Proceedings of CONF-MLA 2024 Workshop: Mastering the Art of GANs: Unleashing Creativity with Generative Adversarial Networks
DOI: 10.54254/2755-2721/111/2024CH0111

Application of Improved Visual SLAM in Intraoperative Navigation of Robotic-Assisted Laparoscopic Surgery

Ziying Lin

School of Mechanical Engineering, University of Manchester, England, UK

ziying,lin-2@student.manchester.ac.uk

Abstract. In robot-assisted laparoscopic surgery, precise intraoperative navigation is essential to enhance the accuracy, safety, and efficiency of the procedure. By extracting feature points from the environment, Visual Simultaneous Localization and Mapping (V-SLAM) technology has been widely employed to accomplish self-localization and map generation. In laparoscopic surgery, V-SLAM offers high-precision real-time positional data for surgical robots, thereby improving the accuracy of operative navigation. Nonetheless, the performance of conventional V-SLAM algorithms is significantly compromised by the extremely dynamic characteristics of surgical environments. Widely used visual SLAM systems, such as ORB-SLAM, are susceptible to inaccurate feature point tracking in the presence of dynamic objects and variations in lighting, hence compromising navigation precision and surgical safety. This study assessed the efficacy of visual V-SLAM technology in enhancing intraoperative navigation during robot-assisted laparoscopic surgery, pointing out its contribution to surgical precision, safety, and operational efficiency. The research enhanced the ORB-SLAM2 algorithm to adapt to the dynamically changing conditions encountered during surgical procedures. The experiment validated the efficacy of the enhanced method within a simulated laparoscopic surgery setting and assessed the impact of three variables: light intensity, instrument occlusion, and dynamic tissue deformation, on trajectory error (ATE and RPE) and mapping accuracy. The findings indicated that under situations of low illumination, significant occlusion, and substantial deformation, ATE and RPE mistakes escalated by over 50%, while map accuracy diminished by approximately 40%, leading to heightened potential surgical risks.

Keywords: V-SLAM, robotic-assisted surgery, laparoscopic surgery.

1. Introduction

Robotic-assisted surgery (RAS) has transformed into minimally invasive surgery (MIS), providing substantial benefits in precision, patient recovery, and overall surgical results. The utilization of robotic systems facilitates improved dexterity and control, which is especially crucial in intricate and sensitive processes. Using RAS, surgeons may execute complex procedures with enhanced precision, markedly diminishing the likelihood of postoperative problems and expediting patient recovery time [1]. Nonetheless, the dynamic characteristics of human tissues and the limited operational space during surgery present significant hurdles for accessing and manipulating delicate anatomical areas. The problems of tissue deformation, tool occlusions, and illumination fluctuation can undermine surgical precision and, in certain instances, jeopardize patient safety [2].

To surmount these problems, more sophisticated technologies are requisite. A viable option is the application of Visual Simultaneous Localization and Mapping (V-SLAM), which facilitates real-time 3D reconstruction and localization through algorithms and sensor data [3]. V-SLAM can improve surgical navigation by precisely mapping the operative environment, so enabling surgeons to navigate through the complex and dynamic anatomy of the human body with increased accuracy and safety [4]. This technique may mitigate the danger of inadvertent harm to critical structures, augment intraoperative decision-making, and enhance overall surgical results.

The primary aim of this study is to assess the effectiveness of V-SLAM in delivering precise, real-time navigation during robotic-assisted laparoscopic surgery. This work seeks to evaluate the adaptability and performance of V-SLAM in a controlled laparoscopic environment by modelling prevalent obstacles, including restricted peripheral vision, inadequate illumination, and dynamic tissue movement. The study will entail the alteration of environmental variables, the collection of pertinent performance data, the analysis of the effects of various surgical situations on V-SLAM performance, and the identification of critical parameters that affect its accuracy and reliability. The findings are anticipated to yield significant insights into the capacity of V-SLAM to revolutionize robot-assisted surgery, presenting a novel viewpoint on improving surgical navigation and accuracy in minimally invasive techniques.

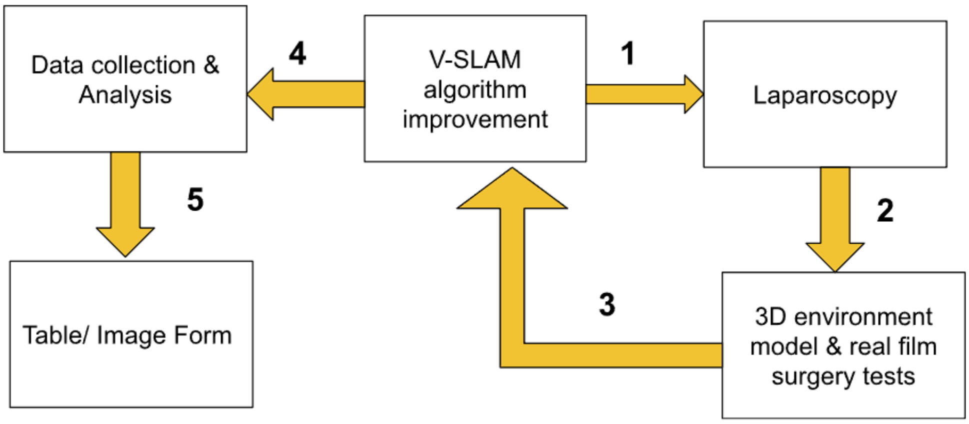

Figure 1. Method Description Flow.

2. Method

Figure 1 shows the process of the overall method. First, the tracking module of the open-source ORB-SLAM algorithm is improved. Then, the improved algorithm is used to simulate the surgical environment in 3D and real-time surgical practice videos, and the environmental variables (such as light intensity, instrument occlusion, dynamic tissue deformation, etc.) are changed to simulate the laparoscope to obtain test data. Finally, the obtained data is analyzed in charts to obtain the conclusion.

This research has improved the "Tracking:: Track() "function in ORB-SLAM, especially for use in dynamic environments. The key is to enhance the system's ability to handle dynamic objects, adapt to lighting changes, and reduce tracking errors.

2.1. Dynamic Object Detection and Removal

In laparoscopic surgery, moving organs and surgical tools are common dynamic elements. If these dynamic objects are included in the feature tracking process, they can lead to errors. So, this research uses optical flow function to calculate the movement of feature points between frames. If certain feature points move beyond a set threshold, they can be considered dynamic objects, such as surgery tools, and removed from tracking.

DOI: 10.54254/2755-2721/111/2024CH0111

2.2. Robust Feature Tracking

In tracking function, replacing traditional ORB feature extraction with deep learning-based feature extraction, convolutional neural networks, which are more robust to lighting changes and blur. The “ConvertToTrackablePoints” function maps the deep learning model's output features to feature points in the image, used to extract features in the surgical scene, potentially offering better robustness to lighting and texture variations.

2.3. Adaptive *Keyframe* Insertion

In laparoscopic surgery, camera angles can change rapidly. Improve the feature tracking part by dynamically adjusting the keyframe insertion rate when rapid viewpoint changes are detected.

2.4. Lighting-Invariant Feature Descriptors

introducing lighting-invariant feature descriptors as SURF or deep learning-based descriptors to improve robustness to lighting changes and noise, making it more suitable for surgical environments.

2.5. Adaptive Feature Extraction Threshold

The complexity of surgical environments can cause significant variation in the number of feature points in different regions. Dynamically adjust the ORB feature extraction threshold to ensure enough feature points in low-texture areas.

3. Experimental evaluation

3.1. System Setup

3.1.1. Camera

A V-SLAM system employs one or more cameras to collect visual data from the surgical environment. These cameras can be immediately integrated into endoscopic instruments or the surgical robot system. Typically, operating requirements dictate the choice of monocular, stereo, or RGB-D cameras. The accuracy of depth data and the overall reliability of the SLAM process are influenced by the choice of camera that the system employs. The predominant camera utilized in laparoscopic surgery is a monocular camera, so this work employed monocular camera technology [5].

3.1.2. Algorithmic processing

Robot-assisted surgery offers a wide range of VSLAM algorithms, typically selected based on the intricacy of the surgical environment and the specifications for real-time performance and precision [6]. The ORB-SLAM2 method is well recognized as one of the most prominent VSLAM algorithms [7]. The compatibility with various input camera requirements and the features of the ORB feature extractor have successfully addressed the challenge of limited feature points in the laparoscopic surgical setting [8]. Furthermore, this approach exhibits robust real-time performance and offers smooth integration of custom needs. It is the predominant algorithm used in laparoscopy. Therefore, this study sets ORB-SLAM2 as the major algorithm to test.

3.2. Performance Metrics

3.2.1. Accuracy

The key performance measure for V-SLAM in RALS is the precision of the system in locating the surgical tools and mapping the surroundings [9]. A common method of evaluating accuracy is by comparing the positions generated by V-SLAM with the ground truth data acquired from more precise tracking systems or fiducial markers. Achieving sub-millimeter accuracy is a prerequisite for V-SLAM systems to be deemed suitable for surgical applications in practice. Elements such as the resolution of the camera, the effectiveness of the algorithm, and the quality of the data fusion process are crucial in establishing the global accuracy of the system [10].

DOI: 10.54254/2755-2721/111/2024CH0111

3.2.2. Robustness
Robustness pertains to the capacity of the system to sustain precise localization and mapping amid diverse demanding circumstances, including swift instrument motions, occlusions, variations in illumination, and the existence of highly reflecting surfaces [11]. A common method for assessing the resilience of V-SLAM systems is to subject them to stress testing in simulated surgical settings that accurately reproduce these difficulties [12]. In order to maintain dependable feedback to the surgeon, a V-SLAM system must possess the capability to promptly recover from transient visual data losses.

3.2.3. Computational efficiency and real-time capability
To ensure the effectiveness of V-SLAM in RALS, it is imperative that it functions in real time, offering prompt feedback to both the robotic system and the surgeon. The system's real-time capabilities are commonly assessed by computing latency (the time delay between data acquisition and feedback) and processing speed (the system's capacity to process frames per second). Fast processing rates and low latency are crucial for enabling the V-SLAM system to effectively handle the dynamic nature of surgical operations [13].

3.2.4. Adapting to dynamic environments
For robotics systems, the capacity of a SLAM algorithm to manage dynamic objects in the environment is a crucial factor to consider. This metric evaluates the algorithm's capacity to identify, monitor, and integrate dynamic objects and moving barriers into the mapping procedure [14]. This study centers on the algorithm's capacity to facilitate the robot in efficiently managing these objects and promptly reacting throughout the continuous SLAM procedure. An algorithm's capacity to adapt and react in real-time applications should be guaranteed by a dynamic architecture. Continuous adaptation is essential for systems functioning in dynamic contexts characterized by rapid changes, such as interactive robotics applications.

3.2.5. Impact on surgical outcomes
For robotics systems, the capacity of a SLAM algorithm to manage dynamic objects in the environment is a crucial factor to consider. This metric evaluates the algorithm's capacity to identify, monitor, and integrate dynamic objects and moving barriers into the mapping procedure [15]. This study centers on the algorithm's capacity to facilitate the robot in efficiently managing these objects and promptly reacting throughout the continuous SLAM procedure. An algorithm's capacity to adapt and react in real-time applications should be guaranteed by a dynamic architecture. Continuous adaptation is essential for systems functioning in dynamic contexts characterized by rapid changes, such as interactive robotics applications.

3.3. Assumptions

3.3.1. D environment and model creation
In this work, pre-existing components from SOFA are created to simulate a laparoscopic surgery environment. The camera used to mimic the movement of a monocular camera, which captures the real-time image data from the simulated scene. A non-rigid viscoelastic deformation model is established in the dynamic environment of the abdominal cavity.

3.3.2. V-SLAM implementation
The visual data obtained from the SOFA simulation are inputted into the ORB-SLAM2 software to analyze and produce both the trajectory of the camera and a three-dimensional representation of the surrounding environment.

3.3.3. *Surgical scenario simulation*

This experiment uses the developed model to replicate standard laparoscopic surgery. The three primary assessment criteria for this investigation are the illumination variables in the actual surgical environment, the lens occlusion variables during surgery, and the real-time changes in dynamic tissues [16].

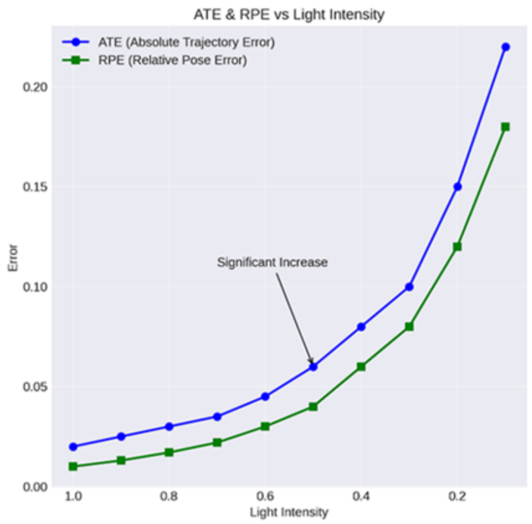

Figure 2. Comparison between light intensity and ATE & PRE.

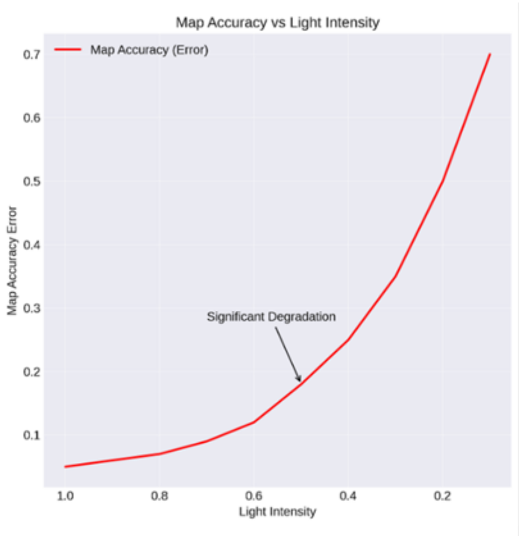

Figure 3. Comparison between light intensity and map accuracy.

4. Results

The evaluation criteria for the SOFA simulation model and ORB-SLAM2 algorithm examined in this work are absolute trajectory error (ATE), relative posture error (RPE), and map correctness. The accuracy of the camera's calculated trajectory for known ground parameters is quantified by ATE and RPE. Furthermore, map accuracy assesses the precision of the map produced by computing the discrepancy between the reconstructed points and the established locations in the generated environment.

4.1. *Varying degrees of lighting*

The dominant determinant of the algorithm's utility in the RAS is the intensity of light [17]. In this study, the light intensity of the surgical environment is methodically reduced at 10 different levels. At each level, ORB-SLAM2 analyses the acquired images and records the resulting Average Travel Time (ATE), Remote Point Estimation (RPE), and map accuracy [17].

The comparisons between the recorded trajectory errors (ATE and RPE) and map accuracy shown in figure 2 and figure 3 demonstrate that as the light intensity decreases, both ATE and RPE increase. This indicates a notable rise in error when the light intensity falls below 0.5, suggesting a decrease in accuracy in trajectory planning. Furthermore, the accuracy of the map decreases when the light intensity decreases, as shown in figure 3. The surgical impact is particularly difficult in recreating the intraperitoneal environment in settings of inadequate illumination.

4.2. *Instrument occlusion*

This parameter analyses the response of the V-SLAM system when surgical instruments partially or completely block the camera's field of vision [18]. According to the provided figures 4 and 5, there is a positive correlation between the percentage of instruments occlusion and the error level of three indicators (ATE, RPE, and map accuracy). This correlation is particularly evident at around 50% occlusion.

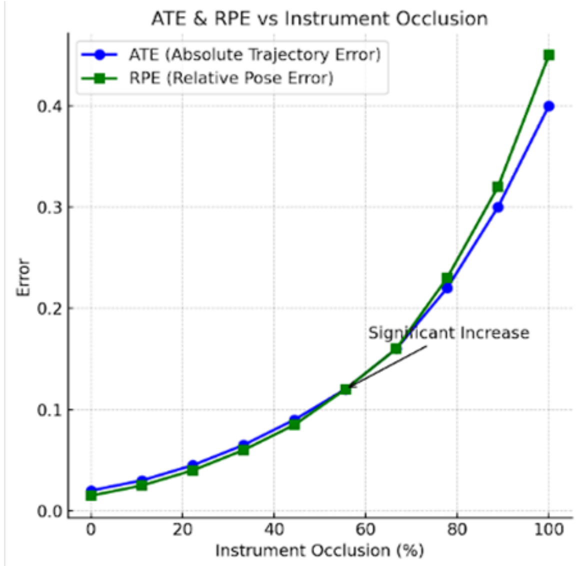

Figure 4. Comparison between instrument occlusion and ATE & PRE.

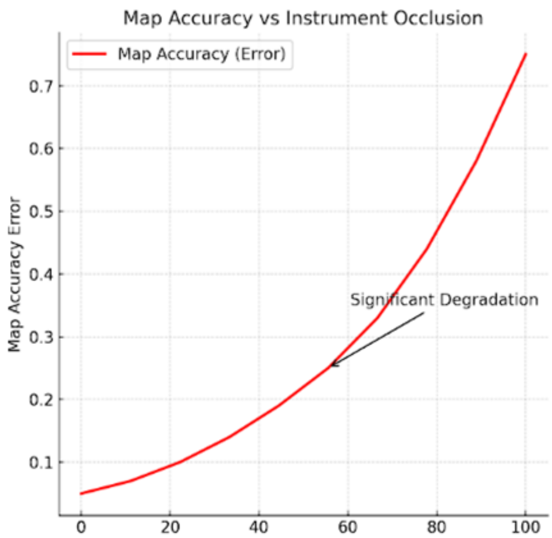

Figure 5. Comparison between instrument occlusion and map accuracy.

4.3. Dynamic tissue deformation

This work mimics the biological processes of tissue movement and deformation that occur naturally in real surgical settings [19]. This evaluates the V-SLAM system's capacity to adjust to dynamic anatomical structures, therefore assuring an accurate tracking and mapping of the surroundings even in the presence of a changing physical landscape. The dynamic tissue deformation set is a non-rigid dynamic model used in surgical environment simulation. The dynamic nature of deformation necessitates the quantification of this standard by dividing it into five levels, ranging from 0 to 5. This range represents no deformation to extremely considerable amounts.

Figure 6 presents a comparison of feature points of laparoscopic surgery in real films using the ORB-SLAM2 algorithm at various deformation levels. The results indicate that the number of important points identified in the highly deformed frame is significantly lower than the other one. Table 1 provides a visual representation of the progressive increase in results up to level 3. Such evidence suggests that the localization accuracy of ORB-SLAM2 is negatively impacted by dynamic tissue deformation, particularly when the distortion is severe.

Table 1. Comparison of deformation level changes with ATE,RPE and Map accuracy error.

Deformation level	ATE	RPE	Map accuracy error
0	0.023	0.015	0.05
1	0.047	0.032	0.09
2	0.083	0.059	0.23
3	0.159	0.126	0.46
4	0.306	0.253	0.78

DOI: 10.54254/2755-2721/111/2024CH0111

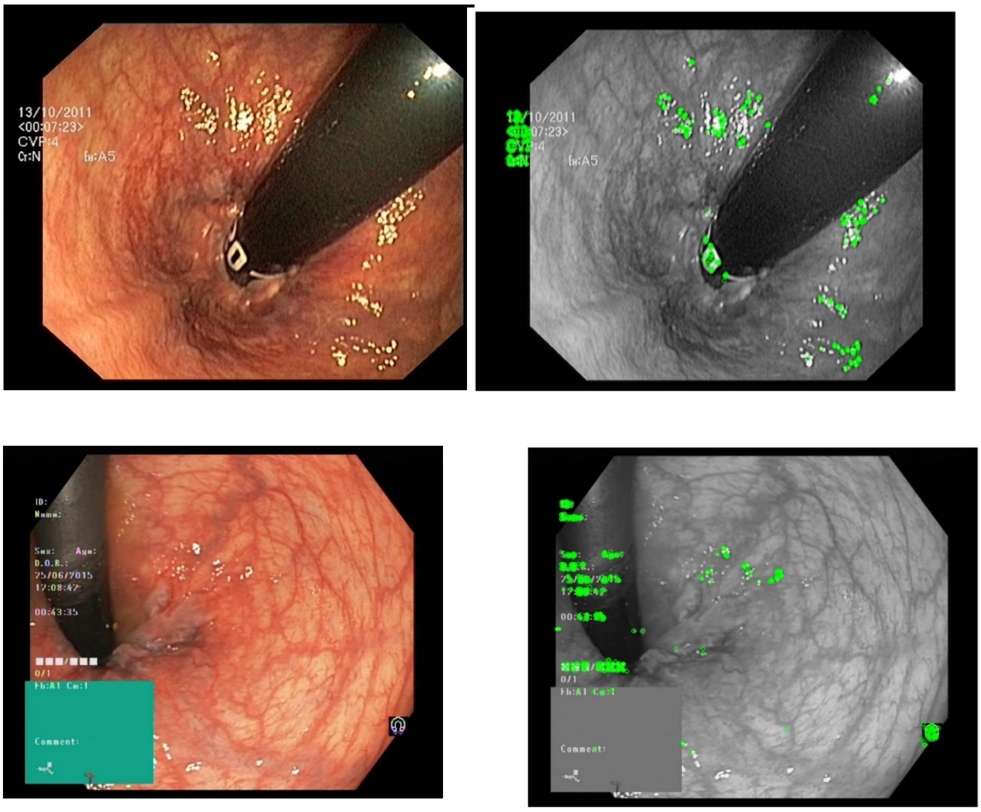

Figure 6. comparison of feature points of laparoscopic surgery in real-films using the ORB-SLAM2 algorithm at various deformation levels.

5. Discussion

Given these factors, the complex issues of dynamic tissue deformation, instrument obstruction, and fluctuating light intensity are among the most crucial and challenging problems that must be resolved [20]. In situations characterized by significant tissue deformation and inadequate lighting, the SLAM system may face even greater challenges if crucial characteristics are also obscured by surgical equipment. Cumulatively, these problems may result in substantial errors that undermine the precision of the SLAM-based navigation system, hence potentially impacting surgical results.

To tackle these difficulties, future research should investigate multi-modal SLAM methods that use supplementary data sources (such as depth sensors and ultrasonic techniques) and sophisticated algorithms that can compensate for non-rigid deformations and occlusions [11]. The DSO direct approach in the latest state-of-the-art (SOTA) SLAM algorithm exhibits superior accuracy and increased adaptability, making it particularly well-suited for low-texture environments. In comparison to the ORB algorithm, the multi-sensor fusion approach exhibits superior scalability and increased resilience. Nevertheless, the constraints of SOTA are the substantial computer resource demands and the expense associated with development and troubleshooting [2].

6. Conclusion

This work demonstrates that the performance of ORB-SLAM2 declines markedly under challenging surgical conditions, including inadequate lighting, substantial occlusion of lenses by surgical instruments, and pronounced tissue deformation. In these unfavorable conditions, the algorithm fails to sustain dependable tracking and precise mapping, leading to significant reductions in navigation accuracy. Experimental findings indicate that in low-light settings, the relative pose error (RPE)

DOI: 10.54254/2755-2721/111/2024CH0111

escalates by 50%, whilst map accuracy diminishes by approximately 40%. Excessive blockage from surgical instruments markedly diminishes the system's capacity to accurately find and map the surroundings, hence exacerbating the mistake rate. Tissue deformation, commonly encountered in laparoscopic surgery, amplifies these mistakes, resulting in erratic tracking of feature locations and markedly diminishing the robustness of the SLAM system.

These findings hold significant implications for the implementation of SLAM-based navigation in robotic-assisted surgery. Compromised accuracy of the navigation system can elevate surgical hazards, including potential harm to critical structures and diminished task execution precision. Despite ORB-SLAM2's foundation on a robust algorithm, its efficacy in intricate surgical settings is significantly lacking. This underscores the necessity for more refinements to the algorithm to augment its adaptability to dynamic and intricate elements in surgical contexts.

Subsequent research and development must persist in addressing these constraints to ensure that SLAM-based navigation systems can adeptly adapt to fluctuating ambient circumstances during surgical procedures. Enhancements may involve the implementation of sophisticated feature extraction techniques, including deep learning methodologies, and the formulation of intricate algorithms to effectively address issues such as variations in lighting, occlusions, and tissue deformation. Continuous optimization of these systems is anticipated to enhance surgical outcomes and mitigate dangers associated with robot-assisted surgery.

References

[1] Taheri, H., & Xia, Z. C. (2021). SLAM; definition and evolution. Engineering Applications of Artificial Intelligence, 97, 104032.

[2] Al-Tawil, B., Hempel, T., Abdelrahman, A., & Al-Hamadi, A. (2024). A review of visual SLAM for robotics: evolution, properties, and future applications. *Frontiers in Robotics and AI*, *11*, 1347985.

[3] Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2024). Nr-slam: Non-rigid monocular slam. *IEEE Transactions on Robotics*.

[4] Bao, G. (2014). *On Simultaneous Localization and Mapping Inside the Human Body (Body-SLAM)* (Doctoral dissertation, Worcester Polytechnic Institute).

[5] Ni, J., Wang, L., Wang, X., & Tang, G. (2023). An Improved Visual SLAM Based on Map Point Reliability under Dynamic Environments. *Applied Sciences*, *13*(4), 2712.

[6] Tourani, A., Bavle, H., Sanchez-Lopez, J. L., & Voos, H. (2022). Visual slam: What are the current trends and what to expect?. *Sensors*, *22*(23), 9297.

[7] Yang, X., Li, H., Zhai, H., Ming, Y., Liu, Y., & Zhang, G. (2022, October). Vox-fusion: Dense tracking and mapping with voxel-based neural implicit representation. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 499-507). IEEE.

[8] Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., ... & Pollefeys, M. (2022). Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12786-12796).

[9] Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2021). Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, *37*(6), 1874-1890.

[10] Joo, S. H., Manzoor, S., Rocha, Y. G., Bae, S. H., Lee, K. H., Kuc, T. Y., & Kim, M. (2020). Autonomous navigation framework for intelligent robots based on a semantic environment modeling. *Applied Sciences*, *10*(9), 3219.

[11] Sousa, R. B., Sobreira, H. M., & Moreira, A. P. (2023). A systematic literature review on long-term localization and mapping for mobile robots. *Journal of Field Robotics*, *40*(5), 1245-1322.

[12] Nakamura, T., Kobayashi, M., & Motoi, N. (2023). Path planning for mobile robot considering turnabouts on narrow road by deep Q-network. *IEEE Access*, *11*, 19111-19121.

[13] Macario Barros, A., Michel, M., Moline, Y., Corre, G., & Carrel, F. (2022). A comprehensive survey of visual slam algorithms. *Robotics*, *11*(1), 24.

DOI: 10.54254/2755-2721/111/2024CH0111

[14] Wu, W., Guo, L., Gao, H., You, Z., Liu, Y., & Chen, Z. (2022). YOLO-SLAM: A semantic SLAM system towards dynamic environment with geometric constraint. *Neural Computing and Applications*, 1-16.

[15] López, J., Sanchez-Vilarino, P., Cacho, M. D., & Guillén, E. L. (2020). Obstacle avoidance in dynamic environments based on velocity space optimization. *Robotics and Autonomous Systems*, *131*, 103569.

[16] Li, S., Zheng, P., Liu, S., Wang, Z., Wang, X. V., Zheng, L., & Wang, L. (2023). Proactive human–robot collaboration: Mutual-cognitive, predictable, and self-organising perspectives. *Robotics and Computer-Integrated Manufacturing*, *81*, 102510.

[17] Mazumdar, H., Chakraborty, C., Sathvik, M. S. V. P. J., Jayakumar, P., & Kaushik, A. (2023). Optimizing pix2pix gan with attention mechanisms for ai-driven polyp segmentation in iomt-enabled smart healthcare. *IEEE Journal of Biomedical and Health Informatics*.

[18] Zheng, S., Wang, J., Rizos, C., Ding, W., & El-Mowafy, A. (2023). Simultaneous localization and mapping (slam) for autonomous driving: concept and analysis. *Remote Sensing*, *15*(4), 1156.

[19] Altawil, B., & Can, F. C. (2023). Design and analysis of a four dof robotic arm with two grippers used in agricultural operations. *International Journal of Applied Mathematics Electronics and Computers*, *11*(2), 79-87.

[20] Kuang, Z., Wei, W., Yan, Y., Li, J., Lu, G., Peng, Y., ... & Shang, W. (2022). A real-time and robust monocular visual inertial slam system based on point and line features for mobile robots of smart cities toward 6g. *IEEE Open Journal of the Communications Society*, *3*, 1950-1962.