

An Accurate and Efficient 3D Reconstruction Method for Wireless Capsule Endoscopy Based on Harris Corner Detection and Fixed Template Matching

Tigang Cai

School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, China

Abstract: *To overcome the limitation of wireless capsule endoscopy (WCE) in providing only two-dimensional intestinal images—which constrains accurate diagnosis of intestinal lesions—this study proposes an accurate and efficient three-dimensional detection method. The approach applies 3D reconstruction to WCE images by computing corresponding points across successive frames and estimating the capsule's position and orientation. A critical step involves the selection of highly specific feature points and the determination of their correspondences. Accordingly, this paper combines a Harris corner-based feature extraction algorithm with a fixed matching template algorithm to achieve robust feature point matching in WCE intestinal images, substantially improving both matching efficiency and robustness. Using the obtained feature matching points, an intestinal model is reconstructed via 3D reconstruction. Experimental results demonstrate that the proposed method offers advantages in efficiency and reconstruction quality compared with conventional 3D reconstruction approaches, indicating its potential to enhance WCE-based lesion localization and diagnostic accuracy.*

Keywords: Wireless capsule endoscopy (WCE); 3D reconstruction; Harris corner detection; Feature point matching; Fixed template matching; Intestinal lesion diagnosis.

1. Introduction

The incidence of gastrointestinal diseases in the world is generally higher than 10% [1]. Colorectal cancer is the third most common cancer in the world and requires minimally invasive means for pathological diagnosis. Since the advent of capsule endoscopy, gastrointestinal Innovation has been carried out in the examination to eliminate the pain and safety hazards caused by traditional methods. By swallowing to follow the texture and peristalsis of the gastrointestinal tract in the human body [1], it also does active control work. A part of the digestive tract is photographed by the camera of the capsule endoscope, and the captured picture is wirelessly transmitted to the external receiving device. Doctors can observe and diagnose the obtained pictures, which greatly reduces the pain of patients during the examination process [1]-[2]. The shortcoming is that the lesion area can not be focused on, but the blind area at the bend cannot be observed, and the viewing angle is limited. A large number of pictures are extremely heavy to read manually, and the differences between different doctors may produce different results for the same picture. Therefore, the research of computer-aided diagnosis using machine learning and image processing technology reduces the work intensity of doctors. The three-dimensional reconstruction of capsule endoscope images is based on the image information of the endoscope, which can focus on lesions and specific positions for precise observation and precise positioning, shortening the inspection time [2]; compared with traditional endoscopes, the wireless capsule endoscope (WCE) has expanded The advantages of visual field, easy operation, safety and hygiene, and strong tolerance, although the research of wireless capsule endoscope has made great progress [2], but there are still some shortcomings. The current 2D endoscopic images lack three-dimensional information of anatomical structure.

At present, most endoscopes have only one lens, which is a monocular vision system, and cannot obtain information on three-dimensional scenes such as depth and size [3]. Surgeons can only By moving the endoscope to obtain continuous two-dimensional images, imagine the three-dimensional structure in the current field of view, and because of the huge amount of image data obtained, it is not practical to rely on the doctor to carefully judge the images one by one [3]. At the same time, because the observed dynamic environment image is a combination of intestinal peristalsis and capsule endoscopy, the observation of the doctor will interfere with the doctor's judgment [1-3], reducing the accuracy of diagnosis. Then, in order to realize the adjustable zoom in and zoom out observation of the lesion site [1], three-dimensional reconstruction of the endoscopic image is needed. When the image returned by WCE shows a focal area, it is usually only measured by measuring the size of the tissue in the two-dimensional image. This will cause a large error, so three-dimensional reconstruction of the intestine is required. In this paper, two sequence images returned by WCE are used for 3D reconstruction. The method adopted in this paper is to improve the feature matching algorithm to find better feature matching points, and then combine with the traditional 3D reconstruction algorithm to achieve the effect of reconstruction of the intestine.

2. Feature Matching Algorithm

Because the capsule endoscope has no power, it travels slowly in the intestine. The two sequence images taken by the travel will contain the same position information [2], then you can pass the sequenced two images returned from the intestine endoscope. For feature matching point detection, the existing algorithm can be used to calculate the motion parameters of the endoscope, and then the intestinal model is reconstructed using the 3D reconstruction algorithm. The first step in 3D

reconstruction is to perform image feature matching, and it is necessary to find feature points. Then for the intestinal image of WCE, there may be many smooth areas, and the location information generated by matching in these places usually produces large errors, so it is necessary to find feature points with high specificity. The feature point extraction in the intestine usually has Gradient method and Harris corner detection method [4], then according to the needs of this article, Harris corner detection [5] is used to extract these features.

2.1 Feature Point Extraction

Harris corner detection is to use a fixed matching window to pan on the picture at random, and then compare the change of pixel gray before and after translation. If the pixel gray gradient difference is large [5], it is determined that there is a corner in this matching window. For the gray change of the pixels in the matching window before and after translation, as shown in formula (1):

$$E(x, y) = \sum_{a,b} w(a, b) [I(a+u, b+v) - I(a, b)]^2 \quad (1)$$

$E(x, y)$ is the offset of the window, (a, b) is the pixel coordinate position of the picture, and the amount of its position is consistent with the size of the window, $w(a, b)$ is a window function. After Taylor's [6] expansion of the above formula, we get a matrix expression as shown in (2) (3), using the gradient of each pixel in the x and y directions for statistical analysis.

$$E(x, y) = [x, y] H \begin{bmatrix} x \\ y \end{bmatrix} \quad (2)$$

$$H = \left(\sum_{a,b} w(a, b) \begin{bmatrix} I_a^2 & I_a I_b \\ I_a I_b & I_b^2 \end{bmatrix} \right) \quad (3)$$

$$R = \det H - k(\text{trace} H)^2 = \lambda_1 \lambda_2 - K(\lambda_1 + \lambda_2)^2, K = 0.03 \quad (4)$$

Then the corner points are obtained according to the formula, where H is the covariance matrix, and the eigenvalues are obtained for λ_1 and λ_2 . Through experiments, we find that there are only corner points in the area when the eigenvalues are relatively large. By using the above corner detection [6], the following feature point maps of the WCE intestinal image were obtained, and roughly 144 coarse feature points with high specificity were initially obtained under the fixed threshold. Figure 3 is Feature point detection results.

2.2 Image Feature Matching

There are three main aspects of image matching, including data type matching, measurement function similarity, and search strategy [7]. These three aspects determine the real-time and accuracy of matching. In our research, we found that some traditional feature matching methods are not suitable for feature matching of WCE images (such as SIFT), because the features of WCE intestinal images are not very obvious, and the inner wall of intestinal images is very smooth [8-11]. As for its characteristic points, it is difficult to determine. Based on the corner point extraction of the highly specific feature points of the above WCE intestinal image, we use a fixed-size template matching algorithm to match the feature points of the

two intestinal images [12],[13]. During the matching process, the The template image N is translated on the intestinal image. By calculating the similarity function $B(a, b)$ of the template image of the sub-graph relative to its coordinate position, the matching image area can be found. Usually, the template matching method can determine the matching point through the correlation function. In the actual feature point matching application, one of the following two measurement formulas such as (5) and (6) can be used to measure the similarity between the template N (the size is M_x, M_y) and the matching image $S^{a,b}$; to achieve the purpose of obtaining matching points.

$$B(a, b) = \sum_{i=1}^{M_x} \sum_{j=1}^{M_y} [S^{a,b}(i, j) - N(i, j)]^2 \quad (5)$$

$$B(a, b) = \sum_{i=1}^{M_x} \sum_{j=1}^{M_y} [N(i, j)]^2 - 2 \sum_{i=1}^{M_x} \sum_{j=1}^{M_y} [S^{a,b}(i, j) \cdot N(i, j)] + \sum_{i=1}^{M_x} \sum_{j=1}^{M_y} [S^{a,b}(i, j)]^2 \quad (6)$$

Among them, the middle term on the right side of formula (6) represents the relationship between the template and the matching image. When N and $S^{a,b}$ matches, the term gets the maximum value. So we can transform the above formula to determine a related function and normalize it to formula (7). We can write formula (7) into a matrix expression format as shown in formula (8). By normalizing the formula, we can get out, only when the template N is similar to the matching image area $S^{a,b}$, the function $P(i, j)$ can obtain the maximum value, and finally achieve the feature matching point.

$$P(i, j) = \frac{\sum_{i=1}^{M_x} \sum_{j=1}^{M_y} [S^{a,b}(i, j) \cdot N(i, j)]}{\left(\sum_{i=1}^{M_x} \sum_{j=1}^{M_y} [N(i, j)]^2 \right)^{1/2} \left(\sum_{i=1}^{M_x} \sum_{j=1}^{M_y} [S^{a,b}(i, j)]^2 \right)^{1/2}} \quad (7)$$

$$P(i, j) = \frac{n^T s(a, b)}{(t^T)^{1/2} (s^T(a, b) s(a, b))^{1/2}} \quad (8)$$

We can transform formula (6) to determine a related function and normalize it to formula (7), and then transform formula (7) into a matrix expression format, such as formula (8).

The intestinal images returned by WCE are 2 to 3 images per second, and we use two consecutive images to match the features of the images [13]. Through experiments, we found that there is a difference in the accuracy of feature point matching under matching templates of different sizes. As shown in Figure 2, we can see that the matching template has the best matching rate at 26×26 , so the matching in this experiment the template selects a template of this size for feature matching. At the same time, we find that corner features do not exist everywhere in the intestinal image, so choosing different matching search paths will achieve different matching efficiency, as shown in Figure 1, this article uses the path with the highest match rate. In addition, we can see that the two WCE images in the sequence have some differences. By using the above feature matching algorithm, we can get the matching results. As shown in Figure 3, the experimental results show that there are about 100 coarse matching results.

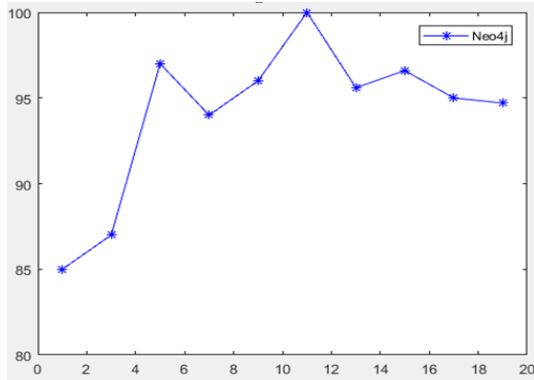


Figure 1: Matching rates for different search ranges

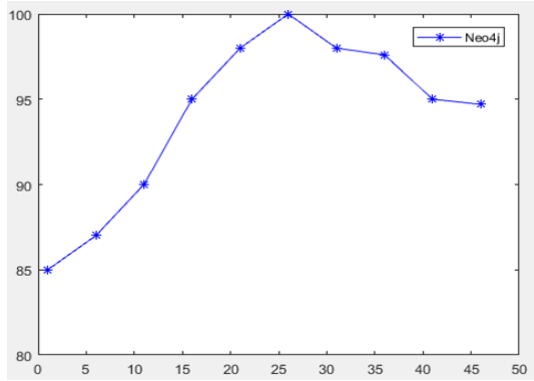


Figure 2: matching accuracy of different templates

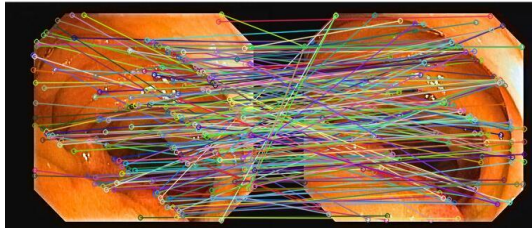


Figure 3: Matching results

Through the analysis of Figure 3, we find that the algorithm composed of corner feature point extraction and template feature point matching in this paper will have some mismatched feature matching points. In order to be more accurate in the later 3D reconstruction of the intestine, we adopt Random sampling consistency (RANSAC) filters out the mismatch points [12], and the results are shown in Figure 4. We found that 50 pairs of highly specific feature matching points were finally obtained, which will provide data support for the subsequent 3D reconstruction of the intestine.

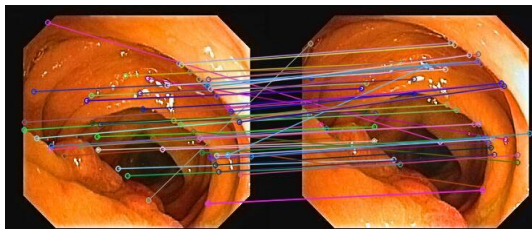


Figure 4: Optimized feature matching graph

3. Intestinal 3D Reconstruction Process

Use the detected multiple feature matching points to extract

the sparse point cloud, and then use the extracted feature matching points to calculate the basic matrix in stereo vision, calculate the camera rotation angle and direction of motion [10]-[13]; then use the visual mechanism Triangulation [13]

can calculate the relative three-dimensional coordinates of each feature point and the world three-dimensional coordinates to obtain a dense point cloud. Since it is impossible to obtain all the three-dimensional points of the image, it is necessary to reconstruct the 3D points and finally edit the texture to achieve the purpose of restoring the intestine. In this paper, based on the extracted high-specific feature matching points of the intestine, avoiding the traditional SIFT feature point matching method [14], the reconstruction is improved based on the 3D reconstruction method to achieve the reduction of the intestinal scene, summarized That is: image acquisition, feature matching, sparse point cloud, camera parameter estimation, dense point cloud, surface reconstruction, texture mapping [14], the specific process is shown in Figure 5. The input image maintains continuity, and any two adjacent images are deemed to require 3D reconstruction. In this paper, the EXIF information that comes with the image is used to determine the internal parameters of the camera. In the subsequent reconstruction, the camera will be Parameter optimization.

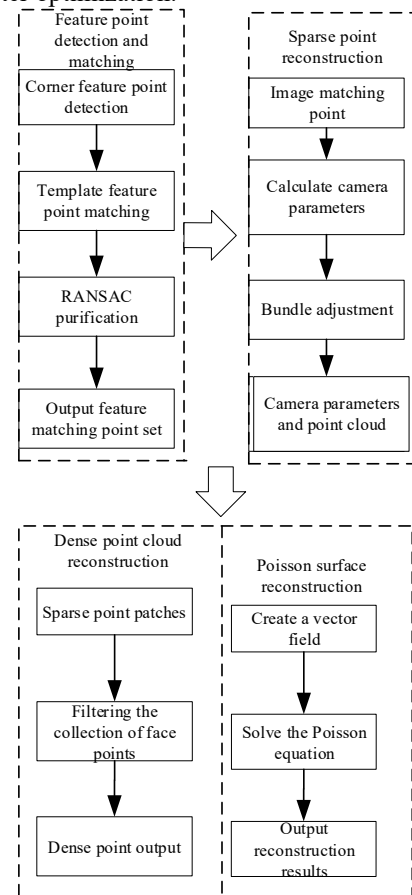


Figure 5: 3D reconstruction process

Under the image feature matching points obtained in this paper, using the camera parameters carried by the intestinal image itself, first run the triangulation method to reconstruct the sparse point cloud. According to the corresponding coordinates (x, y, l) and (x', y', l) in the two pictures and the camera matrices P and P' in the two pictures, the coordinate X of the space point is solved according to the relationship $x = PX$. As shown in formula (9), where p^{iT} is the P column of i .

$$x(p^{3T}X) - (p^{1T}X) = 0 \quad (9)$$

$$y(p^{3T}X) - (p^{2T}X) = 0 \quad (10)$$

There are two types above, two sets of data $x = PX$, $x' = P'X$ can be written as:

$$AX = 0 \quad (11)$$

$$A = \begin{bmatrix} xp^{3T} - p^{1T} \\ yp^{3T} - p^{2T} \\ x'p'^{3T} - p'^{1T} \\ y'p'^{3T} - p'^{2T} \end{bmatrix} \quad (12)$$

The coordinates of the three-dimensional point can be obtained from equation (12).

After obtaining the preliminary three-dimensional points, you need to perform global optimization operations on these points, remove the points with large back projection errors, repeat the optimization steps, and remove the points with high anomalies. After all the sequence images are reconstructed, you can get sparse point cloud. Finally, the obtained sparse point cloud is input into PMVS [14]-[16], through feature matching and filtering operations, a dense point cloud is obtained, and finally a 3D model of the intestine can be obtained by surface rendering and reconstruction.

4. Experiment Analysis

Before the Harris corner detection of two consecutive intestinal images [16], we can find that even though the two intestinal images of the left and right sequences look the same, there are still some differences, which can be obtained by observing the two histograms. There are differences between two adjacent intestinal images, as shown in the front and back sub-pictures (a) (b) of Figure 6. Compared with the traditional method, the number of feature points extracted from the smooth bowel by Harris corner detection is larger than that of the traditional method. Also. The 3D reconstruction of the highly specific intestinal feature matching point set obtained from the above is shown in the figure. This algorithm can obtain depressions with different shapes and sizes. The wrinkles are more obvious in the 3D picture, and the convex parts of the shape get a better 3D display. Compared with the traditional 3D reconstruction after the feature matching algorithm, this algorithm has obtained better results. The comparison and display are shown in Table 1, and the effect of intestinal 3D reconstruction performed by the feature matching algorithm in this paper is shown in Figure 7, Figure 7 sub-graph b is the result of reconstruction using traditional SIFT feature matching.

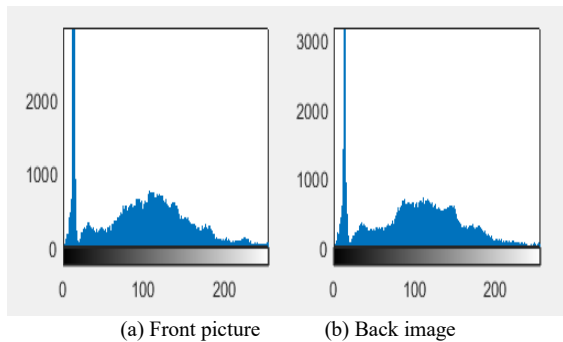


Figure 6: WCE adjacent picture difference

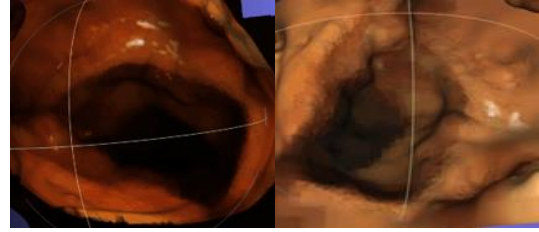


Figure 7: Two algorithm reconstruction results

The 3D reconstruction of the WCE intestine performed by the method used in this article has some obvious advantages over the traditional method, as shown in the comparison of the reconstruction effect in Table 3; at the same time, through the time comparison analysis of Table 1 and Table 2, we can see that for the WCE intestine The 3D reconstruction of the road image is faster than the traditional methods in the feature point matching, sparse point cloud acquisition and other reconstruction processes, especially in the feature point extraction and matching; the intestinal feature matching algorithm used in this paper is highly specific The number of feature matching points is better than the traditional algorithm.

Table 1: Feature point matching time comparison

Algorithm	Total elapsed time /(m)	Feature point matching /(m)	Sparse point cloud /(m)
The algorithm of this paper	4.5	2.4	2.1
Traditional algorithm	9.7	5.2	4.5

Table 2: Comparative analysis of reconstruction process time

Experiment time	Dense match /(m)	Texture editing time /(m)	Number of matching points /(Couple)
The algorithm of this paper	0.91	0.86	55
Traditional algorithm	1.89	1.36	2

Table 3: Reconstruction effect evaluation form

Evaluation index	Edge profile	Cavity profile	Inner wall surface	Texture color
Algorithm	Relatively clear and smooth	Clear	Relatively smooth	Clearer color
Traditional algorithm	Glitch is obvious	Not clear	Not smooth	distortion

5. Conclusions

Based on the corner detection algorithm, this paper applies the feature point extraction of WCE intestinal images. Compared with the traditional feature point extraction (such as SIFT), corner detection is used to extract highly specific feature points from the smoother intestine. It looks more robust. After

the feature points are extracted, the template matching method is used to perform feature matching and matching point optimization on the two intestinal images in the sequence, and finally a better feature matching point set is obtained. Finally, the extracted high-specificity feature matching points are used for 3D reconstruction; for comparison, this article lists the 3D reconstruction results of the feature matching points obtained by the traditional feature matching algorithm, and the experiment draws the conclusion: The 3D reconstruction of the tract is more robust than the traditional intestinal 3D reconstruction in terms of edge contour, inner wall surface and other aspects of reconstruction. It also determines the practicability of the proposed method for WCE intestinal 3D reconstruction.

References

- [1] Hu Chao, Cai Zhenyu, You Xiaohe, et al., 3D Recovery of Gastrointestinal Tract Based on Capsule Endoscopic Images [J], Journal of Transduction Technology, 2017, 30(5): 708-714.
- [2] Cao Yanling, Three-dimensional reconstruction of capsule endoscopic images based on spline interpolation [D]. Harbin Institute of Technology, 2014.
- [3] Shi Zhan, Liao Hong'en, Colon 3D reconstruction from endoscope images based on colon fold Bezier curve interpolation [J], Beijing Biomedical Engineering, 2016, 35(6): 576-581
- [4] Ren Wei, Research on multi-view stereo real-time imaging and imaging key technology in ultra-high definition mode [D], Hefei University of Technology, 2017.
- [5] Yang Wenbo, Dense scene reconstruction of 3D spatial point selection under double constraints [D], Harbin University of Science and Technology, 2018.
- [6] Ujwala Ravale, Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function [J], Procedia Computer Science, 2015, 45 (39):428-435.
- [7] Yang Zhe, Real-time 3D reconstruction based on monocular camera [D], 2017.
- [8] Shi Ying, Wang Wenjian and Bai Xuefei, 3D Dense Reconstruction Method Based on Multiple Features [J], Journal of Frontier of Computer Science and Technology, 2015, 9(5): 594-603.
- [9] Long M, Lan Z, Xie X, et al., Image Enhancement Method Based on Adaptive Fraction Gamma Transformation and Color Restoration for Wireless Capsule Endoscopy[C]//2018 IEEE Biomedical Circuits and Systems Conference (BioCAS). IEEE, 2018: 1-4.
- [10] Prasath V B S and Kawanaka H, Near-light perspective shape from shading for 3D visualizations in endoscopy systems[C]//2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2017: 2293-2295.
- [11] Long M, Lan Z, Xie X, et al., Image Enhancement Method Based on Adaptive Fraction Gamma Transformation and Color Restoration for Wireless Capsule Endoscopy[C]//2018 IEEE Biomedical Circuits and Systems Conference (BioCAS). IEEE, 2018: 1-4.
- [12] Rattanalappaiboon S, Bhongmakapat T and Ritthipravat P, Fuzzy zoning for feature matching technique in 3D reconstruction of nasal endoscopic images [M], Pergamon Press, Inc. 2015.
- [13] Zheng E, Ji D, Dunn E, et al., Sparse Dynamic 3D Reconstruction from Unsynchronized Videos[C]// IEEE International Conference on Computer Vision, 2016.
- [14] Ma Wenchao, Study on 3D reconstruction technology based on sequence image [D], Harbin Engineering University, 2018.
- [15] Fu Wancheng, Research and application of multi-scale image matching algorithm based on feature points [D], Southeast University, 2017.
- [16] Ren Wei, Research on multi-view stereo real-time imaging and imaging key technology in ultra-high definition mode [D].