Autonomous Localization, Navigation and Haustral Fold Detection for Robotic Endoscopy

J. Micah Prendergast, *Student Member, IEEE,* Gregory A. Formosa, *Student Member, IEEE,* Christoffer R. Heckman, Mark E. Rentschler, *Senior Member, IEEE*

***Abstract*- Capsule endoscopes have gained popularity over the last decade as minimally invasive devices for diagnosing gastrointestinal abnormalities such as colorectal cancer. While this technology offers a less invasive and more convenient alternative to traditional scopes, these capsules are only able to provide observational capabilities due to their passive nature. With the addition of a reliable mobility system and a real-time navigation system, capsule endoscopes could transform from observational devices into active surgical tools, offering biopsy and therapeutic capabilities and even autonomous navigation in a single minimally invasive device. In this work, a vision system is developed to allow for autonomous lumen center tracking and haustral fold identification and tracking during colonoscopy. This system is tested for its ability to accurately identify and track multiple haustral folds across many frames in both simulated and *in vivo* video, and the lumen center tracking is tested onboard a robotic endoscope platform (REP) within an active simulator to demonstrate autonomous navigation. In addition, real-time localization is demonstrated using open source ORB-SLAM2. The vision system successfully identified 95.6% of Haustral folds in simulator frames and 70.6% in *in vivo* frames and false positives occurred in less than 1% of frames. The center tracking algorithm showed *in vivo* center estimates within a mean error of 6.6% of physician estimates and allowed for the REP to traverse 2 m of the active simulator in 6 minutes without intervention.**

I. Introduction

Gastrointestinal (GI) diseases are diagnosed and treated primarily through the use of a traditional endoscope or colonoscope. While less invasive alternatives such as wireless capsule endoscopes (CEs) are becoming more prevalent, these devices are limited, via their passive nature, to observational techniques and offer nothing in the way of therapeutic capabilities [1]. Recent research efforts have focused on the development of a mobility and sensory feedback system for such a device, to enable autonomous or semi-autonomous motion control and navigation *in vivo*. These efforts support the goal of designing a truly robotic capsule endoscope (RCE) or colonoscope to improve this minimally invasive procedure.

The development of our scaled robotic endoscope platform (REP) [2] and the modular endoscopy simulation apparatus (MESA) [3] has enabled the rapid development and testing of a multitude of sensing and navigation strategies. This paper presents an upgraded platform, REP-2, as well as a novel vision system able to track the lumen center for autonomous navigation while also identifying and tracking haustral folds to enable careful observation by the physician. In addition, this paper presents the use of a monocular localization strategy using ORB-SLAM2 within the MESA simulator.

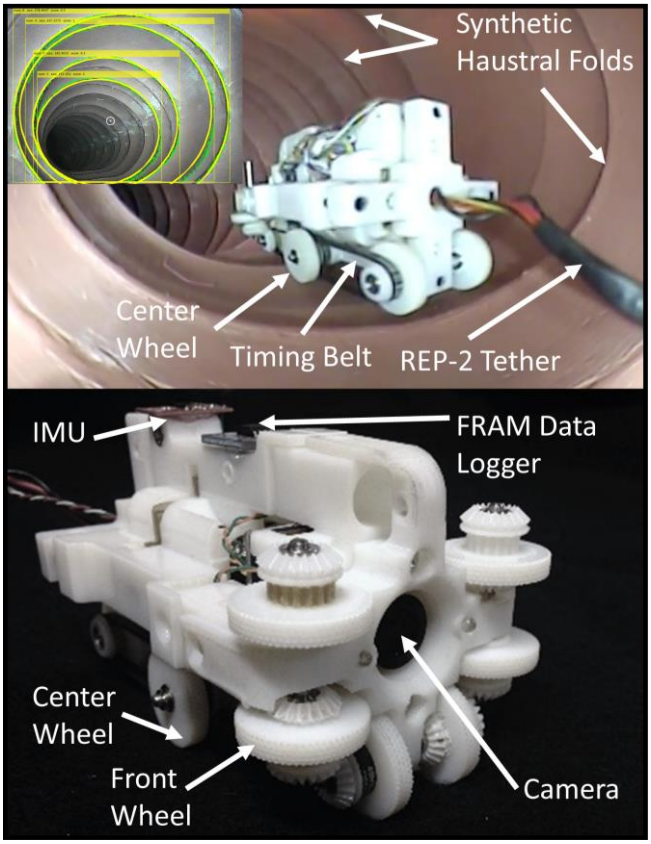

Figure 1. Robotic endoscope platform (REP) inside of the Modular Endoscopy Simulation Apparatus (MESA) with inset showing haustral fold identification and tracking (top) and front view of the REP-2 showing the additional center wheel to improve mobility over obstacles (bottom). Top wheels and belts are not shown as these were not used during testing.

A. Motivation

Although colorectal cancer (CRC) is treatable with a relative five-year survival rate of 90% if detected in the early and localized stage, it remains the third leading cause of cancer deaths in the U.S. for both men and women [4]. The

J. M. Prendergast is with the Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309 USA (Phone: 828-230-0833; e-mail: Joseph.Prendergast@colorado.edu).

G. A. Formosa is with the Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309 USA (email: Gregory.Formosa@colorado.edu)

C. R. Heckman is with the Department of Computer Science, University of Colorado, Boulder, CO 80309 USA (email: Christoffer.Heckman@colorado.edu)

M. E. Rentschler is with the Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309 USA (email: Mark.Rentschler@colorado.edu).

most effective method of early detection is regular screenings, which serve not only to identify CRC, but also to prevent it, by removing potentially precancerous polyps [5]. The colonoscopy remains the most effective screening method for CRC, allowing a surgeon to locate and remove suspect polyps by inserting a scope into a patient's anus, rectum and colon [5]. However, despite recommendations from the American Cancer Society that adults over 50 years of age receive a colonoscopy once every ten years, a 2010 study showed that only 56% of adults 50 years or older had actually had the procedure within that timeframe [4].

Colonoscopies can be uncomfortable for both patient and physicians. It is estimated that up to 90% of patient pain during colonoscopies is caused by looping, where the colonoscope continues to advance into the colon without a simultaneous progression of the tip. This leads to colon distension and in severe cases, can cause tissue damage and perforation [6]. In an attempt to both decrease the invasiveness of this procedure and improve the screening procedure as a whole, capsule devices have been developed.

B. Capsule Endoscopy

Wireless capsule endoscopes offer a less invasive alternative to traditional endoscopes. Once this pill-sized device is swallowed by a patient onboard cameras are able to transmit images of the patient's GI tract to a wireless receiver where they can then be analyzed by a physician. While several wireless capsule endoscopes are currently on the market, all of these devices rely on passive locomotion to progress through the GI tract [1]. Due to their passive nature, these devices are unable to interact with the tissue they observe, lacking the ability to take biopsies or administer therapeutics [7]. Furthermore, because the positioning and orientation of these devices is not controlled, important diagnostic information (*i.e.*, precancerous or cancerous polyps) can be missed, resulting in a diagnostic accuracy less than that of conventional scopes [7]. Implementation of a robust and reliable mobility system for a robotic capsule would help to avoid many of these issues, enabling controllability and making way for more detailed imaging as well as biopsy and therapeutic capabilities.

While a wide range of locomotion methods for RCE's have been attempted including legged [8-11], earthworm [12], inchworm [13-14], treaded/wheeled [15-16] and magnetically linked [17-18] prototypes, these devices have yet to achieve satisfactory mobility *in vivo* and typically lack the sensory feedback that would allow for closed-loop control, thus, potentially complicating an already difficult procedure. Our group has had success with *in vivo* wheeled and treaded robotic mobility using micro-pillared polydimethylsiloxane (PDMS) materials [19-20]. Recent efforts have focused on the benchtop scaling of these devices to larger dimensions for ease of manufacturing and reduced cost [2], while enabling the addition of a multitude of sensors and actuators for improved feedback and automation capabilities.

The device featured in this work incorporates four DC motors with encoders and current sensors, a 6 degree-of-freedom (DOF) inertial measurement unit (IMU), a 3 DOF magnetometer, a monocular vision system and an onboard, Wi-Fi capable microcontroller, all housed in a prototype scaled to be approximately 2x the size of the device from previous work [15]. The many features of this device allow for a diverse range of benchtop tests and experiments to assess the value of different sensing and state estimation strategies as well as the implementation of various control laws to improve mobility and navigation. While the device itself is too large for *in vivo* testing, it serves as an invaluable benchtop testing platform for the assessment of device designs.

C. Sensing and Navigation

The unique GI environment presents sensing and navigation challenges not addressed in conventional field robotics. Although attempts have been made at localization and navigation of *in vivo* devices, these techniques have typically relied on external hardware to track endoscopic devices which can be very difficult given the lack of line-of-sight to the device [21-22] and the necessity of large external hardware. Alternatively, onboard navigation and localization via computer vision often relies on the detection of clear and discrete features such as corners and edges which are typically non-deformable and ever present in the world we experience. While vision in the *in vivo* environment has only recently become an active area of research, the deformable nature of this environment and sparseness of features may pose significant challenges to conventional feature detection approaches [23]. Despite these difficulties, attempts at detecting and tracking using images from within the lumen have been done using techniques such as optical flow and shape from shading [24-29]. 3D reconstruction of stomach simulator images and *ex vivo* stomach tissue has also been shown using sparse-then-dense feature tracking [30]. Similarly, structure from motion has been attempted using images from conventional endoscopy [31-32] and simulated environments [33] as well as using fused visual inertial estimates from a capsule endoscope in a GI simulator [34]. While success has varied, in general, these attempts have been computationally very expensive and most have not been done in real-time or tested on actual devices.

A crucial goal of the REP/MESA system is to evaluate and improve the way future robotic endoscopic devices sense and respond to the deformable GI environment. To improve endoscopy procedures, RCE's must be designed with the capacity to maintain objects of interest (*i.e.*, polyps or diseased tissue) in view, avoid progression into diverticula or into the GI wall (which can cause distension or even perforation), and respond safely (without inducing tissue damage) under the presence of unanticipated disturbances (due to patient movement or peristalsis). In addition, these devices have the potential to significantly simplify user input and reduce training barriers to physicians via autonomous or semi-autonomous navigation, localization and control and may even offer disease identification and environmental mapping.

To achieve these goals, the use of advanced sensing and environmental feedback are requisite additions to these devices. Fortunately, advances in robotic vision may offer many of these capabilities [35]. In addition to autonomous navigation, visual cues allowing for localization within the colon will enable physicians to know not only the location of an RCE during a procedure, but potentially, where the RCE found polyps or took biopsies in previous procedures. Typically, the removal of suspect polyps and other biopsy locations is identified via a small tattoo applied to the colon wall. While this visual cue allows a physician to carefully screen this region in future procedures, these tattoos can be difficult to find. A more robust method for localizing within the colon may better enable a physician to locate these important regions during repeated inspections. In addition, as haustral folds often occlude suspect polyps, real-time information informing a physician about which haustral folds have already been screened (front and back) during a procedure may also prove useful. To accomplish this task, haustral fold tracking across multiple image frames is necessary.

The device featured in this work utilizes a computationally inexpensive method for identifying the haustral folds of the colon lumen and tracking them across multiple image frames. This allows for accurate identification of the lumen's center in real-time, enabling the robot to utilize this information in its feedback loop, and providing a method for closed-loop navigation/visual-servoing through the colon. In addition, monocular Simultaneous Localization and Mapping (SLAM) is applied using the open source ORB-SLAM2 library. This feature based monocular SLAM approach utilizes key-frames, and bundle adjustment to compute pose and sparse point reconstruction up to a scale [36], allowing for real-time localization of the REP-2.

II. Design

Design of the REP-2 is divided into four sections. The first section describes the updated REP-2 chassis and drivetrain and is followed by a brief description of the MESA. Finally, the design of the vision system is described in detail.

A. A Scaled Robotic Endoscope Platform

The REP was designed to be a 2x scaled version of the RCE from [15]. The original REP featured four independently controlled motors to allow for tank steering regardless of the device's orientation. This device, detailed in [2], housed a wireless microcontroller, two dual motor drivers, a 6 DOF IMU and 3 DOF magnetometer, current sensors, motor encoders, FRAM memory module for data logging, and a camera for visual feedback. This device unfortunately was prone to roll-overs and high-centering in the presence of obstacles and thus the drive system of this device has been improved to accommodate the more challenging task of traversing large haustral folds. In addition, the camera has been updated to improve image processing capabilities as described below. This second iteration of the REP, (REP-2) is otherwise identical to the original (electronics, data flow and general operation).

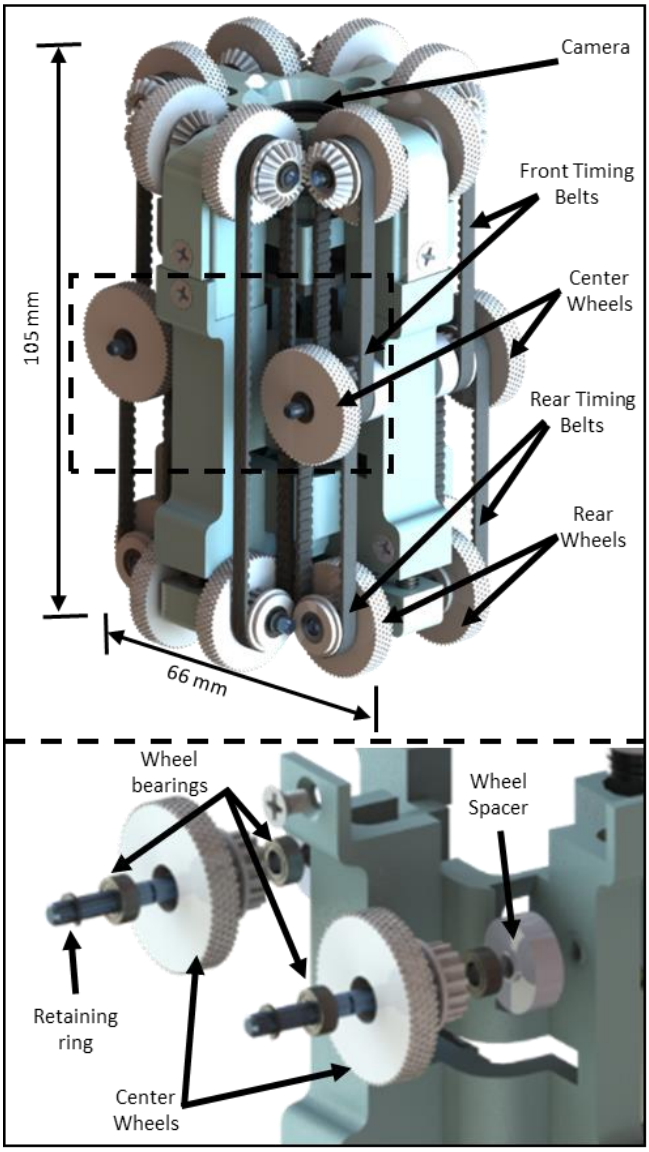

Figure 2. 3D rendering of the REP-2. Updated drive train and clamshell housing are shown (top). Exploded view of the center wheel upgrade is shown in detail (bottom). These modifications improve stability and reduce high centering over simulated haustral folds.

To enable the REP to traverse large simulated haustral folds the original drive system was modified to prevent high centering and to improve stability. An updated clamshell housing allows for the addition of two center shafts on the top and bottom of the device. One center wheels supported by wheel bearings fits over each shaft and is locked in place by a small retaining ring. These center wheels interface via timing belts to their respected driving wheel in the front of the device as well as to the rear wheels of the device. This modification allows the REP-2 to maintain traction even as its center reaches the peak of large haustral folds. In addition, these center wheels increase the overall wheel base width of the device from 21 mm to 46 mm, improving the device's stability substantially and reducing its propensity to roll in the presence of obstacles. Finally, to improve the device's ability to utilize robotic vision, the original CCD interlaced scanning camera was replaced with a 1080p CMOS progressive

scanning camera (ELP-USBFHD01M-L2, Ailipu Technology Co., Shenzhen, Guangdong, China). This camera demonstrated significantly improved performance from the original, particularly in the presence of the rapid motions experienced by the device while moving over the haustral folds.

B. MESA Benchtop Simulator

The MESA system used in this study is a scaled active GI simulator environment and has the ability to induce significant deformation in the silicone synthetic tissue (more than 15 cm in 0.5 seconds). This simulator also has the ability to induce peristalsis type deformation of the colon at wave propagation speeds of 7.7 cm/s. As noted in [3] the synthetic colon on the MESA features haustral fold obstacles designed to replicate the size, shape and distribution of typical folds found in the human colon. The synthetic, highly deformable colon is also insufflatable and is approximately twice the length and diameter of the human colon. The MESA serves as a scaled testing environment for the REP-2, allowing for the assessment of navigation and localization strategies on a low cost platform that offers repeatability and directly measurable ground truth, while also providing the chance to increase or reduce the difficulty of the testing environment via changing geometries or adding disturbances such as external motion, or peristalsis simulation.

C. Sensing and Vision System

The vision system for the REP-2 is similar to the lumen tracking strategy featured in [2] however while this previous strategy was adequate for identifying static folds and estimating the lumen center, it was hampered by false positives, offered no means for tracking specific folds across multiple frames and produced poor center estimates in non-circular and deformable environments. To address these issues, a more advanced edge detection algorithm has been adopted for frame segmentation and significant functionality has been added to this system to allow for haustral fold tracking across multiple frames as well as ellipse fitting rather than circle fitting to more appropriately describe the shape of the identified haustral folds and more accurately estimate the lumen center. In addition, localization capabilities have been incorporated into this system to allow for real-time location estimates using the open source ORB-SLAM2 monocular vision system.

The process to identify haustral folds (Fig. 3) is as follows: (1) Prevalent haustral folds in each frame are detected using Canny edge detection. Each RGB frame is first converted to gray scale where it is then smoothed using a Gaussian filter. Canny's intensity gradient approach is then used to detect the edges in the image. This first step is relatively liberal to ensure that as many complete haustral folds as possible are identified. (2) Once edge detection is complete, all segments below a fixed size threshold (500 pixels) are removed. This step is important in reducing noise and false positives, ensuring only good haustral fold estimates are selected. The threshold selected is based on the size of the original RGB image so as to allow for the use of a range of cameras and

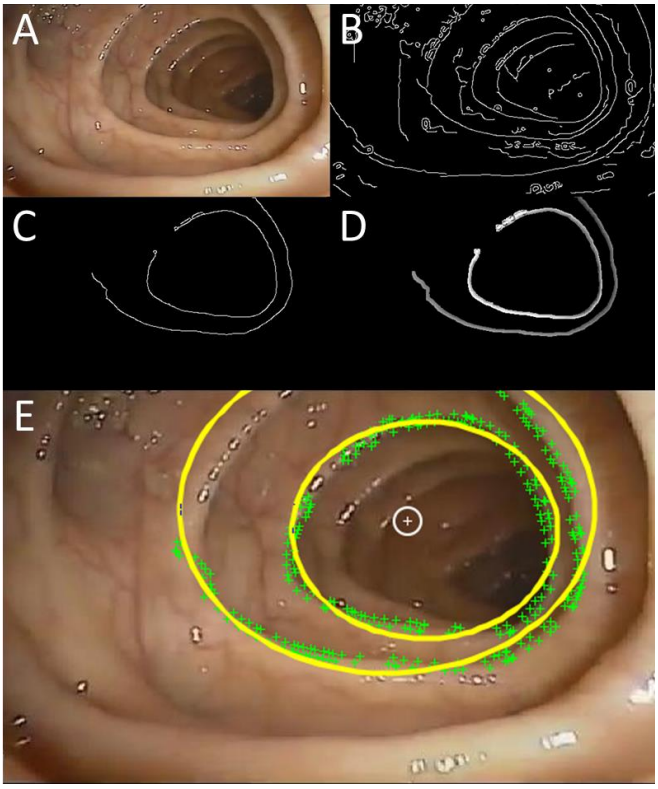

Figure 3. Haustral fold tracking algorithm applied to an image from a human colonoscopy procedure. The original image (A) is processed via Canny edge detection (B). Segments below a set pixel threshold are removed (C) and ellipses are fit to the remaining segments. All segments to which ellipses have been successfully fit are dilated to form an image mask (D). These dilated segments are then passed to a KLT tracking algorithm to be tracked across frames (E).

previously recorded video. (3) An ellipse fit is done on all remaining segments. The mean center of these ellipses serves as the center estimate of the lumen and the standard deviation of this estimate is used to determine the accuracy of this estimate. Any segments to which an ellipse cannot be fit are removed and not used. (4) The segments identified in step 2 are next used as a mask to determine relevant points to track for each segment. Each segment is dilated via a disk structuring element with a radius of 3 pixels, and each of these dilated segments is used to mask off the original image such that only pixels within the dilated segment are used. Minimum eigenvalue feature detection is used to detect features on these masked images and a multi-object tracker based on the Kanade-Lucas-Tomasi algorithm [37] is used to determine which segments have been seen in previous frames and which segments represent new haustral folds to be tracked in subsequent frames. New segments are added to the tracker, and previously seen segments are updated with any new points that may have been found. To provide an estimate on the persistence of each segment, segments that are seen are scored. Initially a segment receives a score of 1, if this same segment is seen in the subsequent frame its score is increased by 1 and so on. When a segment is not seen its score is decreased by 0.5. If a segments score falls below 0 it is removed and no longer tracked. Once segments are identified and added to the tracker, a bounding box is used to identify each segment. These segments are numbered consecutively to

provide a sense of when they were first viewed, and their size is estimated based on the ellipse fit. Each segment's score is also displayed in the bounding box caption. In addition, by resizing the incoming images from the camera to 25% of their original size, this process can be completed at speeds up to 25 Hz without any noticeable detriment to performance on a consumer grade laptop using an Intel Core I5 Processor.

III. Methodology

Methodology is divided into three sections: First the lumen center identification and haustral fold tracking system is tested using the onboard camera of the REP-2 in both *in vivo* images and simulator images taken from the MESA. Next, REP-2 localization using ORB-SLAM2 is tested. Finally, the vision system is used to autonomously navigate the REP-2 through the MESA simulator.

A. Center Estimate Evaluation and Haustral Fold Tracking

To test the vision system's ability to detect and track both the lumen center and haustral folds, the REP-2's camera was used within the colon simulator to record video at 10 fps for analysis. This video was broken into 1000 frames (640x480 resolution) and each frame was passed into the vision system. In addition, video from actual colonoscopy shot at 30 fps (480 x320 resolution) was broken into 2758 frames and analyzed using the identification and tracking algorithm. The output of the algorithm including detected haustral folds, tracked haustral folds (tracked for > 3 frames), false positives and tracked false positives (tracked for > 3 frames) were counted and compared to ground truth (manually counted haustral folds from the same video) to determine the number of misses, good hits, and false hits over the duration of the video.

In addition, the center tracking algorithm output was compared to the estimated center as determined by four both inexperienced users and three gastroenterologists. A simple game was designed to allow a user to select the lumen center in 148 images from both the MESA simulator and actual colonoscopy. Each image was displayed at random and users were instructed to estimate the center of the lumen (described as the point in the image they would immediately head towards to stay in the center of the enclosed space). This selected pixel position was recorded and compared to the estimated center as determined by the lumen center estimator. Users were allowed to see their selection but were given no prior information about what good center estimates might look like beyond the verbal explanation.

B. REP Localization Using ORB-SLAM2

To assess the ability of ORB-SLAM2 to provide useful and accurate pose information, the camera of the REP was first calibrated in MATLAB. The resulting camera parameters were setup in ORB-SLAM2 and the raw camera images from the REP were then published via a Robotic Operating System (ROS) node and subscribed to via ORB-SLAM2. The REP camera was then manually progressed into the MESA simulator while its global position (distance of progression down the tube) was manually verified and recorded. The simulated tissue was placed in multiple geometric

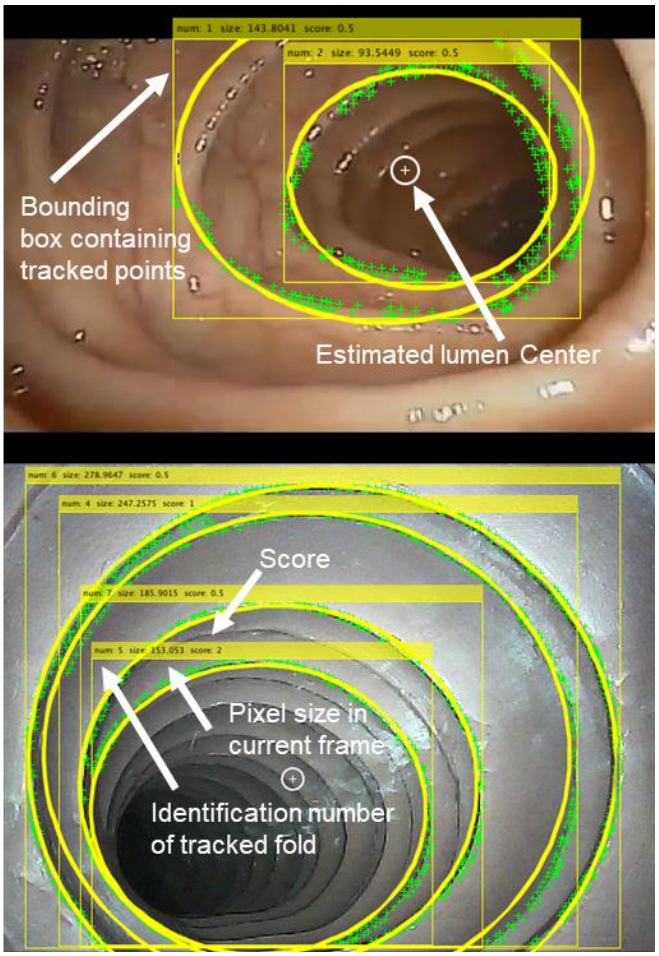

Figure 4. Multi-Haustral Fold tracking is shown in both an actual colon (top) and the simulated colon used with the MESA simulator (bottom). Each bounding box shows the estimated size of the haustral fold in pixels, the ID number of the fold, and a score to indicate how prevalent each fold has been.

configurations to allow for a visual comparison and confirmation of the resulting point cloud and trajectory from ORB-SLAM2. Following this initial test, the REP was manually driven down the full length and back to the start to visually compare the resulting pose plot to the actual MESA geometry.

C. Autonomous Navigation of MESA Simulator

To assess the feasibility of using this algorithm for navigation purposes, the vision system was used to autonomously navigate the REP-2 through the synthetic colon. A dual-mode control strategy similar to that used in [2] was applied to the lumen center estimates from the vision system and both video and run-time were recorded. In addition, any necessary manual intervention was recorded. A total of three trials were conducted to demonstrate the feasibility of this navigation method.

To navigate the MESA, center estimates from each image frame are calculated using the vision algorithm presented previously in [2]. A pixel error E is calculated by taking the difference between the true center pixel position F_{center} and the X pixel position of the center estimate Est_{Center}. The

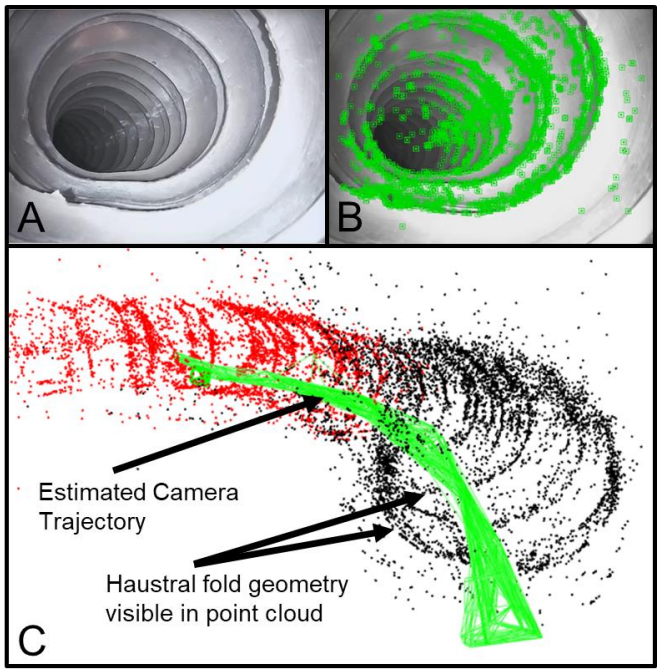

Figure 5. ORB-SLAM2 computes both camera/REP-2 pose and world points up to a scale. An input monocular image from the simulator is shown (A). Orb features are detected (B) and over many frames the estimated trajectory and point cloud is constructed (C). Note the clearly visible haustral folds present in (C).

pixel error in the x-direction is sent to the REP-2 for use in the control loop.

$$E = \begin{cases} F_{center} - Est_{Center} \\ E_{t-1} \end{cases} \quad (1)$$

The dual-mode control strategy uses the lumen center feedback for accurate REP navigation. This controller consists of a proportional, integral, derivative (PID) control mode for errors greater than ±30 pixels, and allows the REP-2 to use tank steering (one wheel forward, one driving in reverse) to correct for center tracking errors. As shown in equations (2) and (3), wheel speeds ω_1 and ω_2 (which are proportional to the PWM motor inputs) are adjusted based on error E and PID controller D_{PID}.

$$\omega_1 = \begin{cases} E \cdot D_{PID}, & |E| > 30, \ E > 0 \\ -E \cdot D_{PID}, & |E| > 30, \ E < 0 \end{cases} \quad (2)$$

$$\omega_2 = \begin{cases} -E \cdot D_{PID}, & |E| > 30, \ E > 0 \\ E \cdot D_{PID}, & |E| > 30, \ E < 0 \end{cases} \quad (3)$$

For errors less than ±30 pixels, simple proportional control, K_p, is used to adjust the speed of one wheel while the other is driven at a set max speed.

$$\omega_1 = \begin{cases} \omega_{max}, & |E| < 30, \ E > 0 \\ \omega_{max} - E \cdot K_p, & |E| < 30, \ E < 0 \end{cases} \quad (4)$$

$$\omega_2 = \begin{cases} \omega_{max} - E \cdot K_p, & |E| < 30, \ E > 0 \\ \omega_{max}, & |E| < 30, \ E < 0 \end{cases} \quad (5)$$

With this method for controlling the REP-2, the device was allowed to progress down the insufflated MESA simulator on its own while being timed by the operator who also ensured that the REP-2's tether was continuously fed into the simulator. The REP-2 was monitored for rollovers and tether snags which were corrected by the operator manually who could squeeze the insufflated deformable simulator externally to reach and adjust the REP-2 if necessary. Any interventions like these were recorded but the REP-2 was allowed to continue to the end of the simulator to complete the trial.

IV. Results & Discussion

By manually comparing the identified haustral folds to those of the detection/tracking algorithm, it was found that this algorithm detected and tracked 95.6% of haustral folds in the simulator images, with a false positive rate of 0.34% (34 false IDs in 1000 frames). While haustral folds were rarely missed, the reliance on image segmentation often results in a single haustral fold being split and tracked as separate pieces. This is particularly true as the camera comes closer to the haustral folds (splitting the folds in the image frame). Despite this, the algorithm was very successful at identifying and tracking the vast majority of haustral folds over many image frames (Table 1).

TABLE I. Haustral Fold Identification & Tracking

	Total	Identified	Tracked	FP % [a]	TFP % [a]
Simulator	23	22	22	0.34	0.08
In vivo	51	36	14	0.10	0.03

a. False positive rate and tracked false positive rate are shown as a percentage of total frames.

Haustral fold tracking in the colonoscopy video proved more difficult likely due to the irregular shape of actual haustral folds and the less discrete edges present. Despite these problems, the algorithm was able to identify 70.6% of folds with a very low occurrence of false positives, 0.10% (27 false IDs in 2758 frames). As with the simulated folds, double counting of folds was common, and improvements made to allow for better merging of these segments may prove useful in improving overall identification and tracking. Tracking did prove difficult *in vivo*, likely due to fast movements and visual

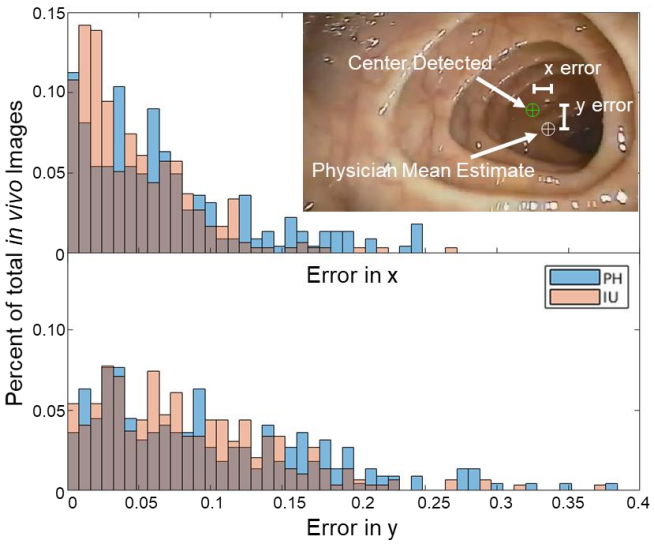

Figure 6. Error scaled as a percent of image size (in x and y) is shown for the center detection algorithm in comparison to both physicians (blue) and inexperienced users (red) for the *in vivo* colon images.

occlusions less present in the simulated frames and only 27.5% of folds were tracked across three or more image frames (Table I).

While haustral fold ID and tracking were somewhat hindered in colonoscopy images, center identification proved to be successful in both simulated and real *in vivo* environments. Pixel error was calculated as both a magnitude distance and a horizontal and vertical distance in X and Y. Error was then calculated as a percentage of the total image width/height in pixels to account for different image sizes. Estimated centers between all users differed widely, however the overall mean center estimate error of the algorithm across the 74 *in vivo* images was 6.6 % when compared to the three physicians (PH) and 4.6% when compared with the inexperienced users (IU) in the x-direction. The simulator center estimates proved less accurate, with overall mean errors in X of 15% when compared to the physicians and less than 9.4% when compared to the inexperienced users (Table II and Fig. 6). In general, the large difference in scale and depth between simulator and *in vivo* images may have resulted in this discrepancy as across the board, physicians and inexperienced users tended to choose points further down the lumen then the algorithm selected.

TABLE II. MEAN PIXEL ERRORS IN LUMEN CENTER DETECTION

	In vivo % Error	*In vivo* x[a] % Error	Simulator % Error	Simulator x[a] % Error
PH #1	6.2 ± 3.8	4.4 ± 3.6	11.8 ± 4.6	9.4 ± 5.2
PH #2	7.6 ± 4.8	5.0 ± 4.2	16.3 ± 5.5	13.5 ± 6.6
PH #3	14.9 ± 6.1	10.5 ± 6.8	24.5 ± 5.1	22.2 ± 6.1
IU #1	7.6 ± 4.6	4.5 ± 4.0	11.6 ± 5.5	9.3 ± 6.1
IU #2	5.9 ± 3.0	3.6 ± 2.6	6.2 ± 4.9	5.4 ± 4.9
IU #3	7.2 ± 4.4	4.0 ± 3.4	8.7 ± 4.5	4.9 ± 4.6
IU #4	9.0 ± 5.3	6.4 ± 5.3	20.3 ± 7.1	17.9 ± 7.6

a. Errors in x are shown separately from overall magnitude as this was the direction being controlled in autonomous navigation.

Localization via ORB-SLAM2 proved to be very successful. Final visual comparison of the resulting pose map and point cloud demonstrated excellent geometric similarity to the MESA simulator as seen in Fig. 7. In addition, haustral fold locations were readily discerned from the resulting point cloud and the REP-2 location was easily determined from the map, allowing for easy manual verification of its location during operation.

Autonomous navigation of the REP-2 within the MESA simulator also proved successful. Center tracking resulted in fast movement throughout the deformable simulator with the fastest trial taking approximately 6 minutes. While minor manual intervention was required on two of the three runs to correct tether snagging on the final sharp curve of the simulator, in general no other assistance was necessary to enable REP progression to the end of the simulator.

V. CONCLUSIONS & FUTURE WORK

An updated REP was designed fabricated and tested within the MESA anatomical simulator. An improved lumen center tracking approach was tested for its ability to detect and track

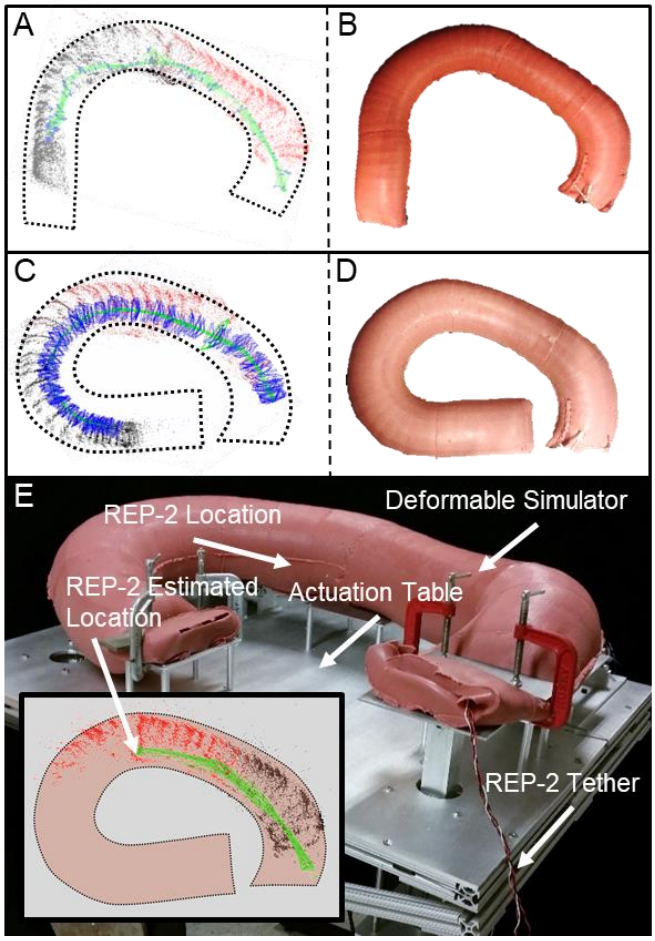

Figure 7. ORB-SLAM2 utilized in two different simulator configurations, with the resulting point clouds (A,C) compared against the geometry of the simulator as photographed from above (B,D). Dotted lines around the point clouds show this simulator geometry superimposed. A view of the MESA table with the estimated REP-2 location as predicted by ORB-SLAM2 is also shown (E). The configuration shown in (C,D) was also the configuration used to test autonomous navigation of the REP-2.

haustral folds and for its ability to determine the lumen center in both colonoscopy video and video from the MESA simulator. Lumen center tracking was then used as feedback to allow for autonomous navigation within the simulator. Finally, an open source monocular SLAM approach (ORB-SLAM2) was applied within the MESA simulator to allow for REP localization and sparse mapping of the simulator environment. Lumen tracking and localization using ORB-SLAM2 both proved effective, allowing for autonomous navigation and localization of this device in the MESA simulator. In addition, the haustral fold detection and tracking algorithm also showed significant potential when applied to actual *in vivo* colonoscopy images.

This preliminary investigation into visual strategies for navigation and localization of the REP, demonstrates the potential for robotic vision to improve the capabilities of robotic capsule endoscopes. While the device featured in this work is currently not a suitable size for *in vivo* applications, we believe these same techniques will prove useful on a scaled down version for animal testing currently under development.

Future work will focus on building on the tools presented here for improved haustral fold detection/tracking in actual colonoscopy video and to allow for monocular SLAM in the more difficult deformable *in vivo* environment. In addition, parallel work is currently focused on a device scaled back to the size of [17]. *In vivo* testing of this device will be conducted to further evaluate the utility of these visual tools.

VI. Acknowledgment

This research was supported in part by the National Science Foundation (NSF) (Grant No. 1235532), the CU-Boulder Innovative Seed Grant Program, and the Colorado Office of Economic Development and International Trade. This research was also conducted with Government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a.

VII. References

[1] Given Imaging Ltd. [Online]. Available: http://www.givenimaging.com/

[2] 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. 2670-77, July 2018. DOI: 10.1109/LRA.2018.2827168.

[3] G.A. Formosa, J.M. Prendergast, M.E. Rentschler, "A Modular Endoscopy Simulation Apparatus (MESA) for Robotic Medical Device Sensing and Control Validation," *IEEE Robotics and Automation Letters*, accepted for publication.

[4] R. L. Siegel, *et al.*, "Cancer statistics, 2016: Cancer Statistics, 2016," CA: A Cancer Journal for Clinicians, vol. 66, no. 1, pp. 7–30, Jan. 2016.

[5] American Cancer Society, "Colorectal cancer facts and figures 2017-2019," Atlanta: American Cancer Society, 2017.

[6] S. G. Shah, *et al.*, Patient pain during colonoscopy: an analysis using realtime magnetic endoscope imaging," *Endoscopy*, vol. 34, no. 6, pp. 435–40, 2002.

[7] P. Valdastri, *et al.*, "Advanced technologies for gastrointestinal endoscopy," *Annu Rev Biomed Eng*, vol. 14, pp. 397–429, 2012.

[8] P. Gao, *et al.*, "Microgroove cushion of robotic endoscope for active locomotion in the gastrointestinal tract: A smart device for the robotic endoscope," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 8, no. 4, pp. 398–406, Dec. 2012.

[9] P. Glass, *et al.*, "A legged anchoring mechanism for capsule endoscopes using micropatterned adhesives," *IEEE Trans Biomed Eng*, vol. 55, no. 12, pp. 2759–67, Dec 2008.

[10] E. Buselli, *et al.*, "Evaluation of friction enhancement through soft polymer micro-patterns in active capsule endoscopy," *Meas. Sci. Technol.*, vol. 21, no. 10, p. 105802, Sep 2010.

[11] D. Lee, *et al.*, "A simple and reliable reel mechanism-based robotic colonoscope for high mobility," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, p. 095440621772394, Jul. 2017.

[12] D. Glozman, *et al.*, "A Self-Propelled Inflatable Earthworm-Like Endoscope Actuated by a Single Supply Line," *IEEE Trans Biomed Eng*, vol. 57, no. 6, pp. 1264–72, Dec 2010.

[13] C. C. Y. Poon, *et al.*, "Design of wormlike automated robotic endoscope: dynamic interaction between endoscopic balloon and surrounding tissues," *Surgical Endoscopy*, vol. 30, no. 2, pp. 772–778, Feb. 2016.

[14] J. E. Bernth, *et al.*, "A Novel Robotic Meshworm With Segment-Bending Anchoring for Colonoscopy," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1718–1724, Jul. 2017.

[15] L. Sliker, *et al.*, "Surgical evaluation of a novel tethered robotic capsule endoscope using micro-patterned treads," *Surg Endosc*, vol. 26, no. 1, pp. 2862–9, Oct 2012.

[16] D. Lee, *et al.*, "An elastic caterpillar-based self-propelled robotic colonoscope with high safety and mobility," *Mechatronics*, vol. 39, pp. 54–62, Nov. 2016.

[17] P. Valdastri, *et al.*, "Magnetic air capsule robotic system: proof of concept of a novel approach for painless colonoscopy," *Surg. Endosc.*, vol. 26, no. 5, pp. 1238–46, May 2012.

[18] A. W. Mahoney and J. J. Abbott, "Generating Rotating Magnetic Fields With a Single Permanent Magnet for Propulsion of Untethered Magnetic Devices in a Lumen," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 411–420, Apr. 2014.

[19] L. Sliker, M. E. Rentschler, "The Design and Characterization of a Testing Platform for Quantitative Evaluation of Tread Performance on Multiple Biological Substrates," *IEEE Trans Biomed Eng*, 59(9): 2524-2530, 2012.

[20] L. Sliker, *et al.*, "An Automated Traction Measurement Platform and Empirical Model for Evaluation of Rolling Micro-Patterned Wheels," *IEEE/ASME Transactions on Mechatronics*. 20(4): 1854-1862, 2015.

[21] M. Li, S. Song, *et al.*, "A novel method of 6-DoF electromagnetic navigation system for surgical robot," in *Intelligent Control and Automation (WCICA), 2010 8th World Congress on*, 2010, pp. 2163–2167.

[22] C. Di Natali, *et al.*, "Real-time pose detection for magnetic medical devices," *Magnetics, IEEE Transactions on*, vol. 49, no. 7, pp. 3524–3527, 2013.

[23] D. Scaradozzi, *et al.*, "Simultaneous localization and mapping (SLAM) robotics techniques: a possible application in surgery," Shanghai Chest, vol. 2, pp. 5–5, Jan. 2018.

[24] C. Di Natali, *et al.*, "Real-time pose detection for magnetic medical devices," *Magnetics, IEEE Transactions on*, vol. 49, no. 7, pp. 3524–3527, 2013.

[25] J. Liu, *et al.*, "An optical flow approach to tracking colonoscopy video," *Comput. Med. Imaging Graph.*, vol. 37, no. 3, pp. 207–223, Apr. 2013.

[26] N. Van Der Stap, *et al.*, "A feasibility study of optical flow-based navigation during colonoscopy," 2012.

[27] N. Masson, *et al.*, "In vivo comparison of real-time tracking algorithms for interventional flexible endoscopy," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, 2009, pp. 1350–1353.

[28] G. Ciuti, *et al.*, "Intra-operative monocular 3D reconstruction for image-guided navigation in active locomotion capsule endoscopy," in *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on*, 2012, pp. 768–774.

[29] G. M. L. Bao, *et al.*, "A Computer Vision based Speed Estimation Technique for Localizing the Wireless Capsule Endoscope inside Small Intestine," in *IEEE International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.

[30] M. Turan, *et al.*, "Sparse-then-dense alignment-based 3D map reconstruction method for endoscopic capsule robots," Machine Vision and Applications, vol. 29, no. 2, pp. 345–359, Feb. 2018.

[31] S. Mills, *et al.*, "Hierarchical Structure from Motion from Endoscopic Video," 2014, pp. 102–107.

[32] M. A. Armin, *et al.*, "Automated visibility map of the internal colon surface from colonoscopy video," International Journal of Computer Assisted Radiology and Surgery, vol. 11, no. 9, pp. 1599–1610, Sep. 2016.

[33] A. Koulaouzidis et al., "Novel experimental and software methods for image reconstruction and localization in capsule endoscopy," Endoscopy International Open, vol. 06, no. 02, pp. E205–E210, Feb. 2018.

[34] Y. Abu-Kheil, *et al.*, "Vision and inertial-based image mapping for capsule endoscopy," in Information and Communication Technology Research (ICTRC), 2015 International Conference on, 2015, pp. 84–87.

[35] F. Nobre, *et al.*, "Drift-correcting self-calibration for visual-inertial SLAM," in Robotics and Automation (ICRA), 2017 IEEE International Conference on, 2017, pp. 6525–6532.

[36] R. Mur-Artal, *et al.*, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[37] C. Tomasi. T. Kanade, "Detection and Tracking of Point Features," Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.