2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Endoscopic Navigation Based on Three-dimensional Structure Registration

Minghui Han, Yu Dai, and Jianxun Zhang

Abstract—**Surgical navigation is challenging on complicated multi-branch structures such as intrarenal collecting systems or bronchi. The objective of this work is to help surgeons quickly establish the corresponding relationship between intraoperative endoscopic images and preoperative CT data. An endoscopic navigation method is proposed based on three-dimensional structure registration. It mainly includes three sections. First, a reconstruction method is presented to obtain three-dimensional information of porous structures from endoscopic images. It combines image enhancement, structure-from-motion and template matching. Second, a hole search strategy based on slicing is given for detecting and extracting three-dimensional porous structures from CT data. Third, a similarity measurement algorithm is developed for registering endoscopic images to CT data. The performance of this work is evaluated on the data from the ureteroscopic holmium laser lithotripsy and the results show its accuracy, robustness and time cost.**

I. Introduction

Endoscopy plays an important role in modern surgery. Its introduction enables the diagnosis and the treatment of diseases in a noninvasive or minimally invasive way. So that the efficiency of the operation is improved and the suffering of patients is relieved. Surgeons manipulate the endoscope into the body for diagnosis and surgery. Although endoscopes are widely used in clinical examination, it is still difficult to operate them for novices. The main reasons are as follows: (1) Due to the size limitation, the visual field of the endoscope is narrow and it is easy to miss information during medical examination [1]; (2) Surgeons could not estimate the corresponding relationship between endoscopic images and preoperative CT images because of the existence of porous structures and the similarity of images. As a result, it's challenging to reach lesions accurately and quickly.

Surgical navigation can provide practical assistance for surgeons based on the fusion of the preoperative information and intraoperative feedback. Related works mainly follow two lines of approaches: the external device-based navigation and the image registration-based navigation.

The external device-based navigation typically uses external devices or sensors to locate the endoscope. For example, electromagnetic-based navigation methods [2-6] use electromagnetic sensors and external electromagnetic transmitters to track. Stereovision-based navigation methods [7, 8] use two or more external cameras to perceive depth information and stereo structures. Optical navigation methods [9] usually require the optical position sensor including an optical lens, light source, and digital signal processor to track the motion of the endoscope. Although there have been many achievements about the external device-based navigation, it also suffers from several bottlenecks. On the one hand, the installation of external sensors is difficult because of space limitation. On the other hand, for the widely-used electromagnetic tracking, various metal medical devices in the external environment will affect its stability.

The image registration-based navigation avoids the dependence on external devices in general. Deguchi et al.[10] reported a selective image similarity measurement to register images based on the extraction of characteristic structures. Luo et al.[11] proposed a tracking method for bronchoscopic navigation combining epipolar geometry analysis, Kalman filtering, and intensity-based image registration. Shen et al.[12] introduced the depth information to two-dimensional image registration, which improved its robustness. They also proposed an airway descriptor based on shape context to encode both the structural characteristics of bifurcations and their spatial distribution [13]. The basic thought of the above methods is to register endoscopic images and virtual endoscopic images. However, it remains many challenges. Firstly, it requires a continuous video stream and assumes that the endoscope is in a constant state of acceleration. But it is difficult to satisfy these conditions in practical application. Secondly, bubbles, bleeding and floating impurities in the viscera will seriously affect the accuracy of intensity-based image registration. Thirdly, depth-based or shape-based image registration frameworks improve the accuracy of registration but are more computationally expensive.

For a multi-branch structure similar to the intrarenal collecting system, the essential task of navigation is to help doctors distinguish the corresponding relations between the holes in intraoperative endoscopic images and the holes on the preoperative CT model. In this paper, a navigation method is proposed based on three-dimensional (3D) structure registration. Porous structures are reconstructed from CT data and endoscopic images respectively. Then a registration method is designed to establish the correspondence between them. This method does not require any external tools and performs well on multi-branch structures.

II. 3D Reconstruction from Endoscopic Images

The purpose of 3D reconstruction is to obtain structural information of holes from intraoperative endoscopic images. The process of reconstruction proposed in this paper can be divided into three major steps as follows: (a) image preprocessing; (b) structure from motion; and (c) extraction of porous structures.

*Research supported by the National Key R&D Program of China (Grant No. 2017YFB1302803), the Natural Science Foundation of Tianjin (Grant No. 18JCYBJC18800) and the Fundamental Research Funds for the Central Universities, Nankai University (Grant No. 63201193).

The authors are with the College of Artificial Intelligence, Nankai University, Tianjin 300350, China (e-mail: hanmh@mail.nankai.edu.cn; daiyu@nankai.edu.cn; zhangjx@nankai.edu.cn;).

A. Image Preprocessing

Since the low resolution of the endoscope and sparse texture of the surface in the human viscera make it difficult to extract feature points, a multi-scale image enhancement method is proposed to preprocess endoscopic images based on weighted guided filtering. The flow of the algorithm is shown in Fig. 1.

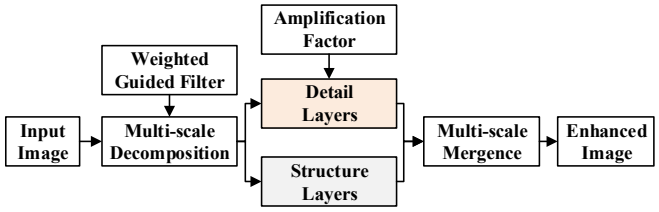

Figure 1. Flowchart of image enhancement.

The main steps of image preprocessing are as follows:

Multi-scale Decomposition: Weighted guided image filter (WGIF) proposed in [14] is used because it has the advantage of adaptive edge-preserving. The structure layer L_1 and the detail layer B_1 are obtained by decomposition on the first scale

$$\begin{cases} L_1 = WGIF(I) \\ B_1 = I - L_1 \end{cases}, \tag{1}$$

where $WGIF(.)$ is the weighted guide filter function, and I is the original input image. By continuing the same decomposition operation, structure layers and detail layers on multi scales can be obtained:

$$\begin{cases} L_{i+1} = WGIF(L_i) \\ B_{i+1} = L_i - L_{i+1} \end{cases}, i = 1, 2, \cdots, N_D, \tag{2}$$

where i denotes the decomposition scale, and N_D is its maximum value.

Multi-scale Mergence: Multi-scale detail enhancement can be achieved as

$$Z = \alpha \sum_{i=1}^{N_D} B_i + L_{N_D}, \tag{3}$$

where Z is the enhanced image, and α is an amplification factor.

B. Structure from Motion

The classical structure-from-motion (SFM) algorithm is introduced to reconstruct the sparse point cloud. It mainly consists of three steps: scale-invariant feature transform (SIFT) [15] based feature extraction and matching, camera motion recovery and the sparse point cloud generation.

C. Extraction of Porous Structures

Bleeding and vascular textures are easy to interfere with hole recognition and template matching because they also present low intensities in grayscale images. However, on the red channel of the image, the intensities of red areas are high, which is obviously different from the dark areas where the holes are located. Therefore, the following steps are performed on the red channel of the image.

Hole recognition: Keyframes containing hole information are selected according to the low intensities of holes in imaging. A pixel set of a hole's edge is extracted from images that are processed based on the OTSU algorithm[16].

Epipolar constrained search: Each pixel p of a hole's edge is searched along the epipolar line on neighbor keyframes based on template matching with the range of inverse depths [d_{min}, d_{max}]. The pixel with the maximum similarity value based on zero-mean normalized cross-correlation (ZNCC) [17, 18] is selected as the best match of p. The epipolar line is computed from the fundamental matrix recovered by SFM. $d_{min}=d-2\sigma$ and $d_{max}=d+2\sigma$ are adaptively defined for each keyframe, in which d and σ are depth mean and variance of its sparse point cloud.

Reconstruction: Under the condition that all the edge points and their matching points are known, the 3D coordinates of the edge points of the hole can be obtained through triangulation.

Fitting: A plane is fitted to the 3D point cloud of each hole using the least-squares method. A circle is fitted to the 2D points projected onto the fitted plane. The radius of the fitted circle is regarded as the radius of the hole; the center of the fitted circle is regarded as the center of the hole; the normal vector of the fitted plane is regarded as the normal vector of the hole.

III. 3D Reconstruction from CT Data

In order to extract the 3D information of key holes from CT data, a method is given based on slicing. The operation of slicing begins at a location within the cavity of the viscus in multiple directions. The position and size of each hole can be determined by analyzing the content of each slice image.

A. Search Initialization

The initialization steps are as follows:

[Step 1] The 3D multi-branch model is reconstructed from the CT sequence $\{T_1, T_2, \ldots, T_N\}$ using the 3D region growing segmentation and the marching cubes algorithm [19].

[Step 2] A point Q is picked up on the cavity surface of the 3D multi-branch model by mouse interaction.

[Step 3] Assuming that Q is included in T_i, the sizes of the regions in T_{i-1}, T_i, and T_{i+1} are calculated.

[Step 4] The center point of the region with the maximum size is extracted, and its corresponding 3D coordinates on the multi-branch model are set as the initial position Q_{init} of hole search.

B. Hole Search

The search strategy proposed in this paper is based on slicing the 3D multi-branch model in multiple directions. Six directions are used: the positive and negative directions of the X, Y and Z axes in the coordinate system of the CT model. For each direction, slicing marches with fixed step length. Fig. 2 takes the search in the positive direction of the X-axis as an example to show details. The main steps are as follows:

Start slicing: Slicing begins from the initial position Q_{init} in the positive direction of the X-axis in fixed step length.

Extract regions: All regions are extracted from the i-th and (i+1)-th slice images and their sizes are stored according

to their corresponding relations. E_i^k denotes the size of the region stored at the position ID k from the i-th slice image.

Search for holes: The process shown in Fig. 2 is followed to find out each hole. The size of the largest region from the slice image at the initial position Q_{init} is recorded as G. μ_1 and μ_2 ($\mu_1<\mu_2$) are defined, which means that the radius of the hole is limited to a certain range during searching. μ_3 is defined to determine whether the difference between E_i^k and E_{i+1}^k is small enough to be recognized as a hole. The search at position ID k will be finished if E_{i+1}^k is too small or a hole has been detected here.

For the other five directions, processes similar to Fig. 2 are adopted. After completing searches in all six directions, the normal vectors and radii of all holes in the 3D multi-branch model can be obtained eventually.

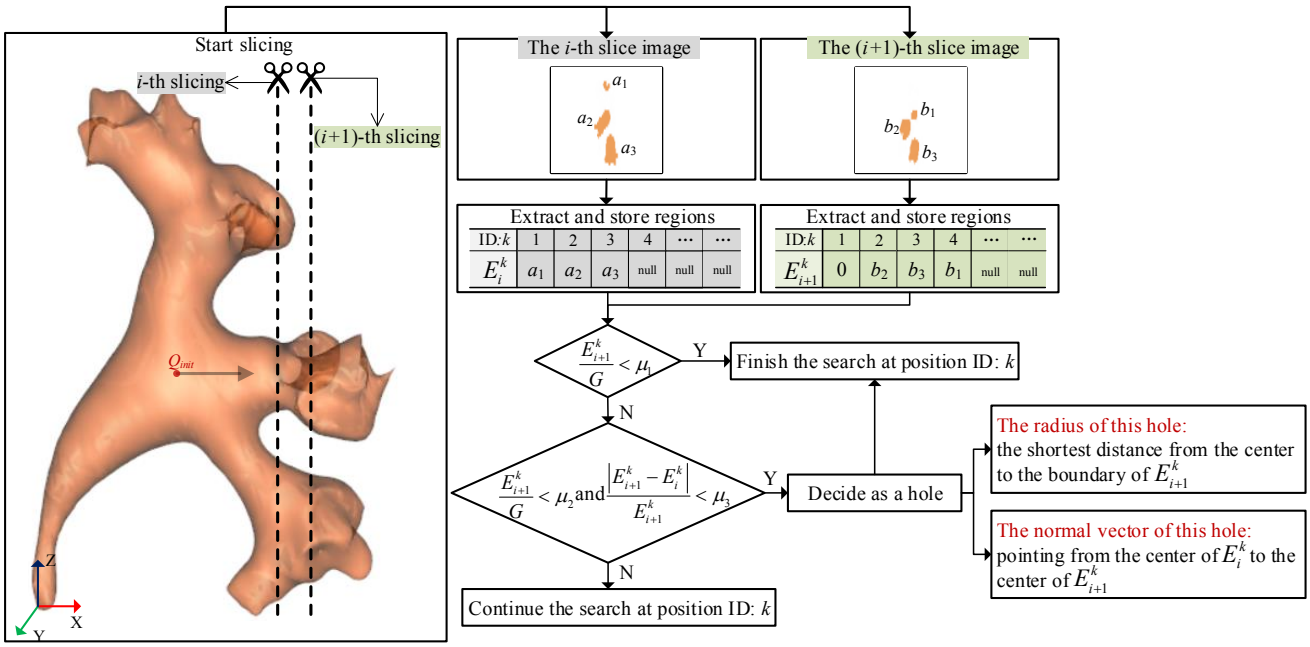

Figure 2. Flowchart of hole search.

IV. 3D Structural Characteristics-based Hole Registration

The objective of endoscopic navigation in this work is to register the holes from endoscopic video to the CT data. A 3D structural descriptor is proposed in this section to characterize both the sizes of holes and the location association between them. Since the reconstruction based on SFM is scale-free, the registration method proposed in this paper is based on some coefficients that are not affected by the scale factor. The reconstruction of a single hole cannot obtain enough information, so it is needed to reconstruct at least two holes from endoscopic images for the purpose of navigation. The following contents are based on the condition that two holes have been reconstructed in section II.

A. Descriptor

Fig. 3 is a schematic diagram of a double-hole structure. A descriptor V is defined to describe its structural characteristics:

$$V=[S,A,D], \tag{4}$$

$$\begin{cases} S=\dfrac{\max(r_1,r_2)}{\min(r_1,r_2)} \\ A=\theta \\ D=\dfrac{l}{|r_1-r_2|} \end{cases}, \tag{5}$$

where S, A, and D respectively describe the size, angle, and distance information of the double-hole structure, r_1 and r_2 are radii of the two holes, θ is the angle between the two normal vectors, l is the perpendicular distance between the two normal vectors. V_{Video}=[S_{Video}, A_{Video}, D_{Video}] is computed from endoscopic video, and V_{CT}^i=[S_{CT}^i, A_{CT}^i, D_{CT}^i] is computed from CT data. The superscript i (i=1,2,…,M) represents the i-th pair of holes extracted from CT data. Assuming that N_h holes are extracted from CT data totally, $M=N_h(N_h-1)/2$ is calculated based on "permutation and combination".

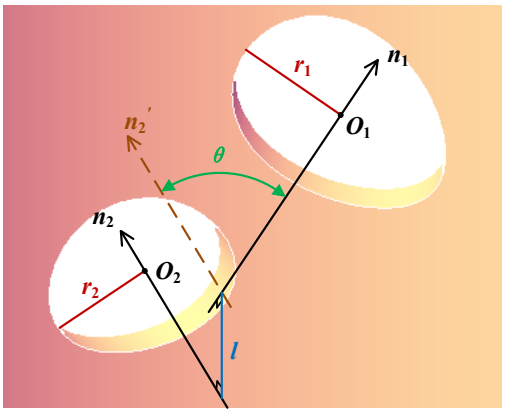

Figure 3. 3D structure of a pair of holes. O_1, O_2 are centers of the two holes. Two normal vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$ are in different planes. The auxiliary line $\boldsymbol{n}_2'$ is parallel to $\boldsymbol{n}_2$ and intersecting with $\boldsymbol{n}_1$. The angle between $\boldsymbol{n}_1$ and $\boldsymbol{n}_2'$ represents the angle between two normal vectors.

B. Similarity Measurement

The similarity measurements of the structural characteristics matching based on the size, angle, and distance information are separately defined as

$$\phi(S_{CT}^i, S_{Video}) = 1 - \frac{\left|S_{CT}^i - S_{Video}\right|}{\max\limits_{j=1,2...M}\left\{\left|S_{CT}^j - S_{Video}\right|\right\}}$$

$$\phi(A_{CT}^i, A_{Video}) = 1 - \frac{\left|A_{CT}^i - A_{Video}\right|}{\max\limits_{j=1,2...M}\left\{\left|A_{CT}^j - A_{Video}\right|\right\}}, \quad (6)$$

$$\phi(D_{CT}^i, D_{Video}) = 1 - \frac{\left|D_{CT}^i - D_{Video}\right|}{\max\limits_{j=1,2...M}\left\{\left|D_{CT}^j - D_{Video}\right|\right\}}$$

where each similarity measurement ϕ is within the range of [0,1]. The closer the value of ϕ is to 1, the higher the similarity between the two double-hole structures is. Then the total similarity between two double-hole structures is estimated by

$$\varphi(V_{CT}^i, V_{Video}) = \frac{\phi(S_{CT}^i, S_{Video}) + \phi(A_{CT}^i, A_{Video}) + \phi(D_{CT}^i, D_{Video})}{3}, \quad (7)$$

where φ is also within the range of [0,1]. The match with the maximum φ is taken as the best match and eventually used as the output of the registration process.

V. Experimental Results

The proposed endoscopic navigation framework was implemented using C++ and its performance was evaluated by using the real data from the ureteroscopic holmium laser lithotripsy. The pixel resolution of the Olympus URF-V2 ureteroscope used in the surgery is 640×480. Accuracy, robustness and time cost of the method are discussed in detail as follows.

A. Accuracy Analysis

Fig. 4 shows the process and results of reconstruction from endoscopic images. Through testing, α=2 and N_D=3 are determined as suitable parameters of the image enhancement algorithm. Compared with the original image, the number of feature points on the enhanced image is increased by about 7 times. It is proved that the image enhancement algorithm in this paper plays an important role in feature extraction. In order to show the reconstruction results, we fit the surface of the extracted double-hole structure, which is shown in the subgraph "Surface Fitting" of Fig. 4. It can be seen that the reconstructed 3D hole structure is very similar to the structure shown in the endoscopic video.

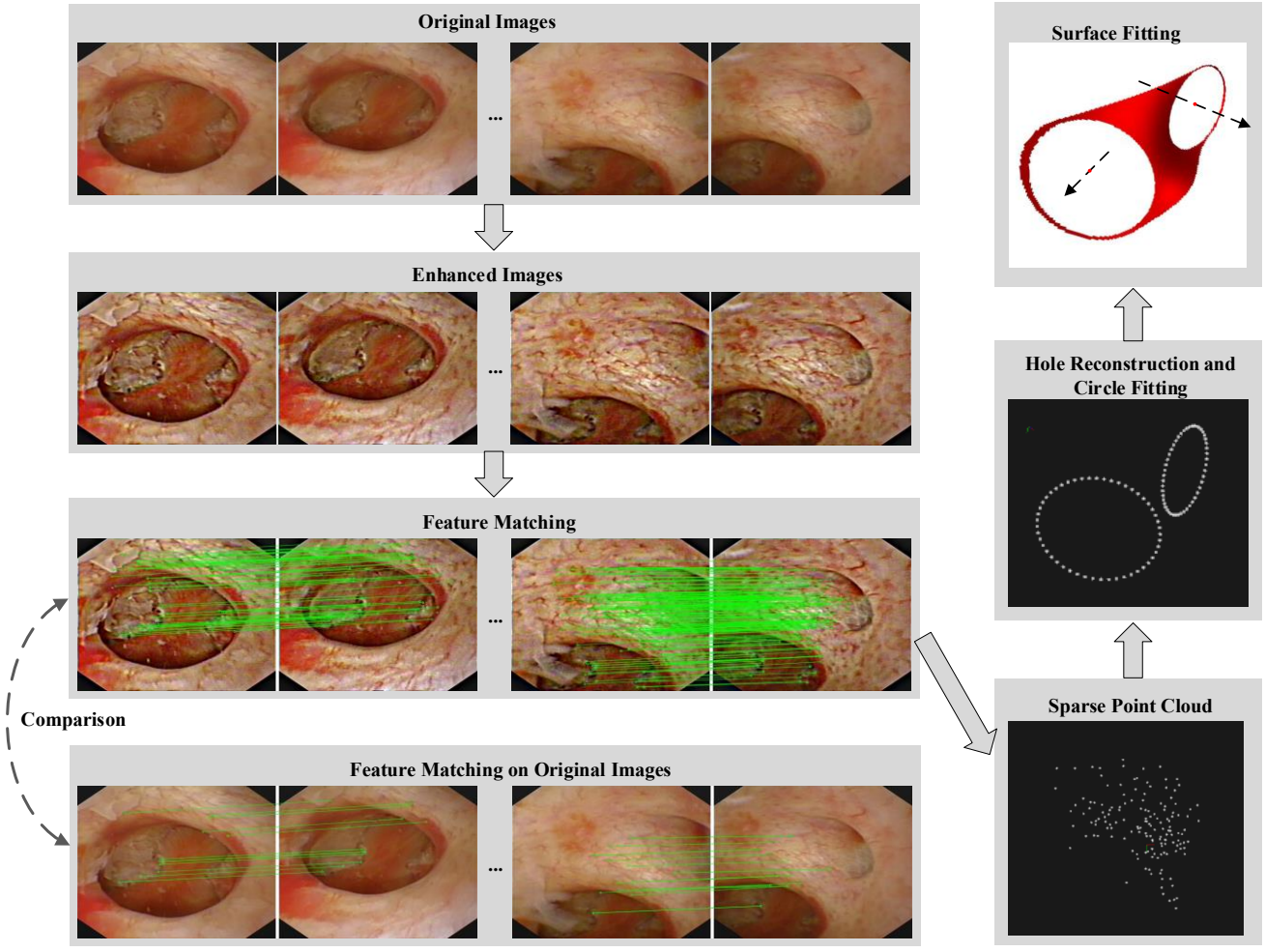

Figure 4. Process and results of reconstruction from endoscopic images.

Fig. 5 shows the results of reconstruction and hole search from CT data. The hole search algorithm was performed on the renal pelvis with parameters μ_1=0.1, μ_2=0.25, and μ_3=0.15. Four red lines on the 3D CT model represent the normal vectors of the four holes separately. The number of holes extracted N_h=4, so M=6.

Table I shows the results of the similarity measurements during the 3D structural characteristics-based registering. The first column of the table denotes the combinations of 2 from 4 holes searched from CT data. The second column of the table gives the similarity φ calculated by our proposed method. It can be seen that the similarity of "2nd and 3rd holes" is maximum, so it is selected as the output of the registration algorithm. The result of the registration is displayed visually in Fig. 6. It can be seen clearly that the registration result is accurate.

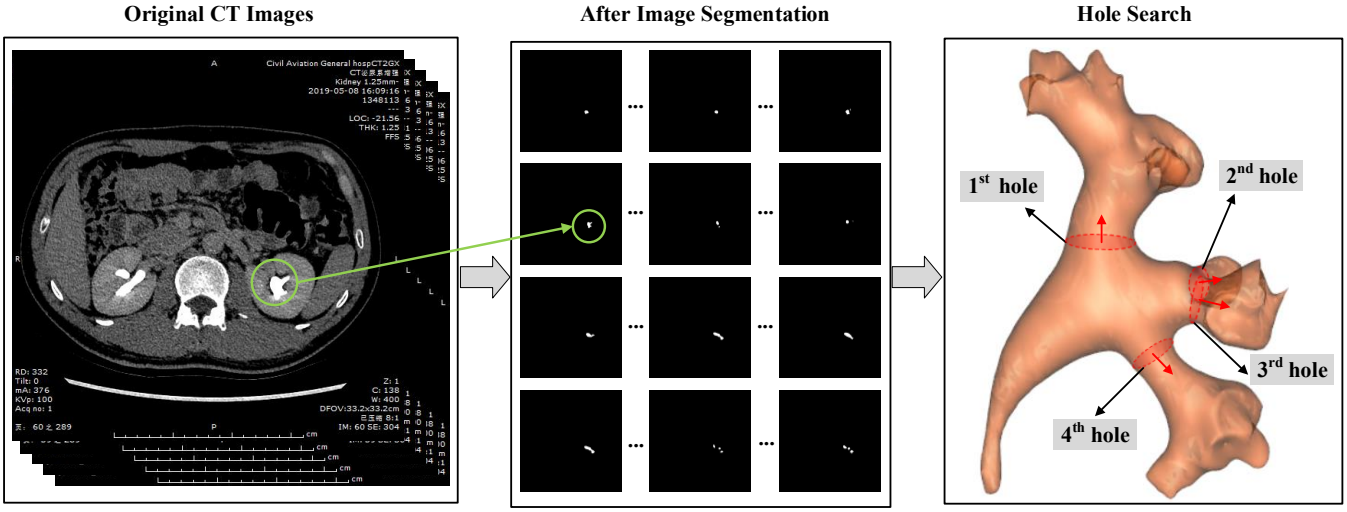

Figure 5. The results of reconstruction and hole search from CT data.

TABLE I. RESULTS OF MATCHING

Combinations of 2 from 4 in CT	Similarity
1st and 2nd holes	0.45
1st and 3rd holes	0.78
1st and 4th holes	0.58
2nd and 3rd holes	0.92
2nd and 4th holes	0.49
3rd and 4th holes	0.15

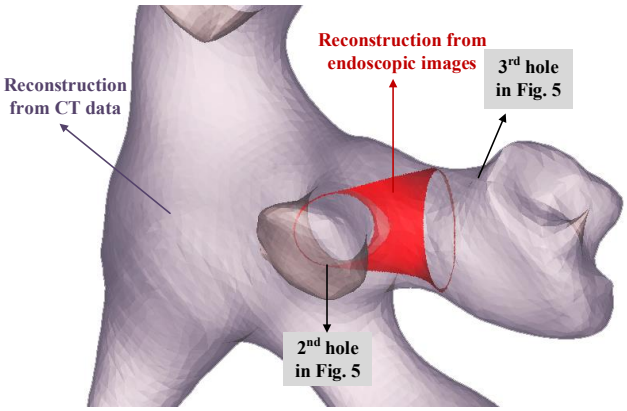

Figure 6. Display of registration result.

B. Robustness Analysis

A set of experiments using noisy images have been done to test the robustness of our method. Pepper and salt noises with different densities were added to endoscopic images. Then the mean and variance of similarity measurements were computed through multiple experiments. Fig. 7 shows the experimental results. It can be seen that the proposed method can consistently obtain the correct output when the noise density is less than 0.010.

C. Time Cost Analysis

In order to prove the real-time property, the time cost of the proposed method has been compared to that of the methods in [12] and [13]. The time required to complete registration was used as an indicator for comparing. For methods in [12] and [13], registration occurs per frame. But for the method in this paper, registration occurs after the reconstruction of a pair of holes.

Table II lists the average time cost of each method without code optimization. It can be seen that our proposed method costs less time even in the case of the lower computer configuration. For the methods in [12] and [13], registration requires multiple interactions with CT data. However, in our method, the reconstruction and hole searching from CT data can be completed offline before the operation, so that the time cost is reduced.

TABLE II. COMPARISON OF TIME COST

Method	PC configuration	Time cost(s)
The method in [12]	3 GHz Intel Core i7 Processor 8 GB Memory	53.30
The method in [13]	3.40 GHz Intel Core i7 Processor 8 GB Memory	3.70
Our method	3.30 GHz Intel Core i5 Processor 8 GB Memory	0.96

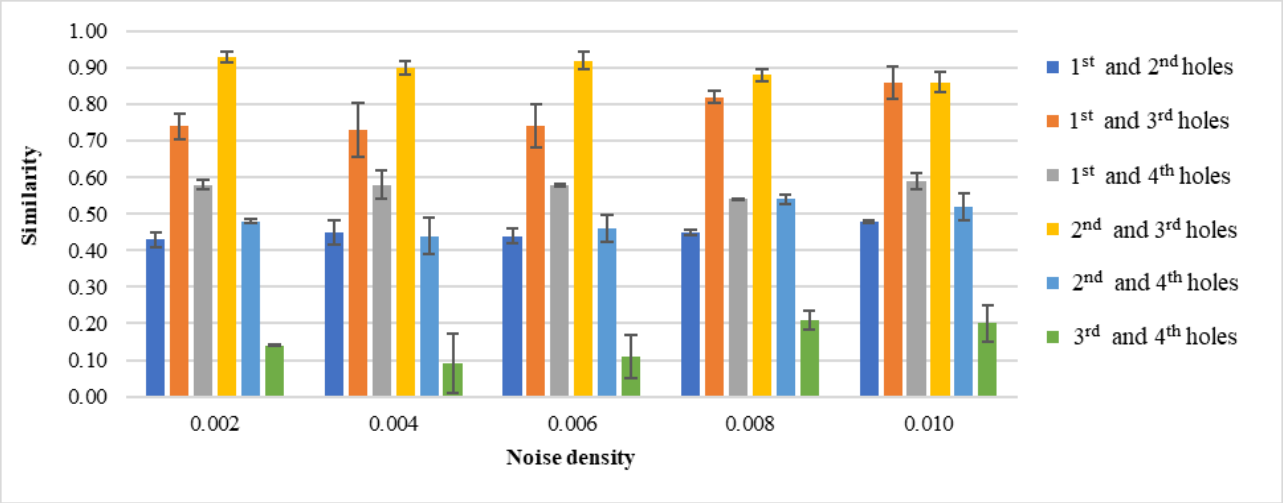

Figure 7. Experimental results based on images with noise densities of 0.002, 0.004, 0.006, 0.008 and 0.010.

VI. Conclusion

In this paper, a novel endoscopic navigation method is proposed based on 3D structure registration. In the method, endoscopic images are registered to CT data by the 3D structural characteristics-based similarity computation. The main innovations of the method are as follows: (1) An algorithm of reconstructing 3D porous structures from low resolution and weak texture endoscopic images is proposed. (2) A hole search strategy is proposed to extract the 3D structure information of holes in CT data. (3) A descriptor and a similarity measurement formula are defined for 3D structure registration. Finally, the performance of the method has been demonstrated based on the real data from the ureteroscopic holmium laser lithotripsy.

References

[1] T. Bergen and T. Wittenberg, "Stitching and surface reconstruction from endoscopic image sequences: a review of applications and methods," *IEEE Journal of Biomedical and Health Informatics,* vol. 20, no. 1, pp. 304-321, 2014.

[2] D. Makris, A. Scherpereel, S. Leroy, B. Bouchindhomme, J.-B. Faivre, J. Remy, P. Ramon, and C.-H. Marquette, "Electromagnetic navigation diagnostic bronchoscopy for small peripheral lung lesions," *European Respiratory Journal,* vol. 29, no. 6, pp. 1187-1192, 2007.

[3] D. Deguchi, M. Feuerstein, T. Kitasaka, Y. Suenaga, I. Ide, H. Murase, K. Imaizumi, Y. Hasegawa, and K. Mori, "Real-time marker-free patient registration for electromagnetic navigated bronchoscopy: a phantom study," *International Journal of Computer Assisted Radiology and Surgery,* vol. 7, no. 3, pp. 359-369, 2012.

[4] A. C. Mehta, K. L. Hood, Y. Schwarz, and S. B. Solomon, "The evolutional history of electromagnetic navigation bronchoscopy: state of the art," *Chest,* vol. 154, no. 4, pp. 935-947, 2018.

[5] K. Luo, Y. Lin, X. Lin, X. Yu, J. Wen, K. Xi, P. Lin, and L. Zhang, "Localization of peripheral pulmonary lesions to aid surgical resection: a novel approach for electromagnetic navigation bronchoscopic dye marking," *European Journal of Cardio-Thoracic Surgery,* vol. 52, no. 3, pp. 516-521, 2017.

[6] M. Turan, Y. Almalioglu, E. P. Ornek, H. Araujo, M. F. Yanik, and M. Sitti, "Magnetic-visual sensor fusion-based dense 3d reconstruction and localization for endoscopic capsule robots," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1283-1289: IEEE.

[7] Z. Baum, T. Ungi, A. Lasso, and G. Fichtinger, "Usability of a real-time tracked augmented reality display system in musculoskeletal injections," in *Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, 2017, vol. 10135, p. 101352T: International Society for Optics and Photonics.

[8] F. Sauer, A. Khamene, and S. Vogt, "An augmented reality navigation system with a single-camera tracker: System design and needle biopsy phantom trial," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2002, pp. 116-124: Springer.

[9] N. Koizumi, K. Sumiyama, N. Suzuki, A. Hattori, H. Tajiri, and A. Uchiyama, "Development of three-dimensional endoscopic ultrasound system with optical tracking," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2002, pp. 60-65: Springer.

[10] D. Deguchi, K. Mori, M. Feuerstein, T. Kitasaka, C. R. Maurer Jr, Y. Suenaga, H. Takabatake, M. Mori, and H. Natori, "Selective image similarity measure for bronchoscope tracking based on image registration," *Medical Image Analysis,* vol. 13, no. 4, pp. 621-633, 2009.

[11] X. Luó, M. Feuerstein, D. Deguchi, T. Kitasaka, H. Takabatake, and K. Mori, "Development and comparison of new hybrid motion tracking for bronchoscopic navigation," *Medical image analysis,* vol. 16, no. 3, pp. 577-596, 2012.

[12] M. Shen, S. Giannarou, and G.-Z. Yang, "Robust camera localisation with depth reconstruction for bronchoscopic navigation," *International Journal of Computer Assisted Radiology and Surgery,* vol. 10, no. 6, pp. 801-813, 2015.

[13] M. Shen, S. Giannarou, P. L. Shah, and G.-Z. Yang, "Branch: Bifurcation recognition for airway navigation based on structural characteristics," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 182-189: Springer.

[14] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Transactions on Image Processing,* vol. 24, no. 1, pp. 120-129, 2014.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.

[16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics,* vol. 9, no. 1, pp. 62-66, 1979.

[17] C. Lin, Y. Li, G. Xu, and Y. Cao, "Optimizing ZNCC calculation in binocular stereo matching," *Signal Processing: Image Communication,* vol. 52, pp. 64-73, 2017.

[18] Z. Cai, C. Hu, W. Lin, and W.-a. Yang, "A feature point extraction and matching algorithm for Wireless Capsule Endoscope image," in *2013 IEEE International Conference on Information and Automation (ICIA)*, 2013, pp. 608-613: IEEE.

[19] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM Siggraph Computer Graphics,* vol. 21, no. 4, pp. 163-169, 1987.