RESEARCH ON KEY TECHNOLOGIES OF LUNAR PATROL PERCEPTION IN SPARSE SCENES

Luyuan WANG[1,2], Jiyang YU[2*], Miaomiao TIAN[2]

Research on key technologies of lunar patrol perception in sparse scenes is proposed, including the fusion of vision and laser of real-time modeling and autonomous navigation during lunar patrols. Object detection and scene segmentation are used to enhance the scene awareness of astronauts during high-speed movement. First, surround view stitching is designed to integrate 360-degree scene information around the lunar rover, which greatly improves the rover's perception ability. Then, the multi-sensor Simultaneous Localization and Mapping (SLAM) technology (vision and laser) is implemented to reconstruct experimental scenes. The global map built before is used to plan the original path, while the safe arrival of the driving task is ensured through the optimization of the local map and dynamic obstacle avoidance. Last but not least, real-time object detection and scene segmentation based on the rover's deep learning algorithm can guide autonomous navigation. All of the various technologies above are applied to the laboratory's prototype of the lunar rover principle.

Keywords: multi-sensor SLAM; autonomous navigation; scene awareness

1. Introduction

The manned lunar rover's main role is to assist astronauts to carry out a large range of lunar surface exploration and use their scientific instruments to help complete the scientific investigation of the lunar surface. To ensure the personal safety of the astronauts, ensure the smooth completion of the patrol work, and assist the efficient development of scientific exploration missions, the lunar rover needs to have an intelligent perception of obstacles, autonomous positioning, and navigation functions, which can improve the perceptual sensitivity of the astronauts. Meanwhile, it also needs to have the safety guarantee functions such as autonomous driving and one-click navigation back.

The existing positioning and navigation schemes for lunar rovers usually use positioning architecture based on information such as astronomical, visual, and radio ranging [1]. After walking autonomously for a certain distance away from the lander, the rover can determine its distance coordinates through two-way

[1] University of Science and Technology of China, Hefei, 230026, China
[2] China Academy of Space Technology, Beijing Institute of Spacecraft System Engineering, Beijing, 100094, China
*Corresponding author: JIYANG YU, yujiyang@spacechina.com

radio communication with the lander, and can also take sequence images with its camera, and determine its position coordinates [2] through feature matching. Due to the camera resolution and field of view, the scheme is only suitable for the rover to explore in a small range around the lander. In addition, because the target location has been selected, most of the rover's navigation process has been preset, with low autonomy and adaptability to different mission scenarios.

When the manned rover moves at a speed of 10 *km/h*, its single exploration mission can reach tens of kilometers. During the mission, ensuring astronauts' safety and improving scientific exploration efficiency are the main goals to develop the intelligent perception and control system for the manned lunar rover. Considering the functional requirements of autonomous exploration and one-click driving for future lunar rovers, real-time positioning, scene interpretation, and autonomous navigation are the optimal choices for manned patrol missions.

In recent years, with the rapid development of visual information processing technology, the positioning results of vision-based Simultaneous Localization and Mapping (SLAM) have been significantly improved [3]. Existing visual SLAM methods mainly include the direct method based on pixel gradient and the indirect method based on feature point matching. Due to the small variation of gray values and single texture features of sandy soil in the lunar surface environment, the neighborhood pixel gradient-based methods cannot extract effective features [6] in the lunar image, while the indirect method based on feature point detection operator has strong stability and robustness to the brightness change and geometric transformation of the image, and is suitable for the lunar surface scene.

Different from the objects with significant features such as trees, flowers, and roads in the ground environment, the lunar surface terrain is unstructured terrain, with sparse features and few effective features. Therefore, it is difficult to obtain accurate image-matching information by using vision as the only input information to the system. Currently, the widely adopted way combines the information of vision and lidar in the SLAM system, which is assisted by an Inertial Measurement Unit (IMU) as the error correction of positioning results to further improve the positioning accuracy. In the existing work, Li *et al*. [7] proposed to use the matching method of 2D laser scanning lines (similar to the contour line) and 3D elevation maps. This method makes full use of the advantage of lidar to collect dense point cloud data, which improves the ability of real-time positioning and mapping for the lunar rover and achieves a good convergence effect. Shang *et al*.[8] realized the incremental high-precision pose optimization method under limited resources by extracting salient features in point clouds and adjusting moderate matching parameters according to the accuracy of different pose estimations.

Deep learning-based scene interpretation methods [9] detect non-flat road conditions such as lunar rocks, lunar craters, and lunar surface slope in the lunar environment, and then segment various types of semantic information in the scene to assist astronauts in road condition discrimination and path selection during the high-speed progress of the manned lunar rover. In the process of real-time positioning and autonomous navigation, scene interpretation can also take the interpretation results as the prior information to optimize the modeling results of the 3D environment. Pang *et al.* [12] proposed C-Moon-Net, a network for the detection of multi-scale lunar craters, which improved the detection accuracy and was suitable for craters of various scales. Li *et al.* [13] proposed a segmentation method for lunar rocks and lunar craters and achieved satisfactory results by using the two-dimensional maximum inter-class variance method based on particle groups.

Based on the above contents, this paper carries out research on key technologies of lunar surface patrol perception under sparse scenes. By integrating the environment perception information with the SLAM system, stereo scene modeling and autonomous navigation are realized. At the same time, object detection and semantic segmentation network were used to enhance the scene perception ability of manned lunar rovers. Specifically, the 360-degree scene information around the lunar rover is first integrated by the surround view stitching technology, which greatly improves the perception ability of the lunar rover and expands the perception range of scene interpretation. Then, the deep neural networks of object detection and semantic segmentation are used to realize real-time detection and scene segmentation of the experimental scene. And the detection and segmentation results can guide real-time obstacle avoidance in the autonomous navigation process of the lunar rover. Finally, multi-sensor SLAM technology of vision sensor (RGB-D camera) and LIDAR is used to construct the global map of the experimental scene, and the data collected by lidar is used to optimize the local map and avoid obstacles dynamically in the process of path planning. In summary, the contributions of this paper are as follows: (1) Surround-view stitching can integrate 360-degree scene information around the lunar rover, which greatly improves the rover's perception ability. (2) The multi-sensor SLAM technology (vision and laser) can effectively improve positioning accuracy, and object detection and scene segmentation can guide autonomous navigation, which is important for safe arrival. (3) The experimental results with the lunar prototype vehicle show that the multi-sensor fusion SLAM system has accurate modeling results, and timely dynamic obstacle avoidance response, which can provide real-time driving path guidance for astronauts.

The remaining paper is structured as follows. Section 2 is to introduce the principle of the algorithm, which describes the implementation process of key technologies in the lunar surface patrol perception task, including visual/laser

SLAM technology, surround view stitching technology, object detection technology, and scene segmentation technology. Section 3 verifies the effectiveness of lunar patrol perception technology on existing experimental scenarios and lunar prototype vehicles and gives the experimental results in autonomous navigation obstacle avoidance tasks, object detection tasks, and scene segmentation tasks. Finally, section 4 concludes the whole paper.

2. Key technologies of lunar patrol perception

This section mainly introduces key technologies involved in lunar patrol perception, including visual/lidar SLAM, surround view stitching, object detection, and scene segmentation.

2.1 Visual/Lidar SLAM

In this paper, we use a visual SLAM scheme to generate a global map and use LIDAR to build a local map and achieve localization, navigation, and dynamic obstacle avoidance. Vision sensors have the advantages of small size, low cost, and strong ability to collect environmental information, which are the main sensors used in SLAM systems. However, limited by the narrow field of the visual sensor (RGB-D camera, binocular camera), the car cannot perceive the scene in all directions with the local map, so we introduce lidar as an auxiliary sensor to obtain a wider detection range and detection angle, which enables dynamic obstacle avoidance.

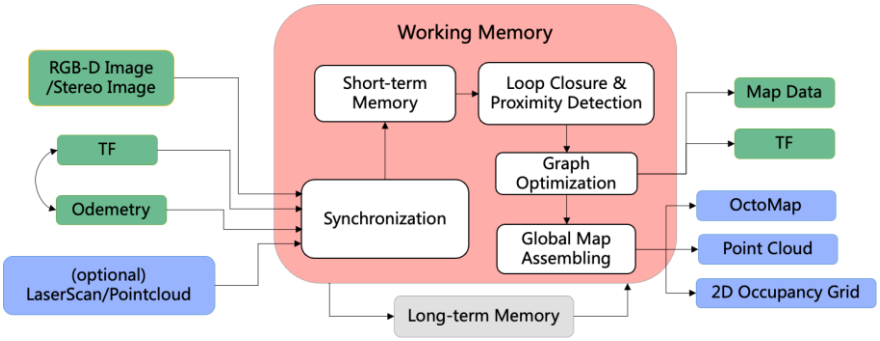

Fig. 1. Visual SLAM based on RGB-D and stereo images

Visual SLAM based on RGB-D and binocular cameras. We use the binocular camera and RGB-D camera for simultaneous appearance-based localization and mapping. The overview of the framework is shown in Fig.1. Visual SLAM technology is mainly composed of front-end (visual odometry), back-end optimization, loop detection, and map construction. The specific descriptions are as follows: 1) Visual odometry. The front-end part based on the vision sensor is mainly to match image features and then complete the pose and

depth map point estimation based on geometric cues. 2) Backend graph optimization. The back-end optimization aims to optimize the output of the visual odometry and use the filtering algorithm or nonlinear optimization method to obtain the optimal pose estimation and global consistency map. 3) Loop closure. Loop closure detection uses the bag-of-words method [14], which is to detect whether the system passes through the same position repeatedly. If it is found that the trajectory of the robot system has a closed loop, the information will be sent to the backend for rectification to eliminate the trajectory drift caused by the cumulative error of continuous estimation. 4) Map generation. After obtaining the back-end optimized pose and estimated map point, the mapping module calculates the 3D coordinates of the map point by combining the robot system's pose and map points' depth and then finishes map construction.

Lidar-based dynamic obstacle avoidance. This paper uses LIDAR for localization and local map generation to achieve dynamic obstacle avoidance. This method uses the Gmapping algorithm based on RBPF filtering [15]. The basic principle of this particle filtering algorithm is that the robot continuously obtains the surrounding environment information through motion and observation, and gradually reduces the uncertainty of its position to obtain accurate positioning results. For localization, we use the AMCL (Adaptive Monte-Carlo Localization) algorithm. The input of the AMCL algorithm is lidar data and odometry data, and the output is the pose of the robot on the map. At the same time, the algorithm realizes the kinematics conversion between the map, odometer, and the base point of the car, as shown in Fig.2.

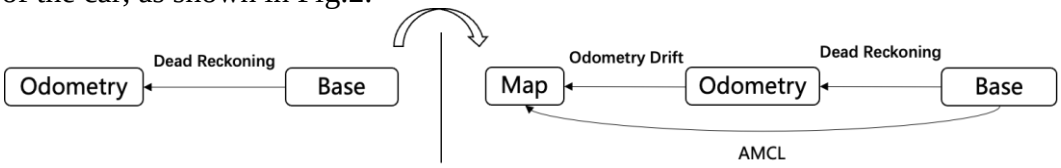

Fig. 2. AMCL algorithm in kinematic transformation

For the generated local map and imported global map, as shown in Fig.3, this paper uses the Movebase motion control module to control the motion of the car. To achieve dynamic obstacle avoidance, this paper adopts the TEB local planner [16] to generate local paths. The initial path is generated by the global path planner, since the environment may be dynamic and may change due to local, incomplete maps or moving obstacles, the TEB local planner takes into account the dynamic constraint correction of the car motion in time Robot trajectory for dynamic obstacle avoidance.

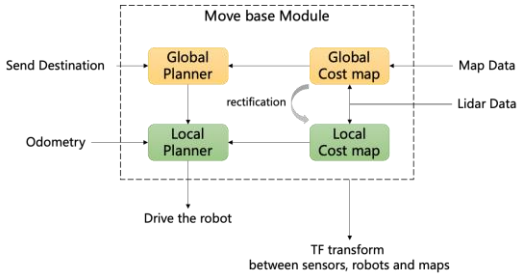

Fig. 3. Movebase control module

2.2 Surround-view stitching

The overall flow of the surround view stitching algorithm is shown in Fig.4. Firstly, robust feature points are extracted from the input image, and the feature points are matched according to the feature descriptor. Then, the positional relationship between adjacent images is obtained according to the matched feature point pairs to perform image registration. Since direct image registration will destroy the consistency of the field of view, the image is first projected on a spherical or cylindrical surface. Finally, the seams of adjacent images are calculated and the fusion of overlapping areas is completed to obtain the final panoramic image.

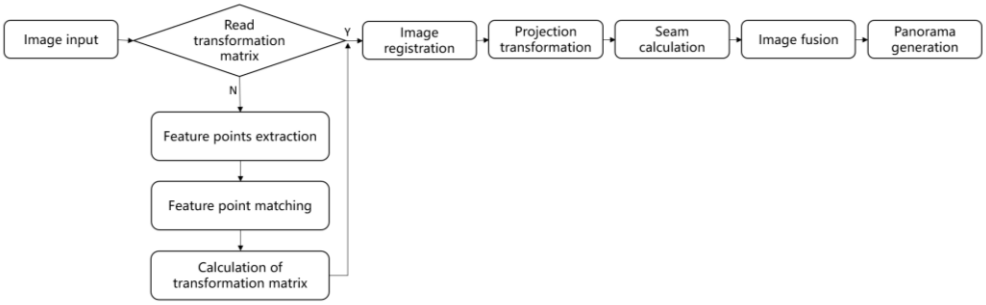

Fig. 4. Surround-view stitching process

Feature points extraction and matching. The feature keypoint extraction algorithm has the following characteristics: a large number of feature points can be extracted in different scenarios; uniqueness is good, to facilitate the matching of feature points; anti-rotation, anti-brightness change, anti-scale scaling, etc. Commonly used feature point extraction algorithms include SIFT [17], SURF [18], and so on. Then, feature descriptors are obtained for the extracted feature points, and matching is performed. Feature matching is obtained by comparing the Euclidean distance of feature descriptors between feature points.

Image registration. Image registration is to calculate the homography matrix of two images according to the matching pairs obtained in the feature matching step, and then fuse multiple images into one image. Considering that the

feature point pair obtained by feature vector matching may have mismatches, the RANSAC [19] algorithm is used to solve it. In addition, considering that the above algorithm solves the position between two images in isolation if the stitching of multiple images is performed directly, errors will accumulate. Therefore, the beam adjustment method is used for joint optimization, and multiple camera parameters are optimized at the same time to obtain a more accurate image location.

Image projection. All images are projected onto a cylindrical or spherical surface through projection transformation to maintain the consistency of the field of view and make the stitched panorama coherent. The choice of the projection plane is related to the way the camera shoots. Generally speaking, spherical projection and cylindrical projection are the most commonly used projection methods. The cylindrical projection is used in this paper, and the schematic diagram can be seen in Fig.5.

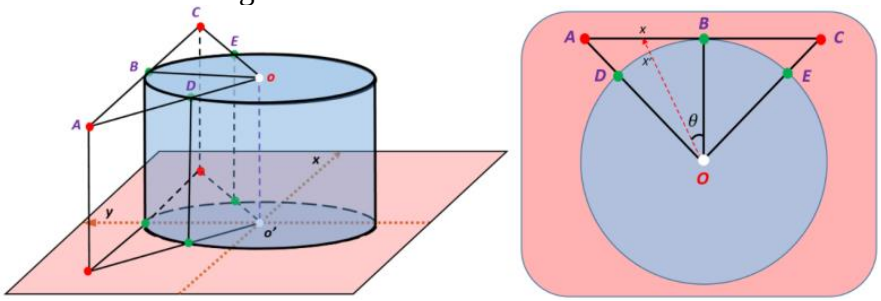

Fig. 5. Schematic diagram of the cylindrical projection

Exposure compensation. The above steps are the basic steps of image stitching, but the results obtained by stitching still have obvious bright and dark changes, some dislocations, and obvious transition marks in the overlapping area between images. To solve these problems, exposure compensation needs to be set so that the overall brightness of different photos is consistent. In this paper, we adopt two commonly used exposure compensation methods: gain compensation [20] and block compensation [21], which can effectively improve the overall brightness of the image.

Seam calculation and image fusion. As shown in Fig.6, the seam refers to the most similar line in the overlapping area of the images. Using a fusion algorithm for several pixels near the seam can effectively remove the misalignment and artifacts between the images. In this paper, the dynamic programming method [22] is used to find seams, and the feathering method [23] is used for image fusion to obtain the final stitching result.

Fig. 6. Seam calculation and image fusion

2.3 Multi-scale feature learning-based object detection

Object detection can detect objects of interest in the scene, that is, obtain the category and position information of the target simultaneously. We adopt a multi-scale feature learning-based object detection algorithm, which is a single-stage object detection model. We have added some new improvement ideas based on the previous object detection algorithm so that its speed and accuracy have been greatly improved. The method mainly consists of three modules: (1) feature extraction module; (2) multi-scale feature fusion enhancement module; (3) object detection head module. The overall technical route is shown in Fig.7, and the detailed introduction of each module is as follows:

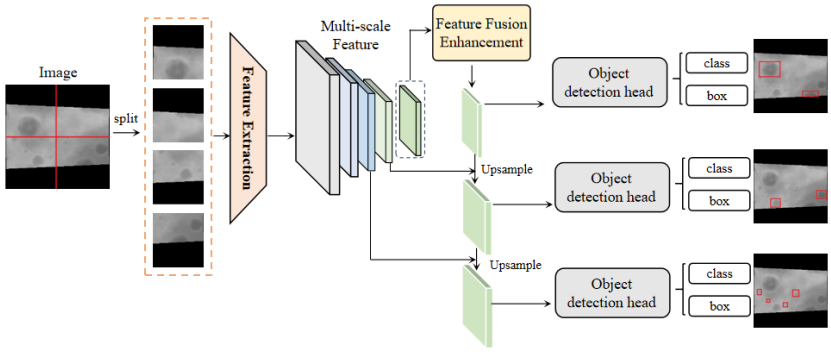

Fig. 7. Object detection based on multi-scale feature

Feature extraction module. We use Darknet53 for feature extraction. In addition, we use a module Focus [11] for initial feature aggregation. To put it simply, the pixels of each channel of an input image are divided into 4 parts and then spliced in the channel dimension. After that, the number of channels is adjusted by a 3×3 convolution. Finally, the size of the feature map is not changed by filling. The structure is shown in Fig.8. The biggest benefit is that it can downsample with minimal information loss.

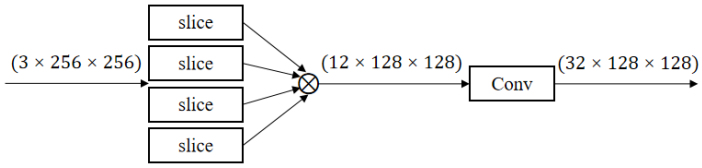

Fig. 8. Focus structure

Multi-scale feature fusion enhancement module. The spatial pyramid structure is derived from SPP-Net [24], which uses kernels of different sizes to implement pooling operations and perform feature fusion. This structure can aggregate the information of multiple receptive fields and enrich the information of the deepest feature map. Relevant experiments show that the spatial pyramid structure has little effect on the running speed of the entire model, but the effect is significantly improved. Whereas, multi-scale information fusion utilizes an improved version of the feature pyramid structure. By performing multi-scale feature fusion, the circulation of feature information at different scales can be accelerated, and the features of small targets can be effectively fused.

Object detection head module. The object detection head module consists of three detection output heads, and the down-sampling ratios are 8, 16, and 32 respectively. The output feature maps sampled at small magnifications detect small objects, and the output feature maps sampled at large magnifications detect large objects. Finally, automatic detection and identification of objects of interest can be achieved.

2.4. Adaptive Prototype for Scene Segmentation

Our proposed framework for scene segmentation based on adaptive prototypes is shown in Fig.9. In this framework, the method first inputs the scene images to the feature extractor to extract corresponding features, and these features are fed into the pixel relation-based encoder to aggregate scene pixels with similar appearances. Then the enhanced features are fed into an adaptive prototype-based decoder, where each prototype identifies one specific scene. And the identified different parts are fused into the more accurate scene segmentation. The specific procedure is as follows:

Pixel Relation-based Encoder. For the scene image, we adopt the feature extractor to extract corresponding features and utilize a convolution layer to reduce the channel dimension of the feature map to a smaller dimension. And then the corresponding query, key, and value can be obtained, which are expressed as Q_n, K_n, V_n

$$Q_n = FW_n^Q, K_n = FW_n^K, V_n = FW_n^V \tag{1}$$

Where $n = 1, 2, ..., N$ denotes the head in the multi-head attention mechanism and W_n is linear projection. Then, we calculate the attention weight S_n,

$$S_n = soft\max(\frac{Q_n K_n^T}{\sqrt{L/8}}) \tag{2}$$

Where is a $\sqrt{L/8}$ scaling factor, we can get the output of the head H_n by adaptively blending values,

$$H_n = S_n V_n \tag{3}$$

We concatenate all single head outputs along the channel dimension and obtain the final output through a projection matrix,

$$H = \text{Concat}(H_1, ..., H_n) W^O \tag{4}$$

Through the self-attention operation, the pixels of the lesion region with similar appearance can be gathered.

Adaptive Prototype-based Decoder. In order to recognize different parts of the scene, different class-specific prototypes need to be learned. First, we utilize a self-attention mechanism [25] to further incorporate context information from other filters to increase their discrepancies. Then we propose a cross-attention mechanism that takes the prototypes $\tilde{P}$ as queries and the enhanced feature maps $\tilde{F}$ as keys and values. Then we have similarities S_n between the enhanced feature map and the enhanced prototype-aware filters. And the prototype-aware activation map M can be calculated as $M = \frac{1}{N}\sum_{n=1}^{N} S_n$. Each scene part activation map denotes the spatial distribution of one specific scene, that is to say, the activation map has high response values at the pixels belonging to the corresponding scene. Finally, the prototype-aware features X can be obtained by adaptively blending and feeding the feed-forward network.

Since not all prototypes have the same importance for scene segmentation, the importance learning mechanism is proposed to learn the importance t^k of different prototypes. For each image obtained from the camera, we can get fused activation map A by $A = \sum_{k=1}^{K} t^k M_k$. Finally, the activation map A is gradually upsampled through the segmentation head to obtain the final segmentation result.

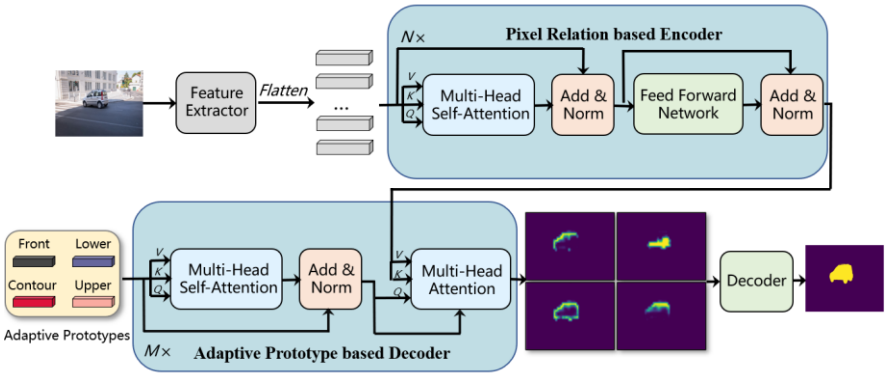

Fig. 9. Scene segmentation framework

3 Experimental validation

3.1 Lunar Rover Simulation Platform

According to the characteristics of the lunar rover performing scientific exploration tasks on the lunar surface, a set of lunar rover prototype vehicles was built for technical verification. Its simulation model is shown in Fig.10. The chassis of the prototype vehicle adopts four-wheel differential drive, and independent motors are used to adjust the torque of the four wheels respectively, so that the prototype vehicle has a good obstacle-surmounting ability in the face of complex terrain, and can be used on slopes, pits, sand, and other complex terrains.

The prototype vehicle is equipped with a variety of sensors to meet the needs of tasks such as autonomous map navigation and scientific detection. Among them, the vision sensors include a binocular vision sensor, RGBD vision sensor (Kinect DK camera), and 360 surround view system, containing a set of binocular vision sensors from two coaxial monocular cameras on the front and another three monocular fisheye cameras on the left, right and rear. In addition to the visual sensor, it is also equipped with lidar and inertial sensors as an aid, of which the lidar is the Ouster series 64-line lidar, whose detection distance can reach 120m and the vertical field of view angle is 45 degrees. The inertia sensor can provide speed, acceleration, and other information in the direction of the XYZ axis. The lunar rover prototype computing platform adopts the Nvidia AGX Xavier industrial module, including an ARM-based 8-core Nvidia Carmel CPU, an Nvidia Volta-based GPU with 512 Nvidia CUDA cores and 64 Tensor cores, reaching 20TOPS hash rate, which better supports end-side calculations. On the Nvidia AGX Xavier platform, the prototype uses Ubuntu18.04 as the operating system with ROS Melodic as the robot system, the data transmission follows the ROS data standards, and the algorithm is written in C++ and Python languages.

Fig. 10. Lunar rover prototype

3.2 Scene reconstruction and path planning

This paper adopts the scheme of lidar and vision camera fusion to achieve the map construction. As shown in Fig.11, we show the results of estimating the depth using a binocular camera and achieving a local 3D reconstruction using the estimated depth. In this paper, the depth estimation method based on stereo images can meet the time and performance requirements of the lunar rover

prototype, and finally realize the visual SLAM map construction, and the 3D map construction results are shown in Figure 12. As shown in Figure 13, we combine the reconstructed global map and local obstacle maps obtained by lidar to do the path planning and dynamic obstacle avoidance. As shown in Figure 14, we have repeated the above experiments in the lunar surface simulation test site, and the experimental results show that the autonomous navigation and positioning algorithm proposed in this paper can also efficiently build maps and independently plan the path of scientific exploration in large-scale sites. As shown in Table 1, our method generally surpasses other methods with prominent margins on average translational error, which proves that the high precision of the predicted odemetry.

Table 1 Average translational error (%) results for the KITTI sequences.

Methods	KITTI sequences										
	00	01	02	03	04	05	06	07	08	09	10
F2F	0.85	2.38	1.01	0.90	**0.35**	0.49	1.25	0.62	1.56	1.24	1.71
F2M	0.68	2.04	0.97	0.77	0.45	0.38	0.57	0.56	1.17	1.38	**0.49**
Fovis	9.09	-	-	1.79	2.22	4.26	6.95	3.65	5.39	14.8	10.6
Viso2	2.38	5.92	4.19	1.94	0.66	1.85	4.60	1.04	2.82	1.68	1.93
Ours	**0.67**	**0.96**	**0.75**	**0.62**	0.50	**0.35**	**0.48**	**0.53**	**1.06**	**0.87**	0.54

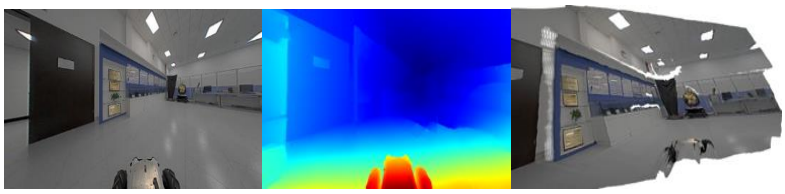

Fig. 11. Stereo Image (left), depth estimation (middle), 3D reconstruction result (right)

Fig. 12. Slam map construction of lunar rover prototype

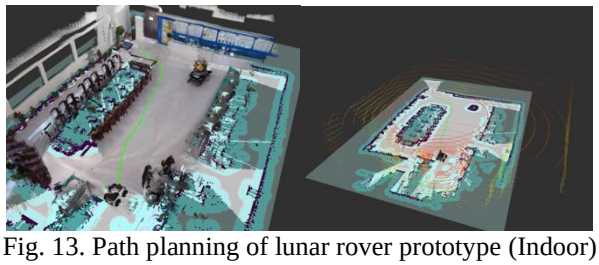

Fig. 13. Path planning of lunar rover prototype (Indoor)

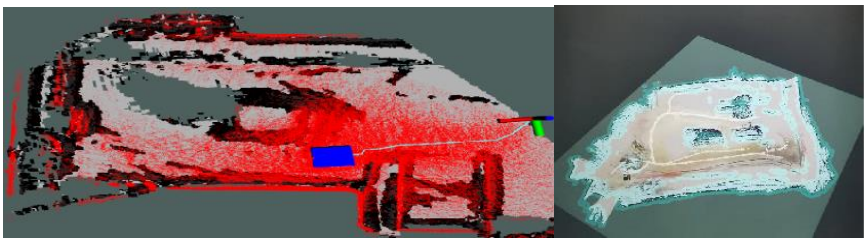

Fig. 14 Path planning of lunar rover prototype (Simulated lunar surface)

3.3 Scene interpretation

Surround-view stitching. As shown in Fig.15, we input images taken from four (front, back, left, and right) views at the same time. The resulting stitching image as shown in Fig.15 can be obtained by the surround view stitching algorithm. From the stitching results, it can be seen that the fusion degree of seams, the integrity of the panorama, and the consistency of brightness can meet the requirements.

Fig. 15 Input images for surround view

Fig. 16 Surround-view stitching result

Object detection. After the image taken by the camera is input, the object detection result shown in Fig.17 can be obtained through the object detection network. It can be seen from the detection results that the designed algorithm can achieve good detection results both for targets with very small imaging sizes and for those with extremely weak texture information. In addition, the proposed algorithm can realize real-time detection on embedded devices. As shown in Table 1, our methods significantly outperform other object detection algorithms, with surprising findings presented on the real-time performance.

Table 2

Comparison of Object detection performance.

Networks	VOC 2007	Picasso		People-Art
	AP	AP	Best F1	AP
R-CNN	54.2	10.4	0.226	26
DPM	43.2	37.8	0.458	32
Poselets	36.5	17.8	0.271	-
D&T	-	1.9	0.051	-
Ours	**59.2**	**53.3**	**0.590**	**45**

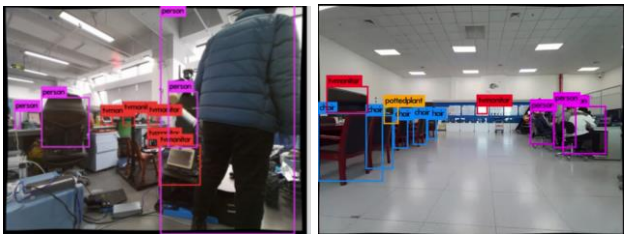

Fig. 17 Object detection result

Scene segmentation. Input the image taken by the camera, and the segmentation result shown in Fig.18 can be obtained through the scene segmentation network. It can be seen from the segmentation results that the integrity of large-area object segmentation, the local details of small-area object segmentation, and the continuity of regional boundary segmentation can all meet the requirements.

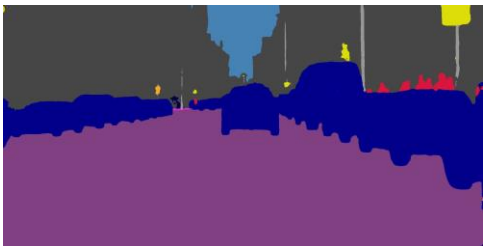

Fig. 18 Scene segmentation result

Table 3

Segmentation performance comparison on SceneParse150

Networks	Pixel Acc.	Mean IoU
FCN-8s	71.32%	0.2939
SegNet	71.00%	0.2164
DilatedVGG	73.55%	0.3231
DilatedResNet-34	76.47%	0.3277
DilatedResNet-50	76.40%	0.3385
Cascade-SegNet	71.83%	0.2751
Cascade-DilatedVGG	74.52%	0.3490
DilatedResNet-101 + PPM	80.91%	0.4253
Ours	**81.01%**	**0.4266**

As shown in Table 2, the proposed method outperforms previous methods by a large margin on the pixel accuracy and mean IoU, which provides important support for semantic-level lunar surface environment perception.

4. Conclusions

In this paper, a compact point cloud model of a 3D scene is established by a SLAM system that integrates visual images and lidar data, and the autonomous navigation and dynamic obstacle avoidance functions of the manned lunar rover are realized on this basis. The application of surround view stitching, lunar surface object detection, and semantic segmentation technology has greatly improved the perception ability of astronauts in the process of lunar rover driving, which plays an important role in ensuring the safety of astronauts and assisting astronauts in carrying out lunar operations. The experimental results with the lunar rover prototype show that the multi-sensor fusion system has accurate modeling results for the experimental scene, effective autonomous path planning, and timely dynamic obstacle avoidance response, which can provide real-time driving path guidance for astronauts. In addition, the algorithms in scene interpretation are all lightweight, which are easy to be directly transplanted into embedded devices. Under the premise of satisfying the astronaut experience, the proposed method has significant advantages in the lunar surface application scenarios with harsh environmental conditions and limited hardware resources.

REFERENCES

[1]. NING X L, CAI H W, WU W R, et al. INS/CNS integrated navigation method for lunar rover. Systems Engineering and Electronics, 2011, 33 (8): 1837 – 1844.

[2]. Liu Chuankai, Wang Baofeng, Wang Jia, et al. Integrated INS and vision-based orientation determination and positioning of CE-3 lunar rover. Journal of Spacecraft TT&C Technology, 2014, 33(3):250-257.

[3]. Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE transactions on robotics, 2015, 31(5): 1147-1163.

[4]. Khan M U, Zaidi S A A, Ishtiaq A, et al. A comparative survey of lidar-slam and lidar based sensor technologies. 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC). IEEE, 2021: 1-8.

[5]. Roy P, Chowdhury C. A survey of machine learning techniques for indoor localization and navigation systems. Journal of Intelligent & Robotic Systems, 2021, 101(3): 1-34.

[6]. A Survey of Simultaneous Localization and Mapping on Unstructured Lunar Complex Environment. Journal of Zhengzhou University (Engineering Science), 2018, 39(03): 45-50.

[7]. Li Xiuzhi, Ju Hehua, Wang Huan. A fast lunar rover simultaneous localization and map-building method in unstructured environment. Journal of Astronautics, 2010, 31(9): 2145-2149.

[8]. SHANG Tianxiang, WANG Jingchuan, DONG Lingfeng, CHEN Weidong. 3D lidar SLAM technology in lunar environment. ACTA AERONAUTICAET ASTRONAUTICA SINICA, 2021, 42(1): 524166-524166.

[9]. Zhao, Zhong-Qin, et al. Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 2019, 30(11): 3212-3232.
[10]. Hao, Shijie, Yuan Zhou, et al. A brief survey on semantic segmentation with deep learning. Neurocomputing, 2020: 302-321.
[11]. Zhao Y, Shi Y, Wang Z. The Improved YOLOV5 Algorithm and Its Application in Small Target Detection. International Conference on Intelligent Robotics and Applications. Springer, Cham, 2022: 679-688.
[12]. PANG Chengcheng, ZHANG Huachun，ZHANG Yanyan. An End-to-End Multi-Scale Lunar Craters Detection Method. Radar Science and Technology, 2022, 20(1): 65-73.
[13]. LI Dongbin, GAO Hongwei, WANG Hui. Research on image segmentation algorithm based on "Jade Rabbit" data. Transactions of Shenyang Ligong University, 2017, 36(6): 61-67.
[14]. Labbe M, Michaud F. Appearance-based loop closure detection for online large-scale and long-term operation. IEEE Transactions on Robotics, 2013, 29(3): 734-745.
[15]. Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters. IEEE transactions on Robotics, 2007, 23(1): 34-46.
[16]. Rösmann C, Feiten W, Wösch T, et al. Trajectory modification considering dynamic constraints of autonomous robots. ROBOTIK 2012; 7th German Conference on Robotics. VDE, 2012: 1-6.
[17]. Lowe D G. Object recognition from local scale-invariant features. Proceedings of the seventh IEEE international conference on computer vision. Ieee, 1999, 2: 1150-1157.
[18]. Bay H, Tuytelaars T, Gool L V. Surf: Speeded up robust features. European conference on computer vision. Springer, Berlin, Heidelberg, 2006: 404-417.
[19]. Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 1981, 24(6): 381-395.
[20]. Pye S D, Wild S R, McDicken W N. Adaptive time gain compensation for ultrasonic imaging. Ultrasound in medicine & biology, 1992, 18(2): 205-212.
[21]. Orchard M T, Sullivan G J. Overlapped block motion compensation: An estimation-theoretic approach. IEEE Transactions on Image Processing, 1994, 3(5): 693-699.
[22]. Avidan S, Shamir A. Seam carving for content-aware image resizing. ACM SIGGRAPH 2007 papers. 2007: 10-es.
[23]. Yue H, Wang R, Zhang J, et al. Fusion of gray and grads invariant moments for feather quill crease recognition. Proceedings of 2013 Chinese Intelligent Automation Conference. Springer, Berlin, Heidelberg, 2013: 121-128.
[24]. He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.
[25]. Cheng B, Misra I, Schwing A G, et al. Masked-attention mask transformer for universal image segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 1290-1299.