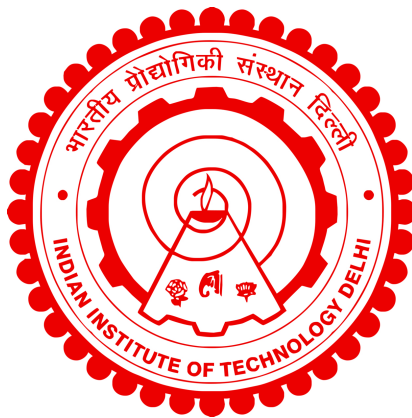


**VIRTUAL ENDOSCOPY TO GENERATE BENCHMARK
DATASETS TO AUGMENT ENDOSCOPIC VIDEO
ANALYSIS**

SARITA SINGH



DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2025

© Indian Institute of Technology Delhi (IITD), New Delhi, 2025.

**VIRTUAL ENDOSCOPY TO GENERATE BENCHMARK
DATASETS TO AUGMENT ENDOSCOPIC VIDEO
ANALYSIS**

by

SARITA SINGH

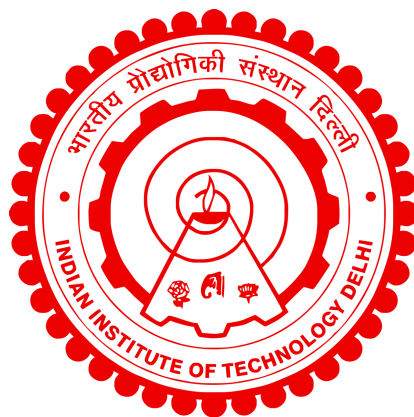
Department of Electrical Engineering

Submitted

in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2025

Certificate

This is to certify that the thesis entitled “**Virtual endoscopy to generate benchmark datasets to augment endoscopic video analysis**”, submitted by **Sarita Singh** to the Department of Electrical Engineering, Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy** in July 2025, is a record of the original, bona fide research work carried out by her under our supervision and guidance.

The matter embodied in this thesis has not been submitted in part or in full to any other University or Institute for the award of any degree or diploma.

Prof. Basabi Bhaumik

Professor,
Department of Electrical Engineering,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi - 110016,
INDIA.

Prof. Shouribrata Chatterjee

Professor,
Department of Electrical Engineering,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi - 110016,
INDIA.

Acknowledgement

First and foremost, I would like to express my gratitude to my supervisor, Prof. Shouribrata Chatterjee who believed in me and gave me a chance to pursue this degree. I am profoundly grateful to him for his invaluable support, guidance, and all the required resources throughout the research journey.

I would express my sincere gratitude to my supervisor, Prof. Basabi Bhaumik for her indispensable guidance, support, and inspiration throughout these years. She has consistently provided unwavering mentorship, offering solutions for both professional and personal challenges. I cannot count the number of times Prof. Bhaumik has offered insightful discussions and advices, helping me navigate through the challenges encountered in my research. I feel truly fortunate to have had the opportunity to work under her guidance over these years. My work and accomplishments were possible because of her consistent support and encouragement. I would also like to express my gratitude to Prof. Pankaj Sharan for the expert advice on the mathematical formulations of a section presented in this work.

I express my deep gratitude to mu SRC members Prof. Mukul Sarkar, Prof, Swades De and Prof. Ankur Gupta, for their insightful feedback and invaluable guidance during my SRC presentation. I am profoundly grateful to Prof. Jayadeva for providing the necessary resources, the servers equipped with GPUs, for successfully executing the machine learning processes.

I express my gratitude to Dr. Deepak Gunjan of All India Institute of Medical Sciences India for providing wireless capsule endoscope records data from his lab for the study. I would also like to acknowledge the contributions of my colleagues from the Indian Institute of Technology Delhi, Sarthak Gupta and Vaibhav Soni, for their support during the initial stages of this research project.

I am immensely grateful to my family members, particularly my father and mother, and my husband, for their blessings, and unwavering assistance during the course of my research. I am deeply indebted for their moral support and handling of numerous social responsibilities, which helped me remain focused on my work. This research contribution would not have been possible without their assistance, and words alone are not sufficient to express my gratitude.

(Sarita Singh)

Abstract

The current state of medical technology has demonstrated a convergence towards minimally invasive diagnosis and treatment. In an effort to mitigate the patient's distress during the screening of gastrointestinal tract using traditional wired endoscopy, researchers have introduced the wireless capsule endoscopes (WCEs). The capsule is a pill-shaped swallowable device equipped with a camera, the patient ingests. It travels through the gastrointestinal tract by virtue of the natural mechanism of peristalsis and captures images of the inner lining. An accurate localization method is required for the decision of drug administration and other therapeutic procedures as it depends on the location of the ailment.

The image-based localization techniques estimate the movement of the wireless capsule endoscope by processing the image sequence captured by the capsule camera without utilizing any additional device. Deep learning based techniques have an immense potential to improve endoscopic video analysis, specifically the pose estimation of endoscope camera and simultaneous localization and mapping. At present, the absence of large, varied, and accurately annotated datasets of the gastrointestinal tract limits the efficient learning of deep learning-based models and their effective performance evaluation.

To address the lack of accurately annotated datasets of the gastrointestinal tract, we have developed SimIntestine, the virtual models of two major sections of the gastrointestinal tract: small intestine and large intestine. The virtual models closely mimic the anatomical structure and motility of functioning intestines inside a living human being. The models incorporate the distinctive characteristics of these organs: the intricate shape with various degrees of bends, plicae circulares, villi, deformation of the walls due to peristalsis, brush border-like appearance of the small intestine, haustral folds in the large intestine, their unique surface textures, etc. We have also developed virtual capsule endoscopes for both small and large intestines. These endoscopes navigate through the virtual models of the intestines and provide color images and depth maps. The purpose of implementing various physiological properties of the organs in detail is to produce virtual capsule endoscope images that closely resemble and are almost indistinguishable from real endoscope images. This allows the synthetic dataset to be used for transfer learning in supervised as well as unsupervised learning-based depth and ego-motion detection algorithms for real endoscope image datasets.

Using SimIntestine, we trained the state-of-the-art unsupervised learning based network architectures, Endo-SfMLearner[1] and Monodepth2 [2] for depth estimation using the synthetic image dataset; and assessed its performance on images from KID [3] dataset and REAL-Colon [4] dataset. The results show that the models trained on our synthetic datasets works remarkably well on real endoscope images. We also used the techniques Endo-SfMLearner [1] and Monodepth2 [2] for the evaluation of pose estimation using the synthetic images from the virtual small and large

intestine models. At the end, we used Endo-Depth-and-Motion [5] for the 3D reconstruction of the small intestine from the synthetic images. The results demonstrate the utilization of synthetic datasets for evaluating depth and pose estimation, and 3D reconstruction techniques.

In addition, we modeled some of the health conditions, polyp, bleeding, and ulcerative colitis, for rendering images of affected regions of the intestines to aid disease detection and classification algorithms. The proposed framework provides a comprehensive and physically realistic annotated synthetic dataset benchmark of intestines which can be used to improve endoscopic video analysis, specifically in the domain of pose estimation and simultaneous localization and mapping which is challenging to obtain using real endoscope unannotated dataset.

This research is intended to help physicians locate the wireless capsule endoscope's location inside the gastrointestinal tract and find the location of the ailment detected in the endoscopic images. Our datasets with ground truth position in virtual small intestines is expected to promote the development of real-time feedback systems to locate the capsule endoscope's position inside the gastrointestinal tract and determine the location of the abnormalities detected. The comprehensive and physically realistic annotated synthetic dataset captures the deformation of the wall in the gastrointestinal tract due to peristalsis and provides a benchmark which can be used to improve the ego-motion estimation techniques for endoscopic videos.

सार

चिकित्सा प्रौद्योगिकी की वर्तमान स्थिति ने न्यूनतम आक्रामक निदान और उपचार की ओर अभिसरण प्रदर्शित किया है। पारंपरिक वायर्ड एंडोस्कोपी द्वारा जठरांत्र संबंधी मार्ग की जाँच के दौरान रोगी की परेशानी को कम करने के प्रयास में, शोधकर्ताओं ने वायरलेस कैप्सूल एंडोस्कोप (WCE) विकसित किए हैं। यह कैप्सूल एक गोली के आकार का निगलने योग्य उपकरण है जिसमें एक कैमरा लगा होता है। यह क्रमाकुंचन की प्राकृतिक क्रियाविधि द्वारा जठरांत्र संबंधी मार्ग से होकर गुजरता है और आंतरिक परत के चित्र लेता है। दवा के प्रशासन और अन्य चिकित्सीय प्रक्रियाओं के निर्णय के लिए एक सटीक स्थानीयकरण विधि की आवश्यकता होती है क्योंकि यह रोग के स्थान पर निर्भर करता है।

छवि-आधारित स्थानीयकरण तकनीकें, बिना किसी अतिरिक्त उपकरण का उपयोग किए, कैप्सूल कैमरे द्वारा कैप्चर किए गए छवि अनुक्रम को संसाधित करके वायरलेस कैप्सूल एंडोस्कोप की गति का अनुमान लगाती हैं। डीप लर्निंग-आधारित तकनीकों में एंडोस्कोपिक वीडियो विश्लेषण, विशेष रूप से एंडोस्कोप कैमरे की मुद्रा का अनुमान और समकालिक स्थानीयकरण एवं मानचित्रण, को बेहतर बनाने की अपार क्षमता है। वर्तमान में, जठरांत्र संबंधी मार्ग के बड़े, विविध और सटीक रूप से एनोटेट किए गए डेटासेट का अभाव डीप लर्निंग-आधारित मॉडलों के कुशल शिक्षण और उनके प्रभावी प्रदर्शन मूल्यांकन को सीमित करता है।

जठरांत्र संबंधी मार्ग के सटीक रूप से एनोटेट किए गए डेटासेट की कमी को दूर करने के लिए, हमने सिमइंटेस्टाइन विकसित किया है, जो जठरांत्र संबंधी मार्ग के दो प्रमुख भागों: छोटी आंत और बड़ी आंत के आभासी मॉडल हैं। ये आभासी मॉडल जीवित मानव के अंदर कार्यशील आंतों की शारीरिक संरचना और गतिशीलता की हूबहू नकल करते हैं। ये मॉडल इन अंगों की विशिष्ट विशेषताओं को समाहित करते हैं: विभिन्न स्तरों के मोड़ों वाला जटिल आकार, गोलाकार प्लिके, विली, क्रमाकुंचन के कारण दीवारों का विरूपण, छोटी आंत का ब्रश बॉर्डर जैसा रूप, बड़ी आंत में हस्ट्रल सलवटें, उनकी अनूठी सतही बनावट, आदि। हमने छोटी और बड़ी दोनों आंतों के लिए आभासी कैप्सूल एंडोस्कोप भी विकसित किए हैं। ये एंडोस्कोप आंतों के आभासी मॉडलों में नेविगेट करते हैं और रंगीन चित्र और गहराई मानचित्र प्रदान करते हैं। अंगों के विभिन्न शारीरिक गुणों को विस्तार से लागू करने का उद्देश्य आभासी कैप्सूल एंडोस्कोप चित्र तैयार करना है जो वास्तविक एंडोस्कोप चित्रों से लगभग मिलते-जुलते और लगभग अप्रभेद्य हों। इससे सिंथेटिक डेटासेट को वास्तविक एंडोस्कोप छवि डेटासेट के लिए पर्यवेक्षित और साथ ही अपर्यवेक्षित शिक्षण-आधारित गहराई और अहं-गति पहचान एल्गोरिदम में स्थानांतरण शिक्षण के लिए उपयोग करने की अनुमति मिलती है।

सिमइंटेस्टाइन का उपयोग करते हुए, हमने सिंथेटिक इमेज डेटासेट का उपयोग करके गहराई का अनुमान लगाने के लिए अत्याधुनिक अनसुपरवाइज्ड लर्निंग-आधारित नेटवर्क आर्किटेक्चर, एंडो-एसएफएमलर्नर[1] और

मोनोडेप्थ2 [2] को प्रशिक्षित किया; और केआईडी [3] डेटासेट और रियल-कॉलन [4] डेटासेट की छवियों पर इसके प्रदर्शन का आकलन किया। परिणाम दर्शाते हैं कि हमारे सिंथेटिक डेटासेट पर प्रशिक्षित मॉडल वास्तविक एंडोस्कोप छवियों पर उल्लेखनीय रूप से अच्छा काम करते हैं। हमने आभासी छोटी और बड़ी आंत के मॉडल से सिंथेटिक छवियों का उपयोग करके पोज़ अनुमान के मूल्यांकन के लिए एंडो-एसएफएमलर्नर [1] और मोनोडेप्थ2 [2] तकनीकों का भी उपयोग किया। अंत में, हमने सिंथेटिक छवियों से छोटी आंत के 3डी पुनर्निर्माण के लिए एंडो-डेप्थ-एंड-मोशन [5] का उपयोग किया।

इसके अलावा, हमने रोग पहचान और वर्गीकरण एल्गोरिदम में सहायता के लिए आंतों के प्रभावित क्षेत्रों की छवियों को प्रस्तुत करने हेतु कुछ स्वास्थ्य स्थितियों, जैसे पॉलीप, रक्तस्राव और अल्सरेटिव कोलाइटिस का मॉडल तैयार किया। प्रस्तावित ढाँचा आंतों का एक व्यापक और भौतिक रूप से यथार्थवादी एनोटेटेड सिंथेटिक डेटासेट बेंचमार्क प्रदान करता है जिसका उपयोग एंडोस्कोपिक वीडियो विश्लेषण को बेहतर बनाने के लिए किया जा सकता है, विशेष रूप से पोज़ अनुमान और समकालिक स्थानीयकरण और मानचित्रण के क्षेत्र में, जिसे वास्तविक एंडोस्कोप अननोटेटेड डेटासेट का उपयोग करके प्राप्त करना चुनौतीपूर्ण है।

इस शोध का उद्देश्य चिकित्सकों को जठरांत्र संबंधी मार्ग के अंदर वायरलेस कैप्सूल एंडोस्कोप की स्थिति का पता लगाने और एंडोस्कोपिक छवियों में पाई गई बीमारी का स्थान निर्धारित करने में मदद करना है। आभासी छोटी आंत में वास्तविक स्थिति वाले हमारे डेटासेट से जठरांत्र संबंधी मार्ग के अंदर कैप्सूल एंडोस्कोप की स्थिति का पता लगाने और पाई गई असामान्यताओं का स्थान निर्धारित करने के लिए रीयल-टाइम फीडबैक सिस्टम के विकास को बढ़ावा मिलने की उम्मीद है। व्यापक और भौतिक रूप से यथार्थवादी एनोटेटेड सिंथेटिक डेटासेट, क्रमाकुंचन के कारण जठरांत्र संबंधी मार्ग की दीवार के विरूपण को दर्शाता है और एक बेंचमार्क प्रदान करता है जिसका उपयोग एंडोस्कोपिक वीडियो के लिए अहंकार-गति अनुमान तकनीकों को बेहतर बनाने के लिए किया जा सकता है।

Contents

Certificate

Acknowledgement

Abstract

Contents

List of Figures

List of Tables

Abbreviations

Symbols

1	Introduction and Literature Review	1
1.1	Introduction to endoscopy	2
1.1.1	Wired endoscopy	3
1.1.1.1	Limitations of wired endoscopy	4
1.1.2	Wireless capsule endoscopy	5
1.1.2.1	Advantages of wireless capsule endoscopy	5
1.1.2.2	Localization of wireless capsule endoscope	6
1.2	Review of localization techniques for wireless capsule endoscope	7
1.2.1	Electromagnetic wave based localization technique	8
1.2.1.1	Limitations of electromagnetic wave-based localization technique for wireless capsule endoscope in the human body	9
1.2.1.2	Modeling of RF signal propagation and attenuation in human body	10
1.2.2	Magnetic based localization technique	13
1.2.2.1	Categories of Magnetic-Based Localization Approaches	13
1.2.3	Image based localization technique	15
1.2.3.1	Process outline for image-based localization:	17

1.2.3.2	Challenges in traditional image-based pose estimation for capsule endoscopy	21
1.2.4	Learning based methods for image based localization technique	22
1.3	Thesis Organization	25
2	Related Work and Research Motivation	27
2.1	Synthetic dataset: addressing real-world challenges	27
2.2	Review of available datasets	28
2.3	Challenges with developing synthetic dataset for small intestine	35
2.4	Significance of simulated dataset of gastrointestinal tract	37
2.5	Literature gaps identified	38
2.6	Motivation for the present work	39
2.7	Objectives of the present work	41
3	Modeling of Small Intestine and Large Intestine	43
3.1	The anatomy of small and large intestine	44
3.1.1	Anatomy of the small intestine	44
3.1.2	Anatomy of the large intestine	46
3.2	Modeling of the intestines	46
3.2.1	Determining the intestinal medial axis	47
3.2.1.1	Obtaining the structure of the virtual intestines from different models	47
3.2.1.2	Determining the dimensions of different sections of virtual models	48
3.2.1.3	Determining the spatial coordinates of points along the medial axis of virtual models	49
3.2.1.4	Smoothing of the curve at the junction of the two segments	51
3.2.1.5	The distribution of points on the medial axis	52
3.2.2	Twisted tubular structure of the intestine	52
3.2.3	Deformation due to peristalsis	54
3.2.3.1	Implementation of segmentation wave	57
3.2.3.2	Implementation of peristalsis wave	59
3.2.3.3	Migrating motor complex wave	61
3.2.3.4	Peristalsis wave in virtual large intestine	62
3.2.3.5	Limitations in pathological peristalsis modeling	62
3.2.4	Plicae circulares in small intestine and haustrations in large intestine	64
3.2.4.1	Implementation of plicae circulares	65
3.2.4.2	Haustral folds in large intestine	67
3.2.5	Deformation of plicae circulares due to peristalsis	69
3.2.5.1	Peristalsis-induced translation of the plicae circulares	69
3.2.5.2	Peristalsis-induced flexural patterns in plicae circulares	71
3.3	Importing the mesh/mesh sequence in Blender	74
3.3.1	Setting the environment in Blender	76
3.3.1.1	Difference in coordinate systems of MATLAB and Blender	76
3.3.2	Exporting the time-varying effect in MATLAB	78

3.3.3	Importing the time-varying effect in Blender	79
4	Modeling of Virtual Capsule Endoscopes and Synthetic Dataset	81
4.1	Villi structure in small intestine	82
4.1.1	Utilizing geometry node to place villi	83
4.2	Texture of small and large intestine in virtual images	85
4.2.1	Implementation of texture of small intestine	86
4.2.2	Implementation of texture of large intestine	87
4.2.3	Implementation of texture of regions affected by gastrointestinal dis- eases	89
4.3	Modeling of virtual capsule endoscopes	90
4.3.1	Structure of PillCam capsules	91
4.3.1.1	Structure of PillCam SB3 capsule	91
4.3.1.2	Structure of PillCam COLON capsule	92
4.3.2	Modeling of virtual capsule endoscopes	92
4.3.3	Movement of virtual capsule and rendering of images and depth maps	94
4.3.3.1	Movement of capsule endoscope in real small and large intestines	94
4.3.3.2	Modeling the movement of virtual capsule endoscope	95
4.3.3.3	Ground truth depth maps of virtual intestines	95
4.3.4	Deformation of walls of intestine at the location of capsule	99
4.4	Impact of light attenuation on virtual endoscopic images	100
4.5	Modeling of abnormalities in gastrointestinal tract	101
4.5.1	Polyp	101
4.5.2	Hemorrhage	102
4.5.3	Ulcerative colitis	102
4.6	Source of wireless capsule endoscopy dataset for reference	103
4.7	The original capsule endoscope images and virtual capsule endoscope images	104
5	Depth and Pose Estimation: Evaluation and Results	107
5.1	Unsupervised learning based ego-motion estimation	108
5.2	Overview of self-supervised learning based ego-motion detection techniques	110
5.2.1	SfmLearner	110
5.2.1.1	Loss function in SfmLearner	111
5.2.1.2	Architecture in SfmLearner	113
5.2.2	SC-SfmLearner	115
5.2.2.1	Loss function in SC-SfmLearner	115
5.2.2.2	Architecture in SC-SfmLearner	116
5.2.3	Monopdeth2	117
5.2.3.1	Loss function in Monodepth2	117
5.2.3.2	Architecture in Monodepth2	119
5.2.4	Endo-SfmLearner	119
5.2.4.1	Loss function in Endo-SfmLearner	120
5.2.4.2	Architecture in Endo-SfmLearner	121

5.2.4.3	EndoSLAM dataset	121
5.2.5	Endo-Depth-and-Motion	124
5.2.5.1	Photometric tracking and pose estimation	124
5.2.5.2	Hamlyn dataset	126
5.3	Evaluation of depth estimation techniques	127
5.3.1	Depth maps as a constraint in pose estimation	127
5.3.2	Evaluation of depth estimation using Endo-SfMLearner and Monodepth2 techniques	128
5.4	Evaluation of pose estimation techniques	132
5.4.1	Pose estimation in self-supervised learning	132
5.4.2	Evaluation of pose estimation using Endo-SfMLearner and Monodepth2 techniques	133
5.4.2.1	Performance metrics of Endo-SfMLearner and Monodepth2 for pose-estimation evaluation	135
5.4.3	Evaluation of pose-estimation results based on Similarity Index	137
5.5	Evaluation of pose estimation with non-rigid deformation due to peristalsis	138
5.5.1	Evaluation of Endo-SfMLearner for pose-estimation with non-rigid deformation	140
5.5.2	Performance metrics of Endo-SfMLearner and Monodepth2 for pose-estimation evaluation with non-rigid deformation	141
5.5.3	Evaluation of pose-estimation results with peristalsis based on Similarity Index	141
5.5.4	Addressing the non-rigid deformation	142
5.6	Evaluation of 3D reconstruction technique	144
5.6.1	TSDF based 3D reconstruction	144
5.6.2	Evaluation of Endo-Depth-and-motion technique for 3D reconstruction of virtual small intestine	145
5.7	Evaluation of depth and pose estimation techniques using large intestine synthetic dataset	146
5.7.1	Evaluation of depth estimation techniques	146
5.7.2	Evaluation of pose estimation techniques	149
5.7.2.1	Performance metrics of Endo-SfMLearner and Monodepth2 for pose-estimation evaluation	150
5.7.3	Evaluation of performance metrics of pose-estimation based on the depth estimation evaluation metrics	151
5.7.3.1	Metrics for evaluating the depth estimation algorithms	153
5.7.3.2	Generation of ground truth depth maps	155
5.7.3.3	Evaluation of estimated depth maps against ground truth	157
5.7.4	Challenges in pose estimation of complete virtual large intestine trajectory	158
6	Conclusion and Future Scope	161
6.1	Conclusion	161
6.2	Future scope	166

A The synthetic dataset from virtual small intestine for pose estimation	169
Bibliography	171
List of Publications	187
About the Author	188

List of Figures

1.1	Different types of gastrointestinal disorders in humans.	2
1.2	Traditional wired endoscopy (source:[6]).	3
1.3	Wireless capsule endoscopy system (a) A patient swallows a PillCam capsule, which travels through the gastrointestinal tract, capturing and transmitting images (source [7]), (b) wireless capsule endoscope (source [8]), (c) the components of a wireless capsule endoscope.	6
1.4	(a) Scatter plot of path loss versus distance for deep tissue implant. (b)Distribution of shadow fading for deep tissue implant to body surface (source:[9]).	12
1.5	(a) A view of the small intestine captured during a capsule endoscopy procedure using the RAPID software (© 2024, Medtronic), and (b) localization map in the RAPID software (© 2024, Medtronic).	18
1.6	The setup of epipolar geometry: two cameras located at O_1 and O_2 observing the 3D point P. The projections of P on the image planes of O_1 and O_2 are p_1 and p_2 respectively, the gray region defined by camera locations and P is epipolar plane, orange line between camera centers is the baseline, locations where baseline intersects the two image planes called epipoles are e_1 and e_2 and blue lines defined by intersection of epipolar plane and image planes are epipolar lines.	20
1.7	The steps in feature detection and matching for pose estimation based localization technique, (a) Consecutive input frames from PillCam video, (b) Input frames after corner clipping and grayscale conversion, (c) Features extracted with ASIFT technique, (d) Feature tracking using KLT technique.	21
1.8	(a) Sequence of images, (b) convolutional neural network, and (c) path traced by the camera.	23
2.1	Examples of (a) phantom, (b) surgery and (c) synthetic datasets [10].	28
2.2	Consecutive frames from the artificial dataset showing forward motion and the rotation (source [11]).	29
2.3	Emulation hardware and virtual test beds (source [12]).	30
2.4	(a) Robotic system for moving capsule endoscope, (b) the lifelike bowel phantom with the landmark pins (source [13]).	31
2.5	Sample images of the Hamlyn dataset (source [5]).	32
2.6	(a) Organs fixed to scaffolds cut in different shapes and (b) images from ENDOSLAM dataset and VR-Caps synthetic dataset (source [1]).	33
2.7	3D models reconstructed from real patient CT scans and the rendered images (source [14]).	36

3.1	Illustration of the (a) small intestine and (b) large intestine. Taken from Anatomy and Physiology. OpenStax (Licensed: CC-BY).	45
3.2	The structure of (a) small intestine and (b) large intestine referred for de-veloping the medial axis of the virtual intestine model.	48
3.3	(a) The medial axis of the virtual small intestine made up of 75 smaller segments, and (b) the medial axis of the virtual large intestine made up of 4 segments.	49
3.4	(a) The small intestine model made up of 75 smaller segments. (b) Smoothening the junction of adjacent curves.	51
3.5	(a) The small intestine model made up of 75 smaller segments, (b) Smoothening the junction of adjacent curves, (c) Cylindrical structure along the axis of a segment.	54
3.6	Illustration of (a) segmentation wave and (b) peristalsis wave on a small intestine segment. Taken from Anatomy and Physiology. OpenStax (Li-censed: CC-BY).	56
3.7	Implementation of (a) segmentation wave and (b) peristalsis wave on a section of small intestine.	61
3.8	(a) A diagram showing plicae circulares and villi on small intestine walls. (Copyright © 2001 Benjamin Cummings, an imprint of Addison Wesley Longman, Inc.), (b) Small intestine visualized by capsule endoscopy (Pill-Cam®) (Source: [15]).	64
3.9	Modeling the plicae circulares and their characteristic features (a) the gaus-sian curve representing a single plicae and its associated geometrical pa-rameters along the longitudinal axis (b)implementation of plicae employing shape of an asymmetric ovoid along the cross section of the intestine model (c) final implementation of plicae circulares on the inner wall of small in-terestine model.	67
3.10	(a) Implementation of plicae circulares in MATLAB and (b) their effect on the rendered images.	68
3.11	(a) A diagram showing the haustration in large intestine and (b) Large intestine visualized by capsule endoscope (Source: [3]).	68
3.12	(a) Implementation of hustral folds in MATLAB and (b) The cross-section of intestine modified In shape of triangle, and (c)their effect on the rendered images.	69
3.13	(a) Bessel function graph as a circular wave pattern, (b)(c)(d) different patterns of deformation of intestine walls and plicae circulares implemented using Bessel function.	71
3.14	(a) Shape of the plicae circularis before contraction, (b) shape of the plicae circularis after contraction with the segment between A and B deformed due to compression.	72
3.15	(a) The section AB in unwrapped form, (b) the section AB convoluted into a curve.	72
3.16	(a) captured image of small intestine plicae circulares before implementation of deformation model, (b) captured image of small intestine plicae circulares after implementation of deformation model.	74
3.17	(a) Small intestine model (b) large intestine model (c) segment model. . . .	76

4.1	(a) Villi in small intestine visualized by capsule endoscopy (PillCam®) (Source: [15]), (b) modeling of villi structure of variable shapes in Blender.	83
4.2	(a) Implementation of villi in Blender using Geometry Node, (b) rendered image without villi, (c) rendered image with villi.	85
4.3	Implementation of texture of small intestine.	87
4.4	Implementation of texture of large intestine.	88
4.5	The texture of (a) original images from real capsule endoscope and (b) synthetic images from virtual capsule endoscope.	88
4.6	Material texture of (a) large intestine, (b) small intestine, (c) polyp, (d) area affected by ulcerative colitis and (e) bleeding.	90
4.7	Modeling of virtual capsule endoscopes of small intestine and large intestine.	92
4.8	Movement of the virtual capsule endoscope along the axis of small intestine and the rendered images.	96
4.9	Deformation of the walls of the intestine due to capsule modeled using cast modifier in Blender.	99
4.10	Virtual capsule endoscope images (a) without mist and (b) with mist.	100
4.11	Reference images with diseases from KID [3] dataset and their corresponding implementation in synthetic dataset.	103
4.12	(a) original capsule endoscope images and (b) synthetic images from virtual PillCam SB3 capsule endoscope for small intestine.	105
4.13	(a) original colonoscope images and (b) synthetic images from virtual PillCam COLON capsule endoscope for large intestine.	106
5.1	SfMLearner network architecture for the depth, pose and explainability prediction modules. (a) single-view depth network (DispNet [16] architecture with multi-scale side predictions) (b) pose network and explainability mask source[[17]].	113
5.2	(a) EndoSLAM LowCam small intestine images and (b) EndoSLAM High-Cam colon images.	124
5.3	Sample images from Hamlyn dataset	127
5.4	Ground truth depth maps corresponding to sample images from the proposed synthetic dataset.	130
5.5	Depth maps generated by the Endo-SfMLearner and Monodepth2 trained on synthetic images for (a) real capsule endoscope images, (b) virtual capsule endoscope images.	131
5.6	Ground truth and the estimated trajectory of a section of virtual small intestine using (a) EndoSfMLearner and (b) Monodepth2.	134
5.7	Evaluation of Patch SSIM along with ground truth vs. estimated trajectory plot and error metrics for (a) segment 6 of virtual small intestine, (b) segment 10 of virtual small intestine.	139
5.8	Ground truth and estimated trajectory of a section of virtual small intestine with non-rigid deformation of walls due to peristalsis, using (a) EndoSfMLearner and (b) Monodepth2 technique.	140
5.9	Evaluation of Patch SSIM along with ground truth vs. estimated trajectory plot with peristalsis and error metrics for (a) segment 6 of virtual small intestine, (b) segment 10 of virtual small intestine.	143

5.10	3D map reconstruction of 11 segments using the TSDF technique.	146
5.11	Ground truth depth maps corresponding to sample images from the proposed large intestine synthetic dataset.	147
5.12	Depth maps generated by the Endo-SfMLearner and Monodepth2 trained on large intestine model synthetic images for (a) real capsule endoscope images, (b) virtual capsule endoscope images.	148
5.13	Ground truth and the estimated trajectory of various sections of virtual large intestine using Endo-SfMLearner.	150
5.14	Ground truth and the estimated trajectory of various sections of virtual large intestine using Monodepth2.	150
5.15	Error metrics for depth estimation techniques, EndoSfMLearner and Monodepth2 for different segments of the virtual large intestine model.	158
5.16	Ground truth and the estimated trajectory of complete virtual large intestine using (a) EndoSfMLearner and (b) Monodepth2.	159

List of Tables

1.1	List of commercial capsule endoscopes (source [18]).	7
1.2	Parameters for the statistical path loss model (source:[9]).	12
3.1	Structure of the small intestine and large intestine.	46
3.2	Parameters of segmentation and peristalsis wave in small intestine.	56
5.1	Comparison of Endo-SfMLearner and Monodepth2 for self-supervised pose estimation, highlighting differences in target applications, network architecture, and handling of dynamic scenes and occlusions.	133
5.2	Error metrics of pose estimation results for virtual small intestine. The lower values of ATE, Trans. RPE, and Rot. RPE indicate more accurate pose estimation.	136
5.3	Error metrics of pose estimation results for virtual small intestine segment with peristalsis. The lower values of ATE, Trans. RPE, and Rot. RPE indicate more accurate pose estimation.	141
5.4	Error metrics of segment wise pose estimation results for virtual large intestine.	151
5.5	Error metrics of pose estimation results for virtual large intestine.	159

Abbreviations

GI	G astro I ntestinal
CRC	C olo R ectal C ancer
WCE	W ireless C apsule E ndoscopy
CMOS	C omplementary M etal O xide S emiconductor
LED	L ight E mitting D iode
UHF	U ltra H igh F requency
FDA	F ood and D rug A dministration
CCD	C harge C oupled D evice
RF	R adio F requency
SLAM	S imulatneous L ocalization A nd M apping
FCC	F ederal C ommunication C ommission
MICS	M edical I mpant C ommunication S ervices
DoA	D irection of A rrival
ToA	T ime of A rrival
TDoA	T ime D ifference of A rrival
RFID	R adio F requency I Dentification
RSSI	R eceived S ignal S trength I ndicator
IMU	I ntertial M easurement U nit
HFSS	H igh F requency S tructure S imulator
SAR	S pecific A bsorption R ate
SIFT	S cale- I nvariant F eature T ransform
ASIFT	A ffine S cale- I nvariant F eature T ransform
SURF	S peeded U p R obust F eatures
FLANN	F ast L ibrary for A pproximate N earest N eighbors
ORB	O riented F AST and R otated B RIEF
MSAC	M -estimator S Amples C onsensus
RANSAC	R andom S Amples C onsensus
KLT	K anade- L ucas- T omasi

SfS	S hape from S hading
SfM	S tructure from M otion
MVS	M ulti V iew S tereo
ICP	I terative C losest P oint
RAPID	R eporting A nd P rocessing of I mages and D ata
CNN	C onvolutional N eural N etwork
RNN	R ecurrent N eural N etwork
SB	S mall B owel
MDR	M odel of D eformable R ings
PVC	P olyvinyl C hloride
CT	C omputed T omography
DICOM	D igital I maging and C ommunications in M edicine
VR	V irtual R eality
SimInt	S imulated I ntestine
MMC	M igrating M otor C omplex
PC	P licae C irculares
STL	S tereolithography
DispNet	D isparity E stimation N etwork
PoseNet	P ose E stimation N etwork
BSDF	B idirectional S cattering D istribution F unction
LIDAR	L ight D etection A nd R anging
RMSE	R oot M ean S quare E rror
Abs Rel	A bsolute R elative
MAE	M ean A bsolute E rror
log RMSE	l ogarithmic R oot M ean S quare E rror
GIANA	G astrointestinal I mage A NAlysis
DOF	D egrees O f F reedom
ATE	A bsolute T rajectory E rror
trans RPE	t ranslational R elative P ose E rror
rot RPE	r otational R elative P ose E rror
SSIM	S tructural S imilarity I ndex M easure
PSSIM	P atch S tructural S imilarity I ndex M easure
SC	S cale C onsistency
ESAB	E ndoSfMLearner S patial A ttention B lock
SO	S pecial O rthogonal G roup
SE	S pecial E uclidean G roup
TSDF	T runcated S igned D istance F unction

Symbols

a	radius of the tube
A, B	points along the circumference of plicae circularis
$A_{mod}(t)$	amplitude of segmentation wave function $r_{seg}(d, t)$
$A_p(t)$	amplitude of the gaussian functions at time t
$A_{peris}(t)$	amplitude of the peristalsis wave
A_{seg}	amplitude of segmentation wave
B	magnetic flux intensity
c	velocity of wave
$d(n, k)$	distance from pylorus to k^{th} point of the n^{th} segment
$d_s(x, y, z)$	signed distance from voxel at (x, y, z) to the nearest surface
$D_t(p)$	predicted depth at pixel p
E	essential matrix
\hat{E}_s	explainability prediction network
$f_r(d, t)$	sinusoidal waveforms at distance d at time t
F	fundamental matrix
$g_r(d, t)$	sinusoidal waveforms (half wavelength apart) at distance d at time t
\tilde{h}	radial displacement of walls from center line
I_s	source image
I_t	target image
\hat{I}_s	synthesized image generated after warping source image I_s
$J_n(x)$	zeros of bessel function
$kx(i), ky(i), kz(i)$	coordinates of beginning node of i^{th} segment
K	camera intrinsics matrix
l_{AB}	length of section of plicae circularis between points A and B
l_s	distance between two adjacent points on the medial axis
L	path loss
L_{geo}	geometry consistency loss
$L_{\min\text{-reproj}}$	minimum reprojection loss

$L_{\text{multi-scale}}$	multi-scale loss
L_{photo}	photometric loss
L_{reg}	regularization term
L_{smooth}	smoothness loss
L_{vs}	view synthesis loss
\vec{m}	magnetic dipole moment vector
n	total number of frames
nk	order of bessel function
n_s	number of segments in model
$n_{p(i)}$	number of points in i^{th} segment
N	number of folds in plicae circularis
p	pixel
p'	pixel after reconstruction
$p(t)$	controls the checkered appearance of segmentation wave
pk	number of zeros of bessel function
P_i^{est}	estimated position at i^{th} frame
P_i^{gt}	ground truth position at i^{th} frame
$q(d)$	controls the placement of plicae circulares
r_i	radius of inner tubular structure of intestine
R_i	radius of outer tubular structure of intestine
r'_i	radius of inner tubular structure of intestine after contraction
R'_i	radius of outer tubular structure of intestine after contraction
r_1	radius of outer two quarter loops of plicae circularis after contraction
r_2	radius of inner semicircle loop of plicae circularis after contraction
\vec{r}	position vector
$r_{pc}(d)$	radial displacement of intestine due to plicae circulares
$r_{peris}(d, t)$	radial displacement of intestine due to peristalsis wave
$r_{seg}(d, t)$	radial displacement of intestine due to segmentation wave
$r_{si}(d)$	radial displacement of intestine due to peristalsis and segmentation
R	rotation matrix
R_i^{est}	estimated rotation matrix at i^{th} frame
R_i^{gt}	ground truth rotation matrix at i^{th} frame
s	scale of source image
s_i	spacing between two plicae circulares of i^{th} segment
S	shadowing factor
t	time
t_{pc}	thickness of plicae circularis
\tilde{t}	time

t_i^{est}	estimated translation vector at i^{th} frame
t_i^{gt}	ground truth translation vector at i^{th} frame
T	translation matrix
T_s	time period of the segmentation wave
$T_{t \rightarrow t+1}$	6-DOF relative pose matrix from image t to $t + 1$
T_β	affine brightness transformation
v_p	velocity of peristalsis wave
$x(i, j), y(i, j), z(i, j)$	coordinates of j^{th} point of i^{th} segment
$x_{\text{arc}}(i), y_{\text{arc}}(i), z_{\text{arc}}(i)$	magnitudes of arc along x, y, z axes of i^{th} segment
\tilde{x}	axial coordinate
X	total section of plicae circularis where there is no fold
w_i	width of plica circularis of i^{th} segment

Greek Symbols

α_t	contrast of image
β_t	brightness of image
γ	path loss component
δ	truncation threshold
λ	wavelength of wave
λ_e	weight for explainability mask regularization
λ_p	spatial width of peristalsis wave
λ_s	wavelength of segmentation wave
λ_{smooth}	weight for depth smoothness loss
μ_0	magnetic permeability of air
∇_x	spatial gradients along x direction
∇_y	spatial gradients along y direction
$\Omega_{pc}(d)$	Gaussian curve for modeling plicae circulares
π	projection function
ϕ	amplitude of the wave
$\phi_1(d, t), \phi_2(d, t)$	spatially displaced Gaussian functions
σ_p	standard deviation of gaussian functions, $\phi_1(d, t)$ and $\phi_2(d, t)$
σ_{pc}	standard deviation of gaussian function, $\Omega_{pc}(d)$
σ_s	standard deviation of shadowing factor
θ_{s1}, θ_{s2}	points along the circumference of plicae circularis