

VISION BASED OBSTACLE DETECTION USING ROVER STEREO IMAGES

Y. Wang¹, M. Peng¹, K. Di¹, W. Wan^{1,*}, Z. Liu¹, Z. Yue¹, Y. Xing², X. Mao², B. Teng²

¹ State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China - (wangyx716, pengman, dikc, wanwh, liuzq, yuezy)@radi.ac.cn

² Beijing Institute of Control Engineering, Beijing, China – xingyan_bice@163.com, cicely_mao@163.com, tengweb@hotmail.com

Commission III, ICWG III/II

KEY WORDS: Obstacle detection, Stereo camera, Point cloud, Image segmentation, Rover

ABSTRACT:

Vision based obstacle detection using stereo images is an essential way for hazard avoidance and path planning in planetary rover missions. However, due to light condition changes and topographic relief, only partial or sparse three-dimensional points may be derived by image matching and triangulation reconstruction, which is not sufficient for recognizing obstacles. In this paper, we developed a strategy to detect obstacles using rover stereo images by combining both image grayscale information and sparse 3D point information. Experiments were carried out using stereo images captured by navigation cameras mounted on the Yutu rover of Chang'e-3 mission. Moreover, how obstacle localization accuracy affected by the parameters are analysed and discussed.

1. INTRODUCTION

Planetary rover exploration is the most direct way of exploring a planet's surface and subsurface, and because rover safety is a priority, obstacle detection is an essential task in support of rover activities, such as target localization, hazard avoidance, terrain traversability estimation and path planning (Ghosh and Biswas, 2017). Sensors that can detect and localize obstacles include laser (Simmons et al., 1996; Nefian et al., 2017), lidar (Manduchi et al., 2005; Loh et al., 2005), and grayscale, color, depth (Bellone et al., 2013; Zhang et al., 2015), or infrared cameras (Nath et al., 2018). Considering the power consumption and weight limit of payloads for planetary rover missions, optical cameras are usually preferred.

Optical camera based obstacle detection methods can be divided into three types according to the dimensionality of the information used for object detection: (1) two-dimensional (2D) image information acquired by a monocular camera; (2) three-dimensional (3D) information obtained by stereo images; and (3) a combination of both 2D and 3D information.

In past works, obstacle detection using 2D information has mainly been applied to rock detection, and most crater detection work was done using orbiter images. Methods for obstacle detection using 2D images are mainly edge-, region-, or machine learning-based. (a) Edge-based method. In edge-based methods, edges are first detected and then connected to obtain obstacles contours. Castano et al. (2005) used multiple denoising methods and bilateral filtering to remove false textures, then applied Sobel or Canny operators to detect edges and joined them to obtain closed contours. This procedure was implemented on a multi-scale pyramid image to detect several scales of stones. Improving upon this approach (Castano et al., 2007), the researchers first identified and eliminated the sky at the top of the image, applied edge detection and joining algorithms and then used the flood fill technique on the closed

contours to get the extraction results. In another approach, Gulick et al. (2001) built a rock position model based on known solar azimuth information and assuming perfectly spherical rocks and applied the model together with edge fragments gained by Canny to predict the locations of detected rocks. (b) Region-based method. Among the region-based methods, Bajracharya (2002) implemented K-means clustering to segment a 2D grayscale image into regions and then classified the regions by texture and intensity into ground, shadow and hazard regions. The shadows were further assessed based on the sun angle to generate a final hazard map. (c) Machine Learning method. Viola and Jones (2001) established cascaded sample templates and trained them with Adaboost to construct a classifier for rock detection. Using a similar strategy, rock windows were manually collected and labelled as training samples to establish a cascade classifier (Thompson and Castano, 2007). However, these methods could not outline the true boundary of the rocks and provided only a roughly rectangular boundary of the rock for subsequent image processing. As an alternative machine-learning-based method, Thompson and Castano (2007) also investigated a pixel-wise classification algorithm using support vector machine based on the local intensity values of each pixel in the image. However, these results yielded individual points that were mostly on the rocks and not the contours.

In the Mars rovers Spirit, Opportunity and Curiosity and the lunar rover Yutu, stereo cameras have provided critical input for rover traversing tasks from data acquisition to decision support. Using triangulation principle, 3D information can be obtained from stereo images. Matthies et al. (2008) used stereo images to obtain a subpixel-range map from which slopes and rocks could be detected. Lagisetty et al. (2013) matched features extracted from stereo images, reconstructed 3D obstacle information and implemented obstacle avoidance based on the kinematic model of the rover and the assumption that obstacles have regular shapes. In support of Yutu rover

* Corresponding author

operations, Liu et al. (2015) developed algorithms that automatically generate high-resolution digital elevation model (DEM) from the rover's navigation cameras (Navcam) images, and produce obstacle map subsequently considering all factors obstructing the traverse of rover such as slope, aspect, and elevation difference. Some studies have also combined stereo information with data from other sources. Simmons et al. (1996) reconstructed 3D points using stereo images, transformed them into the global coordinate system and then down-sampled to obtain a terrain grid map. The terrain map was then used with hazard detection results from laser data for path planning.

Although 3D information provides range information, it loses local texture and grayscale information. Therefore, the combination of 2D and 3D information would be expected to yield more reliable results. Some previous works have combined 2D image and 3D range information. Gor et al. (2001) used such an approach to detect different sizes of rocks: for small rocks, an edge-based method was used, connecting selected edges that were extracted based on statistical criteria, and for large rocks, a range map was used to identify a plane, and the distances from all points to the plane were clustered by K-means, with some post-processing to identify the large rock region. Huertas et al. (2008) used 3D points reconstructed from stereo images to fit a plane from which slopes could be extracted, and used monocular 2D images to extract rocks based on entropy information, whereby the height of the rocks could be derived from the fitted plane residuals. Fox et al. (2002) first fitted a plane using 3D range data to extract the rocks and then projected them onto 2D images to segment the images; the contours were then fitted into ellipse shapes and passed through several criteria to obtain final rock extraction results. Snorrason et al. (1999) used a histogram-based thresholding method to segment 2D grayscale images, obtained an obstacle mask by morphological dilation and applied a DEM to project the mask map into a top-down view, where morphology and smoothing processes were then executed to complete the obstacle map. Di et al. (2013) developed a rock-detection algorithm based on an object-oriented combination of image intensity and 3D point cloud data; the 2D image was first segmented using a mean-shift algorithm, then the segmented objects isolated from the background were divided into small and large objects; and finally plane fitting and shadow analysis were applied to identify rocks from among all the object candidates.

Under poor imaging conditions, only incomplete or partial 3D point clouds may be obtained, resulting in shadow-like areas that lack textural features, which can cause matching failure and thus cannot be used to restore a DEM based on dense 3D points cloud. Therefore, in this scenario, we propose a method that combines gray-scale image information from 2D images with local 3D points cloud information to identify and label obstacles in the image.

The rest of this paper is structured as follows: Section 2 presents and specifies the proposed method; Experimental result is presented in Section 3 and evaluation and analysis are described in Section 4. Finally, conclusions are given in Section 5.

2. METHODOLOGY

Figure 2 outlines our strategy for detecting obstacles that combined 2D and 3D information. Firstly, 3D points are generated by dense matching and triangulation for the stereo

pair. Meanwhile, the left image is segmented using mean-shift algorithm. With the external and internal parameters of the cameras, the calculated 3D points are back-projected to the image and the projection points on the image are considered as seed points. Based on the segmentation result, after eliminating the background, the seed points are counted in each segmented region to mark candidate regions. Then, several metrics are calculated inside these regions using both 3D points information and image grayscale information. The obstacles are recognized by comparing the calculated values and the pre-set thresholds of the metrics. Finally, morphological processing and hole filling are applied to obtain the final obstacle distribution map.

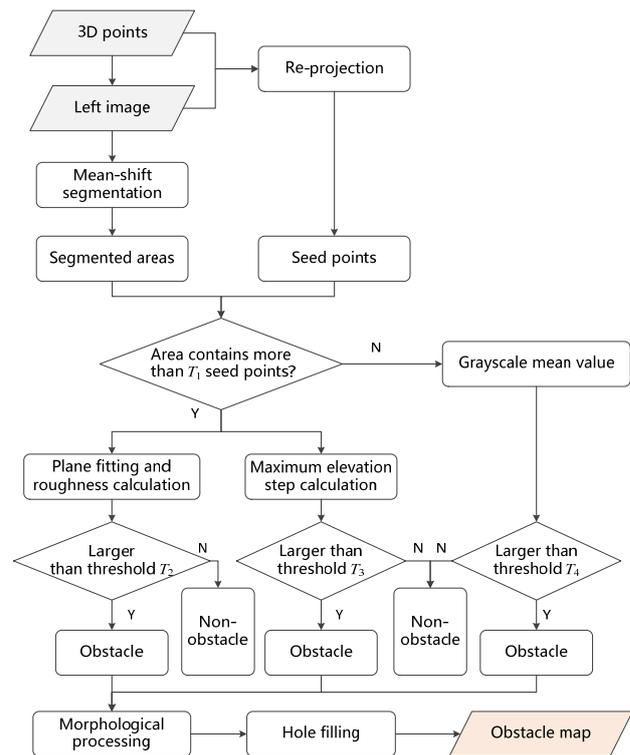


Figure 1. Flowchart of the obstacle detection strategy combining 2D and 3D information

2.1 Generation of 3D information

First, epipolar images based on the intrinsic and extrinsic parameters of the stereo cameras are generated. Based on these epipolar images, we extract feature points using Förstner operator, whose high reliability and localization precision has been previously demonstrated (wan et al., 2017). Extracted feature points are then matched using a correlation coefficient approach whereby pairs of points that have a correlation value larger than a set threshold. To obtain uniformly distributed feature points on the images, we improve the Förstner algorithm by dividing the image into a number of grids that are used in the extraction and matching processes to obtain widely distributed point results. In addition, RANSAC is then performed to eliminate false matchings. The dense matching is then guided by the matched feature point. In order to obtain accurate terrain, least squares matching is then followed to achieve sub-pixel precision. After dense matching, 3D coordinates of the points can be calculated based on triangulation principle.

2.2 Grayscale image segmentation

Meanwhile, the left epipolar image of the pair is segmented using mean-shift algorithm. With reasonable thresholds, the mean-shift algorithm can divide the images into separated regions considering both spatial and grayscale information of the pixels. The main principles and steps of the algorithm are as follows.

First, the line number, column number and grayscale value of the image are converted into a three-dimensional feature vector and then normalize it. For each image pixel, we then calculate the center position of its corresponding feature vector using a kernel function. For a $w \times w$ pixel-sized window, the kernel function g computes the response of the center position as:

$$y_{j+1} = \frac{\sum_{i=1}^n x_i g \left[\left\| \frac{x - x_i}{h} \right\|^2 \right]}{\sum_{i=1}^n g \left[\left\| \frac{x - x_i}{h} \right\|^2 \right]}, j = 1, 2, \dots \quad (1)$$

where n = the number of the pixels in the current window
 x_i = 3D feature vector of each pixel in the current window
 w = bandwidth of space in the kernel function (window size)
 h = bandwidth of color in the kernel function

The kernel function is defined as:

$$g(x) = \begin{cases} 1 & \|x\| \leq 1 \\ 0 & \|x\| > 1 \end{cases} \quad (2)$$

We mark all the convergence points and then connect them into regions. Given a minimum region threshold that defines the minimum number of pixels of a segmented region, we merge regions that are smaller than the threshold into adjacent regions. The final segmentation results are gained after eliminating the background region.

2.3 Seed points

From these segment results, we consider a point P with homogeneous coordinates in the world coordinate system $\tilde{P}_w = (X_w, Y_w, Z_w, 1)^T$; its corresponding 2D image coordinates are $\tilde{P}_u = (x_u, y_u, 1)^T$. According to the imaging model, the process of projecting a 3D point onto a 2D image is expressed as:

$$s\tilde{p}_u = A[R \quad T]\tilde{P}_w, \quad A = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where A = intrinsic parameter matrix of the camera
 (f_x, f_y) = focal length in the x and y directions
 (u_0, v_0) = image coordinates of the principle point
 R = rotation matrix from world to camera coordinate system
 T = translation vector from world to camera coordinate system

The projected 2D points are sub-pixel, and therefore the final coordinates of a seed point are determined by the nearest pixel. After re-projection, we mark the regions that contain no less than T_1 seed points as seed regions.

2.4 Obstacle detection

For the seed region, we calculate the roughness and maximum elevation step values. Roughness reflects the extent of terrain fluctuation of the window area. The elevation step value is the maximum drop height of an certain area. The details are as follows:

(1) Roughness.

Firstly fit a plane using the 3D information in each given $w \times w$ pixel sized window:

$$Ax + By + Cz + D = 0 \quad (4)$$

The normal vector of the plane is $[A, B, C]$, then compute the distances from each point to the fitted plane. The roughness for the central point of the current window is defined as the average value of the distances:

$$V(x, y) = \frac{1}{N} \sum_{i \in W} \left[\frac{Ax_i + By_i + Cz_i + D}{\sqrt{A^2 + B^2 + C^2}} \right] \quad (5)$$

where (x_i, y_i, z_i) = 3D coordinates corresponding to each pixel in the current window
 W = the window area
 N = the number of pixels in the current window

And note that if the corresponding elevation value of a pixel is invalid, then it does not participate in the calculation and counting.

Roughness values shows the dissociation in the distances between the spatial coordinates of each pixel and the fitting plane of the window region. The larger the roughness value is, the greater the terrain fluctuation to some extent.

(2) Elevation step.

In the given window, elevation of each pixel that contains 3D information is compared to obtain the maximum and minimum elevation values. The elevation step value is defined as the subtraction of the minimum from the maximum elevation values.

The elevation step value shows elevation abruptness extent of a certain space, which is an important indicator to ensure the traversal safety of a rover.

We set two thresholds, T_2 and T_3 , for these two metrics, respectively. If either value exceeds the corresponding threshold, the region is considered an obstacle. For the non-seed regions, i.e., the regions containing less than T_1 seed points, the 3D information is not sufficient to make a determination with confidence, and we therefore calculate the grayscale mean value of the region and compare it with threshold T_4 . If the mean value is smaller than the threshold, the region is also considered an obstacle.

Morphological post-processing and hole filling are then executed to obtain closed contour regions and generate the obstacle map.

3. EXPERIMENT

To verify our proposed methods, the binocular images taken by navigation cameras (Navcams) on the Chang'e-3 mission's Yutu Lunar Rover were used as experimental data. The Navcams were mounted at a baseline of 0.27 m and 1.5 m above the ground. The resolution of each Navcam image is 1024×1024 pixels. The focal length of the cameras is 17.7 mm, with a field of view of 46.4° .

At each waypoint along a traverse, the rover usually captured seven pairs of stereo images. To assess performance of our proposed method, we selected pairs that are under bad imaging conditions or surfer occlusion. One of the stereo image pairs taken by the Navcam at Site A are shown in Figure 2. Because of the rover's geometry and ground fluctuations, the bottom left of each image is covered by the rover itself, and some of the shadow areas are lack of texture.

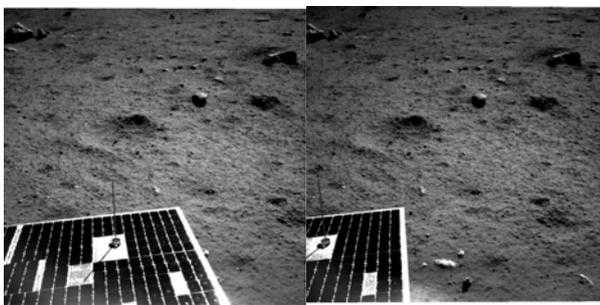


Figure 2. One pair of stereo images captured by the Yutu Lunar Rover's Navcams at Site A

Figure 3 shows the Förstner feature point extraction results (yellow dots) and the matched points after RANSAC (red dots). After dense matching, some of the areas still lacked matching pairs. The 3D restore results are shown in Figure 3, in which the shadow areas and places far away from the camera lack dense restored 3D points. After that, we projected these 3D points back to the left image to obtain seed points.

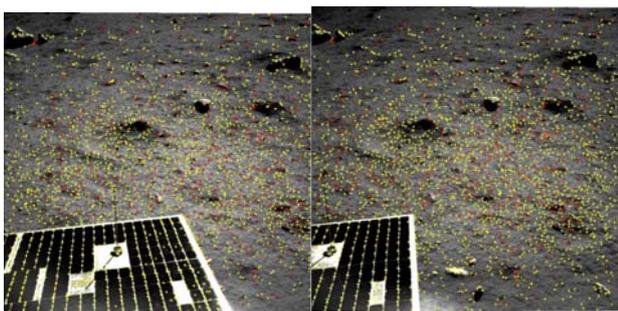


Figure 3. Feature point extraction and matching results

We segmented the left image using the mean-shift algorithm, as shown in Figure 5. The seed points projected on the left image were counted in the segmented regions. If there were more than three points in each region, the two 3D metrics were calculated; otherwise, grayscale statistical information was used to determine whether the region contained obstacle pixels, as described in Section 2. Following this strategy, the final obstacle map was derived from the combined 2D and 3D information, as shown in Figure 6.

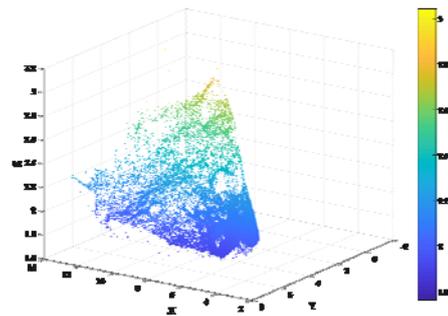


Figure 4. Local 3D information restored from the stereo pair

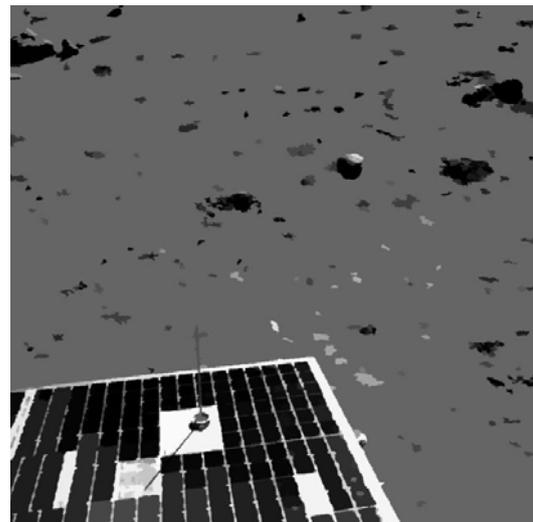


Figure 5. Mean-shift segment result of the left image

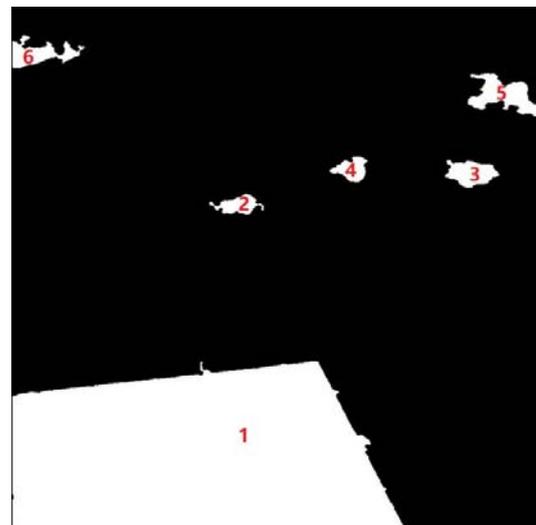


Figure 6. Obstacle map derived by combining 2D and 3D information

4. EVALUATION AND SENSITIVITY ANALYSIS

Here we discuss the effects on obstacle localization of employing different parameters for our proposed method.

The detection results are a list of contour point coordinates, and therefore we employed the center of the contour points as the obstacle localization result. Ground truth values were calculated based on the contours outlined manually referring to the 2D images and the corresponding 3D data. The results obtained using different parameter and threshold settings were assessed based on localization error and its distance deviation. Localization error was analysed by calculating the distance from the center of the detected obstacle to the ground truth. Distance deviation was computed from the deviations of all localization results.

For each parameter or threshold, we fixed the other parameters and thresholds, and investigated reasonable intervals to calculate the center coordinates of the detected obstacles at the different values of the tested parameter.

The three mean-shift parameters determine the center location of the detected areas, and the thresholds only confirm whether these areas are obstacles so that not related to localization accuracy; i.e., the mean-shift parameters segment the image into regions, and the thresholds decide which regions are obstacles. Therefore, we varied only these three mean-shift parameters: window size, color bandwidth and minimum region area. Localization error and distance deviation results of the six obstacles illustrated in Figure 6 are shown in Figures 7 and 8, respectively.

As shown in Figures 7 and 8, the solar panel area denoted as Obstacle 1 has the smallest fluctuation in both criteria because of its unique texture. However, because the rock and crater images suffer from overexposure and shadows, the localization errors in those regions vary when different parameters are used, and the distance deviations fluctuate accordingly. As we can see from Figure 7(a), a relatively smaller window size produced less localization errors. From the diagram, the window size of 5×5 or 7×7 pixels can be seen to be the most appropriate options for this scenario. As shown in Figure 7(b), small color bandwidth values generate large changes in the results, and when the value is in the range from 7 to 9, the results change only a little. The results obtained by changing the minimum region area, illustrated in Figure 8(c), show stability when the threshold is larger than approximately 50 pixels.

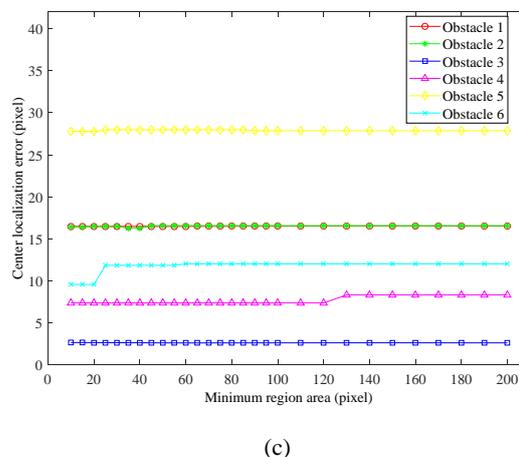
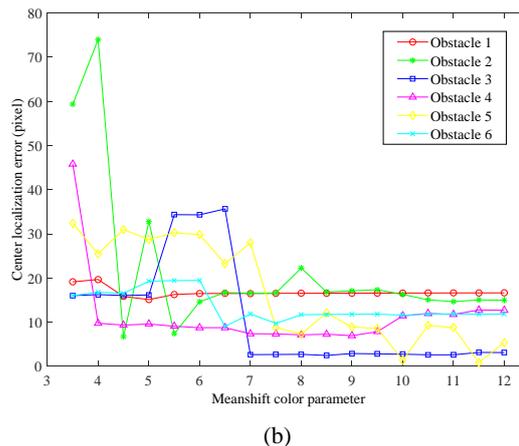
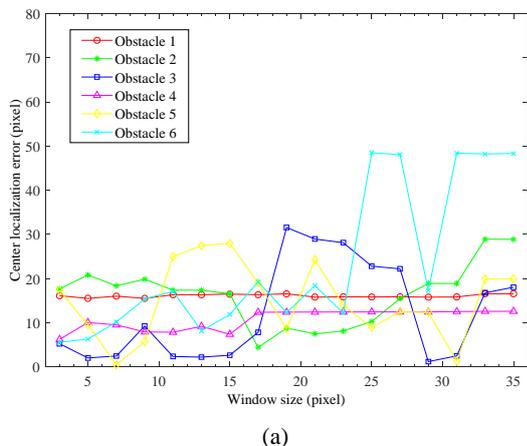
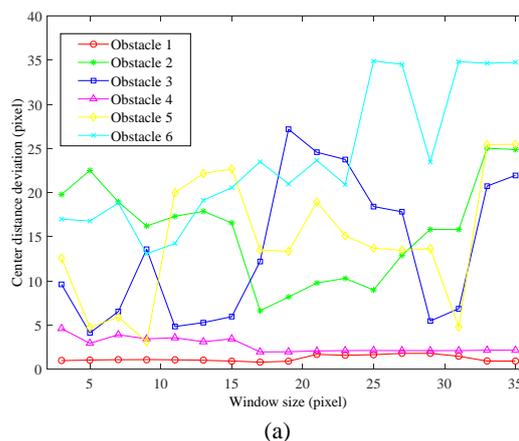


Figure 7. Localization results obtained using different parameter settings: (a) window size; (b) color bandwidth; (c) minimum region area



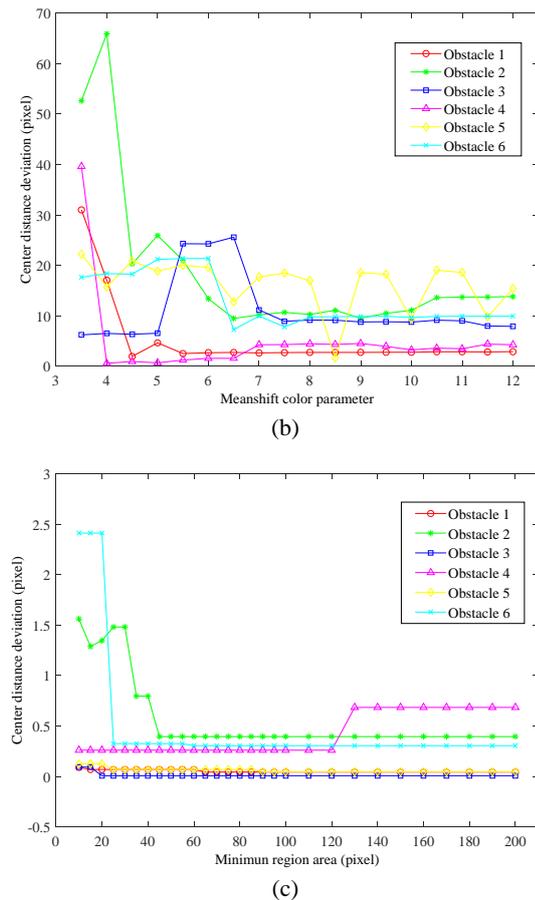


Figure 8. Deviations of the localization results obtained using different parameter settings: (a) window size; (b) color bandwidth; (c) minimum region area

5. CONCLUSIONS

In this paper, we proposed strategies for obstacle detection from planetary images under unfavourable imaging conditions. We combined 2D grayscale information together with 3D point statistical information from a single pair of images to obtain a region-based obstacle map. Stereo images captured by the Navcams mounted on the lunar rover Yutu were used to test the proposed method. The experimental results show that the developed method is effective and reliable. In addition, we analysed how obstacle localization accuracy was affected by the parameters used in this method. The developed method can be applied in future planetary rover missions for obstacle detection, hazard avoidance, path planning, and other rover activities.

ACKNOWLEDGEMENTS

This research is funded by National Natural Science Foundation of China (grant numbers 41701489, 41771488, and 41671458)

REFERENCES

Bajracharya, M., 2002. Single image based hazard detection for a planetary lander. In Proceedings of the 5th Biannual World Automation Congress, Orlando, FL, USA, 585-590.

Bellone, M., Messina, A., Reina, G., 2013. A New Approach for Terrain Analysis in Mobile Robot Applications. In Proceedings of IEEE International Conference on Mechatronics, Vicenza, Italy, 225-230.

Castano, R., Judd, M., Estlin, T., Anderson, R. C., Gaines, D., Castafio, A., Bornstein, B., Stough, T., Wagstaff, K., 2005. Current results from a rover science data analysis system. In Proceedings of IEEE Aerospace Conference, Big Sky, MT, USA, 356-365.

Castano, R., Estlin, T., Gaines, D., Chouinard, C., Bornstein, B., Anderson, R.C., Burl, M., Thompson, D., Castano, A., Judd, M., 2007. Onboard Autonomous Rover Science. In Proceedings of IEEE Aerospace Conference, Big Sky, MT, USA, 1-13.

Di, K., Yue, Z., Liu, Z., 2013. Automated Rock Detection and Shape Analysis from Mars Rover Imagery and 3D Point Cloud Data. *Journal of Earth Science*, 24(1), 125-135.

Fox, J., Castano, R., Anderson, R. C., 2002. Onboard Autonomous Rock Shape Analysis For Mars Rovers. In Proceedings of IEEE Aerospace Conference, Big Sky, MT, USA, 5-2052.

Ghosh, S., Biswas, J., 2017. Joint Perception And Planning For Efficient Obstacle Avoidance Using Stereo Vision. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 1026-1031.

Gor V., Manduchi R., Anderson R. C., Mjolsness E., 2001. Autonomous rock detection for Mars terrain. In Proceedings of AIAA Space Conference and Exposition Albuquerque, NM, USA, AIAA-01-4597.

Gulick, V. C., Morris, R. L., Ruzon, M. A., Roush, T. L., 2001. Autonomous image analyses during the 1999 Marsokhod rover field test. *J. Geophys. Res.* 106, 7745–7763.

Huertas, A., Cheng, Y., Matthies, L. H., 2008. Automatic Hazard Detection for Landers. In Proceedings of 9th International Symposium on Artificial Intelligence, Los Angeles, CA, USA.

Lagisetty, R., Philip, N. K., Padhi, R., Bhat, M. S., 2013. Object Detection and Obstacle Avoidance for Mobile Robot using Stereo Camera. In Proceedings of IEEE International Conference on Control Applications, Hyderabad, India, 605-610.

Loh, J. E., Elkaim, G. H., Curry, R. E., 2005. Roughness Map for Autonomous Rovers. In Proceedings of the American Control Conference, Washington, DC, USA.

Manduchi, R., Castano, A., Talukder, A., Matthies, L., 2005. Obstacle Detection and Terrain Classification for Autonomous Off-Road Navigation. *Autonomous Robots*. 2005, 18, 81–102.

Matthies, L., Huertas, A., Cheng, Y., Johnson, A., 2008. Stereo vision and shadow analysis for landing hazard detection. In Proceedings of IEEE International Conference on Robotics and Automation, Pasadena, CA, USA, 2735-2742.

Nath S. S., Sathish A., Kalaivani C. T., Archana T., 2018. Autonomous Terrain Navigation for Planetary Rovers. *TAGA Journal*, 14, 686-690.

Nefian, A., Wong, U. Y., Dille, M., Bouyssounouse, X., Edwards, L., To, V., Deans, M., Fong, T., 2017. Structured light-based hazard detection for planetary surface navigation. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, BC, USA, 2665-2671.

Simmons, R., Henriksen, L., Chrisman, L., Whelan, G., 2017. Obstacle Avoidance and Safeguarding for a Lunar Rover. In Proceedings of AIAA Forum on Advanced Developments in Space Robotics, Madison, WI, USA.

Snorrason, M., Norris, J., Backes, P., 1999. Vision based obstacle detection and path planning for planetary rovers. In Proceedings of SPIE 3693, Unmanned Ground Vehicle Technology, Orlando, FL, USA.

Thompson, D. R., Castano, R., 2007. Performance Comparison of Rock Detection Algorithms for Autonomous Planetary Geology. In Proceedings of IEEE Aerospace Conference, Big Sky, MT, USA, 1-9.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA.

Wan, W., Peng, M., Xing, Y., Wang, Y., Liu, Z., Di, K., Teng, B., Mao, X., Zhao, Q., Xin, X., 2017. A Performance Comparison of Feature Detectors for Planetary Rover Mapping and Localization. In Proceedings of International Symposium on Planetary Remote Sensing and Mapping, Hong Kong, China, 149-154.

Zhang, H., Tan, Y., Cao, Q., 2015. Terrain Data Real-Time Analysis Based on Point Cloud for Mars Rover. In Proceedings of the Chinese Intelligent Automation Conference, Berlin, Heidelberg, 99-107.