

Article

3D Texture Reconstruction of Abdominal Cavity Based on Monocular Vision SLAM for Minimally Invasive Surgery

Haibin Wu ¹, Ruotong Xu ¹, Kaiyang Xu ¹, Jianbo Zhao ¹, Yan Zhang ¹, Aili Wang ^{1,*}  and Yuji Iwahori ² 

¹ Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China; woo@hrbust.edu.cn (H.W.); 1920610059@stu.hrbust.edu.cn (R.X.); 1920610061@stu.hrbust.edu.cn (K.X.); 2020600002@stu.hrbust.edu.cn (J.Z.); 2020610056@stu.hrbust.edu.cn (Y.Z.)

² Department of Computer Science, Chubu University, Kasugai 487-8501, Japan; iwahori@isc.chubu.ac.jp

* Correspondence: aili925@hrbust.edu.cn

Abstract: The depth information of abdominal tissue surface and the position of laparoscope are very important for accurate surgical navigation in computer-aided surgery. It is difficult to determine the lesion location by empirically matching the laparoscopic visual field with the preoperative image, which is easy to cause intraoperative errors. Aiming at the complex abdominal environment, this paper constructs an improved monocular simultaneous localization and mapping (SLAM) system model, which can more accurately and truly reflect the abdominal cavity structure and spatial relationship. Firstly, in order to enhance the contrast between blood vessels and background, the contrast limited adaptive histogram equalization (CLAHE) algorithm is introduced to preprocess abdominal images. Secondly, combined with AKAZE algorithm, the Oriented FAST and Rotated BRIEF(ORB) algorithm is improved to extract the features of abdominal image, which improves the accuracy of extracted symmetry feature points pair and uses the RANSAC algorithm to quickly eliminate the majority of mis-matched pairs. The medical bag-of-words model is used to replace the traditional bag-of-words model to facilitate the comparison of similarity between abdominal images, which has stronger similarity calculation ability and reduces the matching time between the current abdominal image frame and the historical abdominal image frame. Finally, Poisson surface reconstruction is used to transform the point cloud into a triangular mesh surface, and the abdominal cavity texture image is superimposed on the 3D surface described by the mesh to generate the abdominal cavity inner wall texture. The surface of the abdominal cavity 3D model is smooth and has a strong sense of reality. The experimental results show that the improved SLAM system increases the registration accuracy of feature points and the densification, and the visual effect of dense point cloud reconstruction is more realistic for Hamlyn dataset. The 3D reconstruction technology creates a realistic model to identify the blood vessels, nerves and other tissues in the patient's focal area, enabling three-dimensional visualization of the focal area, facilitating the surgeon's observation and diagnosis, and digital simulation of the surgical operation to optimize the surgical plan.



Citation: Wu, H.; Xu, R.; Xu, K.; Zhao, J.; Zhang, Y.; Wang, A.; Iwahori, Y. 3D Texture Reconstruction of Abdominal Cavity Based on Monocular Vision SLAM for Minimally Invasive Surgery. *Symmetry* **2022**, *14*, 185. <https://doi.org/10.3390/sym14020185>

Academic Editor: José Carlos R. Alcántud

Received: 19 December 2021

Accepted: 13 January 2022

Published: 18 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: 3D texture reconstruction; simultaneous localization and mapping (SLAM); contrast limited adaptive histogram equalization (CLAHE); Oriented FAST and Rotated BRIEF(ORB); bag-of-words model



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Minimally invasive surgery is a new surgical technology that uses modern medical instruments and equipment to pass through small wounds on the surface of the human body and perform multiple actions with human hand-eye cooperation in the human body [1]. Compared with traditional surgery or early minimally invasive surgery, modern minimally invasive surgery has the advantages of accurate operation, less bleeding and faster postoperative recovery. It is increasingly welcomed by patients and widely used in internal cavity surgery. However, surgeons are prone to disorientation and occasional

hand–eye imbalance when they perform complex surgery through the 2D visual display of endoscopic video stream, and it is difficult to determine the lesion location by empirically matching the endoscopic visual field with the preoperative image, which is easy to cause intraoperative errors.

In recent years, minimally invasive surgery has been gradually integrated with computer three-dimensional (3D) reconstruction technology. For example, surgeons use surgical experience and image processing technology to stereo locate the lesion area using endoscope system, breaking the limitations of traditional surgery [2]. To help surgeons simulate actual surgical operations, the digital 3D reconstruction model can be printed out in equal scale through 3D printing technology [3]. Additionally, the 3D reconstruction model allows the surgeon to explain the patient's condition and surgical plan visually [4], facilitating smooth communication between the surgeon and patient and enhancing the patient's confidence in treatment. At present, researchers at home and abroad have proposed different kinds of methods based on computer vision to restore the three-dimensional surface structure of surgical scene in minimally invasive surgery, which are mainly based on laser, coded structured light, time camera and video camera. Among them, the surface reconstruction technology based on endoscope video has many obvious advantages. Specifically, this method provides intraoperative information without destroying the internal structure of the human body, and there is no need to introduce additional hardware into the current surgical platform. Although endoscopic video provides on-site feedback information for surgeons during surgery, video information has limitations and cannot meet the needs of doctors. First, there is no clear depth information in two-dimensional images, so surgeons must estimate the depth according to their experience. In addition, the field of vision of the endoscope is very narrow, and it is difficult for the surgeon to accurately locate the position and direction of the endoscope and surgical instruments. More importantly, due to the limitation of the complex environment of human lumen, the number of cameras to obtain the surface information of lumen tissue also affects the real-time and anti-interference ability in the process of 3D reconstruction.

Monocular vision is a three-dimensional reconstruction technology that uses a camera to capture the image of the target object. There are two main ways to realize the three-dimensional modeling of monocular vision in the lumen environment. One way is to use the information of the lumen image itself to obtain the three-dimensional feature information of the lumen through a specific algorithm. Another way is to calibrate the camera parameters of the endoscope system to obtain the depth information of the measured point. Because monocular vision method has the advantages of simple equipment structure, convenient use and easy data processing, most of the research at home and abroad is to use a monocular vision algorithm to reconstruct the inner cavity.

In order to improve the accuracy of 3D reconstruction, Wu et al. [5] proposed combining shape from shading (SFS) method and motion shape restoration method for the inner cavity 3D reconstruction in 2010. This method combines the iterative nearest point algorithm to reduce the error of coordinate system conversion in multiple artificial spine images, improve the matching rate and recover the bone boundary line.

In 2012, Ciuti et al. [6,7] proposed a complete set of SFS calibration methods. Assuming that the light source is close to the organ surface and far away from the optical center, the spatial three-dimensional coordinates are obtained by triangulating the part of the organ surface with specular highlights. Without any preoperative data, the endoscope device performs 3D measurement according to the calculated trajectory and finally realizes the automatic navigation of the capsule. However, the magnetic levitation capsule cannot reach the ideal state in the process of movement, and the calibration accuracy needs to be further improved. In the same year, Tokgozoglu et al. [8] proposed an SFS method based on color projection, which can minimize the intensity changes caused by different surface characteristics. In 2015, Goncalves et al. [9] proposed a perspective shape from shading (PSFS) algorithm based on near light source perspective mapping to solve radial distortion and reduced image edge resolution. This method establishes the radial distortion

model, compensates the problem of reduced image edge resolution, and completes the three-dimensional reconstruction of knee bone. In 2016, Lei et al. [10] proposed perspective mapping SFS method based on photometric calibration method to reconstruct organ surface. This method, combined with optical flow method, changes the relative change of gray gradient field into absolute change, which improves the stability of organ surface reconstruction. In 2018, Turan et al. [11] used the above method for gastrointestinal surface reconstruction. However, the gastrointestinal surface is not a smooth area. The uneven surface will make the gradient vector change rate higher, and the obtained gray value will also be lower than the real value, resulting in large reconstruction error.

To sum up, the difficulty of SFS algorithm in the process of three-dimensional reconstruction of inner cavity is that there are multiple mappings between a two-dimensional image and the surface shape. At the same time, there is only one formula in the brightness equation, but there are two variables. Therefore, the direction of the object surface cannot be determined only by the brightness equation. However, SFS algorithm is easy to combine with other methods and complement each other for 3D reconstruction. At the same time, SFS algorithm can perform dense calculation on smooth surfaces. Since the 1980s, SLAM has been proposed for the first time, which specifically refers to the technology that the subject equipped with a specific sensor moves in an unknown environment, locates itself and constructs an incremental map [12], which is widely used in real-time reconstruction of endoscope scenes.

In 2015, Lin et al. [13] proposed to restore the surface structure of three-dimensional scene of abdominal surgery based on SLAM, improved the texture characteristics of lumen image, the selection of green channel and the processing of reflective area, and studied a new type of image features, namely branch points in blood tube features. After detecting the vascular feature points, the branch segments are jointly detected and matched to match the vascular features in the image. Finally, three-dimensional blood vessels are recovered from each frame of image, and three-dimensional blood vessels from different perspectives are integrated through blood vessel matching to obtain a global three-dimensional blood vessel network.

In 2016, Yang [14] proposed endoscopic localization and construction of gastrointestinal feature map based on monocular SLAM. In this method, Oriented FAST and Rotated BRIEF (ORB) algorithm is selected for feature points detection from the perspective of efficiency and matching accuracy. Combined with local pose optimization algorithm and triangulation measurement with minimum geometric distance, a large amount of data redundancy is processed through reselection of key frames and screening of feature points. However, because the environment is the intestinal tract with non-closed endoscopic trajectory, and the local part tends to be straight, it is different from the closure of most lumen environments.

In 2019, Mahmoud et al. [15] proposed dense three-dimensional reconstruction of abdominal cavity based on monocular ORB-SLAM. Firstly, the camera pose of key frames is estimated by using the detection and matching process of sparse ORB-SLAM, and the selection of key frames is determined according to the parallax criterion. Then, the variational method combining zero mean normalized cross-correlation (ZNCC) and gradient-robust kernel norm regularizer is used to calculate the dense matching between key frames in parallel. This method uses monocular video input and does not need any reference point or external tracker. It has been verified and evaluated on pig abdominal video sequence, which shows that it is robust to serious illumination changes and different scene textures. The main limitation of the system is that the texture feature description of soft tissue surface is not representative, and there is texture distortion after reconstruction.

In the same year, Xie et al. [16] combined with the measurement data of endoscope in gastrointestinal tract and introduced the local pose optimization algorithm and triangulation algorithm with minimum geometric distance in terms of pose optimization and spatial point positioning. In 2021, LaMarca et al. [17] first proposed the tracking and mapping of deforming scenes from single sequences algorithm, which can run in real time in the

deformed scene, and divide the calculation into two parallel threads. The deformation tracking thread is used to estimate the camera pose and the deformation of the scene. The deformation mapping thread is applied to the pose estimation of the endoscope to better adapt to the lumen deformation scene, so as to generate an accurate 3D model of the human lumen. However, it is easy to be affected by uneven illumination, resulting in poor visual texture, and is not suitable for non-equidistant deformation lumen reconstruction.

In minimally invasive surgery, human tissue will deform and bleed, and often do not have strong edge characteristics, so it shows the characteristics of highlight and specular reflection. Facing this complex minimally invasive surgery environment, monocular SLAM method has high robustness and can process soft tissue sequence images in real time. Thus, 3D texture reconstruction of abdominal cavity based on monocular vision SLAM for minimally invasive surgery is proposed in this paper. The rest of this paper is organized as follows: Section 2 briefly introduces the proposed methods relevant and improvements in this paper. Sections 3–5 describe improved abdominal cavity feature tracking, mapping and optimization, Poisson surface reconstruction and texture mapping, and the experimental results and analysis are given. Section 6 summarizes the conclusions and future work.

2. Proposed Methods

Aiming at the abdominal environment's lack of features and specular reflection area, this paper realizes three-dimensional reconstruction of abdominal cavity based on monocular SLAM. The system flow chart is shown in Figure 1. The system can be divided into the following five modules: sensor data reading, abdominal cavity image preprocessing, abdominal cavity feature tracking, local abdominal cavity map construction, loop detection and map construction. Firstly, the abdominal image is preprocessed to distinguish the specular reflection area and blood vessels to reduce the influence of the former. Secondly, a SLAM system is established, which includes the following three parts: tracking thread, local beam adjustment optimization thread and global pose loop detection. Finally, aiming at the deficiency that the abdominal cavity point cloud is sparse and cannot fully and intuitively describe the lumen environment, the multi frame abdominal cavity images of SLAM are used to provide three-dimensional node data and texture information at the same time, and a three-dimensional rotating lumen model that can be observed from multiple angles is constructed.

Our proposed method is more suitable for the narrow, humid environment of an abdominal cavity environment lacking in features. In this paper, we introduce the CLAHE algorithm to enhance the vascular details in low-illumination abdominal images and adjust the histogram to enhance the contrast of the images. This paper generates BoW models specifically for medical images by extracting visual features of medical images in a way that reduces the time required to solve for abdominal image similarity and reduces the accumulated errors in the construction of 3D maps of the human abdominal cavity. As the abdominal cavity map constructed by the SLAM system is a sparse structure, this paper uses Poisson surface reconstruction to construct a dense mesh surface and superimposes the abdominal cavity texture image onto the mesh to generate a smooth inner wall texture on the surface. By forming a 3D-visualized abdominal cavity texture model, it provides the surgeon with more intuitive information to make more accurate diagnosis. The simulation of surgical operations based on the 3D reconstructed textured model allows the surgeon to have a clear grasp of the surgical procedure before the operation, facilitating the surgeon's assessment of surgical risks and the planning of surgical design plans in advance.

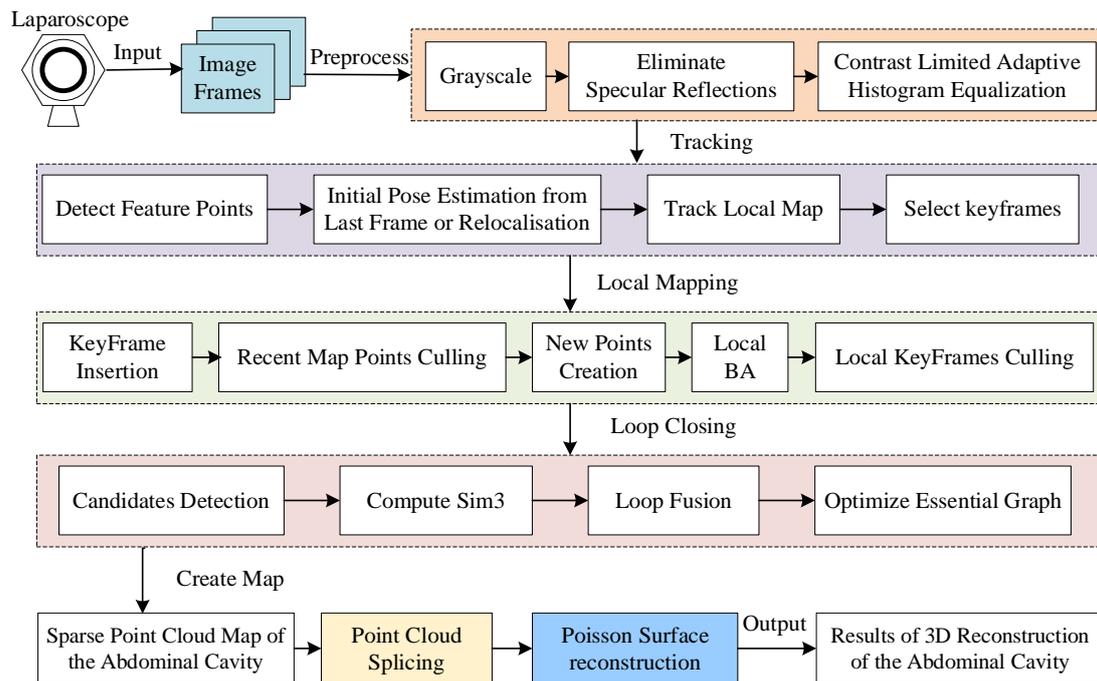


Figure 1. Monocular SLAM based 3D reconstruction system for the human abdominal cavity.

3. Abdominal Cavity Feature Tracking in Monocular SLAM

Compared with images taken in indoor and outdoor environments, abdominal images are usually low striation and include low illumination and specular reflection areas because they are taken in the human abdomen with smooth and wet tissue surface. In order to more accurately and truly reflect the abdominal structure and spatial relationship under different viewing angles and different lighting conditions, this paper uses abdominal image preprocessing to eliminate specular reflection and improve image contrast, and then uses the feature description algorithm in monocular SLAM system to detect and match the feature points of abdominal image.

3.1. Image Preprocessing

Due to the lack of features, repetition and noise in the abdominal cavity image taken by laparoscopy, a sudden change of illumination may occur during the operation, resulting in the reduction in visual recognition of abdominal image, as shown in Figure 2. Therefore, before laparoscopic pose estimation, it is necessary to preprocess the abdominal image collected by laparoscopy in order to extract more feature information from the abdominal image.

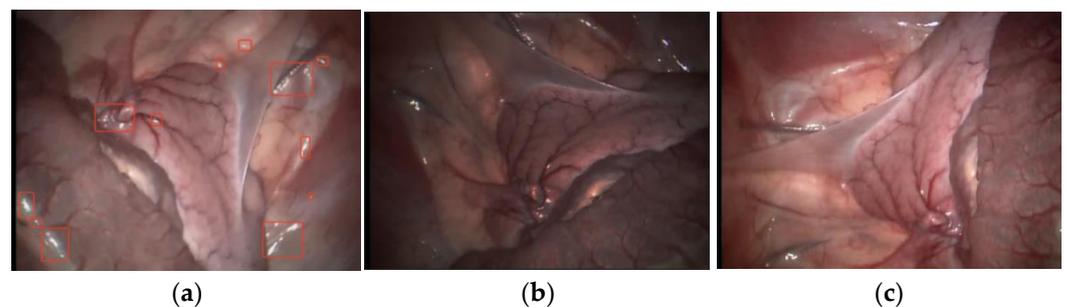


Figure 2. Abdominal images with different illumination. (a) Specular reflection; (b) Low-light image; (c) Partially low illumination.

The abdominal cavity images acquired by the laparoscope are RGB images. In order to extract feature information quickly and accurately, this paper uses grey-scale processed abdominal images, which are less noisy, and converts the abdominal images into an HSV color model, using the S channel to remove specular reflections from the abdominal images which contains saturation information.

In order to improve the contrast of abdominal images, Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [18] is selected to preprocess the low illumination images. Firstly, the abdominal cavity image with low illumination is divided into 8×8 sub blocks of the same size, and these sub blocks do not overlap each other. Calculate the gray histogram of each sub block as follows:

$$N_{AVG} = \frac{n_x n_y}{N_g} \quad (1)$$

where N_{AVG} represents the average number of pixels allocated to each gray level, n_x is the number of pixels in the horizontal direction, n_y is the number of pixels in the vertical direction, and N_g is the number of gray levels in the sub block.

Set the interception coefficient β of the number of gray level pixels, and calculate the interception threshold T as:

$$T = \beta \times N_{AVG} \quad (2)$$

After determining the interception threshold T , the pixels exceeding the threshold in the gray histogram of each sub block are cut, and the intercepted pixels are evenly distributed to each gray level, as shown in Formula (3):

$$N_{re} = \frac{S_c}{N_g} \quad (3)$$

where N_{re} is the number of pixels allocated to each gray and S_c is the total number of intercepted pixels.

The preprocessing experiment is carried out in the human abdominal cavity image in Hamlyn medical image database, as shown in Figure 3. It can be seen that after preprocessing, the specular reflection area in the abdominal image is reduced, the contrast between the blood vessel and the background is enhanced, and the image texture is clearer.

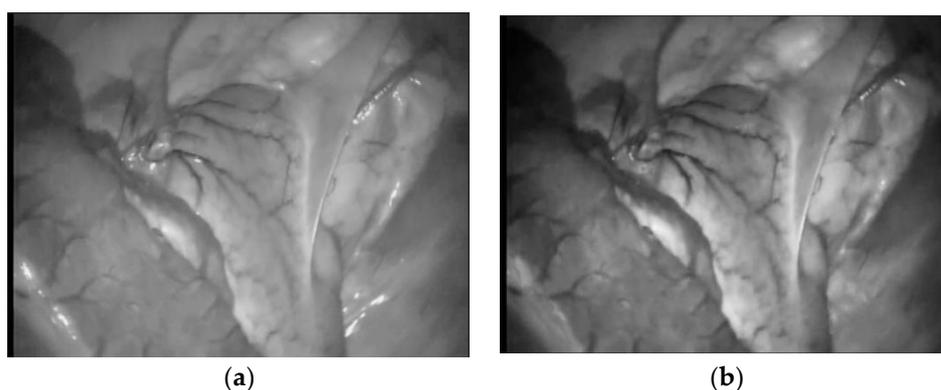


Figure 3. Image pre-processing results. (a) The normal image; (b) The pre-processed images.

3.2. Feature Points Extraction and Matching

In computer vision, feature points contain relevant information in an image. In the case of geometric position, feature points are usually corner points, that is, the points that change in both directions or axes. Monocular SLAM algorithm creates a sparse map in the abdominal scene, usually by extracting points containing characteristic information such as blood vessels as scene features.

ORB [19] algorithm uses features from accelerated segment testfast (FAST) [20] corner extraction algorithm and Binary Robust Independent Elementary Features (BRIEF) [21] descriptor to describe these feature points, which has strong robustness to rotation and scaling, and good invariance to camera automatic gain, automatic exposure and illumination changes. However, in the process of corner detection, the repetition rate of feature points obtained by fast algorithm is low. AKAZE [22] algorithm detects feature points in the scale space established by nonlinear diffusion filtering, which can adaptively adjust and retain the edge area information according to the details of the local area of abdominal image, so as to improve the repeatability and uniqueness of feature points. Combined with AKAZE algorithm, this paper improves the ORB algorithm, named as AKAZE-ORB algorithm. Firstly, AKAZE algorithm is used to extract the feature points of abdominal image to improve the repeatability of feature point extraction, and then the BREF descriptor is used to describe the detected feature points. The AKAZE-ORB algorithm improves the number of feature points extracted, and the registration effect is verified by comparative experiments.

AKAZE implements fast explicit diffusion (FED) [23] embedded in a pyramidal framework that enhances the speed of feature detection in nonlinear scale space. Key points are located by finding the extrema of the second-order derivatives of the image over the nonlinear multi-scale pyramid built from the principle of image diffusion [24]. The FED expression is shown in Formula (4):

$$\mathbf{L}^{i+1,j+1} = \left(\mathbf{I} + \tau_j \mathbf{A}(\mathbf{L}^i) \right) \mathbf{L}^{i+1,j}, \quad j = 0, 1, \dots, n - 1 \tag{4}$$

$$\tau_j = \frac{\tau_{\max}}{2 \cos^2 \frac{\pi(2j+1)}{4n+1}} \tag{5}$$

where $\mathbf{A}(\mathbf{L}^i)$ is the conduction matrix of the image L_i and τ is a constant time step. \mathbf{I} is the identity matrix. Where n represents the dominant diffusion step, τ_j is the corresponding step size, and τ_{\max} is the maximum step size when the dominant diffusion stability condition is met.

Next, this paper uses the BRIEF to establish the feature descriptor. The BRIEF descriptor uses the binary string as the description vector, which describes the feature points by performing gray scale test on the pixels in the neighborhood of the key points, as shown in Formula (6):

$$\varepsilon(p; x, y) = \begin{cases} 1 & p(x) < p(y) \\ 0 & p(x) \geq p(y) \end{cases} \tag{6}$$

where $p(x), p(y)$ are the gray values of points and pixels, respectively. Through comparison, n binary code strings are obtained to form an n -dimensional binary vector:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \varepsilon(p; x_i, y_i) \tag{7}$$

Feature matching is an important step in the monocular SLAM abdominal 3D reconstruction system. The most basic method is the violent matching, that is, measure the distance between each feature point and all descriptors, and take the closest distance as the matching point by sorting. Because the BRIEF descriptor is in binary form, the distance measurement of the descriptor usually relies on the Hamming distance. In terms of eliminating mismatches, RANSAC algorithm [25] is used to randomly extract the sub data sets in the noisy data sets in an iterative way to establish a mathematical model, and then use the parametric model to evaluate and test the remaining non extracted sub data sets.

The effect of traditional ORB algorithm and AKAZE-ORB algorithm on extracting feature points of abdominal image frame is shown in Figure 4. The effect of AKAZE-ORB algorithm on abdominal image frame feature point matching is shown in Figure 5.

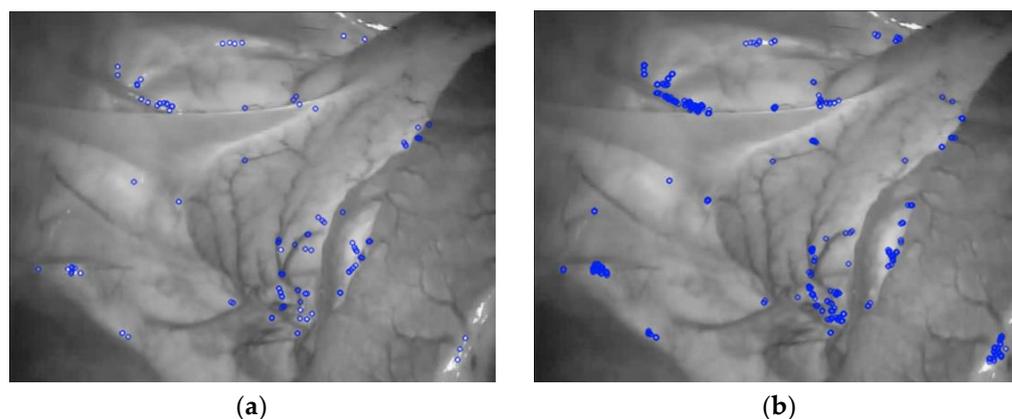


Figure 4. Comparison of ORB and AKAZE-ORB algorithm for feature points extraction. (a) ORB; (b) AKAZE-ORB.

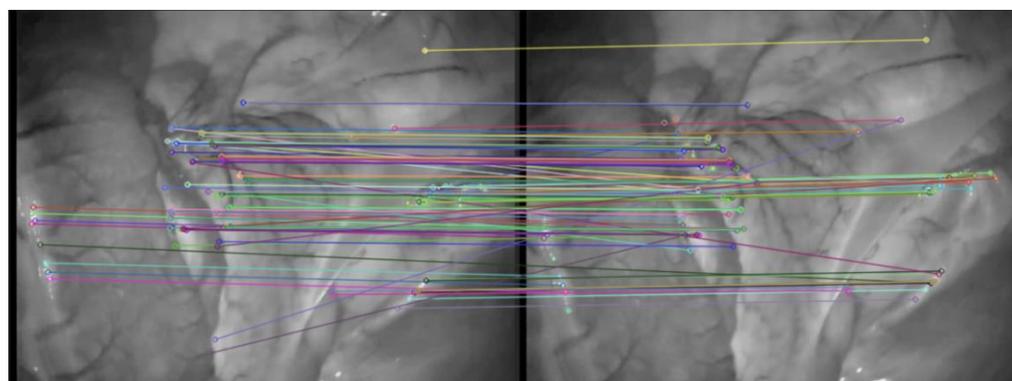


Figure 5. Feature points matching results obtained by AKAZE-ORB.

In order to improve the accuracy of feature matching, this paper uses the symmetry of feature points to quickly eliminate the majority of mis-matched pairs combined with the RANSAC algorithm, as shown in Figure 6.

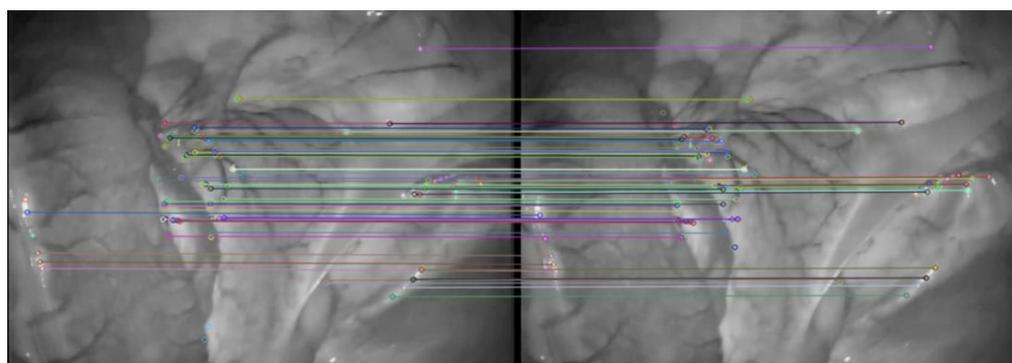


Figure 6. The results of RANSAC algorithm to identify incorrect matches.

In this paper, ORB algorithm and AKAZE-ORB algorithm are used to extract feature points, respectively, as follows: the number of extracted feature points, the number of matching point pairs, the matching rate of abdominal images and running time in the 50th, 100th, 200th, 300th and 450th frames are counted for Dataset1 (uniform interval) and 140th, 280th, 365th, 490th, 540th are counted for Dataset2 (random interval) as shown in Tables 1 and 2, respectively. It can be seen from Table 1 that the running time of AKAZE-ORB algorithm is similar to that of ORB algorithm, but in terms of the number of feature

points extracted, the total number of feature points detected by AKAZE-ORB algorithm is about 1.5 times that of ORB algorithm. In terms of the number of matching point pairs of feature points, AKAZE-ORB algorithm is also higher than ORB algorithm, and the repeatability of feature points is high. In terms of matching accuracy, after eliminating the wrong matching point pairs, the matching rate of AKAZE-ORB is about 10% higher than that of ORB algorithm. After the evaluation of four indicators, it can be concluded that AKAZE-ORB is more suitable for feature extraction of abdominal environment.

Table 1. Comparison of ORB and AKAZE-ORB feature point extraction algorithms for Dataset1.

Frame	Methods	Time (ms)	Initial Match	Correct Match	Matching Rate (%)
50	ORB	76.33	456	305	66.89
	AKAZE-ORB	75.98	531	407	76.72
100	ORB	73.76	450	297	66.00
	AKAZE-ORB	74.19	524	395	75.49
200	ORB	70.23	457	293	64.11
	AKAZE-ORB	69.54	519	380	73.27
300	ORB	72.67	455	301	66.15
	AKAZE-ORB	70.11	521	384	74.76
450	ORB	73.42	457	298	65.21
	AKAZE-ORB	74.92	511	379	74.26

Table 2. Comparison of ORB and AKAZE-ORB feature point extraction algorithms for Dataset2.

Frame	Methods	Time (ms)	Initial Match	Correct Match	Matching Rate (%)
140	ORB	74.18	169	112	66.27
	AKAZE-ORB	72.34	221	156	70.58
280	ORB	70.03	136	88	64.71
	AKAZE-ORB	67.09	172	120	69.77
365	ORB	67.97	131	90	68.70
	AKAZE-ORB	66.61	170	125	73.53
490	ORB	71.54	211	145	69.19
	AKAZE-ORB	68.13	283	207	73.14
540	ORB	69.67	129	86	66.67
	AKAZE-ORB	67.54	172	123	71.51

After extracting the abdominal cavity image feature information, depth cannot be recovered from a single image. The relative depth of the abdominal cavity images needs to be obtained through the continuous motion of the laparoscope to form the parallax angle. The monocular SLAM system uses initialization to estimate the initial position of the laparoscope as the initial value to obtain abdominal point cloud depth information and construct a local abdominal 3D point cloud map.

4. Abdominal Cavity Mapping and Optimization

Compared with the traditional three-dimensional reconstruction method using multi frame static abdominal images, the monocular SLAM system has the ability to optimize pose and eliminate cumulative error. By selecting key frames, using a bag-of-words model and BA optimization, the system reduces the accumulated error in the process of abdominal cavity map construction, and obtains the sparse three-dimensional point cloud on the abdominal cavity surface, which lays the foundation for dense reconstruction.

4.1. Construction of Abdominal Cavity Bag-of-Words Model

Bag of words (bow) [26] is a technology that uses a visual dictionary to convert images into sparse vectors, which enables this paper to process large image data sets more efficiently. Words in the visual dictionary refer to the descriptors of ORB features. A word represents a subset of descriptors of multiple similar features, and the dictionary contains all words. In the SLAM system, the feature is extracted from each key frame and the descriptor is calculated. All the features of the current frame are searched in the dictionary, a word vector is constructed and added to the image database for query. When querying two images, we mainly consider the similarity between them, that is, the spatial distance of word vector. Usually, for the latest key frame, a series of key frames with high similarity are found as loopback candidate frames, and then the key frames with good quality are retained after verification and screening.

In this paper, 1500 sequence images of human body are extracted from the endoscopic video database of Hamlyn, a large number of feature points are generated according to the image data, organized and clustered according to a certain structure, and a vocabulary specially used for minimally invasive surgery is trained. The fork tree structure is simple and practical. It is the best choice to represent the word bag. It has logarithmic query efficiency. It can also query directly from a certain layer according to some known information to improve the query efficiency. Figure 7 shows the structure of the K-ary tree dictionary. Starting from the root node, each layer node is divided into k nodes downward until the set depth d is reached. The leaf nodes stored in the d th layer are clustered words. If you build a dictionary tree with k bifurcation and d depth, the specific process is as follows:

1. The root node represents the set of all features, K-means algorithm is used to cluster into k classes.
2. In the first layer, continue to cluster use K-means algorithm and separate k nodes to get the next layer.
3. On the new layer, cycle the second step until the depth of the tree reaches the d th layer.

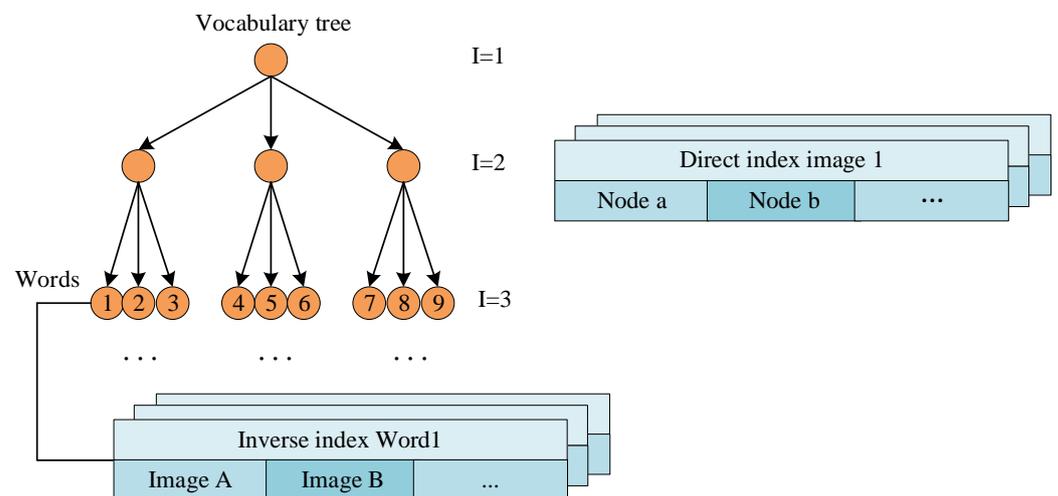


Figure 7. K-tree dictionary structure model.

In the whole tree structure, the leaf layer node is a word, and the intermediate node (cluster center) generated in the process of establishing a dictionary can be used to query words quickly. Each word includes parent node number, whether it is a leaf node, description of sub vector, weight and semantic label. The vocabulary words are the leaf nodes of the tree. The inverse index stores the weight of the words in the images in which they appear. The direct index stores the features of the images and their associated nodes at a certain level of the vocabulary tree.

The word bag vector is sparse and only needs to store the index and value of the non-zero element of the vector. If the sum of two word bag vectors V_1 and V_2 is given, a score D in the interval is obtained by using the L_1 norm, which is defined as the similarity of the two vectors:

$$D(V_1, V_2) = 1 - \frac{1}{2} \left| \frac{V_1}{|V_1|} - \frac{V_2}{|V_2|} \right| \quad (8)$$

In the Formula (16), the greater the value of D , the more similar of V_1 and V_2 . Therefore, by comparing the similarity of word bag vector, if the similarity score reaches the set threshold, it can be considered that the two abdominal images are similar.

When the number of feature points extracted by ORB algorithm is 392 and 456 matches are generated, the violent matching takes 46.62 ms to complete the matching, and the BoW matching takes 40.23 ms. When the number of feature points extracted by AKAZE-ORB algorithm is 587 and 531 matches are generated, the violent matching takes 41.52 ms to complete the matching, and the BoW matching takes 36.18 ms, which means that BoW matching can greatly reduce the time of feature matching.

4.2. BA Optimization

In the process of constructing the abdominal cavity 3D point cloud map, in order to avoid feature information tracking failure, when the current frame extracts less feature information, has low correlation with the historical frame, a new abdominal keyframe needs to be inserted as soon as possible to update the visual correlation map. To ensure abdominal feature tracking steadily, the system in this paper picks redundant keyframes in the process of local abdominal map construction to improve the speed of the 3D texture model. With the continuous addition of the key frames of the abdominal cavity image, the error will be larger and larger when calculating the camera pose and 3D point coordinates of the abdominal cavity space of adjacent frames. In this paper, the BA algorithm [27] is used to construct the least squares problem and solve it iteratively to reduce the cumulative error and realize the optimization of local map.

There are m three-dimensional points in abdominal space, of which the coordinates of a point P_n are $\mathbf{P}_n = [X_n, Y_n, Z_n]^T$ and the pixel coordinates of its projection are $\mathbf{u}_n = [u_m, v_m]^T$. Then, the relationship between pixel position and spatial point position is shown in Formula (9):

$$\mathbf{S}_n \begin{bmatrix} u_m \\ v_m \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{T}_l \begin{bmatrix} X_n \\ Y_n \\ Z_n \\ 1 \end{bmatrix} \quad (9)$$

where \mathbf{S}_n is the lie algebra of the depth Z_n of the point P_n and \mathbf{T}_l is the lie algebra of the camera pose. After conversion to matrix form, Formula (9) is as follows:

$$\mathbf{S}_n \mathbf{u}_n = \mathbf{K}\mathbf{T}_l \mathbf{P}_n \quad (10)$$

There are errors in solving the equation due to the noise of camera observations and unknown pose. Therefore, in this paper, the error summation is transformed into the corresponding least squares problem, and then the optimal camera pose can be obtained.

$$\mathbf{T}_l^* = \operatorname{argmin}_{\mathbf{T}_l} \frac{1}{2} \sum_{n=1}^m \left\| \mathbf{u}_n - \frac{1}{\mathbf{S}_n} \mathbf{K}\mathbf{T}_l \mathbf{P}_n \right\|_2^2 \quad (11)$$

Local optimization makes the re-projection error infinitely close to 0, so as to obtain the optimal camera parameters and the coordinates of three-dimensional space points. Therefore, the BA algorithm is a method to optimize the position and pose parameters of feature points, which can improve the positioning accuracy in abdominal space.

4.3. Local Configuration of Abdominal Cavity Surface

This paper selects Dataset15 (the 15th video) in the Hamlyn laparoscopic video dataset to verify and analyze the feasibility and effectiveness of the point cloud map construction method designed in this paper. Figure 8 shows the effect of traditional ORB algorithm and AKAZE-ORB algorithm on sparse reconstruction of abdominal surface, where the green mark is the trajectory of laparoscopy, the red mark points represent the map points being reconstructed, and the black points represent the map points after reconstruction. The blue line indicates the pose of the camera at the time of key frames, which constitute the motion trajectory of the camera. It can be seen that the monocular SLAM abdominal 3D reconstruction system can obtain the 3D reconstruction point cloud based on abdominal feature points and the motion trajectory of laparoscopy, but the obtained point cloud is very sparse. The AKAZE-ORB algorithm obtains a denser point cloud effect than the original system, but it is still unable to obtain a dense abdominal point cloud map.

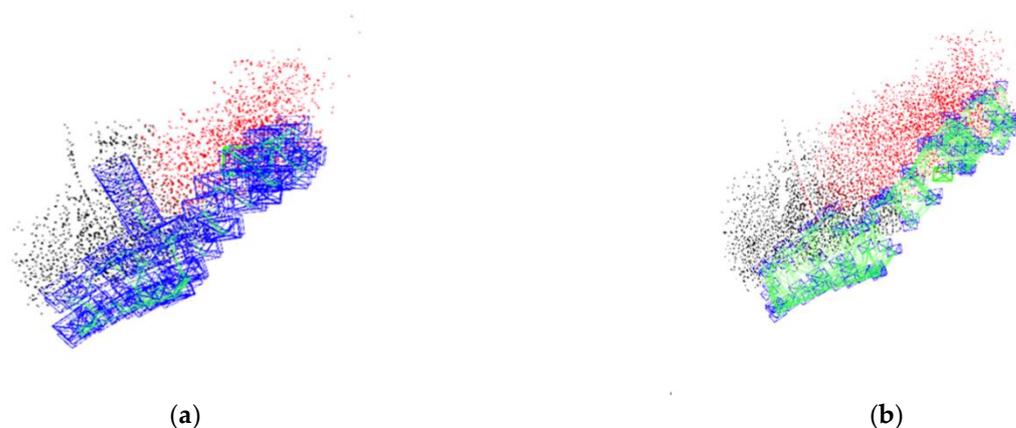


Figure 8. Local configuration of the abdominal surface. (a) ORB; (b) AKAZE-ORB.

5. Poisson Surface Reconstruction and Texture Mapping

Although the 3D reconstruction of an abdominal cavity surface based on the SLAM system can obtain real-time endoscope motion trajectory and 3D reconstruction point cloud based on feature points, sparse 3D point cloud surface cannot obtain dense 3D reconstruction effect. Therefore, the dense abdominal cavity map is obtained by Poisson surface reconstruction and texture mapping.

The approach of Poisson surface reconstruction [28] is based on the observation that the (inward pointing) normal field of the boundary of a solid can be interpreted as the gradient of the solid's indicator function [29]. Thus, given a set of oriented points sampling the boundary, and construct the Poisson equation $\Delta \tilde{x} = \nabla \cdot \nabla \tilde{x} = \nabla \cdot \vec{v}$. For the problem of uncertain position, the projection on the function space can best approximate the projection, and then the minimum value \tilde{x} of the following equation can be obtained.

$$\sum_{o \in O} \left\| \left\langle \Delta \tilde{x} - \nabla \cdot \vec{V}, F_o \right\rangle \right\|^2 = \sum_{o \in O} \left\| \left\langle \Delta \tilde{x}, F_o \right\rangle - \left\langle \nabla \cdot \vec{V}, F_o \right\rangle \right\|^2 \quad (12)$$

Finally, the surface model reconstruction is obtained by extracting the isosurface from the indicator function. The position of the isosurface should be close to the position of the input sample, and then the Poisson surface can reflect the real surface of the point cloud model to be reconstructed.

In order to verify the performance-feature extraction and matching of AKAZE-ORB algorithm proposed in this paper, a Hamlyn laparoscopic video data set is used to construct 3D sparse point cloud map. Figures 9 and 10 use Dataset1 and Dataset2, respectively, to reconstruct the abdominal Poisson surface obtained by different algorithms. Poisson surface reconstruction is to make all points as close to the implicit equation as possible. Therefore,

it changes the original vertex data in the process, which is robust to external points and the generated surface is very smooth. Figures 8 and 9 show the 3D reconstruction results of abdominal cavity of the classical SLAM system, and (b) show the 3D reconstruction results of abdominal cavity of the improved SLAM (ISLAM) system in this paper. From the reconstruction results, it can be seen that the abdominal mesh model reconstructed by the classical SLAM system has some holes, some surface mesh errors, and the reconstructed surface has uneven parts. For example, the areas in red are sparse and sunken parts of the mesh which leaves obvious gaps in the reconstructed abdominal model. However, the reconstructed model surface by our ISLAM system is smooth, and the relevant contour details are retained, which reduces the generation of holes in the reconstructed surface. The meshes are more dense in the red area, which characterize the geometry of the abdominal surface better, makes the abdominal model more real, smooth and delicate, and realizes more accurate reconstruction of the abdominal model.

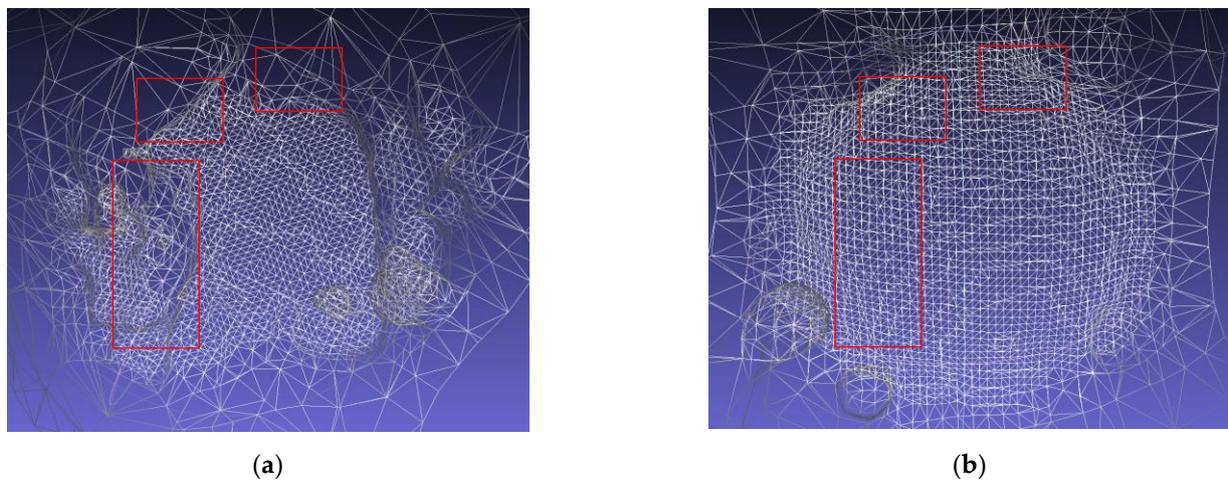


Figure 9. Results of Poisson surface reconstruction with different algorithms for Dataset1. (a) SLAM; (b) ISLAM.

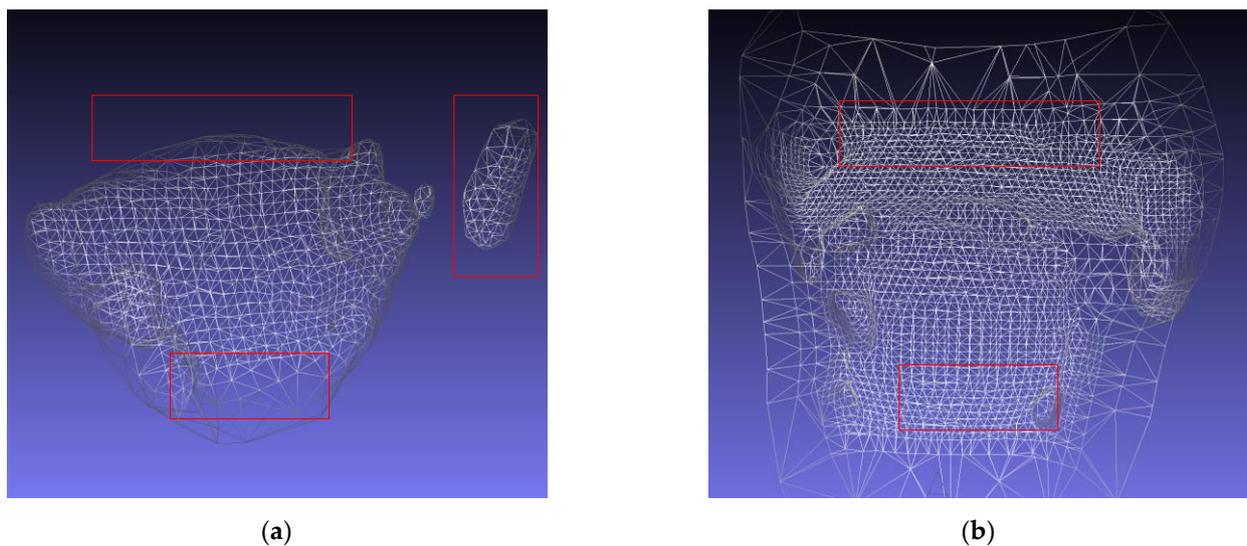


Figure 10. Results of Poisson surface reconstruction with different algorithms for Dataset 2. (a) SLAM; (b) ISLAM.

Figures 11 and 12 are the abdominal reconstruction results of two data sets, respectively. From the texture mapping results, it can be seen that the reconstruction algorithm after feature extraction and matching by using the AKAZE-ORB algorithm, the effect of texture

mapping is better than that of classical SLAM system. The reconstruction effect integrity of classical SLAM system is low, and it is difficult to characterize the features of blood vessels and tissues, while the reconstruction surface of ISLAM system is smooth, natural and realistic with fewer holes for three-dimensional visualization of the abdominal model. Additionally, the texture after mapping is also transitional and natural, with strong realism.

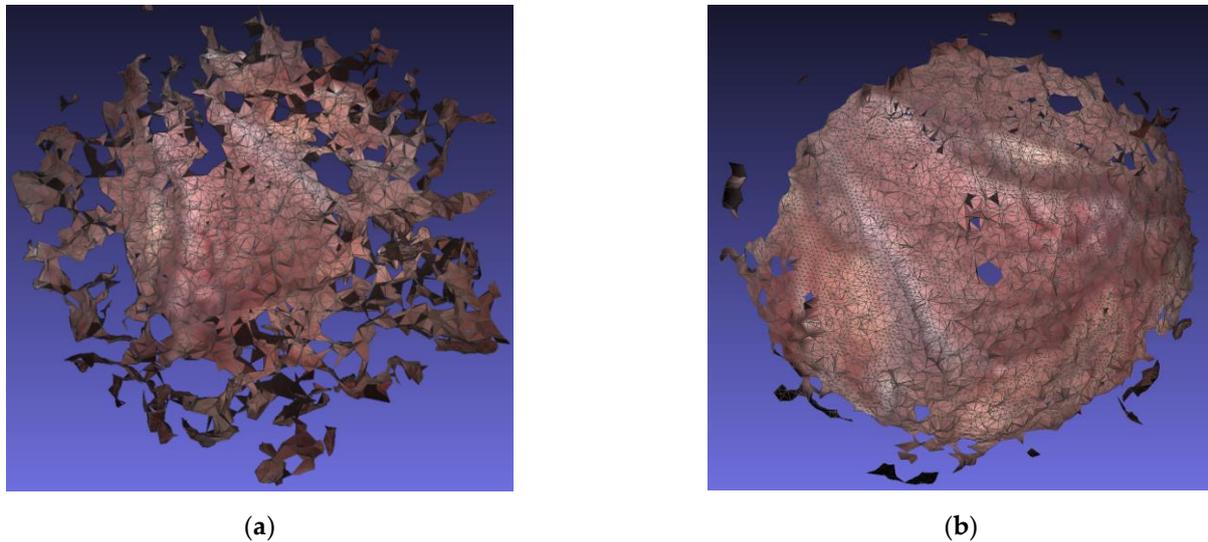


Figure 11. 3D reconstruction of the abdominal cavity for Dataset 1. (a) SLAM; (b) ISLAM.

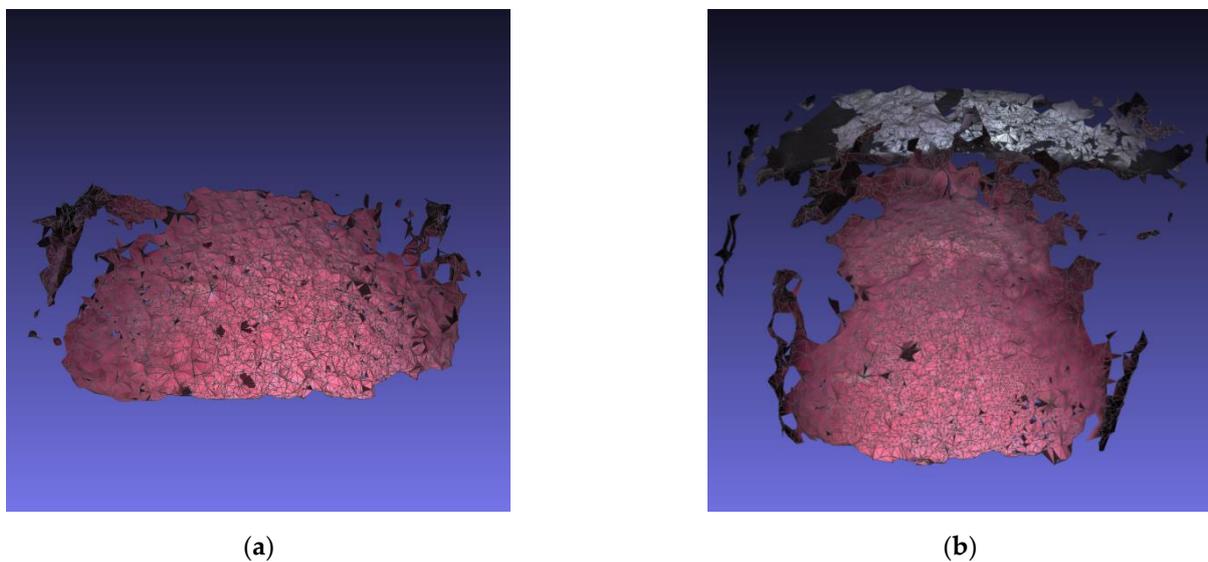


Figure 12. 3D reconstruction of the abdominal cavity for Dataset 2. (a) SLAM; (b) ISLAM.

6. Conclusions

This paper proposed a novel 3D texture reconstruction of abdominal cavity based on monocular vision SLAM for minimally invasive surgery. CLAHE algorithm is introduced into abdominal image preprocessing to enhance the contrast between blood vessels and background. In the aspect of feature points extraction and matching, AZAKE-ORB algorithm improves the registration accuracy and density. Combined with Poisson surface reconstruction algorithm, the surface of abdominal cavity 3D model is smooth and has a strong sense of reality. In addition, the visual features of medical images are extracted and used to generate a bag-of-words containing abdominal feature information, which makes the comparison of similarity between abdominal images easier and improves the robustness

and real-time performance of loop detection. This paper uses the Hamlyn dataset medical image database to design the registration accuracy and densification evaluation experiment. The experimental results show that compared with the classical slam system, the system improves the registration accuracy of feature points, improves the densification, and the visual effect of dense point cloud reconstruction is more realistic.

In the 3D texture reconstruction of abdominal space, the proposed system cannot deal with the movement and deformation of internal organs and eliminate vibration effects of heartbeat, respiratory and surgical factors on non-rigid internal cavity surfaces. Therefore, the future work would study how to combine the prior knowledge of the internal cavity and make full use of various medical imaging techniques to build a high-precision 3D texture reconstruction of the abdominal cavity in a dynamic environment.

Author Contributions: Conceptualization, A.W. and H.W.; methodology, software, validation, R.X., K.X., J.Z. and Y.Z.; writing—review and editing, H.W. and A.W.; supervision, Y.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China under Grant NSFC-61671190.

Data Availability Statement: The data presented in this study are available on <http://hamlyn.doc.ic.ac.uk/vision/> (accessed on 5 December 2021).

Acknowledgments: We thank Kaiyuan Jiang for his valuable comments and discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Rosen, M.; Ponsky, J. Minimally invasive surgery. *Endoscopy* **2001**, *33*, 358–364. [[CrossRef](#)] [[PubMed](#)]
- Fei, B. The Method and Development of Computer-Assisted Surgery. Foreign Medical Sciences. *Biomed. Eng. Fascicle* **2015**, *20*, 624–631.
- Fatima, S.; Haleem, A.; Bahl, S.; Javaid, M.; Mahla, S.; Singh, S. Exploring the significant applications of Internet of Things (IoT) with 3D printing using advanced materials in medical field. *Mater. Today Proc.* **2021**, *45*, 4844–4851. [[CrossRef](#)]
- Ashima, R.; Haleem, A.; Bahl, S.; Javaid, M.; Mahla, S.; Singh, S. Automation and manufacturing of smart materials in Additive Manufacturing technologies using Internet of Things towards the adoption of Industry 4.0. *Mater. Today Proc.* **2021**, *45*, 5081–5088. [[CrossRef](#)]
- Wu, C.; Narasimhan, S.G.; Jaramaz, B. A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes. *Int. J. Comput. Vis.* **2010**, *86*, 211–228. [[CrossRef](#)]
- Ciuti, G.; Visentiniscarzarella, M.; Dore, A.; Menciassi, A.; Dario, P.; Yang, G. Intra-operative monocular 3D reconstruction for image-guided navigation in active locomotion capsule endoscopy. In Proceedings of the Fourth IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics, Roma, Italy, 24 June 2012; pp. 768–774.
- Visentini-Scarzanella, M.; Stoyanov, D.; Yang, G. Metric depth recovery from monocular images using Shape-from-Shading and specularities. In Proceedings of the IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September 2012; pp. 25–28.
- Tokgozoglul, H.N.; Meisner, E.M.; Kazhdan, M.; Hager, G.D. Color-based hybrid reconstruction for endoscopy. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16 June 2012; pp. 8–15.
- Goncalves, N.; Roxo, D.; Barreto, J.; Rodrigues, P. Perspective shape from shading for wide-FOV near-lighting endoscopes. *Neurocomputing* **2015**, *150*, 136–146. [[CrossRef](#)]
- Lei, L.; Li, J.; Liu, M.; Hu, X.; Zhou, Y.; Yang, S. Shape from shading and optical flow used for 3-dimensional reconstruction of endoscope image. *Acta Otolaryngol.* **2016**, *136*, 1190–1192. [[CrossRef](#)] [[PubMed](#)]
- Turan, M.; Pilavci, Y.Y.; Ganiyusufoglu, I.; Araujo, H.; Konukoglu, E.; Sitti, M. Sparse-then-Dense Alignment based 3D Map Reconstruction Method for Endoscopic Capsule Robots. *Mach. Vis. Appl.* **2018**, *29*, 345–359. [[CrossRef](#)]
- Hu, Q.; Cai, W.; Lu, D.; Jiang, M. Survey on Monocular Visual Inertial SLAM Algorithms. *Softw. Guide* **2020**, *19*, 275–280.
- Lin, B.; Sun, Y.; Sanchez, J.E.; Qian, X. Efficient Vessel Feature Detection for Endoscopic Image Analysis. *IEEE. Trans. Biomed. Eng.* **2015**, *62*, 1141–1150. [[CrossRef](#)] [[PubMed](#)]
- Yang, B. Gastrointestinal Endoscope Localization and Feature Map Reconstruction Based on Monocular SLAM. Master's Thesis, Nanchang University, Nanchang, China, 28 May 2016.
- Mahmoud, N.; Collins, T.; Hostettler, A.; Soler, L.; Doignon, C.; Montiel, J.M.M. Live Tracking and Dense Reconstruction for Handheld Monocular Endoscopy. *IEEE Trans. Med. Imaging* **2019**, *38*, 79–89. [[CrossRef](#)] [[PubMed](#)]

16. Xie, C.; Yao, T.; Wang, J.; Liu, Q. Endoscope localization and gastrointestinal feature map construction based on monocular SLAM technology. *J. Infect. Public Health* **2020**, *13*, 1314–1321. [[CrossRef](#)] [[PubMed](#)]
17. Lamarca, J.; Parashar, S.; Bartoli, A.; Montiel, J.M.M. DefSLAM: Tracking and Mapping of Deforming Scenes from Monocular Sequences. *IEEE Trans. Robot.* **2020**, *37*, 291–303. [[CrossRef](#)]
18. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. In *Graphics GEMS IV*, 1st ed.; Morgan Kaufmann: Pittsburgh, PA, USA, 1994; pp. 474–485. [[CrossRef](#)]
19. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6 November 2011; pp. 2564–2571.
20. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the European Conference on Computer Vision, Berlin, Germany, 5 July 2006; pp. 430–443.
21. Calonder, M.; Lepetit, M.; Ozuysal, M.; Trzcinski, T.; Strecha, C.; Fua, P. BRIEF: Computing a local binary descriptor very fast. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1281–1298. [[CrossRef](#)] [[PubMed](#)]
22. Alcantarilla, P.F.; Solutions, T. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1281–1298.
23. Grewenig, S.; Weickert, J.; Bruhn, A. From box filtering to fast explicit diffusion. In Proceedings of the Dagm Conference on Pattern Recognition, Berlin, Germany, 22–24 September 2010; pp. 533–542.
24. Sharma, S.K.; Jain, K. Image Stitching using AKAZE Features. *J. Indian Soc. Remote* **2020**, *48*, 1389–1401. [[CrossRef](#)]
25. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
26. Gálvez-López, D.; Tardos, J.D. Bags of binary words for fast place recognition in image sequences. *IEEE Trans. Robot.* **2012**, *28*, 1188–1197. [[CrossRef](#)]
27. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—a modern synthesis. In Proceedings of the International Workshop on Vision Algorithms, Berlin, Germany, 12 April 2002; pp. 298–372.
28. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson surface reconstruction. In Proceedings of the Fourth Eurographics Symposium on Geometry Processing, Sardinia, Italy, 26 June 2006; pp. 61–70.
29. Kazhdan, M.; Hoppe, H. Screened poisson surface reconstruction. *ACM Trans. Graph.* **2013**, *32*, 1–13. [[CrossRef](#)]