

APPLICATION OF RGB-D SLAM IN 3D TUNNEL RECONSTRUCTION BASED ON SUPERPIXEL AIDED FEATURE TRACKING

Rongchun Zhang^a, Shang Shi^a, Xuefeng Yi^{b,*}, Meiru Jing^a

^a School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China, 210023 - (rongchunzhang, 1021173517, 1021173516)@njupt.edu.cn

^b School of Earth Sciences and Engineering, Hohai University, Nanjing, China, 211100 - hhuyxf@126.com

Commission II, WG II/5

KEY WORDS: Azure Kinect DK, RGB-D SLAM, Superpixel, Line Feature, Tunnel Reconstruction.

ABSTRACT:

In large-scale projects such as hydropower and transportation, the real-time acquisition and generation of the 3D tunnel model can provide an important basis for the analysis and evaluation of the tunnel stability. The Simultaneous Localization And Mapping (SLAM) technology has the advantages of low cost and strong real-time, which can greatly improve the data acquisition efficiency during tunnel excavation. Feature tracking and matching are critical processes of traditional 3D reconstruction technologies such as Structure from Motion (SfM) and SLAM. However, the complicated rock mass structures on the tunnel surface and the limited lighting environment make feature tracking and matching difficult. Manhattan SLAM is a technology integrating superpixels and Manhattan world assumptions, in which both line features and planar features can be better extracted. Rock mass discontinuities including traces and structural planes are distributed on the inner surface of tunnels, which can be extracted for feature tracking and matching. Therefore, this paper proposes a 3D reconstruction pipeline for tunnels, in which, the Manhattan SLAM algorithm is applied for camera pose parameters estimation and the sparse point cloud generation, and the Patch-based Multi-View Stereo (PMVS) is adopted for dense reconstruction. In this paper, the Azure Kinect DK sensor is used for data acquisition. Experiments are proceeded and the results show that the proposed pipeline based on Manhattan SLAM and PMVS performs good robustness and feasibility for tunnels 3D reconstruction.

1. INTRODUCTION

This paper uses the Azure Kinect DK camera to collect data and conduct investigations in the tunnel of the Bailongjiang Water Diversion Project in Gansu Province, China (Figure 1). The Azure Kinect DK camera is an inexpensive RGBD data acquisition device that can be used to collect RGBD data. Sparse reconstruction is an indispensable part of completing 3D reconstruction. In this paper, sparse reconstruction is achieved by using SLAM, which greatly reduces the time required for sparse reconstruction (Taheri and Xia, 2021). It can effectively improve the efficiency of construction. Since the experimental scene is a tunnel that has just been excavated and has not yet been supported by shotcrete, the follow-up construction will pay more attention to the shape of the rock wall.



Figure 1. The environment for data collection in this experiment was taken by Azure Kinect DK.

Therefore, there is no strict requirement for the surface texture of the modeled tunnel, and the undulation of the tunnel rock wall is more important. The rapid extraction of rock wall morphology is the focus of this paper. Using the conventional SfM method to complete the sparse reconstruction needs to take a photo first, and then spend a long time performing the sparse reconstruction. The SLAM method can achieve sparse reconstruction in real-time, which significantly improves the reconstruction efficiency. Saving the results of sparse reconstruction can provide a good data foundation even if there is a need for dense reconstruction later. Due to the particularity of the tunnel environment, it is difficult for the conventional ORB feature extraction method to complete the tracking of the SLAM algorithm (Ruble et al., 2011). This paper uses the Manhattan SLAM (Yunus et al., 2021) method to complete the entire data acquisition and sparse reconstruction process. The final results include camera poses, movement paths, keyframe numbers, and light point clouds. In geological work, for the safety of subsequent construction, it is necessary to survey the rock wall to prevent deformation, expansion, landslide, water penetration, and other problems. The texture information of the 3D reconstruction can provide rock formation and rock type information. Completing the three-dimensional reconstruction of the tunnel environment can reduce the workload of tunnel surveying and use the three-dimensional model to formulate a construction plan for the rock wall of the tunnel. Although the method of first taking pictures from different angles and then performing sparse reconstruction can meet the needs of the subsequent dense reconstruction, taking photographs and sparse reconstruction will take more time, and the pose of the camera needs to be solved by the photographs.

* Corresponding author

Using SLAM instead of SfM can complete pose calculations and sparse reconstruction when collecting data, saving a lot of time for engineering practice and improving work efficiency.

2. RELATED WORK

With the development of science and technology, more and more methods can be used for 3D reconstruction, and the cost of the equipment used is also getting lower and lower. How to use fewer data to complete 3D reconstruction that meets the needs of use is one of the current development directions.

In recent years, SLAM has been applied in many fields. For different application environments, the performance of different algorithms will vary greatly. Applications in different areas need to choose the appropriate SLAM method or modify the existing SLAM method to balance efficiency and accuracy.

ORB SLAM2 is a SLAM method that can use binocular and RGBD data and does not include a save module, which cannot save the obtained point cloud results (Mur-Artal and Tardós, 2017). LIO-SAM (Shan et al., 2020) using lidar fused with inertial odometry can provide high positioning accuracy and can also achieve high-precision mapping, but requires expensive lidar. TANDEM (Koestler et al., 2021) can achieve high modeling accuracy, but the neural network needs to be trained in advance, and the application environment needs to be re-executed. DSP-SLAM is a semantic SLAM method mainly used in outdoor scenes, but the tunnel environment is extraordinary and cannot be well adapted. Visual-Laser-Inertial SLAM uses a laser fringe projector to calculate the depth based on RGB, which can obtain good modeling accuracy. For a single image, Self-supervised Mesh Reconstruction can be used to complete the modeling by labeling the object's outline through training (Qian et al., 2021). In the tunnel scene, Xue used the method of SFM and direct linear transformation (DLT) to complete the three-dimensional modeling in the tunnel scene (Xue et al., 2021).

Without training and introducing prior knowledge, using conventional SLAM technology to complete modeling in complex scenarios requires optimization and improvement in the feature extraction part. Although ORB-SLAM2 introduces RGB-D sensors, after testing, in a scene full of many contours, such as a tunnel, if the number of feature points is increased, the operation efficiency will be significantly reduced and it is difficult to achieve real-time 3D reconstruction. Even if the data is collected first and then processed, the problem of tracking loss will occur due to too many similar feature points on the rock wall. If the number of feature points is reduced, it will make feature point tracking more difficult. As the camera moves, the trackable feature points will gradually decrease, and eventually, the tracking will fail. The Manhattan world hypothesis solves this problem nicely. There are a lot of planes and straight lines on the artificially excavated rock walls. Based on the point tracking of ORB SLAM, adding Manhattan feature tracking, that is, line features and surface features can effectively increase the elements involved in tracking and complete the matching of adjacent frames. Manhattan SLAM uses the simple linear iterative clustering (SLIC) superpixel algorithm in the face feature extraction part, which increases the efficiency of matching and improves the accuracy of matching. SLIC is very simple (Achanta et al., 2012), but it can find the boundary as well as other super-pixel methods. At the same time, it has faster speed, higher memory efficiency, improved segmentation performance, and can be directly extended to superpixel generation. In the depth image, it can meet the needs well. Due to the lack of natural lighting in the tunnel, artificial lighting is mainly used. The

handheld lighting equipment will change the light intensity of the rock wall. If superpixels are not used, the drastic changes in brightness will also affect the matching.

3. METHODOLOGY

In this research, 3D reconstruction of the tunnel processing includes two stages: sparse reconstruction and dense reconstruction. Figure 2 shows the principle flowchart of the proposed pipeline. Both camera pose estimation and sparse reconstruction are finished by Manhattan SLAM, the results of which are integrated into PMVS for dense reconstruction.

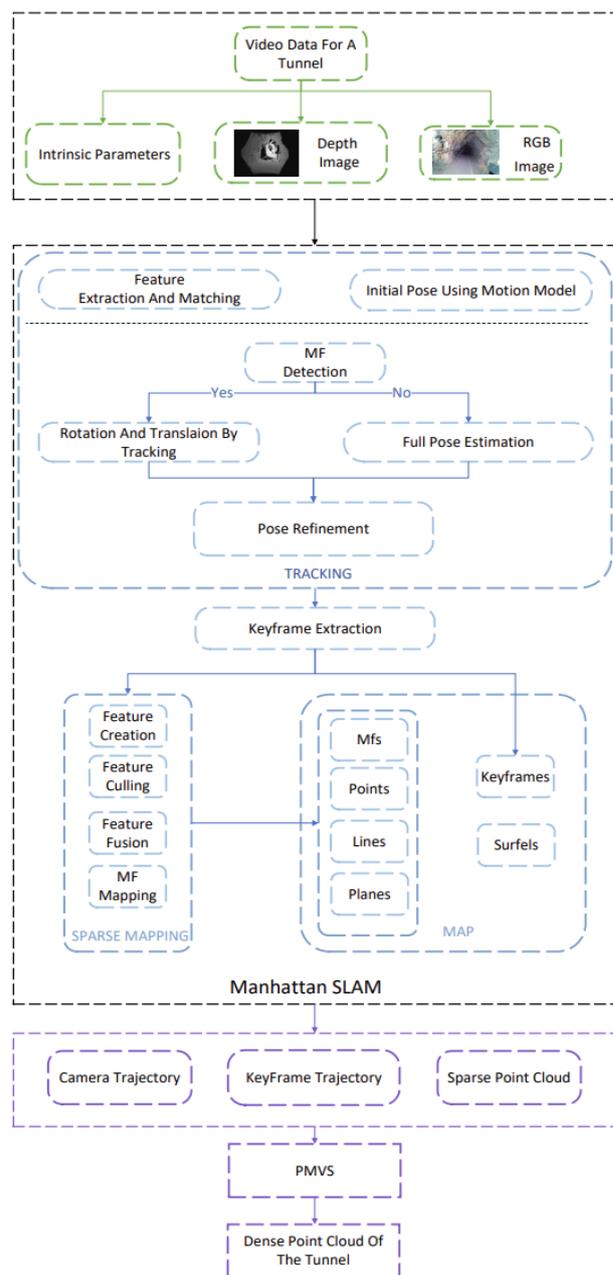


Figure 2. The principle flowchart of the proposed pipeline.

3.1 Data processing

Similar to ORB SLAM2, points and lines are extracted from RGB images, and planes are extracted from depth images during

data processing. The initial pose estimate is obtained by utilizing the isokinetic motion model and further optimized by feature correspondence and structural rules. For points and lines, the last frame's guided search is used to match features, and planes are directly compared on the global map. We then detect Manhattan Frame (MFs) to determine whether the current scene is an MW scene or a non-MW scene, using the respective pose estimation strategies. As an additional step in both cases, we track features in the local map of the current frame to further refine the pose estimation. If the current structure observes less than 90% of the points marked in the previous frame, a new keyframe is created.

3.2 Manhattan SLAM

ORB-SLAM2 is a SLAM method based on binocular or RGBD cameras. It mainly includes three threads: 1) localization of the camera at each frame by finding features that match the local map and using the moving features to perform beam adjustment to minimize the reprojection error; 2) local map management and optimize the local map, perform a local beam method adjustment; 3) Loop closure detection to detect the entire route and correct the accumulated drift by performing pose map optimization. This thread then starts the fourth to perform a full BA after pose graph optimization to compute the best structural and kinematic solutions. The Manhattan SLAM has been improved based on ORB-SLAM2. The tracking thread is modified to a certain extent, the map reconstruction part is added, and the loopback detection part is deleted.

RGBD images provide good texture and depth information, which can be used to feature extraction. The depth image performs well in planar feature extraction. The point and line features are easier to be recognized in the RGB images. Firstly, the camera is assumed to be uniformly moving, and then the initial pose of the camera is calculated according to the displacement of the feature points on the image. Next, the pose is further calculated and optimized by tracking the position changes of the features. Among them, the tracking of point and line features is matched in the current last frame, and the matching of surface features is based on the map formed by all the current frames. Through these two different pose estimation methods, the Manhattan frame is detected to determine whether MF exists in the current frame. In addition, to enhance the accuracy of pose estimation, the features in the local map saved in the current online path will be used for further pose correction. A new keyframe would be created again if the accuracy is not accepted.

Figure 3 is an example with 90% accuracy. In the feature tracking step, point feature, line feature, and planar feature are combined for the tracking task in a complex environment.

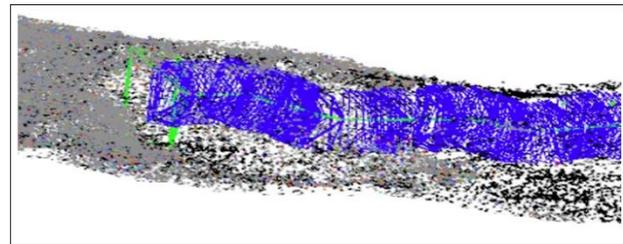


Figure 3. The movement route of the camera during the Manhattan SLAM operation. The blue shapes denote keyframes; the green line along the tunnel center axis represents the tracks of the camera; the black points show the matching feature points.

FAST is a useful feature detection algorithm, but it fails to rotation invariance. Feature point rotation is unavoidable in dynamic tracking, so FAST algorithm is not appropriate for SLAM. In the tunnel environment, the road is not flat, and the angle deviation is inevitable when collecting data. Therefore, the ORB algorithm provides a better alternative (Rublee et al., 2011). In this algorithm, the three-dimensional coordinate of the point is expressed as $P = (X, Y, Z)$. The point P is projected into a 2D plane for feature matching. The coordinates are expressed as $p_{obs} = (u, v)$. When matching, the Hamming distance between the strings corresponding to the descriptors of the feature points projected on the plane is compared. When the similarity exceeds the set threshold, it will be considered as the same feature point, so as to complete the matching. The detection method of line features is LSD (VonGioi et al., 2012). LSD algorithm uses LBD descriptor to describe line features. LSD performs better efficiency than Hough transform in line feature extraction. For a straight line either in 3D or 2D, it can be expressed by the coordinates of the two endpoints of this line. Similar to point features, line features in 3D need to be projected into a plane for matching. The two endpoints are described as (P_{start}^l, P_{end}^l) and (P_{start}^l, P_{end}^l) , respectively. The line scanned on the plane is represented by $l_{obs} = P_{start}^l \times P_{end}^l / ||P_{start}^l|| ||P_{end}^l|| = (a, b, c)$ (Yunus et al., 2021). Figure 4 shows an example of point and line feature extraction in Manhattan SLAM, which are denoted as green points and red lines, respectively.

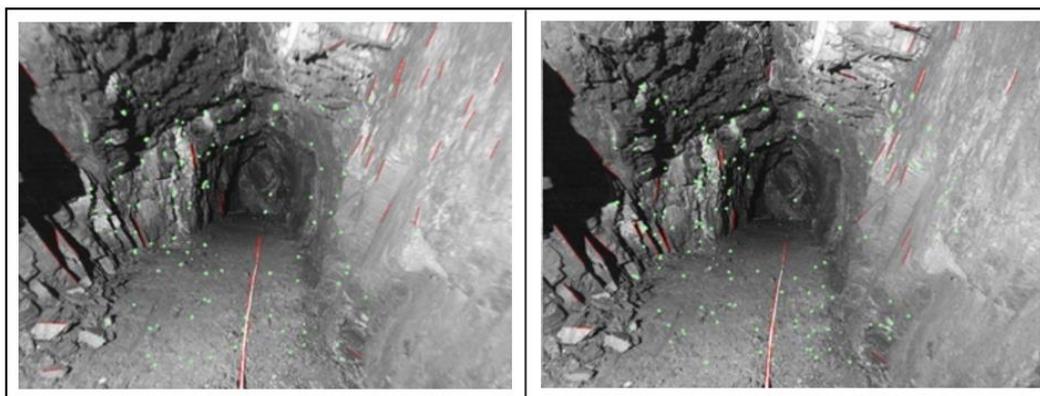


Figure 4. Feature matching results for a pair of keyframes.

The LBD descriptor is used here to project the line of three-dimensional space onto the two-dimensional plane and to complete the matching between endpoints to extract the plane

features. To improve the sampling efficiency, SLIC superpixel is used to extract the plane features of the depth image (Wang et al., 2019), and match them. The original depth image will have

noises and anomalies due to resolution. Superpixel extraction can avoid the noise and extract the plane features. The superpixel uses the SLIC superpixel using the K-means clustering method. By initializing the clustering center first, and then continuously circulating, the pixels are clustered according to the position and brightness of the pixels. This step is only carried out on the depth image. (Wang et al., 2019).

Frames are described by normal lines projected from the map onto the plane, and each frame can be described by normal lines. Through the detection of intra-frame normals, if three mutually perpendicular normals (n_1^k, n_2^k, n_3^k) are found, this frame is the Manhattan frame. At this point, a rotation matrix is used to represent the projection of Manhattan frame in camera coordinate system. However, if only two perpendicular normal lines are detected, they can also be restored to the Manhattan framework after processing. The processing method adopted is to take the cross product of these two vertical normals and calculate another normal, so as to calculate the Manhattan frame on the basis of these two planes. In the actual situation, there will be some errors in different locations due to hardware reasons, resulting in certain noise. In order to deal with the column non-orthogonality of the matrix caused by noise, the singular value decomposition method is used to approximate the matrix to the rotation matrix $R_{c|m_k}$. In order to save all the Manhattan frames in the whole map, the Manhattan map is created in the system to save the Manhattan frames found.

3.3 3D dense reconstruction for tunnels

Three-dimensional reconstruction is comprised of sparse reconstruction and dense reconstruction. In the tunnel environment, the rock mass surface morphology plays an important role during geological analysis. In this paper, the sparse reconstruction results generated by SLAM lack sufficient texture features and need to be processed by the dense reconstruction method. The sparse reconstruction is finished by Manhattan SLAM, and the results including camera pose parameters and the sparse point cloud are used for the dense point cloud generation with PMVS method. The PMVS method is mainly to project a pixel on the image back to a specified surface in the space, including three steps, which are initial feature matching, surface slice generation and surface slice filtering. Firstly, a group of sparse patches are obtained by the initial feature matching, and then the final result is obtained by continuously encrypting and deleting multiple patches.

Define a mask P when encrypting a mask, and then expand it by Equation (1) (Furukawa and Ponce, 2009).

$$C(p) = \{C_i(x', y') | p \in Q_i(x, y), |x - x'| + |y - y'| = 1\} \quad (1)$$

In the process of reconstruction, those planars with low precision will be filtered according to visual consistency.

4. EXPERIMENTS AND RESULTS

4.1 Data acquisition and pre-processing

The data used in this experiment comes from handheld Azure Kinect DK camera. Due to the lack of natural lighting in the

tunnel, the lighting comes from the handheld lighting equipment, so the illumination of the rock wall will change during the process of data collection, and the brightness between different frames will vary. Completing the entire experimental process requires reasonable handling of scene changes to track the same feature.

Due to the handheld device during data collection, shaking will inevitably occur, resulting in blurred image quality. Therefore, when collecting data, select the sampling mode with frame rate priority, reducing the impact of shaking in the final data.

A total of 1 minute and 30 seconds of data were captured in the tunnel of the entire test section, and 1501 RGB and depth maps were extracted respectively. The data was collected three times in total, and the one with the smallest shaking was selected for photo extraction. When extracting photos, the camera's internal parameters encapsulated in the data are also extracted synchronously.

The experimental area is a tunnel under construction, and there is no interference from moving objects such as pedestrians when collecting data. To test the performance of the method, the analysis of dynamic scenes was done by changing the lighting conditions. During the experiment, the original engineering lighting system was removed first, and hand-held lighting equipment was used for lighting. When collecting data, the angle and position of the handheld lighting device will change, resulting in changes in the brightness of the rock wall, making it possible to obtain dynamic scenes and increasing the difficulty of tracking. In practical engineering applications, artificial lighting is generally used in the tunnel environment. Compared with natural light, artificial lighting is difficult to achieve uniform coverage. In the experiment, the hand-held lighting device can reasonably simulate the lighting conditions under challenging conditions in practical engineering applications to achieve the purpose of testing the method.

4.2 The 3D reconstruction of the tunnel

For ORB SLAM2, only using feature point tracking will be disturbed by too many rock blocks on the rock wall, and there are too many similar edges and corners, and it is easy to lose tracking. The Manhattan SLAM used in this paper adds line features based on point features to make up for the problem of tracking loss. The number of lines on the rock wall is much less than on the edges and corners, and it is much less challenging to complete continuous tracking. After pre-adjusting the number of extracted feature points before the experiment, feature tracking can still be completed at 30fps.

Figure 5 shows the sparse reconstruction results using Manhattan SLAM. Figure 5 (a), (b) and (c) respectively shows the visualization of the entrance, the side wall and the excavation face of the tunnel. Both camera pose parameters and the sparse point cloud obtained by Manhattan SLAM are further used for PMVS dense reconstruction. The visualization of the dense point cloud of the tunnel in different perspectives is shown in Figure 6.

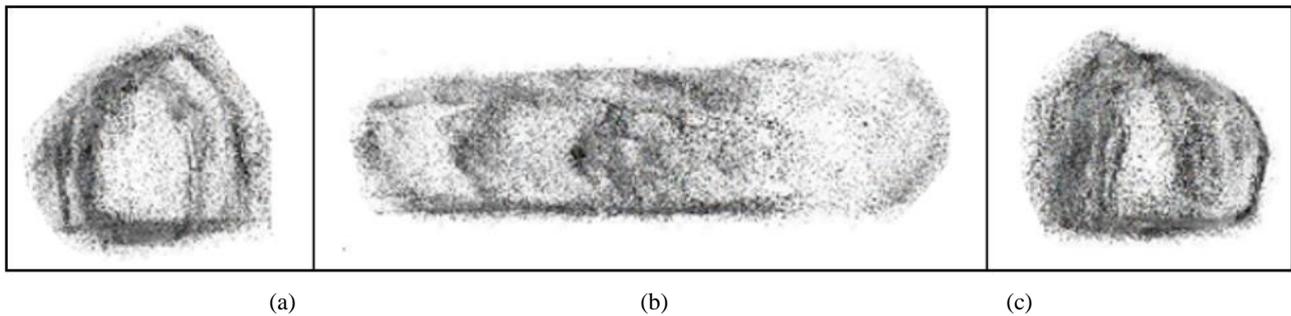


Figure 5. The visualization of the sparse point cloud generated by Manhattan SLAM: (a) the entrance of the tunnel; (b) the side wall of the tunnel (c) the excavation face of the tunnel.

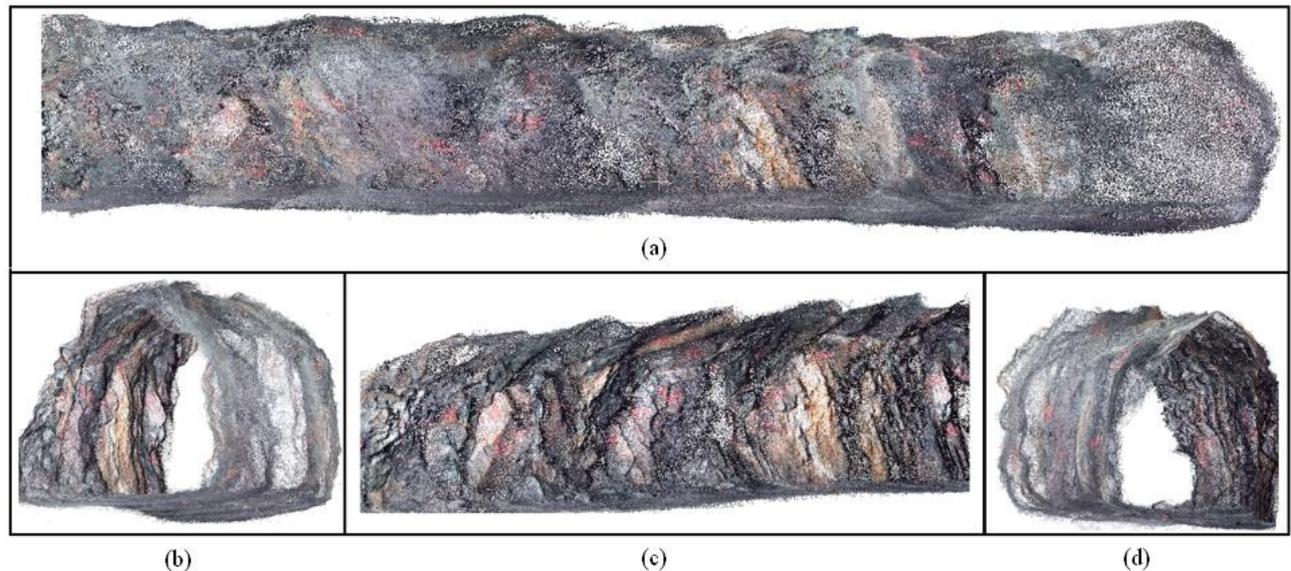


Figure 6. Dense reconstruction results: (a) the 3D dense point cloud visualization of the tunnel; (b) the view orthogonal to the central axis of the tunnel (left to right); (d) the 3D dense point cloud visualization of the side wall of the tunnel; (c) the view orthogonal to the central axis of the tunnel (right to left)

4.3 Analysis and discussion

The hardware environment used in this investigation is AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz, 16G RAM. The operating system deployed on the laptop used in the investigation is 64-bit Ubuntu 20.04. The time consume involving in data acquisition and experimental process demonstrates that the current efficiency can meet the requirements of real-time reconstruction. Thirty frames per second can be processed during operation. Significantly, the stability of the device contributes to keep the good quality of the video during data acquisition.

Different from the general modeling environments such as buildings, tunnel modelling requires data collection with higher efficiency because of the short engineering project schedule as well as poor security. Manhattan SLAM integrates line features and planar features for matching, which are considered as a supplementary when feature points is insufficient for feature tracking during camera pose parameters estimation and the sparse point cloud generation. Superpixels performs well in expressing planar feature, which is easier to be recognized and extracted from the depth images. In order to improve the efficiency, this paper extracts RGB and depth images with the high frame rate from the video data of the tunnel for experiments, the results proves that the navigation and modeling of Manhattan SLAM in environments of tunnel excavation surface are feasible.

5. CONCLUSIONS

This paper proposes a 3D dense reconstruction pipeline for tunnels. It finishes camer pose estimation and the sparse point cloud generation using the Manhattan SLAM algorithm, and achieves the 3D tunnel dense reconstruction using PMVS. The feasibility of the pipeline is proved by experiments.

The proposed pipeline in this paper has the following advantages: (1) Both the line features and the planar features are combined with the point features for features tracking and matching, which effectively improves the accuracy of the results. (2) The data acquisition and modeling process can be done in real-time, which is essential for the stability analysis of tunnels under the premise of safety. (3) The RGBD camera is a portable and inexpensive device for tunnel 3D reconstruction, which has good application prospects.

The further research will concentrate on 3D tunnel reconstruction using images and video data with higher resolution. ROS robots will be considered for path finding and data acquisition, which has broad development prospects.

ACKNOWLEDGEMENTS

This research was funded by the National Natural Science Foundation of China (Grants No. 41901401), the Natural Science

Foundation of Jiangsu Province (Grants No. BK20190743), and the China Postdoctoral Science Foundation (Grants No. 2021M691653).

REFERENCES

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11), 2274-2282.
- Cheng, D., Shi, H., Xu, A., Schwerin, M., Crivella, M., Li, L., Choset, H., 2021: Visual-Laser-Inertial SLAM Using a Compact 3D Scanner for Confined Space. *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5699-5705.
- Du, Z. L., Ma, Y., Li, X. L., Lu, H. M., 2019: Fast scene reconstruction based on improved SLAM. *Computers, Materials and Continua*, 61(1), 243-254.
- Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Fitzgibbon, A., 2011: Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 559-568.
- Karkus, Peter, Shaojun Cai, David Hsu ,2021: Differentiable slam-net: Learning particle slam for visual navigation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2815-2825.
- Koestler, L., Yang, N., Zeller, N., Cremers, D, 2022: TANDEM: Tracking and Dense Mapping in Real-time using Deep Multi-view Stereo. *Conference on Robot Learning*. PMLR. 2022, 34-45.
- L. Ling, I. S. Burrent and E. Cheng, 2012, A Dense 3D Reconstruction Approach from Uncalibrated Video Sequences, *IEEE International Conference on Multimedia and Expo Workshops* 587-592, doi.org/10.1109/ICMEW.2012.108.
- L. Li, H. Kim, S. Jiang, Y. Kim and T. Kuc, 2018: Feature saliency based SLAM of mobile robot, *International Conference on Electronics, Information, and Communication (ICEIC)*, 1-3, doi.org/10.23919/ELINFOCOM.2018.8330615.
- Tölgyessy, M., Dekan, M., Chovanec, E., Hubinský, P., 2021: Evaluation of the azure Kinect and its comparison to Kinect V1 and Kinect V2. *Sensors*, 21(2) 413.
- Mur-Artal R., Tardós J. D., 2017: ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics* 33(5), 1255-1262.
- Newcombe, R. A., Davison, A. J., 2010: Live dense reconstruction with a single moving camera, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1498-1505.
- Pan, H., Guan, T., Luo, Y., Duan, L., Tian, Y., Yi, L., Yu, J. 2016: Dense 3D reconstruction combining depth and RGB information. *Neurocomputing*, 175, 644-651.
- Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall, 2012: LSD: a Line Segment Detector, *Image Processing On Line*, 35–55. doi.org/10.5201/ipol.2012.gjmr-lsd
- R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, 2015: ORB-SLAM: A Versatile and Accurate Monocular SLAM System, *IEEE Transactions on Robotics*, 31(5), 1147-1163, doi.org/10.1109/TRO.2015.2463671.
- Rusu, Radu Bogdan, Steve Cousins, 2011: 3d is here: Point Cloud Library (PCL), *IEEE international conference on robotics and automation*. IEEE, 1-4.
- Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011: ORB: An efficient alternative to SIFT or SURF. *2011 International conference on computer vision*. IEEE, 2564-2571.
- Schonberger, J. L., Frahm, J. M. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104-4113.
- Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., and Rus, D, 2020. LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5135-5142
- Straub, J., Rosman, G., Freifeld, O., Leonard, J. J., & Fisher, J. W., 2014: A mixture of manhattan frames: Beyond the manhattan world. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3770-3777
- Sun, J., Xie, Y., Chen, L., Zhou, X., Bao, H. ,2021: NeuralRecon: Real-time coherent 3D reconstruction from monocular video. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15598-15607.
- Taheri, H., Xia, Z. C. ,2021: SLAM; definition and evolution. *Engineering Applications of Artificial Intelligence*, 97, 104032.
- Tang J, Folkesson J, Jensfelt P., 2019: Sparse2dense: From direct sparse odometry to dense 3-d reconstruction, *IEEE Robotics and Automation Letters*, 4(2), 530-537.
- Wang K, Gao F, Shen S., 2019: Real-time scalable dense surfel mapping, *International Conference on Robotics and Automation (ICRA)*. IEEE, 6919-6925.
- Weilharther R, Fraundorfer F. Highres-mvsnet, 2021: A fast multi-view stereo network for dense 3d reconstruction from high-resolution images. *IEEE Access*, 9, 11306-11315.
- Xue, Y., Zhang, S., Zhou, M., Zhu, H, 2021: Novel SfM-DLT method for metro tunnel 3D reconstruction and Visualization. *Underground Space* 6(2), 134-141.
- Y. Furukawa and J. Ponce, 2009: Accurate, Dense, and Robust Multiview Stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1362-1376, doi.org/10.1109/TPAMI.2009.161.
- Yunus R, Li Y, Tombari F, 2021: ManhattanSLAM: Robust Planar Tracking and Mapping Leveraging Mixture of Manhattan Frames. *IEEE International Conference on Robotics and Automation (ICRA)*, 6687-6693. doi.org/10.1109/ICRA4850-6.2021.9562030.
- Yeh, Chia-Hung, Min-Hui Lin, 2021: Robust 3D Reconstruction Using HDR-Based SLAM. *IEEE Access* 9 ,16568-16581.