Robotica: (2021) volume 39, pp. 334–345. © The Author(s), 2020. Published by Cambridge University Press.
doi:10.1017/S0263574720000399

Depth Estimation for Local Colon Structure in Monocular Capsule Endoscopy Based on Brightness and Camera Motion

Lei Xu†, Jing Li†*, Yang Hao‡, Peisen Zhang‡, Gastone Ciuti†¶, Paolo Dario†¶ and Qiang Huang†‡

†Advanced Innovation Center for Intelligent Robots and Systems, Beijing Institute of Technology, Beijing, China. E-mail: 7420180002@bit.edu.cn
‡School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China. E-mails: 3120150088@bit.edu.cn, zpsbit@foxmail.com, qhuang@bit.edu.cn
¶The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy. E-mails: gastone.ciuti@santannapisa.it, paolo.dario@santannapisa.it

(Accepted May 3, 2020. First published online: May 27, 2020)

SUMMARY

We present a 3D reconstruction method using brightness and camera motion estimation for registering local colon structure in colonoscopy. The proposed method is based on reverse projection from 2D fold contours to 3D space, motion estimation from 3D reconstructed points between neighboring frames, and model registration to reconstruct the fold structure. On the synthetic colon, the average percentages of the reconstructed depth error and circumference error are about 14.2% and 15.2%, respectively. The accuracy is enough for the navigation and control in capsule robot. This work demonstrates that the proposed method is superior to the methods using single-frame-based brightness intensity.

KEYWORDS: 3D reconstruction; Motion estimation; Colon fold contours; Model registration.

1. Introduction

In the past 10 years, colorectal cancer is one of the most common cancers in the world. The death rate of colorectal cancer accounts for 15% of cancers, which is only lower than that of lung cancers.[1] It has been recognized that adenoma-type polyps may become cancerous. If the polyps that cause colorectal cancer can be detected and removed, the occurrence of colorectal cancer can be effectively prevented. At the same time, if colorectal cancer can be detected at an early stage, the cure rate will be more than 90%. So medical experts recommend that early screening is essential to the people over 40 years old, for whom with family history of cancer the screening should be done 10 years earlier.

Optical colonoscopy is the main method to examine and detect intestinal lesions nowadays. The method consists of introducing an instrument called endoscope which has a light source and a camera mounted on it to observe the internal mucosa of the colon. During traditional flexible colonoscopy, images generated by the camera on the head of the endoscope are displayed on a monitor for real-time analysis by the endoscopist. During the insertion phase of colonoscopy, the endoscopist adjusts the angle by controlling the operation handle at the end of the endoscope to align the head of the endoscope with the intestinal tract, so that the endoscope can be inserted into the intestine smoothly.[2]

* Corresponding author. E-mail: 10902016@bit.edu.cn

Because the end of the flexible endoscope is rigid and large size, it may cause pain to patients, or even intestinal perforation.

There is a recent trend on developing active systems for endoscopy which aim at providing more flexible control to the colonoscopy by external driving force. One kind of the active endoscopes is called capsule endoscope, which has a shape and size similar to capsule pill. Because of its smaller size, a smooth operation can be performed by the endoscopist during the procedure of colonoscopy using the capsule endoscope with function of navigation.[3] Although the endoscopist has a general concept about the possible shape of the colon, the certain shape and structure of a specific patient's colon can vary widely from this general concept. Therefore, the endoscopist does not have prior knowledge of a specific patient's colon structure before colonoscopy. Capsule endoscope only has a monocular camera which results in missing depth information of colon structure during colonoscopy, so it is necessary to reconstruct 3D structure of colon based on images for navigation of colonoscopy.

Existing 3D reconstruction methods for general objects such as multi-view stereo,[4,5] structure from motion (SfM)[6–8] and shape from shading (SfS)[9–12] determine the 3D structure of objects from 2D images of surface view of objects. Multi-view stereo is to generate 3D objects using 2D images from multiple cameras. Because capsule endoscope has only one fish-eye camera, reconstruction from stereo views is not feasible for our situation. SfM is to find correspondence of the feature points between images to estimate the relative motion between the object and the camera, then the 3D shape of the object is calculated. There is a limitation of the method that the target object must be rigid, which makes it difficult for SfM method to perform well in colonoscopy. SfS is to estimate the surface normals of the target object by observing it under different lighting conditions, it is based on the fact that the amount of light reflected by a surface is dependent on the orientation of the surface in relation to the light source and the observer to compute the shape of the object surface. But these existing work focus on colon surface reconstruction using SfS,[13] this does not meet our target to reconstruct the real-time local 3D colon structure for the endoscopist indicating the navigation path. Recently, 3D colon reconstruction techniques using Convolutional neural network and deep learning have have been proposed.[14–16] They train a model by a large number of tagged images that observes sequences of images and aims to explain its observations by predicting camera motion and colon structure, but the demanding samples are not easy to obtain in our application. So 3D shape reconstruction of local colon structure from images for navigation and real-time control is the problem that we need to solve.

This paper addresses a reconstruction method to estimate colon folds depth based on brightness intensity and optimize the results based on camera motion estimation. The accuracy of depth estimation based on our method is higher than that based on brightness intensity only,[17] which is a good exploration in the similar applications. First, we present a lumen detection method that can be used for a wide range of endoscopic images. Then, we can obtain colon fold segments in the endoscopic images by edge detection method[18] and fit these segments to closed contours. Reverse projection model based on camera intrinsic parameters can be established to project 2D coordinate of points on the fold contour to 3D space, as been done in the work.[17] Camera motion matrix is calculated by two consecutive frames of images with common fold contours, which are tracked by optical flow algorithm. Finally, local 3D reconstruction structure can be transformed to global coordinate system by using optimal camera motion matrix to reduce mean square error. Figure 1 shows the flow chart of the reconstruction method. Compared with existing SfM and SfS methods, the proposed method can reconstruct 3D shape of colon structure, which is more suitable for navigation. Meanwhile, compared with reconstruction method based on single frame, the accuracy of reconstruction result could be improved effectively.

2. Depth Estimation Based on Brightness

2.1. Lumen detection

In order to build imaging model of appearance of the lumen, we draw a graphical scheme of how colonoscopy image is generated. As illustrated in Fig. 2, the amount of light falling on the colon surface decreases approximately by the square of the distance between the light source and the point of surface. The farthest parts from the light source such as lumen are poorly lighted, which characterizes the lumen center as the dark region of the image. The image is segmented by rectangles of size $n \times n$ based on the idea that the region of lumen is connected and closed. In order to decide whether two adjacent rectangles belong to the same region only the gray value of their center points is used. If

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press

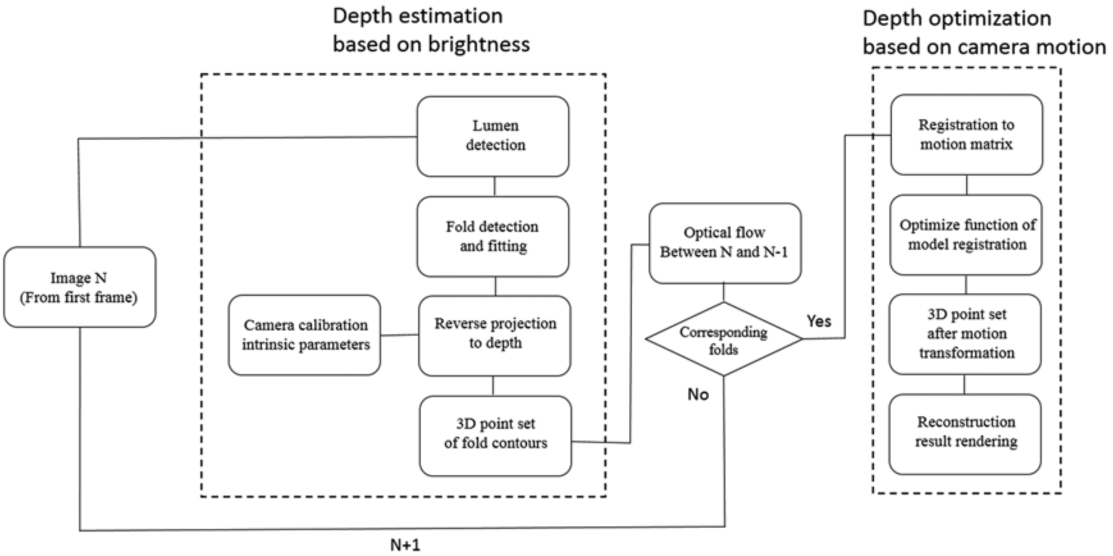
Depth estimation based on brightness
Depth optimization based on camera motion
Lumen detection
Fold detection and fitting
Image N (From first frame)
Camera calibration intrinsic parameters
Reverse projection to depth
3D point set of fold contours
Optical flow Between N and N-1
Corresponding folds
Yes
No
Registration to motion matrix
Optimize function of model registration
3D point set after motion transformation
Reconstruction result rendering
N+1

Fig. 1. Flow chart of the reconstruction method based on brightness and camera motion.

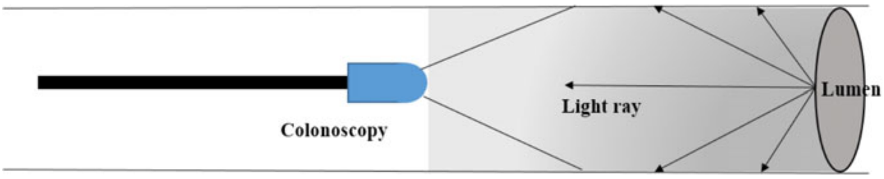

Fig. 2. Imaging model of appearance of the lumen based on illumination.

the gray value difference is less than or equal to the tolerance ϵ which can be adjusted, the rectangles are merged into one region. If g_1 and g_2 are two gray values to be examined, they are merged into the same region. Hence, we have Eq. (1).

$$|g_1 - g_2| \leq \epsilon \tag{1}$$

In the candidate regions, those regions that do not conform to the characteristics of lumen such as those with regions of no lumen shape feature and no holes are rejected. We define that the shape feature is named as structure factor F, it can be obtained according to (2).

$$F = \frac{1}{A}\sum\left(\|p - p_i\| - \frac{1}{A}\sum\|p - p_i\|\right)^2 \tag{2}$$

In Eq. (2), A is the area of the region, p is the center coordinate of the region, and p_i is the point in the contour of region. F means deviation distance between the contour of the region and center of the region. Because of lumen shape similar to circular, F is not greater than the setting threshold f_{Th}. We determine the lumen region that has minimal mean gray value of region from the rest in the smoothed image (Fig. 3(c)) using a gaussian kernel to filter noise. Figure 3(b) shows the result of merged regions, the selected result from merged regions is the lumen region marked as the red contour in Fig. 3(d).

2.2. *Fold contour extraction and fitting*

Because of the problems of occlusions among folds and specular reflection of light source, contours obtained by edge detection algorithm are discontinuous and incomplete. Meanwhile, the noise generated on the imaging causes false edges. The remaining edges are grouped into a set after filtered to remove the tiny ones. To connect all the contours belonging to the same fold in the set, we design an identification method as shown in Fig. 4(b): The central point o of lumen can be known by calculating the center of lumen region obtained in the previous step. The contour l_1 will be selected from the set if the distance between the central point and it is shortest. The other contour l_2 will be selected

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press

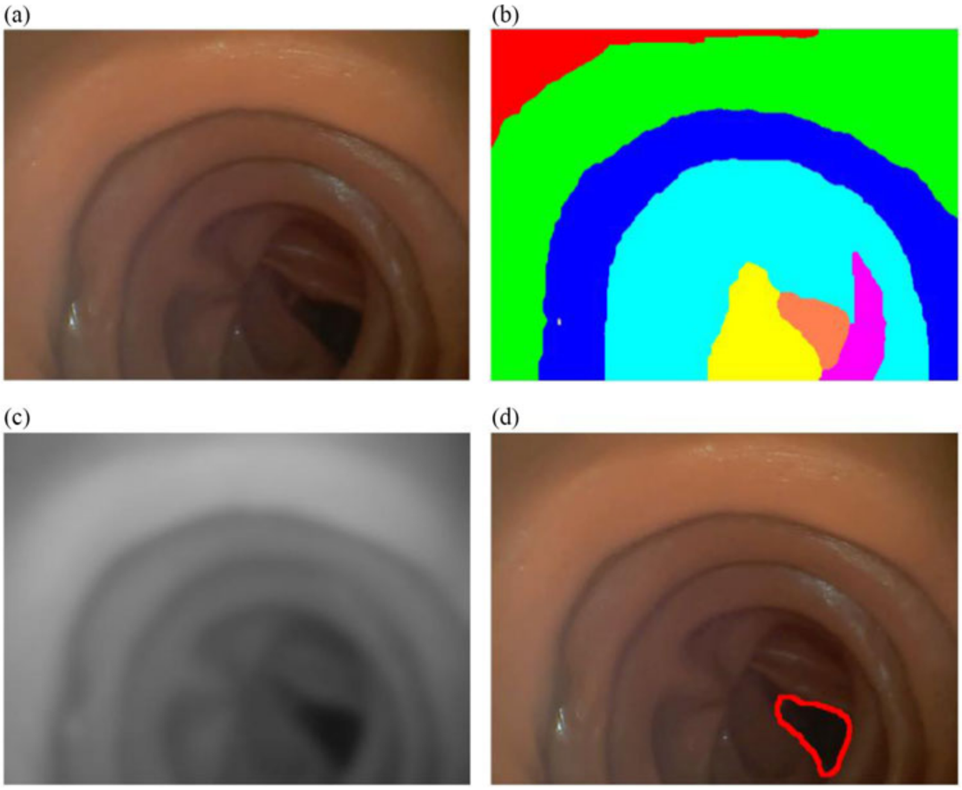

Fig. 3. (a) Colonoscopy image. (b) Merged regions. (c) Smoothed gray image (a). (d) Lumen region.

with the shortest distance from endpoints of l_1 to its endpoints in the set. Then, the two endpoints from l_1 and l_2, respectively, are marked p_1 and p_2. If the angle θ_0 formed by the connection line of p_1o and the connection line of p_2o, the angle θ_1 formed by the connection line of p_1o and the tangent line of the point of p_1, the angle θ_2 formed by the connection line of p_2o and the tangent line of the point of p_2 are all less than θ_m, and the distance $d_{p_1p_2}$ between p_1 and p_2 is less than λ, where θ_m and λ are parameters that can be set in specific applications, it is considered that the two contours l_1 and l_2 belong to the same fold. That means the two contours are merged if they fulfill the following condition in Eq. (3), $0 \leq p \leq 1$ is the weighting factor.

$$\frac{\mathrm{d}_{p_1p_2}}{\lambda} * p + \frac{\max\{\theta_0, \theta_1, \theta_2\}}{\theta_m} * (1 - p) < 1 \tag{3}$$

Then two endpoints p_1 and p_2 are smoothly connected to shape one contour. If the synthesized contour is not closed, it could be treated as l_1 to continue the step above until the set is empty. Figure 4 shows the results of each step of the extraction and fitting process.

2.3. 3D reconstruction of fold contours

The imaging model of the camera can be established by obtaining the matrix of intrinsic parameter from camera calibration.[19] With the given imaging model, the pixels on the fold contours in the image are projected to 3D space along the light ray passing through optical center. As illustrated in Fig. 5, the pixel of point u can be projected into 3D space vector $\boldsymbol{v_u}$ by reverse imaging model. As long as the depth information d_u of the pixel of point u is calculated, the 3D coordinates $(x(v_u), y(v_u), z(v_u))$ of it can be obtained according to (4).

$$\begin{cases} x\,(v_u) = d_u * x\,(u)\,\sin\delta_u / r_u \\ y\,(v_u) = d_u * y\,(u)\,\sin\delta_u / r_u \\ z\,(v_u) = d_u * \cos\delta_u \end{cases} \tag{4}$$

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press

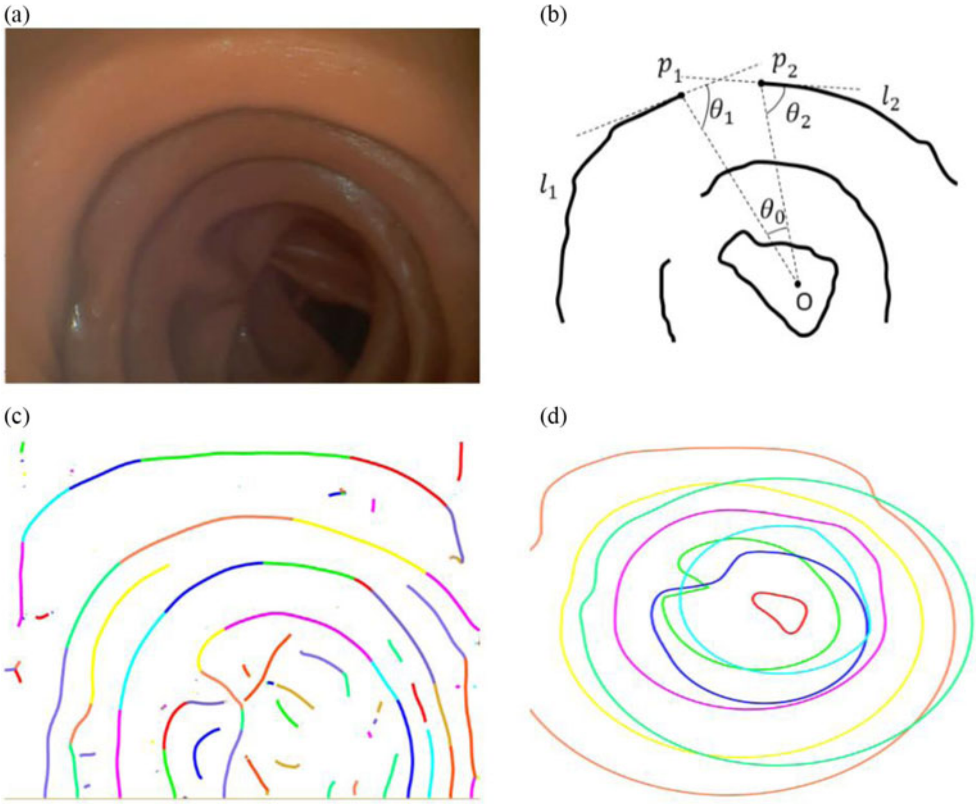

Fig. 4. (a) Colonoscopy image. (b) Sketch of two fold contour segments. (c) Result of edge detection. (d) Completed fold contours.

$(x(u), y(u))$ is 2D coordinates of the pixel u in the image; δ_u is the angle between the reverse projection vector and the axis of Z; and r_u is the Euclidean distance from the pixel u to the image center. To compute the value of δ_u, we assume that the camera follows the equidistance projection. Then we can obtain Eq. (4) as follows:[20]

$$\delta_u = r_u/f \tag{5}$$

where f is focal length of the lens, let R_0 be the image diameter, which is the maximum value of r_u. According to Fig. 5, when r_u is at the maximum, δ_u is also at the maximum, the value is half of the ψ, and ψ is field of view angle of the camera. Thus, $f = 2R_0/\psi$, we can obtain Eq. (6).

$$\delta_u = r_u * \psi/2R_0 \tag{6}$$

For estimation of depth from brightness intensity, we assumed that the colon surface is Lambertian except at specular spots. According to Lambertian cosine law,[21,22] at the same slant angle, the surface further away from the camera is darker than the one closer to the camera. At the same distance from the camera, the surface with more slant angle is darker than the one with less slant angle. For calculating the depth information d_u of the pixel u in 3D space, Eq. (7) is as follows:

$$d_u = \frac{C\left(1 - w_u + w_u \cos\left(\arcsin\left(\delta_u\right)\left(1 - h_u\right)\right)\right)}{i_u} \tag{7}$$

In Eq. (7), C is a constant, w_u is the normalized distance between the pixel u and the brightest nonspecular reflection spots in the fixed region along the normal direction of the pixel, and $0 \leq w_u \leq 1$; h_u is normalized height of the fold contour where the pixel u is located. Along the perpendicular direction of the tangent line of the fold contour on the pixel u, the normalized value of the difference between the average gray value of the 10 neighboring pixels of u inside the contour and the average

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press

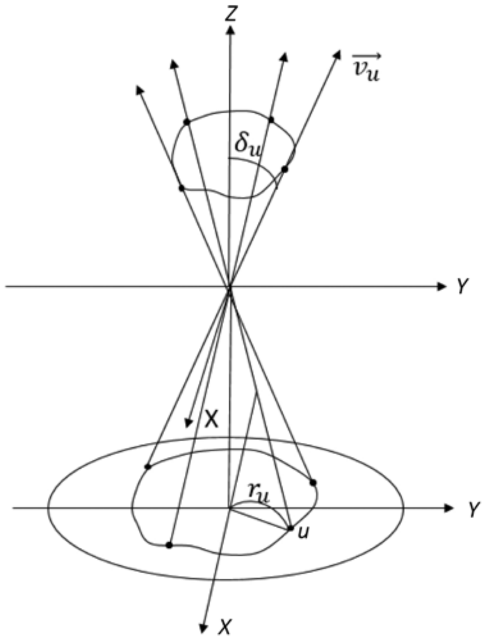

Fig. 5. Imaging model of colonoscopy.

gray value of the 10 neighboring pixels of u outside the contour is calculated, and the result is assigned to h_u; i_u is the gray value of the pixel u. Finally, we can calculate the 3D coordinates v_u from vertex $u(x(u), y(u))$ on a fold contour of an input image. This process will be applied to all the points on each fold contour.

3. Depth Optimization Based on Camera Motion

Because the real colonoscopy does not fully obey Lambertian cosine law, the estimation error of the depth information is obvious. In order to improve the accuracy of estimation of depth information, we present a method to achieve the optimal estimation by registering 3D points set of the corresponding fold contours among multiple frames to obtain the corresponding transformation matrix, which is used for minimizing the function of mean square error. The first image is used as a reference, and the target image is selected from the second image in turn, and the fold contours are extracted from the reference and the target. Optical flow is calculated to obtain movement vector between the two images.[23–25] We calculate the optical flow as the minimizer of a suitable energy functional. In general, the energy functional has the following form:

$$\{E(\boldsymbol{w}) = E_D(\boldsymbol{w}) + \alpha E_S(\boldsymbol{w}) \tag{8}$$

where $\boldsymbol{w} = (u, v, 1)$ is the optical flow vector field to be determined, $E_D(\boldsymbol{w})$ denotes the data term, $E_S(\boldsymbol{w})$ denotes the smoothness term, and α is a regularization parameter that determines the smoothness of the solution. Taking into account the below assumptions: constancy of the gray values and the spatial gray value derivatives between corresponding pixels in consecutive images and preservation of discontinuities in the flow field. Equation (8) can be written as

$$\begin{aligned} E(w) = \int \varphi_s \left(|f(x+\boldsymbol{w}) - f(x)|^2 + \gamma |\nabla_2 f(x+\boldsymbol{w}) - \nabla_2 f(x)|^2 \right) drdc \\ + \alpha \int \varphi_s \left(|\nabla_2 u(x)|^2 + |\nabla_2 v(x)|^2 \right) drdc \end{aligned} \tag{9}$$

here, $\varphi_s\left(s^2\right) = \sqrt{s^2 + \varepsilon^2}$ is a linear penalty function with $\varepsilon = 0.001$, which provides the desired preservation of edges in the movement in the flow field to be determined, α is the regularization parameter, and γ is the gradient constancy weight. Using movement vector w, the corresponding fold contours of the two images can be determined. In the reference image, the extracted contours are labeled in order. If there is a common fold contour between the reference and the target, it will be labeled as 1 in the reference and the target, then considering the third image as the target frame, we repeat the previous step to estimate movement vector. If there is the fold contour in the target

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press

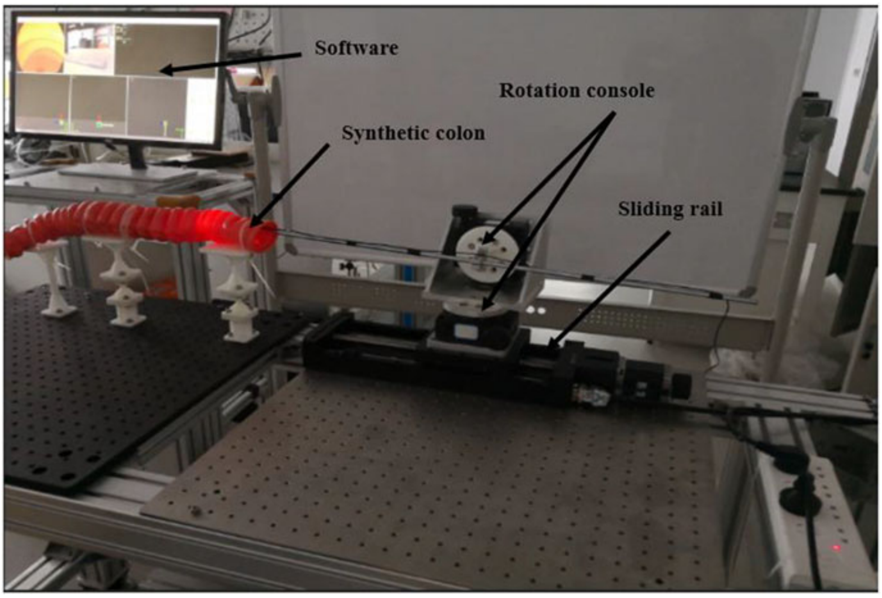

Fig. 6. Experimental platform for simulating camera motion.

which belongs to the same contour labeled as 1# in the reference, the corresponding contour in the third image is also labeled as 1#, and so on. The contours with the same label are selected from the reference and the target to obtain the corresponding 3D points set of reconstructed fold contours. The corresponding 3D points set with the same contours in the reference and the target are put into two point set (Q$_x$, Q_y, Q_z) and (P$_x$, P_y, P_z), respectively, then movement matrix $\mathbf{\Omega}$ is obtained by solving the Eq. (10).

$$\mathrm{MIN} = \sum_i \left\| \begin{pmatrix} Q_x \\ Q_y \\ Q_z \\ 1 \end{pmatrix} - \mathbf{\Omega} * \begin{pmatrix} P_x \\ P_y \\ P_z \\ 1 \end{pmatrix} \right\|^2 \tag{10}$$

where $\mathbf{\Omega}$ is 4×4 matrix, its expression can be written as

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0_{1\times3} & 1 \end{bmatrix} \tag{11}$$

In Eq. (11), $\mathbf{R}$ is 3×3 rotation matrix, and $\mathbf{t}$ is 3×1 translation matrix. Finally, we apply the above method of 3D model registration to multi-frame images, that is, from the reference and the second image as the target to calculate movement matrix $\mathbf{\Omega}_0$, from the reference and the image n as the target to calculate movement matrix $\mathbf{\Omega}_{n-2}$. In the set of movement matrices $\{\mathbf{\Omega}_0, \ldots, \mathbf{\Omega}_{n-2}\}$ obtained by model registration, the optimal movement matrix is obtained by the Eq. (12).

$$\mathbf{\Omega}_G = \min \left\| \begin{pmatrix} Q_x \\ Q_y \\ Q_z \\ 1 \end{pmatrix} - \mathbf{\Omega}_i * \begin{pmatrix} P_x \\ P_y \\ P_z \\ 1 \end{pmatrix} \right\|^2 \tag{12}$$

The reconstructed point set P can be obtained from the following Eq. (13):

$$P = \mathbf{\Omega}_G * P_R \tag{13}$$

P_R is the reconstructed point set for reference frame.

4. Experimental Results

In order to be useful for validating our algorithm of 3D reconstruction, the experimental platform with a synthetic colon model is built as illustrated in Fig. 6. There is a 3-DoF motor-controlled

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press

Table I. Depth and circumference of reconstructed results based on single image.

	Ground truth (mm)		Reconstructed result (mm)		Absolute difference (mm)	
	Depth	Circf	Depth	Circf	Depth	Circf
1#	53	301	45	271	8	30
2#	69	270	57	226	12	44
3#	90	284	73	225	17	59
			MAD		12.3	44.3

Table II. Camera motion estimation results based on image sequence.

	Ground truth	Reconstructed result	Absolute difference
	Trans (mm) Rotat (deg)	Trans (mm) Rotat (deg)	Trans Rotat
1#	(0,0,5)	(0,0,4.2)	16%
Video	(0,2.0,0)	(0,1.74,0)	13%
2#	(0,0,10)	(0,0,8.1)	19%
Video	(0,0,2.0)	(0,0,1.69)	15.5%
3#	(0,0,5)	(0,0,4.1)	15.5%
Video	(0,2.0,1.0)	(0,1.62,0.83)	18%
4#	(0,0,10)	(0,0,7.9)	21%
Video	(0,1.0,2.0)	(0,0.81,1.64)	18.5%

sliding rail on the platform, which can precisely control the pitch angle, the yaw angle, and the translation distance of the camera. The resolution of the image that camera acquires is 960×720 pixels. The software is developed in C++ with OpenCV3.0 and rendered in OpenGL4.0. We evaluated the effectiveness of our method on estimating the reconstructed colon structures based on brightness intensity and the optimal reconstruction result based on camera motion (*BAMforshort*).

Because there are no real 3D data of the synthetic colon during the experiment, we manually measured the depth of folds and the fold circumference ground truth from the outside of the synthetic colon. Although the measurements taken from the outside of the synthetic colon could be different from those taken from the inside of the synthetic colon, these measurements are attainable. We selected a straight section of the synthetic colon that has four obvious colon folds to give good results for evaluation. We use the mean absolute difference of measurements (MAD for short) between the reconstructed result and the ground truth for evaluation. The 0# fold is selected to confirm the physical scaling factor that gave the least MAD against the ground truth and scaled the three folds in the reconstructed model to the ground truth using the same scaling factor, which is to scale the reconstructed model in relative units to in millimeters. Table I shows the evaluation results for reconstruction based on brightness intensity from a single image in the selected straight section. The average depth MAD and circumference MAD on the synthetic colon are 12.3 and 44.3 mm, respectively.

For the evaluation of the influence of camera motion transformation to reconstructed result, we recorded four pieces of video with 45, 51, 49, and 55 frames, respectively. In these videos, the camera takes the 3-DoF movements, which are accomplished by controlling the sliding rail. Table II shows that the average percentage of translation error and rotation error on the synthetic colon are 18.5% and 16.3%, respectively.

Table III shows the reconstruction result of the fold contours in Table I by using estimated camera motions matrix to register local models. We observed a better accuracy of reconstructed fold contours in the synthetic colon than the result that was reconstructed from a single image. Furthermore, the reconstruction result of the fold contours with higher accuracy of motion estimation is better. The average depth MAD and circumference MAD to the same fold contours in Table I on the synthetic colon that reconstructed from 1# video are 9 and 42 mm, respectively. The same parameters from 2# video and 3# video are 10.3 and 43.6 mm, 11 and 44.3 mm. Only the result from 4# video is worse than the reconstruction result based on s single image. The average percentage of the depth

Table III. Depth and circumference of reconstructed results based on multi-frame motion estimation from 1# to 4# videos.

	GT. (mm)		1# video (mm)		2# video (mm)		3# video (mm)		4# video (mm)	
	Dep	Cir	Dep	Cir	Dep	Cir	Dep	Cir	Dep	Cir
1#	53	301	48	271	48	270	45	269	46	265
2#	69	270	61	230	59	228	60	227	56	223
3#	90	284	76	228	74	226	74	226	70	229
	MAD		9	42	10.3	43.6	11	44.3	13.3	46

GT. denotes ground truth.
Dep denotes depth, Cir denotes circumference.

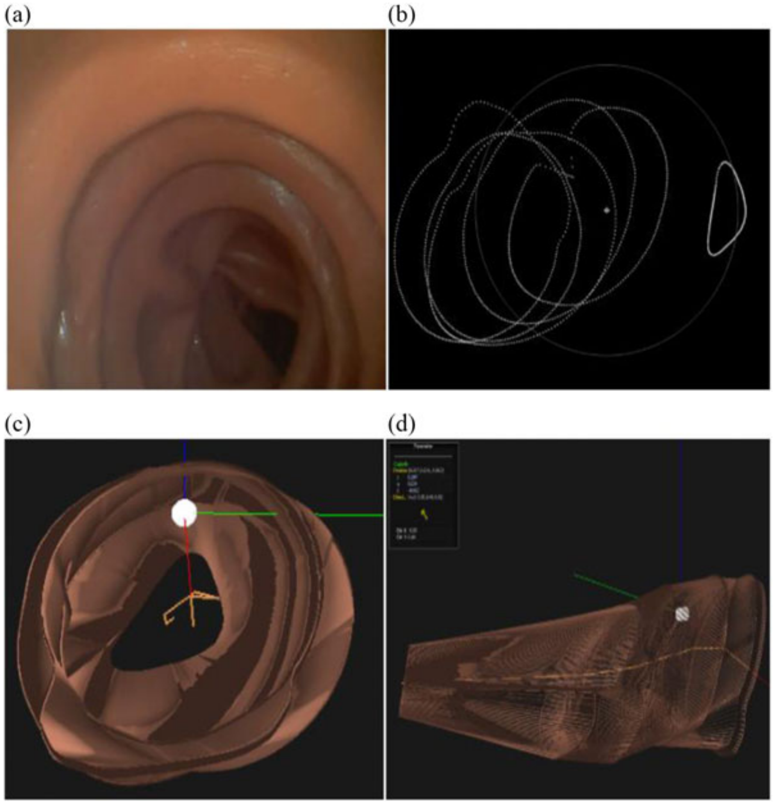

Fig. 7. (a) One of input image; (b) 3D reconstructed fold contours; (c) front view of rendered model; (d) side view of rendered model.

error is about 17.4% in the reconstructed result based on single image, and the error is about 14.2% in the reconstructed result based on multi-frame motion estimation. And the average percentage of the circumference error is about 15.5% in the reconstructed result based on single image, the error is about 15.2% in the reconstructed result based on multi-frame motion estimation. Figure 7 shows an example of reconstruction on a sequence of images from a synthetic colon where some colon folds have cylindrical shape.

SfM is also a motion-based technique that can be potentially explored to reconstruct the colon surface.

Table IV shows comparison to reconstruction results of depth between SfM and BAM. Because of more optimization iterations, the average accuracy of results based on SfM is improved by about 1.5% from 1# video and 2# video compared with BAM. However, the reconstruction time of BAM is about 0.3 s, and SfM takes more than five times of the time. In 3# video and 4# video, fold occlusion and specular reflection make feature points for 3D reconstruction based on SfM insufficient, which result in failures. However, BAM can still obtain satisfying reconstruction in these cases. Figure 8

Table IV. Reconstructed results of depth based on SfM and BAM from 1# to 4# videos.

	GT.(mm)	1# video		2# video		3# video		4# video	
	Dep	SfM	BAM	SfM	BAM	SfM	BAM	SfM	BAM
1#	53	49	48	48	48	–	45	–	46
2#	69	62	61	61	59	–	60	–	56
3#	90	76	76	77	74	–	74	–	70
MAD		8.3	9	8.6	10.3	–	11	–	13.3
Times(s)		1.53	0.28	1.61	0.31	–	0.33	–	0.31

GT. denotes ground truth.
Dep denotes depth. – denotes no results.

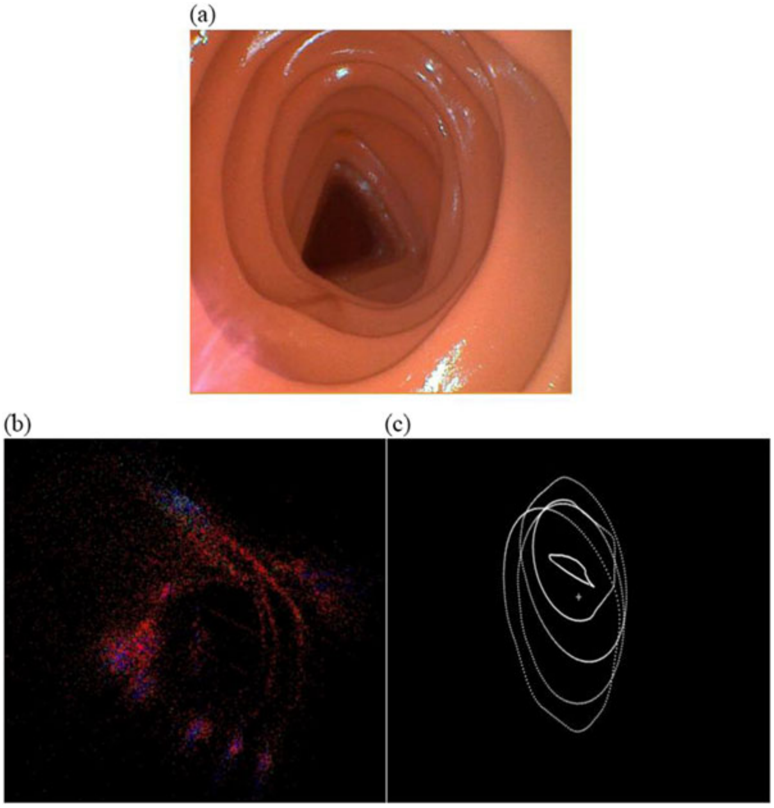

Fig. 8. (a) One of input image; (b) 3D reconstruction based on SfM; (c) 3D reconstruction based on BAM.

shows reconstructed result of SfM is a sparse point cloud with large number of noise points and cannot provide visual colon structure in a straightforward way, while results of BAM show 3D position of fold contours. The reconstructed result based on BAM is more suitable for navigation. In addition, in capsule endoscopy, camera motion is not constant, and the number of prominent feature points between images is usually insufficient, which makes it quite difficult for existing SfM methods to perform well.

To evaluate the performance of BAM in biological intestinal environment, we applied our method to the pig's colon. A section of pig's colon was fixed on the platform, some air was injected into the colon to expand the internal space of it (simulating the endoscopic scenarios). Then the capsule endoscopy recorded video during movement. Because the ground truth value of the pig's colon structure information is unknown, we compare visually the image of our reconstructed result with the captured image to evaluate BAM. Figure 9 shows that rendered model is similar to the captured image, which suggests that the proposed method is suitable for the colon reconstruction with real endoscopic images.

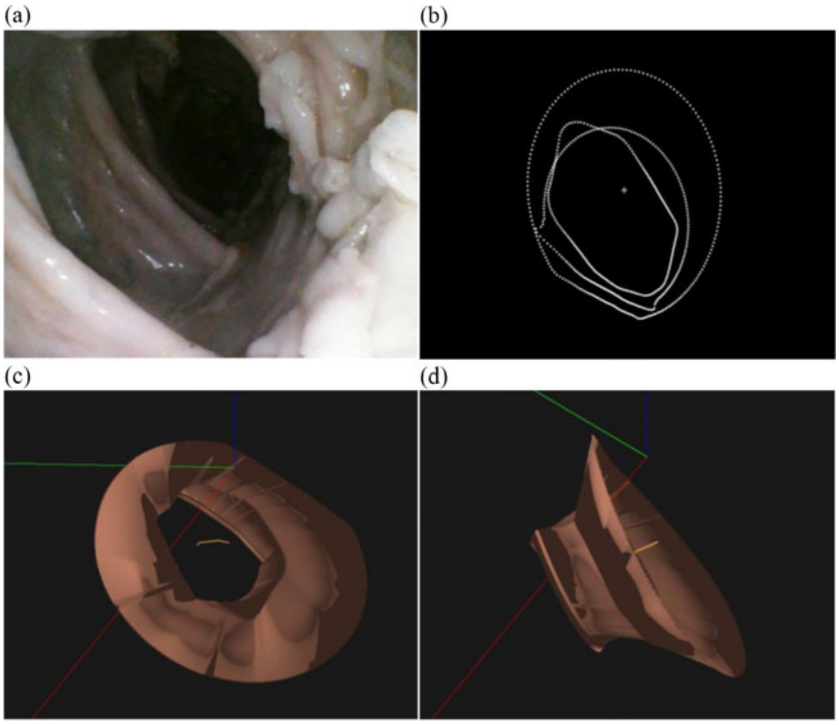

Fig. 9. (a) Image of real pig's colon; (b) 3D reconstructed fold contours; (c) front view of rendered model; (c) side view of rendered model.

5. Conclusions

The 3D reconstruction of the colon structure is useful for several applications, such as navigation path and helping in computer-aided diagnosis. This paper has described a reconstruction method of colon structure based on camera motion from image sequence for screening colonoscopy. We registered 3D points set of the corresponding closed fold contours between multiple frames to obtain the corresponding transformation matrix and calculated the reconstruction result by the matrix. For the evaluation of the reconstruction result, we compared the reconstruction result with a single image and the result with image sequence. The average percentage of the depth error is reduced by 3.2% in the image sequence, and the circumference error which represents the accuracy of surface reconstruction is also reduced by 0.3%. We also compared the reconstructed result with that based on SfM, the difference between the depth errors is not significant, but SfM takes more than five times of the consuming time, even if the results can be achieved, as long as BAM. At the same time, BAM also achieves good reconstruction result in the pig's colon. The results show that the method is superior to the one using single-frame-based brightness intensity. So the proposed method is effective and feasible to 3D reconstruction of the colon structure in colonoscopy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61603045) and the Key Laboratory of Biomimetic Robots and Systems (School of Mechatronical Engineering, Beijing Institute of Technology), Ministry of Education, China.

References

1. L. Smith, "Screening for colorectal cancer: Surveillance after resection of a colorectal cancer, and the removal of large adenomas," *Endoscopy* **17**, 98–102 (1987).
2. G. C. Harewood, "Relationship of colonoscopy completion rates and endoscopist features," *Digestive Dis. Sci.* **50**(1), 47–51 (2005).
3. G. Ciuti, P. Valdastri, A. Menciassi and P. Dario, "Robotic magnetic steering and locomotion of capsule endoscope for diagnostic and surgical endoluminal procedures," *Robotica* **28**(2), 199–207 (2010).
4. Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(8), 1362–1376 (2010).

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press

5. D. Cremers and K. Kolev, “Multi-view stereo and silhouette consistency via convex functionals over convex domains,” *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(6), 1161–1174 (2011).
6. C. Kazó and L. Hajder, “Rapid Weak-Perspective Structure from Motion with Missing Data,” *Proceedings of the IEEE International Conference on Computer Vision Workshops* (2011) pp. 491–498.
7. M. Chandraker, “What Camera Motion Reveals about Shape with Unknown BRDF,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)* (2014) pp. 2179–2186.
8. L. Hajder and D. Chetverikov, “Weak-perspective structure from motion for strongly contaminated data,” *Pattern Recognit. Lett.* **27**(14), 1581–1589 (2006).
9. A. H. Ahmed and A. A. Farag, “Shape from Shading for Hybrid Surfaces,” *Proceedings of the IEEE International Conference on Image Processing* (2007) pp. 525–528.
10. O. Ikeda, “Shape-from-Shading Algorithm for Oblique Light Source,” *Proceedings of the International Symposium on Advances in Visual Computing* (2007) pp. 357–366.
11. X. Ming, R. C. Zhao and P. Maria, “Solving Self-Shadow Problem of Shape from Shading in Light Source Projected System,” *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing* (2004) pp. 334–337.
12. M. VisentiniScarzanella, D. Stoyanov and G. Z. Yang, “Metric Depth Recovery from Monocular Images Using Shape-from-Shading and Specularities,” *Proceedings of the IEEE Conference on Image Processing (ICIP)* (2013) pp. 25–28.
13. G. Ciuti, M. VisentiniScarzanella, A. Dore, A. Menciassi, P. Dario and G. Z. Yang, “Intra-Operative Monocular 3D Reconstruction for Image-Guided Navigation in Active Locomotion Capsule Endoscopy,” *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics* (2012) pp. 768–774.
14. M. A. Armin, N. Barnes, J. Alvarez, H. D. Li, F. Grimpen and O. Salvado, “Learning Camera Pose from Optical Colonoscopy Frames Through Deep Convolutional Neural Network (CNN),” **In**:*Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures. 4th International Workshop* (CARE, 2017) pp. 50–59.
15. F. Mahmood and N. J. Durr, “Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy,” *Med. Image Anal.* **48**, 230–243 (2018).
16. A. Rau, P. Edwards, O. Ahmad, P. Riordan, M. Janatka, L. Lovat and D. Stoyanov, “Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy,” *Int. J. Comput. Assist. Radiol. Surg.* **14**(7), 1167–1176 (2019).
17. D. Hong, W. Tavanapong, J. Wong, J. Oh and P. D. Groen, “3D Reconstruction of virtual colon structures from colonoscopy images,” *Comput. Med. Imaging Graphics* **38**(1), 22–33 (2014).
18. J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698 (1986).
19. R. Carsten and A. Steger, “A comprehensive and versatile camera model for cameras with tilt lenses,” *Int. J. Comput. Vis.* **123**(2), 121–159 (2017).
20. H. Bakstein and T. Pajdla, “Panoramic Mosaicing with a 180 Degree Field of View Lens,” *Proceedings of the IEEE Workshop on Omnidirectional Vision* (2002) pp. 60–68.
21. A. Kaufman and J. Wang, “3D surface reconstruction from endoscopic videos,” **In:** *Visualization in Medicine and Life Sciences* (J. Encarnação, eds.) (Springer Berlin Heidelberg, 2008) pp. 61–74.
22. M. Kazhdan, M. olitho and H. Hoppe, “Poisson Surface Reconstruction,” *Proceedings of the Symposium on Geometry Processing* (2006) pp. 61–70.
23. A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger and C. Schnorr, “Variational optical flow computation in real-time,” *IEEE Trans. Image Process.* **14**(5), 608–615 (2005).
24. H. H. Nagel and W. Enkelmann, “An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences,” *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(5), 565–593 (1986).
25. T. Brox, A. Bruhn, N. Papenberg and J. Weickert, “High Accuracy Optical Flow Estimation Based on a Theory for Warping,” *Proceedings of the European Conference on Computer Vision(ECCV)* (2004) pp. 25–36.

https://doi.org/10.1017/S0263574720000399 Published online by Cambridge University Press