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

Comparing Visual Odometry Systems in Actively Deforming Simulated Colon Environments

Mitchell J. Fulton[1], J. Micah Prendergast[1], Emily R. DiTommaso[2], and Mark E. Rentschler[1]

Abstract— **This paper presents a new open-source dataset with ground truth position in a simulated colon environment to promote development of real-time feedback systems for physicians performing colonoscopies. Four systems (DSO, LSD-SLAM, SfMLearner, ORB-SLAM2) are tested on this dataset and their failures are analyzed. A data collection platform was fabricated and used to take the dataset in a colonoscopy training simulator that was affixed to a flat surface. The noise in the ground truth positional data induced from the metal in the data collection platform was then characterized and corrected. The Absolute Trajectory RMSE Error (ATE) and Relative Error (RE) metrics were performed on each of the sequences in the dataset for each of the Simultaneous Localization And Mapping (SLAM) systems. While these systems all had good performance in idealized conditions, more realistic conditions in the harder sequences caused them to produce poor results or fail completely. These failures will be a hindrance to physicians in a real-world scenario, so future systems made for this environment must be more robust to the difficulties found in the colon, even at the expense of trajectory accuracy. The authors believe that this is the first open-source dataset with groundtruth data displaying a simulated *in vivo* environment with active deformation, and that this is the first step toward achieving useful SLAM within the colon. The dataset is available at `www.colorado.edu/lab/amtl/datasets`.**

I. INTRODUCTION

Colonoscopy procedures have remained largely unchanged for the past several decades despite a steady rise in colon cancer diagnoses. Colorectal cancer (CRC) is now the second most deadly cancer with over fifty-three thousand people projected to die from colon cancer in 2020 [1]. However, it has a 90% five-year survival rate if caught in early stages [2]. Of these, approximately 4.3% of diagnoses result from interval colorectal cancer (I-CRC), or CRC that is diagnosed within five years of a negative colonoscopy [3]. This suggests a lack of quality in CRC screenings. The further development of algorithms that could improve real-time feedback to physicians performing colonoscopies could greatly increase the quality of colonoscopy procedures. However, algorithm development for this purpose is often hindered by a lack of applicable test datasets of colon-like environments. In this paper we present a dataset of videos taken in a simulated colon environment with translational ground truth to be used for localization and mapping in the colon.

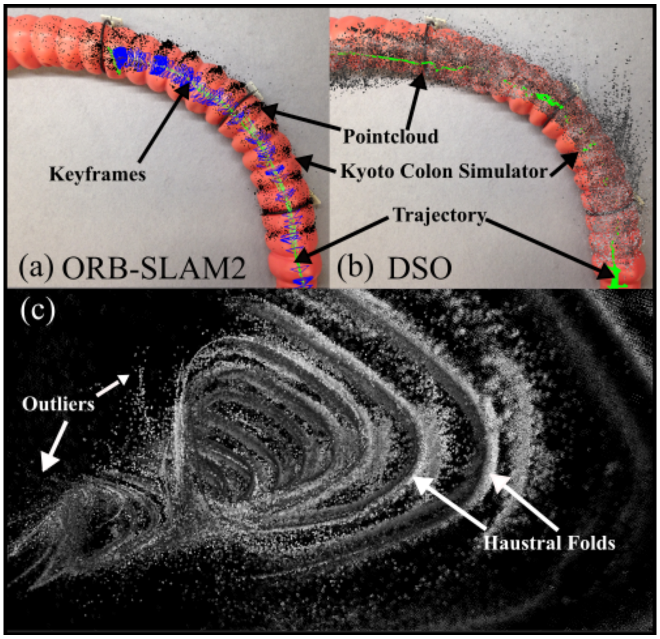


Fig. 1: A new dataset was taken in a Kyoto Kagaku colon simulator and the performance of current SLAM systems was evaluated. Shown are point clouds and trajectories for (a) ORB-SLAM2 and (b) DSO, with (c) showing a first-person view of the map created with DSO. Outliers and haustral folds, the structures within the colon, can be seen.

A. Motivation

The quality of a colonoscopy is most often measured by the adenoma detection rate (ADR) of a colonoscopist. This rate describes how frequently the medical professional finds an adenoma in patients over 50. While the ADR has been extensively studied and shown to have a 3% decrease in cancer rates for every 1% increase in ADR [4], this metric is imperfect. It is significantly affected by patient demographics and the type of screenings done, and does not consider the number of adenomas found in each patient, which could affect the efficacy of this measure [5]. The ADR also only yields feedback over a large number of patients

*This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1849357. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

Mark Rentschler is a co-founder of Aspero Medical, Inc., a University of Colorado spin-out company that is focused on commercializing balloon overtube products for use in enteroscopy.

[1] Mitchell J Fulton, J Micah Prendergast, and Mark Rentschler are with the Department of Mechanical Engineering, University of Colorado Boulder, Boulder, CO 80309, USA `mitchell.fulton@colorado.edu` `joseph.prendergast@colorado.edu` `mark.rentschler@colorado.edu`

[2] Emily R DiTommaso is with the Department of Electrical, Computer, and Energy Engineering, University of Colorado Boulder, Boulder, CO 80309, USA `emily.ditommaso@colorado.edu`

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

after the colonoscopies have taken place. Therefore, this does not provide any feedback in real-time to the physician performing the procedure. Enabling a physician to explore and visualize the endoscopes pose within a 3D tissue map in real-time could help to ensure accurate diagnoses with a more complete screening from a colonoscopy.

B. Related Works

To improve the quality of diagnosis and treatment of diseases within the colon, new technologies have been developed and tested within the gastrointestinal tract [6], [7], [8]. In all of these, a method of localization is needed for effective inspection, treatment, and control. However, because many of these algorithms require specific novel hardware and use additional sensors beyond the single camera found on a typical colonoscope, they are not immediately deployable to assist medical professionals. Though these would provide more robust estimations, we have omitted these algorighms as their use also depends on the adoption of new hardware to replace the standard colonoscope. To augment the current colonoscopy procedures for localization and mapping to assist in inspection and treatment, a system using solely monocular video input must be used. The problem of monocular Simultaneous Localization And Mapping (SLAM) has long been studied, and is often considered a solved problem for static, well-textured environments. While monocular SLAM systems have been developed to handle dynamic environments, few that can explicitly handle large deformation are able to run in real-time [9]. To achieve real-time performance, one group used an extended Kalman filter-based SLAM system, but because of the computational burden the number of points that can be tracked is extremely limited [10]. To create a map of the colon, several relatively successful offline attempts have been made, but only one group has been able to reconstruct sections of the colon successfully in real-time [11]. However, they have yet to handle any active deformation, occlusion, large camera motions, or longer video sequences as would be present in a colonoscopy procedure. Despite the existence of these systems for localization and visual odometry, to our knowledge they have not yet been tested in an actively deforming environment similar to the colon.

There have been several datasets made publicly available aimed toward aiding medical technology in the colon. However, none of these datasets focus on translation or mapping, but on polyp segmentation [12], [13], [14] or disease classification and tracking [12], [15], [16], [17], [18]. In addition, few of these datasets have video, and of these the videos are either unlabeled or have bounding boxes for disease tracking. A new video dataset with active deformation and ground truth data in an environment similar to the body would aid in the development of novel SLAM systems for use in the body.

II. METHODS

In this paper we investigate the performance of four popular real-time SLAM and Visual Odometry (VO) systems in a simulated colon environment, both with and without active deformation.

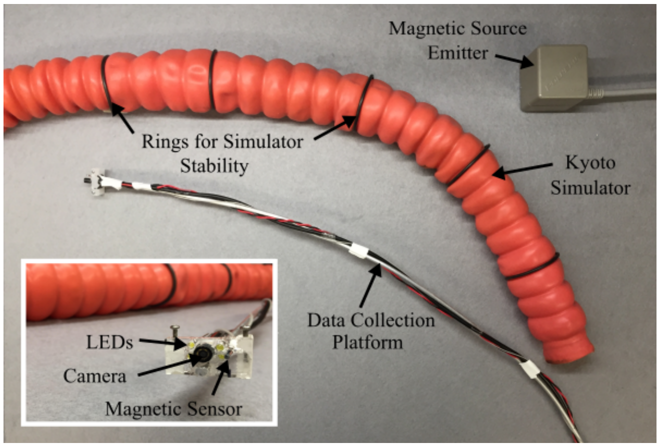


Fig. 2: To take data in the colon simulator a custom data collection platform was designed and fabricated. The camera was then progressed within the colon simulator while a magnetic tracker took ground truth measurements. Both the colon simulator and the magnetic source for ground truth were affixed to a flat backing to ensure consistently accurate measurements.

A. Chosen SLAM and VO Systems

Four popular open-source SLAM or VO systems were chosen for comparison in the colon simulators: ORB-SLAM2 [19], DSO [20], LSD-SLAM [21], and SfMLearner [22]. Each of these systems represent a different approach to the SLAM problem. ORB-SLAM2 is a sparse indirect method. It first finds unique features in the scene and matches them between frames. Based on these feature correspondences, the SLAM problem can then be solved. DSO is a direct and sparse method, where the points chosen are sampled throughout the frames directly from all available pixels, rather than a chosen type of feature. These points are then jointly optimized to solve the SLAM problem. LSD-SLAM is a direct dense method, which performs direct image alignment over all of the pixels with sufficient gradient. These are then all used in a probabilistic model to eliminate noise and solve the SLAM problem. Finally SfMLearner, an unsupervised learning based method with a loss function similar to LSD-SLAM, performs VO directly over the whole image while learning complex features as it trains. A supervised learning method was not chosen due to the difficulty of acquiring adequate ground truth data in an *in vivo* environment, though further research in this area could prove to be valuable.

B. Data Collection

The dataset was collected in the environment of the Kyoto Kagaku Colonoscope Training Model (Kyoto Kagaku Co. Ltd, Kyoto, Japan) using a small rolling shutter endoscope camera (Kyzee 5.5 mm Wireless Endoscope, Wuzhou Jin Zhengyuan Technology Co. Wuzhou, China) to mimic a real colonoscopic procedure. To be able to effectively traverse the simulated colon, a data collection platform was created by mounting the camera onto an aluminum rod with multiple LEDs for lighting (Fig 2). The simulated colon was insufflated to allow clearer views of the simulated tissue and enable smoother motion of the camera, and was mounted by Velcro straps to a felt board. Before taking data, the cameras were calibrated using Kalibr [24]. The camera was then progressed through the colon while recording frames at approximately 30 Hz. For the sequences where deformation was included (Sequences 3, 4, 6, and 7), force was manually applied externally to the simulator throughout the video sequence. The level of deformation was not quantitatively measured, as its prominence in the frame is dependent not only on the magnitude of the deformation, but also the distance from the camera and configuration of the colon. Deformation also only affects the SLAM systems based on what areas that are in view are currently being tracked, which has many unpredictable internal algorithmic influences, including recent camera motion and a random initialization.

TABLE I: Sequence Characteristics

Seq.	Straightened	Deformation	Difficulty	Trajectory Length[m]
1	Yes	No	Easy	1.9254
2	Yes	No	Easy	1.5543
3	Yes	Yes	Medium	2.2933
4	Yes	Yes	Medium	1.8194
5	No	No	Hard	5.6494
6	No	Yes	Hard	2.5606
7	No	Yes	Hard	3.8232

To provide ground truth data, we used a Polhemus Patriot HS magnetic tracker (Polhemus Inc., Colchester, VT, USA). We selected magnetic tracking as we required a system that would not need line of sight to measure position. Other measurement systems that could achieve this (radio frequency tracking, GPS) do not have fine enough resolution for this application, have too much attenuation from the body, or cannot be used due to the size constraints of this environment. In addition, due to size constraints a depth camera or LIDAR system was not able to be used for a ground truth depth measurement, limiting the analysis to position and orientation only.

To effectively synchronize all of the data, it was published to and recorded in ROS [25]. Using ROS allowed us to inform the camera movement by running a SLAM system in real time during the data acquisition process. We chose to do this in Sequences 1-4 using ORB-SLAM2 for this purpose as it demonstrated the slowest initialization and easiest loss of tracking. The full list of sequence characteristics is shown in Table I.

C. System Parameters and Training

Parameter tuning was performed on each of the systems and focused on improving stability and robustness as well as a good initialization. However, in each of the geometric SLAM systems it was found that this parameter tuning yielded little to no improvement over the default values, so nearly all default values were kept. For ORB-SLAM2 the number of features extracted was increased to 2000, with all other values kept the same. For LSD-SLAM all parameters were kept at default values. DSO was run with camera intrinsics calibration only and using maximum accuracy settings at the expense of some speed.

To train the SfMLearner, training data was collected from two colon simulators: the MESA simulator [23] and the Kyoto Kagaku Colonoscope Training Model (Kyoto Kagaku Co. Ltd, Kyoto, Japan). Videos were taken with two different cameras: a small rolling shutter endoscope camera (Kyzee 5.5 mm Wireless Endoscope, Wuzhou Jin Zhengyuan Technology Co. Wuzhou, China) and a larger, high definition global shutter camera (Kayeton Global Shutter High Speed 120 fps HD 720P Webcam, Shenzhen Kayeton, Shenzhen, China) which was restricted to use in the MESA due to its size. The focal length was changed and the cameras were calibrated in Kalibr [24] before taking data. The data collection platform shown in Fig 2 and a similar counterpart for the larger camera was used to traverse each colon. The simulated colons were insufflated to allow clearer views of the simulated tissue and enable smoother motion of the camera, and were mounted by Velcro straps to a felt board. In each video taken the simulator was in a different path configuration to help vary the data and alleviate overfitting of the learning-based method. It was also ensured that the Kyoto Kagaku simulator was not in a similar configuration to the ground truth sequences used as the test data. Initially, the system was coarsely tuned with the KITTI dataset [27] and fine tuned with colon images, but this was found to be inferior to only using the colon simulator training data. Using only the colon training data, the best results were achieved with using a learning rate of 0.0002 with $\beta_1 = 0.9$ and $\beta_2 = 0.99$ a mask parameter of 0.2, a smoothness parameter of 0.1, and a batch size of 3. Convergence was achieved in approximately 42,000 iterations.

D. Noise Characterization

To ensure accurate ground truth readings from the magnetic tracker, a noise characterization was performed for the data collection platform. Because the magnetic tracker operates by measuring a constantly emitted magnetic field, interference arises from both ferrous and non-ferrous metals. Metals with paramagnetic effects, including aluminum and stainless steel, can induce magnetic interference while within a magnetic field. Because the data collection platform incorporates aluminum and stainless steel, it was necessary to characterize the impact of these elements on sensor accuracy prior to relying on this system for ground truth.

1) Noise Data Collection: A custom noise characterization test fixture was designed and built to collect noise data for the magnetic tracking system. Generally speaking, data taken within each simulator was done within a relatively constant xy plane, as each simulator was fixed to a flat surface. For this reason, noise characterization was limited to xy position changes. Shown in Fig. 3, the test fixture has a laser etched ground truth grid with the magnetic field emitter

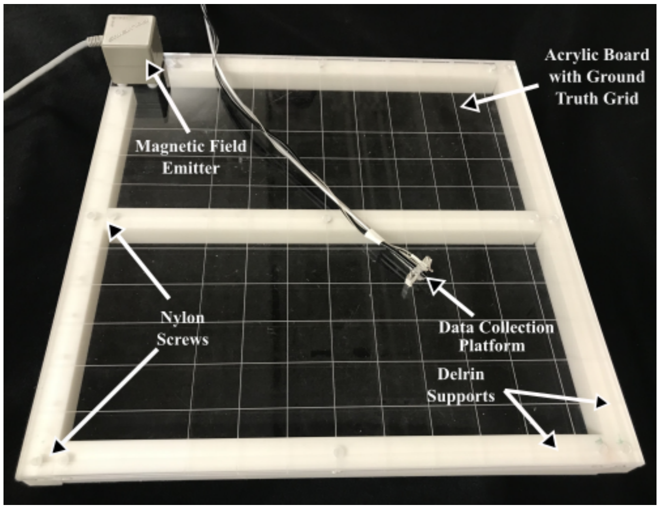


Fig. 3: The test fixture to collect positional and orientational noise data consists of a flat acrylic board with a ground truth positioning grid, Delrin supports, and nylon screws. The data collected was used to characterize the noise arising from the ferro- and paramagnetic components of the data collection platform.

fixed in one corner. The emitter's position can be adjusted to calibrate its magnetic center with the origin of the grid on the board. The board is made of an acrylic sheet, Delrin supports, and nylon screws so that no magnetic interference is added. It was then placed on a plastic cart and moved away from any other metal or magnetic sources.

To collect data, the position of the magnetic source on the board was calibrated with only the sensor in the field. The sensor was then inserted into the data collection platform and thirty seconds of measurements were taken at approximately 60 Hz at each grid point on the board. This was repeated four times with the sensor facing the 0, $\frac{\pi}{2}$, π, and $\frac{3\pi}{2}$ angles in the xy plane, and the data was stored for analysis.

2) Noise Data Analysis: It was observed that the metal of the data collection platform introduced a bias in the positional readings significantly depending on the proximity of the sensor and source to the data collection platform. Because the data collection platform is rigid, the proximity to the sensor and source can be described by two terms, the radial distance r and the coincidence angle of the sensor θ with respect to the source. The radial distance used was a simple euclidean distance, while the coincidence angle was found by

$$\theta_c = \arctan_2(y_{pos}, x_{pos}) - \psi. \tag{1}$$

This assumes planar motion, considering only x and y positions (y_{pos}, x_{pos}) and the rotation around the z axis (ψ). However, it was found that when the colon simulator was on a flat surface adding the remaining rotational had a negligible impact for this data.

Using the reduced dimensionality, a curve was fit to the noise data to correct any induced biases in the experimental data. The equation of the form

$$x_i = a_0 r_i \sin(a_1\theta_c + a_2) + a_3 \tag{2}$$

was used, where a_i is a coefficient to be tuned for the cartesian dimension x_i. Using this equation, a nonlinear least squares regression with an arctangent loss function was performed to find the optimal parameter values. With these values, the positional data was corrected to remove the induced bias from the magnetic interference of the data collection platform. The worst case bias was reduced from over ± 3.5 cm to under ± 0.5 cm.

III. RESULTS AND DISCUSSION

After the simulator data was collected and adjusted for noise, an analysis was performed using the toolbox from [26] to find the absolute and relative errors of the final trajectories. For the SfMLearner, because each pose transformation is an independent estimation, each measurement was scaled individually and then connected to form a single continuous trajectory. All of the systems' trajectories were then transformed into a common format of [timestamp, translation, rotation quaternion], and the analysis was performed.

TABLE II: Translational Absolute Trajectory RMSE Error $[cm]$

	DSO	LSD-SLAM	ORB-SLAM2	SfMLearner
Seq. 1	0.779	4.699	1.594	1.821
Seq. 2	0.722	3.740	0.794	1.149
Seq. 3	9.470	10.038	1.363	1.664
Seq. 4	0.590	8.620	0.632	1.585
Seq. 5	x	x	x	16.172
Seq. 6	x	x	x	15.137
Seq. 7	x	x	x	11.893

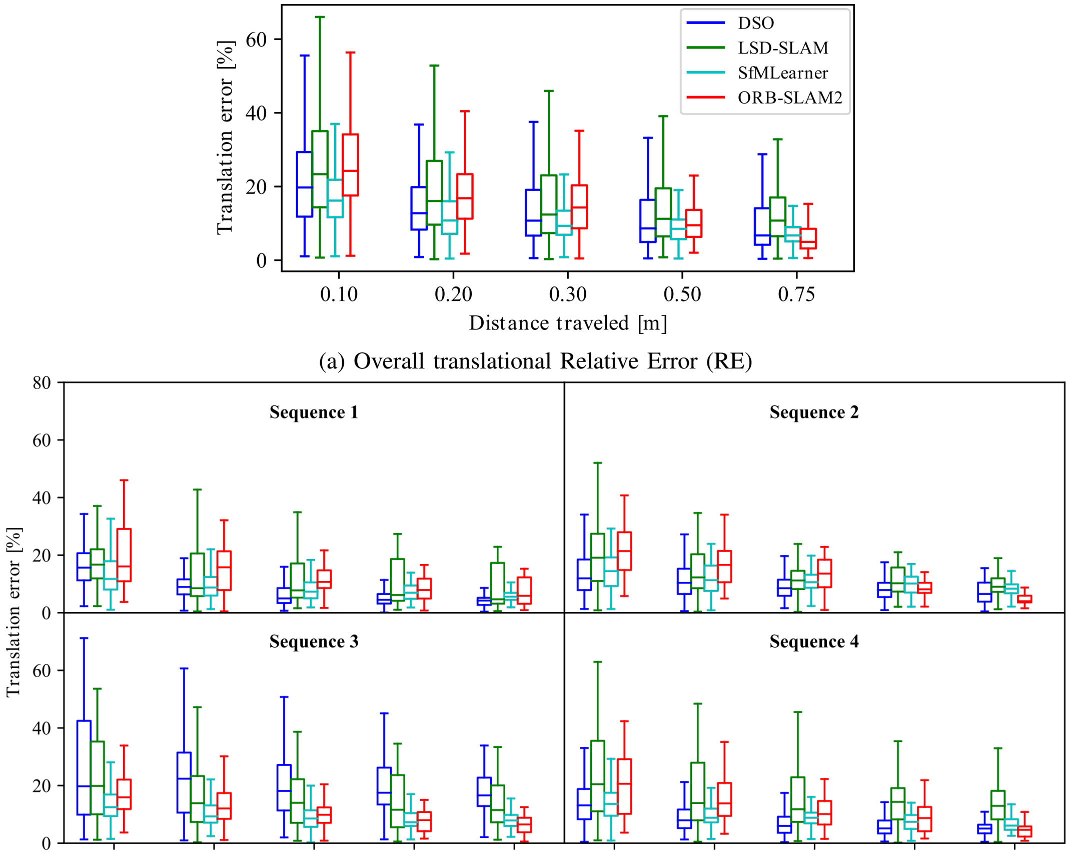
A. Translation

Traditionally trajectory estimations have been measured using only translational error. This is most often measured using Absolute Trajectory RMSE Error (ATE), defined by

$$ATE_{pos} = (\frac{1}{N}\sum_{i=0}^{N-1} ||\Delta\mathbf{p}_i||^2)^{\frac{1}{2}} \tag{3}$$

where $\Delta\mathbf{p_i}$ is the error in the positional change at step i after the starting points of the ground truth and estimate have been aligned. Because the RMSE is calculated for the entire path, this metric is a measure of the overall, or global, accuracy of the estimated trajectory. Table II shows the ATE in centimeters for the seven test datasets taken in the Kyoto Kagaku simulator, with an 'x' representing system failure on the dataset.

However, the ATE is sensitive to the time at which errors occur [27] so this analysis focuses instead on the Relative Error (RE). The RE is also calculated using Eq. 3, except it is only measured over shorter subsections of the overall trajectory. One benefit of using RE to assess trajectory accuracy is that the error is calculated many times



(a) Overall translational Relative Error (RE)

(b) Per-sequence translational Relative Error (RE)

Fig. 4: These overall (a) and per-sequence (b) plots show the translational Relative Error (RE) as a percentage of tested sequence length over five different sequence lengths for each of the four tested systems. The RE for the learning based method is consistently low due to scaling each step in the sequence before the global optimization. Because the scaling factor is optimized over the entire trajectory and applied directly only to the translational components of the trajectory, shorter sequences have a higher relative error. Sequences 5-7 are not shown due to a lack of initialization across all systems.

for each chosen subsection length, and other values (standard deviation, mean, etc.) can be derived from this collection of errors. In addition, different sub-trajectory lengths can be used to evaluate accuracy over different scales throughout all tested video sequences.

As can be seen in Fig. 4 there is a very high potential for translational error over short sequences in all of the tested systems, but this diminishes for longer sections up to 0.75 m, nearly half the longest trajectory length. Because the scale and alignment was optimized over the global trajectory and is applied directly to the translation, longer sub-sequences perform better as more of the optimized data are present.

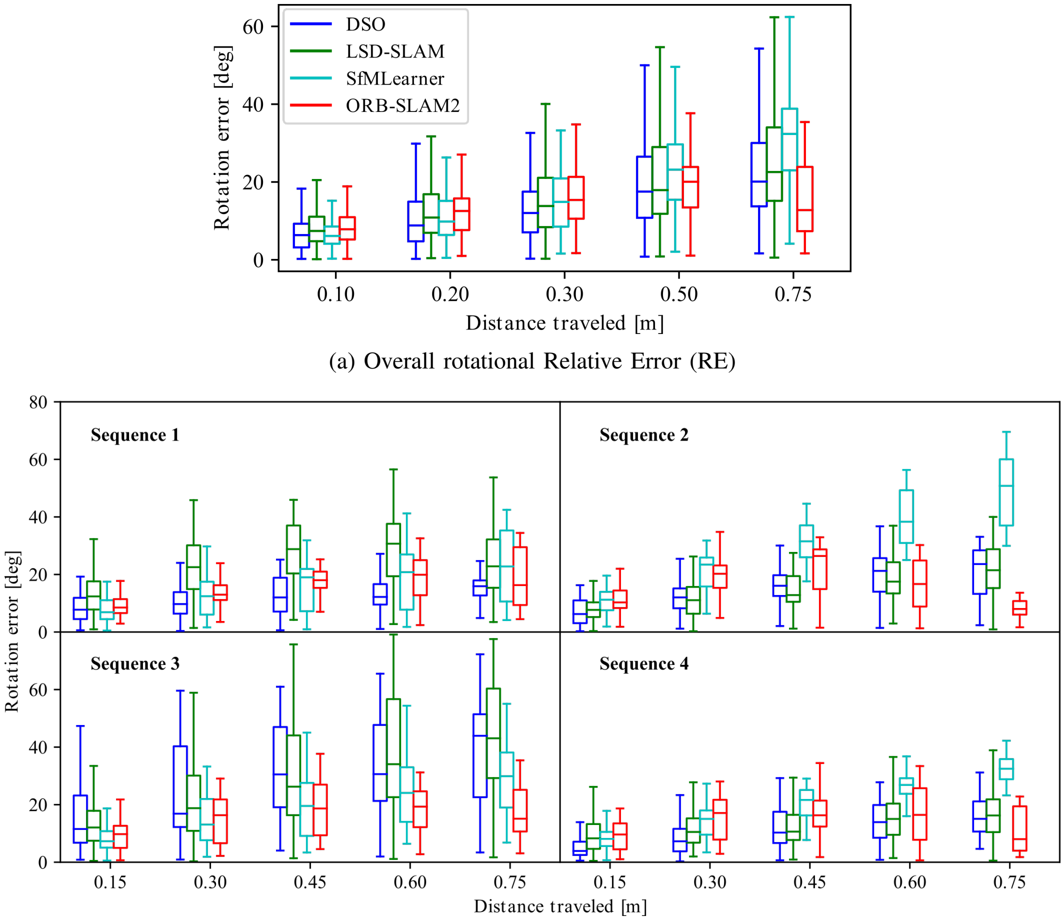
B. Rotation

The rotational error can be measured in a similar way to the translational error of (3). However, $\mathbf{p}_i$ is replaced by $\Delta\mathbf{R}_i$, the difference in rotation matrices between ground truth and the estimation. The axis angle representation of this resulting matrix is used as the error [26]. Substituting this into (3), the rotational ATE becomes

$$ATE_{rot} = (\frac{1}{N}\sum_{i=0}^{N-1} ||\Delta\angle(\mathbf{R}_i)||^2)^{\frac{1}{2}}. \tag{4}$$

Like the positional ATE, the rotational ATE is subject to a bias from the time of the error. Therefore we exclusively used the RE, or the ATE over shorter sequences, for our rotational accuracy as we felt the rotational ATE did not accurately convey the global error. While a noise characterization was not done for the rotational degrees of freedom, it was observed that the worst case biases spanned less than ten degrees.

Unlike the positional RE which tends to shrink over longer sub-sequences, the rotational RE tends to grow with sub-sequence length. Since the rotation between frames is often much smaller in magnitude than the translation, it is easy for it to drift over time. In addition, the trajectory is scaled and aligned via the estimated and ground truth poses, but the



(a) Overall rotational Relative Error (RE)

(b) Per-sequence rotational Relative Error (RE)

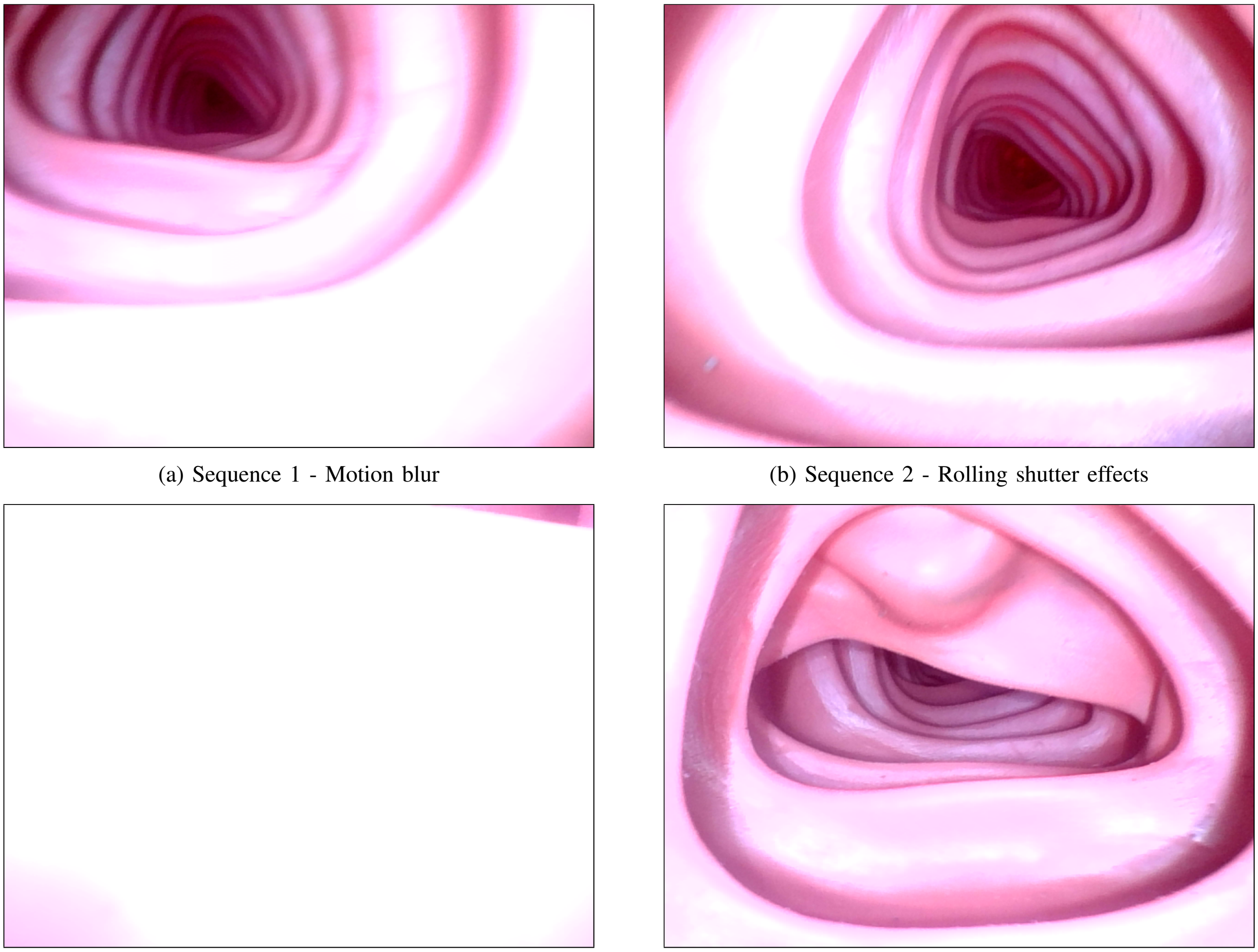
Fig. 5: These overall (a) and per-sequence (b) rotational RE plots show all four sequences' rotational accuracy as a function of sequence length. The RE for the rotation grows over time, as the globally optimized scale is not applied to the rotational component of the trajectory. It can be seen for ORB-SLAM2 that in longer sections, as trajectory optimization is performed, the error goes down. For the learning-based method the rotational error grows more than the other methods as each step is independent and not optimized together. Sequences 5-7 are not shown due to a lack of initialization across all systems.

rotation matrix is not scaled. Therefore small errors build up over time, as the rotation matrices do not have the alignment and scaling optimization applied directly to them. This can be seen in Fig. 5, as the growing sub-trajectory length increases the rotational error. It can also be seen that the SfMLearner accumulates more rotational error than the other SLAM systems. This is due to SfMLearner independently scaling each estimate, yielding more rotational drift over time. The other systems have trajectory optimization throughout the sequence, causing the rotational drift to be greatly reduced. As can be seen for ORB-SLAM2 in Fig. 5 specifically, optimization occurs after a longer distance is traveled.

C. Failures

In addition to the accuracy of the systems, the failures must also be considered. Generally the failures for each system fell into one of three categories: a failure to initialize, a poor estimation, or a loss of tracking. All three categories of failure were caused by difficulties within the environment. The effects of these failures can be seen in Figs. 4,5 with high upper extremes when the estimates were very poor or in Table II when the systems were not able to complete the trajectory. Even with a good view of many colon structures, the environment has repetitive and sparse texture, making features hard to track and accurately match. This view is often complicated further by motion blur, rolling shutter effects, occlusion, deformation, and spectral highlights (or areas where reflections cause false feature extraction and matching) as shown in Fig. 6.

1) DSO: DSO was the most robust geometric SLAM system with regard to continuous tracking, but had the least accurate initialization. While DSO would always initialize quickly, the solution it reached was usually incorrect. Therefore, every time it was run the initialization had to be manually checked and restarted if it was incorrect. This failure in initialization often caused problems later in the sequence and greatly affected the accuracy of the estimated trajectory. This dependence on the initialization caused the



(a) Sequence 1 - Motion blur

(b) Sequence 2 - Rolling shutter effects

(c) Sequence 3 - Occlusion

(d) Sequence 4 - Deformation and spectral highlights

Fig. 6: The most common causes of failure in Sequences 1-4 are motion blur (a), rolling shutter effects (b), occlusion (c), and deformation and spectral highlights (d), respectively. Each of these failure modes are present in nearly all sequences but are most prevalent in the sequences shown. Sequences 5-7 are omitted as the most common failure in each is a lack of initialization.

system's poor estimation failures to vary widely even when run multiple times on the same sequence. Even though the trajectory estimate could be inaccurate, DSO was the best system when it came to avoiding a loss of tracking. It does not have a tracking failure recovery system and was the only geometric SLAM system to not lose tracking on Sequences 1-4. However, if it did lose tracking, it was not able to recover and thus could not complete Sequences 5-7.

2) LSD-SLAM: LSD-SLAM was the least accurate of all the systems but was moderately robust against all failures. Though the initialization method is shared between LSD-SLAM and DSO and a similar manual checking process was used, LSD-SLAM's initialization more often converged to a correct value. Despite this correct initialization, the reprojections were often very noisy, causing poor trajectory estimates. Poorer trajectory estimates and losses in tracking were then caused by difficulties in the environment (Fig. 6). LSD-SLAM has failure recovery and loop closure, which helps with the robustness of the program even after a loss in tracking, but still has poor accuracy in the colon environment.

3) SfMLearner: SfMLearner makes independent estimates at each step so did not need to initialize and did not lose tracking. However, since each estimate was independent, the scale of the trajectory drifted greatly. Therefore each step is scaled independently to produce a full trajectory estimate. This yields accurate results but only when a ground truth trajectory is known. Like the other systems, SfMLearner produces poor tracking when large deformations, spectral highlights, and occlusions are present. Even though it did not fail in Sequences 5-7 (Table II), the estimates were very inaccurate.

4) ORB-SLAM2: ORB-SLAM2 achieved the best performance within the colon simulator (Figs. 4,5) but also had the most failures. Out of all the systems tested its initialization was the slowest and required the most specific movements, prompting its real-time operation during data collection. In addition, the initialization was not consistent and would not always converge when performed. However, when an initialization was reached it was accurate and did not cause any later failures. The failure of a poor estimation was encountered the least often in ORB-SLAM2. The estimations that are made are accurate and the most sparse with good keyframing logic, failure recovery, and loop closure. However, rather than making poor estimations and continuing, ORB-SLAM2 will lose tracking most readily of the systems tested. While the failure recovery mode is very good, it often requires the user to stop or move the camera to a previous position, hindering the progression through the colon.

IV. Conclusions and Future Work

This paper presents an analysis of four popular SLAM systems (ORB-SLAM2, DSO, LSD-SLAM, and SfMLearner) in a new, open-source dataset available at `www.colorado.edu/lab/amtl/datasets` The authors believe this is the first dataset available with ground truth and large active deformations as would be seen in the colon. Even though the sequences in the dataset are somewhat idealized, the challenges presented in them can cause these programs to fail. These failures would not be acceptable in a real-world scenario as they would hinder the physician from focusing on the colonoscopy procedure. In an effort to improve robustness of SLAM systems, future work will combine traditional SLAM techniques with machine learning to provide the stability and constant predictions of learning-based methods with the continuity and consistently scaled trajectories of SLAM methods. This will enable better localization and mapping *in vivo* to give physicians real-time procedure feedback as well as improve localization and mapping accuracy in other actively deforming and dynamic environments.

References

[1] American Cancer Society *Key Statistics for Colorectal Cancer*, Jan. 8, 2020. Accessed on: Jan. 23, 2020 [Online]. Available:cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html

[2] American Cancer Society *Survival Rates for Colorectal Cancer*, Jan. 8, 2020. Accessed on: Jan. 23, 2020 [Online]. Available:cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html

[3] Lee, Yu Mi, and Kyu Chan Huh, "Clinical and biological features of interval colorectal cancer," Clinical endoscopy vol. 50, no. 3, 2017, pp. 254.

[4] D. A. Corley et al., Adenoma Detection Rate and Risk of Colorectal Cancer and Death, N Engl J Med, vol. 370, no. 14, pp. 12981306, Apr. 2014.

[5] J. C. Anderson and L. F. Butterly, Colonoscopy: Quality Indicators:, Clinical and Translational Gastroenterology, vol. 6, no. 2, p. e77, Feb. 2015.

[6] 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 Trans. Robot., pp. 18, 2019, doi: 10.1109/TRO.2019.2949466.

[7] Yim, Sehyuk, and Metin Sitti. "Design and rolling locomotion of a magnetically actuated soft capsule endoscope." IEEE Transactions on Robotics vol. 28, no. 1, 2011, pp. 183-194.

[8] Pham, Lan N., and Jake J. Abbott. "A Soft Robot to Navigate the Lumens of the Body Using Undulatory Locomotion Generated by a Rotating Magnetic Dipole Field." In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1783-1788. IEEE, 2018.

[9] M. R. U. Saputra, A. Markham, and N. Trigoni, Visual SLAM and Structure from Motion in Dynamic Environments: A Survey, ACM Computing Surveys, vol. 51, no. 2, pp. 136, Feb. 2018, doi: 10.1145/3177853.

[10] O. G. Grasa, J. Civera, and J. M. M. Montiel, EKF monocular SLAM with relocalization for laparoscopic sequences, in 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, May 2011, pp. 48164821, doi: 10.1109/ICRA.2011.5980059.

[11] Ma, Ruibin, Rui Wang, Stephen Pizer, Julian Rosenman, Sarah K. McGill, and Jan-Michael Frahm. "Real-time 3D reconstruction of colonoscopic surfaces for determining missing regions." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 573-582. Springer, Cham, 2019.

[12] H. Borgli et al. Hyper-kvasir: A comprehensive multi-class image and video dataset for gastrointestinal endoscopy, Dec 2019. [Online]. Available: osf.io/mkzcq

[13] D. Jha et al. Kvasir-seg: A segmented polyp dataset, in Proceedings of the International Conference on Multimedia Modeling (MMM). Springer, 2020. [Online]. Available: https://datasets.simula.no/kvasir-seg/

[14] J. Bernal et al. Comparative validation of polyp detection methods in video colonoscopy: Results from the miccai 2015 endoscopic vision challenge, IEEE Transactions on Medical Imaging, vol. 36, no. 6, pp. 12311249, 2017.

[15] K. Pogorelov et al. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection, in Proceedings of the 8th ACM on Multimedia Systems Conference, ser. MMSys17. New York, NY, USA: ACM, 2017, pp. 164169. [Online]. Available: http://doi.acm.org/10.1145/3083187.3083212

[16] S. Ali et al. Endoscopy artifact detection (EAD 2019) challenge dataset, CoRR, vol. abs/1905.03209, 2019. [Online]. Available: http://arxiv.org/abs/1905.03209

[17] M. Ye, S. Giannarou, A. Meining, and G.-Z. Yang, Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations, Medical Image Analysis, vol. 30, pp. 144 157, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841515001449

[18] P. Coelho et al. A deep learning approach for red lesions detection in video capsule endoscopies, in Image Analysis and Recognition, A. Campilho, F. Karray, and B. ter Haar Romeny, Eds. Cham: Springer International Publishing, 2018, pp. 553561.

[19] R. Mur-Artal and J. D. Tardos, ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras, IEEE Transactions on Robotics, vol. 33, no. 5, pp. 12551262, Oct. 2017.

[20] J. Engel, V. Koltun, and D. Cremers, Direct Sparse Odometry, IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 3, pp. 611625, Mar. 2018.

[21] J. Engel, T. Schps, and D. Cremers, LSD-SLAM: Large-Scale Direct Monocular SLAM, in Computer Vision ECCV 2014, vol. 8690, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 834849.

[22] T. Zhou, M. Brown, N. Snavely, and D. G. Lowe, Unsupervised Learning of Depth and Ego-Motion from Video, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 66126619.

[23] 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 Robot. Autom. Lett., vol. 3, no. 4, pp. 40544061, Oct. 2018, doi: 10.1109/LRA.2018.2861015.

[24] P. Furgale, T. D. Barfoot, and G. Sibley, Continuous-time batch estimation using temporal basis functions, in 2012 IEEE International Conference on Robotics and Automation, St Paul, MN, USA, 2012, pp. 20882095, doi: 10.1109/ICRA.2012.6225005.

[25] M. Quigley et al., ROS: an open-source Robot Operating System, p. 6.

[26] Z. Zhang and D. Scaramuzza, A Tutorial on Quantitative Trajectory Evaluation for Visual(-Inertial) Odometry, in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, 2018, pp. 72447251, doi: 10.1109/IROS.2018.8593941.

[27] A. Geiger, P. Lenz, and R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 33543361, doi: 10.1109/CVPR.2012.6248074.