

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/298808511>

Stereo Vision Technologies for China's Lunar Rover Exploration Mission

Article in International Journal of Robotics and Automation · March 2016

DOI: 10.2316/Journal.206.2016.2.206-4440

CITATIONS

3

READS

801

7 authors, including:



Minglei Li

Nanjing University of Aeronautics & Astronautics

19 PUBLICATIONS 225 CITATIONS

[SEE PROFILE](#)



Shaochuang Liu

Chinese Academy of Sciences

34 PUBLICATIONS 154 CITATIONS

[SEE PROFILE](#)



Youqing Ma

Chinese Academy of Sciences

25 PUBLICATIONS 60 CITATIONS

[SEE PROFILE](#)



Hao Ma

5 PUBLICATIONS 24 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Stereo vision [View project](#)



Planetary rover' SLAM [View project](#)

STEREO VISION TECHNOLOGIES FOR CHINA'S LUNAR ROVER EXPLORATION MISSION

Minglei Li,* Zezhou Sun,** Shaochuang Liu,* Youqing Ma,* Hao Ma,* Changming Sun,*** and Yang Jia**

Abstract

In December 2013, China's Chang'e-3 system successfully landed on the moon and started carrying out rover exploration. Stereo vision technologies have played important roles in terrain mapping and rover localization from beginning to end in this mission. This article mainly discusses two stereo vision implementations. First, using stereo images, we propose a novel solution for terrain reconstruction that combines both geometry and appearance information within a joint model. This model could provide accurate terrain maps for the rover navigation and virtual reality simulations of the mission. Then, we present a binocular visual odometry algorithm which leverages an image-network-based bundle adjustment to obtain rover's position and orientation from cross-site images. In addition, both field experiments and onboard results are presented to verify the effectiveness of the proposed algorithms. These technologies have contributed substantially to the Chang'e-3 mission and help to lay a foundation for future autonomous missions.

Key Words

Chang'e-3 system, lunar rover exploration, stereo vision mapping, visual odometry

1. Introduction

On the 14th of December 2013, China's Chang'e-3 system carrying two kinds of probes, the lander and the rover, arrived at the landing site of "Rainbows Bay" on the moon, then the lunar rover Yutu (jade rabbit) started walking on the moon surface. The attempts of this mission include testifying soft landing capability, using lunar rover to realize scientific targets observation and preparing for eventual human exploration. Extraterrestrial rover exploration missions require accurate terrain maps and reliable localization capabilities for robot navigation and scientific

targets determination [1], [2]. The stereo vision system mounted on the mast of the rover is used to satisfy the engineering demands of this exploration.

Vision-based technologies in space exploration have been proved very efficient and successful because of low power, low cost and high accuracy. During the 1980s and the 1990s, Moravec *et al.* [3], [4] developed a series of vision-based algorithms for mobile robot perception, localization and navigation at Carnegie Mellon University. Institutes such as JPL, Stanford University and elsewhere also carry out extensive research on robot rovers autonomous navigation, using various kinds of sensors and onboard or offboard computers [5]–[7]. These works established the basis for vision-aided space rover exploration. In 2003 and 2011, NASA, respectively, launched Mars Exploration Rover (MER) mission and Mars Science Laboratory (MSL) mission on Mars. In those missions, stereo vision mapping and cross-site localization methods were developed to serve for the rovers' navigation and positioning tasks [8]–[10].

While scouting on moon surface, a real, remote and hazard environment, Yutu rover realizes the mapping and localization tasks using the same stereo images captured by its stereo vision system. Because of the limitation of the field of view and the special illumination conditions, a novel mapping algorithm is proposed to generate a joint terrain model. This model combines both geometry and appearance information at different shots to provide adequate terrain information. Simultaneously, accurate localization results are derived from cross-site stereo images by a binocular visual odometry (VO) algorithm which takes use of Affine-SIFT (ASIFT) matching method and an image-network-based bundle adjustment (BA) method.

The organization of the paper is as follows. In Section 2, we will describe the rover stereo vision mapping algorithm and the binocular VO algorithm, respectively, with a quantitative localization error analysis, and then both field tests and practical results on moon are presented in Section 3 with conclusions in Section 4.

2. Algorithms

In this section, we give a concise review for both mapping and VO algorithms. Figure 1(a) shows a picture of the

* Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, China; e-mail: mingleili87@gmail.com, liusc@irs.ac.cn, {mayq, mahao}@radi.ac.cn

** Chinese Academy of Space Technology, China; e-mail: {zezousun, nanshan1919}@sohu.com

*** CSIRO Computational Informatics at Sydney, Australia; e-mail: Changming.Sun@csiro.au

Recommended by Prof. Gian Luca Foresti
(DOI: 10.2316/Journal.206.206.206-4440)

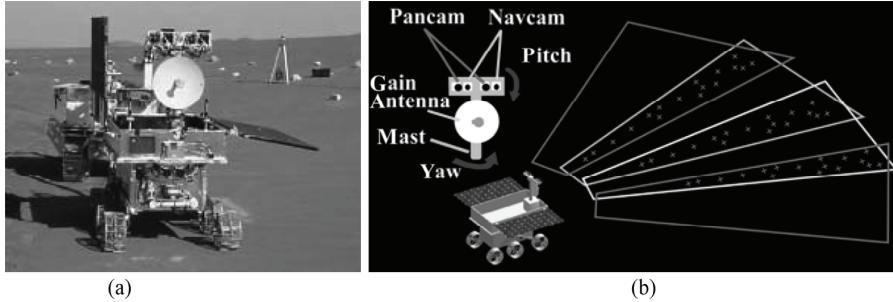


Figure 1. Chang'e-3 Yutu rover: (a) a prototype of Yutu rover and (b) schematic diagram of imaging strategy at one site, where the environment is encoded in some image pairs which make up a curved arc shape. Cross icons denote features detected in the overlapping region between adjacent image pairs.

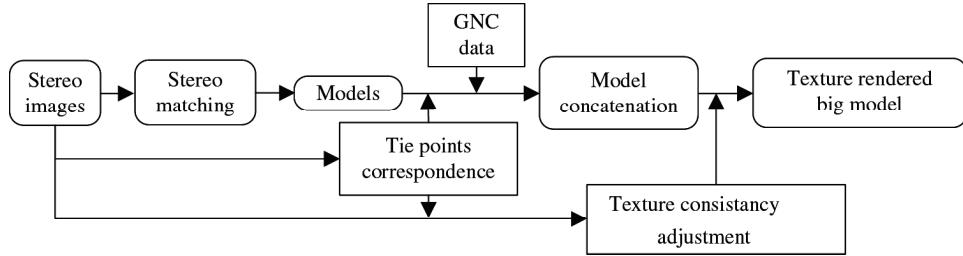


Figure 2. Mapping algorithm flow.

rover prototype in the test base. The rover consists of a payload cab and six active wheels that are able to traverse a wide variety of soil structures. It is equipped with two solar panels that provide the rover with power. The stereo vision system of Yutu rover consists of two navigation cameras (Navcam) and two panoramic cameras (Pancam) on the mast, as well as two hazard avoidance cameras installed on the lower front portion of the rover. Both Navcam and Pancam are mounted on a bar of the rover mast, which stands about 1.5 m above the lunar surface. The bar has 2 degrees of freedom allowing the Navcam to accomplish yaw and pitch motions. The bar cannot have roll motions except for moving along with the rover.

In the Chang'e-3 mission, the control is traded between the rover and ground managers. A scheme of the imaging strategy is shown in Fig. 1(b). While scouting on moon surface, the lunar rover needs to split trajectory into many sites for every 5–7 m advance to obtain the surrounding environments for path planning procedures. Because of the limitation of the field of single image view, the imaging environment is encoded as a curved arc that is divided into segments according to the pan angle of the system.

2.1 Mapping Strategy and Algorithm

The vicinity terrain map is critical for path selection and scientific target detection in unmanned environments. At every site, the Navcam takes 4–7 shot pairs with the mast rotating for several steps along the pan axis. Each pair of images corresponds to a single terrain model. During the mission, the rover uses a Guidance, Navigation and Control System (GNC) which processes the inertial measurement unit (IMU) and wheel odometer data to obtain the initial localization and orientation results. Using the

onboard pose estimates from GNC, each individual model can be displaced within a reasonable precision. However, the mast motion parameters contain small drifts which can cause misalignments between adjacent models. If we concatenate models directly by their measured position, there will be obvious offset in their overlapping regions. Here, we propose a joint modelling method to minimize the corresponding points' reprojection errors and to obtain texture consistency. The algorithm used herein is depicted in Fig. 2.

In the pre-processing stage, considering the special illumination conditions on the moon, we use Wallis filter [11] to handle significant areas of bright and dark tones in the images. The Wallis filter, a locally-adaptive contrast enhancement, can effectively adjust brightness values in local areas so that the local mean and standard deviation match specified target values. This filter can greatly improve the dense matching quality, compared with some typical global contrast enhancement, such as linear and normalized, which cannot simultaneously produce good local contrast at both ends of the brightness range.

The method of extracting depth map from stereo matching is a relatively mature technology. Now, we give a review for the stereo matching processes. Prior to launch, intrinsic and extrinsic calibration parameters of the stereo system have been obtained through calibration processes on ground, depending on known calibration objects with well-defined optical characteristics in the scene. Using the prior calibration parameters, the images are transformed so that the epipolar lines are aligned horizontally. Then the correspondence search is reduced to match along each scanline on the remapped images. Now a semi-global matching method based on Hirschmuller [12] is applied to search the dense correspondences to obtain a depth map [13].

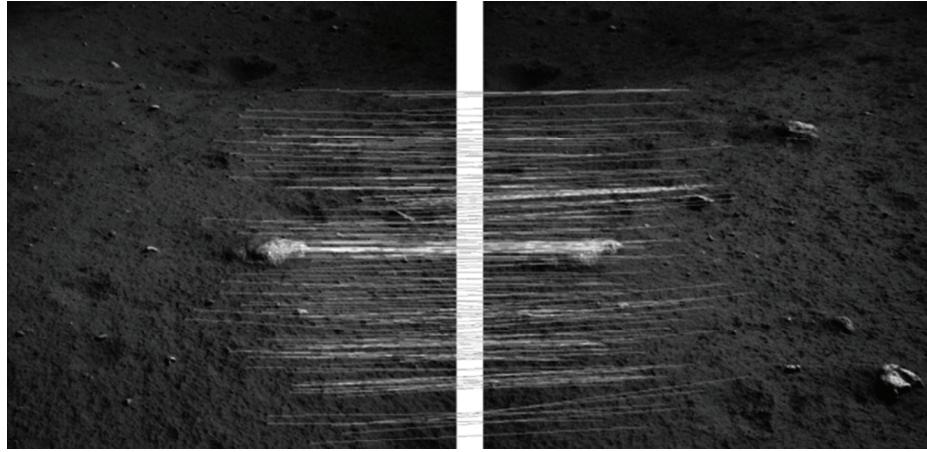


Figure 3. An example of feature correspondences between two adjacent left images.

Using initial poses estimated from GNC system, the terrain model calculated from depth map is registered in the lunar local surface coordinate frame. Once every single model of different shots is generated, a procedure is then executed to concatenate these models into a big model.

Given that the motion of the mast between two successive frames only consisted of a typical yaw rotation of 20°, the overlap of adjacent images is over 40%. To find the corresponding features between two images automatically, Harris corner detector [14] is applied to the images. This yields a set of salient features which are very well suited for matching. The extracted corners are matched automatically between pairs of images using normalized cross-correlation (NCC) score. Then mismatches are removed using a robust matching scheme: random sample consensus (RANSAC). An example of feature correspondences in adjacent images is shown in Fig. 3. Thereafter, in an adjacent model pair, we can find two sets of 3D correspondences with respect to their 2D projections.

The innovation of our mapping method is that both geometry and appearance information at different shots of one site are considered in some iterative optimization solutions. The onboard GNC outputs are used as an initial value for a nonlinear minimization, which finds back the pose values of \mathbf{R} and \mathbf{t} that minimize the sum of all differences between corresponding points in the overlap. In general, assume that n 3D points are projected in m images. With respect to all the 3D points and camera parameters, the optimization of the energy function can be written in the form of:

$$\min \left\| \sum_{i=1}^n \sum_{j=1}^m (\mathbf{P}_j \mathbf{X}_i - \mathbf{u}_{ij}) \right\|^2 \quad (1)$$

where \mathbf{X}_i denotes a vector of 3D point i , and \mathbf{u}_{ij} denotes the position of the i -th point in image j . \mathbf{P}_j is the projection matrix of camera j , composed by intrinsic camera matrix \mathbf{A}_j and a rotation-translation matrix $[\mathbf{R}_j | \mathbf{t}_j]$. Hence, unknown parameters include all parameters describing the projection matrix and the n 3D points. Then,

the minimization problem is solved by using a Levenberg–Marquardt (LM) nonlinear least squares algorithm [15] to simultaneously refine the 3D structure and camera poses to obtain a consistent reconstruction:

$$(\mathbf{J}^T \Sigma^{-1} \mathbf{J} + \mu \mathbf{I}) \delta = \mathbf{J}^T \Sigma^{-1} (\hat{\mathbf{u}}_{ij} - \mathbf{u}_{ij}) \quad (2)$$

where \mathbf{J} is the Jacobian of a projection matrix, and matrix Σ is the covariance matrix for measured vector of 3D points. In our implementation, the covariance matrix is assumed to be identity matrix. Parameter μ is a damping item depending on some user-defined parameters and contributes to the convergence.

The next step is to generate a consistent texture map for the terrain model. As different shots at one site are taken within a short time interval, we can assume that the colour inconsistency between each image pair is only caused by shooting direction and can be corrected by a scale factor which should be close to 1. Hence, to obtain texture consistency, the pixel values of the corresponding points are used to calculate a stretch factor for every image.

For each space tie point \mathbf{X}_i , $g(\mathbf{u}_{ij})$ denotes its pixel grey value in the original image j . Then, the factor λ_j of image j can be found by minimizing a cost function in the form of:

$$\sum_{j=0}^m \sum_{j'=j+1}^m \sum_{i=0}^n |\lambda_j \cdot g(\mathbf{u}_{ij}) - \lambda_{j'} \cdot g(\mathbf{u}_{ij'})| \quad (3)$$

where j and j' denote two different images. Finally, texture rendered model is generated by building Delaunay triangulated irregular network (TIN) and mapping texture on triangle plane units. Both reconstruction accuracy assessment and onboard results are presented in Section 3.1.

2.2 Visual Odometry Algorithm

Besides 3D perception, Navcam is also used for rover localization. VO is the process of determining the position and orientation of an autonomous vehicle by analysing the associated camera images. Although the onboard position estimates, derived solely from the IMU and the wheel

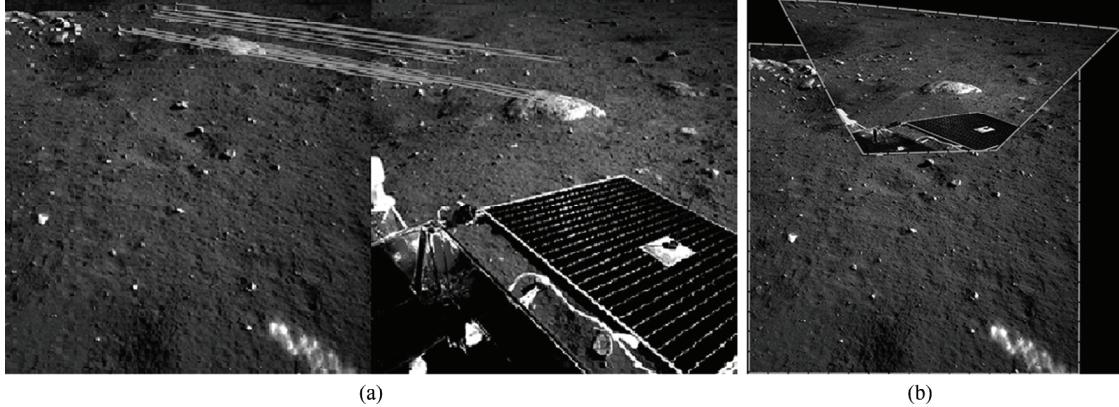


Figure 4. ASIFT matching between two sites: (a) feature correspondences and (b) overlapping region.

odometer data, can achieve a reasonable accuracy in a short term, over time its error increases exponentially and the result will be unacceptable [16]–[19]. There are two types of VO methods: the first one is dense optical flow which has a high computational cost and does not suit for the Chang'e-3 mission, in which images are discontinuously taken at different sites; another method is similar to structure from motion to recover the camera trajectory using sparse features correspondences from two or more images, and this is the method used in the Chang'e-3 mission. In our implementation, a binocular VO algorithm is developed to recover the camera trajectory, depending on geometric constraints formed by corresponding tie points between cross-site images. Compared with traditional VO algorithms, the proposed method takes use of ASIFT [20] matching algorithm to find stable tie point correspondences and then uses an image-network-based BA to obtain the accurate motion estimation. The main steps of the binocular VO algorithm are as follows:

First, reliable corresponding features should be extracted in successive stereo image pairs. Considering the distortion between cross-site images, restrict search regions are extracted to guide feature matching to accelerate the matching speed and to improve accuracy. By prior knowledge of the approximate motion from the GNC system, the image regions in the previous site are calculated by back-projection to the ground coverage of the current site. So the search regions in the current images are specified by extending the back-projected regions.

We leverage an ASIFT method proposed by Morel *et al.* [20] to perform the detection and matching of local interest points. ASIFT has the ability to cope with the apparent deformation of the object image caused by the change of camera positions. After feature matching, 3D coordinates of these correspondences are calculated by stereo triangulation, using initial motion estimates. Then, outliers are detected by comparing their coordinate deviations which should be within a threshold, as their positions calculated from different sites should be the same. An example of ASIFT matching result is shown in Fig. 4.

In two-site case, after triangulation, there will be two sets of 3D correspondences. Let X_i^C and X_i^P represent the i -th observed positions at the current and previous sites.

While the transformation between two positions is noted by Θ , we have:

$$f_i = X_i^C = \Theta \cdot X_i^P + e_i \quad (4)$$

To decrease the error accumulation in motion estimation, an image-network-based BA is proposed to provide high-precision localization results. By linking all images into a network, our BA optimization can minimize the re-projection error of the tie points in all the camera frames. To obtain the linear equation, formula (4) are taken first-order expansion with respect to the motion parameters and 3D observations, the problem is transformed to minimize:

$$E(\mathbf{R}, \mathbf{t}) = \sum_i \left\| \frac{\partial f_i}{\partial \Theta} \cdot \delta_\Theta + \frac{\partial f_i}{\partial X_i^c} \cdot \delta_{X_i^c} - \mathbf{L}_i \right\|^2 \quad (5)$$

$$\mathbf{L}_i = X_i^C - \hat{X}_i^C = X_i^C - \Theta^0 \cdot X_i^P \quad (6)$$

where δ_Θ and $\delta_{X_i^c}$ are the residues of the initial estimations, and \hat{X}_i^C is the calculated position using the initial motion parameters, which can be obtained from the GNC system. Let Θ_j denote the pose of the camera at time j , then we can determine an estimate by solving:

$$\left[\sum_{i,j} \mathbf{J}_{i,j}^T \mathbf{J}_{i,j} \right] [\delta_{\Theta_j} \quad \delta_{X_{i,j}}]^T = \sum_{i,j} \mathbf{J}_{i,j}^T \mathbf{L}_i \quad (7)$$

where $\mathbf{J}_{i,j}$ is the Jacobian of the parameters for the i -th observation at the site j . Every time, the new estimate is calculated by adding residuals for the next iteration until convergence. As we have a very good observation from stereo triangulation, it only takes a few iterations before convergence.

Compared with a closed form least square solution which only takes the quality of the 3D observations as a weight factor, this method fully incorporates observation errors into the estimation, which greatly improves the accuracy of the final motion estimate.

2.3 Localization Error Analysis

Given a set of 3D corresponding points, the covariance matrix Σ_X of scene point observation is defined by the

stereo vision error. Once feature correspondences are established, the coordinates (x, y, z) of the points are computed with the triangulation formula: $z = B \cdot f/d$, $x = B \cdot x_L/d$ and $y = B \cdot y_L/d$, where (x_L, y_L) are the image coordinates of 3D point projection on the left image plane, (x_R, y_R) are the correspondent image coordinates on right image plane, B is the stereo baseline, f is the focal length, and d denotes the disparity, that is, $d = x_L - x_R$. As the error distribution of the disparity of any given pixel can be well approximated by a Gaussian with a standard deviation σ_d , the measurement error of depth can be derived as: $\sigma_z = z^2/Bf \cdot \sigma_d$. Then covariance matrix of the point observation is calculated as [2]: $\Sigma_X = E[\Delta X \cdot \Delta X^T]$.

To analyse the uncertainty of localization estimation, the parameters of normal function (7) are split into motion parameters θ and scene point parameters X . According to error propagation principle [20], we can take partial derivatives of energy function F with respect to motion components of θ to form the gradient vector $g = \frac{\partial F}{\partial \theta}$. Now taking a Taylor series expansion of g around (X, θ) to obtain a first-order approximation:

$$g(X + \Delta X, \theta + \Delta \theta) = g(X, \theta) + \frac{\partial g(X, \theta)}{\partial X} \Delta X + \frac{\partial g(X, \theta)}{\partial \theta} \Delta \theta \quad (8)$$

As $F(X + \Delta X, \theta + \Delta \theta)$ gains the extreme at $(X + \Delta X, \theta + \Delta \theta)$, $g(X + \Delta X, \theta + \Delta \theta) = 0$. Also, (X, θ) is the optimization result of $F(X, \theta)$, thus $g(X, \theta) = 0$. Now from formula (8), we can get:

$$\Delta \theta = -\left(\frac{\partial g(X, \theta)}{\partial \theta}\right)^{-1} \frac{\partial g(X, \theta)}{\partial X} \Delta X = -\left(\frac{\partial g}{\partial \theta}\right)^{-1} \frac{\partial g}{\partial X} \Delta X \quad (9)$$

Table 1

Standard Deviation (SD) Values of Two VO Methods

| SD Error | Δx (m) | Δy (m) | Δz (m) | Relative Error |
|----------|----------------|----------------|----------------|----------------|
| BA | 0.24 | 0.21 | 0.09 | 2.76% |
| No BA | 0.37 | 0.26 | 0.11 | 3.87% |

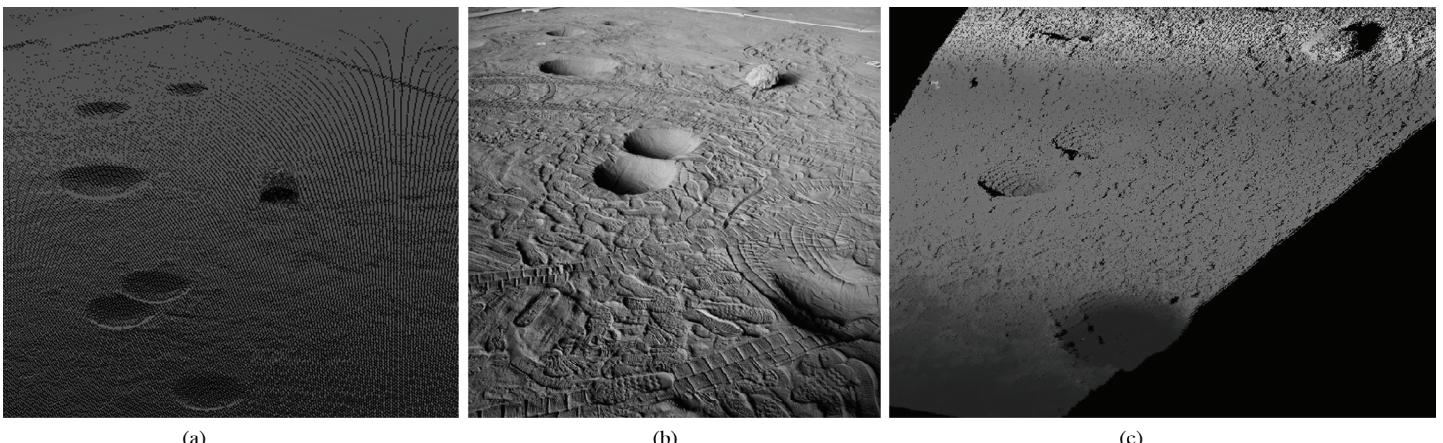


Figure 5. Mapping result in field test: (a) laser scanning data; (b) an original left Navcam image; and (c) point cloud result generated by mapping algorithm.

Finally, the covariance estimate Σ_θ of transformation parameters is obtained in the form of:

$$\Sigma_\theta = E[\Delta \theta \cdot \Delta \theta^T] = \left(\frac{\partial g}{\partial \theta}\right)^{-1} \left(\frac{\partial g}{\partial X}\right)^T \Sigma_X \frac{\partial g}{\partial X} \left(\left(\frac{\partial g}{\partial \theta}\right)^T\right)^{-1} \quad (10)$$

In our simulation, considering the stereo cameras configuration of the rover, f is set to 1,189 pixels and B is 0.27 m. According to statistics, σ_d is set to 1/3 pixel and feature accuracy of ASIFT matching is better than 3 pixel. We use a maximum moving interval of 12 m to test the accuracy of our algorithm. In Table 1, the theoretical localization accuracy of BA is compared with a non-BA method which only uses a closed form least square to estimate motion parameters without refining 3D observations. From the comparison, we can see that the BA method achieves the significant improvement in error values, and the result could meet the requirement of this mission. An accuracy test using real data is described in Section 3.2.

3. Experimental Results

3.1 Mapping Results

Prior to landing, many experiments have been taken on earth with a prototype of Yutu rover in a simulated field with artificial craters and rocks. To analyse the accuracy of the mapping result, precise 3D terrain information of the simulated field was obtained by a laser scanner which has a millimetre level precision. A mapping result from the field experiment is shown in Fig. 5, where (a) is the scanning result, (b) is an image captured by left Navcam, and (c) is a 3D point cloud generated from stereo mapping algorithm. Both the craters and stone have been recovered clearly, including a fresh vehicle track in the closest crater. Statistics of the deviation between visual mapping result and scanning data is depicted in Fig. 6, which indicates that the deviation increases with the distance of the observation, and the deviation along

Z direction is slightly larger than X and Y directions. On the whole, within 10 m observation, the mapping accuracy is better than 0.12 m and satisfies the mission requirement.

Furthermore, a comparison between our joint modelling method and a simple method is shown in Fig. 7, where we can see that the misalignments between adjacent

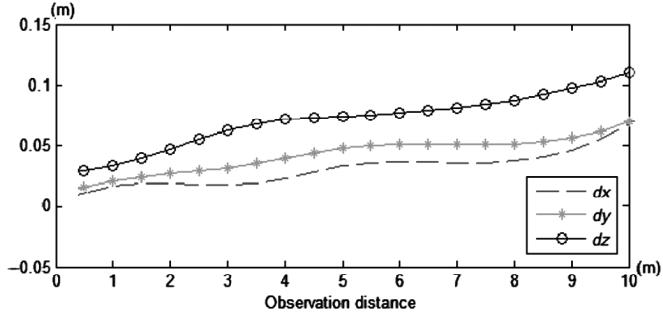


Figure 6. Statistics of mapping deviation in X , Y and Z directions.

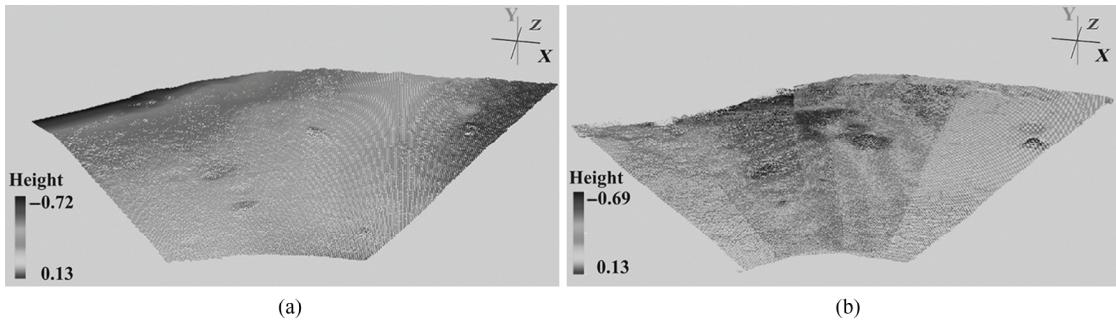


Figure 7. Mapping comparison tests: (a) a mapping result from joint modelling method and (b) a result without joint modelling.

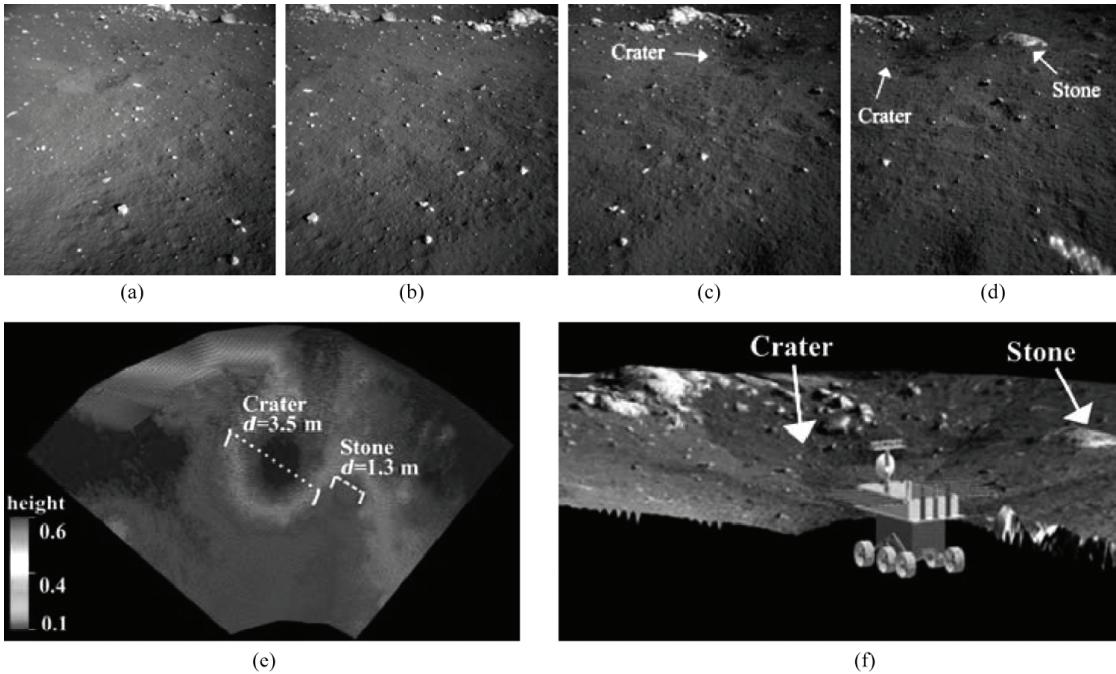


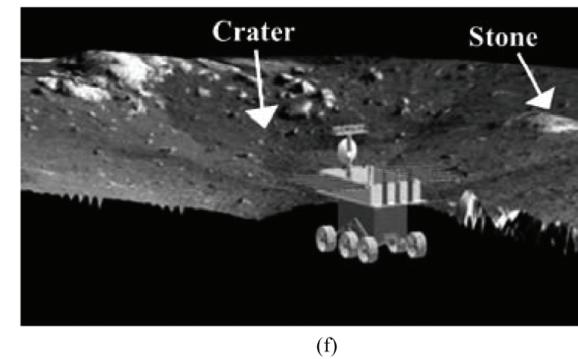
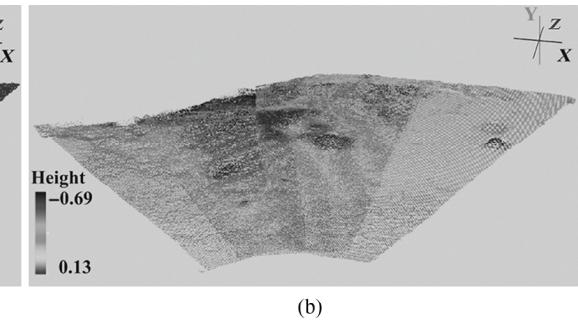
Figure 8. Mapping results on moon: (a)–(d) are original left images; (e) is the reconstructed 3D point cloud; and (f) is texture rendered model.

models have been fixed after joint modelling and our final result has a better performance.

The onboard results of the lunar surface images are shown in Fig. 8. Images were taken at site A by rotating the mast four times. In this test, the maximum recovered region is 17.3 m width and 11.2 m length. As a whole, the vicinity area of site A is relatively flat with many small non-obstacle rocks. The 3D model of the terrain is used for virtual reality simulations of the mission, wherein the model is texture mapped as shown in Fig. 8(f). This reconstructed terrain model has a high-level precision with respect to a near-range scale which is suitable for the rover navigation and obstacle avoidance.

3.2 Visual Odometry Results

A test of the VO system was also performed in the simulated field before launch. While the vehicle was moving on the ground, we took use of a radio based indoor positioning system (IPS) to provide precise localization



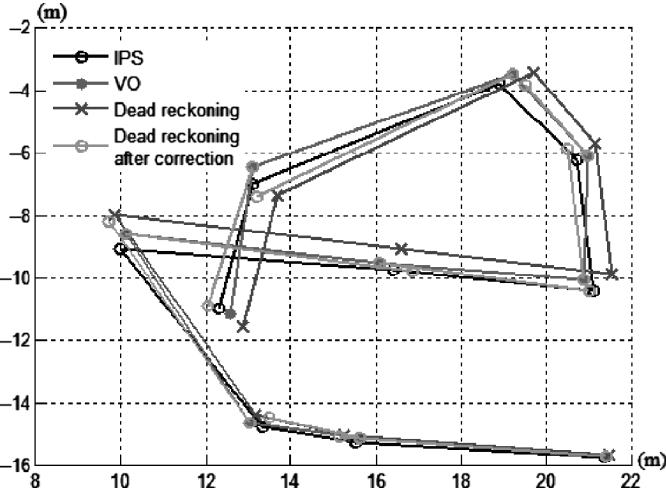


Figure 9. Comparison of different localization outputs.

Table 2
Onboard Slip Rate Detected by Visual Odometry

| Route | Design Length (m) | Diff (m) | Slip Rate (%) |
|-------|-------------------|----------|---------------|
| X-A | 2.658 | 0.1604 | 6.00 |
| A-B | 9.678 | 0.5447 | 5.62 |
| B-C | 10.733 | 0.1478 | 1.37 |
| C-D | 9.655 | 0.7126 | 7.38 |
| D-E | 10.297 | 1.9308 | 18.75 |

outputs for localization accuracy analysis. At the same time, the binocular VO localization procedure was launched to compute the positions of the rover, using cross-site images. The comparison result can be seen in Fig. 9, where the black line is the output from IPS, the red line is the VO localization result, the blue line is dead-reckoning result without any correction, and the green line is the output by making VO correction to the dead-reckoning at each site. If only depending on dead-reckoning, the accumulated localization error can reach up to 14%. After the VO correction procedure, the accuracy could be improved to higher than 5%, which meets the requirement of this mission.

On the first lunar day (a lunar day takes over 27 earth days) after separation from the lander, the lunar rover travelled about 43 m around its landing site with six designated sites (site X, A, B, C, D and E). In routine operations, onboard position estimation is carried out by dead-reckoning with gyros and the wheel encoders, with occasional heading updates by sun sensing with the Pancam. As wheels will slip, this estimation will be prone to error. Therefore, long distance localization is carried out on earth using VO algorithm as described in Section 2.2.

To evaluate the slip rate of the rover, the output of the onboard odometry is compared with the positioning results of VO. Table 2 shows the positioning differences between the onboard GNC output and VO output. Although the

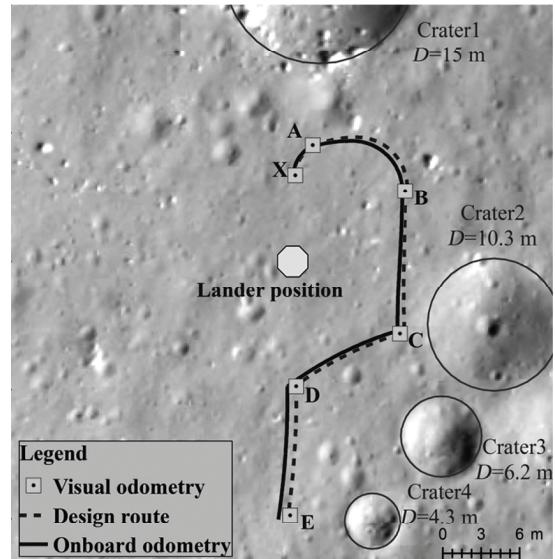


Figure 10. First lunar day localization results of six sites.

landing site is mostly flat with many occasional obstacles, the result indicates that VO reveals the ego-motion performance based on gyros and wheel encoder producing at most 20% error after traversing from one site to another.

The overall localization results of the Yutu rover on first lunar day is shown in Fig. 10. The base map is an orthophoto map generated from descent images captured by the Chang'e-3 lander. As the descent images provide a critical link between orbital images and surface images, it could be used to locate the landing site in a global coordinate frame, as well as to provide high-resolution maps for mission planning. In Fig. 10, the solid red line represents positioning output from the onboard odometry, the dotted line denotes the designed motion routes and the VO detection results are identified by the green icons on the map. The lander position and many big craters have been marked on the map. It confirms that the route planning results depending on the vision based mapping models have successfully avoided potential traverse threatens. At the same time, VO implementation has detected that there exist 4.4% motion estimate error from the onboard odometry after 43 m travel.

4. Conclusion

In this paper, a stereo vision system used in the Chang'e-3 lunar rover exploration mission is described. We have presented a novel and complete mapping method for the reconstruction of a photorealistic terrain model, which combines both geometry and appearance information into a high-performance model. Using the same images, a binocular VO algorithm is developed to recover the relative motions of the rover in its trajectory. By the use of an image-network-based BA, the localization results achieved the significant improvement. This system has been successfully used in the Chang'e-3 program, and the performance analyses show that this system has been competent for the lunar rover's navigation and positioning tasks. Precise mapping and position information is a critical component

for many continuous missions on lunar rover exploration. To improve the level of autonomy ability, our next study emphasis will be vision-based onboard odometry and slip perception for automatic driving in future space exploration mission.

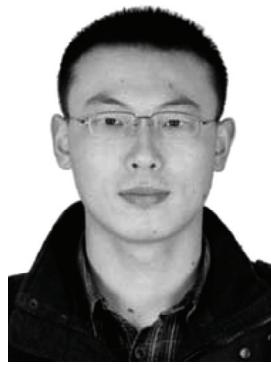
Acknowledgement

This work is supported in part by Chinese Academy of Space Technology.

References

- [1] Z. Sun, H. Zhang, and Y. Jia, Technological advancements and promotion roles of Chang'e-3 lunar probe mission, *Science China Technological Sciences*, 56(11), 2013, 2702–2708.
- [2] W. Wan, Z. Liu, K. Di, B. Wang, and J. Zhou, A cross-site visual localization method for Yutu rover, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-4, 2014, 279–284.
- [3] H. Moravec, The Stanford cart and the CMU Rover, in I.J. Cox and G.T. Wilfong (eds.), *Autonomous Robot Vehicles*, (New York, NY: Springer-Verlag, 1990), 407–419.
- [4] C. Thorpe, M. Hebert, T. Kanade, and S. Shafer, Toward autonomous driving: The CMU Navlab, *IEEE Expert*, 6(4), 1991, 31–52.
- [5] L. Matthies, Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation, *International Journal of Computer Vision*, 8(1), 1992, 71–91.
- [6] S. Goldberg, M. Maimone, and L. Matthies, Stereo vision and rover navigation software for planetary exploration, *Aerospace Conference Proceedings*, 5, 2002, 2025–2036.
- [7] R. Volpe, T. Estlin, S. Laubach, C. Olson, et al., Enhanced Mars rover navigation techniques, *Proc. IEEE Conf. on Robotics and Automation*, San Francisco, CA, 2000, 926–931.
- [8] C. Olson, L. Matthies, J. Wright, R. Li, et al., Visual terrain mapping for Mars exploration, *Computer Vision and Image Understanding*, 105(1), 2007, 73–85.
- [9] L. Matthies, M. Maimone, A. Johnson, Y. Cheng, et al., Computer vision on Mars, *International Journal of Computer Vision*, 75(1), 2007, 67–92.
- [10] J. Wright, F. Hartman, S. Maxwell, B. Cooper, et al., Updates to the rover driving tools for Curiosity, *Proc. 8th IEEE Conf. on System of Systems Engineering*, Maui, HI, 2013, 147–152.
- [11] C. Fraser, E. Baltzavias, and A. Gruenb, Processing of Ikonos imagery for submetre 3D positioning and building extraction, *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(3), 2002, 177–194.
- [12] H. Hirschmuller, Accurate and efficient stereo processing by semi-global matching and mutual information. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, 807–814.
- [13] Y. Xiong and L. Matthies, Error analysis of a real-time stereo system, *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, San Juan, 1997, 1087–1093.
- [14] C. Harris and M. Stephens, A combined corner and edge detector, *Proc. 4th Alvey Vision Conference*, 1988, 147–151.
- [15] M. Lampton, Damping–undamping strategies for the Levenberg–Marquardt nonlinear least-squares method, *Computer in Physics Journal*, 11(1), 1997, 110–115.
- [16] M. Vergauwen, M. Pollefeyns, and L. Van Gool, A stereo-vision system for support of planetary surface exploration, *Machine Vision and Applications*, 14(1), 2003, 5–14.
- [17] P. Ignacio, S. Angel, and V. Ljubo, Robust visual odometry for complex urban environments, *IEEE Intelligent Vehicles Symposium Proc.*, Eindhoven, Netherlands, 2008, 440–445.
- [18] S. Atsushi, T. Yuya, and K. Yoji, Efficient solution to 6DOF localization using unscented Kalman filter for planetary rovers, *Proc. IEEE/RSJ International Conf. on Intelligent Robots and Systems*, St. Louis, MO, 2009, 4154–4159.
- [19] R. Li, K. Di, A. Howard, L. Matthies, et al., Rock modeling and matching for autonomous long-range mars rover localization, *Journal of Field Robotics*, 24(3), 2007, 187–203.
- [20] J. Morel and G. Yu, ASIFT: A new framework for fully affine invariant image comparison, *SIAM Journal on Imaging Sciences*, 2(2), 2009, 438–469.

Biographies



Minglei Li is currently a Ph.D. student in the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. In 2013, he worked at the Chinese Academy of Space Technology, implementing software for lunar rover. He used to be a visiting student at the Visual Computing Center, King Abdullah University of Science and Technology, from Aug. 2014 to Feb. 2015. His research interests include machine vision, remote sensing and 3D reconstruction.



Zezhou Sun is the chief designer of Chang'e-3 system at Chinese Academy of Space Technology. He received his Ph.D. degree in electronic engineering from Nanjing University of Aeronautics and Astronautics. Now he is a professor of Nanjing University of Aeronautics and Astronautics.



Shaochuan Liu is a professor in the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. He earned his Ph.D. degree from the School of Geodesy and Geomatics at Wuhan University in 1996. His researches relate to planetary exploration and remote sensing. He has led a number of projects for lunar rover exploration sponsored by Chinese Academy of Space Technology.



Youqing Ma is a research assistant at Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. He received his Ph.D. degree in photogrammetry and computer vision from Wuhan University in 2014. His research interests include stereo vision, 3D processing and image matching.



Hao Ma received his Bachelor degree from Wuhan University in 2012. Now he is a graduate student in the University of Chinese Academy of Sciences. His research interests include GIS analysis, pattern recognition and machine learning.



Yang Jia is a chief designer of China's Chang'e-3 system at Chinese Academy of Space Technology. He is responsible for the development and optimal design of Yutu rover.



Changming Sun is a principal research scientist at CSIRO Computational Informatics at Sydney, Australia. He received his Ph.D. degree from Imperial College London in 1992. His research interests include stereo vision, digital photogrammetry and image matching.