

RESEARCH ARTICLE

A novel precise pose prediction algorithm for setting the sleeping mode of the Yutu-2 rover based on a multiview block bundle adjustment

Song Peng¹, Youqing Ma^{2,3,*} , Xinchao Xu⁴, Yang Jia¹ and Shaochuang Liu^{2,3}

¹Beijing Institute of Spacecraft System Engineering, Beijing, 100094, China, ²International Research Center of Big Data for Sustainable Development Goals, Beijing, 100094, China, ³Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China, and ⁴School of Geomatics, Liaoning Technical University, Fuxin, Liaoning, 123000, China

*Corresponding author. E-mail: mayq@radi.ac.cn

Received: 3 August 2021; **Revised:** 21 March 2022; **Accepted:** 30 March 2022; **First published online:** 2 May 2022

Keywords: Yutu-2 rover, sleeping mode, pose prediction, terrain reconstruction, multiview block bundle adjustment, image network

Abstract

To set the sleeping mode for the Yutu-2 rover, a visual pose prediction algorithm including terrain reconstruction and pose estimation was first studied. The terrain reconstruction precision is affected by using only the stereo navigation camera (Navcam) images and the rotation angles of the mast. However, the hazard camera (Hazzcam) pose is fixed, and an image network was constructed by linking all of the Navcam and Hazzcam stereoisimages. Then, the Navcam pose was refined based on a multiview block bundle adjustment. The experimental results show that the mean absolute errors of the check points in the proposed algorithm were 10.4 mm over the range of L from 2.0 to 6.1 m, and the proposed algorithm achieved good prediction results for the rover pose (the average differences of the values of the pitch angle and the roll angle were -0.19 degrees and 0.29 degrees, respectively). Under the support of the proposed algorithm, engineers have completed the remote setting of the sleeping mode for Yutu-2 successfully in the Chang'e-4 mission operations.

1. Introduction

The lunar surface modes of Yutu-2 [1, 2] are divided into a working mode and a sleeping mode, as shown in Fig. 1. During its operation, Yutu-2 will experience the lunar night (lasting approximately two Earth weeks) during which the temperatures will fall to as low as -180°C . Due to the lack of illumination during the lunar night, the solar cell on the wings cannot provide sufficient power. To avoid these super low-temperature and nonillumination issues, engineers must set the sleeping mode for Yutu-2 before entering the lunar night. In the actual setup process of the sleeping mode for Yutu-2, the rover pose (the yaw angle γ_y , the pitch angle γ_p and the roll angle γ_r) plays a key role. There are two major considerations with respect to the application performance for the equipment on the Yutu-2 rover: the solar wing and the radioisotope heater unit (RHU) [3].

1. In the thermal testing of Yutu-2 on Earth, the left solar wing operated well in situations when the angle between the light and the plane of the left solar wing (ϑ) was more than 15.0 degrees; then, at that point the rover will be awakened. Hence, to ensure that the requirement of the angle ϑ is satisfied, the roll angle of the rover must be adjusted. When Yutu-2 wakes up, the temperature of the equipment must not be either too low or too high. If the rover leans too far to the left, it will be awakened too early, just after the sun rises. In that case, some of the equipment cannot work properly due to the low temperature. If the rover tilts too much to the right, it will be awakened

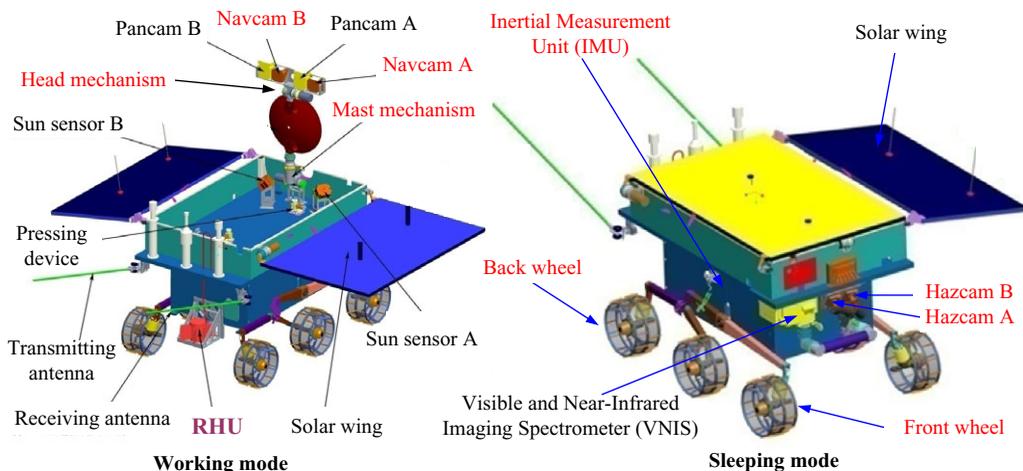


Figure 1. Lunar surface modes of Yutu-2.

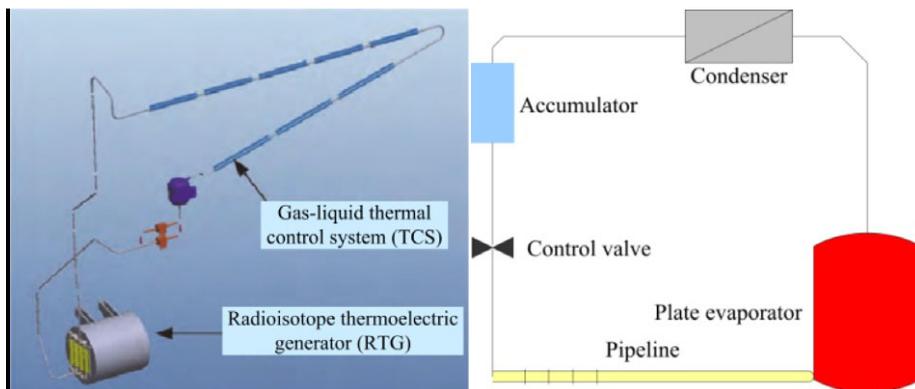


Figure 2. Left: 3D model of the heat and electricity cogeneration system; right: Composition of the TCS.

when the sun has risen too much. In that case, the temperature could be beyond a limit, and the equipment on board would not work properly.

- Yutu-2 has a RHU that provides 120 W of thermal power [3]. An RHU is a small device that provides heat through radioactive decay. RHUs are especially useful on the Moon because of its lengthy and cold two-week night. Yutu-2 needs to survive on the lunar surface for a long time, and this requires that the temperature control task during the lunar night be completed successfully. In the Chang'e-4 (CE-4) mission, a gas–liquid thermal control system (TCS) was placed in Yutu-2 to transmit the thermal energy to various parts of Yutu-2 with maximum efficiency, as shown in Fig. 2.

Figure 2 shows how the waste heat of the radioisotope thermoelectric generator (RTG) is centralized and dispersed. The TCS is connected to the RTG through a plate evaporator, which collects waste heat and transmits it into the cabin. If there is a rotation angle deviation in Yutu-2, the TCS will be adversely affected. As shown in Fig. 2, there is a height difference between the accumulator and the plate evaporator. Hence, the liquid flow in the TCS is driven by the one-sixth of the gravity on the moon's surface. The ammonia circulates through a control valve, a pipeline, into the plate evaporator, in which the liquid ammonia is vaporized, and then to a condenser, where the gaseous ammonia is converted to a liquid

and the heat is reabsorbed by the cabin. If the accumulator is not higher than the plate evaporator due to the incline angle of the rover body, the liquid ammonia cannot flow to the plate evaporator. Hence, the Yutu-2 rover must be placed on level ground to ensure that the TCS is positioned approximately horizontal.

As described above, due to the requirements of the solar wing and [3] the requirements of the TCS, the rover pose must satisfy certain limits prior to entering the Yutu-2 rover sleeping mode. As the lunar night approaches, the rover must stop working. If, at that the moment, the measured angles of Yutu-2 in the current position do not meet the above three conditions, the result will be an extremely large drain on the energy supply. Hence, the rover must adjust its pose before entering the lunar night. Due to the construction characteristics of the rover, the servo system cannot change the shape or reduce the height of one side of the wheels. In that case, the rover cannot adjust the pose by using the servo system. However, there are two other methods to adjust the rover pose. The first method is that the rover digs a hole in the lunar surface at the current position by using the wheels. The second method is by moving to a new location where the rover pose can match the required attitude conditions. Due to work efficiency and equipment security, the second method is the better choice. Hence, the problem of searching for a rover pose that exactly meets the requirements is the key to the rover's remote operating system. Accurate pose prediction of Yutu-2 before entering the sleeping mode is critical for short-term path planning [4], attitude control of the rover, and instrumentation setting in the lunar mission, which will save the Yutu-2 rover's energy, improve the work efficiency and reduce the risk of the lunar exploration project.

Due to the narrow scope of this area of interest, research on the pose prediction method for a planetary rover has not been studied before, while research has mainly been focused on simultaneous visual localization and mapping (SLAM) [1, 5–9], path planning [4, 10–13] and visual prediction of terrain occlusion [2]. In a real planetary environment, the research of the planetary rover's pose prediction has mainly been focused on how to realize a rover's perception and positioning. To accomplish the goal of setting the sleeping mode for the rover, the surrounding information obtained from various sensors is a key factor of pose prediction. However, pose prediction for a planetary rover is a challenging problem due to the lack of available data in a certain environment on the planetary surface.

2. Related work

Multiple sensors on planetary rovers, such as star sensors, radar sounders [14], laser altimeters [15] and various imaging sensors, have been used to detect the external environment. These various engineering data may be used to make a pose prediction for the planetary rover. The onboard star sensor is used only in the initial orientation of the rover and does not work when the rover moves continuously. In other words, the star sensor is only used to measure the rover pose in its current position, while it cannot be used to predict the rover pose in the next position. The Lunar Radar Sounder (LRS) onboard the Kaguya spacecraft (SELENE) has been exploring the lunar subsurface [14]. The LRS uses the HF band (5 MHz), which enables subsurface data to be obtained to a depth of several kilometers. The range measurement precision of the Lunar Orbiter Laser Altimeter (LOLA) [15, 16] has a 10 cm standard deviation. The image data provide abundant structural shape and texture information on landscape objects, which radar sounders and laser altimeters cannot provide. The imaging sensors mainly include various cameras on the orbiter, lander and rover. At the same time, due to the resolution limits, the images from the orbiter's cameras cannot be readily used to predict the pose of the rover. The images from the lander's descent camera also have limitations in terms of image coverage [1]. In practice, the detection zone of the planetary rover is beyond the scope of the image coverage of the descent camera after a long movement. Synthesizing the measurement precision, resolution limit and measurement range, the images from the rover's cameras are the best option for the rover's environmental perception and pose prediction.

To set the sleeping mode for the planetary rover, visual localization, precise terrain reconstruction and rover pose estimation are crucial aspects of pose prediction. The Yutu-2 rover is one of the robot systems that can perceive the obstacle laden external environment and self-determine its own instructions to allow

it to move to an object in the planetary environment. Similar to the Chinese Chang'e-3 (CE-3) Yutu lunar rover, the Yutu-2 rover's onboard navigation system consists mainly of an inertial measurement unit (IMU), an odometer and several stereo cameras, including Navcams, Pancams (Panoramic camera) and Hazcams, to satisfy the exploration requirements, as shown in Fig. 1. In our previous work [1], a precise visual localization method was used to obtain a precise rover pose for the current position. In the terrain reconstruction procedure, the key points are dense matching, forward intersection and digital elevation model (DEM)/digital orthophoto map (DOM) generation, which have been successfully applied in the photogrammetry area [4, 6, 7]. However, those methods do not consider the measurement errors of the rotation angles of the head and the mast mechanism.

The movement mechanism systems in the CE-3 rover, CE-4 rover or other lunar rovers belong to a kind of multijoint rotary platform. Other technologies have also been proposed for precise visual terrain reconstruction tasks using multijoint rotary platforms. One key problem is the conversion of the camera frame and the rover frame. In the Mars Exploration Rover (MER) mission, the CAHV model from the JPL, which used four vectors of C, A, H and V to describe the transformation from the object space to the image plane, was proposed [17]. By augmenting the vectors O and R, which describe the radial distortion, the CAHV model was upgraded to the CAHVOR model. The CAHVOR model was applied to both Spirit and Opportunity. The engineering cameras of the Mars Science Laboratory (MSL) included a stereo pair of navigation cameras and two pairs of hazard avoidance cameras. The camera design of the MSL was the same as that of the MER except for the baseline of the Navcam system, which was increased to 42.6 cm. The position accuracy in the depth direction increased by more than twice after the baseline increased. The camera model of the MSL was the same as that of the MER [18]. However, the CAHV model does not consider the measurement errors of the head and mast mechanism on the rover. In ref. [19], the authors proposed an integrated self-calibration method with two additional constraints. The first constraint was an error model of the mast mechanism, and the second constraint was an error model of the IMU. In this method, the actual value of the twist must be given and measured repeatedly. However, the onboard navigation system of the rover cannot provide the actual rotation angles of the mast mechanism when the rover is in lunar service. In ref. [2], a Loop-closed block bundle adjustment (BA) model employing the 3D coordinates of the feature points and the distance between the corners of the solar cells was proposed. Thus, the measurement errors of the rotation angles of the head and the mast mechanism can be avoided to a certain extent. However, in most cases, only seven or eight pairs of stereo Navcam images in the angle range of +60 degrees (left) to -60 degrees (right) along the front direction, and the rear of the rover will not be in the FOV of the Navcam. Thus, the Loop-closed block BA method can only avoid the measured yaw angle of the mast. In other words, the measured errors of the expand angle of the mast and the pitch angle of the head are not deleted. To improve the terrain reconstruction precision, some other data should be used. DOM matching between the topography product from the rover's cameras and the topography product from the descent camera were applied in the CE-3 mission [8]. However, only low height descent images should be used to retain the precision of the DOM matching. Thus, the FOV of the descent camera will make the DOM matching technology work less well. It is noted that the DOM matching between the Navcam images and the descent images can be used for path planning only in the initial work procedure of the rover. When the rover is not in the FOV of the descent camera, the DOM matching method cannot work. However, this method inspires us to use multiviews from different cameras to avoid errors in the rotation angles of the head and the mast mechanism and improve the precision of the terrain reconstruction. Hence, a novel multisensor terrain reconstruction method should be studied instead of the classical terrain reconstruction method with stereo Navcam images.

Although there are no related studies of pose estimation for planetary rovers, researchers have performed a large amount of research on robot systems, and most of the research on autonomous vehicles has been conducted on uneven terrains [20]. In ref. [21], the pose estimation problem is formulated as a rigid body contact problem for a given location of the vehicle's center of mass over the terrain and a given yaw angle. The authors [22] describe a method for pose estimation of four-wheeled vehicles, which utilizes the fixed resolution of the DEMs to generate a detailed vehicle model. In ref. [23], a

geometric method was used with terrain geometry, robot geometry and stability criteria. This combination of features was also used in the area of planetary rovers. These methods represent the terrain as a DEM, which is split into overlapping, approximately rover-sized patches. In addition, when those methods were applied in the Yutu-2 rover, the adaptability and further application of the proposed algorithm should be considered.

In this paper, a visual prediction method of the rover pose is presented for the first time, which includes precise visual localization, high-precision terrain reconstruction and pose prediction based on terrain products. There are two primary motivations for visual pose prediction:

1. Visual pose prediction technology is more useful than a technology using star sensors, radar sounders and/or laser altimeters. The proposed algorithm can continuously provide more accurate pose information for the planetary rover in lunar service.
2. Due to the rotation angle errors of the head and the mast mechanism on the Yutu-2 rover, the precision of the terrain reconstruction by using the stereo Navcam images decreases in classical methods [1, 2, 4, 6–8, 17, 19]. In this paper, a multiview block bundle adjust (BA) model based on Navcam/Hazcam stereoisimages was first proposed for Navcam pose estimation. It will meet the needs of the required mapping precision and improve the calculation efficiency of the path planning.

As stated above, the proposed algorithm yields the following contributions:

1. From the practical aspect of technological innovation, the proposed visual pose estimation algorithm was first developed to set the sleeping mode of Yutu-2 before entering the lunar night. By foreseeing the pose estimation results for Yutu-2 in the current position, engineers can make informed decisions about the following procedures, including short-term path planning, attitude control of the rover and instrumentation setting in the lunar mission.
2. The theoretical innovation of this paper is to propose a multiview block BA algorithm for the precise terrain reconstruction problem of a multijoint rotary platform. Compared with other classical methods [1], [2], [4], [6]–[8], [17], [19], the Navcam/Hazcam stereoisimages are used to construct an image network, and the refined Navcam pose can be obtained. In other words, the measured rotation angle errors of the multijoint rotary platform are removed to some extent, and the low-precision phenomenon of visual terrain reconstruction can be avoided. Hence, the proposed algorithm could not only be used in planetary rovers but could also be used in other multijoint rotary platforms, which provides a novel solution to the visual precise terrain reconstruction problem. In the proposed multiview block BA algorithm, the exterior orientation parameters of the Navcam/Hazcam stereoisimages are the initial values and are viewed as pseudo-observations. Meanwhile, a much more complete measurement error formula of the feature points' 3D coordinates is given.
3. Extensive comparisons with other image processing methods have been made [1, 2, 4, 6–8, 17, 19]. The results reveal that the proposed algorithm is more useful and can continuously provide precise terrain information for the rover in lunar service. In addition, a much more complete formula of the measurement errors for the 3D coordinates is given. The proposed algorithm is somewhat simpler than previous pose estimation methods [21–23], which have been successfully used in CE-4 mission operations.

The remainder of this paper is structured as follows. Section 3 is the topographic mapping capability analysis of the Navcam and Hazcam stereo pairs. Section 4 describes the proposed algorithm. In Section 5, experimental results are presented to demonstrate the effectiveness and precision of the proposed algorithm compared to the classical methods [1, 2, 4, 6–8, 17, 19]. Section 6 concludes the paper.

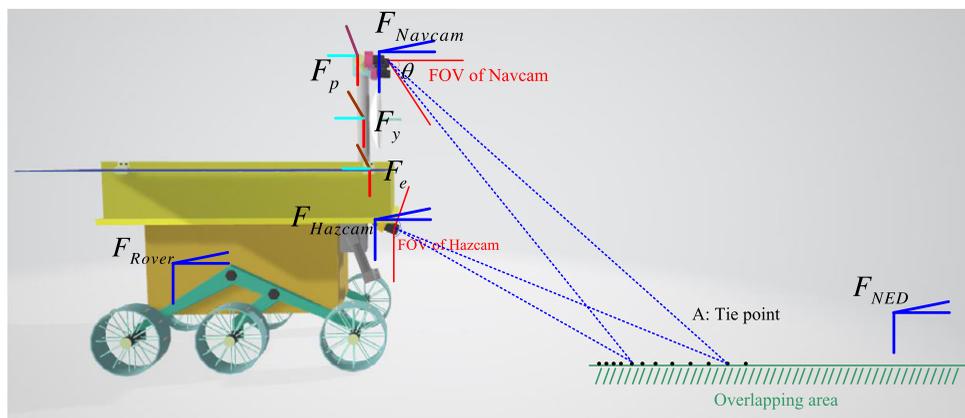


Figure 3. Multisensor terrain reconstruction concept: by matching the multidegree (e.g., four degrees) overlapping image points (tie points) between the Navcam images and the Hazcam images, the proposed algorithm produces a refined camera pose and more precise in-site mapping (side view).

3. Topographic mapping capability analysis of the Navcam and Hazcam stereo pairs

The IMU measurement data, the rotation angles (the expand angle θ_e , pitch angle θ_p and yaw angle θ_y) of the head and the mast mechanism and the stereo-images are compressed into telemetry data without loss of quality, which are then transformed into the rover's telemetry operating system on the ground (earth). The stereo Navcams and Pancams are mounted on the head of the mast mechanism. The stereo Hazcams are mounted on the front of the rover, and their geometric relationship with the rover is fixed. The mast mechanism is mounted on the rover. The Navcams, which have an effective depth of field from 0.5 m to a large value (much larger than 0.5 m), are the most important stereo-imaging systems for determining the Yutu-2 rover's localization and mapping research activities. The pixel count of the stereo Navcams is 1024×1024 pixels with a field of view (FOV) of 46.9 degrees, a stereo base of 27 cm and a focal length of 8.72 mm. The Hazcams are used for obstacle avoidance. The pixel count of the stereo Hazcams is also 1024×1024 pixels with an FOV of almost 180 degrees, a stereo base of 10 cm and a focal length of 7.28 mm. The intrinsic and extrinsic (inner and exterior orientation) parameters of these stereo cameras can be acquired by using a self-calibrating bundle adjustment (SCBA) [24] model with control points on the earth.

The involved frames include the north-east-down (NED) frame F_{NED} , pitch frame F_p , yaw frame F_y , expand frame F_e , Navcam frame F_{Navcam} , Hazcam frame F_{Hazcam} and rover frame F_{Rover} . Considering the FOV, the image coverage of the Navcam/Hazcam and the overlapping area is as shown in Fig. 3.

Figure 3 shows how the precise terrain reconstruction problem can be solved by using the tie point A between the Navcam/Hazcam images. However, the baselines of the stereo Navcams/Hazcams are short, and the intersection angles are smaller when the features are further away from the cameras. Thus, we should analyze the reconstruction precision of the features at different distances. In that case, the selection strategy of the having several A tie points plays an important role in the new multisensor terrain reconstruction method. To simplify the analysis, the stereo Navcam/Hazcam images are viewed as a normal case stereo pair, which indicates that the optical axes of the stereo cameras are parallel to each other and perpendicular to the baseline B. Figure 4 shows the configuration of the normal case stereo pair.

In Fig. 4, f is the focal length of the stereo cameras, p denotes the parallax, and O_l and O_r are the projective centers of the stereo cameras. Suppose that the left camera frame is the reference frame and the image coordinates of the left camera are (x, y) . Then, the 3D coordinates of the tie points A, denoted by (X_A, Y_A, Z_A) in the left camera frame, can be calculated using the following parallax equations [25]:

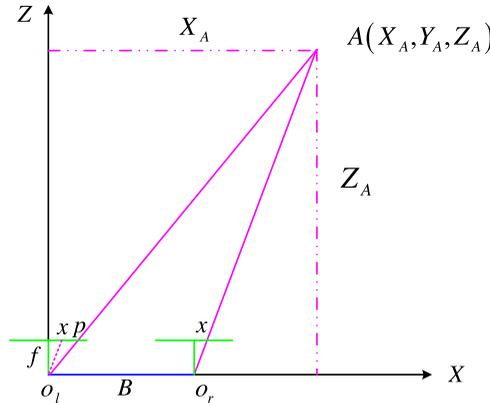


Figure 4. Configuration of the normal case stereo pair.

$$X_A = \frac{B}{p}x = \frac{Y_A}{f}x, Z_A = \frac{B}{p}f, \text{ and } Y_A = \frac{B}{p}y = \frac{Z_A}{f}y. \tag{1}$$

Through error propagation, the measurement errors (standard errors $\sigma_x, \sigma_y, \sigma_z$) of the 3D coordinates (X_A, Y_A, Z_A) can be calculated as follows:

$$\sigma_x = \sqrt{\left(\frac{Z_A^2}{Bf}\right)^2 \left(\frac{x}{f}\right)^2 \sigma_p^2 + \left(\frac{Z_A}{f}\right)^2 \sigma_x^2 + \left(\frac{x}{p}\right)^2 \sigma_B^2 + \left(\frac{x}{f}\right)^2 \left(\frac{B}{p}\right)^2 \sigma_f^2}; \tag{2a}$$

$$\sigma_y = \sqrt{\left(\frac{Z_A^2}{Bf}\right)^2 \sigma_p^2 + \left(\frac{B}{f}\right)^2 \sigma_f^2 + \left(\frac{f}{p}\right)^2 \sigma_B^2}; \tag{2b}$$

$$\sigma_z = \sqrt{\left(\frac{Z_A^2}{Bf}\right)^2 \left(\frac{y}{f}\right)^2 \sigma_p^2 + \left(\frac{Z_A}{f}\right)^2 \sigma_y^2 + \left(\frac{y}{p}\right)^2 \sigma_B^2 + \left(\frac{y}{f}\right)^2 \left(\frac{B}{p}\right)^2 \sigma_f^2}; \tag{2c}$$

where σ_p is the measurement error of the parallax and σ_x , and σ_y are the measurement errors of the image coordinates in the horizontal and vertical directions, respectively. σ_B denotes the measurement errors of the baseline, which can be obtained using the following parallax equation:

$$\sigma_B = \frac{(X_l - X_r) (\sigma_{X_l} - \sigma_{X_r}) + (Y_l - Y_r) (\sigma_{Y_l} - \sigma_{Y_r}) + (Z_l - Z_r)(\sigma_{Z_l} - \sigma_{Z_r})}{B}, \tag{3}$$

where (X_l, Y_l, Z_l) and (X_r, Y_r, Z_r) represent the projective centers of the stereo cameras in the rover frame F_{Rover} , and $(\sigma_{X_l}, \sigma_{Y_l}, \sigma_{Z_l})$ and $(\sigma_{X_r}, \sigma_{Y_r}, \sigma_{Z_r})$ are the corresponding measurement errors.

The emulation and data analysis of the topographic mapping capability indicate that the expected measurement error is reasonable (<1.4 mm) within a range of 1.5 m for the Navcams and 0.6 m for the Hazcams, less than 4.1 mm within a range of 2.5 m for the Navcams and 1.0 m for the Hazcam sand that it is less than 5.8 mm within a range of 3.0 m for the Navcams and 1.2 m for the Hazcams. Since the range error is the dominant error, we use the range error to discuss the selection of the tie points in the feature matching procedure. In other words, the results indicate that the proposed algorithm based on the Navcam and Hazcam stereoimages has theoretical feasibility. However, the above 3D coordinates (X, Y, Z) of those feature points are all in the left camera frame (F_{Navcam}, F_{Hazcam}) .

In the terrain reconstruction procedure based on the stereo Navcam images, the feature points in F_{Navcam} must be translated into F_{NED} by using the coordinate system transformation framework and the rover pose [1]. It is noted that the coordinate system transformation framework includes F_p, F_y and F_e , which connect F_{Rover} and F_{Navcam} . The measurement precisions of the rotation angles of the head and the mast mechanism are 0.3 degrees. When the visual pose prediction method is applied, the measurement

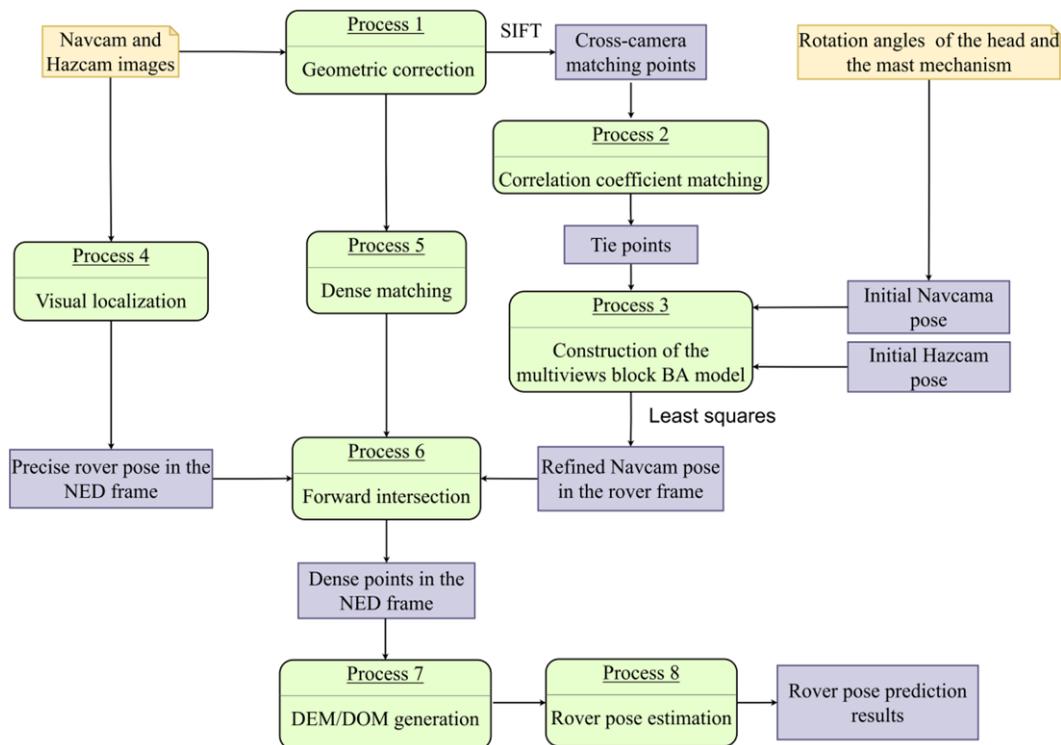


Figure 5. Workflow of the proposed visual pose prediction algorithm.

errors of the rotation angles of the head and the mast mechanism will decrease the precision of the rover's localization and mapping, which is noteworthy. Hence, this low-precision terrain reconstruction phenomenon should be avoided, which could inspire researchers to seek out other methods, such as using multisensor images.

4. Proposed method

4.1. Workflow of the proposed visual pose prediction algorithm

In this paper, we present the proposed visual pose prediction algorithm using the Navcam and Hazcam stereoimages. Figure 5 shows the workflow of the proposed algorithm.

After geometric correction, the geometric distortion of the Navcam images and Hazcam images is deleted. Because the scale-invariant feature transform (SIFT) [26] is the most invariant to scale, rotation and viewpoint changes, SIFT is used to match the left images from Navcam and Hazcam, and the cross-camera matching points are obtained. Based on those points, the correlation coefficient matching method is applied in the stereo matching procedure. Then, the tie points (e.g., four degrees) in the stereo image sequence from different cameras can be obtained. The initial Navcam pose can be obtained from the rotation angles of the head and the mast mechanism. Next, an image network based on the multiviews block BA model with the initial Navcam/Hazcam pose and the tie points is established. The ordinary least squares solution is used, and the following parameters can be obtained and are considered to be known: the refined Navcam pose in F_{Rover} and the error of the camera pose. At the same time, a precise visual localization method in our previous work [1] is used to obtain a precise rover pose in F_{NED} . Finally, dense matching [27], forward intersection, DEM/DOM generation and rover pose estimation are used to obtain the rover pose prediction results in F_{NED} .

Geometric constructions formed by the tie points between the cross-camera images are the basis for the Navcam pose estimation solution.

4.2. Refined Navcam pose by using multisensor images

4.1.1. The basic imaging model of the multiview block BA for the stereo Navcam/Hazcam system

To clarify the theory of the multiviews block BA, the collinearity equation is used [1] as follows:

$$\begin{cases} x_a + \delta x + x_0 = -f \frac{a_1 (X_A - X_S) + b_1 (Y_A - Y_S) + c_1 (Z_A - Z_S)}{a_3 (X_A - X_S) + b_3 (Y_A - Y_S) + c_3 (Z_A - Z_S)} \\ y_a + \delta y + y_0 = -f \frac{a_2 (X_A - X_S) + b_2 (Y_A - Y_S) + c_2 (Z_A - Z_S)}{a_3 (X_A - X_S) + b_3 (Y_A - Y_S) + c_3 (Z_A - Z_S)} \end{cases} \tag{4}$$

where

(f, x_0, y_0) denotes the principal distance and principle point;

(X_A, Y_A, Z_A) denotes the 3D coordinates of tie point A in F_{Rover} ; (x_a, y_a) denotes the 2D coordinates of the corresponding image point a ;

$(X_S, Y_S, Z_S, \varphi_S, \omega_S, k_S)$ denotes the exterior orientation parameters (EOPs) of the Navcam/Hazcam in F_{Rover} ; ${}^l_{Rover}r = (X_S, Y_S, Z_S)$;

$$R_l^{Rover} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} \cos \varphi_S & 0 & -\sin \varphi_S \\ 0 & 1 & 0 \\ \sin \varphi_S & 0 & \cos \varphi_S \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega_S & -\sin \omega_S \\ 0 & \sin \omega_S & \cos \omega_S \end{bmatrix} \begin{bmatrix} \cos k_S & -\sin k_S & 0 \\ \sin k_S & \cos k_S & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

$(\delta x, \delta y)$ represents the lens distortion of each image point.

4.1.2. The EOPs of the Navcam/Hazcam

When the multiview block BA is applied in the image network, the unknowns include $(X_S, Y_S, Z_S, \varphi_S, \omega_S, k_S, X_A, Y_A, Z_A)$. Given appropriate initial values of the unknowns, the BA will converge, and the optimal EOPs of the Navcam will be calculated.

The telemetry data in the CE-4 mission include the rotation angles of the head and the mast mechanism. In that case, the rotation matrix $R_{nav_I}^{Rover}$ and the vector ${}^{nav_I}_{Rover}r$ can be obtained by using the geometric relationship between F_{Navcam} and F_{Rover} , as shown in Formulas (5a) and (5b).

$$R_{nav_I}^{Rover} = R_e^{Rover} R_{expand} R_Y^e R_{yaw} R_p^Y R_{pitch} R_C^P R_{nav_I}^C \tag{5a}$$

$${}^{nav_I}_{Rover}r = R_e^{Rover} R_{expand} (R_Y^e R_{yaw} (R_p^Y R_{pitch} (R_C^P T_l^C + T_C^P) + T_P^Y) + T_Y^e) + T_e^{Rover} \tag{5b}$$

where

$$R_{expand} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_e & -\sin \theta_e \\ 0 & \sin \theta_e & \cos \theta_e \end{bmatrix} \text{ denotes the rotation matrix between } F_{Rover} \text{ and } F_e;$$

$$R_{yaw} = \begin{bmatrix} \cos \theta_y & 0 & -\sin \theta_y \\ 0 & 1 & 0 \\ \sin \theta_y & 0 & \cos \theta_y \end{bmatrix} \text{ denotes the rotation matrix between } F_e \text{ and } F_y;$$

$$R_{pitch} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_p & -\sin \theta_p \\ 0 & \sin \theta_p & \cos \theta_p \end{bmatrix} \text{ denotes the rotation matrix between } F_y \text{ and } F_p;$$

R_e^{Rover} , R_Y^e , R_p^Y , R_C^P , and $R_{nav_I}^C$ denote the known rotation matrix, and the corresponding angle measurement accuracy is $1.0''$; and T_l^C , T_C^P , T_P^Y , T_Y^e and T_e^{Rover} denote the known transformation vector.

The Hazcam pose in F_{Rover} , including the rotation matrix $R_{haz_l}^{Rover}$ and the vector $r_{haz_l}^{Rover}$, will be presented as:

$$r_{haz_l}^{Rover} = R_C^{Rover} T_{haz_l}^C + T_C^{Rover}, \tag{6a}$$

$$R_{haz_l}^{Rover} = R_C^{Rover} R_{haz_l}^C, \tag{6b}$$

where

$T_{haz_l}^C$ and T_C^{Rover} denote the transformation vector among the Hazcam frame, the cube mirror frame and F_{Rover} , respectively;

R_C^{Rover} denotes the rotation matrix between the latter two frames, where the angle measurement error is $1.0''$, and $R_{haz_l}^C$ denotes the rotation matrix between the Hazcam frame and the cube mirror frame.

4.1.3 Least squares solution of the multiview block BA model

To improve the robustness of the proposed method, the geometry constraint condition of the stereo Navcams and Hazcams should be adopted in the multiview block BA. It is noted that the relative pose of the stereo Navcams/Hazcams is obtained in advance during the camera calibration procedure.

Using Taylor’s formula, the linearization equation of Formula (4) can be derived in Gauss–Markov mode:

$$\begin{cases} v_{x_a} - \frac{\partial x_a}{\partial X} V_X - \frac{\partial x_a}{\partial Y} V_Y - \frac{\partial x_a}{\partial Z} V_Z = x_a^0 - x_a + \frac{\partial x_a}{\partial X_S} \delta X_S + \frac{\partial x_a}{\partial Y_S} \delta Y_S + \frac{\partial x_a}{\partial Z_S} \delta Z_S + \frac{\partial x_a}{\partial \varphi_S} \delta \varphi_S + \frac{\partial x_a}{\partial \omega_S} \delta \omega_S + \frac{\partial x_a}{\partial k_S} \delta k_S \\ v_{y_a} - \frac{\partial y_a}{\partial X} V_X - \frac{\partial y_a}{\partial Y} V_Y - \frac{\partial y_a}{\partial Z} V_Z = y_a^0 - y_a + \frac{\partial y_a}{\partial X_S} \delta X_S + \frac{\partial y_a}{\partial Y_S} \delta Y_S + \frac{\partial y_a}{\partial Z_S} \delta Z_S + \frac{\partial y_a}{\partial \varphi_S} \delta \varphi_S + \frac{\partial y_a}{\partial \omega_S} \delta \omega_S + \frac{\partial y_a}{\partial k_S} \delta k_S \end{cases} \tag{7}$$

where x_a^0, y_a^0 denotes the calculated values of the 2D coordinates of the image points, which can be obtained by using Formula (4);

$\frac{\partial \cdot}{\partial \cdot}$ is the derivative, and δ denotes the corrections of the EOPs;

v_{x_a}, v_{y_a} represent the correction of the 2D image points in F_{xy} ;

V_X, V_Y, V_Z represent the correction of the 3D coordinates of the tie point.

Then, Formula (7) can be described as a generalized model:

$$V_x = a_x t + B_x M - L_x \tag{8}$$

where $V_x = [v_{x_a} v_{y_a}]^T$, $t = (\delta X_S, \delta Y_S, \delta Z_S, \delta \varphi_S, \delta \omega_S, \delta k_S)^T$, $M = (\delta X, \delta Y, \delta Z)^T$, and $L_x = [x_a - x_a^0 y_a - y_a^0]^T$;

$$a_x = \begin{bmatrix} \frac{\partial x_a}{\partial X_S} & \frac{\partial x_a}{\partial Y_S} & \frac{\partial x_a}{\partial Z_S} & \frac{\partial x_a}{\partial \varphi_S} & \frac{\partial x_a}{\partial \omega_S} & \frac{\partial x_a}{\partial k_S} \\ \frac{\partial y_a}{\partial X_S} & \frac{\partial y_a}{\partial Y_S} & \frac{\partial y_a}{\partial Z_S} & \frac{\partial y_a}{\partial \varphi_S} & \frac{\partial y_a}{\partial \omega_S} & \frac{\partial y_a}{\partial k_S} \end{bmatrix}, B_x = \begin{bmatrix} \frac{\partial x_a}{\partial X} & \frac{\partial x_a}{\partial Y} & \frac{\partial x_a}{\partial Z} \\ \frac{\partial y_a}{\partial X} & \frac{\partial y_a}{\partial Y} & \frac{\partial y_a}{\partial Z} \end{bmatrix}.$$

The stochastic model for Formula (8) is:

$$V_x \sim N(0, \sigma_0^2 P_x^{-1}), \tag{9}$$

where P_x is the corresponding weight matrix and σ_0 is the prior variance component.

The EOPs of the Navcams and Hazcams can be acquired from Formulas (5)-(6) before the BA is applied. Hence, the EOPs should be viewed as pseudo-observations [1, 28] for the multiview block BA:

$$\begin{cases} v_{X_S} = \delta X_S \\ v_{Y_S} = \delta Y_S \\ v_{Z_S} = \delta Z_S \\ v_{\varphi_S} = \delta \varphi_S \\ v_{\omega_S} = \delta \omega_S \\ v_{k_S} = \delta k_S \end{cases} \tag{10}$$

Formula (10) can be viewed as:

$$V_{po} = It. \tag{11}$$

The stochastic model for Formula (11) is:

$$V_{po} \sim N(0, \sigma_0^2 P_{po}^{-1}), \tag{12}$$

where P_{po} is the corresponding weight matrix, $V_{po} = [v_{X_S} v_{Y_S} v_{Z_S} v_{\varphi_S} v_{\omega_S} v_{k_S}]^T$.

In this case, the number of observation values will increase, which further improves the estimation precision of the EOPs in the multiview block BA.

When a least squares solution is applied, the Lagrange target function can be written as $\varnothing(\epsilon, \rho) = V_x^T P_x V_x + V_{po}^T P_{po} V_{po} + 2\rho^T (Gt + W)$. In which ρ denotes the vector of the unknown Lagrange multipliers. When the first-order condition equations $\frac{\partial \varnothing}{\partial \epsilon}, \frac{\partial \varnothing}{\partial \rho}$ are set as zero, the unknowns can be viewed as:

$$\epsilon = u + (\gamma^T P_x \gamma + I_0^T P_{po} I_0)^{-1} G_0^T \left(G_0 (\gamma^T P_x \gamma + I_0^T P_{po} I_0)^{-1} G_0^T \right)^{-1} (-W - G_0 \mu), \tag{13}$$

where

$$\begin{aligned} \mu &= (\gamma^T P_x \gamma + I_0^T P_{po} I_0)^{-1} \gamma^T P_x L_x; \\ \gamma &= [a_x B_x], G_0 = [G \ 0 \ 0 \ 0], \epsilon = [tM]; \end{aligned}$$

In a single image, $I_0 = \begin{bmatrix} I_{6 \times 6} & 0 \\ 0 & 0_{3 \times 3} \end{bmatrix}$;

where

$G = [G_i^l \ G_i^r \ G_j^l \ G_j^r]$ denotes the coefficient matrices, which are partial derivatives for each variable. The subscripts i , and j represent the Navcams and Hazcams, respectively, and the superscripts l , and r represents the left camera and the right camera, respectively;

$W = -F^0 - E$, where F and E represent constant terms, and F^0 represents the constant vector when the initial extrinsic parameters are all set to zero, which can be seen in ref. [1].

It is noted that $P_{po} = \sigma_0^2 / \sigma_{po}^2 I$, and σ_{po} represents the precision of the pseudo-observations. In the case of the Navcams, the pseudo-observations are primarily influenced by the errors in the rotation angles of the head and the mast mechanism. Thus, σ_{po} can be set to σ_r , where σ_r denotes the precision (0.3°) of the rotation angles. Here, $\sigma_0 = \sigma_p$ denotes the precision value (0.2 pixels) of the matching image point. In the case of the Hazcams, the pseudo-observations are primarily influenced by the estimated precision of the camera pose. Here, σ_{po} is set to σ_m , where σ_m denotes the calculated precision ($\sigma_{\varphi_S}, \sigma_{\omega_S}, \sigma_{k_S} = 0.1^\circ, 0.06^\circ, 0.06^\circ$) of the Hazcam pose in the camera calibration procedure.

Formulas (8) and (11) can be solved based on the least squares solution, and the root mean square error (RMSE) σ_0 can be expressed as:

$$\sigma_0 = \sqrt{(V_x^T P_x V_x + V_{po}^T P_{po} V_{po}) / t_0}, \tag{14}$$

where t_0 is the redundancy. The RMSE of the unknowns is given as:

$$\sigma_u = \sigma_0 \sqrt{(\gamma^T P_x \gamma + I_0^T P_1 I_0)^{-1}}, \tag{15}$$

where $\sigma_u = (\sigma_{X_S}, \sigma_{Y_S}, \sigma_{Z_S}, \sigma_{\varphi_S}, \sigma_{\omega_S}, \sigma_{k_S})$, $P_1 = \begin{bmatrix} P_{po} & 0 \\ 0 & I_{3 \times 3} \end{bmatrix}$.

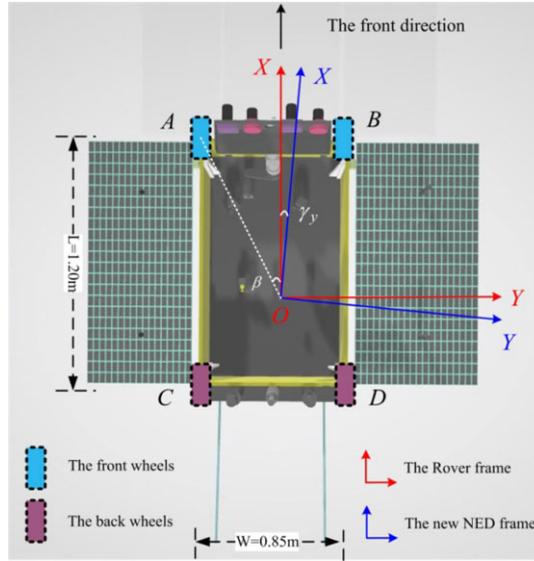


Figure 6. Rover pose estimation concept (top view).

4.3. Rover pose prediction based on the topographic product

The topographic products (DEM/DOM) in the current position are all generated with the stereo Navcam images in the rover’s telemetry operating system on the ground (earth). As the lunar night approaches, searching for the rover pose in the next station must be studied based on those topographic products. To simplify the analysis, the rover pose prediction is formulated as a rigid body contact problem for a given location of the rover’s center over the DEM and a given attitude angle γ_y . First, the plane positions of the front and back wheels are calculated by using the given location of the rover’s center O and the actual geometrical sizes. Second, the nearest points in the DEM are obtained based on the above plane positions, and the elevation of the contact points between the wheels (the front wheels A, B and the back wheels C, D) and ground can be interpolated by the kriging method with some nearby points. Third, the other attitude angles (γ_p, γ_r) of the rover will be approximately calculated through the 3D coordinates XYZ of those contact points, which are as follows:

$$\begin{cases} \gamma_p = -\text{atan} \left((Z_{AB} - Z_{CD}) / \sqrt{(X_{AB} - X_{CD})^2 + (Y_{AB} - Y_{CD})^2} \right) \\ \gamma_r = \text{atan} \left((Z_{BD} - Z_{AC}) / \sqrt{(X_{AC} - X_{BD})^2 + (Y_{AC} - Y_{BD})^2} \right), \end{cases} \tag{16}$$

where atan denotes the arctangent function, $\tau_{AB} = (\tau_A + \tau_B) / 2$, $\tau_{CD} = (\tau_C + \tau_D) / 2$, $\tau_{AC} = (\tau_A + \tau_C) / 2$, $\tau_{BD} = (\tau_B + \tau_D) / 2$, and $\tau \in \{XYZ\}$.

As shown in Fig. 6, $\beta = 35.3$ degrees, the blue lines denote the new NED frame F'_{NED} , where the origin is the origin O of F_{Rover} . Given a yaw angle γ_y and the plane location X_O, Y_O of the rover, the 2D coordinates (XY) of the centers of the wheels A, B, C, D can be obtained in F'_{NED} .

When the rover’s attitude angles (γ_p, γ_r) meet the set requirement, the rover must stop working and be ready to enter the sleeping mode in the corresponding given location.

The proposed algorithm is applied in the full DEM image search. The column and the row of each pixel in the DEM present a plane location of the rover. Thus, the above procedure should be repeated. In practice, the parallel processing technique of multithreading is applied to search for the potential location for the rover’s sleeping mode in the DEM. In the processing of the rover pose prediction based on the DEM, there are usually difficulties in that many potential results are obtained. All of the associated factors (the distance, the terrain trafficability, energy, and so on) shall be considered prior to selecting the most suitable pose (location and attitude angles) for the rover from those potential results.

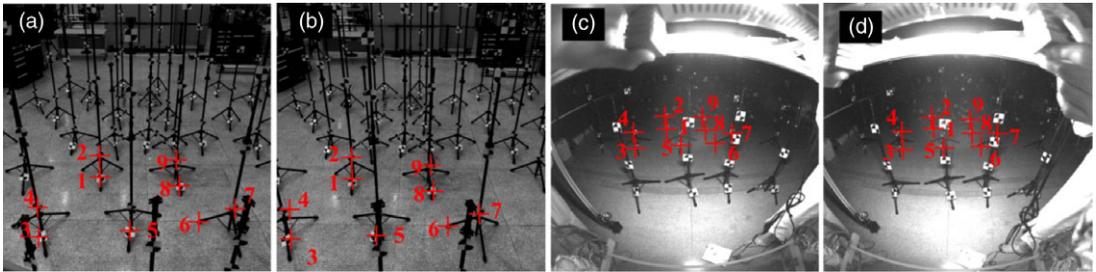


Figure 7. (a), (b) Distributions of the nine selected tie points of stereoimages in the test field. (c) The stereo Navcam image; (d) the stereo Hazcam image. The red crosses indicate the locations of the tie points, and the red numbers indicate the number of each tie point.

5. Experimental results

5.1. Experimental results of visual pose estimation for the Yutu-2 rover in the indoor field

5.1.1. Terrain reconstruction

In the terrain reconstruction procedure, the proposed method (referred to as proposed) is compared with the classical terrain reconstruction method (referred to as Classical-TR) [4, 6–8], the BA method based on the ordinary least squares (referred to as BA + Geometry) [1], the CAHVOR method [17] and the integrated self-calibration method [19] (referred to as ISC). In Classical-TR, the Navcam pose is directly obtained by using the coordinate system transformation framework [1]. In BA + Geometry, Formula (8) is invoked, while the matrix I is set as a zero matrix. Formulas (5) and (6) are only used to obtain the initial Navcam/Hazcam pose. In CAHVOR, the transformation from the object space to the image plane is obtained, which is almost the same as Classical-TR. In ISC, the measured rotation angles of the mast mechanism are viewed as the actual values and are used as inputs. In those methods, the rover localization results [1] and the dense matching process [27] are the same, which are both self-programmed in C++.

Before the CE-4 mission began, the proposed algorithm was tested with a model rover, which has the same size and functions as Yutu-2, in a simulated indoor field with artificial craters and rocks. The rover pose, the 3D coordinates of the feature points in F_{Rover} and the cube mirror frame F_C were measured, which can be viewed as the true data (the angular precision was $2.0''$ and coordinate measurement precision was 0.12 mm). It is noted that F_C was mounted on top of the left Navcam. Theoretically, the tie points should be evenly distributed in the overlapping areas between the Navcam/Hazcam FOV. However, the 3D measurement errors of the stereo Navcam/Hazcam images are another factor that affects the Navcam pose estimation precision. In other words, the tie points should be selected from the lower part of the above overlapping areas. The tie points were extracted using the Hough line detection method and the manual matching process, whose precision was almost 0.5 pixels. Figure 7 shows the distribution of the selected tie points, and the rotation angles ($\theta_e, \theta_y, \theta_p$) at the current station are 0.0 degrees, 0.09 degrees and -29.93 degrees. The distance between the various tie points and the front wheel ranges from 1.3 to 2.2 m.

To test the relationships between the left Navcam pose estimation based on different methods and the number of tie points, the following four cases were considered:

- Case 0. nine tie points (“1, 2, 3, 4, 5, 6, 7, 8, 9”);
- Case 1. Eight tie points (“1, 2, 3, 4, 5, 7, 8, 9”);
- Case 2. six tie points (“1, 3, 4, 5, 7, 8”);
- Case 3. Five tie points (“1, 2, 4, 8, 9”).

In those cases of the proposed algorithm, the number of iterations was set to 6, and the results are presented in Table I.

The column “ σ_0/pixels ” of Table I, which indicates the RMSE of the reprojections and the left Navcam pose [9], is different in the four cases. In column “ δ ” of Table I, the symbol $\delta =$

Table I. The left Navcam pose estimated results using different methods for four different cases.

	Camera position/mm				Camera altitude angles/rad							σ_0	Times(s)
	\hat{X}_s	\hat{Y}_s	\hat{Z}_s	δ	$\hat{\varphi}_s$	$\hat{\omega}_s$	\hat{k}_s	$\delta\varphi$	$\delta\omega$	δk			
Classical-TR	-134.7	-1249.0	548.1	6.7	-0.006247	-0.521284	0.025314	0.009993	0.005105	0.024229			
proposed	Case 0	-139.8	-1254.9	550.1	2.5	-0.022659	-0.543598	0.0285702	0.026405	0.027419	0.027485	0.51	1.38
	Case 1	-139.9	-1254.5	550.1	2.1	-0.003449	-0.513311	0.000684	0.007195	0.002868	0.000401	0.50	1.37
	Case 2	-139.8	-1255.6	550.1	3.2	-0.027448	-0.548024	0.031408	0.031194	0.031845	0.030323	0.68	1.26
	Case 3	-139.8	-1255.9	550.1	3.5	-0.027351	-0.549005	0.031018	0.031097	0.032826	0.029933	0.77	1.25
BA +	Case 0	-139.8	-1254.9	550.0	2.5	-0.022871	-0.543477	0.028552	0.026617	0.027297	0.027467	0.59	1.36
Geometry	Case 1	-139.8	-1254.7	550.0	2.3	-0.003409	-0.513298	0.000217	0.007155	0.002882	0.000868	0.59	1.36
	Case 2	-139.9	-1255.7	550.1	3.3	-0.027787	-0.548111	0.031553	0.031533	0.031931	0.030468	0.83	1.25
	Case 3	-139.9	-1256.2	549.9	3.8	-0.027378	-0.549098	0.031007	0.031124	0.032918	0.029922	0.97	1.21
CAHVOR		-134.7	-1249.1	548.1	6.7	-0.006248	-0.521280	0.025317	0.009994	0.005100	0.024232		
ISC		-134.6	-1249	548.2	6.7	-0.006250	-0.521271	0.025351	0.009996	0.005091	0.024266		10.31

The test environment was an Intel(R) Core(TM) i5@2.50 GHz, 2 GB windows system.

Table II. The reconstruction precision of all selected tie points.

Number	σ_x/mm	σ_y/mm	σ_z/mm	σ_{xyz}/mm
1	3.2	9.1	14.0	17.1
2	2.9	9.8	15.3	18.4
3	2.1	6.8	10.5	12.7
4	1.0	7.0	12.1	14.0
5	1.8	8.1	11.4	14.1
6	5.7	14.1	21.7	26.5
7	5.3	6.3	11.5	14.1
8	1.5	9.5	14.8	17.7
9	5.8	9.2	11.4	15.8

$\sqrt{(\bar{X}_s - \hat{X}_s)^2 + (\bar{Y}_s - \hat{Y}_s)^2 + (\bar{Z}_s - \hat{Z}_s)^2}$ denotes the absolute error value of the left Navcam’s position estimation. The symbols $(\hat{X}_s, \hat{Y}_s, \hat{Z}_s, \hat{\varphi}_s, \hat{\omega}_s, \hat{k}_s)$ represent the left Navcam pose estimation based on the above five methods. The symbols $(\bar{X}_s, \bar{Y}_s, \bar{Z}_s, \bar{\varphi}_s, \bar{\omega}_s, \bar{k}_s)$ correspond to the true data of the left Navcam pose, which are $-139.8, 1252.5, -550.7$ mm, $0.003746, -0.516179,$ and 0.001085 rad, respectively. It is noted that the left Navcam pose $(X_s, Y_s, Z_s, \varphi_s, \omega_s, k_s)$ in F_{Rover} can be acquired by using the SCBA method with 59 control points. Here, σ_0 in the SCBA method was 0.18 pixels, while σ_t of the estimated camera pose was 0.3, 0.6, 1.3 mm, 0.001178, 0.000910 and 0.000539 rad. Hence, the camera pose estimation was viewed as the true data, which was denoted by $(\bar{X}_s, \bar{Y}_s, \bar{Z}_s, \bar{\varphi}_s, \bar{\omega}_s, \bar{k}_s)$. The symbols $\delta\varphi = |\hat{\varphi}_s - \bar{\varphi}_s|, \delta\omega = |\hat{\omega}_s - \bar{\omega}_s|, \delta k = |\hat{k}_s - \bar{k}_s|$ denote the absolute error value of the left Navcam’s altitude angles. The results in Table I indicate the following:

1. The Navcam’s position parameters by using the proposed algorithm were much closer to the true data than those of BA + Geometry, CAHVOR, Classical-TR and ISC. In CAHVOR, Classical-TR and ISC, the parameter δ was 6.7 mm. In case 0, case 1, case 2 and case 3 of the proposed algorithm, the parameters δ values were 63.08%, 68.80%, 52.87% and 48.47% of those obtained by CAHVOR, respectively. In cases 0-3 of BA + Geometry, the parameters δ values were 62.69%, 65.54%, 51.38% and 43.48% of those obtained by CAHVOR, respectively. For example, the absolute errors of the left Navcam’s altitude angles $(\delta\varphi, \delta\omega, \delta k)$ in case 1 were 0.007195 0.002868 and 0.000401 rad, respectively, while the corresponding errors in Classical-TR were 0.009993 0.005105 and 0.024229 rad. Thus, the proposed algorithm was able to correct the Navcam pose in case 1.
2. To improve the Navcam pose estimation precision, the tie points’ distribution should be uniform and more concentrated in the lower part of the image’s overlapping areas in the proposed algorithm and BA + Geometry. Here, the $\delta, \delta\varphi, \delta\omega, \delta k$ values in case 1 were higher than those of case 0, case 2 and case 3. Through the thirteenth column of Table I, σ_0 is 0.51 pixels, 0.50 pixels, 0.68 pixels and 0.77 pixels in case 0, case 1, case 2 and case 3 of the proposed algorithm, respectively. In general, the greater the number of tie points and the more concentrated the tie points’ distribution, the higher the Navcam pose estimation precision in case 1, case 2 and case 3. However, this phenomenon did not occur in case 0. In case 0, the forward intersection method was used to calculate the 3D coordinates of the nine tie points with the Navcam stereo pair and Hazcam stereo pair, and the reconstruction precision result is listed in Table II.

In Table II, the symbols $\sigma_x, \sigma_y, \sigma_z$ represent the RMSE of the tie points’ coordinates, where the symbol σ_{xyz} equals $\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$. The σ_{xyz} value of tie point “6” is 26.5 mm, while σ_{xyz} values of

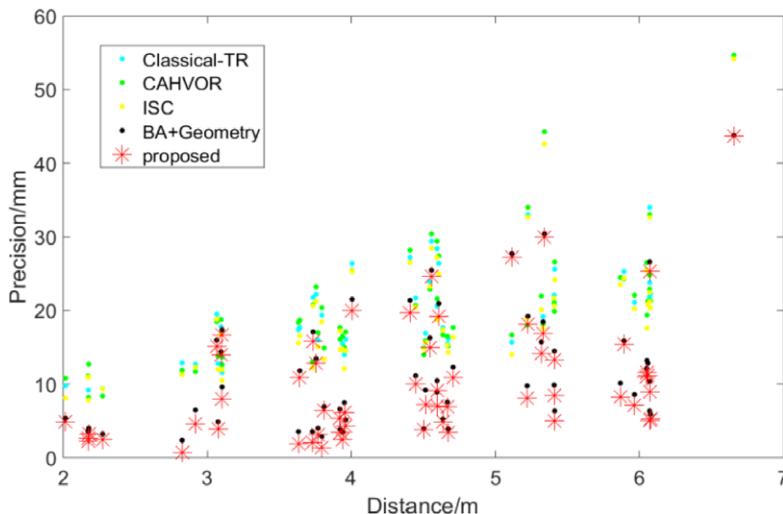


Figure 8. Precision of the 3D coordinates of the checkpoints.

the other eight tie points range from 12.7 to 18.4 mm. Because tie point “6” was selected manually, the corresponding image points have a lower extraction precision. Hence, a low-precision extracted image point will result in a low reconstruction precision. In case 1, all of the tie points except for “6” were used to estimate the left Navcam pose, and the estimated camera pose was much closer to the true data than in case 0, case 2 and case 3. The other tie points (“1, 2, 3, 4, 5, 7, 8, 9”) had similar image extraction precisions, which further indicates that more tie points can promote the visual reconstruction precision.

From a theoretical standpoint, the reconstruction precisions of the feature points determined by using the Navcam stereo pair were slightly higher than those determined by using the Hazcam stereo pair within a certain range. However, due to the measurement errors of the rotation angles of the head and the mast mechanism, the reconstruction precision of the feature points in the Navcam stereo pair decreased. The mean absolute error of the feature points by using the stereo Navcam images and the coordinate system transformation framework decreased to 36.2 mm in the range 0.8–2.2 m. However, the mean absolute error of the feature points determined by using the Hazcam images pair in F_{Rover} was 4.6 mm in the range 0.8–2.2 m. When the left Navcam pose in F_{Rover} was refined with the tie points by using the proposed algorithm, then the corresponding reconstruction precision of those tie points and other check points was increased.

The reconstruction precisions of the check points determined by using the proposed algorithm, BA + Geometry, CAHVOR, Classical-TR and ISC are shown in Fig. 8. The 3D coordinates of the check points in the left Navcam frame F_{Navcam} were first calculated by the forward intersection method with the stereo Navcam pair. Second, those check points were translated to frame F_{Rover} with the corresponding estimated camera pose by the above five methods. Third, the absolute errors of those check points were calculated with the measured 3D coordinates (X, Y, Z).

In Fig. 8, the red star points, cyan points, green points, yellow points and black points represent the absolute errors of the 3D coordinates in the proposed algorithm, BA + Geometry, CAHVOR, Classical-TR and ISC, respectively. Through statistical analysis, the mean absolute errors of the check points in the proposed algorithm, BA + Geometry, CAHVOR, Classical-TR and ISC were 10.4, 11.3, 19.3, 20.4 and 20.2 mm in the range L from 2.0 to 6.1 m, respectively. Thus, the reconstruction precisions of the proposed algorithm were better than those of BA + Geometry, CAHVOR, Classical-TR and ISC. Meanwhile, the precisions of the 3D coordinates of the check points in CAHVOR, Classical-TR and ISC were almost the same.

It was noted that $\delta\varphi$, $\delta\omega$, δk in case 0, case 2 and case 3 of the proposed algorithm were higher than those of CAHVOR, Classical-TR and ISC, while the parameter δ was lower. This shows that the stability of the attitude angles was lower than the position in the least squares solution of the proposed algorithm.

1. The proposed algorithm in case 0 took more time than it took in cases 1–3. To assess the calculating efficiency, the calculation times of cases 0–3 from the process of inputting the matching image points to the calculation of the Navcam pose are listed in the column “Times (ms)” of Table I.

To validate the terrain reconstruction precision in the overlapping area, 3D point cloud processing software (CloudCompare) was used. For more details, see the website <https://www.danielgm.net/cc/>. CloudCompare provides a set of basic tools for manually editing and rendering 3D point clouds and triangular meshes. It also offers various advanced processing algorithms, among which are methods for performing distance computations (cloud-cloud or cloud-mesh nearest neighbor distance). The mean value and the standard deviation (std. dev.) in a Gaussian model are two of the important indices used to judge point cloud registration. The proposed algorithm, Classical-TR, BA + Geometry and the Loop-closed BA method [2] were used to obtain the terrain information. Because the reconstruction precisions of the check points in CAHVOR, Classical-TR and ISC are almost the same, CAHVOR was chosen as the method that would be used for making comparisons. Taking Station A as an example, 8 pairs of stereo Navcam images and one pair of Hazcam images were obtained, and the rotation angles ($\theta_x, \theta_y, \theta_z$) of the head and the mast mechanism at the current station were 0.066 degrees, $-60\sim+60$ degrees and -23.998 degrees, respectively. In the Loop-closed BA, the control points on the model rover cannot be selected in the Navcam images, and a total of 158 matching points were extracted. There were 5 iterations. Using the Loop-closed block BA model based on the TLS solution, σ_0 is 1.5 pixels.

Using a Faro 3D laser scanner, we acquired the point cloud data of the real indoor field. The measured precision was approximately 1.2 mm. Hence, these point cloud data from the Faro 3D laser scanner were viewed as the true data. The terrain products from the above four methods can be compared with the point cloud data obtained by using CloudCompare software. Figure 9 shows the terrain reconstruction results that were obtained by using the above four methods.

As shown in Fig. 9, the mean values in the proposed algorithm, Loop-closed BA, BA + Geometry and Classical-TR methods were 8.8, 15.6, 15.7 and 76.6 mm, respectively, while the std. dev. Values were 6.0, 13.6, 13.1 and 22.8, respectively. The results indicate that the proposed algorithm can improve the reconstruction precision with the stereo Navcam pair and the stereo Hazcam pair.

5.1.2. Visual pose prediction

The movement path of the model rover can be viewed in [1]. At Stations 1 to 6, seven or eight pairs (a backward pair was added) of Navcam stereoisimages and one pair of Hazcam stereoisimages were captured, which validated the robustness and performance of the developed software. At Stations 2 to 6, the measurement positions and attitude angles of the rover in $F_{control}$ were measured, which can be viewed as the true data. At the same time, the visual pose estimation results for the rover were also given using the BA + pseudo method [1]. With the visual pose estimation of the rover as input, the topographic products were obtained through the terrain reconstruction procedure. Note that the rover pose and the topographic products (DEM/DOM) were all in $F_{control}$. In the analysis, given the measured yaw angle γ_y and the plane location (X_O, Y_O) of the rover, the pitch and roll angle of the rover was calculated by the proposed algorithm, which are denoted as ($\hat{\gamma}_p, \hat{\gamma}_r$). The measured angles (γ_p, γ_r) and the calculated angles ($\hat{\gamma}_p, \hat{\gamma}_r$) of the rover at five stations are given in Table III.

In Table III, $\delta\gamma_p = \gamma_p - \hat{\gamma}_p, \delta\gamma_r = \gamma_r - \hat{\gamma}_r$. The estimated pitch and roll angles of the rover at Stations 2 to 6 were obtained using the input data from Stations 1 to 5, respectively. From the ninth and tenth columns of Table III, the average values of $\delta\gamma_p, \delta\gamma_r$ were -0.19 degrees and 0.29 degrees, respectively. The results indicate that the predicted values of γ_p were closer to the true data than γ_r , while both the

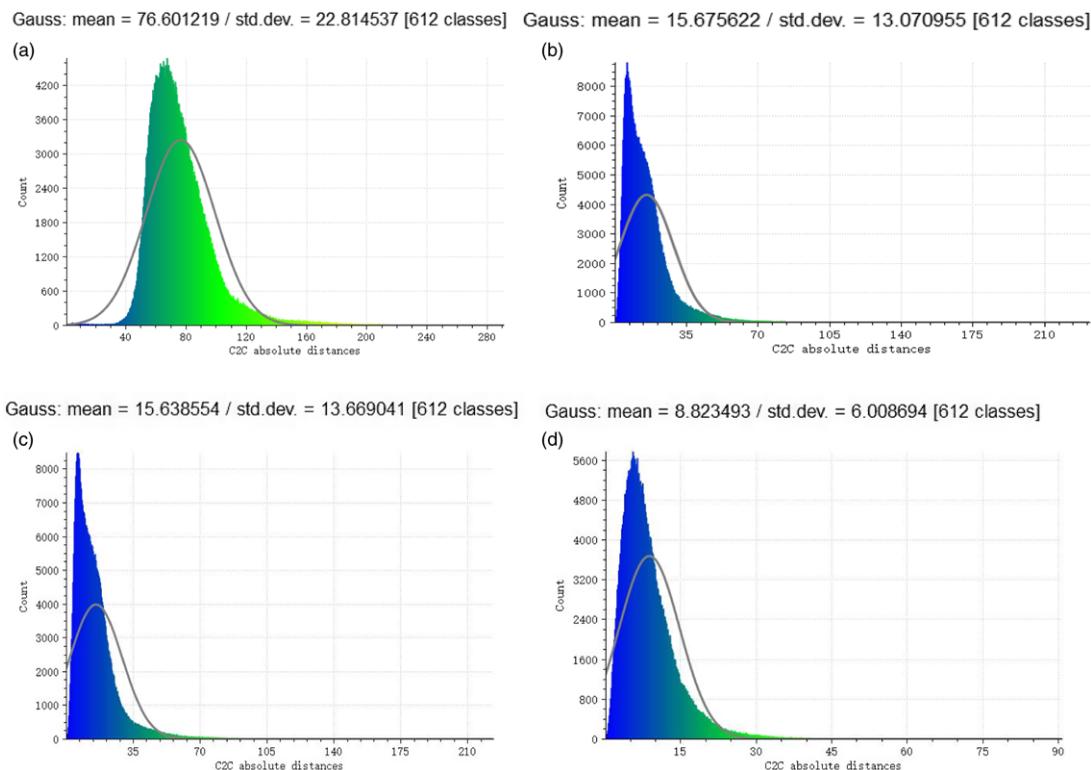


Figure 9. Terrain reconstruction results at Station D. (a): Classical-TR; (b): BA + Geometry; (c): Loop-closed BA; (d): the proposed algorithm.

pitch and roll angle met the set requirements for entering the sleeping mode of the rover. In practice, the angles γ_y , γ_p , γ_r of the Yutu-2 rover must be set as 161 degrees to 169 degrees, -4 degrees to 14 degrees and -2 degrees to 1.5 degrees, respectively. The specific numerical ranges of the rover pose directly originate from the ground thermal testing of Yutu-2.

5.2. Applications of visual pose estimation for the Yutu-2 rover of the CE-4 mission

On 30 March 2019, the service design goal (three months) of Yutu-2 was reached. During the operation cycle, Yutu-2 had gone through the lunar night three times, and the locations for entering the sleeping mode of the rover were Stations S1, LE00210 and LE00309. It is noted that the attitude angles of the rover were measured by the Sun sensor, whose angular precision is 1.0 degrees (3σ). To predict the rover pose at those stations, the Navcam images, Hazcam images and other telemetry data at Stations D, LE00209 and LE00308 were, respectively.

5.2.1. Terrain reconstruction

To validate the terrain reconstruction procedure based on the proposed algorithm, Classical-TR, BA + Geometry and Loop-closed BA, the Navcam/Hazcam images and other telemetry data at Stations D, LE00306 and LE00308 were used. By comparing the mean values and the standard deviations of the above terrain descriptions from the above-mentioned four methods and the terrain description based on the stereo Hazcam images, we were able to establish whether the above algorithms worked and precisely how well.

Table III. Measured angles and calculated angles of the rover at five stations.

Station	X_o/mm	Y_o/mm	$\gamma_y/\text{degrees}$	$\hat{\gamma}_p/\text{degrees}$	$\gamma_p/\text{degrees}$	$\hat{\gamma}_r/\text{degrees}$	$\gamma_r/\text{degrees}$	$\delta\gamma_p/\text{degrees}$	$\delta\gamma_r/\text{degrees}$
2	-1704.1	22640.8	-62.07	0.64	0.39	0.43	0.69	-0.25	0.26
3	174.0	13836	-133.96	0.58	0.47	1.11	1.40	-0.11	0.29
4	-6528.6	4766.3	78.88	0.00	-0.12	-0.76	-0.57	-0.12	0.20
5	-4731.4	13148.9	80.10	1.40	1.27	-1.19	-0.80	-0.13	0.39
6	-5615.1	11339	65.92	13.63	13.32	-0.94	-0.61	-0.32	0.32

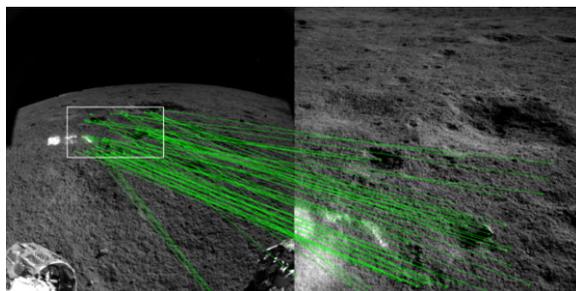


Figure 10. Initial image matching results at Station D: Left Hazcam image and the ninth left Navcam image.

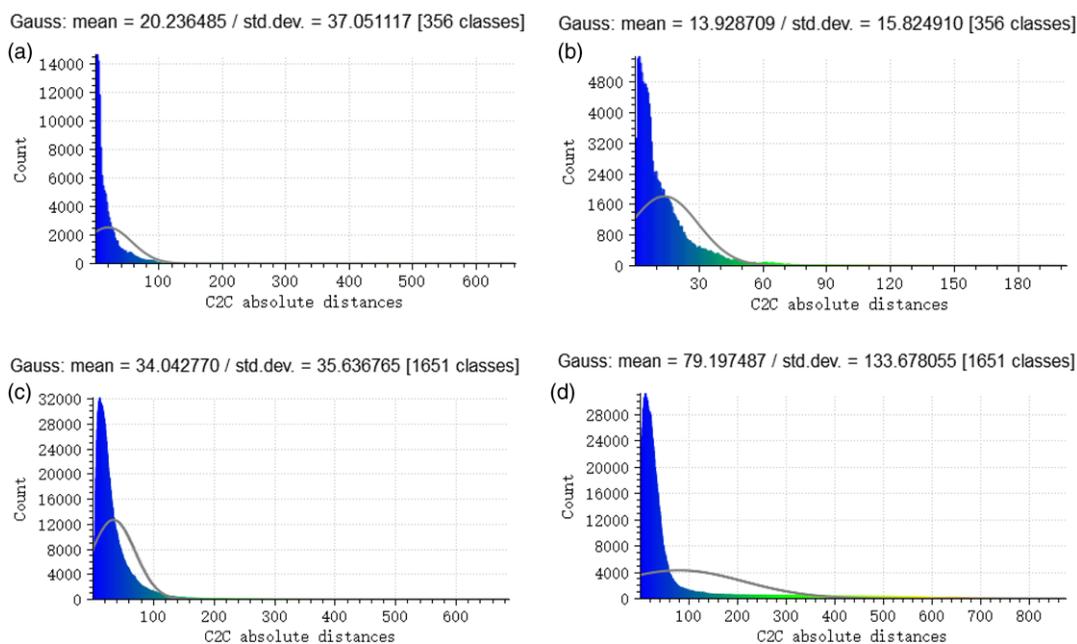


Figure 11. Terrain reconstruction results at Station D. (a): BA + Geometry; (b): proposed; (c): Loop-closed BA; (d): Classical-TR.

Taking Station D as an example, 16 pairs of stereo Navcam images and one pair of Hazcam images were obtained. Figure 10 shows the initial image matching results between the different images by using SIFT.

In Fig. 10, the number of initial matching points was 45, and the distance to the rover ranged from 0.9 to 2.3 m. When the tie point extraction method presented in this paper was used, the number of good tie points was 26. The white rectangle in Fig. 10 represents the overlapping image area between the left Hazcam image and the ninth left Navcam image. It was noted that the rotation angles (θ_e , θ_y , θ_p) in the case of the ninth left Navcam pose were 0.109 degrees, 10.108 degrees and -29.929 degrees, respectively.

To validate the terrain reconstruction precision in the overlapping area, the proposed algorithm, Classical-TR, BA + Geometry and the Loop-closed BA method were used to obtain the terrain description. Figure 11 shows the terrain reconstruction results obtained by using the above-mentioned methods. In “Loop-closed BA” [2], 19 control points on the structure of the model rover were selected, and a total

Table IV. The compared results of the terrain descriptions at three stations based on the different methods.

	D		LE00306		LE00308	
	Mean mm	Std. Dev.	Mean mm	Std. Dev.	Mean mm	Std. Dev.
Classical-TR	79.1	133.7	56.4	42.7	137.5	199.6
Loop-closed BA	34.0	35.6	–	–	72.2	104.0
BA + Geometry	20.2	37.1	34.4	39.4	54.6	102.7
Proposed algorithm	13.9	15.8	27.0	35.1	28.5	36.2

of 235 matching points were extracted, of which 216 had four-degree overlapping image points. The distance between the corners of the solar cells in the left to right direction of the model rover was chosen as the constraint condition of the image points, and there were 21 image points. There were 5 iterations. Using the Loop-closed block BA model based on the TLS solution, σ_0 is 0.96 pixels, and the camera pose precisions of the ninth left Navcam at Station D angles were 0.255 degrees, 0.209 degrees and 0.153 degrees.

In Fig. 11, the mean values in the proposed algorithm, BA + Geometry, Loop-closed BA and Classical-TR methods were 13.9, 20.2, 34.0 and 79.1 mm, respectively, while the standard deviations were 15.8, 37.1, 35.6 and 133.7, respectively.

Comparisons of the results above-mentioned methods at Stations LE00306 and LE00308 were also made, and the results are presented in Table IV.

Loop-closed BA could not be used at Station LE00306 because there was only one pair of Navcam images. The results in Table IV indicate that the proposed algorithm can improve the reconstruction precision with the stereo Navcam pair and the stereo Hazcam pair.

5.2.2. Visual pose prediction

In practice, there are a large number of visual pose prediction results for the Yutu-2 rover at any one station based on the terrain products. After accounting for the moving distance, obstacles and planning the path, a suitable pose from the above-noted prediction results can be selected. However, due to the slippage phenomenon caused by traversing on loose soil terrain or against steep slopes, Yutu-2 may not be able to move to the selected position. Hence, the accurate pose for Yutu-2 at the moved position should be viewed as inputs, which can be obtained by [1], as shown in columns “ X_o ” and “ Y_o ” of Table V. It is noted that Yutu-2’s attitude angles are obtained from the telemetry data (Sun sensor), as shown in columns “ γ_y ”, “ γ_p ” and “ γ_r ” of Table V. Then, the rover’s pitch angle $\hat{\gamma}_p$ and roll angle $\hat{\gamma}_r$ can be recalculated by using the proposed algorithm with the terrain products, as shown in columns “ $\hat{\gamma}_p$ ” and “ $\hat{\gamma}_r$ ” of Table V. The path planning results for Stations S1, LE00210 and LE00309 are shown in Figs. 12, 13 and 14.

From columns “ γ_p ” and “ γ_r ” of Table V, the rover pose at the current position meets the set requirement for entering the sleeping mode. In other words, the Yutu-2 rover must stop working and be ready to enter the sleeping mode at the corresponding given location. From the columns “ $\hat{\gamma}_p$ ” and “ $\hat{\gamma}_r$ ” or “ $\delta\gamma_p$ ” and “ $\delta\gamma_r$ ” of Table V, the predicted visual pose for Yutu-2 can be relatively stable. In Figs. 12, 13 and 14, the green line represents the moving path, the red line represents the planning path with the red crosses at the potential locations for the rover’s sleeping mode and the white cross as the selected location.

It is noted that the sleeping setup procedure of the Yutu-2 rover does not require timeliness of the calculation result because to access the calculating efficiency, the calculation times of the proposed algorithm from the process of inputting the stereo image to the calculation of the Yutu-2 rover pose are approximately 2.5 to 4.3 min. The calculation environment is an Intel(R) Core(TM) i5@2.50 GHz, 2 GB windows system.

Table V. Measured angles and calculated angles of the rover at three stations.

Station	X_0/mm	Y_0/mm	$\gamma_s/\text{degrees}$	$\hat{\gamma}_p/\text{degrees}$	$\gamma_p/\text{degrees}$	$\hat{\gamma}_r/\text{degrees}$	$\gamma_r/\text{degrees}$	$\delta\gamma_p/\text{degrees}$	$\delta\gamma_r/\text{degrees}$
S1	14579.0	-9393.1	162.63	-2.01	-2.38	0.55	-0.12	0.37	0.67
LE00210	19092.3	-63607.5	164.46	-1.09	-1.68	-0.38	-1.67	0.59	1.29
LE00309	29644.7	-96758.9	164.08	0.14	-1.08	-1.00	0.18	1.22	-1.18

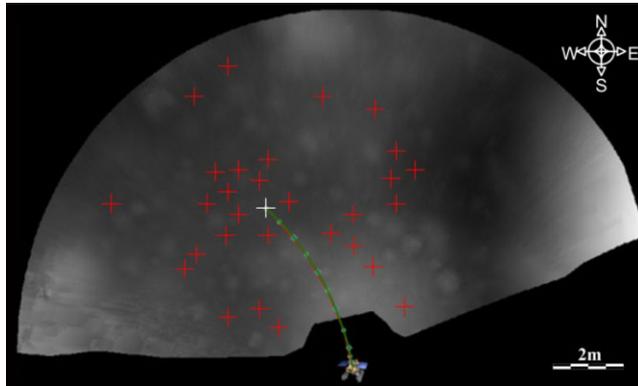


Figure 12. Path planning results for position S1 (background: DEM).

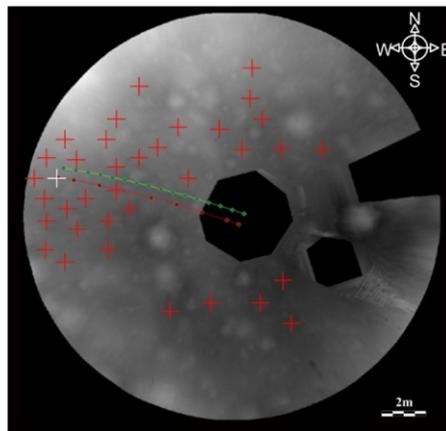


Figure 13. Path planning results for position LE00210 (background: DEM).

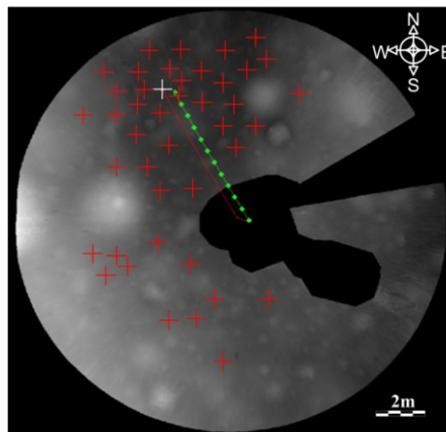


Figure 14. Path planning results for position LE00309 (background: DEM).

6. Conclusions

A visual pose prediction method to set a sleeping mode for the Yutu-2 rover is proposed in this paper. The core of the paper is a multiview block BA model, which can solve the precise terrain reconstruction problem by using multisensor data. In this procedure, the precision loss phenomenon for terrain reconstruction with the Navcam images and the rotation angles of the head and the mast mechanism can be avoided when the hazard cameras system is used. The movement mechanism systems in the Yutu-2 rover belonged to a kind of multijoint rotary platform. The multiview block BA model based on the multisensor images can be considered to be a new solution to the precise terrain reconstruction problem on the multijoint rotary platform. Meanwhile, the exterior orientation parameters of the Navcam/Hazcams can be viewed as pseudo-observations, and a much more complete formula of the measurement errors for the 3D coordinates in a stereo camera system is given in this work, which can be used in further study on the terrain reconstruction problem. On this basis, a pose prediction method for the Yutu-2 rover based on terrain products is given. The experimental results in the simulation field and in the lunar mission demonstrated the following:

- The prediction precision of the rover pose could meet the set requirement for entering the sleeping mode of the Yutu-2 rover. The average difference values of γ_p and γ_r were -0.19 degrees and 0.29 degrees, respectively. The proposed algorithm has been successfully used in CE-4 mission operations.
- The mean absolute errors of the check points in the proposed algorithm, BA + Geometry [1], CAHVOR [17], Classical-TR [4, 6–8] and ISC [19] were 10.4 11.3 mm, 19.3 20.4 and 20.2 mm over the range of L from 2.0 to 6.1 m, respectively. Thus, the proposed algorithm achieved more precise terrain results than BA + Geometry, CAHVOR, Classical-TR and ISC.

The proposed algorithm's assumption is that the front and back wheels are stressed uniformly, which is appropriate in most cases. However, the proposed algorithm will not work well when there is a slippage phenomenon caused by traversing loose soil terrain or against steep slopes (especially when in craters) or in a shadow or overexposed location. In future research, the computing efficiency of the visual pose prediction method should be improved and it should be combined with other tools, such as a soil attribute prediction and classification system and soil mechanics theory. In addition, the estimated Navcam pose parameters based on the proposed algorithm will be affected, or even severely distorted, when the tie points are contaminated by gross errors or outliers. Hence, future studies of the proposed algorithm are also needed to detect outliers under certain weighting conditions [29] and the total least squares solution [30].

China plans to send other rovers to the moon in future missions, such as Chang'E-7 and Chang'E-8. These rovers will also be equipped with a stereo vision system. The proposed algorithm is not only used on the CE-4 Yutu-2 rover but can also be used for setting up sleeping mode for the other lunar rovers.

Financial support. This work was supported in part by the National Natural Science Foundation of China under Grants 42071447 and 41601494, in part by the Key Laboratory of Earth Observation and Geospatial Information Science under Grant 201901, and in part by the Natural Science Foundation Guidance Program of Liaoning Province under Grant 20180550849.

Competing interests declaration. 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

Author contributions. Youqing Ma contributed to the conception of the study; Song Peng and Youqing Ma performed the experiments; Xinchao Xu contributed significantly to the analysis and manuscript preparation; Song Peng, Youqing Ma and Xinchao Xu performed the data analyses and wrote the manuscript; Yang Jia and Shaochuang Liu helped perform the analysis with constructive discussions.

References

- [1] Y. Q. Ma, S. C. Liu, B. Sima, B. Wen, S. Peng and Y. Jia, "A precise visual localisation method for the Chinese Chang'e-4 Yutu-2 rover," *Photogramm. Rec.* **35**(169), 10–39 (2020). doi: [10.1111/PHOR.12309](https://doi.org/10.1111/PHOR.12309).
- [2] Y. Q. Ma, S. Peng, Y. Jia and S. C. Liu, "Prediction of terrain occlusion in Chang'e-4 mission," *Measurement*, **152** 107368, 1–18 (2020). doi: [10.1016/j.measurement.2019.107368](https://doi.org/10.1016/j.measurement.2019.107368).
- [3] Q. Wang and J. Liu, "A Chang'e-4 mission concept and vision of future Chinese lunar exploration activities," *Acta Astronaut.* **127**, 678–683 (2016). doi: [10.1016/j.actaastro.2016.06.024](https://doi.org/10.1016/j.actaastro.2016.06.024).
- [4] Y. X. Wang, W. H. Wan, S. Gou, Z. Q. L. M. Peng, K. C. Di, L. C. Li, T. Yu, J. Wang and X. Cheng, "Vision-based decision support for rover path planning in the Chang'e-4 mission," *Remote Sens.* **12**(624), 1–16 (2020). doi: [10.3390/rs12040624](https://doi.org/10.3390/rs12040624).
- [5] D. Geromichalos, M. Azkarate, E. Tsardoulas, L. Gerdes, L. Petrou and C. Pulgar, "SLAM for autonomous planetary rovers with global localization," *J. Field Robot.* **37**(5), 1–18 (2020). doi: [10.1002/rob.21943](https://doi.org/10.1002/rob.21943).
- [6] R. X. Li, S. W. Squyres, R. E. Arvidson, B. A. Archinal, J. Bell, Y. Cheng and M. Golombek, "Initial results of rover localization and topographic mapping for the 2003 mars exploration rover mission," *Photogramm. Eng. Remote Sens.* **71**(10), 1129–1142 (2005). doi: [10.14358/PERS.71.10.1129](https://doi.org/10.14358/PERS.71.10.1129).
- [7] K. C. Di, F. Xu, J. Wang, S. Agarwal, E. Brodyagina, R. Li and L. Matthies, "Photogrammetric processing of rover imagery of the 2003 mars exploration rover mission," *Isprs J. Photogramm. Remote Sens.* **63**(2), 181–201 (2008). doi: [10.1016/j.isprsjprs.2007.07.007](https://doi.org/10.1016/j.isprsjprs.2007.07.007).
- [8] Z. Q. Liu, K. C. Di, M. Peng, W. H. Wan, B. Liu, L. C. Li and H. Chen, "High precision landing site mapping and rover localization for Chang'e-3 mission," *Sci. China-Phys. Mech. Astron.* **58**(1), 1–11 (2015). doi: [10.1007/s11433-014-5612-0](https://doi.org/10.1007/s11433-014-5612-0).
- [9] M. Y. Qing, L. S. Chuang, W. Bo, Z. Shuo, L. M. Lei and P. Song, "Weighted total least squares for the visual localization of a planetary rover," *Photogramm. Eng. Remote Sens.* **84**(10), 605–618 (2018). doi: [10.14358/PERS.84.10.605](https://doi.org/10.14358/PERS.84.10.605).
- [10] J. Zhang, Y. Xia and G. Shen, "A novel learning-based global path planning algorithm for planetary rovers," *Neurocomputing* **361**(1), 69–76 (2019). doi: [10.1016/j.neucom.2019.05.075](https://doi.org/10.1016/j.neucom.2019.05.075).
- [11] K. Otsu, G. Matheron, S. Ghosh, O. Toupet and M. Ono, "Fast approximate clearance evaluation for rovers with articulated suspension systems," *J. Field Robot.* **37**(5), 768–785 (2020). doi: [10.1002/rob.21892](https://doi.org/10.1002/rob.21892).
- [12] M. Maimone, J. Biesiadecki, E. Tunstel, Y. Cheng and C. Leger, "Surface Navigation and Mobility Intelligence on the Mars Exploration Rovers," **In: Intelligence for Space Robotics** (A. Howard and E. Tunstel, eds.) (2006) pp. 45–69.
- [13] P. Krüsi, P. Furgale, M. Bosse and R. Siegwart, "Driving on point clouds: motion planning, trajectory optimization, and terrain assessment in generic nonplanar environments," *J. Field Robot.* **34**(5), 940–984 (2017). doi: [10.1002/rob.21700](https://doi.org/10.1002/rob.21700).
- [14] T. Ono, A. Kumamoto, H. Nakagawa, Y. Yamaguchi, S. Oshigami, A. Yamaji and H. Oya, "Lunar radar sounder observations of subsurface layers under the nearside maria of the Moon," *Science* **323**(5916), 909–912 (2009). doi: [10.1126/science.1165988](https://doi.org/10.1126/science.1165988).
- [15] H. Riris, X. Sun, J. F. Cavanaugh, L. Ramos-Izquierdo, P. Liiva, G. B. Jackson and D. E. Smith, "The Lunar Orbiter Laser Altimeter (LOLA) on Nasa's Lunar Reconnaissance Orbiter (LRO) Mission," **In: Conference on Lasers and Electro-Optics** (p. CMQ1) (Optical Society of America, San Jose, California, United States, 2008). doi: [10.1109/CLEO.2008.4551271](https://doi.org/10.1109/CLEO.2008.4551271).
- [16] B. Wu, F. Li, H. Hu, Y. Zhao and H. Zhang, "Topographic and geomorphological mapping and analysis of the chang'e-4 landing site on the far side of the moon," *Photogramm. Eng. Remote Sens.* **86**(4), 247–258 (2020). doi: [10.14358/PERS.86.4.247](https://doi.org/10.14358/PERS.86.4.247).
- [17] K. C. Di and R. X. Li, "CAHVOR camera model and its photogrammetric conversion for planetary applications," *J. Geophys. Res.-Planets* **109**(E4), 1191 (2004). doi: [10.1029/2003JE002199](https://doi.org/10.1029/2003JE002199).
- [18] J. Maki, D. Thiessen, A. Pourangi, P. Kobzeff, T. Litwin, L. Scherr, S. Elliott, A. Dingizian and M. Maimone, "The Mars science laboratory engineering cameras," *Space Sci. Rev.* **170**(1-4), 77–93 (2012). doi: [10.1007/s11214-012-9882-4](https://doi.org/10.1007/s11214-012-9882-4).
- [19] S. Zhang, S. Liu, Y. Ma, C. Qi, H. Ma and H. Yang, "Self calibration of the stereo vision system of the chang'e-3 lunar rover based on the bundle block adjustment, ISPRS-J," *Photogramm. Remote Sens.* **128**(5), 287–297 (2017). doi: [10.1016/j.isprsjprs.2017.04.004](https://doi.org/10.1016/j.isprsjprs.2017.04.004).
- [20] T. Ni, W. Li, H. Zhang, H. Yang and Z. Kong, "Pose prediction of autonomous full tracked vehicle based on 3D sensor," *Sensors* **19**(23), 5120 (2019). doi: [10.3390/s19235120](https://doi.org/10.3390/s19235120).
- [21] Y. Ma and Z. Shiller, Pose estimation of vehicles over uneven terrain(March 2019), <https://arxiv.org/abs/1903.02052>.
- [22] J. Jordan and A. Zell, "Real-Time Pose Estimation on Elevation Maps for Wheeled Vehicles," **In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)** (IEEE, Vancouver, BC, Canada, 2017) pp. 1337–1342, doi: [10.1109/IROS.2017.8202311](https://doi.org/10.1109/IROS.2017.8202311).
- [23] P. Papadakis, "Terrain traversability analysis methods for unmanned ground vehicles: A survey," *Eng. Appl. Artif. Intell.* **26**(4), 1373–1385 (2013). doi: [10.1016/j.engappai.2013.01.006](https://doi.org/10.1016/j.engappai.2013.01.006).
- [24] S. Zhang, Y. Jia, S. Peng, B. Wen, Y. Q. Ma, C. Qi and S. C. Liu, "Self-Calibration of the stereo vision system of the Chang'e-4 lunar rover based on the points and lines combined adjustment," *Photogramm. Eng. Remote Sens.* **86**(3), 169–176 (2020). doi: [10.14358/PERS.86.3.169](https://doi.org/10.14358/PERS.86.3.169).
- [25] K. C. Di and R. X. Li, "Topographic Mapping Capability Analysis of Mars Exploration Rover 2003 Mission Imagery," **In: Proceedings of the 5th International Symposium on Mobile Mapping Technology**, MMT, Padua, Italy, 2007).
- [26] D. G. Lowe, "Distinctive image features from Scale-Invariant keypoints," *Int J. Comput. Vis.* **60**(2), 91–110 (2004). doi: [10.1023/b:visi.0000029664.99615.94](https://doi.org/10.1023/b:visi.0000029664.99615.94).
- [27] A. Geiger, M. Roser and R. Urtasun, "Efficient Large-Scale Stereo Matching," **In: Asian Conference on Computer Vision** (Springer, Berlin, Heidelberg, 2010) pp. 25–38, doi: [10.1007/978-3-642-19315-6_3](https://doi.org/10.1007/978-3-642-19315-6_3).
- [28] X. Yuan, "Quality assessment for gps-supported bundle block adjustment based on aerial digital frame imagery," *Photogramm. Rec.* **24**(126), 139–156 (2009). doi: [10.1111/j.1477-9730.2009.00527.x](https://doi.org/10.1111/j.1477-9730.2009.00527.x).

- [29] Y. Q. Ma, S. C. Liu and Q. Li, “An advanced multiple outlier detection algorithm for 3D similarity datum transformation,” *Measurement* **163**(7), 107945 (2020). doi: [10.1016/j.measurement.2020.107945](https://doi.org/10.1016/j.measurement.2020.107945).
- [30] T. Gu, Y. Tu, D. Tang and T. Luo, “A robust moving total least squares fitting method for measurement data,” *IEEE Trans. Instrum. Meas.* **69**(10), 7566–7573 (2020). doi: [10.1109/TIM.2020.2986106](https://doi.org/10.1109/TIM.2020.2986106).