Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/35/20230364

Surgical robot navigation based on SLAM technology

Wenxin Lai

Department of Robot Engineering, Harbin Institute of Technology(Weihai), Weihai, China

2201350202@stu.hit.edu.cn

Abstract. With the widespread application of surgical robots and the development of computer vision, SLAM-applicated surgery is receiving increasing attention. However, due to the unique surgical environment, SLAM faces some challenges. Two key issues will be discussed in this article: dynamic object detection and image segmentation, as well as scene reconstruction under data scarcity. Firstly, dynamic object detection and image segmentation is an important issue in SLAM applications. During the surgical process, doctors often use surgical instruments, which may partially or completely obscure the object, making it difficult to detect the target. Methods based on traditional feature matching may not be able to accurately detect dynamic targets perform image segmentation. Therefore, this article will combine semantic networks for analysis to improve the performance. In addition, scene reconstruction under data scarcity is another challenge in SLAM applications. Traditional SLAM methods typically rely on a large amount of feature points or map data. But in surgery, due to the complexity of occlusion and geometric structure, reliable data may not be easily obtained. This article will develop with the steps of reconstruction and analyze feasible methods that can improve the accuracy and stability of reconstruction. To conclude, this article will concentrate on these two issues, analyze recent papers, and ultimately summarize some feasible solutions, providing ideas and references for other researchers in this field.

Keywords: SLAM, image segmentation, data scarcity.

1. Introduction

With the growing usage rate of laparoscopic and robotic procedures in surgery, more attention has been gained and incredible advances have been achieved in this area [1].

Nowadays machine vision is integrated with surgical robots to expand real-time vision [2]. The use of three-dimensional reconstruction technology can accurately depict the surgical scene to assist doctors to better perform surgery. Depth information can be used to directly conduct modelling. However, the ways to obtain the depth information (such as TOF, Structured light) is easily influenced by the scene [3]. Method based on SFM (Structure from Motion) applying the principle of multi view geometry can deduce 3D information from a group of 2D images. But average-blow performance might exist and offline processing is needed when facing low-resolution images under the lumen circumstances [4]. As for using a visual odometer, it relies on the clear visual textures and is not suitable for intracavitary tissue.

In addition, another method called SLAM is a technique for constructing environmental maps and estimating the state of robot movement [5]. The uniqueness lies in the simultaneous recursive

DOI: 10.54254/2755-2721/35/20230364

completion of the mapping and positioning [6]. It can provide timely feedback to surgeons and is widely used in surgery. Figure 1 shows the various applications of SLAM in surgical procedures.

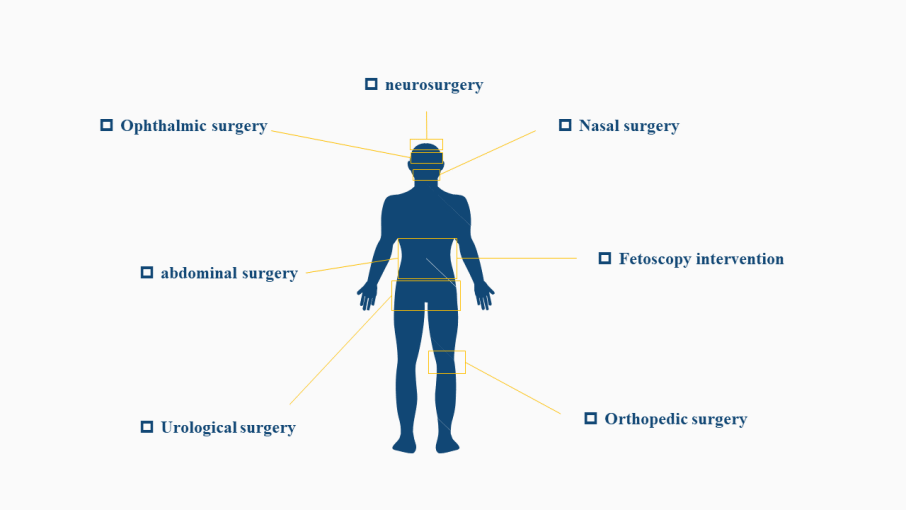


Figure 1. Various applications of SLAM in surgical procedures.

However, some challenges remain in this field. Moving objects, such as surgical instruments and people in the operating room, will lead to a bad performance and declining robustness of SLAM system. Therefore, how to strengthen the detection and segmentation of dynamic targets is a topic worth discussing. Another challenge lies in the data scarcity due to the uniform and textureless surface. The lack of required data will have a fatal impact on the application of SLAM. Thus how to carry out iterative reconstruction with data lost is another topic that should be explored further. This paper will focus on these two issues, and summarize the solutions in recent years.

The rest of this article is organized as follows: Section 2 will analyse the solutions to dynamic object detection and image segmentation problems. The third part will discuss solutions to the problem of data scarcity. Comparative analysis will be proposed in the fourth section. Finally, the conclusion is drawn in Section 5.

2. Mainbody

2.1. Dynamic object detection and image segmentation

In recent years, some feasible solutions to this problem have been proposed. Figure 2 shows the research method of this article. Four schemes and the impact of the order of segmentation and reconstruction on SLAM application will be discussed in this section.

DOI: 10.54254/2755-2721/35/20230364

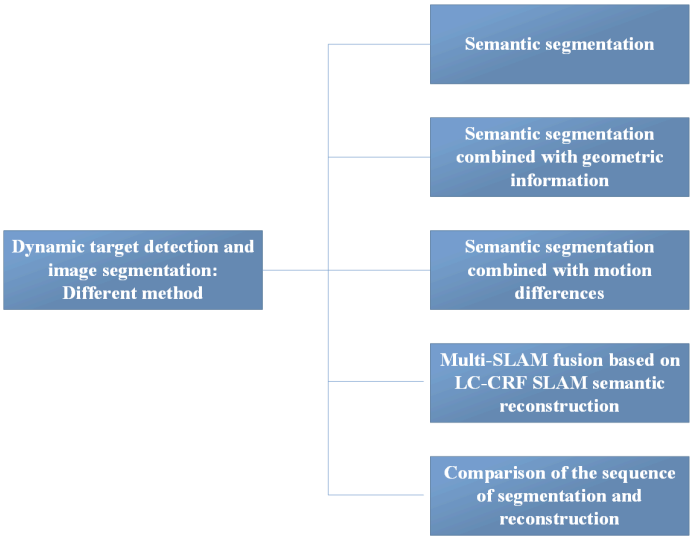


Figure 2. Research method for dynamic objection detection and image segmentation.

This work proposes a U-Net convolutional neural network based on symmetric encoder-decoder structure [4]. They use a pre-trained VGG16 encoder in a semantic segmentation network to obtain binary masks of surgical instruments. It weakens the impact of dynamic feature points brought by surgical instruments on iterative reconstruction.

Some researcher uses a semantic perception framework combined with semantic segmentation and geometric information [7]. The core idea is to use a surfel described by five parameters to track the intracavitary tissue. Parameters are updated based on the observation depth, normal, segmentation map and related evaluation functions after initialization.

This work combines semantic segmentation networks with motion differences [8]. Figure 3 shows the entire process of the model. For active moving objects (eg.people), similarity test is performed between the $mask_i$ clustered from deep image and $mask_c$ extracted by MaskRCNN algorithm. Finally the mask of the active moving object is determined by the similarity. The passive moving object will enter the state recognition module. And the state will be used to identify the subordinate optical flows that help segment passive moving objects in the current frame.

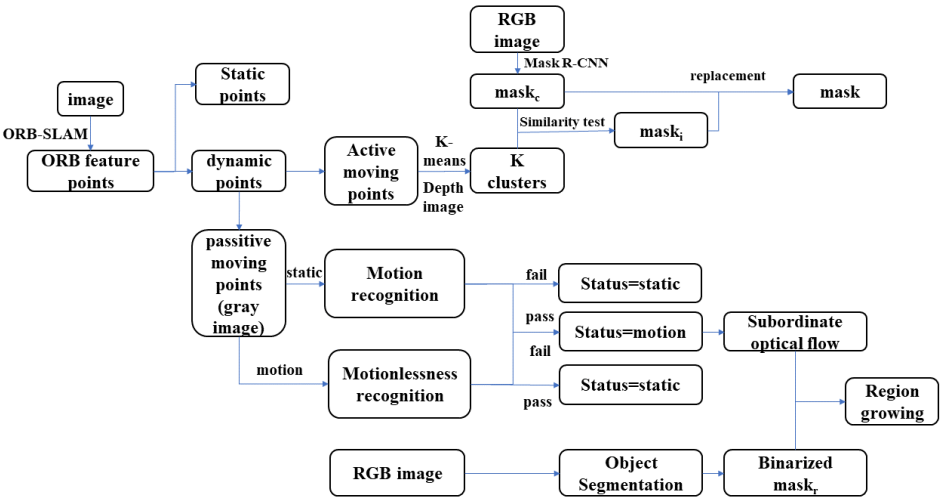


Figure 3. Flow chart of motion-difference-combined RGBD-SLAM model [8].

DOI: 10.54254/2755-2721/35/20230364

This study proposes a multi-SLAM fusion technology based on LC-CFR-improved semantic reconstruction, fully utilizing information from various spatial perspectives [9]. The mask is formed from the infrared information and optimized by morphological optimization. They use the mask to filter the established point cloud and update the point cloud with the filtered and processed ones. Then, the point clouds from two cameras are fused to update the global point cloud. Eventually the reprojection of global point will be used to optimize segmentation.

In addition, the two strategies of segmentation before Iterative reconstruction (strategy 1) and reconstruction before segmentation (strategy 2) are compared in [10].

2.2. Reconstruction under data scarcity

Due to the irregular target shape, uniform and textureless surface, and tissue deformation, the effective feature points required for reconstruction are insufficient. Figure 4 shows the steps for the iterative reconstruction and to solve the problem, the first three steps will be chosen for analysis.

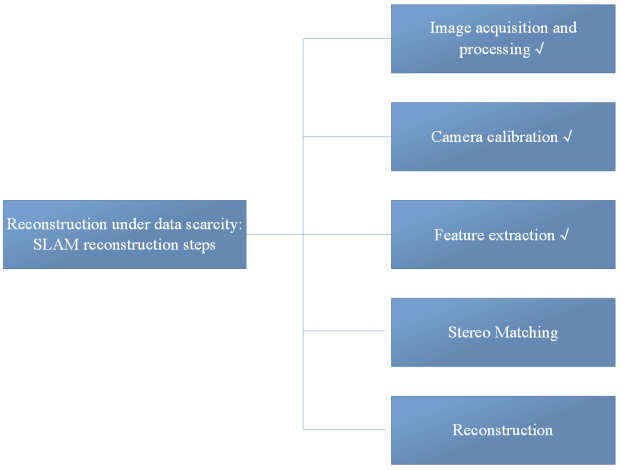


Figure 4. Research method for reconstruction under data scarcity.

A learning-based single frame stereo depth estimation method has been proposed in to enhance the robustness to textureless surfaces [11]. They add a spatial pyramid pooling and a 3D convolution-based feature decoder into the HSM network. And SERV-CT tissue data, SCARED data are utilized to supervised fine-tune the expert model while KITTI dataset is used to pretrain the model. Finally unsupervised methods is introduced to build the more precise depth estimation network.

This work introduces two adaptive weights to measure the contribution of 2D and 3D residuals and forms an evaluation function [12]. The camera pose is estimated by the function and the depth, optical flow information. Among them, adaptive weights are learned through back-propagation of Deep Declaration Networks (DDN). The whole algorithmic logic is shown in Figure 5.

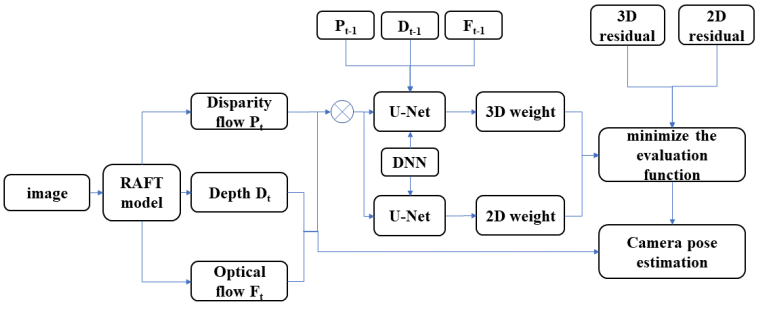


Figure 5. Flow chart of adaptive pixel weights model [12].

DOI: 10.54254/2755-2721/35/20230364

This study first adjust the ORB-SLAM parameters to make them suitable for feature extraction in the oral environment and then project a laser to artificially generate more features [13].

3. Discussion

3.1. Dynamic object detection and image segmentation

3.1.1. Semantic segmentation. TernausNet-16 performs better than the traditional U-Net network under the quantitative evaluation of IoU and Dice coefficient. The result in Figure 8 (1) indicates that the addition of pretrained encoder in U-Net increases the accuracy of segmentation.

3.1.2. Semantic segmentation combined with geometric information. As is shown in the Figure 6, Semantic-SuPer performs better than original SuPer and NoSoftLabel-semantic-SuPer under the quantitative evaluation of reprojection error. DefSLAM and SD-DefSLAM are two SLAM algorithm that track scenes based on relatively sparse feature matching. Compared to these two algorithm mentioned above, semantic-SuPer outperforms DefSLAM and achieves performance comparable to or better than SD-DefSLAM.

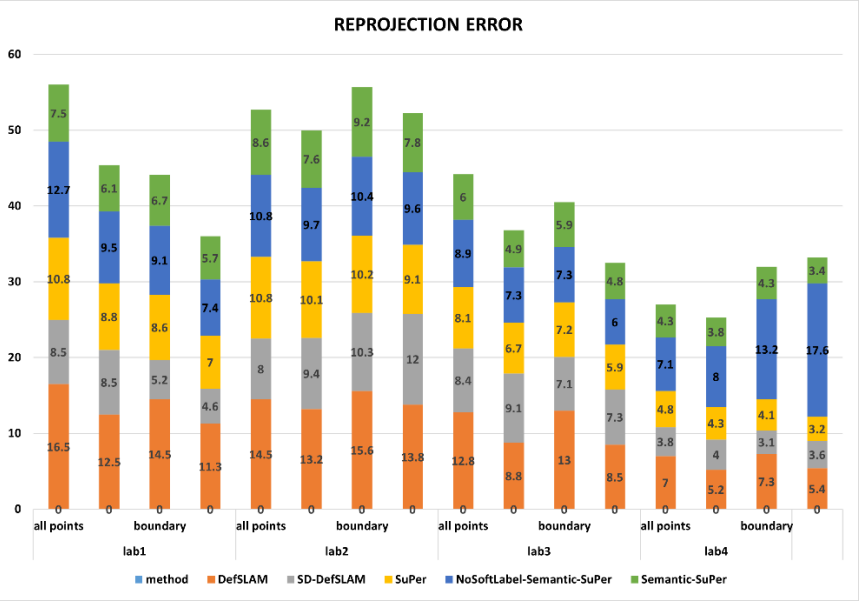
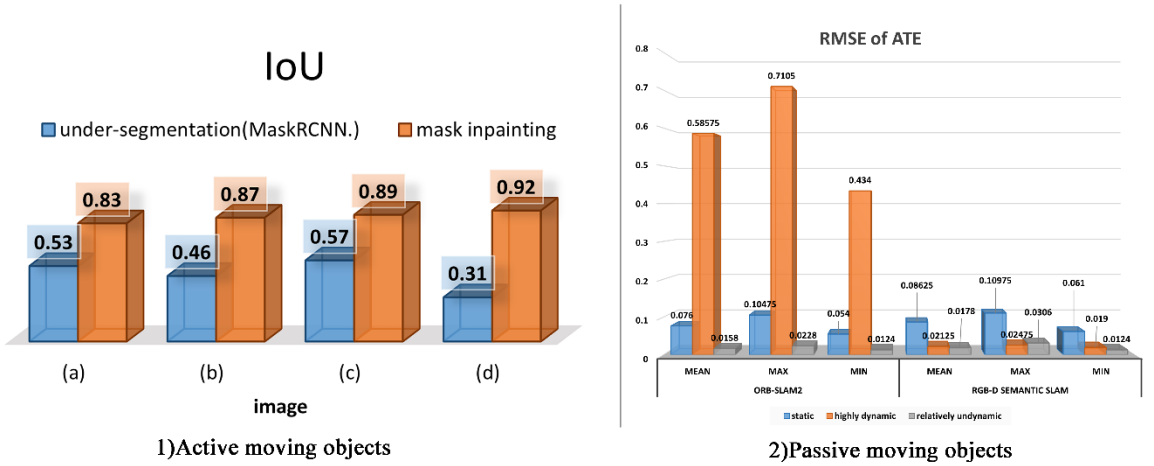


Figure 6. Quantitative comparison results of Semantic-SuPer and related models.

3.1.3. Semantic segmentation combined with motion differences. For active feature points, the segmentation effect through mask inpainting is better than that solely through MaskRCNN. It's clear that mask inpainting significantly improves the segmentation effect and solves the problem of insufficient segmentation to some extent.

Compared with ORB-SLAM2, RGB-D semantic SLAM has no superiority in static scene, but performs better in highly dynamic scenes. However, in relatively undynamic scenes, the positioning accuracy cannot be significantly improved because some static feature points around dynamic feature points are also removed. Quantitative comparison results are represented in Figure 7.

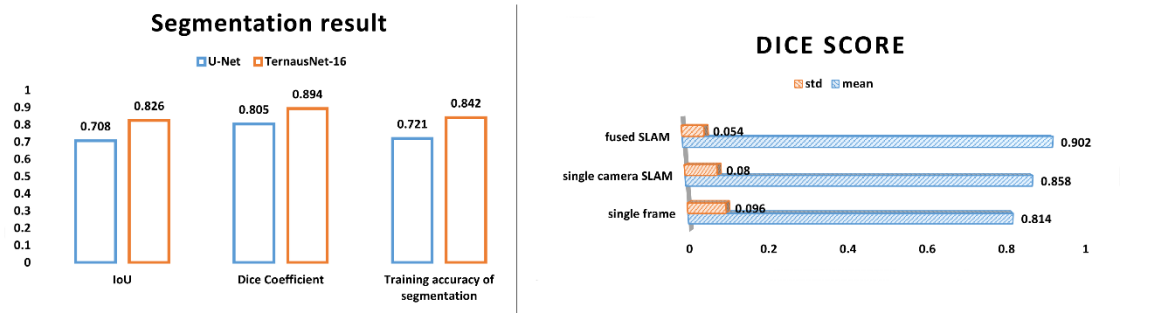
DOI: 10.54254/2755-2721/35/20230364



1)Active moving objects 2)Passive moving objects

Figure 7. Results of motion-combined semantic segmentation model.

3.1.4. Multi-SLAM fusion based on LC-CRF SLAM semantic reconstruction. Different from Single frame SLAM with no historical information and single camera SLAM with limited capture range, the fused SLAM with different perspectives information is superior according to the (2) in Figure 8. It enhances consistency and robustness for segmentation, and also improves the efficiency of obtaining spatial perception.



1)Semantic segmentation 2)multi-SLAM fusion based on LC-CRF SLAM semantic reconstruction

Figure 8. Results of semantic segmentation and multi-SLAM fusion.

3.1.5. Comparison of the sequence of segmentation and reconstruction. From the analysis of the results, strategy 1 saves time and shows higher accuracy with camera pose known, and strategy 2 is more robust and reliable under the condition of unknown camera pose. The reason is that in strategy 1 the reconstruction on the local area and feature enhancement both contribute to the time and accuracy while under unknown camera pose conditions, the local reconstruction of strategy 1 may easily lead to fatal errors.

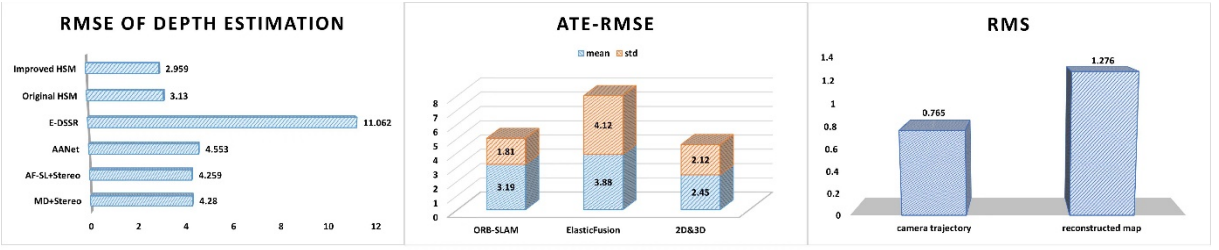
3.2. Reconstruction under data scarcity

3.2.1. Image acquisition and processing. When under the data scarcity, the ground truth depth is always insufficient. However, it can be easily deduced from Figure 9(1) that the optimized HSM net can process high-resolution RGB images, generate more disparity candidate values, and estimate depth information more accurately. That's what truly makes improved HSM perform better than others in depth estimation.

DOI: 10.54254/2755-2721/35/20230364

3.2.2. Camera calibration. The adaptive weights can be adjusted based on the contribution of different regions of the input image. It utilizes the complementary feature of 2D and 3D residuals to enhance the robustness of camera pose estimation and adapt to scenes with scarce data. Thus the 2D&3D model has superior performance compared to ORB-SLAM and ElasticFusion model and the result is shown in Figure 9(2).

3.2.3. Feature extraction. The use of lasers can accelerate initialization efficiency as well as generate more feature points. And it also makes feature points evenly distributed on areas with sufficient and insufficient features. From the perspective of camera trajectory and map reconstruction in Figure 9(3), the RMS deviation of laser model is small.



1) Image acquisition and processing 2) Camera calibration 3) Feature extraction

Figure 9. Quantitative comparison results of solutions to reconstruction under data scarcity.

3.3. limitations

When the processed data amount is large, all four methods of image segmentation face real-time requirements. Due to the presence of RGB-D sensors, performance of [8] and [9] vary under different lighting conditions and scene structures

Due to the existence of training network, including the deep learning network, the first two methods ([4] and [7]) of image segmentation and all three solutions to data scarcity lack the ability to generalize in different scenarios.

4. Conclusion

This paper focuses on the problem of dynamic targets' segmentation and reconstruction under data scarcity in the navigation of surgical robots using SLAM. It reviews relevant papers in recent years and obtains some feasible methods through comparative analysis, providing ideas and references for other researchers in this field. In response to the issue of segmentation, of dynamic targets, several methods have been studied, including semantic segmentation, semantic segmentation combined with geometric information, semantic segmentation combined with motion differences, and multi-SLAM fusion based on LC-CRF SLAM semantic reconstruction. These methods can improve the accuracy and effectiveness of image segmentation, thereby achieving more accurate object boundaries and semantic segmentation results in the field of computer vision. In response to the problem of data scarcity, research has been conducted from several aspects, including image acquisition and processing, camera calibration, and feature point extraction. The analysis helps overcome the challenges posed by insufficient data and provide effective solutions for other researchers. In the future, how to establish universal and more robust methods to solve image segmentation and data scarcity is a topic that can be paid attention to.

References

[1] Peters B S, Armijo P R and Krause C 2018 Review of emerging surgical robotic technology. *Surg Endosc* 32, 1636–1655.

[2] Afifi A, Takada C, Yoshimura Y and Nakaguchi T 2021 Real-time expanded field-of-view for minimally invasive surgery using multi-camera visual simultaneous localization and mapping. *Sensors*, 21(6), 2106.

DOI: 10.54254/2755-2721/35/20230364

[3] Wang Z. 2020 Review of real-time three-dimensional shape measurement techniques. *Measurement*, 156, 107624.
[4] Wu H, Zhao J, Xu K, Zhang Y, Xu R, Wang A and Iwahori Y 2022 Semantic SLAM based on deep learning in endocavity environment. *Symmetry*, 14(3), 614.
[5] Singandhupe A and La H M 2019 A review of SLAM techniques and security in autonomous driving. 3rd IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 602-607.
[6] Khairuddin A R, Talib M S and Haron H 2015 Review on simultaneous localization and mapping (SLAM). IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 85-90.
[7] Lin S 2023 Semantic-SuPer: a semantic-aware surgical perception framework for endoscopic tissue identification, reconstruction, and tracking. IEEE International Conference on Robotics and Automation (ICRA), London, United Kingdom, 4739-4746.
[8] Xie W, Liu P X and Zheng M 2021 Moving object segmentation and detection for robust RGBD-SLAM in dynamic environments. IEEE Transactions on Instrumentation and Measurement, 70, 1-8.
[9] Gao C, Rabindran D and Mohareri O 2022 RGB-D semantic SLAM for surgical robot navigation in the operating room. Cornel University.
[10] Su Y -H, Huang I, Huang K and Hannaford B 2018 Comparison of 3D surgical tool segmentation procedures with robot kinematics prior. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 4411-4418.
[11] Wei R 2023 Stereo dense scene reconstruction and accurate localization for learning-based navigation of laparoscope in minimally invasive surgery. IEEE Transactions on Biomedical Engineering, 70, 488-500.
[12] Hayoz M, Hahne C and Gallardo M 2023 Learning how to robustly estimate camera pose in endoscopic videos. Int J CARS 18, 1185–1192.
[13] Qiu L and Ren H 2018 Endoscope navigation and 3D reconstruction of oral cavity by visual SLAM with mitigated data scarcity. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2278-22787.