SLAM-based technology to improve the impact of uneven illumination on minimally invasive surgery

Yuxuan Han

School of International Education, Beijing University of Chemical Technology, Beijing, 100029, China

2020090073@buct.edu.cn

Abstract. Due to developments in surgery, minimally invasive surgery (MIS) has become a common choice for surgical patients. Because of the small surgical incision in MIS, its postoperative effects will be less fat liquefaction, infection, splitting, numbness of the incision and weakness of the healing abdominal wall muscles compared to traditional surgery. However, many bottlenecks have been encountered in MIS, such as uneven illumination affecting the recognition of organ images. In other words, such problems are about improving the ability of Simultaneous Localization and Mapping (SLAM) to extract information about the environment. This paper summarises some relevant research on SLAM improvement methods. Scientists are now looking for more suitable methods for constructing maps of the human internal environment, including improvements on existing SLAM techniques. In the context of SLAM, what SLAM is and the history of SLAM is described. Traditional SLAM methods as well as SLAM improvement methods are also mentioned. Finally, its advantages and disadvantages are discussed, and its prospects are predicted.

Keywords: minimally invasive surgery, SLAM, ORB-SLAM.

1. Introduction

Surgical procedures and treatment methods are continuously improving with medical technology and practices advancement. At this point, MIS may substitute computerized endoscopic surveillance for operating with the human eye, considerably enhancing surgical precision. The advent of robot-assisted surgery can help surgeons perform more flexible procedures on delicate areas by overcoming the drawbacks of traditional laparoscopy in surgery, such as poor depth perception, unavoidable hand tremors, and the surgeon's greater susceptibility to fatigue after lengthy surgery [1]. Additionally, the technique minimizes the size of the incision produced on the patient, minimizing harm to organ tissue, minimizing the possibility of patient stress, and expediting the healing process after surgery.

The computer must create a three-dimensional model of the body's interior tissues and other data from the endoscope during MIS [2]. Additionally, to successfully handle the robotic camera and robotic forceps to allow image alignment and organ tracking, console surgeons must thoroughly comprehend the strengths and shortcomings of the AR navigation system [3]. The involvement of console surgeons enables greater communication between the 3D computer graphics model and the real-time altered navigation system pictures [3]. However, challenging visual, perceptual, and dexterous restrictions are usually present with such procedures [4]. Due mainly to the complexity of the surgical environment,

DOI: 10.54254/2755-2721/12/20230292

MIS faces several problems such as high brightness caused by illumination devices, incomplete edge information in internal cavity images, and soft tissue bleeding affecting image acquisition during surgery [2]. This has led to the introduction of improved SLAM techniques into MIS, which allow surgeons to reduce the impact of inaccurate computer reconstructions of 3D structures due to the surgical environment.

This paper first describes what SLAM is and the process of SLAM. This is followed by a history of SLAM and its applications in medicine. The paper describes four of the most fundamental and promising kinds of SLAM: ORB-SLAM, Superpoint, ORB-SLAM-based algorithm for laparoscopic position estimation (OSALPE), and K-Means in combination with Superpoint (KMS). The working mechanism and characteristics of these four methods are introduced respectively, and a comparison of these four SLAM methods is made. Finally, the advantages and disadvantages of the four methods are discussed by comparing the percentage of valid points in feature points and matching points correct rate, and the development prospects of SLAM for medicine are predicted.

2. SLAM and MIS

The most important application of SLAM in the direction of MIS is the robotic processing of medical images. The following will first describe the SLAM and MIS relationship. And with specific types of SLAM, including ORB-SLAM and Superpoint, as examples, their characteristics will be discussed, and their advantages and disadvantages considered and compared.

2.1. Introduction of SLAM

The duty of SLAM can be broken roughly into three steps. First, the robot must finish acquiring images of the new environment information and extracting feature points. Next, the robot must determine its location using the feature information. Finally, the environment must be recreated using the information.

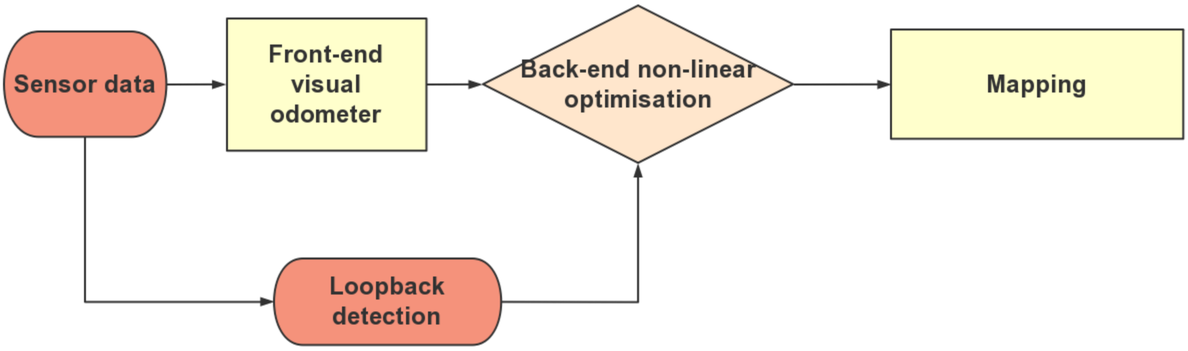

Figure 1. The process of SLAM.

Figure 1 depicts the SLAM procedure in five steps. The process is briefly reading image information, constructing a local map, optimizing for noise or matching errors, loopback detection, and building the map according to the rules.

Sensor data is frequently used in vision SLAM to characterize the visual sensor. Most visual sensors are currently being studied in laboratories and have limited real-world product uses. The visual sensor finds it challenging to directly measure distances concerning the surroundings; instead, it must estimate its positional change using two or more image frames and then determine its current location by adding up the positional change. The visual sensor gathers data about the surroundings, pre-processes it, and then sends it to the visual odometer [5].

The visual odometer estimates the camera pose from the environmental image information delivered by the visual sensor. The feature point approach and the direct method are two methods that the SLAM front-end uses to remove some near-misses.

The noise generated by the robot while it is moving and the buildup of errors in the front-end process are the two key issues that the back-end non-linear optimization has to address [5]. The back-end uses

DOI: 10.54254/2755-2721/12/20230292

the non-linear approach to optimize the overall situation utilizing the positional data that the front-end has given. Edge-to-edge residuals, face-to-face residuals, motion distortion correction, and iterative bit-pose optimization are the major components of the back-end. A filtering approach and a graph optimization method are often used to handle them.

A key component of SLAM is loopback detection, commonly called closed-loop detection. It is employed to ascertain whether the robot has repeatedly reached the same location [5]. A consistent trajectory map is produced by the constraint of loopback detection, which is a more precise and compact constraint than the back end. Loopback detection and global optimization are required to provide a full picture. Offsets caused by trajectory drift can be considerably reduced using loopback detection.

Based on the camera trajectory and picture data, a map is built during the map-building session to satisfy the job requirements, using map types such as raster maps, topological maps, sparse maps, and dense maps [5].

2.2. History of SLAM

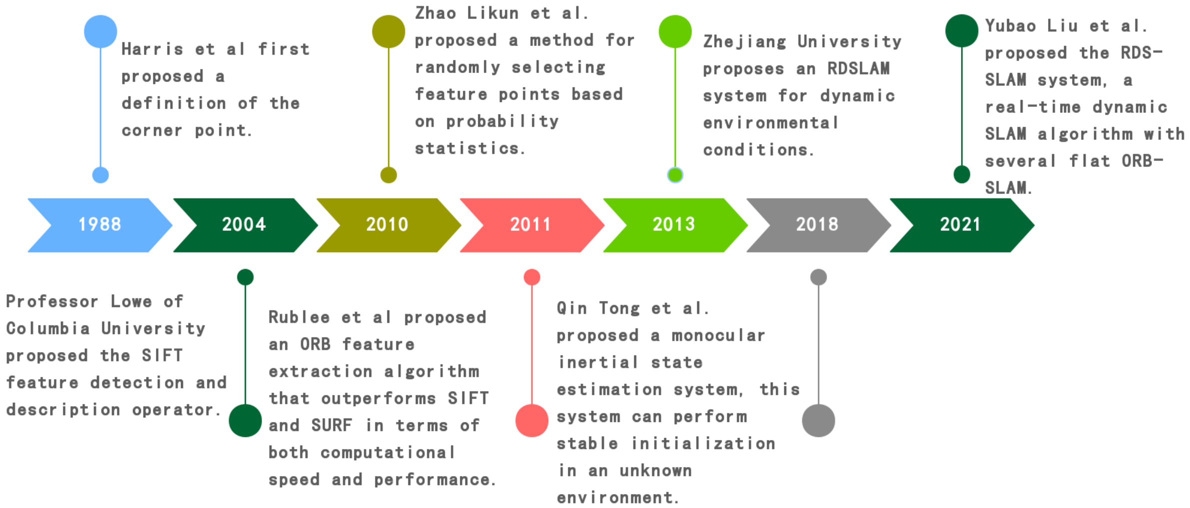

Figure 2. History of SLAM [6-12].

Robotic processing of medical pictures is the most significant use of SLAM in the direction of minimally invasive surgery. The development of SLAM is depicted in the timeline in Figure 2. According to the state of research, the use of visual SLAM in medicine is increasingly gaining traction but is not yet fully developed. There is currently more study on medical image processing techniques, although it is only marginally better than earlier studies, and there is limited research on endoscopic visual SLAM approaches. Therefore, researching visual SLAM techniques in MIS is quite important.

2.3. The SLAM approach to MIS

Table 1 summarizes four solutions to the issue of insufficient feature information extraction brought on by high bright spots.

DOI: 10.54254/2755-2721/12/20230292

Table 1. Comparison of the four SLAM methods.

Methods	Innovation	Advantages	Percentage of valid points (%)	Matching points correct rate (%)	References
ORB-SLAM	None	A high number of successfully matched points; High mismatch rate	100	80.85	[2]
Superpoint	None	Few successful matches; Low mismatch rate	74.72	95.06	[2]
OSALPE	Optimization of ORB feature point extraction; Designing key frame discrimination strategies	A high number of successfully matched points; High mismatch rate	100	85.96	[15]
KMS	K-Means combined with Superpoint; PANSAC algorithm to remove false matches	A high number of successfully matched points; Low mismatch rate	100	95.70	[2]

2.3.1. ORB-SLAM. The first strategy is the ORB-SLAM strategy, which uses three threads for simultaneous localization and map building: tracking, map creation, and loopback detection [13]. This SLAM system is built on a single purpose and calculated around ORB characteristics. The tracking thread determines every picture frame's camera location and when to add a new keyframe [5]. By comparing initial feature points to the preceding picture frame, the camera location is then optimized using motion BA. The position recognition module performs global repositioning if features are missing [5]. A local visual map is extracted using the Covisibility Graph of the keyframes kept by the system after the initial camera pose estimate and feature match are obtained. The current frame is then searched for matches corresponding to the local map points by a reprojection method, and all matches are used to optimize the current camera pose. Ultimately, the tracking thread chooses whether to add a new keyframe [5]. To assure stability, the conventional ORB-SLAM is enhanced with Features by Accelerated Segment Test (FAST) and BRIEF. For the newly added picture frame, ORB employs the FAST algorithm to execute a FAST corner point search, which looks for any pixel locations throughout the whole image frame with a significant difference in grey value from enough pixels in their immediate neighbourhood [13]. The construction, tracking, relocation, and closed-loop detection modules of ORB-SLAM all employ the same features, increasing system efficiency without including further features for closed-loop detection. Although the extracted feature points in the experimental findings have a high proportion of legitimate points, the mismatch rate is significant. It does not produce more accurate results that are acceptable.

2.3.2. Superpoint

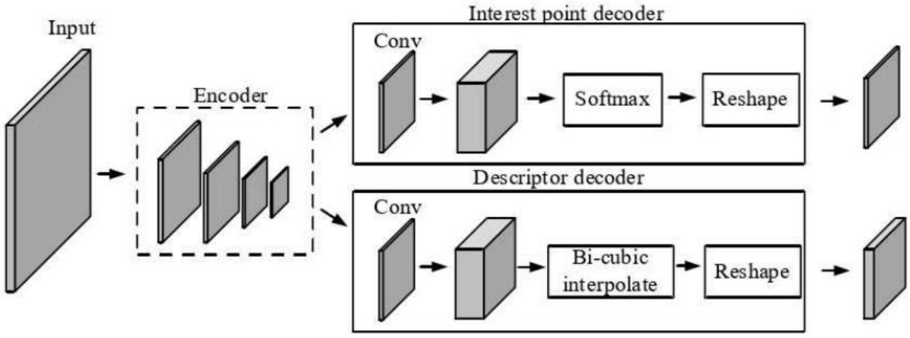

Figure 3. Superpoint network diagram [2].

Superpoint is the second approach. A feature point algorithm with descriptors is the Superpoint algorithm's ultimate output [14]. Figure 3 displays a schematic of the Superpoint network's organization. The network is split into four sections: the loss function, the descriptor detection network, the feature point detection network, and the coding network. Figure 3 illustrates two networks that share some of the same structures in the first half. The shared network in Figure 3 is the coding network. The feature point detection network functions as a decoder since it has an encoding that must be broken down in this situation. Each pixel in an image is given a probability value by the feature detection network, which represents the likelihood that the pixel is a feature point. A decoder is also part of the descriptor detection network. First, a semi-dense descriptor is learned, followed by a bicubic interpolation to acquire the entire descriptor, and finally, an L2-normalizes operation to provide a unit-length descriptor. L_p is the loss function of the feature points, L_d is the loss function of the descriptors and is the coefficient weight to balance L_p and L_d. The loss function is:

$$L(X, X', D, D'; Y, Y', S) = L_p(X, Y) + L_p(X', Y') + \lambda L_d(D, D', S) \quad (1)$$

2.3.3. ORB-SLAM positional estimation improvement algorithm. The third method is the OSALPE [15]. One of the essential elements for three-dimensional reconstruction is the laparoscopic picture that matches the position parameters. ORB-SLAM can only correct small-aperture camera aberration for the features of laparoscopic pictures, such as non-constant lighting and poor texture. It is challenging to follow the estimated laparoscopic image location steadily and precisely [15]. It may be discovered by investigating the procedures and ideas behind ORB feature extraction that the descriptors of feature points are strongly tied to the image's grey scale. The grey scale values of the same item in various photos are inconsistent when impacted by light fluctuations, and the feature-matching approach in ORB-SLAM cannot effectively address these issues [15]. Based on optimizing the ORB feature extraction and matching procedure for laparoscopic pictures in ORB-SLAM, OSALPE introduces image detail improvement, highlights recovery approaches to enhance image texture characteristics, and reduces light leakage. The ORB-SLAM is based on extracting and matching ORB features being optimized. The main benefit of OSALPE is that it will pre-process the picture to screen out unreliable regional feature points and increase the accuracy of feature matching across adjacent frames [15]. To generate the image keyframes and their accompanying pose parameters for depth map extraction and 3D reconstruction, it extracts key frames from the picture sequence by creating a keyframe discrimination method. Then it modifies the key frame posture [15]. Developing the key frame discrimination approach and optimizing the ORB feature extraction technique constitutes the innovation. The accuracy of the depth map created by SLAM is increased with the introduction of picture enhancement and highlighting restoration techniques.

2.3.4. K-Means in combination with Superpoint. The fourth method is KMS [2]. As seen in Table 1, this approach differs from Superpoint in that it combines K-Means with it. To isolate the interior cavity environment from the instruments and prevent potential mismatches caused by uneven lighting and highly reflecting surroundings, the K-Means clustering technique is paired with the Superpoint algorithm [2]. Using a self-supervised method, this network trains a base graph with uncontroversial feature points to obtain an initial feature detector. It then uses a neural network to extract the image feature information, improving extraction stability and decreasing illumination sensitivity [2]. Secondly, the Random Sample Consensus (RANSAC) algorithm is used to remove mismatches [2]. The RANSAC algorithm is an iterative technique that uses the concept of local and external points to reduce the impact of noise on the result while estimating the parameters of a mathematical model [16]. Because there are a lot of feature points and nearly no point pairings that are mismatched in KMS, there are a lot of successful Out of all the methods listed above, it has the largest number of valid points matched and is also the most precise and efficient [2]. Additionally, by successfully separating the interior cavity environment from the instrument during identification, this approach can raise both the proportion of valid points that can be recovered from the feature points and the rate at which points are correctly matched.

3. Conclusion

According to the analysis of the four SLAM techniques in this paper, the K-Means clustering algorithm, followed by the RANSAC algorithm to eliminate false matches of Superpoint is less perfect than KMS, which has the highest accuracy and stable filtering of valid feature information. When comparing the values and image results obtained by each algorithm, the percentage of valid points is increased by 25.28% and the false match rate is decreased by 0.64%. The simulation results obtained using this method are more accurate and stable than the traditional algorithm for feature extraction of internal cavity images. To lessen the influence of environmental conditions like lighting on the screening of feature points, OSALPE builds on ORB-SLAM by optimizing the ORB feature extraction and matching procedure for laparoscopic pictures. For OSALPE, the right rate of successful matching points has grown by 5.11%. Despite having a lower accurate percentage of matching points than KMS, OSALPE has a very high rate of screening genuine feature points. KMS is the most reliable and precise approach combined with the four methods.

SLAM is already an indispensable part of the medical world and the future of SLAM in the medical industry is evident. However, there is also a never-ending stream of issues. For example, developing new and improved sensors for SLAM issues will be key to technological and computer-related growth. Deep learning has become increasingly popular across many scientific domains as big data, cloud computing, and computer technology have advanced. Similarly, to this, deep learning influences the area of SLAM. A module in the SLAM process, such as robot posture estimation, picture feature extraction and matching, etc., is being attempted to be replaced by deep learning. So, combining SLAM with deep learning is another direction in SLAM development.

References

[1] Y. Wang, D. Cao, S.-L. Chen, Y.-M. Li, Y.-W. Zheng, and N. Ohkohchi, “Current trends in three-dimensional visualization and real-time navigation as well as robot-assisted technologies in hepatobiliary surgery,” WJGS, vol. 13, no. 9, pp. 904–922, Sep. 2021, doi: 10.4240/wjgs.v13.i9.904.
[2] Tao Zhang, “Research on visual SLAM method in minimally invasive surgery,” Harbin Institute of Technology, 2022, doi: 10.27063/d.cnki.ghlgu.2022.000216.
[3] A. Hamada et al., “The Current Status and Challenges in Augmented-Reality Navigation System for Robot-Assisted Laparoscopic Partial Nephrectomy,” in Human-Computer Interaction. Multimodal and Natural Interaction, M. Kurosu, Ed., Cham: Springer International Publishing, 2020, pp. 620–629.
[4] L. Yang and K. Etsuko, “Review on vision-based tracking in surgical navigation,” IET cyber-systems robotics, vol. 2, no. 3, pp. 107–121, Sep. 2020, doi: 10.1049/iet-csr.2020.0013.
[5] Bona Li, “A Review of ORB-SLAM Algorithm,” SEA, vol. 12, no. 01, pp. 78–86, 2023, doi: 10.12677/SEA.2023.121009.

[6] C. Harris and M. Stephens, “A Combined Corner and Edge Detector,” in Procedings of the Alvey Vision Conference 1988, Manchester: Alvey Vision Club, 1988, p. 23.1-23.6. doi: 10.5244/C.2.23.

[7] L. Yang and K. Etsuko, “Review on vision-based tracking in surgical navigation,” IET cyber-systems robotics, vol. 2, no. 3, pp. 107–121, Sep. 2020, doi: 10.1049/iet-csr.2020.0013.

[8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in 2011 International Conference on Computer Vision, Barcelona, Spain: IEEE, Nov. 2011, pp. 2564–2571. doi: 10.1109/ICCV.2011.6126544.

[9] Wei Tan, Haomin Liu, Z. Dong, G. Zhang, and H. Bao, “Robust monocular SLAM in dynamic environments,” in 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, Australia: IEEE, Oct. 2013, pp. 209–218. doi: 10.1109/ISMAR.2013.6671781.

[10] T. Qin, P. Li, and S. Shen, “VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator,” IEEE Trans. Robot., vol. 34, no. 4, pp. 1004–1020, Aug. 2018, doi: 10.1109/TRO.2018.2853729.

[11] Y. Liu and J. Miura, “RDS-SLAM: Real-Time Dynamic SLAM Using Semantic Segmentation M -ethods,” IEEE Access, vol. 9, pp. 23772–23785, 2021, doi: 10.1109/ACCESS.2021.305061-7.

[12] Xue R, Li Y, Li M, et al. Characterization study of ORB-SLAM system [J]. High Tech Letters,2022,32(11):1112-1125.

[13] D. DeTone, T. Malisiewicz, and A. Rabinovich, “SuperPoint: Self-Supervised Interest Point Detection and Description,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA: IEEE, Jun. 2018, pp. 337–33712. doi: 10.1109/CVPRW.2018.00060.

[14] Weishan Chen, “3D reconstruction of monocular laparoscopic images based on visual SLAM,” Wuhan University of Technology, 2021, doi:10.27381/d.cnki.gwlgu.2021.000315.

[15] M. A. Fischler and R. C. Bolles, “Random sample consensus,” vol. 24, no. 6, 1981.