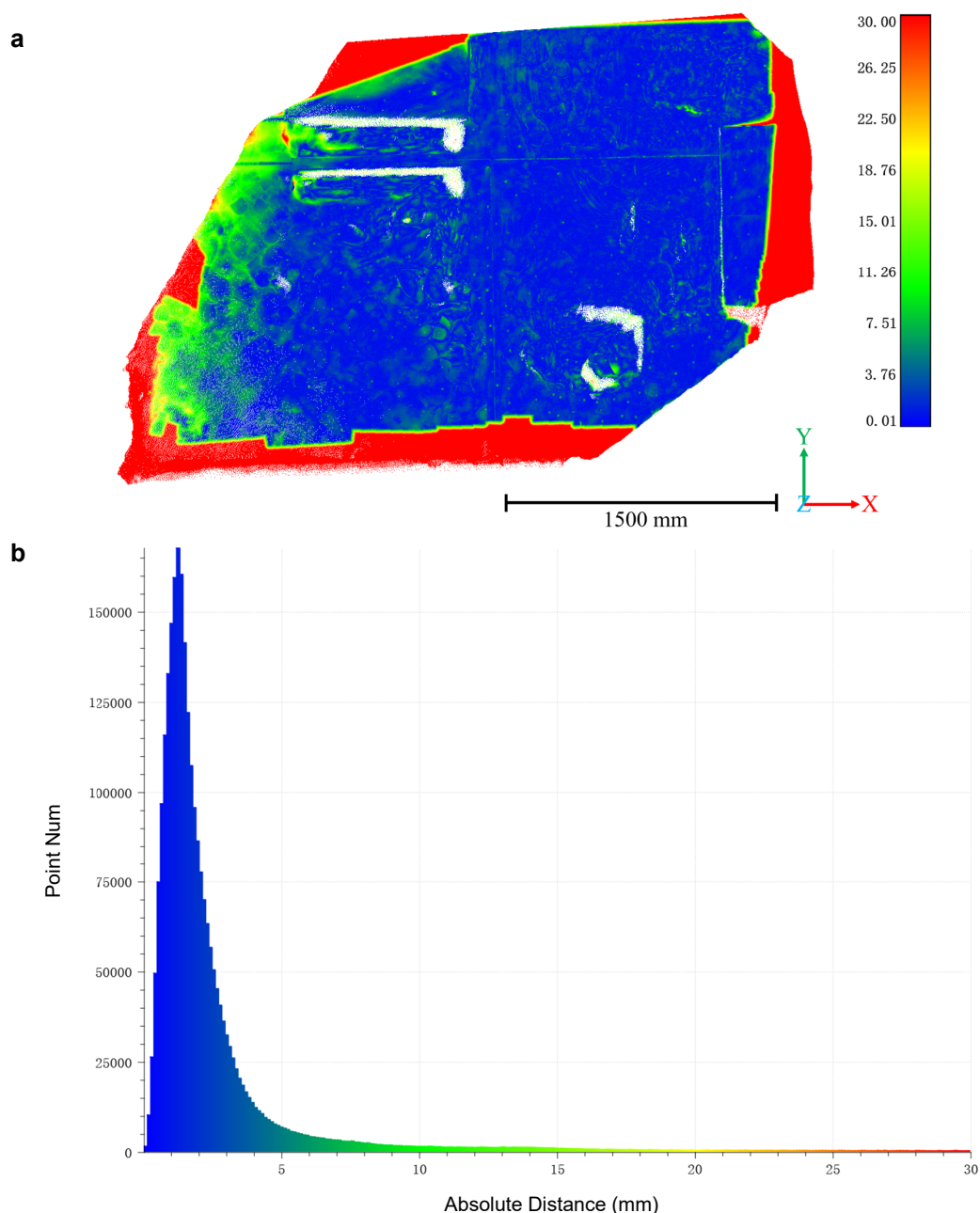
Fig. S3 Evaluation results of 3D reconstruction. a Visualization of the point-wise absolute distance, red denotes non-overlapped region. **b** Number of points at different absolute distances.

End localization of the mechanical arm

The end localization of the mechanical arm was used in the fine-tuning step. In this section, we introduce the calculation of end localization, which provides an additional measurement for the pose of the sampler. The structure of the mechanical arm is shown in Fig. S4a, which has four joints. Fig. S4b provides a detailed illustration, with coordinate systems 0-4 defined for base and each joint. The lengths of the two arms are denoted as a_2 and a_3, while $d_1 - d_4$ represent the lengths of the four joints. The mechanical arm coordinate system is defined as the base coordinate system (coordinate system 0). After the definition of each coordinate system, the transformation matrices between two coordinate systems can be determined using the D-H parameters shown in Tab. S1, and thus the transformation between base coordinate system (coordinate system 0) and end coordinate system (coordinate system 4) is determined:
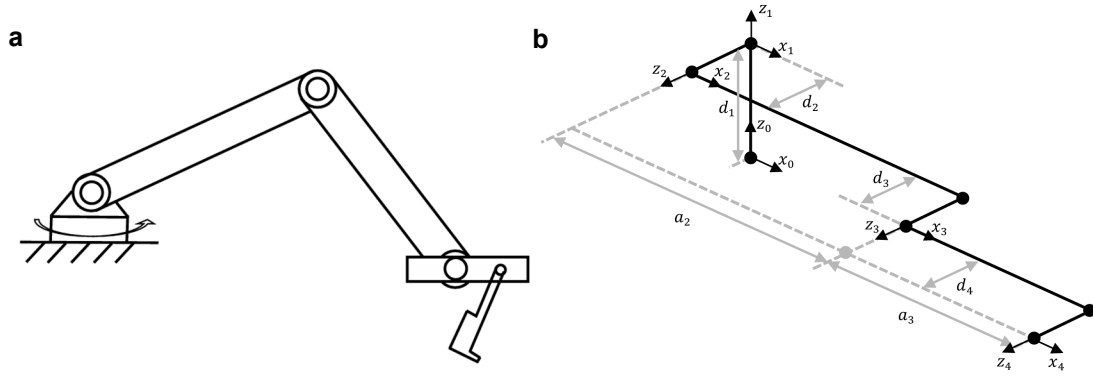
Fig. S4 Structure of the mechanical arm. a Sketch of the mechanical arm. **b** Illustration of the four joints and their coordinate systems.

Tab. S1 The D-H parameters of the mechanical arm

i	θ_i	$\alpha_{i-1}/°$	a_{i-1}/mm	d_i/mm
1	θ_1	0	0	101.0
2	θ_2	90	0	85.5
3	θ_3	0	1970	96.0
4	θ_4	0	1770	93.0

$$\mathbf{T}_1^0 = \begin{bmatrix} cos\theta_1 & -sin\theta_1 & 0 & 0 \\ sin\theta_1 & cos\theta_1 & 0 & 0 \\ 0 & 0 & 1 & d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

$$\mathbf{T}_2^1 = \begin{bmatrix} cos\theta_2 & -sin\theta_2 & 0 & 0 \\ 0 & 0 & -1 & -d_2 \\ sin\theta_1 & cos\theta_1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

$$\mathbf{T}_3^2 = \begin{bmatrix} cos\theta_3 & -sin\theta_3 & 0 & a_2 \\ sin\theta_3 & cos\theta_3 & 0 & 0 \\ 0 & 0 & 1 & d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

$$\mathbf{T}_4^3 = \begin{bmatrix} cos\theta_4 & -sin\theta_4 & 0 & a_3 \\ sin\theta_4 & cos\theta_4 & 0 & 0 \\ 0 & 0 & 1 & d_4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

Therefore, using the joint angles, the position of the end in the base coordinate system is determined. End localization can measure the pose of the sampler. However, as discussed in the main paper, end localization has an error of several centimeters, and is only accurate when measuring small relative displacements.

Reference

1. Ma, Y. Q. et al. Experimental research on a precise terrain reconstruction algorithm based on stereo images for the Chang'E-5 lunar sample return mission (in Chinese). *Sci Sin Tech* **52**, 1509–1520 (2022).

2. Aanæs, H. et al. Large-scale data for multiple-view stereopsis. *International Journal of Computer Vision* **120**, 153-168 (2016).