A Review of Integrating Visual SLAM Technology into Minimally Invasive Surgery Environment to Optimize Surgical Scene

Chang Gao [1]

[1]School of automation, University of Science and Technology Beijing,Beijing,100080, China

U202243344@xs.ustb.edu.cn

Abstract. Minimally invasive surgery (MIS) does offer many advantages over traditional open surgery, including less trauma, shorter hospital stay, less pain, and a lower risk of infection. However, there are some limitations to Mis, especially the technical challenges associated with the surgical scenario. Based on the literature in recent years, the main methods of scene optimization in MIS combined with slam were compared, and three visual slam methods were selected to compare, analyze and discuss them. Through the comparison and discussion of the three methods and the analysis of the limitations and advantages of each technology, it can be seen that there is still a certain space for progress in the field of MIS. Among the three technologies, MIS-SLAM technology can realize large-scale localization and intensive mapping in real time, which makes it have great application potential in the medical field. By transforming and morphing the current model, the technology not only retains vivid texture information, but also gradually integrates new observations to build a more accurate and comprehensive three-dimensional scene.

Keywords:MIS-SLAM, SD-DEFSLAM, RGB-DSLAM, Minimally Invasive surgery.

1 Introduction

1.1 A Subsection Sample

MIS is a highly challenging surgical procedure that requires operating instruments to be manipulated within limited anatomical spaces, and the images obtained by the instruments are often distorted, sometimes even incomplete[1]. Traditional MIS faces a series of problems, including inconvenience in endoscope operation, limited field of view, serious reflection problems, difficulty in acquiring depth information, and challenges in accurately locating the target area. Moreover, prolonged surgical examination can cause discomfort to the patient. In response to these challenges, SLAM systems have been innovatively applied to the field of MIS to solve the difficulties faced by modern MIS.In order to help surgeons simulate the actual

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_28

surgical procedure, digital 3D reconstruction models can be printed at equal scale through 3D printing technology[2][3]. By building a three-dimensional map of the abdominal cavity, the robot can autonomously plan the surgical path and adjust the position and angle of the surgical instruments in real time. This technology not only reduces the workload of doctors but also improves the efficiency and accuracy of the surgery.

In recent years, more and more research has been focused on the application of visual SLAM technology in MIS, with these studies aiming to overcome the challenges faced in MIS. Visual SLAM technology has gained much attention for its advantages in recovering three-dimensional structures and estimating endoscope movements. However, traditional SLAM technology has a serious problem of being heavily dependent on scene rigidity, ignoring the challenges of the dynamic parts of the scene. Additionally, compared to open surgery, MIS faces problems of limited field of view, inaccurate range localization, and limited surrounding information. Due to these limitations, surgeons operate with delicate tools in a confined space, lacking direct three-dimensional visual support.

2 Research Goal

The research content of this paper is the application of visual slam technology in minimally invasive surgery to optimize the surgical scene. After reading a number of literatures, this paper mainly introduces and summarizes three kinds of visual slam technology. It is expected that through the research and summary of a variety of technologies, new ideas and inspiration can be obtained in the future, and further development can be made in the optimization of minimally invasive surgery scene. The development of the application of visual slam in the field of minimally invasive surgery The early visual SLAM was based on filtering theory. Mountney et al. studied the application of VSLAM in MIS, and applied the extended EKF-SLAM framework to MIS environment. However, its nonlinear error model and huge computational workload became the obstacles to its actual implementation. A powerful camera tracking and mapping estimator with excellent camera repositioning capabilities is then proposed, known as the ORBSLAM system[4]. Song[5]presents a learning-driven approach to achieve 3D reconstruction of surgical scenes and preliminary localization of laparoscopy.Wei[6]proposed image-guided laparoscopic localization was obtained through 3D reconstruction of the anatomical structure
Mahmud et al have proposed a dense 3D reconstruction of abdominal cavity based on ORB-SLAM. Compared with SLAM sparse reconstruction, this method reconstructs a denser abdominal cavity model, has clearer texture and higher real-time performance, and is suitable for abdominal cavity images with unequal light.

On the basis of ORB-SLAM technology, the purpose of this study is to search for slam technology in minimally invasive surgery to optimize the surgical scene and improve the surgical accuracy, and to compare and discuss the advantages and disadvantages of these technologies.

3 Principle And Application of Prior Art

3.1 Principle of MIS-SLAM Technology

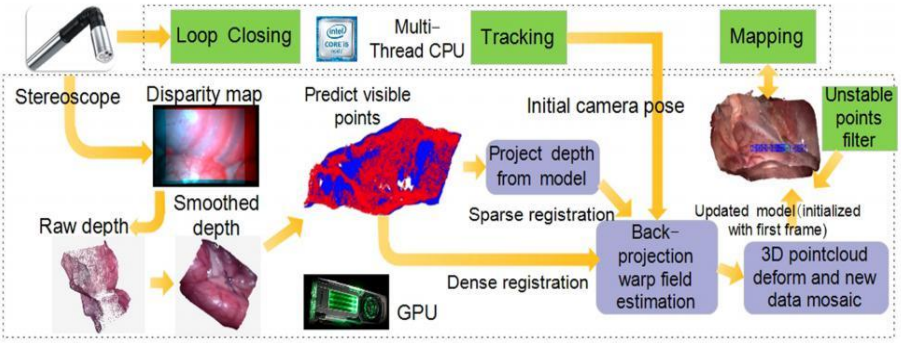

Fig. 1. Framework of MIS-SLAM [7].

The architecture of MIS-SLAM technology mainly consists of two parts: initial tracking and deformable tracking. In the illustration shown in Fig.1, the CPU primarily undertakes the tasks of processing the ORB-SLAM algorithm, uploading data, determining the global pose, and starting the visualization module. Meanwhile, the GPU focuses on functions such as depth estimation, image alignment, data fusion, and visualization presentation.

In the initial tracking phase, the improved version of the ORB-SLAM algorithm is implemented by the CPU. In the deformable tracking phase, the GPU is responsible for implementing dense mapping. Specifically, after the CPU passes the initial global pose information, the GPU uses the first estimated depth data to initialize the model. Every time new observation data is input, the GPU receives the matched ORB function from the CPU, selects the possible visible points from the model, and projects these points onto the two-dimensional depth image. Then, the global pose and non-rigid deformation field are precisely estimated through the registration process.

In order to provide doctors with a better view during surgery, two images can be generated from different angles using a 3D laparoscope or a pair of binoculars, which can create a 3D geometric structure based on parallax in the doctor's mind to better understand the environment. In this paper, a deformable soft tissue dynamic reconstruction system based on stereomicroscope is proposed. By using the deformation field based on the embedded deformation node method, three-dimensional dense shapes can be recovered from stereo images in real-time. With the help of general computing on graphics processing unit (GPGPU), the results can be displayed in real-time.

By introducing an efficient large-scale stereo matching (ELAS) depth estimation method, we can further optimize system performance[8]. Fig.3 shows the generated

depth map and its smoothed effect. To enhance the robustness of the system, we fully utilize the idle CPU resources to run the ORB-SLAM module, thus providing more reliable initial conditions for subsequent pose enhancement. The ORB-SLAM module efficiently utilizes ORB features on the GPU, accurately tracking motion trajectories while reducing GPU computing load even when the camera moves quickly. Rapid motion may cause distortion in model construction, leading to the mismatching of blurred image pseudo-edges and depth information. However, after the global robust algorithm upgrade, the system can still maintain stable global poses and ensure high-quality output results.As shown in Fig.2.

Fig. 2. Examples of depth and smoothing depth [7].

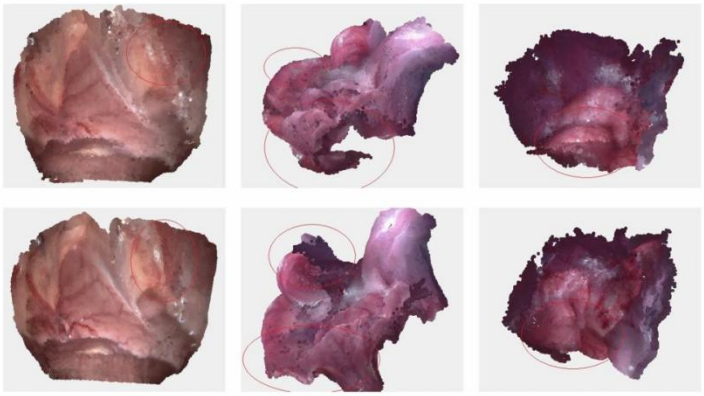

Fig. 3. Comparison of before (first row) and current (second row) [7].

3.2 Principle Of SD-DEFSLAM Technology

To better adapt to the constantly changing environment, SD-DefSLAM adopts a fusion of direct and indirect methods: by enhancing the correlation between data using the Lucas-Kanade optical flow method, it uses geometric bundle adjustment technology to precisely estimate the camera's pose and generate a deformable map. Additionally, it employs a bag-of-features model based on feature descriptors to achieve camera relocalization. Furthermore, the system incorporates a convolutional neural network (CNN) specifically trained for a particular application domain to recognize and segment dynamic objects, thereby further enhancing the system's

practicality and flexibility.As shown in Fig.4.By improving data association and deformation tracking, reducing projection error, improving scene stability, using BOW technology to reposition, recovery tracking using semantic information and training CNN to solve identification errors caused by moving objects in deformable scenes, so that the system can better calculate map deformation.The system utilizes photometric methods to effectively associate short-term and medium-term data, while optimizing the backend to handle geometric errors. By adopting an improved Lucas-Kanade (LK) algorithm, the system tightly connects image sequences, achieving data consistency within a short period. During this process, the tracking thread simultaneously estimates the camera's pose and the surface deformation, striving for the minimum geometric projection error.

When the system loses data for some reason, the missing frames are converted into a Bag-of-Words (BoW) model and queried from the recognition database to obtain some candidate key frames. For each key frame, the system calculates the correlation relationship with the mapping points. Using the PnP (Perspective-n-Point) method, the system can obtain the initial camera pose. It is worth noting that the PnP method has certain limitations in handling rigid problems, but the system is able to effectively address the short-term occlusion problems commonly encountered in endoscopic surgery, demonstrating high robustness and practicality. Using semantic information, CNNS are trained to identify and segment the typical moving pairs in each application domain, cover the corresponding image regions to avoid their feature matching, and solve the identification errors caused by matching moving objects in deformable scenes.

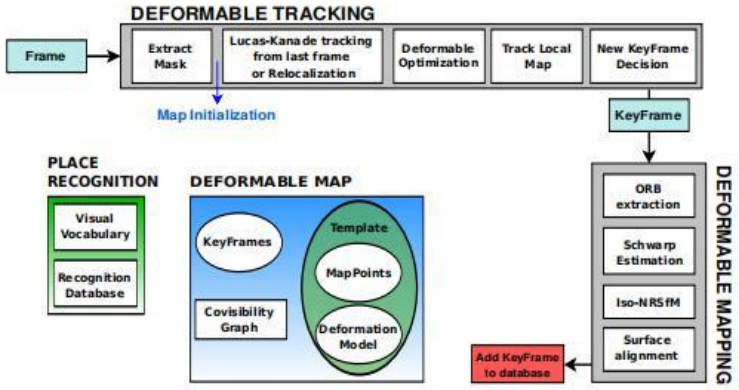

Fig. 4. SD-DefSLAM scheme running concurrently with tracing and mapping threads [9].

3.3 Principle of RGB - DSLAM Technology

This semantic SLAM system greatly improves the robot's perception of the OR environment through accurate trajectory estimation and semantic scene understanding [10]. The system has the potential to aid robot setup, surgical navigation, and improve human-robot interaction.

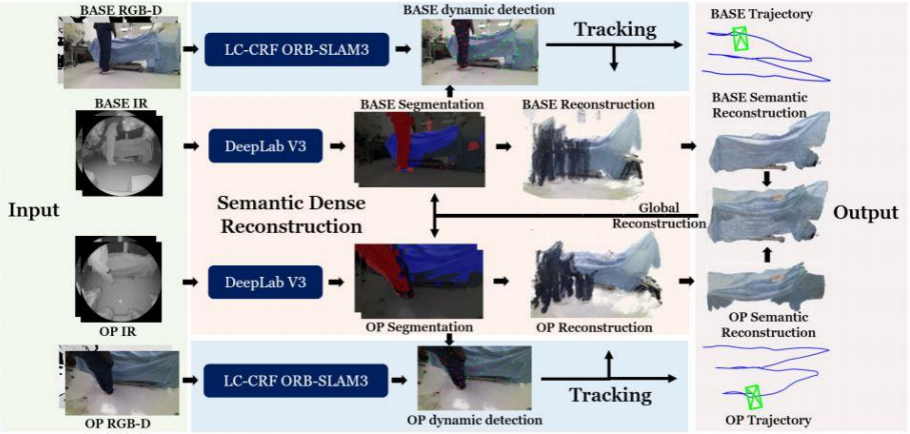

Fig. 5 Overall flow of the system [10].

As shown in Fig.5, the technology aligns the Kinect's two sensors with the robot's motion chain. The data obtained from the ToF depth camera and the RGB camera can be converted into the basic framework of the robot. Semantic refactoring modules from BASE and OPSLAM are then fused together to create global refactoring. The integrated global reconstruction combines information from different angles to make up for the limited shooting range. It greatly improves the robot's ability to perceive OR environment.

4 Discussion

The three techniques mentioned above have some shortcomings, and the biggest problems in MIS-SLAM is texture mixing. ： This lighting problem can lead to systematic differences between the two images. In the future,the work will focus on reducing computational complexity as the model grows and exploring a way to balance textures from different lighting. SD-DEFSLAM's experiments are based on deformable models assuming isometric deformations.This assumption is quite rigid and may not always apply to all situations, which could lead to a decline in the accuracy of deformation estimation, thereby negatively affecting the quality of data association. To address this issue, it is advisable to study and apply new deformation models that can more accurately describe non-uniform deformation, thereby improving the robustness and accuracy of the system.While RGB-D cameras are capable of providing depth information, the stability of RGBD cameras may be compromised due to factors such as changes in illumination, occlusion, blood and body fluids that may be present in the surgical environment, resulting in reduced accuracy in positioning and map construction. In the future, more advanced

calibration techniques and image processing algorithms can be used to solve the problem through research.

Additionally, the challenges encountered in practical operation for different types of surgeries are also varied. For example, in cardiovascular surgery, the focus of future research should be on using endoscopic visual technology to precisely guide and track the cardiovascular endoscopic surgery process. In the field of gastrointestinal navigation system, the core goal is to assist doctors in more comprehensively and accurately detecting and describing the pathological conditions of the stomach and intestines. However, how to improve detection accuracy and enhance scene recognition ability is still a subject that needs to be explored. As for respiratory surgery, although some clinical applications and commercialized products have already been launched, due to the complicated and narrow respiratory branches, the main technical challenge in visual navigation process is how to exclude the interference of complex branches to accurately reach the target. In the operation implementation, the real-time, accuracy, and stability of the system, which can meet the large-scale real-time processing requirements, will also have a decisive impact on the final technical choice.

5 Conclusion

MIS-SLAM and Def SLAM are more suitable for modulo dense deformable data or deformable scenes, while RGB-D SLAM is more used to provide positioning and map information. MIS-SLAM has clear application advantages in minimally invasive surgery, while SD-DEFSLAM and RGB-D SLAM may need further optimization and improvement to be better used in this field. This paper mainly focuses on the application and advantages of the above three advanced visual slam technologies in specific surgery.

With the continuous development of slam in the positioning and mapping of minimally invasive surgery, the future will be to higher positioning accuracy, stronger robustness and stability, small distortion, high-speed real-time image processing direction. In order to improve the efficiency and safety of surgery, efforts are being made to provide doctors with more advanced surgical navigation and positioning technologies. The MIS-SLAM technology being discussed in this article has significant advantages in clinical augmented reality (AR) or virtual reality (VR) applications in fast-moving scenarios. The focus of future research will be on reducing the computational complexity caused by model growth and seeking a way to balance the effects of different lighting and textures. At the same time, research is being conducted on a closed-loop processing mechanism to improve system performance when previously detected shapes are re-detected.In addition, recent research has shown that although preoperative data fusion into MIS navigation can provide valuable guidance to doctors, it cannot solve the problem of tissue deformation during surgery. Although intraoperative data fusion technology has great potential, it still needs to be greatly improved before it can be applied in real-time. In the future, multimodal data fusion technology is expected to be one of the important

directions of development for surgical navigation systems, providing doctors with more precise and real-time surgical support. There are other visual slam techniques not mentioned in this paper that can be applied to minimally invasive surgical scenes, and it is possible to think about integrating the technology in the future to achieve more accurate navigation.

References

1. T. Bergen and T. Wittenberg, "Stitching and surface reconstruction from endoscopic image sequences: A review of applications and methods", IEEE J. Biomed. Health Inform., vol. 20, no. 1, pp. 304-321, Jan. 2016.
2. Fatima, S.; Haleem, A.; Bahl, S.; Javaid, M.; Mahla, S.; Singh, S. Exploring the significant applications of Internet of Things (IoT) with 3D printing using advanced materials in medical field. *Mater. Today Proc.* (2021)
3. Wu, H.: 3D texture reconstruction of abdominal cavity based on monocular vision SLAM for minimally invasive surgery. Symmetry **14**(2): 185(2022)
4. Mountney, P., Stoyanov, D., Davison, A.J., Yang, G.-Z.: Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, Springer, Heidelberg 4190, 347-354 (2006)
5. Song,G.: BDIS-SLAM: a lightweight CPU-based dense stereo SLAM for surgery. International Journal of Computer Assisted Radiology.Surgery (2024)
6. Wei, R.: Stereo-dense scene reconstruction and accurate localization for learning-based laparoscopic navigation in minimally invasive surgery." IEEE Biomedical Engineering Repertoire **70**(2), 488-500 (2022)
7. J. Song, J. Wang, L. Zhao, S. Huang and G. Dissanayake.: MIS-SLAM: Real-Time Large-Scale Dense Deformable SLAM System in Minimal Invasive Surgery Based on Heterogeneous Computing, in IEEE Robotics and Automation Letters **3**(4), 4068-4075 (2018)
8. Song, G.: Dynamic reconstruction of deformable soft-tissue with stereo scope in minimal invasive surgery. IEEE Robotics and Automation Letters (2017)
9. J. J. Gómez-Rodríguez, J. Lamarca, J. Morlana, J. D. Tardós and J. M. M. Montiel.: SD-DefSLAM: Semi-Direct Monocular SLAM for Deformable and Intracorporeal Scenes, 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 5170-5177 (2021)
10. Gao, C.: Inesh Rabindland ,Omid Muhammad"RGB-D Semantic Slam (2022)