Proceedings of the 2023 International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/12/20230300

Endoscope 3D reconstruction based on ORB-SLAM

Baichuan Zhang

The College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China

Birchard-Z@nuaa.edu.cn

Abstract. Minimal invasive surgery (MIS) is the mainstream trend in developing surgical technology. As the endoscope is a significant tool in the surgical process, whether it can track the inner cavity and realize accurate real-time 3D reconstruction has a vital impact on the smooth progress of MIS. However, there are still problems in the endoscopic environment, such as severe image distortion, the effect of lighting conditions, and the inability to extract lumen textures. Orinted fast and rotated brief simultaneous localization and mapping (ORB-SLAM) is currently a relatively advanced simultaneous localization and mapping (SLAM) method with better performance. The ORB-SLAM-based endoscope 3D reconstruction method can improve performance and overcome the challenge of endoscope 3D reconstruction. This paper will first introduce several existing endoscope 3D reconstruction methods based on ORB-SLAM and analyze the limitations and issues of these methods through their experimental results. Then the paper will explore the solutions to the defects in this method from other methods and compare the characteristics and the result of experiments. Secondly, through the summary of the above methods and the introduction of the integration of the ORB-SLAM-based methods and other current advanced technologies, the future development trend and huge development potential of ORB-SLAM-based endoscopic 3D reconstruction are introduced. This paper will be of profound affection to further improve the optimization and application of the ORB-SLAM-based endoscope 3D reconstruction method.

Keywords: orinted fast, rotated brief simultaneous localization, mapping, endoscope, 3D reconstruction.

1. Introduction

Compared with the general robot, the equipment in MIS usually has no extra sensors and distance-measuring tools. It is harder to utilize SLAM in a Monocular handle camera like an endoscope camera than in others because of its high-velocity movement, bad stability and complex working environment. Due to the effect of circumventing external interferences such as optics and magnets, increasingly scholars have committed to the research about SLAM-Based technology in tracking, relocalization and 3D reconstruction.

In 2007, the image from the monocular endoscope was processed by Wu et al. with a proposed structure from motion (SFM) method [1]. However, this method is not adapted to real-time 3D reconstruction scenarios because it needs to handle offline data. In 2006, Mountney et al. utilized simultaneous location and mapping from the visual sensor (VSLAM) in MIS [2]. They applicated and extended the basic structure of extended kalman filter SLAM (EKF-SLAM) from Davison to MIS [3].

DOI: 10.54254/2755-2721/12/20230300

In 2009, the possibility of using EKF-SLAM in a monocular endoscope was researched by Grasa et al., who proved that the 3D reconstruction of the dense map only by the monocular endoscope is dramatically challenging under non-rigid and texture-lacking conditions [4]. In 2010, Mountney and Yang proposed learning the motion paraments of the periodical changes, like hepatic, to improve the estimation of VSLAM [5]. In 2011, Rublee et al. presented an ORB feature extraction algorithm [6]. This method exceeded scale-invariant feature transform (SHIFT) in computing velocity and robustness, which solved the problem where Binary Robust Independent Elementary Features (BRIEF) was not rotation invariant [7]. In 2013, Lin et al. applied the parallel tracking and mapping (PTAM) algorithm proposed by Klein et al. to stereoscopic endoscopy [8, 9]. PTAM is the first system based on keyframes. SLAM technology saw a breakthrough when it was initially separated into front-end and back-end concepts, paving the way developing numerous following SLAM systems. In 2015, Mur-Artal raised the ORB-SLAM system. Figure 1 organizes the ORB-SLAM system framework [10]. The system inherits the PTAM framework and uses ORB binary feature points to quickly and reliably complete positioning. The ORB-SLAM system is a proven advanced SLAM system. Combining many new technologies, such as the ORB algorithm, local keyframe mapping, establishment of covisiblility graph, word bag algorithm relocation, etc., the system can robustly track the camera and accomplish the 3D reconstruction by extracting feature points. Based on ORB-SLAM, ORB-SLAM2 proposed in 2017 added a new global optimization after the local map optimization of loopback detection [11]. ORB-SLAM3 manages a series of sub-maps, which share a bag of words, enabling operations such as relocation and loopback [12]. Figure 2 sorts out the timeline proposed by ORB-SLAM.

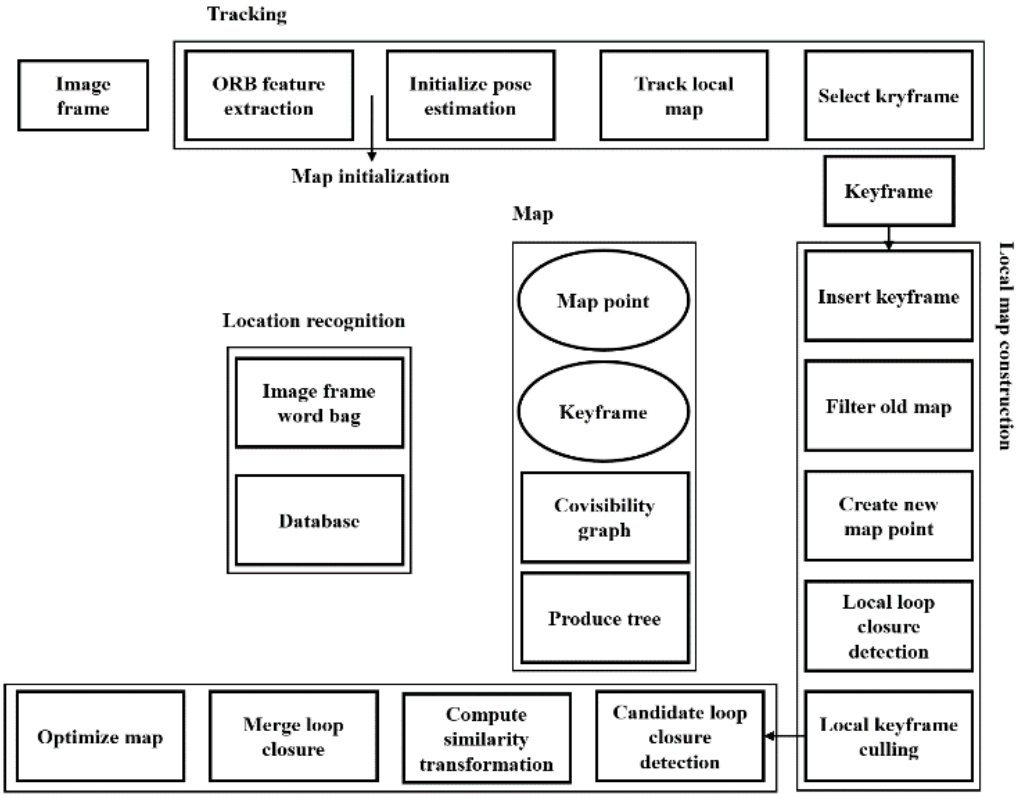

Figure 1. ORB-SLAM system overview.

DOI: 10.54254/2755-2721/12/20230300

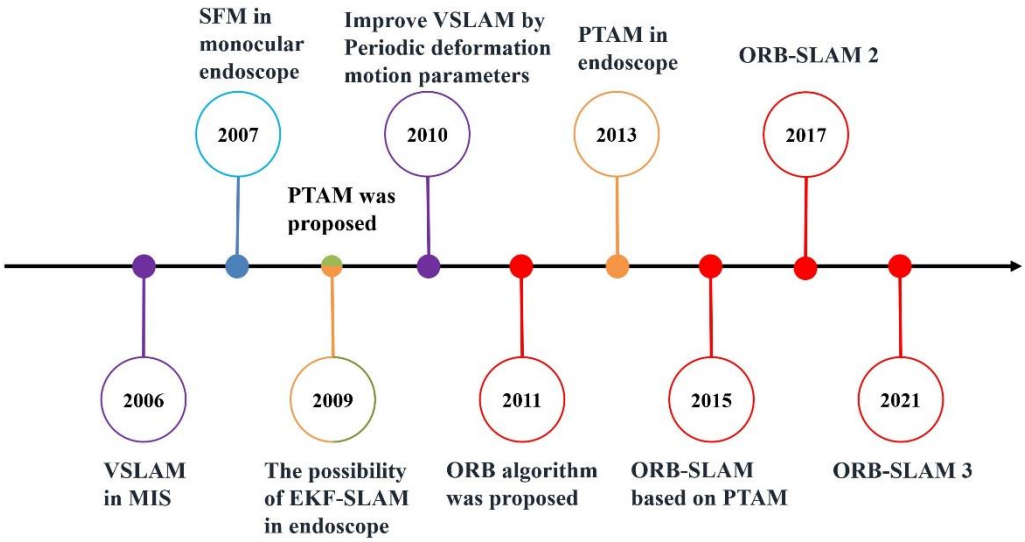

Figure 2. The timeline of ORB-SLAM (The same color means related research direction).

2. Applications of ORB-SLAM in endoscope 3D reconstruction

With the development of computer vision, algorithms based on computer vision have gradually attracted the attention of researchers because, with the assistance of such algorithms, endoscope position tracking and 3D reconstructions in three-dimensional space can be successfully performed. However, in MIS, a single hand-held camera (such as an endoscope camera) is usually used, so some of the proposed methods [1] are unsuitable for MIS.

To track the position of the endoscope in MIS scenarios and provide real-time 3D reconstruction based on the image provided by the monocular endoscope as the only input, Mahmoud et al. proposed an endoscope 3D reconstruction method based on ORB-SLAM, achieving by following methods [13]. (1) Tracking: Endoscope position is tracked in every live video frame. (2) Mapping: Keyframes are screened and matched between them; map point 3D positions and keyframe 3D poses are computed by bundle adjustment (BA). The system will initially over-initialize map points and keyframes, and only keyframes with more information are retained after the second round of strict screening. In order to keep the scale and rotation constant, the method initializes the map in detecting ORB features of different image scales (3) Endoscope relocation: For possible situations of abduction cameras, this method stores all keyframes in the Bag, which has Binary Words, index database combined with the covisibility graph. As long as valid endoscopic poses are found, the system responds quickly and resumes tracking. In addition, the method further densifies the map. The initialization process is prolonged to the second stage, where a cross-correlation guided by epipolar geometry is applied to all unmatched ORB points in the keyframe to reconstruct the image's sparse region as a semi-dense map.

The performance of ORB-SLAM in 3D reconstruction and the error of the reconstructed map is evaluated through experiments. The results show that the endoscope tracking based on ORB-SLAM is high quality, and the accuracy and robustness are also very good. Although the number of matching map points is very small, the system can still construct a sparse map, and the results of the pig lung experiment also show that the system has a certain ability to estimate motion trajectories. Even if the endoscope is removed and replaced during observation, the system can be repositioned within 3 seconds. To assess the error of the 3D reconstruction point cloud, the researchers used computed tomography (CT) as the ground standard to evaluate the system's reconstructed map. Its root mean square error (RMSE) reached 3-4.1 mm. Figure 3 shows the reconstructed map obtained after the experiment [13]. In (e-f), the diaphragm’s back-and-forth movement can be estimated because of the pig’s breathing.

DOI: 10.54254/2755-2721/12/20230300

This study has demonstrated that this tuned ORB-SLAM system is a robust method which means that it can be used in monocular endoscope tracking and 3D scene reconstruction with images acquired by the endoscope as the only input. However, simultaneously, some limitations of ORB-SLAM have also been discovered through research. Due to the following reasons, ORB-SLAM-based endoscopy cannot detect repeatable points on some soft organs, and many regions cannot track map points in some scenes, making 3D dense reconstruction challenging [14]. (1) Because of the limitation of monocular vision, the image distortion of the endoscope is serious. (2) The surface texture of the organism is limited, which means that the feature points available for the system to scan and extract are limited. (3) Endoscopes often use cold light and are highly susceptible to other uneven lighting conditions.

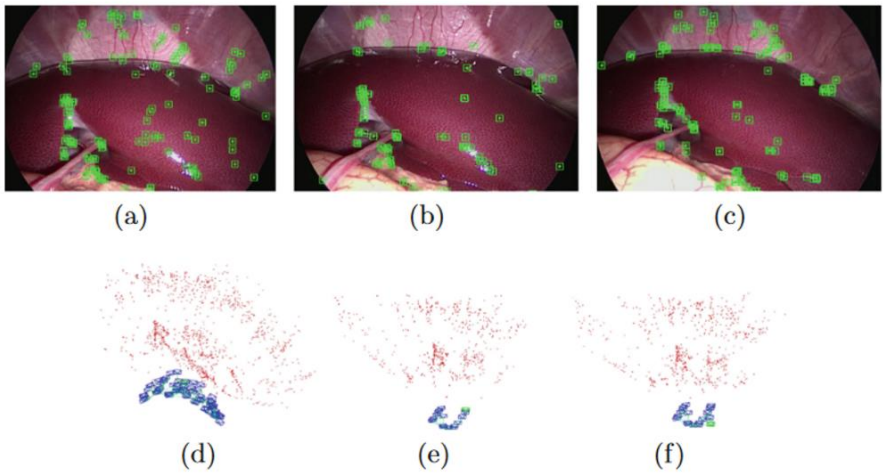

Figure 3. The performance of ORBSLAM [13].((a–c) images with reprojected map points (green points). (d) Reconstructed map points (red) and keyframes (endoscope tip trajectory). (e–f) Breathing motion, current endoscope location is shown as a green rectangle. (e) and (f) during inhale and exhale, respectively. (Color figure online))

To solve the problems in the above studies, Chen et al. proposed an improved method [14]. Matching adjacent frames using local features for endoscope pose estimation and keyframe selection. In improving the ORB-SLAM system, the system uses the PnP algorithm to match the ORB features to estimate the endoscope pose and further optimize the keyframe pose; uses triangulation and Bayesian probability methods to measure depth reliability; uses filters to reduce noise. In terms of tracking, through preprocessing, the researchers first integrated the fisheye model into the system and eliminated distortion by checkerboard calibration; used grey levels to distinguish target areas; removed features not in the grey level of the target area to obtain local features of endoscope images. Regarding dense depth map estimation, researchers are inspired by to transform the depth prediction problem of reference frames into a Bayesian probability estimation problem and reduce the depth map noise through smoothing filters [15]. Figure 4 respectively shows (a) the original image, (b) the reconstruction of binocular d and (c) the reconstruction of the improved method [14]. Experimental results on the data set collected by the liver in vitro show that this method can retain the overall effectiveness of the results and obtain a good dense point cloud effect even in dim light conditions. The researchers also conducted comparative experiments using Hamlyn in vivo videos, further verifying the feasibility and effectiveness of improved ORB-SLAM in monocular endoscopy and the superiority of reconstructed point clouds compared with the binocular method in density and smoothness.

DOI: 10.54254/2755-2721/12/20230300

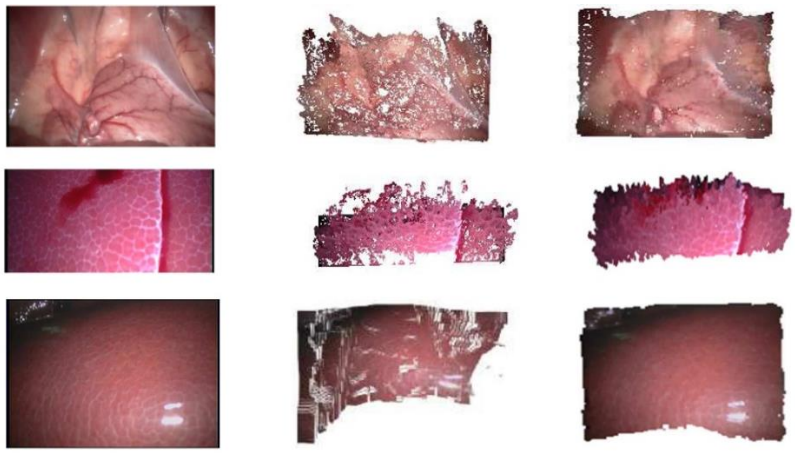

Figure 4. The reconstruction result of the Hamlyn in vivo videos [14]. ((a) Presents the original images. (b) shows the point clouds construct by a binocular method proposed in Wang's method [15]. (c) shows the results from our method.)

However, in the process of this research, due to the improvement of the algorithm, the complexity of depth estimation is high, and the number of calculations increases. Depth estimation and smoothing also require GPU acceleration and thus cannot provide real-time 3D reconstruction.

Other methods optimize the ORB-SLAM system from different perspectives. The following two methods improve the method in terms of algorithm and hardware, respectively.

Nader et al. proposed an adjustment method for the ORB-SLAM system, which computes dense matching Between parallel cluster frames by a variational method combining zero-mean normalized cross-correlation and a gradient Huber norm regularizer [16]. Experiments show that there is indeed a gain in computing time after adjustment. Huo et al. made a breakthrough in the 3D reconstruction of ORB-SLAM from a binocular endoscope [17]. The disparity map is obtained by the traditional SGBE method. Real-time matching is used to obtain the depth sequence, and Stereo Net is used to rebuild the left and right image sequences of the binocular endoscope. The left-view RGB-D camera's map is simultaneously based on the image sequences. It not only fixes the issue that conventional binocular SLAM cannot achieve real-time dense 3D reconstruction, but it also extracts the greatest amount of physical data from the image to guarantee the processing speed. The experiment in pig stomachs shows that this method improves the computational efficiency and point cloud density while basically fulfilling the real-time requirements. The RMSE is only 1.620mm lower than the one in the above methods. It can be inferred from Table 1 that the adjusted method makes the calculation method drop significantly. In the 3D reconstruction image in Figure 5, the point cloud on the binocular reconstruction image is less discrete.

Table 1. The performance of Nader's method [16].

Image Resol.	720×288	960×260
Cluster Selection (s)	0.17	0.21
BA (s)	1.3	2
Inverse Depth Discretiz (s)	0.00036	0.0039
Cost Volume (s)	3.4	5.2
Variational Minimiz. (s)	6.2	8.4
Depth maps realignment (s)	0.38	0.47

DOI: 10.54254/2755-2721/12/20230300

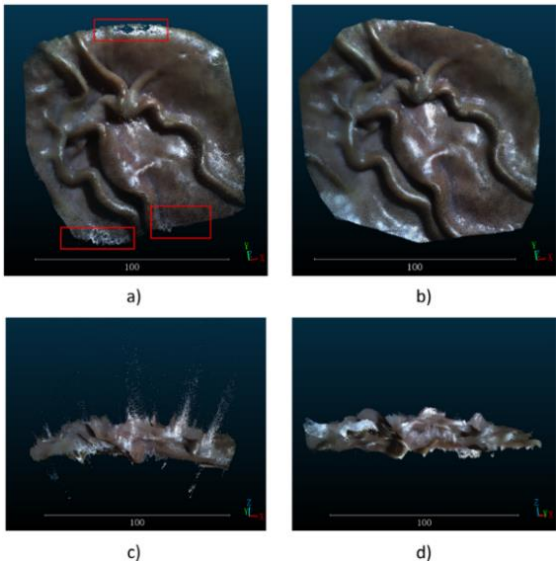

Figure 5. The result of Huo's method [17].

The relevant information about the four methods mentioned above is summarized. These methods' performance indicators and characteristics are compared, as shown in Table 2, which shows that the subsequent enhanced algorithm performs better in RMSE and non-rigid conditions, but it compromises real-time tracking and relocation capabilities and is subject to certain environmental constraints, such as light source conditions.

Table 2. Summary of the ORB-SLAM-based endoscope 3D reconstruction.

	Estimation	Relocalization	Mono	Stereo	Motion	light source conditions	non-rigid deformation	Real-time	RMSE (mm)
ORB-SLAM	BA	3s	✓	×	✓	-	×	✓	3-4.1
Improved ORB-SLAM	BA	-	✓	×	-	-	-	×	-
Dense SLAM Method	BA	-	✓	×	-	effect by shadow	small deformation	×	1.10
Binocular Method	StereoNet	-	×	✓	-	white light	×	×	1.26

3. Future works and potentials

The ORB-SLAM system is a powerful tracking and mapping tool with excellent relocalization abilities. Compared with other SLAM algorithms, ORB-SLAM is more suitable for extracting physical information in biological lumens, but it still has some limitations to be solved in future works. To improve the capability of scanning and extracting features on organisms with limited texture, the current

DOI: 10.54254/2755-2721/12/20230300

methods either filter the keyframes and only keep the keyframes with plenty of information, which makes the available matching map points less and affects subsequent dense 3D reconstruction, or use complex algorithms to optimize keyframe poses, and use more complex methods to extract denser point clouds from endoscopic images, which increases the number of calculations and even makes the system unable to achieve basic real-time reconstruction. Therefore, in future works, how to balance the number of keyframes and computing time while ensuring the performance of system feature extraction is overwhelmingly significant and indispensable. Currently, ORB-SLAM 2 with binocular cameras and ORB-SLAM 3 with fisheye cameras have been proposed, providing new possibilities for the further optimization of ORB-SLAM in the 3D reconstruction of endoscopes. Besides, the researchers also opened up new directions for the improvement of ORB-SLAM from the perspective of technology combination, such as using deep learning and neural network like convolutional neural network (CNN) to strengthen the information processing ability and Enable faster real-time response; using laser marking to generate more features on lumen surfaces to alleviate data scarcity; Utilizing the characteristics of ORB-SLAM with small error to improve the position accuracy of augmented reality (AR) annotations, labels, 3D model alignment and other information during surgery [18, 19]. The examples of the combination between the above technologies and the ORB-SLAM show that the system has broad development space and great potential.

4. Conclusion

In the paper, the high feasibility of utilizing this method in endoscopic 3D reconstruction is illustrated by an overview of the origin and development of ORB-SLAM. This paper mainly introduces two existing ORB-SLAM-based endoscopic 3D reconstruction methods and analyzes the limitations and issues to explore the solutions in other reconstruction methods. In addition, two other methods are listed from the aspects of algorithm and hardware, which provide a research direction for further improving the 3D reconstruction method based on ORB-SLAM. The paper also compares some performance indicators such as estimation, relocalization, non-rigid deformation, real-time, RMSE etc. of the four methods and, more intuitively, show some differences between the methods. Then, by discussing the advantages and disadvantages of several methods mentioned in the article, the following conclusions are drawn: It is currently difficult to balance calculation time and 3D reconstruction quality in applying ORB-SLAM in endoscopic 3D reconstruction. In order to realize real-time reconstruction, it can only be filtered multiple times to remove keyframes with large uncertainty and preserve the one that contain more information. Although the calculation time is reduced, the reconstruction scope is shrunk. In order to improve the reconstruction accuracy and expand the range, more complex algorithms need to be used to obtain denser point clouds and reduce the degree of dispersion of the point on the reconstructed 3D model. Although dense 3D reconstruction is achieved this way, the real-time function is sacrificed, and even the image must be processed offline. Finally, from the aspect of technology integration, it is clarified that utilizing ORB-SLAM in endoscope 3D reconstruction has a high degree of matching with the current advanced technology such as CNN, laser marking, AR etc., which means that it has the possibility of iteration and further improvement and has broad application scenarios and great development potential. After continuous research and improvement, the endoscope 3D reconstruction system based on ORBSLAM will achieve smaller errors, faster response and stronger robustness. And through integration with other technologies, the system will also achieve more functional expansion, which greatly facilitates the use of surgeons.

References

[1] Chia-Hsiang Wu, Yung-Nien Sun, & Chien-Chen Chang. (2007). Three-Dimensional Modeling From Endoscopic Video Using Geometric Constraints Via Feature Positioning. IEEE Transactions on Biomedical Engineering, 54(7), 1199-1211. https://doi.org/10.1109/TBME.2006.889767

[2] Mountney, P., Stoyanov, D., Davison, A., & Yang, G.-Z. (2006). Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery. In R. Larsen, M. Nielsen,

& J. Sporring (Eds.), Medical Image Computing and Computer-Assisted Intervention-MICCAI 2006 (Vol. 4190, pp. 347-354). Springer Berlin Heidelberg. https://doi.org/10.1007/11866565_43

[3] Davison. (2003). Real-time simultaneous localisation and mapping with a single camera. Proceedings Ninth IEEE International Conference on Computer Vision, 1403-1410 vol.2. https://doi.org/10.1109/ICCV.2003.1238654

[4] Grasa, O. G., Civera, J., Guemes, A., Munoz, V., & Montiel, J. M. M. (n.d.). EKF Monocular SLAM 3D Modeling, Measuring and Augmented Reality from Endoscope Image Sequences.

[5] Mountney, P., & Yang, G.-Z. (2010). Motion Compensated SLAM for Image Guided Surgery. In T. Jiang, N. Navab, J. P. W. Pluim, & M. A. Viergever (Eds.), Medical Image Computing and Computer-Assisted Intervention-MICCAI 2010 (Vol. 6362, pp. 496-504). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15745-5_61

[6] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. 2011 International Conference on Computer Vision, 2564-2571. https://doi.org/10.1109/ICCV.2011.6126544

[7] Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). BRIEF: Binary Robust Independent Elementary Features. In K. Daniilidis, P. Maragos, & N. Paragios (Eds.), Computer Vision - ECCV 2010 (Vol. 6314, pp. 778-792). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15561-1_56

[8] Lin, B., Johnson, A., Qian, X., Sanchez, J., & Sun, Y. (2013). Simultaneous Tracking, 3D Reconstruction and Deforming Point Detection for Stereoscope Guided Surgery. In H. Liao, C. A. Linte, K. Masamune, T. M. Peters, & G. Zheng (Eds.), Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions (Vol. 8090, pp. 35-44). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40843-4_5

[9] Klein, G., & Murray, D. (2007). Parallel Tracking and Mapping for Small AR Workspaces. 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 1-10. https://doi.org/10.1109/ISMAR.2007.4538852

[10] Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, 31(5), 1147-1163. https://doi.org/10.1109/TRO.2015.2463671

[11] Mur-Artal, R., & Tardos, 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. https://doi.org/10.1109/TRO.2017.2705103

[12] Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., & Tardós, J. D. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM. IEEE Transactions on Robotics, 37(6), 1874-1890. https://doi.org/10.1109/TRO.2021.3075644

[13] Mahmoud, N., Cirauqui, I., Hostettler, A., Doignon, C., Soler, L., Marescaux, J., & Montiel, J. M. M. (2017). ORBSLAM-Based Endoscope Tracking and 3D Reconstruction. In T. Peters, G.-Z. Yang, N. Navab, K. Mori, X. Luo, T. Reichl, & J. McLeod (Eds.), Computer-Assisted and Robotic Endoscopy (Vol. 10170, pp. 72-83). Springer International Publishing. https://doi.org/10.1007/978-3-319-54057-3_7

[14] Chen, W., Liao, X., Sun, Y., & Wang, Q. (2020). Improved ORB-SLAM Based 3D Dense Reconstruction for Monocular Endoscopic Image. 2020 International Conference on Virtual Reality and Visualization (ICVRV), 101-106. https://doi.org/10.1109/ICVRV51359.2020.00030

[15] Wang, K., & Shen, S. (2018). Adaptive Baseline Monocular Dense Mapping with Inter-Frame Depth Propagation. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 3225-3232. https://doi.org/10.1109/IROS.2018.8593936

DOI: 10.54254/2755-2721/12/20230300

[16] Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., & Montiel, J. M. M. (2019). Live Tracking and Dense Reconstruction for Handheld Monocular Endoscopy. IEEE Transactions on Medical Imaging, 38(1), 79-89. https://doi.org/10.1109/TMI.2018.2856109

[17] Huo, J., Zhou, C., Yuan, B., Yang, Q., & Wang, L. (2023). Real-Time Dense Reconstruction with Binocular Endoscopy Based on StereoNet and ORB-SLAM. Sensors, 23(4), 2074. https://doi.org/10.3390/s23042074

[18] Turan, M., Almalioglu, Y., Konukoglu, E., & Sitti, M. (2017). A Deep Learning Based 6 Degree-of-Freedom Localization Method for Endoscopic Capsule Robots (arXiv:1705.05435). arXiv. http://arxiv.org/abs/1705.05435

[19] Qiu, L., & 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), 2278-22787. https://doi.org/10.1109/CVPRW.2018.00295