2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

A Track-based Colon Endoscopic Robot with Depth Perception Stereo Cameras for Haustral Fold Detection during Colonic Navigation

Shujing He, Yujie Zhang, Baoyi Huang, Jie Lin, Chaoyang Shi and Chengzhi Hu, *Senior Member, IEEE*

***Abstract*— Colon endoscopic robots represent a promising screening modality for the visualization of colon cancers with high sensitivity. However, current colonoscopy robots are often characterized by intricate and bulky mechanical structures, which pose practical challenges when moving through the complex and narrow environment of the colon. Moreover, these robots are typically equipped with a single camera, limiting their ability to accurately estimate the depth of haustral folds in the colon, which is of great importance for the active colonic navigation of the robots. To address these challenges, we develop a track-based stereoscopic endoscopic robot (TSER) which is equipped with four tracks positioned at the corners of its body. This innovative design maximizes the contact between the tracks and the colon wall, enhancing maneuverability. The tracks are constructed from de-molded polydimethylsiloxane (PDMS) and incorporate micro-patterns on their outer surfaces. We have proposed a straightforward strategy for detecting haustral folds using TSER's stereo camera, which allows for precise identification of their position and depth. The TSER achieves an average motion speed of 9.8 mm/s in a bellows tube that contains silicone oil and a speed of 5.2 mm/s in an *ex-vivo* porcine intestinal segment. Impressively, the TSER boasts an 88.11% accuracy rate in haustral fold depth estimation, surpassing the performance of existing geometric shape fitting methods. These results demonstrate that the TSER holds great potential for effective and efficient movement and inspection within the colon, offering a promising solution for improved colon cancer screening.**

I. Introduction

Colorectal cancer ranks fourth among the common cancers and second among the leading causes of cancer-related deaths worldwide [1]. Objective statistics demonstrate that early detection significantly improves cancer survival rates, which are nearly 100% at Stage 0 and decrease to below 5% at Stage IV. Prompt detection can substantially alleviate the financial burden associated with treatment [2].

Conventional endoscopes remain prevalent in clinical practice, typically featuring narrower diameters and broadened ranges of field of view and rotation angles. However, the utilization of conventional endoscopes has been impeded by several clinical constraints, such as the expense of the procedure, patient discomfort, the need for sedation, restricted maneuverability, and a prolonged learning curve [3]. Various robotic endoscopic systems have been developed to enhance diagnostic methods and reduce operator training time. Over recent decades, numerous studies have concentrated on capsule-sized robotic gastrointestinal diagnostic devices, such as legged robots [4]–[6], paddle-based robots [7], [8], inchworm robots [9], [10], wheeled robots [11], [12], and propeller-based [13], [14] robots. These robotic devices can aid in early cancer detection and can be designed as wireless or tethered systems. The wireless configuration allows the robot to move through the gastrointestinal tract without the hindrance of cables affecting its movement, while the tethered design enhances real-time computational capabilities and ensures an adequate power supply [11], [12]. Among these designs, legged mechanisms typically require a larger size and greater friction against the colon wall to facilitate effective movement [15]. Inchworm mechanisms, on the other hand, are limited in movement within the intestine due to their lack of multiple degrees of freedom. Propeller-type robots have several degrees of freedom but are more advantageous in the stomach environments than in the intestines. Wheeled and tracked robots with particulate designs incorporated on their surfaces have been demonstrated to be effective in the intestinal environment [15], [16]. Tracked robots have an advantage over other designs due to their large contact area with the intestinal wall, which makes them well-suited for use in the intestinal environment [17].

Endoscopic robots typically employ a front camera to relay images from the field of view. Nevertheless, the monocular camera's two-dimensional visual feedback is deficient in binocular parallax, thus requiring operators to rely on limited visual cues, such as texture gradient and dimensional familiarity, to estimate the surgical depth, which can sometimes lead to overshooting the target [18]. Accurate estimation of colon depth is crucial for 3D information reconstruction [19], [20]. When navigating within the colon, the natural shape of the haustral fold is difficult to fit by a precise mathematical formula [21]. Similarly, although endoscopic robot navigation using monocular cameras has been investigated

This work is supported by the open research fund of Guangdong Provincial Key Laboratory of Advanced Biomaterials (2022B1212010003), and the Shenzhen Science and Technology Program (Grant No. JCYJ20190809144013494), and the Science and Technology Program of Guangdong (Grant No. 2021A1515011813 and No. 2022A1515010098), and the Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation ("Climbing Program" Special Funds) (Grant No. pdjh2023c10802). This work is partly supported by the Science, Technology, and Innovation Commission of Shenzhen Municipality under grant no. ZDSYS20200811143601004, and Stable Support Plan Program of Shenzhen Natural Science Fund under Grant 20220815104331001.

S. He, Y. Zhang, B. Huang and C. Hu are with the Shenzhen Key Laboratory of Biomimetic Robotics and Intelligent Systems, Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen, 518055, China, and are also with the Guangdong Provincial Key Laboratory of Human-Augmentation and Rehabilitation Robotics in Universities, Southern University of Science and Technology, Shenzhen, 518055, China. (Corresponding email: hucz@sustech.edu.cn)

J. Lin is with the Guangzhou university of Chinese medicine, Shenzhen, 518034, China.

C. Shi is with the Key Laboratory of Mechanism Theory and Equipment Design of Ministry of Education, School of Mechanical Engineering, Tianjin University, Tianjin, 300072, China.

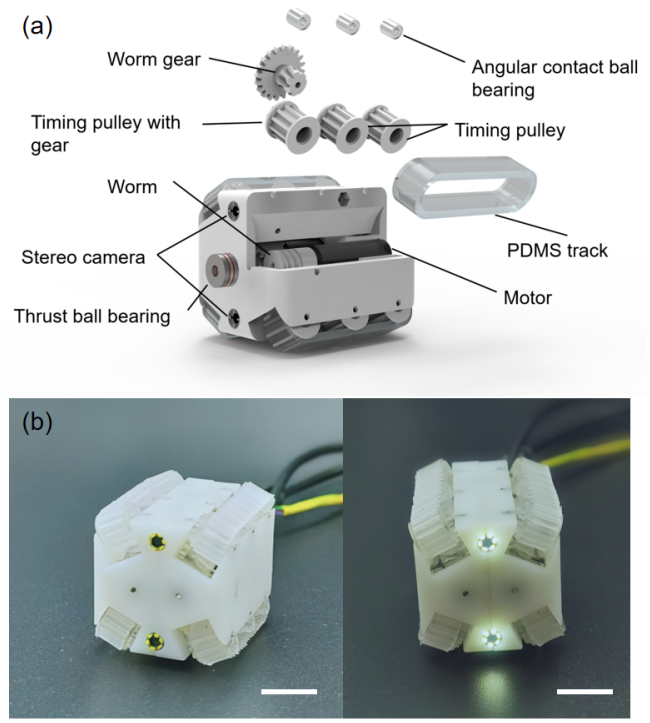

Fig. 1. Structural design of the TSER. (a) The TSER comprises two motors and employs a worm gear mechanism to transmit driving force from the motors to the tracks. To ensure a close fit to the intestinal wall, the polydimethylsiloxane (PDMS) tracks are angled and placed at the corners of the body. Additionally, micro-patterns have been integrated into the PDMS tracks to enhance surface friction. (b) The fully assembled TSER. Scale bar is 1 cm.

in a few studies, the approach often necessitates significant, trustworthy data at the outset for deep learning [20], [22]–[24].

In this paper, we present an optimized design for a TSER system, incorporating two cameras in a compact configuration. We conducted the motion performance test of the device within an *ex-vivo* porcine large intestine. The motility of TSER was assessed in a bellows tube filled with low-viscosity silicone oil to mimic mucus and in an isolated porcine large intestine, achieving speeds of 9.8 mm/s and 5.2 mm/s, respectively. Based on the visual feedback from the stereo camera, we developed a real-time haustral fold detection strategy that uses depth estimation. This strategy exhibits excellent performance in preventing collisions with the intestinal wall through non-learning-based, lightweight, and computationally inexpensive image processing methods. Furthermore, we evaluated the ability of the TSER to perceive haustral folds in an *ex-vivo* porcine large intestine.

II. Mechanism Design

Due to the size constraints of the colon, the design of the TSER necessitates a drive mechanism that is both simple and efficient. Worm gear drives are widely used in many endoscopic robots [7], [15], [17], [25], [26]. This is partly since the worm gear drive has a 90° power transmission, which allows the motors to be placed along the length of the body, thereby reducing the size of the body. Additionally, a single worm gear can effectively drive multiple worm gears simultaneously. Although the utilization of worm gear may cause superimposed axial forces along the worm, it increases the contact between the body and the colon wall and minimizes the risk of drive failure due to loss of contact between parts of the mechanism and the colon wall.

Inspired by a previous design [15], our colon endoscopic robot further reduces the size and increases motion reliability. We decouple the upper and lower tracks on one side and relocate them from the center to the corners of the body to maximize contact between the tracks and the colon wall while leveraging the benefits of the track drive mechanism. The symmetrical design allows the body to maintain contact with the colon wall using tracks placed at the corners of the body, even when navigating a tortuous colon with variable motion directions. Furthermore, this design ensures that the robot remains operational without experiencing motion failures even after flipping. To maintain the TSER's compact size, we have carefully chosen contact angles for the four tracks that deviate from the standard 45°, a modification that does not compromise the robot's mobility after flipping.

The structural design of the TSER is shown in Fig. 1. The TSER is designed to possess rectilinear locomotion capabilities and autonomously navigate and orient itself to advance along its intended path. Accordingly, the TSER has two degrees of freedom by being equipped with two motor drives (ZWPD006006, Shenzhen Zhaowei Machinery & Electronics Co., Ltd., Shenzhen, China). Each motor drives two worm wheels via a worm gear mechanism. These motors are outfitted with gearboxes with a reduction ratio of 136, a no-load speed of 240 r/min, and a rated load of 12.24 mN·m. On the other side of the worm gears, a gear mechanism is integrated to efficiently transfer the driving force to the timing pulleys with minimal mechanical transmission loss while keeping the TSER's height and width compact. The track load is distributed across three timing pulleys. The front timing pulley is equipped with an integrated gear to allow direct power transmission from the worm gears to the track. As previously discussed, the utilization of a single worm to drive multiple worm gears can induce an increased axial force on the worm. To address this, a thrust ball bearing is mounted between the worm and the body. This bearing serves a dual purpose: safeguarding the motor by bearing the axial load and ensuring the stability of component fitting. To avoid any damage to the colonic mucosa, the track is fabricated with PDMS, which is a transparent, biocompatible, and stiffness-tunable silicone elastomer. To increase friction, the outside of the PDMS track was modified with micro-patterns. All parts of the prototype were manufactured using 3D printing. The shell was processed using a Stereo Lithography Apparatus (SLA), and the drive components were processed using multi-jet printing technology (MJP2500Plus, 3D Systems, ROCK HILL, SC) for higher precision.

For a single drive mechanism, which is shown in Fig. 2(a),

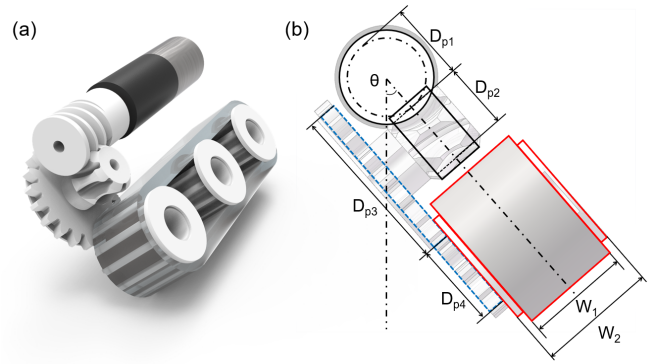

Fig. 2. Analysis of a single transmission mechanism. (a) Illustration of a single transmission mechanism. (b) Description of transmission mechanism parameters.

the angular velocity of the worm gear can be expressed as:

$$\omega_{wg} = \frac{Z_1 \omega_m}{Z_2} \tag{1}$$

where Z_1 is the number of worm leads, Z_2 is the number of worm teeth and ω_m is the motor angular speed. When the worm gear and the gear wheel are integrated, they have the same angular velocity. The angular velocity of a timing pulley, which is also integrated with a gear, can be expressed as follows:

$$\omega_{sw} = \frac{Z_3 \omega_{wg}}{Z_4} \tag{2}$$

where Z_3 is the number of teeth of the gear to be integrated with the worm gear, and Z_4 is the number of teeth of the gear to be integrated with the timing pulley. Thus, the speed V and the maximum driving force F_d of the TSER motion can be expressed as:

$$V = \frac{\omega_m Z_1 Z_3}{Z_2 Z_4} R_t \tag{3}$$

$$F_d = \frac{T_m \eta_{gh} Z_2 Z_4}{Z_1 Z_3 R_t} \tag{4}$$

where T_m represents the stall torque of the motor, η_{gh} represents the efficiency of the gearbox, and R_t is the radius of the timing belt.

Considering the size and manufacturing limitations, we have designed the worm gear drive mechanism with 2 worm leads and 7 worm teeth. This configuration results in a reduction ratio of 3.5, which is favorable for increasing the driving force of the track. The tracks can be made into MXL-size timing belts by de-molded PDMS. To minimize the size of the timing pulley and to ensure that the driving force transmitted to the track is high, the gear integrated with the worm gear is designed with 20 teeth and a module of 0.5. The gear integrated with the timing belt pulley in which it meshes has 10 teeth. The specific parameters of the transmission mechanism are shown in Table I, and the parameters are explained in Fig. 2(b).

At the front of the TSER, a pair of 3.5 mm endoscopes (OV9734, Yichuang Electronics, Shenzhen, China) form a stereo camera system. Each camera has a focal length of 2.8 mm and a field of view of 75°. These cameras provide a stabilized video output at 720P and 30 fps, providing the basis for further stereo-based navigation. Each camera has six LEDs for illumination, which is controlled via USB. The cameras and other cables are tied together from the back of the body. The inertial measurement unit (IMU) (ATK-IMU901, Alientek, Guangzhou, China) is also mounted inside the TSER for motion tracking and orientation estimation.

TABLE I
PARAMETERS USED IN TSER

Parameter	Symbol	Dimension
Width	W	27 mm
Height	H	27 mm
Mass (without tether)	m	22.0 g
Diameter of worm pitch circle	D_{p1}	1 mm
Diameter of worm gear pitch circle	D_{p2}	3.5 mm
Diameter of gear wheel pitch circle	D_{p3}	10 mm
Diameter of pinion pitch circle	D_{p4}	5 mm
Width of track	W_1	6 mm
Width of timing pulley	W_2	7 mm
Contact angle	θ	25.58°

A control board (STM32F103RCT6 72 Mhz ARM Cortex M3) is used and is responsible for motor PWM output and IMU reading to facilitate data acquisition and motor control. The stereo camera is connected directly to the computer via USB for further image processing. The computer communicates with the control board via a serial port.

III. MOTILITY EVALUATION OF THE TSER

To evaluate motility, we conducted the motion tests of the TSER inside a bellows tube with a diameter of 40 mm. The inner surface of the bellows has 3 mm high circular barriers, which allows us to test the capability of the TSER to pass over the obstacles with a tracked drive mechanism without the degree of freedom in the pitch direction. The viscosities of *ex-vivo* and *in-vivo* human colonic mucus were 0.42 ± 0.59 and 0.02 ± 0.20 Pa·s, respectively [27]. To mimic the physiological condition of the human colon, 5 ml of low-viscosity silicone oil (0.65 cSt) was poured into a 40 cm bellows tube. The recorded motion speed of the TSER revealed an average speed of 10.3 mm/s when navigating an empty bellows tube. Upon the introduction of silicone oil, the TSER's speed decreased slightly to 9.8 mm/s, a reduction that remains sufficient to ensure effective propulsion within the colon.

To empirically assess the locomotive capabilities of the TSER within a compliant colon environment, an *ex-vivo* segment of porcine large intestine about 40 cm was selected. As depicted in Fig. 3, the TSER was inserted into the intestine, followed by controlled inflation to mimic examination conditions during clinical endoscopic procedures. Furthermore, an endoscopic apparatus was introduced through the opposite aperture to facilitate real-time monitoring and observation of TSER's locomotion within the intestinal environment. The porcine large intestine did not use any other auxiliary fixation tools or devices. The TSER has two tethers, one connected

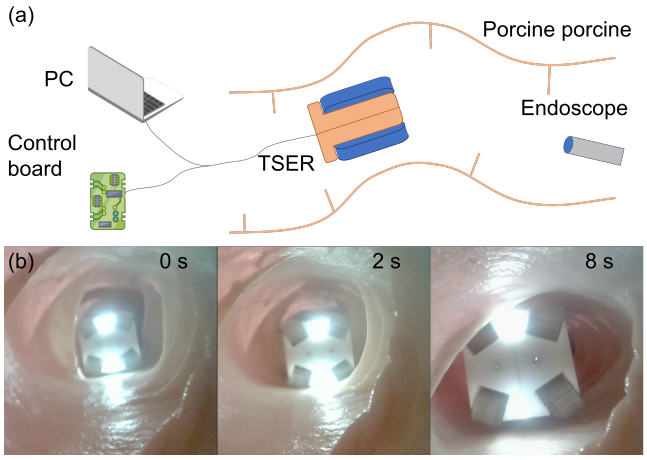

Fig. 3. (a) A schematic diagram of the motility test of the TSER. An additional endoscope is placed within the porcine large intestine to provide visual feedback during the locomotion of the TSER. (b) Motility experiment of the TSER within a segment of *ex-vivo* porcine large intestine.

to the PC for collecting image data and one connected to the control board. The motility of the TSER in the porcine large intestine is shown in Fig. 3(b). The TSER exhibited some slippage during the *ex-vivo* experiment. Due to its lightweight construction, the TSER can be affected by haustral folds, causing the body to tilt. The results indicate that the TSER possesses the ability to move over the soft tissue of the colon. In five porcine large intestine locomotion experiments, successful locomotion was achieved four times, with one failure due to the TSER being obstructed in the haustral segment. In the successful cases, the average speed of the TSER was 5.2 mm/s.

IV. Strategies for Haustral Fold Detection

Accurate depth perception within the colon is needed for fold recognition, endoscopic movement planning, and three-dimensional reconstruction of the colon. Existing endoscopes face limitations in spatial perception, often relying on the operator's estimation of distances in the colon based on the size of the haustral fold contour. The curved nature of the colon and limited illumination have posed significant challenges for operators to accurately discern a navigational path for the endoscope. Moreover, it is commonly considered that the human colon exhibits a triangular haustral fold structure, which is replicated in many colon simulators [28]. However, medical practitioners typically refrain from fully inflating the colon to relieve the patient, making it challenging to maintain a stable triangular shape. Additionally, the inherent complexity of natural systems hinders the mapping of endoscopic navigation with precise mathematical computations [21]. Consequently, a more generalized approach is essential to perceive and execute endoscopic navigation tasks, accommodating individual variations. Our approach leverages depth perception from the stereo camera in conjunction with geometric information from the haustral folds to estimate the current state, enabling the formulation of an appropriate navigation strategy.

Pigs and humans have similar organ physiology and dimensions [29]. Both humans and pigs have sacculations and longitudinal muscle bands in the intestine [30], although the number is not the same. We implemented the strategy for haustral folds detection using an *ex-vivo* porcine large intestine as our experimental model.

A. Stereo Camera System

Cameras in stereo systems are typically positioned in one of two ways: parallel and toed-in [31]. In the case of the TSER, we employ a parallel camera configuration since there is no predefined reference position in the intended usage scenario [32].

The stereo camera is first calibrated using MATLAB. Given that the TSER is equipped with an IMU, the orientation of the camera does not affect the processing of the image algorithms. For simplicity, we assume that the initial state of the TSER camera is horizontally oriented. As the features in the colon are not dense, each image from the left and right cameras is transferred to a PC, and the image size is reduced to 640×480 pixels using bilinear interpolation to speed up the processing rate of the algorithm. Subsequently, the reduced image is rectified, then the image is greyed out, and finally, a Gaussian blur filter is applied to the image to complete the image pre-processing.

The disparity between the left and right images is calculated using the Semi-Global Block Matching (SGBM) algorithm [33], which is a semi-global matching algorithm. To preserve precise haustral fold edge geometry features, an edge extraction process is applied after the binarization of the computed disparity image. This process utilizes the rectified image as a mask to perform a bitwise AND operation on the resultant depth image. This procedure facilitates the integration of robust geometric contour information with depth data, compensating for the loss of reliable edge geometry features in the computed disparity image. Since the depth information of each pixel point from the same haustral fold is closely spaced within the depth image, we utilize a clustering algorithm known as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) in this step. DBSCAN can help to delineate the distribution of depth information within the current field of view, ensuring accurate interpretation of depth data.

After computing contour features at a specific depth, they can be utilized as obstacles along the TSER's trajectory during motion state planning. However, it's crucial to ascertain the orientation of each crescent-shaped fold (left or right). This determination enables the TSER to discern the direction of motion for each haustral fold. This step is done by determining the direction in which the gray gradient decreases within the grayscale image. After outlining the features with rectangles, the depth information and geometric information are merged using the following equation:

$$w_{new} = -0.5D + 1.2w_{rec} + 30, \tag{5}$$

where w_{new} is the width of the rectangle with the same height as the original rectangle, D represents the average

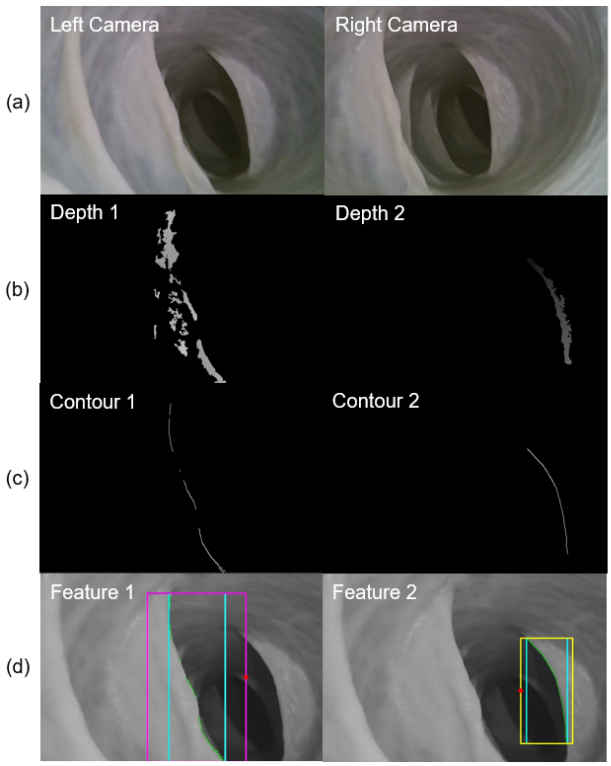

Fig. 4. Processing flow for stereo images. (a) Original images from the left and right cameras. (b) Segmented depth image, and after the bitwise AND operation with the binary image in (c), the features and the key points of the features are found in (d).

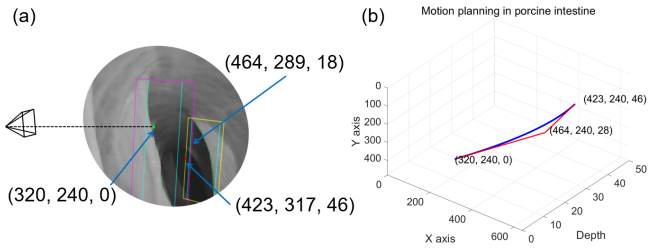

Fig. 5. Motion planning based on Bezier curves. The key points are marked in (a). Since there are no degrees of freedom on the pitch, the key points in the three-dimensional space are reduced to two dimensions, resulting in the key points having the same Y-coordinate value. And the path is optimized (blue curve) using the Bezier curve (b).

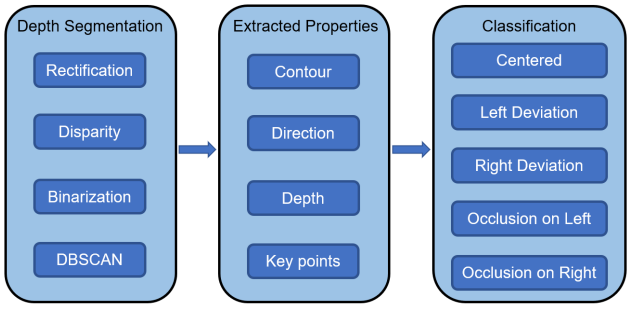

Fig. 6. Flowchat of stereo camera-based feature detection strategy. The process involves segmenting haustral folds using depth information, extracting relevant properties, and classifying the current motion state.

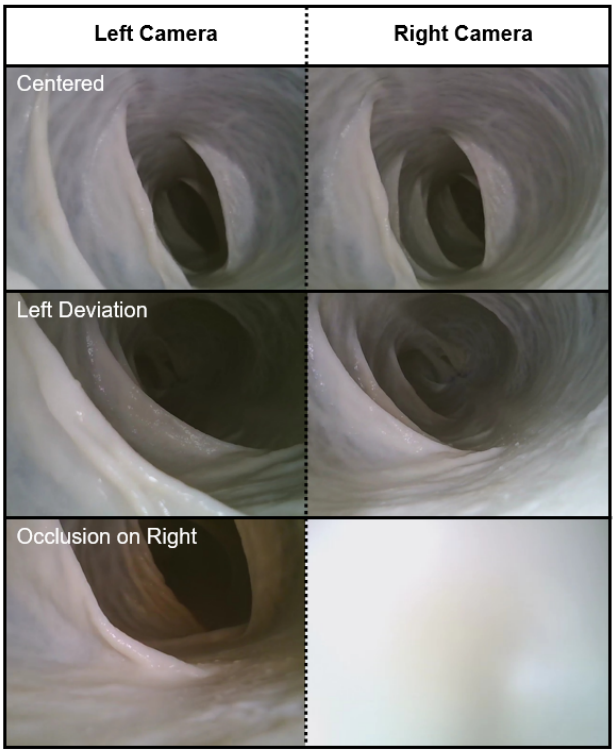

Fig. 7. Classification of motion states using the porcine large intestine as an example. The classification process relies on assessing the presence or absence of contours in the current field of view, along with the directions of the haustral folds.

depth of the feature, and w_{rec} is the width of the original rectangle. This step allows an offset for the movement of the TSER. The closer to the haustral fold, the greater the offset required. As the TSER has two degrees of freedom and lacks pitch control, the center of the border of the newly adjusted rectangle is used as the key point. The complete stereo image processing flow is shown in Fig. 4. Fig. 4(a) is the original left and right camera transmitted image. Fig. 4(b) shows the depth image after computation. At this stage, the image has no reliable geometric edge information. Fig. 4(c) is the contour information at different depths after bitwise AND operation. Fig. 4(d) shows the adjusted feature rectangles and the key points (red points). Based on the key points, Bezier curves are used to optimize the paths, shown in Fig. 5.

B. Classification of the Motion State

As shown in Fig. 6, the current motion state is classified after the features are extracted. Five motion states of the TSER in the porcine large intestine can be classified, as represented in Fig. 7.

- *Centered*: The TSER primarily undertakes motion along the central axis of the colon tract. There are contours in the original image, and there are crescent-shaped haustral folds with opposite directions in the image after the bitwise AND operation.

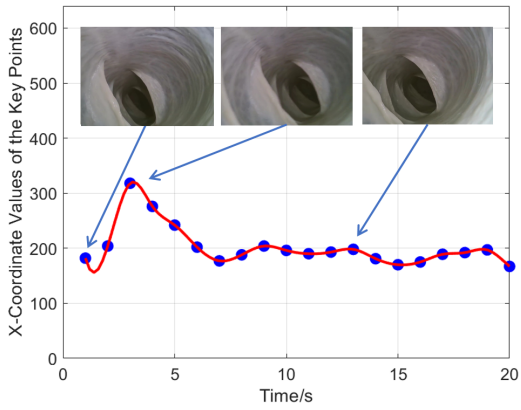

Fig. 8. The X-coordinate values of the tracked haustral fold in the camera coordinate system as the TSER passes through the haustral fold.

- *Left/Right Deviation*: The TSER moves to one side of the colon wall. Contours are present in the original image, but the crescent-shaped fold features of the bitwise AND operation image are oriented in the same direction.
- *Occlusion on Left/Right*: The TSER is occluded in the haustral structure. One side of the camera is obscured, resulting in the absence of contours in the original image.

C. Evaluation of Haustral Fold Detection

For depth perception using the stereoscopic camera within the porcine large intestine, the TSER was placed along the center axis of the intestine with a slight rotation. Out of the 1884 frames of data gathered using the stereo camera, depth information was successfully computed for every frame. Among these, 224 frames exhibited inaccurate depth perception of the haustral folds. We attribute this discrepancy to one side of the camera being close to the intestinal wall. The overall success rate for depth perception reached 88.11%. The processing speed of the image algorithm is 15 Hz (Intel Core i5-8500 CPU).

For the tracking of individual haustral folds, we selected a video recorded by the stereo camera and sliced it to calculate the change in the X-coordinate value of the extracted key point of a specific haustral fold in the camera coordinate system as the TSER traversed it. Since the TSER has no degrees of freedom in pitch, the change in the Y-coordinate value is not recorded. A total of 20 frames were extracted from the video, and the computed results are shown in Fig. 8. The blue dots indicate the estimated X-coordinate values at the critical points of the haustral fold, and the red curve shows the change in the fitted values. The X-coordinate value increases as the TSER approaches the haustral fold, followed by a decrease in the X-coordinate value as the TSER turns to avoid the collision.

V. Conclusions & Future Work

In this paper, we present the track-based stereoscopic endoscopic robot (TSER), a colonoscopy robot equipped with a tracked mechanism and a stereo camera. The TSER is driven by two DC motors and has two degrees of freedom. It has four tracks made of PDMS distributed around the body to increase the reliability of movement in the colon. We detail the design of the TSER and propose a depth-based feature detection strategy for application in the colon, which is envisioned for potential use in autonomous colonoscopy navigation. The motility of the TSER was evaluated in both a bellows tube supplemented with silicone oil and an inflated *ex-vivo* porcine large intestine, simulating clinical endoscopy conditions. The TSER achieved impressive motility speeds of 9.8 mm/s and 5.2 mm/s, respectively. Based on the stereo camera in the TSER, the depth perception of haustral folds was evaluated within the porcine large intestine. The depth perception of the folds reached 88.11%. The experimental results show that the TSER has the ability to move through the complex colon environment and perceive haustral folds for autonomous navigation. Although the TSER is a miniaturized robotic system with a tether, it maintains its functionality effectively. Some slippage was observed as the TSER traversed the porcine large intestine, and its lightweight nature occasionally caused the TSER's tracks to partially suspend when the porcine large intestine was over-inflated. Future work will focus on the design of an adaptive diameter mechanism and investigate how to increase friction to further ensure the stability of TSER motion, in addition to the development of novel navigation algorithms to further enhance the TSER's performance in clinical applications.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] F. Van Erning, L. Van Steenbergen, V. Lemmens, H. Rutten, H. Martijn, D. J. van Spronsen, and M. Janssen-Heijnen, "Conditional survival for long-term colorectal cancer survivors in the netherlands: who do best?" *European Journal of Cancer*, vol. 50, no. 10, pp. 1731–1739, 2014.

[3] A. Alazmani, A. Hood, D. Jayne, A. Neville, and P. Culmer, "Quantitative assessment of colorectal morphology: Implications for robotic colonoscopy," *Medical Engineering & Physics*, vol. 38, no. 2, pp. 148–154, 2016.

[4] S. Gorini, M. Quirini, A. Menciassi, G. Pernorio, C. Stefanini, and P. Dario, "A novel sma-based actuator for a legged endoscopic capsule," in *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006. BioRob 2006.* IEEE, 2006, pp. 443–449.

[5] M. Simi, P. Valdastri, C. Quaglia, A. Menciassi, and P. Dario, "Design, fabrication, and testing of a capsule with hybrid locomotion for gastrointestinal tract exploration," *IEEe/ASME Transactions on Mechatronics*, vol. 15, no. 2, pp. 170–180, 2010.

[6] P. Valdastri, R. J. Webster, C. Quaglia, M. Quirini, A. Menciassi, and P. Dario, "A new mechanism for mesoscale legged locomotion in compliant tubular environments," *IEEE Transactions on Robotics*, vol. 25, no. 5, pp. 1047–1057, 2009.

[7] K. Osawa, R. Nakadate, J. Arata, Y. Nagao, T. Akahoshi, M. Eto, and M. Hashizume, "Self-propelled colonoscopy robot using flexible paddles," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6710–6716, 2020.

[8] K. Makino, H. Watanabe, T. Yamaguchi, H. Terada, and N. Sekiya, "Study on fins knit by two kinds of biodegradable string for a manipulatable endoscope," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE, 2021, pp. 448–451.

[9] J. A. Steiner, L. N. Pham, J. J. Abbott, and K. K. Leang, "Modeling and analysis of a soft endoluminal inchworm robot propelled by a rotating magnetic dipole field," *Journal of Mechanisms and Robotics*, vol. 14, no. 5, p. 051002, 2022.

[10] J. Chen, J. Yang, F. Qian, Q. Lu, Y. Guo, Z. Sun, and C. Chen, "A novel inchworm-inspired soft robotic colonoscope based on a rubber bellows," *Micromachines*, vol. 13, no. 4, p. 635, 2022.

[11] J. M. Prendergast, G. A. Formosa, and M. E. Rentschler, "A platform for developing robotic navigation strategies in a deformable, dynamic environment," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2670–2677, 2018.

[12] J. M. Prendergast, G. A. Formosa, C. R. Heckman, and M. E. Rentschler, "Autonomous localization, navigation and haustral fold detection for robotic endoscopy," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 783–790.

[13] Y. Zhang, W. Wang, W. Ke, and C. Hu, "Optimized design and analysis of active propeller-driven capsule endoscopic robot for gastric examination," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 4689–4695.

[14] Y. Zhang, Z. Li, W. Ke, and C. Hu, "Development of a compact autonomous propeller-driven capsule robot for noninvasive gastric endoscopic examination," in *2022 IEEE International Conference on Cyborg and Bionic Systems (CBS)*. IEEE, 2023, pp. 1–6.

[15] G. A. Formosa, J. M. Prendergast, S. A. Edmundowicz, and M. E. Rentschler, "Novel optimization-based design and surgical evaluation of a treaded robotic capsule colonoscope," *IEEE Transactions on Robotics*, vol. 36, no. 2, pp. 545–552, 2020.

[16] J. M. Prendergast, G. A. Formosa, M. J. Fulton, C. R. Heckman, and M. E. Rentschler, "A real-time state dependent region estimator for autonomous endoscope navigation," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 918–934, 2021.

[17] V. Consumi, L. Lindenroth, J. Merlin, D. Stoyanov, and A. Stilli, "Design and evaluation of the softscreen capsule for colonoscopy," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1659–1666, 2023.

[18] R. Bogdanova, P. Boulanger, and B. Zheng, "Depth perception of surgeons in minimally invasive surgery," *Surgical Innovation*, vol. 23, no. 5, pp. 515–524, 2016.

[19] S. Zhang, L. Zhao, S. Huang, M. Ye, and Q. Hao, "A template-based 3d reconstruction of colon structures and textures from stereo colonoscopic images," *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 1, pp. 85–95, 2020.

[20] K. B. Ozyoruk, G. I. Gokceler, T. L. Bobrow, G. Coskun, K. Incetan, Y. Almalioglu, F. Mahmood, E. Curto, L. Perdigoto, M. Oliveira *et al.*, "Endoslam dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos," *Medical Image Analysis*, vol. 71, p. 102058, 2021.

[21] G. N. Khan and D. F. Gillies, "Vision based navigation system for an endoscope," *Image and Vision Computing*, vol. 14, no. 10, pp. 763–772, 1996.

[22] T. Yang, Y. Yang, P. Wang, Y. Cao, Z. Yang, and H. Liu, "A lumen-adapted navigation scheme with spatial awareness from monocular vision for autonomous robotic endoscopy," *Robotics and Autonomous Systems*, vol. 165, p. 104444, 2023.

[23] M. Yip, S. Salcudean, K. Goldberg, K. Althoefer, A. Menciassi, J. D. Opfermann, A. Krieger, K. Swaminathan, C. J. Walsh, H. Huang *et al.*, "Artificial intelligence meets medical robotics," *Science*, vol. 381, no. 6654, pp. 141–146, 2023.

[24] S.-J. Hwang, S.-J. Park, G.-M. Kim, and J.-H. Baek, "Unsupervised monocular depth estimation for colonoscope system using feedback network," *Sensors*, vol. 21, no. 8, p. 2691, 2021.

[25] L. Sliker, M. Kern, J. Schoen, and M. Rentschler, "Surgical evaluation of a novel tethered robotic capsule endoscope using micro-patterned treads," *Surgical Endoscopy*, vol. 26, pp. 2862–9, 04 2012.

[26] D. Kim, D. Lee, B. Kim, and B.-I. Lee, "A self-propelled robotic colonoscope using elastic caterpillars," in *IEEE ISR 2013*. IEEE, 2013, pp. 1–4.

[27] R. L. Howard, M. Markovetz, Y. Wang, C. Ehre, S. Z. Sheikh, N. L. Allbritton, and D. B. Hill, "Biochemical and rheological analysis of human colonic culture mucus reveals similarity to gut mucus," *Biophysical Journal*, vol. 120, no. 23, pp. 5384–5394, 2021.

[28] G. A. Formosa, J. M. Prendergast, J. Peng, D. Kirkpatrick, and M. E. Rentschler, "A modular endoscopy simulation apparatus (mesa) for robotic medical device sensing and control validation," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4054–4061, 2018.

[29] T. Flisikowska, C. Merkl, M. Landmann, S. Eser, N. Rezaei, X. Cui, M. Kurome, V. Zakhartchenko, B. Kessler, H. Wieland *et al.*, "A porcine model of familial adenomatous polyposis," *Gastroenterology*, vol. 143, no. 5, pp. 1173–1175, 2012.

[30] L. M. Gonzalez, A. J. Moeser, and A. T. Blikslager, "Porcine models of digestive disease: the future of large animal translational research," *Translational Research*, vol. 166, no. 1, pp. 12–27, 2015.

[31] M. Ferre, R. Aracil, and M. A. Sanchez-Uran, "Stereoscopic human interfaces," *IEEE Robotics & Automation Magazine*, vol. 15, no. 4, pp. 50–57, 2008.

[32] M. Simi, M. Silvestri, C. Cavallotti, M. Vatteroni, P. Valdastri, A. Menciassi, and P. Dario, "Magnetically activated stereoscopic vision system for laparoendoscopic single-site surgery," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 3, pp. 1140–1151, 2012.

[33] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2007.