Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2023)
DOI: 10.54254/2753-8818/5/20230373

Researches Advanced in Intubation Based on SLAM

Dongshuo Gao[1,4,†], Zidong Liu[2,†] and Qian Wang[3,†]

[1] Yan Tai University, Yantai, Shan Dong province,264000, China
[2] Northeastern University, Shenyang Liao Ning Province,110000, China
[3] Beijing Technology and Business University, Beijing, 10000, China

[4] 202056501333@s.ytu.edu.cn
[†]There are authors contributed equally

Abstract. Intubation is an emergency medical procedure used to rescue people who are unconscious or unable to breathe on their own. In the process of tracheal intubation, traditional intubation is difficult due to narrow viewing angle, trachea bending, and occlusion of internal structures. Thanks to the rapid development of simultaneous localization and mapping (SLAM) and virtual reality (AR) technologies, intubation aids are also developing towards intelligence and automation. Based on the collection and analysis of a large number of literatures, this paper conducts an in-depth study of the existing SLAM-based intubation-assisted technology. Specifically, on the basis of analyzing the current research on the application of equipment, visualization schemes, and algorithm schemes in tracheal intubation and intubation training, this paper studies the application progress in terms of efficiency, versatility, accuracy, and user feedback. Further, we summarize the existing key issues and discuss future developments.

Keywords: Visual SLAM, Augmented reality, Intubation.

1. Introduction

In medicine, endotracheal intubation is an emergency procedure used to rescue people who are unconscious or unable to breathe on their own. Endotracheal intubation is the intubation of the trachea through the mouth and nose as an open channel for the upper airway, allowing air to freely flow in and out of the lungs, allowing the lungs to ventilate. However, due to its complicated operation and detection, some patients are difficult to intubate, which can easily lead to serious consequences such as long intubation time, low success rate, loss of rescue time, and complications [1-2]. All of the above are in stark contrast to the current wide range of application scenarios and rescue application backgrounds. Therefore, how to lower the technical threshold of intubation technology and improve the success rate and efficiency of intubation is particularly critical for promoting intubation technology in the primary medical system and for more patient rescue time [3].

In recent years, with the rise and development of artificial intelligence and deep learning, many researchers have applied artificial intelligence technology to the field of medical image detection and achieved remarkable results [4-6]. In terms of endotracheal intubation devices, intubation aids are also developing towards intelligence and automation. Thanks to the rapid development of Simultaneous Localization and Mapping (SLAM) technology, assistive technologies and devices for endotracheal intubation have been continuously developed in recent years. SLAM has always been a research hotspot

DOI: 10.54254/2753-8818/5/20230373

in the field of computer vision and robotics, and its basic task is to predict the location of an agent in a scene and perceive its surrounding environment information to construct a map. During the actual tracheal intubation process, medical personnel cannot perceive the internal state of the trachea finely due to narrow viewing angles, tracheal bending, and occlusion of internal structures. SLAM can perfectly solve the technical difficulties of tracheal intubation. It can assist in the completion of intubation, making intubation more versatile, feedback stronger, more accurate and more efficient during intubation [7]. For example, SLAM senses the environment inside the trachea through sensors [8]. Compared with endoscopic images with a smaller viewing angle and sensitive to the distance and size of the target, SLAM can more effectively suppress vision when sputum and airway secretions block the tracheal orifice or esophagus interference, and reduce the technical threshold of endotracheal intubation. At the same time, the continuous maturity and improvement of AR technology and 3D-modeling technology also provide the possibility to build a good visual model [9]. To this end, SLAM-based tracheal intubation-assisted research has gradually attracted the interest of a gigantic number of researchers.

Covid-19 has become the most widely spread and most concerned respiratory infectious disease in the past five years. As one of the important links in the treatment of this disease, endotracheal intubation technology is urgent and important to improve the efficiency and practicability of this technology [10]. Based on the analysis of the current research on the application of equipment, visualization schemes, and algorithm schemes in tracheal intubation and intubation training, this paper studies the application progress in terms of efficiency, universality, accuracy, and user feedback. Finally, we summarize the existing key questions and explore future developments.

2. Methods

We reviewed some intubation experiment reports and endoluminal modeling algorithm schemes. In the following paragraphs, we will first analyze the AR-assisted intubation operation and teaching issues that emerged in these experiments. Secondly, on the basis of consulting some intubation experiment reports and intraluminal modeling algorithms, three representative intraluminal modeling algorithms are introduced and analyzed.

2.1. Semantic SLAM Based on Deep Learning

Firstly, physician observes the actual environment of the lumen through the endoscope to obtain the corresponding lumen image, and then transmit it to the SLAM tracking county and semantic segmentation network. The devices used in the surgery process are segmented through the semantic segmentation network to solve the impact of medical device movement in the image frame in the actual surgery scene. In the tracking thread, ORB feature points with stable geometric features in each new frame are extracted and segmented to determine whether the current feature point is a dynamic feature point. If the feature points are judged to be dynamic, they will be eliminated. Otherwise, it will be judged as the SLAM of static feature points for subsequent tracking and mapping, including the estimation of camera attitude by matching adjacent frames, the depth estimation obtained by triangulation, and the joint optimization of map and camera attitude by using local beam and global beam.

By adding a semantic information to distinguish surgical instruments, the segmentation of dynamic features is completed, and the determined dynamic features are removed as outliers. Using this judgment method, the robustness and accuracy of the system can be improved to a large extent, making it more suitable for the actual environment in the real process. Secondly, the ORB feature points are extracted from each acquired image frame, and the corresponding point pairs are obtained by comparing the descriptors of each feature point among them, so that the endoscopic motion can be estimated according to the corresponding point pair relationship. Random sample consistency (RANSAC) is an excellent algorithm to estimate the relevant parameters of the model in an iterative manner. Random samples are mainly used to select more reliable matching points from a gigantic number of matching points to eliminate outliers. And semantic segmentation is used to precisely exclude relevant dynamic features, so as to obtain effective data in the dataset containing outliers.

DOI: 10.54254/2753-8818/5/20230373

According to the results obtained by these methods, errors found in the operation tool movement can be eliminated. Specific judgment method: whether a point in the feature sequence PIK ∈ SKN is satisfied, the feature point is considered as a dynamic feature point and deleted from the function point sequence. At the same time, the pre selected features are filtered through semantic segmentation. This method not only eliminates the dynamic characteristics of surgical instruments, reduces the impact of surgical instrument movement on detection, but also avoids the negative impact of detection errors on SLAM. In this process, the features of other regions are determined as static features, and the abnormal phenomena are further eliminated through RANSAC algorithm, which lays a good foundation for the introduction of subsequent methods of endoscopic position estimation based on the correct relevant data obtained.

When the initial camera attitude estimation is completed, it is estimated by perspective n-point (PnP) algorithm or iterative closest point (ICP) algorithm. Through the establishment of the endoscope attitude and local map, the matching between the current frame obtained and the map constructed is realized. In the aspect of key frame generation, the method of minimizing re projection is adopted to determine the key frames according to the position, attitude and motion of adjacent frames, and achieve good optimization of position and attitude. Secondly, in order to optimize the local map constructed, the link of filtering the newly generated map points is added in the local mapping thread. The triangulation map points have a high degree of common view, and the execution bundle adjustment (BA) is used to optimize the local map, so as to remove the miscellaneous key frames in the initially constructed local map, Finally, the constructed map is updated in real time through global BA optimization of the global attitude and filtered map points to obtain a better map than the original one.

2.2. SD-Def SLAM

Semi direct deformation SLAM (SD DefSLAM) is a method that does not depend on the constancy of light source and the fixation of scene, so as to ensure reliable operation in low texture areas, realize the correlation of short-term and medium-term data, and optimize the relevant geometric errors. Figure 1 describes the overall situation of SD DefSLAM. We can see that two threads are used, one for deformable mapping, so as to build a growing free map; Another thread is used for deformable tracking, so that the pose of each camera and the deformation of the map can be estimated.

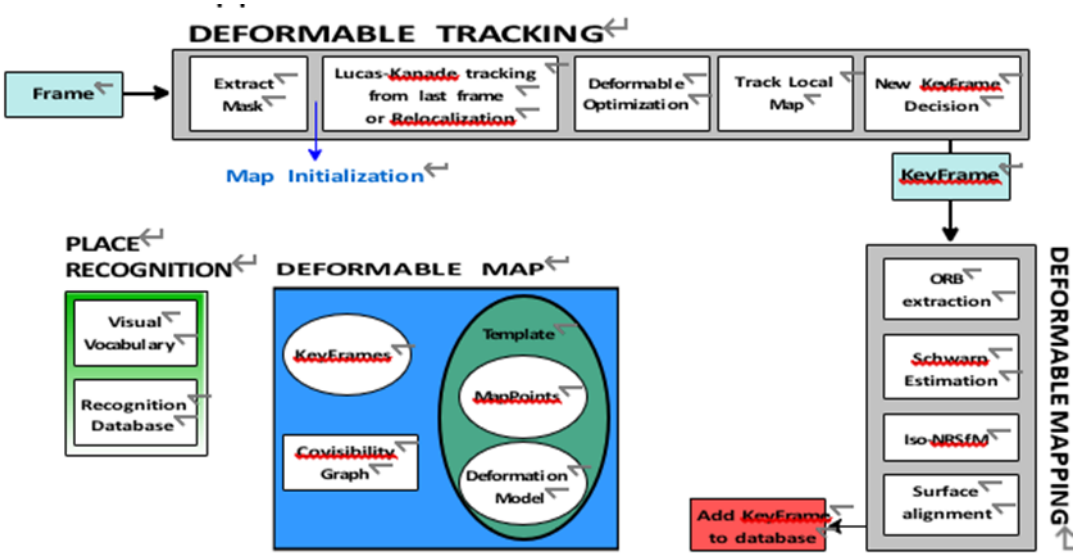

Figure 1. The framework of SD-Def SLAM algorithm.

The formation of the map depends on a set of reference keyframes, which will observe new scene parts and a related surface model during the exploration process. The observation surface of each model has

DOI: 10.54254/2753-8818/5/20230373

a triangle network, representing that it is in a static state, and its vertex is a 3D point of corresponding coordinates. It also includes a set of keyframes for optimizing the model. The model can be created and refined by the deformable mapping thread depending on the key frame rate, and its deformable model is estimated by the deformable tracking thread in the form of frame rate. Key frames are added to the database of position recognition to complete the repositioning after occlusion. The estimation is gradually improved by thinning the key frames after referring to the surface observed by the key frames. The last shift is created to explore new locations and expand the map in equal proportion. The core of deformation mapping is a non rigid moving structure (NRSfM) algorithm based on lamps and infinitesimal flatness. By estimating the normals of key frame points, assume that the points in the surface are smooth relative to the estimated rest of the normals, initialize these points, and refine them by using a new viewing angle. After the estimation discovery, the algorithm of obtaining normal shape estimates the scale shape of the surface matching these normals. Finally, the SE is executed to obtain the correct proportion of the restoration compared to other parts of the region. This new surface will also become one of the templates for deformation tracking.

The deformation tracking thread uses the update rate of the frame to estimate the position of the camera and the deformation degree of the 3D map surface, and codes the static morphological model and the deformed morphological model. In order to obtain a better position of estimated points, we use LK tracker to calculate each estimated point independently, and match the corresponding results with initial assumptions. The initial deformation of the mesh is analyzed using the matching of the initial assumptions, and the robust outliers are obtained. Next, the reinitialized LK tracker will search the map landmarks in the observation area to find a larger baseline to match with the LK tracker. To ensure the stability of tracking, LK tracker can reposition the system in the map through repositioning module in case of tracking loss. Finally, the subdivision tool of CNN can eliminate the matching dynamic non modeling objects for further optimization. [11]

2.3. ORB-SLAM

The sparse keyframes based SLAM system (ORB-SLAM) is extended with a new thread for dense reconstruction. We define the keyframe set as the selected frames used in ORB-SLAM for its Bundle Adjustment (BA) process. And it will be realized through four modules. The authers assume laparosocpe is pre-calibrated with fixed intrinsic through move either the camera or the planar pattern to get a planar pattern from a few different orientations and lens distortion has been eliminated.

In the first module, according to the coverage rate of current intensive reconstruction in the key frame I_r given above, automatically select a subset of all available key frames obtained by camera tracking. The coverage of intensive reconstruction is determined by projecting the current reconstruction to the key frame Ir. If the proportion of reconstructed pixels obtained is less than 50%, return to select I_r for densification again. In the first stage, we search from the key frame Ir to find the frames whose parallax with respect to the key frame Ir exceeds the threshold. In the second stage, frame I_{im} and two adjacent frames I_{im-1} and I_{im+1} are selected. If frame I_{im} and two adjacent frames I_{im-1} are selected, the parallax between Iim+1 is lower than α 2 threshold, the selected frame will be removed from the cluster as the frame with low parallax, so as to reduce the computing cost.

The main purpose of the second module is to use sparse reconstruction to define the depth range used to build the three-dimensional cost volume. First, we detect the reflection of the mirror on the key frame Ir to avoid introducing outliers into the dataset. At the same time, invisible pixels are allocated to zero in the data item. After optimization, all pixels in the above area are cleared, so as to reduce the uncertainty estimation of depth. In the second stage, we introduce the regular term R (u, ρ (u)) to ensure the smoothness of scene reconstruction while maintaining the discontinuity of depth. At the same time, empirical factors are also introduced β Min and β Max, the final interval is $[\beta_{min}\rho_{min}, \beta_{max}\rho_{max}]$ to ensure that extreme points are not wrongly excluded. This range of inverse depths is evenly discretized into ξ sampling points.

$$E(\rho, a) = \int_{\Omega} \left\{ \lambda(u) C\big(u, a(u)\big) + \frac{1}{2\theta} \big(\rho(u) - a(u)\big)^2 + R\big(u, \rho(u)\big) \right\} \quad (1)$$

DOI: 10.54254/2753-8818/5/20230373

In the third module, we use the optimized variational method to intensively reconstruct the key frames selected in the above stages. Compared with the previous variational methods, we use an auxiliary mapping a: Ω→ R to approximate the energy function, which is used to couple the ZNCC data items and Huber norm regularization to obtain a strong local minimum. Iterative solution starts when t=1, where θ stay θ Initialization started at 1 place, ρ (U) And a (U) are initialized by the initial depth map obtained from the second module. During iteration, when θ (t+1)= θ T (1 − t) exceeds the terminal threshold θ End of optimization. In addition, we execute single Newton Step during iteration to obtain sub-pixel accuracy.

$$\arg\min_{a(u)} \lambda(u)C\big(u, a(u)\big) + \frac{1}{2\theta}\big(\rho(u) - a(u)\big)^2 \tag{2}$$

In the fourth module, we can obtain a globally consistent reconstruction by aligning the keyframe depth maps with the sparse SLAM map We combine the gradually dense places obtained in the real-time tracking process into a single coordinate system. Most sparse SLAM points have a one-to-one 3D anchor point in the dense map obtained by densification, and the similarity transformation is performed on each depth map. Through the calculation of formula (3), it is possible to perform a similarity transformation S of the re projection error estimation of point P in a nonlinear minimized sparse SLAM and use K adjacent key frames to share the most feature points.

$$\arg\min_{S\in Sim(3)} \sum_{j\in P, i\in K} \rho_h(\|X_{i,j} - \pi\big(T_i, SX_j\big)\|^2 \tag{3}$$

Xi and j in formula (3) are the image observations of key frames i and point j in sparse SLAM. And Xj is the 3D position of the keyframe Ir in the fourth module.

3. Comparison and performance analysis

Building a non-explicit SLAM visual framework that includes a semantic separation for at least surgical scenarios. The symmetric distribution network was established based on the TernausNet-16 network architecture and improved to operate separately from the surgery tool in the cave image. Then a dynamic point change model added to SLAM to remove specific points in the field of the surgery mask, so the dynamic features identified in the surgery tool and the only possible points are these. The SLAM system activated that the internal image collection should be more stable and more accurate. In the video collection tests, endoscopic, our proposal method [7] reached better results in the section and the mapping of the tool.

A direct half-way [11] based on tracking transparent and non-explorable images, data connections are much better, accurate and measurable. The combination with the CNN section allows the first SLAM system to solve the real challenges of life through medical methods.

The third method described in the second part above proposes a new dense SLAM system. Through this system, the motion of the endoscope can be tracked through the frame rate, and high-quality close range reconstruction of the surgical scene can be completed in a very short time. And the algorithm has been verified in the new data set, and obtained a high evaluation, and has shown excellent robustness in the actual serious lighting changes and various scene textures. On the one hand, it has a high practical accuracy. In the evaluation of surface intensive reconstruction accuracy, the accuracy has reached 1.1 mm. On the other hand, the stereo method provides a similar measurement method for the estimation accuracy of the abdominal posture. In the method, only the size alignment is used, so the relative value of camera rotation and translation and the accuracy of the estimated surface is mainly measured. When the camera returns to the included area in the constructed map, the SLAM algorithm is implemented.

4. Discussion

The first method can effectively solve the problem of dynamic feature points, so as to better build the internal cavity image and reduce the negative impact caused by the movement of surgical instruments. This method enables SLAM system to process lumen image sequences more accurately and faster, thus

DOI: 10.54254/2753-8818/5/20230373

obtaining more accurate results. We believe that if this method is applied to endotracheal intubation, it can effectively help solve the problem of difficult air cavity positioning, and is more convenient for endotracheal intubation.

The second method introduced in the second part can improve the correlation between the acquired data, thereby improving the composition accuracy and reducing scale drift. At the same time, combined with CNN segmentation method, it has a good effect in detecting moving objects, repositioning and so on. Therefore, it can be used to detect moving objects, eliminate moving feature points and deal with trachea obstruction in endotracheal intubation.

The method introduced in Section 3 of Part II can use image features to track the endoscope, so that the internal scene can complete high-quality density reconstruction in a few seconds. At the same time, reconstruction can be realized in bad scenes, and the accuracy of 1.1 mm can be ensured.

5. Conclusion

In the process of tracheal intubation, traditional intubation is difficult due to narrow viewing angle, trachea bending, and occlusion of internal structures. Benefiting from the rapid development of simultaneous localization and mapping (SLAM) technology, assistive technologies and equipment for endotracheal intubation have continued to develop in recent years. SLAM uses sensors to perceive information about its surrounding environment to build a map, making the feedback during intubation stronger, more accurate, and more efficient. Based on the analysis of the current research on the application of equipment, visualization schemes, and algorithm schemes in tracheal intubation and intubation training, this paper studies the application progress in terms of efficiency, versatility, accuracy, and user feedback. Finally, we summarize key existing issues and discuss future developments.

References

[1] Wanigasekara, R. M. R., S. D. M. H. Siyambalapitiya, and S. D. S. H. Dissanayake. *Generate Navigations to Guide and Automate Endotracheal Intubation Process*. Diss. 2021.

[2] Williamson, J. A., et al. "Difficult intubation: an analysis of 2000 incident reports." *Anaesthesia and intensive care* 21.5 (1993): 602-607.

[3] Matek, Jan, et al. "Optical Devices in Tracheal Intubation—State of the Art in 2020." *Diagnostics* 11.3 (2021): 575.

[4] Alismail, Abdullah, et al. "Augmented reality glasses improve adherence to evidence-based intubation practice." *Advances in Medical Education and Practice* 10 (2019): 279.

[5] Sielhorst, Tobias, et al. "Depth perception–a major issue in medical AR: evaluation study by twenty surgeons." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Berlin, Heidelberg, 2006.

[6] Huang, Cynthia Y., et al. "The use of augmented reality glasses in central line simulation:“see one, simulate many, do one competently, and teach everyone”." *Advances in medical education and practice* 9 (2018): 357.

[7] Wu, Haibin, et al. "Semantic SLAM Based on Deep Learning in Endocavity Environment." *Symmetry* 14.3 (2022): 614.

[8] Dias, Patricia L., et al. "Augmented Reality–Assisted Video Laryngoscopy and Simulated Neonatal Intubations: A Pilot Study." *Pediatrics* 147.3 (2021).

[9] Long, Yonghao, et al. "Integrating Artificial Intelligence and Augmented Reality in Robotic Surgery: An Initial dVRK Study Using a Surgical Education Scenario." *arXiv preprint arXiv:2201.00383* (2022).

[10] Yao, Wenlong, et al. "Emergency tracheal intubation in 202 patients with COVID-19 in Wuhan, China: lessons learnt and international expert recommendations." *British journal of anaesthesia* 125.1 (2020): e28-e37.

[11] Rodríguez, Juan J. Gómez, et al. "Sd-defslam: Semi-direct monocular slam for deformable and intracorporeal scenes." *arXiv preprint arXiv:2010.09409* (2020).