8-2017

Image Registration to Map Endoscopic Video to Computed Tomography for Head and Neck Radiotherapy Patients

William S. Ingram

IMAGE REGISTRATION TO MAP ENDOSCOPIC VIDEO TO COMPUTED TOMOGRAPHY

FOR HEAD AND NECK RADIOTHERAPY PATIENTS

by

William Scott Ingram, B.S.

APPROVED:

_____

Laurence Court, Ph.D.
Advisory Professor

_____

Arvind Rao, Ph.D.

_____

Xin Wang, Ph.D.

_____

Richard Wendt III, Ph.D.

_____

Jinzhong Yang, Ph.D.

APPROVED:

_____

Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

IMAGE REGISTRATION TO MAP ENDOSCOPIC VIDEO TO COMPUTED TOMOGRAPHY FOR HEAD AND NECK RADIOTHERAPY PATIENTS

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

William Scott Ingram, B.S.

Houston, TX

August 2017

Dedication

This work is dedicated to my fiancée Reagen DePriest, who has put up with me complaining about graduate school for the past four years.

Acknowledgements

First I want to thank my advisor, Laurence Court, for his unwavering support. His enthusiasm for research fostered a wonderful learning environment, and his confidence was crucial to my success on a difficult project that pushed me far beyond the skill set with which I entered graduate school.

I was very fortunate to be a member of a large research group, and I would like to thank all of those colleagues past and present for their willingness to listen to my ideas, help me with my writing, and suffer through my jokes: Brian Anderson, Carlos Cardenas, Joey Cheung, Xenia Fave, David Fried, Rachel Ger, Kelly Kisling, Rachel McCarroll, Tucker Netherton, Constance Owens, Peter Park, Ryan Williamson, Adam Yock, and Henry Yu.

A special thanks goes to my best friend and colleague Ashley Rubinstein, who has been with me every step of the way since we moved to Houston six years ago. She has made me a better person academically, professionally, and personally, and I'm not sure how I will handle having our desks be more than ten feet apart. But that won't be an issue after our residencies when we are co-PIs, so we should probably start working on our joint CV.

Finally, I would like to thank the members of my advisory committee for their assistance and guidance over the years: Beth Beadle, Arvind Rao, Xin Wang, Richard Wendt III, and Jinzhong Yang. Their expertise and mentorship was an important part of the success of this work.

Abstract

IMAGE REGISTRATION TO MAP ENDOSCOPIC VIDEO TO COMPUTED TOMOGRAPHY FOR HEAD AND NECK RADIOTHERAPY PATIENTS

William Scott Ingram, B.S.

Advisory Professor: Laurence Court, Ph.D.

The purpose of this work was to explore the feasibility of registering endoscopic video to radiotherapy treatment plans for patients with head and neck cancer without physical tracking of the endoscope during the examination. Endoscopy-CT registration would provide a clinical tool that could be used to enhance the treatment planning process and would allow for new methods to study the incidence of radiation-related toxicity.

Endoscopic video frames were registered to CT by optimizing virtual endoscope placement to maximize the similarity between the frame and the virtual image. Virtual endoscopic images were rendered using a polygonal mesh created by segmenting the airways of the head and neck with a density threshold. The optical properties of the virtual endoscope were matched to a calibrated model of the real endoscope. A novel registration algorithm was developed that takes advantage of physical constraints on the endoscope to effectively search the airways of the head and neck for the desired virtual endoscope coordinates.

This algorithm was tested on rigid phantoms with embedded point markers and protruding bolus material. In these tests, the median registration accuracy was 3.0 mm for point measurements and 3.5 mm for surface measurements. The algorithm was also tested on four endoscopic examinations of three patients, in which it achieved a median registration accuracy of 9.9 mm. The uncertainties caused by the non-rigid anatomy of the head and neck and differences in patient positioning between endoscopic examinations and CT scans were examined by taking repeated measurements after placing the virtual endoscope in surface meshes created from different CT scans. Non-rigid anatomy introduced errors on the order of 1-3 mm. Patient positioning had a larger impact, introducing errors on the order of 3.5-4.5 mm.

Endoscopy-CT registration in the head and neck is possible, but large registration errors were found in patients. The uncertainty analyses suggest a lower limit of 3-5 mm. Further development is required to achieve an accuracy suitable for clinical use.

Table of contents

Dedication iii

Acknowledgements iv

Abstract v

Table of contents vii

List of figures xiii

List of tables xvii

Chapter 1: Introduction 1

Chapter 2: Principal hypothesis and specific aims 10

Chapter 3: Image acquisition 12

3.1 Introduction 12

3.2 Endoscopic video 12

3.3 Virtual endoscopy 14

3.3.1 General approach 14

3.3.2 Lighting model 16

3.4 Camera calibration 20

3.4.1 The camera model 21

3.4.2 Methods 25

3.4.3 Results 26

3.4.4 Summary 32

Chapter 4: Image registration methods 33

4.1 Introduction 33

4.2 The prototypical method: frame-to-frame tracking 38
4.2.1 Algorithm description 38
4.2.2 Manual determination of initial endoscope coordinates 39
4.3 The novel method: path-based volumetric search 40
4.3.1 Algorithm description 40
4.3.2 Slicing the surface mesh and clustering seed points 45
4.4 Methods for projective measurements 49
4.4.1 General concepts 49
4.4.2 Fast projective measurements via the world transform 51
4.4.3 Computation of measurement angle from the world transform 53
4.4.4 Computation of the edge mask from the world transform 55
Chapter 5: Image registration in phantoms 57
5.1 Introduction 57
5.2 Methods 58
5.2.1 Phantom design 58
5.2.2 CT acquisition and virtual endoscopy 60
5.2.3 Endoscopic video datasets 61
5.2.4 Frame-to-frame tracking 62
5.2.5 Path-based volumetric search 63
5.2.6 Measurements of registration accuracy 66
5.3 Results 69
5.3.1 Marker phantom 69
5.3.1.1 Evaluation of seed point spacing for path-based volumetric search 69

5.3.1.2 Comparison of the two registration methods 72
5.3.2 Bolus phantom 73
5.3.2.1 Evaluation of seed point spacing for path-based volumetric search 73
5.3.2.2 Comparison of the two registration methods 75
5.3.3 The impact of scene geometry 78
5.4 Discussion 80
5.4.1 Summary 80
5.4.2 Seed point spacing for path-based volumetric search 82
5.4.3 Challenges in the bolus phantom 82
5.4.4 The impact of scene geometry 83
Chapter 6: Image registration in patients 85
6.1 Introduction 85
6.2 Methods 86
6.2.1 Patient cohorts 86
6.2.2 CT acquisition and virtual endoscopy 88
6.2.3 Endoscopic video datasets 89
6.2.4 Optimization of virtual endoscopy lighting parameters 93
6.2.4.1 Creation of structure masks 94
6.2.4.2 Selection of ROI locations 96
6.2.4.3 Histogram comparison metrics 98
6.2.4.4 Optimization methods and results 100
6.2.5 Frame-to-frame tracking 104
6.2.6 Path-based volumetric search 105

6.2.7 Measurements of registration accuracy 107
6.3 Results 109
6.3.1 Comparison of the two registration methods 109
6.3.2 The impact of scene geometry 114
6.4 Discussion 116
6.4.1 Summary 116
6.4.2 Exclusion of patients from the first cohort 118
6.4.3 Anatomical differences between real and virtual endoscopy 119
6.4.4 The impact of manual inputs 121
6.4.5 Computation times 122
6.4.6 Local optima in the Nelder-Mead simplex 122
6.4.7 The role of patient positioning 123
Chapter 7: The influence of patient positioning and non-rigid anatomy 126
7.1 Introduction 126
7.2 Methods 128
7.2.1 Patient cohort 128
7.2.2 Virtual endoscopy 129
7.2.3 Virtual endoscope paths 130
7.2.4 Measurements of projective errors 133
7.2.5 The impact of scene geometry 136
7.2.6 The impact of the interval between CT acquisitions 136
7.3 Results 137
7.3.1 The influence of non-rigid anatomy 137

7.3.2 The influence of patient positioning 140
7.3.3 The influence of scene geometry 142
7.3.4 The impact of the interval between CT acquisitions 144
7.4 Discussion 145
7.4.1 Summary 145
7.4.2 Projective measurements without real endoscopic images 146
7.4.3 The role of deformable CT registration 147
7.4.4 Trends in projective errors 148
7.4.5 The clinical context of projective errors 150
7.4.6 Caveats to the projective errors presented in this study 151
Chapter 8: Image processing parameters 153
8.1 Introduction 153
8.2 Methods 154
8.2.1 Patient dataset 154
8.2.2 Volumetric grid search near the ground truth 155
8.2.3 Similarity measures 158
8.2.4 Gaussian smoothing for virtual images 162
8.2.5 Edge-preserving smoothing for video frames 164
8.2.6 Downsampling 165
8.2.7 Masking 167
8.2.8 Analyses of calibration parameters 167
8.2.8.1 View angle 167
8.2.8.2 Distortion parameters 171

8.3 Results 174
8.3.2 Similarity measures 174
8.3.2.1 Histogram bins 176
8.3.3 Gaussian smoothing for virtual images 177
8.3.4 Edge-preserving smoothing for video frames 178
8.3.5 Downsampling 179
8.3.6 Masking 180
8.3.7 Analyses of calibration parameters 180
8.3.7.1 View angle 180
8.3.7.2 Distortion parameters 182
8.4 Discussion 184
8.4.1 Similarity measures and preprocessing 184
8.4.2 Calibration parameters 186
Chapter 9: Discussion 189
9.1 Specific aim 1 189
9.2 Specific aim 2 192
9.3 Principal hypothesis 194
9.3 Future directions 195
Bibliography 197
Vita 209

List of figures

Figure 1: The endoscope and auxiliary equipment 13

Figure 2: The configuration of lights on the endoscope 17

Figure 3: Schematic of perspective projection 24

Figure 4: The calibration rig 26

Figure 5: An example of distortion removal from an endoscopic video frame 28

Figure 6: Flowchart showing the steps of the path-based volumetric search 45

Figure 7: Example of surface slicing and seed point clustering 48

Figure 8: Illustration of the hexagon lattice used to calculate the number of seed points 49

Figure 9: Illustration of the expected impact of scene geometry on projective uncertainty 55

Figure 10: Example of the edge mask created from the world transform 56

Figure 11: The marker phantom 59

Figure 12: The bolus phantom 60

Figure 13: Virtual endoscope path creation in the marker phantom 64

Figure 14: Projective measurement of CT-space marker positions 67

Figure 15: Projective measurement of CT-space bolus contour 69

Figure 16: Comparison of seed point spacing in the marker phantom for the path based-volumetric search 71

Figure 17: Comparison of the two registration methods in the marker phantom 73

Figure 18: Comparison of seed point spacing in the bolus phantom for the path based-volumetric search 75

Figure 19: Comparison of the two registration methods in the bolus phantom 77

Figure 20: Dependence of point measurement errors on surface angle and distance 79

Figure 21: Dependence of bolus contour measurement on surface distance 80

Figure 22: Differences between patient endoscopic video characteristics 88

Figure 23: Examples of an endoscopic video frame and a virtual endoscopic image in patient MDA2 90

Figure 24: Examples of an endoscopic video frame and virtual endoscopic image in patient MDA3 93

Figure 25: Example of a structural mask used for lighting optimization 96

Figure 26: Example of the ROIs used to calculate image histograms 97

Figure 27: Virtual endoscope path creation in a patient 106

Figure 28: Comparison of the two registration methods for patient MDA1, video sequence 1 111

Figure 29: Comparison of the two registration methods for patient MDA1, video sequence 2 112

Figure 30: Comparison of the two registration methods in patient PMH1 113

Figure 31: Comparison of the two registration methods for patient PMH2 114

Figure 32: Reduction in registration error when distant points are excluded 116

Figure 33: Example of anatomical differences between endoscopic video and virtual endoscopy 120

Figure 34: An example of the anatomical regions used to in the virtual endoscope path 131

Figure 35: An example of the images in the virtual endoscope path 132

Figure 36: A schematic illustrating the measurement of projective errors 135

Figure 37: Median projective errors in the three anatomical regions for each daily CT 139

Figure 38: Projective errors at each virtual endoscope path position for three patients 140

Figure 39: Median projective errors in the three anatomical regions for the diagnostic CTs 141

Figure 40: Reduction is projective errors when distant points are excluded 143

Figure 41: Examples of the Gaussian smoothing kernels applied to virtual endoscopic images 163

Figure 42: Examples of the bilateral filters applied to endoscopic video frames 165

Figure 43: Examples of 2x and 4x downsampled virtual endoscopic images 166

Figure 44: Examples of virtual endoscopic images rendered with variable view angles for the virtual camera 169

Figure 45: The circular ROIs used to sample projective errors for the analysis of calibration parameters 170

Figure 46: Examples of distortion removal with reduced distortion models 173

Figure 47: Projective measurement errors induced by changes to the view angle of the virtual camera 182

Figure 48: Projective measurement errors induced by changes to the distortion model 184

List of tables

Table 1: Summary of the virtual endoscopy lighting model 20

Table 2: Summary of the intrinsic camera parameters..... 24

Table 3: Optimized intrinsic parameters without skew 28

Table 4: Optimized intrinsic parameters for distortion-free frames..... 29

Table 5: Optimized intrinsic parameters for distortion-free, horizontally-downsampled frames..... 31

Table 6: Description of the variables in the path-based volumetric search algorithm ... 44

Table 7: Comparison of seed point spacing in the marker phantom..... 71

Table 8: Comparison of the two registration methods in the marker phantom 72

Table 9: Comparison of seed point spacing in the bolus phantom..... 74

Table 10: Comparison of the two registration methods in the bolus phantom..... 77

Table 11: Lighting parameter ranges for the brute-force optimization.....101

Table 12: Results of the brute-force optimization of the lighting parameters 102

Table 13: Lighting parameter values used to start the simplex optimizations 102

Table 14: Results of the simplex optimizations of the lighting parameters 103

Table 15: Final results of the lighting parameter optimization.....104

Table 16: Comparison of the two registration methods in patients.....110

Table 17: Averages of median projective errors in the three anatomical regions for each daily CT 138

Table 18: Correlation values between projective error and measurement angle and measurement distance.....143

Table 19: Projective errors with and without edge masks applied 144

Table 20: Results of the volumetric grid searches for all similarity measures 175

Table 21: Results of the volumetric grid searches with different numbers of histogram bins 176

Table 22: Results of the volumetric grid searches with Gaussian smoothing of the virtual endoscopic images 177

Table 23: Results of the volumetric grid searches with edge-preserving smoothing of the endoscopic video frames 178

Table 24: Results of the volumetric grid searches with downsampled images 179

Table 25: Results of the volumetric grid searches with different view angles for the virtual camera 181

Table 26: Results of the volumetric grid searches with different models used to remove distortion from the registration frames 183

1

Introduction

Cancer of the head and neck consists of a diverse set of malignancies that develop in the epithelial cells of the nasal cavity, paranasal sinuses, oral cavity, pharynx, and larynx. In the United States, there were an estimated 61,760 new cases of head and neck cancer and 13,190 deaths from it in 2016. These account for 3.7% of all new cancer cases and 2.2% of all cancer deaths, respectively[1]. As with many types of cancer, the use of tobacco and alcohol are two of the major risk factors, and the effects of these two substances are synergistic[2, 3]. There is also a causal association between human papillomaviruses (HPV) and a subset of head and neck cancers[4]. There is some evidence for familial inheritance of the disease, and an association of increased risk has been reported for numerous hereditary cancer syndromes[5].

The anatomy of the head and neck is essential not only for basic physiological functions such as eating and breathing, but also for social interaction. This makes organ preservation and minimization of disfigurement especially important goals in the treatment of head and neck cancer. Surgery and radiotherapy are the primary treatment modalities for early-stage disease, and the addition of concurrent chemotherapy improves outcomes for patients with advanced disease, albeit with increased incidence of toxicity[6]. Head and neck radiotherapy is traditionally delivered

in daily fractions of 2 Gy to a total dose of 60-70 Gy. Many clinical trials have investigated the efficacy of unconventional fractionation schemes, including hyperfractionation and accelerated radiotherapy. In general, these studies have found that both schemes improve survival and locoregional control, but hyperfractionation has the greatest benefit[7].

Radiotherapy treatment plans for head and neck cancer are created using a computed tomography (CT) image set as a 3D representation of the patient on which to define anatomical volumes, including normal tissue structures to avoid and target volumes that encompass the tumor, its range of motion, and uncertainties in patient positioning[8]. This CT image set is referred to as the planning CT or simulation CT, and it is fundamental to the design, delivery, and evaluation of modern radiotherapy. It contains all of the spatial information about the radiation dose distribution within the patient, so it is also used extensively in retrospective studies that seek to evaluate novel treatment planning techniques or to understand the dosimetric characteristics that influence the incidence of radiation-related toxicity. CT is the modality of choice for treatment planning because the value at each voxel is determined by how much it attenuates x-rays in the diagnostic energy range (120-140 kVp on modern CT scanners), so the images can be used to simulate the deposition of dose in the therapeutic energy range (6-18 MV on modern linear accelerators). However, other imaging modalities provide additional information that can be valuable in designing treatment plans or assessing patient response.

One of the most commonly-used modalities for this purpose is positron emission tomography (PET) with ^{18}F-fluorodeoxyglucose (FDG), which allows for imaging of

tumors via the high glucose uptake exhibited by cancer cells.[9] FDG-PET can be used in conjunction with CT to improve target delineation in the lungs[10], head, and neck[11], and there is preliminary evidence supporting its utility in many other disease sites[12]. It also allows for novel treatment planning strategies, such as delineating metabolically-active sub-regions within the tumor to treat more aggressively[13]. Similar to PET is single-photon emission computed tomography (SPECT), which allows for imaging based on a different set of biological functions using a different class of radioactive tracers. It is less commonly used for radiotherapy treatment planning, but there is some experience supporting the use of SPECT lung perfusion images to reduce the dose to regions of the lungs that contribute the most to overall function[14]. Another useful imaging modality is magnetic resonance imaging (MRI), which allows for imaging based on differences in the magnetic resonance relaxation properties of tissues. MRI provides greater contrast between different types of soft tissue than CT, which facilitates the visual identification of a tumor's extent. It is used extensively for target volume delineation in the central nervous system, and it can improve delineation in the head, neck, and pelvis[15]. Despite the utility of these imaging modalities, the calculation and optimization of radiation dose must be done on CT. For this reason, any supplemental imaging modality must still be registered to the simulation CT if it is to be used in an objective and quantitative way to design radiotherapy treatment plans, or to access the dosimetric information that the plans contain.

Image registration is the process of establishing spatial correspondence between all points in a pair of related images. In general, image registration has three components:

1. The similarity measure, a value representing the relative configuration of the two images that is maximized when their alignment is optimal.
2. The transformation, a mathematical expression that defines how the image points are allowed to move.
3. The search strategy, an optimization algorithm that updates the transformation to maximize the similarity measure.

In the case of integrated PET-CT scanners, the only difference between the two images is the known translation of the patient between the two detectors, so image registration is trivial provided that the patient remains motionless on the scanner couch. When the images are acquired on different devices at different points in time, the relationship between the two coordinate systems is unknown, and there may be large differences in patient positioning and anatomical configuration. In this case, registering the two images is more challenging, and the methods to do so have been the subject of extensive research[16]. A concept related to image registration is image fusion, which is the combination of information from multiple images, typically from different modalities, into a single image. The most common example of image fusion in radiation oncology is PET-CT, in which the CT image is displayed in grayscale with the PET image overlaid as a color wash. Image registration is a prerequisite for image fusion, but image fusion is not a necessary output of image registration.

Despite the growing prevalence of supplementary imaging modalities and their demonstrated utility for head and neck radiotherapy, endoscopy has received very little attention in this context. An endoscope is an optical device consisting of a control

section that is manipulated by the operator and a rigid or flexible insertion tube that is used to inspect luminal organs. An external light source is fiber-optically coupled to the distal end of the insertion tube, which can often be angulated via controls on the body. Modern endoscopes have an image sensor at the distal end, which allows digital video to be displayed for multiple viewers in the procedure room and recorded for later use. There are many types of endoscopes specialized for different portions of the respiratory and gastrointestinal tracts, and for organs such as the bladder and kidneys. Insertion tube diameters range from approximately 3 to 15 mm with working lengths up to 2200 mm[17]. Some have instrument channels that allow for retrieval of tissue samples and other procedures, and there are echoendoscopes that have an ultrasound transducer at the distal end for imaging of anatomy beyond the luminal wall.

Registration and fusion of endoscopic video and CT has received considerable attention for guidance of surgical and bronchoscopic procedures. Some of the earliest work used electromagnetic sensors attached to the endoscope and a receiver headset worn by the patient to track the endoscope during endoscope-guided sinus surgeries[18, 19]. This tracking system allowed for the display of coronal, sagittal, and axial CT images corresponding to the endoscope's position. Around the same time, a method was developed to localize the bronchoscope during transbronchial biopsies by registering real bronchoscopic images to virtual bronchoscopic images derived from CT[20]. Since then, several groups have developed image-based methods to track bronchoscope position and provide navigational assistance during bronchoscopic procedures[21–33]. These methods employ a variety of computer vision techniques, including optical flow, detection of corresponding image features, and structure from motion. They include the

development of novel optimization algorithms and image similarity measures to improve tracking performance, and they all share one thing in common: comparison of real endoscopic images to virtual endoscopic images derived from CT. Electromagnetic tracking has also been used for CT-based bronchoscope navigation[34], and an image-based method has been developed to register endoscopic video to CT for guidance during skull base surgery[35]. In this method, the position of a rigid endoscope was calculated by tracking infrared markers on its control section with cameras in the operating room. This position was then refined by detecting matching features extracted from real and virtual endoscopic images.

Registration and fusion of endoscopic video and CT for interventional procedures has been an active field of research for over 20 years, the development of which is catalogued in review articles on endoscopic surgical guidance[36] and endoscopic navigation based on computer vision[37]. However, very little attention has been given to endoscopy-CT registration in the head and neck and its applications for radiotherapy patients. Endoscopic examination is an important tool for the initial evaluation of a large portion of these patients[38]. It provides a clear visual inspection of tumor extent that can reveal early-stage disease or mucosal irregularities that are not appreciable on CT. This information may be valuable for target delineation, but without a method to register the endoscopic video to the planning CT, it can only be used subjectively based on the physician's expertise. Endoscopic examinations are also used to assess patients during and after radiotherapy. In this setting, endoscopy-CT registration could be used to overlay the radiation dose distribution from the planning CT on the endoscopic

video. This would allow the physician to accurately assess the dose delivered to various anatomical structures seen in the video.

Outside of routine clinical use, endoscopy-CT registration could play a role in improving our understanding of the dosimetric factors that influence radiation-related normal-tissue toxicity. Mucositis, which is an inflammation and ulceration of epithelial cells, is a common and often debilitating side effect that occurs for about 80% of head and neck radiotherapy patients[39]. Severe symptoms occur in up to 56% of patients, and it requires hospitalization in up to 32%. However, mucositis is visible only by endoscopy when it occurs outside of the oral cavity. Current studies of the dosimetric factors that influence toxicity generally rely on dose-volume histograms derived from the treatment plan[40], but these histograms, which have inherently limited spatial information, are not ideal for studying mucositis, which has a limited spatial extent that is not visible on CT. Endoscopy-CT registration would provide a method to segment areas of mucositis on the treatment plan, which would allow for more detailed toxicity studies that could improve the quality of life of head and neck radiotherapy patients.

Unlike most forms of medical image registration, endoscopic video and CT have different dimensionality: endoscopic video is a 2D projection of 3D space, and CT is a volumetric representation of 3D space. This disparity is one of the biggest challenges, and there are two broad categories of approaches to overcome it: project CT space to 2D via virtual endoscopy, or reconstruct a 3D surface from endoscopic video. The other major distinction in endoscopy-CT registration methods is whether or not prospective endoscope tracking is used, generally with electromagnetic sensors as previously discussed. In recent years, two groups have studied the registration and fusion of

endoscopic video and CT in the head and neck. One group has used electromagnetic endoscope tracking and virtual endoscopy to improve target delineation[41] and to overlay radiation dose from the planning CT on endoscopic video[42]. The other group does not use prospective tracking, and has published preliminary results using structure-from-motion techniques to reconstruct the 3D surface of the airways and register it to the planning CT[43–45].

Both of these approaches have drawbacks. In order to use electromagnetic tracking, the patient be in exactly the same position for the endoscopic examination and the planning CT in order that the coordinate systems match. In standard clinical practice, the endoscopic examination is a simple procedure performed in the seated position that may take no more than 15 minutes. However, the planning CT is acquired in the supine position with the patient's head, neck, and sometimes shoulders secured in a molded thermoplastic mask for positioning reproducibility throughout the course of radiotherapy. The burden of using a CT couch and thermoplastic mask for the endoscopic examination, as well as the fact that electromagnetic tracking endoscopes require customization with expensive equipment that is not available in most clinics, means that the tracking approach is not suitable for routine clinical use. On the other hand, the 3D reconstruction approach requires multiple views of the surface from different viewpoints, and automatically identifying corresponding points in those views. This may be difficult with endoscopic video in the head and neck, which contains highly-variable illumination and a large degree of muscle motion.

The motivation for the work presented in this dissertation was to develop an endoscopy-CT image registration framework for the airways of the head and neck that

avoids these drawbacks, and to investigate the sources of uncertainty in this poorly-characterized form of image registration. All methods were developed with the explicit goal of requiring only images and equipment that are available in routine clinical practice. This ensures that this work can serve as a foundation to implement endoscopy-CT image registration with the widest possible availability for head and neck cancer patients.

2

Principal hypothesis and specific aims

Principal hypothesis: Endoscopic video in the head and neck can be registered to CT without prospective physical endoscope tracking through the use of virtual endoscopy.

Specific aim 1: Develop, test, and optimize a method to register endoscopic video of the head and neck to CT

Hypothesis: Endoscopic video frames can be registered to CT with an accuracy of 5 mm in rigid phantoms and 10 mm in patients.

To achieve this aim, an algorithm was developed to search virtual endoscope coordinate space for the virtual image that best matches a given endoscopic video. This algorithm was developed initially in rigid phantoms that contain fiducial markers. Its accuracy was tested by mapping video-frame measurements of these markers to CT space and comparing them with the ground truth. Then the algorithm was tested on head and neck radiotherapy patients using endoscopic examinations and CT scans acquired as part of standard clinical practice. The patient data set was

also used to optimize the image processing parameters that influence registration accuracy.

Specific aim 2: Investigate the sources of uncertainty in projective mapping via virtual endoscopy and determine their impact on endoscopy-CT registration.

Hypothesis: Patient positioning will have the largest impact on registration errors.

To achieve this aim, the impacts of daily variations in non-rigid anatomy and patient positioning differences between CT scans and endoscopic videos were investigated using virtual endoscopic measurements on CT scans of the same patients taken on different days and different positions. In addition to these sources of uncertainty, the impacts of the focal length and radial distortion of the endoscope's camera on virtual endoscopic measurements were investigated.

3

Image acquisition

3.1 Introduction

This chapter describes the equipment and software that were used to acquire endoscopic video and virtual endoscopic images. This is largely foundational material that will be referenced throughout the remainder of this dissertation. The endoscope and some characteristics of the recorded videos are described in Section 3.2. Virtual endoscopy and the software used to render virtual images are discussed in Section 3.3. Section 3.4 presents the methods and results of camera calibration, which is the process of measuring the optical characteristics of the endoscope's camera.

3.2 Endoscopic video

Endoscopic videos were acquired using an ENF-VQ rhinolaryngoscope (Olympus America, Center Valley, PA). It is a flexible video endoscope with a 300-mm working length. Its outer diameter is 3.6 mm along the working length, and 3.9 mm at the distal end, which houses the lens and the image sensor. The distal end can be angulated up and down 130 degrees by manipulating a lever on the control section. The endoscope

was operated using the Visera Pro OTV-S7Pro camera control unit and the Visera Pro CLV-S40Pro light source. The videos were recorded using the nStream G3 HD medical digital recording and image management device (Image Stream Medical, Littleton, MA), which produced MPEG-2 video files at a frame rate of 30 frames per second and a resolution of 720 x 486 pixels[2]. These auxiliary devices and a monitor that displays the endoscope's live output are housed in a mobile tower, which is kept in a dedicated procedure room that includes an exam chair for the patient. The endoscope and the procedure room setup are shown in Figure 1.

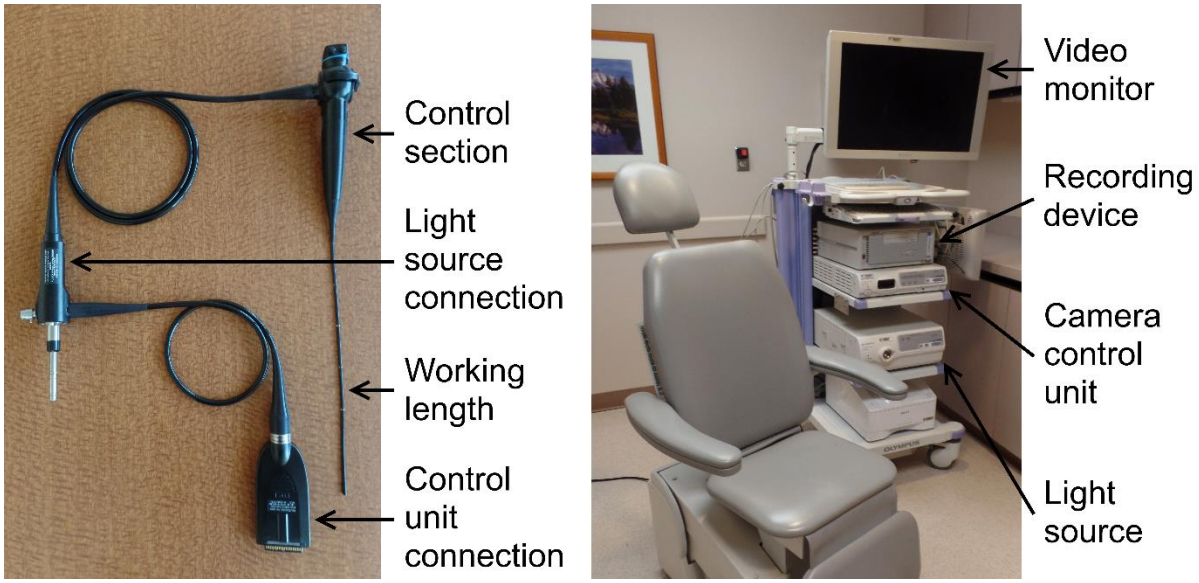

Figure 1: The endoscope and auxiliary equipment. Left: the Olympus ENF-VQ rhinolaryngoscope used to acquire endoscopic videos. Right: The exam chair and endoscope control tower in the head and neck clinic.

3.3 Virtual endoscopy

3.3.1 General approach

Virtual endoscopy is the rendering of 2D images using 3D models generated from CT or MRI, providing images similar to those produced by an endoscope placed inside the anatomy. It was developed in the mid-1990s as a non-invasive diagnostic tool, and it was quickly applied in a variety of settings, including neurosurgical planning[46], training for endoscope operators[47], and anatomical evaluation in the aorta[48], colon[49], and the airways of the head and neck[50–52]. There are two basic approaches to virtual endoscopy: volume rendering and surface rendering[53]. With volume rendering, every voxel is assigned an opacity and a color, and the virtual image is generated by casting rays through the volume. With surface rendering, an explicit geometrical representation of one or more structures of interest are created, typically by segmenting the images with a threshold and applying an algorithm such as marching cubes[54] to generate a polygonal mesh. Surface rendering is more common in virtual endoscopic applications because the anatomy of interest is typically an air-tissue interface, and it is the approach used for the work presented in this dissertation.

Virtual endoscopic images were rendered using the Visualization Toolkit (VTK) (Kitware, Inc., Clifton Park, NY), an open-source software library[55]. VTK is written in the C++ programming language, and the Python programming language binding was used for the work presented in this dissertation. The VTK rendering process is an object-oriented pipeline with a scene containing a `vtkCamera`, `vtkLights`, `vtkActors`

representing the objects to be rendered, and a `vtkRenderWindow` that produces the image. Throughout this dissertation, the terms virtual image, virtual endoscopic image, and virtual frame will be used to refer to the rendered image. The principal input to the rendering pipeline is a .vtk file containing a triangular mesh that represents the surface to be displayed. These meshes will be referred to as surface meshes or virtual endoscopy meshes. They were created with an extension of the class `vtkVoxelContoursToSurfaceFilter` that read .roi files from the Pinnacle[3] radiotherapy treatment planning software (Philips Healthcare, Andover, MA). These files contained the voxel coordinates of CT-based contours representing air-tissue interfaces. Additional details on how these interfaces were segmented are provided in Sections 5.2.2, 6.2.2, and 7.2.2.

When virtual endoscopy is used as a diagnostic tool, the optical properties of the images are required to meet only the user's subjective criteria for adequate visualization. However, for applications to endoscopy-CT image registration, these properties should match those of the real endoscope:

1. Focal lengths of the real and virtual cameras
2. Distortion introduced by the endoscope camera's lens and image sensor
3. Scene lighting, including reflectance properties and attenuation with distance from the camera

Focal length and distortion are discussion in section 3.3, and the lighting model is discussed in section 3.2.2.

3.3.2 Lighting model

The endoscope has two lights on the distal end. They are displaced 1.5 mm laterally from the center of the camera lens. Rather than circular points, they are shaped as small arcs concentric with the lens; this configuration is illustrated in Figure 2. In the VTK rendering pipeline, lighting is handled by creating one or more `vtkLight` objects, setting their intensities and other properties, and placing them in the scene. The endoscope's lights were modeled by placing two lights 1.5 mm left and right of the virtual camera. The lights were set as camera lights, which means that their position and orientation were tied to the virtual camera as it moved around the scene. In preliminary tests, virtual images rendered with this configuration were visually indistinguishable from those rendered with a single light that was coincident with the camera. For this reason, no attempt was made to model the extent of the light arcs above and below the camera.

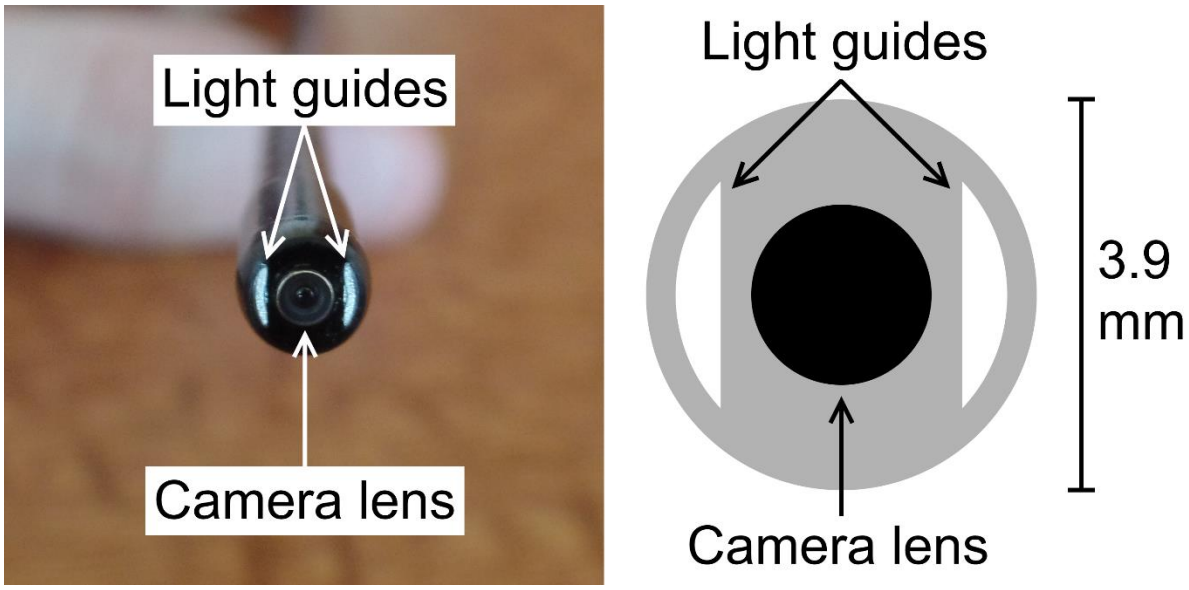

Figure 2: The configuration of lights on the endoscope. The distal ends of the fiberoptic light guides are small arcs concentric with the camera lens.

There are three flavors of light available in VTK: ambient, diffuse, and specular. Ambient light comes from all directions, so all surfaces are lit equally and the brightness does not depend on the orientation of the camera relative to the surface. Diffuse light comes from a single direction, but is reflected equally in all directions. This means that the brightness depends on the orientation of the light relative to the surface, but not on that of the camera. Specular light also comes from a single direction, but the angle of incidence is preserved when it is reflected. Specular brightness depends on the orientations of both the light and the camera relative to the surface, and it produces highlights that give the surface a shiny appearance.

In the work presented in this dissertation, only diffuse lighting was used for virtual images. Ambient lighting was rejected on the basis that the endoscope is inside a dark cavity with only its own source of light. Specular reflections are certainly present in endoscopic video. However, it was determined empirically that diffuse lighting is sufficient to reproduce overall variations in brightness. Furthermore, specular reflections are highly dependent on the local structure and texture of the surface. The virtual endoscopy meshes did not have exactly the same structure as the anatomical surfaces seen in the endoscopic videos, and it was not feasible to reproduce either the tissue textures or the presence of saliva and other fluids in the virtual images, so it was unlikely that specular reflections would coincide in the real and virtual images. In VTK, each flavor of light has its own color defined by red, green, and blue channels. To achieve diffuse lighting only, the `vtkLight` ambient and specular colors were set to $(0,\ 0,\ 0)$ and their diffuse color was set to $(1,\ 1,\ 1)$ to produce grayscale virtual images.

Positional lighting was used for all virtual images, which means that the light rays diverged from the source. This was necessary to reproduce the appearance of endoscopic video, in which the lights were very close to the surface. Within the VTK rendering pipeline, positional lights can be attenuated as the rays travel through space. The light incident on a surface point is attenuated by the following factor:

$$attenuation\ factor = \frac{\cos^{e}(\phi)}{a_c + a_l d + a_q d^2} \tag{1}$$

In the numerator, which governs the spotlight effect, ϕ is the angle between the light's direction and a vector pointing from the light's position to the surface point. The spotlight exponent e can be set by the user to determine how the light falls off towards the edges of the image. In the denominator, d is the distance between the light's position and the surface point. The constant, linear, and quadratic attenuation coefficients a_c, a_l, and a_q can be set by the user to determine how the light falls off in distal regions of the image. Details on how the values for the attenuation parameters were determined are provided in Sections 5.2.2 and 6.2.4. A summary of the virtual endoscopy lighting model used for the work presented in this dissertation is given in Table 1. The VTK rendering pipeline uses the OpenGL software library (Khronos Group, Beaverton, OR), so it uses the same lighting model, which is discussed in detail in the OpenGL Programming Guide[56].

Table 1: Summary of the virtual endoscopy lighting model.

Parameter	Value	Comment
Light type	Camera light	Position and orientation tied to virtual camera
Position	$(\pm 1.5 \text{ mm}, 0, 0)$	Two lights, left and right of virtual camera
Color	Ambient = $(0, 0, 0)$ Diffuse = $(1, 1, 1)$ Specular = $(0, 0, 0)$	Grayscale images with diffuse lighting only
Positional lighting	On	Light rays diverge from source
Intensity	Variable	Different values used for phantom and patient images (see Sections 5.2.2 and 6.2.4)
Spotlight exponent	Variable	See Equation 1. Different values used for phantom and patient images (see Sections 5.2.2 and 6.2.4)
Constant, linear, and quadratic attenuation coefficients	Variable	See Equation 1. Different values used for phantom and patient images (see Sections 5.2.2 and 6.2.4)

3.4 Camera calibration

Virtual endoscopic images are rendered with a perfect pinhole camera, but real endoscopic images have distortion introduced by imperfections in the lens and image sensor. All real cameras have some degree of distortion, but it is particularly

pronounced for endoscopes, which typically use wide-angle “fisheye” lenses to increase the field of view. This distortion must be removed so that the real and virtual images represent the same 3D scene. Doing so requires a set of values called intrinsic parameters, which are defined in sections 3.3.1. The measurement of the intrinsic parameters is known as camera calibration, and it is a foundational procedure in a wide variety of photogrammetric applications[57–62]. The methods used to perform the calibration are described in section 3.3.2, and the results of the calibration are discussed in section 3.3.3.

3.4.1 The camera model

Let $\boldsymbol{W} = (X,\ Y,\ Z)$ be a point in the camera’s reference frame, which is defined such that the camera is looking down the positive z axis with its optical center at the origin, the +X axis points to the right, and the +Y axis points down. This reference frame is also referred to as the camera-centered coordinate system. $\boldsymbol{W}$ is mapped to the image plane by perspective projection, which is simply normalization by the z component:

$$\boldsymbol{w_n} = \begin{bmatrix} X/Z \\ Y/Z \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

This is illustrated in Figure 3. Let $r = \sqrt{x^2 + y^2}$. Distortion is incorporated by displacing the projected point with radial and tangential components:

$$\boldsymbol{w_d} = (1 + c_1 r^2 + c_2 r^4 + c_3 r^6)\boldsymbol{w_n} + \begin{bmatrix} 2c_4 xy + c_5(r^2 + 2x^2) \\ c_4(r^2 + 2y^2) + 2c_5 xy \end{bmatrix} = \begin{bmatrix} x_d \\ y_d \end{bmatrix} \quad (3)$$

In this equation, the coefficients c_1, c_2, and c_3 determine the radial distortion, and the coefficients c_4 and c_5 determine the tangential distortion. Finally, the distorted point is transformed to its pixel address $\boldsymbol{u} = (u,\ v)$ in the image:

$$\boldsymbol{K} \equiv \begin{bmatrix} f_x & \alpha f_x & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$\boldsymbol{K} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (5)$$

The matrix $\boldsymbol{K}$ will be referred to as the calibration matrix. The entries f_x and f_y are the focal length of the camera expressed in units of horizontal and vertical pixels. The skew coefficient α is determined by the angle between the physical X and Y axes of the image sensor. It will be 0 if this angle is 90 degrees or very close to it, which is generally the case with modern image sensors. The entries p_x and p_y are the coordinates of the principal point, which is the projected pixel address of the point $(0,\ 0,\ 0)$. The parameters defined thus far, summarized Table 2, constitute the intrinsic parameters.

Given a point $\boldsymbol{W}' = (X',\ Y',\ Z')$ in a coordinate system that is not camera-centered, it must be rotated and translated into the camera's reference frame before perspective projection and distortion can be applied. This is accomplished by incorporating extrinsic parameters into the camera model, which consist of a 3 x 3 rotation matrix $\boldsymbol{R}$

that defines the camera's orientation and a 3 x 1 translation vector $\boldsymbol{t}$ that gives the position of the origin in the camera's reference frame. These are stacked into a 3 x 4 matrix that transforms the world coordinates to camera-centered coordinates:

$$[\,\boldsymbol{R} \mid \boldsymbol{t}\,]\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_0 \\ r_{10} & r_{11} & r_{12} & t_1 \\ r_{20} & r_{21} & r_{22} & t_2 \end{bmatrix}\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{6}$$

The extrinsic and intrinsic parameters combine to form the camera matrix $\boldsymbol{P}$, which governs projection of points in the world coordinate system onto an image taken by a camera placed in the scene:

$$\boldsymbol{P} \equiv \boldsymbol{K}\,[\,\boldsymbol{R} \mid \boldsymbol{t}\,] \tag{7}$$

$$\boldsymbol{P}\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{8}$$

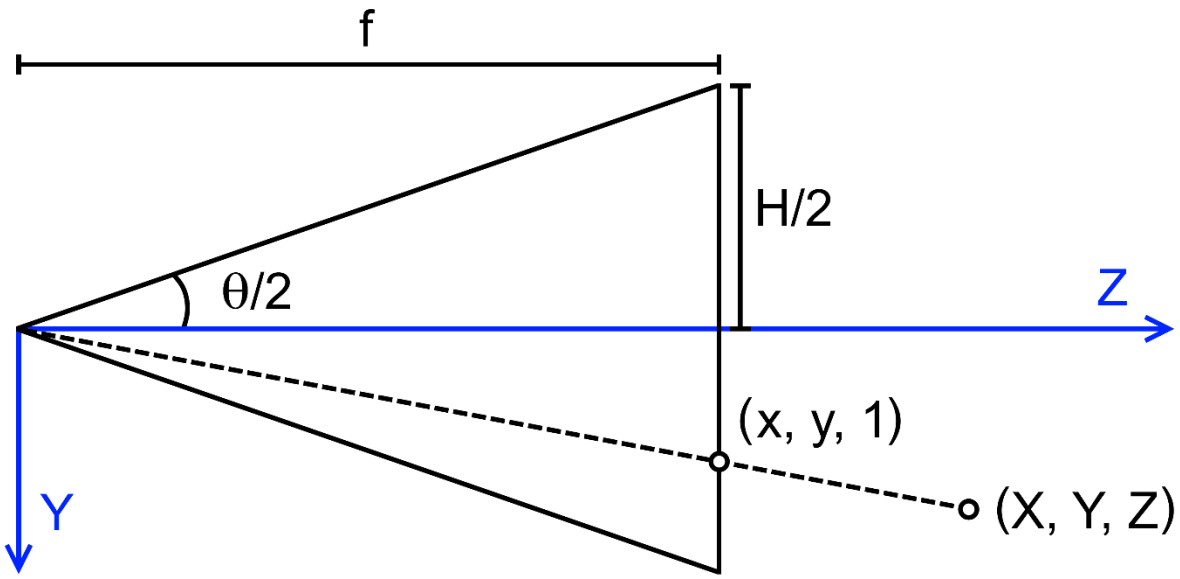

Figure 3: Schematic of perspective projection. In this side view, the black triangle represents camera's field of view and the image plane, and the blue lines are the +Y and +Z axes. The +X axis points out of the page. The camera-centered point $(X,\ Y,\ Z)$ is projected to its position in the image plane $(x,\ y,\ 1)$ by division by the Z component. f and θ are the camera's focal length and angle of view, and H is the height of the image.

Table 2: Summary of the intrinsic camera parameters.

Symbol	**Name**	**Comment**
c_1, c_2, c_3	Radial distortion coefficients	See Equation 3
c_4, c_5	Tangential distortion coefficients	See Equation 3
f_x, f_y	Focal lengths	See Equation 3
α	Skew coefficient	See Equations 4 and 5
p_x, p_y	Principal point	See Equations 4 and 5

3.4.2 Methods

Camera calibration was performed by acquiring a set of endoscopic video frames viewing a planar checkerboard pattern, identifying the locations of the checkerboard corners in each frame, and determining the intrinsic parameters that best model the projection of these corners to the frames. Corner detection and parameter computation were accomplished using the Camera Calibration Toolbox for MATLAB programming environment (The MathWorks, Inc., Natick, MA). This toolbox, which can be found at http://www.vision.caltech.edu/bouguetj/calib_doc/, is a freely-available, user-friendly implementation of the calib3d module of OpenCV, an open-source computer vision software package. The computational details of this module are documented elsewhere[63], so they will not be discussed here. The intrinsic camera model presented in Section 3.3.1 is very general, and the full set of parameters is not always necessary to accurately characterize a camera. For this reason, a series of calibrations was performed to eliminate superfluous parameters, and determine those that are necessary to remove distortion from the video frames and to match the virtual camera to the endoscope's camera.

The calibration rig, shown in Figure 4, was created by printing a grid of 5-mm squares in a checkerboard pattern and fixing it to a piece of acrylic to minimize any physical distortions. The square dimensions were verified manually after printing. To ensure that the endoscope's entire field of view was characterized, the rig was created with a large grid that extended beyond the field of view. An endoscopic video of the rig

was recorded, and 20 frames from a variety of orientations and distances were manually selected.

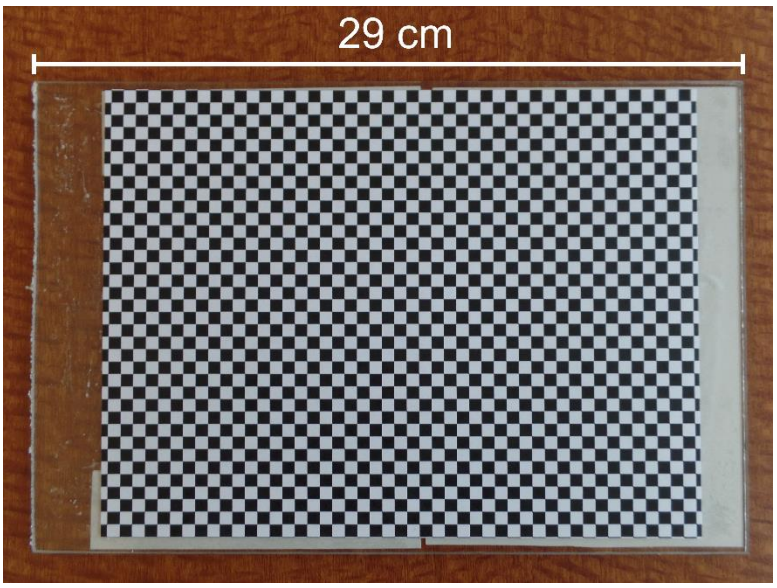

Figure 4: The calibration rig, which consisted of a 5-mm checkerboard grid fixed to a piece of acrylic.

3.4.3 Results

An initial calibration was run optimizing all of the intrinsic parameters (see Table 2). In this run, the value of α was equal to 0 within its uncertainty. This justified removing α from the model and treating the axes of the image sensor as exactly

perpendicular. A second calibration was run with α fixed at 0, and the resulting values are given in Table 3. This model was used with the OpenCV function `undistort` to remove distortion from the calibration frames, and from all endoscopic video frames in the remainder of the work discussed in this dissertation. The effect of this distortion removal is illustrated in Figure 5. Note that the measured principal point $(p_x,\ p_y)$ was displaced from the center of the image $((width-1)/2\,,\ (height-1)/2) = (359.5,\ 242.5)$ by about 5 pixels in both directions. For a perfect pinhole camera, such as the virtual camera, the principal point is located exactly at the center. To account for this, the principal point was shifted to the image center using the `newCameraMatrix` argument of `undistort`.

A third calibration was run after distortion was removed from the calibration frames. The resulting values of the distortion coefficients c_{1-5} were all equal to 0 and the coordinates of the principal point were equal to $(359.5,\ 242.5)$ within their respective uncertainties. This justified removing these parameters from the model. A fourth calibration was run with c_{1-5} fixed at 0 and $(p_x,\ p_y)$ fixed at $(359.5,\ 242.5)$, and the resulting values are given in Table 4.

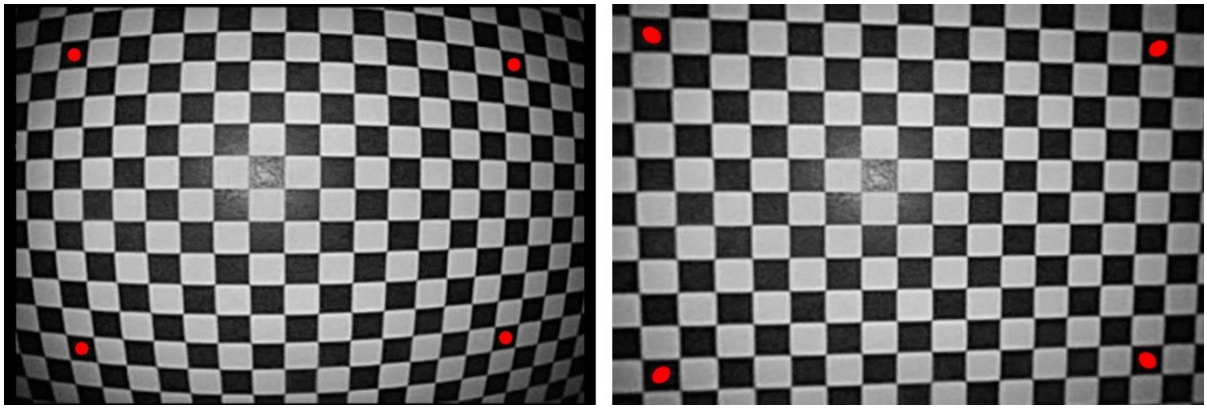

Figure 5: An example of distortion removal from an endoscopic video frame. Left: One of the frames used for calibration. The red dots were added digitally to provide reference points. The black borders were present in all videos produced by the recording device. Right: The same frame with distortion removed using the intrinsic model in Table 3.

Table 3: Optimized intrinsic parameters without skew. This model was used to remove distortion from endoscopic video frames. c_{1-5} and α are unitless. f_x, f_y, p_x, and p_y are expressed in pixels.

Parameter	**Value**
c_1	-0.3858 ± 0.0071
c_2	0.212 ± 0.024
c_3	-0.00115 ± 0.00030
c_4	0.00083 ± 0.00022
c_5	-0.075 ± 0.023
f_x	575.1 ± 1.0
f_y	526.0 ± 0.9
α	0
p_x	353.9 ± 1.0
p_y	238.4 ± 0.8

Table 4: Optimized intrinsic parameters for distortion-free frames. This model was used to determine the focal lengths of the endoscopic camera. c_{1-5} and α are unitless. f_x, f_y, p_x, and p_y are expressed in pixels.

Parameter	**Value**
c_1	0
c_2	0
c_3	0
c_4	0
c_5	0
f_x	575.2 ± 1.0
f_y	526.1 ± 0.9
α	0
p_x	359.5
p_y	242.5

This calibration indicated that the horizontal focal length f_x was about 9% larger than the vertical focal length f_y. There was nothing inherently wrong with this, but it did pose a problem for determining the focal length to use for the virtual camera. In the VTK rendering pipeline, the focal length is not set directly. Instead, the view angle θ of the virtual camera is set by the user. These two properties are mathematically related, as shown in Figure 3:

$$\tan\left(\frac{\theta}{2}\right) = \frac{H}{2f} \tag{9}$$

where H is the height of the image. VTK allows for vertical or horizontal view angles to be specified, but not both independently, so virtual images cannot be rendered with $f_x \neq f_y$. There are a few scenarios that can cause this inequality:

1. Asymmetrical optics in the camera
2. Different pixel dimensions or spacing along the axes of the image sensor
3. Digital modification of the image dimensions when the video file is encoded
4. Some combination of these three scenarios

It is not possible to determine the cause of the inequality from image measurements, but it can be accounted for in any case.

The horizontal and vertical focal lengths of the distortion-free calibration frames were made equal by decreasing their size while maintaining the same view angle. This was accomplished by downsampling them in the horizontal direction from 720 pixels to

$$\frac{f_y}{f_x} \cdot 720 = \frac{526.1}{575.2} \cdot 720 = 659\ pixels \tag{10}$$

According to Equation 9, the corresponding focal length, as measured in pixels, should have decreased by the same factor. To verify that this was the case, a fourth calibration was run on the distortion-free, downsampled calibration frames. As expected, the resulting focal lengths were equal within their uncertainties. A final calibration was run with f_x fixed to equal f_y and $(p_x,\ p_y)$ fixed at the downsampled image center

(329, 242.5), and the resulting values are given in Table 5. This calibration was used with Equation 9 to set the view angle of the virtual cameras used for the remainder of the work discussed in this dissertation:

$$\begin{aligned} virtual\ camera\ view\ angle &= 2 \cdot \arctan\left(\frac{H}{2f}\right) \\ &= 2 \cdot \arctan\left(\frac{486}{2 \cdot 526.5}\right) = 49.6\ degrees \end{aligned} \tag{11}$$

Table 5: Optimized intrinsic parameters for distortion-free, horizontally-downsampled frames. This model was used to determine the view angle for virtual endoscopic images. c_{1-5} and α are unitless. f_x, f_y, p_x, and p_y are expressed in pixels.

Parameter	**Value**
c_1	0
c_2	0
c_3	0
c_4	0
c_5	0
f_x	526.5 ± 1.0
f_y	526.5 ± 1.0
α	0
p_x	329
p_y	242.5

3.4.4 Summary

Camera calibration is an important foundational step for endoscopy-CT registration because it is used to remove distortion from endoscopic video frames and to match the focal length of the virtual endoscope to that of the real endoscope. Calibration was performed by recording an endoscopic video of a planar checkerboard pattern, selecting 20 frames from a variety of orientations and distances, automatically detecting the corners of the checkerboard, and calculating the intrinsic camera parameters that best model their projection onto the image. The outputs of camera calibration were the endoscope's focal length and principal point, as well as five coefficients that describe its radial and tangential distortion.

4

Image registration methods

Parts of this chapter are based on the following publication[64]:

W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

4.1 Introduction

The registration of endoscopic video to CT was carried out in three general steps:

1. Choose an endoscopic video frame to be registered. In a clinical application, this would be a frame containing a structure of interest such as a tumor or an area of mucositis.
2. Find the endoscope's CT-space coordinates in the selected frame. These coordinates consist of position and orientation.

3. Use the endoscope coordinates to establish spatial correspondence between the two modalities via virtual endoscopy.

The frame selected in step 1 will be referred to as the registration frame, and the coordinates found in step 2 will be referred to as the registered endoscope coordinates. In general, registered endoscope coordinates were found by maximizing the similarity between the registration frame and virtual endoscopic images as a function of the virtual endoscope's coordinates. More details on the calculation and maximization of similarity are given later in this section. Spatial correspondence was established by using the virtual endoscopic images to project pixels in the registration frame into the CT-space surface mesh. The computational methods used to do so are discussed in Section 4.4.

One of the most challenging aspects of endoscopy-CT registration is searching for the registered coordinates in a robust and efficient way. The prototypical method is to track the endoscope across the recorded video by updating the coordinates of a virtual endoscope frame-to-frame, either by maximizing image similarity between the frame and the virtual image or by estimating motion based on point correspondences in adjacent frames[21, 22]. This method has the advantage that the search space is quite small at each iteration, given that the endoscope does not travel very far between frames. However, it requires establishing an anchor point from which to start tracking, and it requires registration of frames prior to the desired registration frame. If the virtual endoscope becomes lost at any point in this process, the registration will fail without manual intervention.

The frame-to-frame tracking method was tested on phantom and patient images. Algorithm details for the frame-to-frame tracking method are provided in Section 4.2. It was tested on phantom and patient images. The methods of the phantom tests are presented in Section 5.2.4, and the results are presented in Sections 5.3.1.2 and 5.3.2.2. The methods and results of the patient tests are presented in Sections 6.2.5 and 6.3.1. Preliminary tests suggested that its robustness suffers when applied to patient videos, so a novel registration algorithm was devised that avoids the limitations of frame-to-frame tracking. This method searches the volume of the airways to directly find the registered coordinates for the desired frame, and it does so efficiently by utilizing physical constraints on the endoscope to reduce the size of the search space. This method will be referred to as the path-based volumetric search, and the algorithm details are provided in Section 4.3. It was also tested on phantom and patient images. The methods of the phantom tests are presented in Section 5.2.5, and the results are presented in Sections 5.3.1 and 5.3.2. The methods and results of the patient tests are presented in Sections 6.2.6 and 6.3.1.

Both methods rely on calculations of image similarity between endoscopic video frames and virtual endoscopic images. This similarity was calculated using a combination of mutual information and gradient alignment[65]. The virtual endoscope has a position and orientation in CT space, giving six degrees of freedom:

$$C = (x, \quad y, \quad z, \quad \theta_x, \quad \theta_y, \quad \theta_z) \tag{12}$$

Let F and $V(C)$ denote an endoscopic video frame and the virtual image rendered at the coordinates C. The similarity measure MI_{grad} is defined by

$$MI_{grad}\big(F, V(C)\big) \equiv MI\big(F, V(C)\big) \cdot GW\big(F, V(C)\big) \tag{13}$$

In this equation, $MI\big(F, V(C)\big)$ is the mutual information between the two images[66]. Normalized mutual information[67] is not used because the overlap between the video frames and virtual images never changes. $GW\big(F, V(C)\big)$ is a weighting term that favors alignment of edges in the two images. It is a sum over all pairs of corresponding pixels given by

$$GW\big(F, V(C)\big) = \sum_{(f,v)\in(F,V(C))} \frac{\cos\big(\phi_{f,v}\big) + 1}{2} \cdot \min(|\nabla F(f)|, |\nabla V(v)|) \tag{14}$$

In this equation, f and v denote pixels in the two images. To calculate this value, horizontal and vertical derivative images are created by convolving the images with Sobel filters[68]. These are used to compute the angle between the derivatives $\phi_{f,v}$ and the gradient magnitudes $|\nabla F(f)|$ and $|\nabla V(v)|$. The first term in the sum is maximized when the angle between the derivatives is zero, indicating that both images have an edge in the same direction. The second term ensures that only strong edges that are present in both images are favored. Using Equations 9 and 10, the registered coordinates for a registration frame F are defined by

$$C_{reg} \equiv \underset{C \in \mathbb{R}^6}{\operatorname{argmax}} \left(MI_{grad}\left(F, V(C)\right) \right) \tag{15}$$

MI_{grad} was selected for this application based on the widespread success of mutual information in medical image registration[69] and the observation that structural edges are the most salient features present in both real and virtual endoscopic images.

Both registration methods rely on the Nelder-Mead simplex optimization algorithm to maximize the similarity measure[70]. It is a minimization algorithm, so the negative of MI_{grad} was used as the objective function. A simplex is a geometric figure that has $n + 1$ vertices in n dimensions. For example, a triangle is a 2D simplex. In the Nelder-Mead method, the objective function value is calculated at the simplex vertices, which are updated by a series of reflections, expansions, and contractions to reduce the function value until convergence criteria are met. It was selected because it does not require calculation of the function's Jacobian or Hessian, which is not feasible in this application. It is a local optimization algorithm, and the scale of the search space is determined by a vector containing the distance along each coordinate axis that are used to create the initial simplex. This vector will be referred to as $\Delta_{simplex}$. The Nelder-Mead method is used differently in the two registration methods, which are described in Sections 4.2 and 4.3.

4.2 The prototypical method: frame-to-frame tracking

4.2.1 Algorithm description

In this method, the virtual endoscope is repeatedly moved such that the virtual image matches the next frame in the video. It consists of the following steps:

1. Select a registration frame F_{reg}.
2. Select a starting frame F_0 and place the virtual endoscope at the corresponding coordinates C_0 by mathematically aligning point correspondences.
3. Get the next frame F_1 and use C_0 as the initial guess for the Nelder-Mead method to search for the coordinates C_1 that maximize the similarity between F_1 and the virtual endoscopic image $V(C)$.
4. Repeat step 3 for frame F_2 using C_1 as the initial guess, and then for frame F_3 using the resulting C_2 as the initial guess, and so on until F_{reg} is reached.

Frame-to-frame tracking is computationally straightforward, but the results at each frame depend on the results of the previous frame. Because the Nelder-Mead method is a local optimization algorithm, it can only find the nearest local optimum. This means that if the virtual endoscope gets off track at some point and becomes lost, the process will fail.

4.2.2 Manual determination of initial endoscope coordinates

Manual input is required for step 2 of frame-to-frame tracking. First, the virtual endoscope must be placed at some coordinates C_{guess} that are sufficiently close to the correct coordinates for F_0 that some of the same structures are visible in F_0 and $V(C_{guess})$. Then, F_0 and $V(C_{guess})$ are displayed side by side, and a set of corresponding locations are selected by the user. Let $(u_i,\ v_i)$ and $(u_i',\ v_i')$ denote the pixel addresses of the i[th] correspondence in F_0 and $V(C_{guess})$, respectively. The CT-space mesh coordinates $(X_i',\ Y_i',\ Z_i')$ for each pixel $(u_i',\ v_i')$ are computed using methods described in Section 4.4. The goal then becomes to find the coordinates C_0 that project each $(X_i',\ Y_i',\ Z_i')$ as close as possible to its expected location $(u_i,\ v_i)$ in F_0.

Recall that the projection of 3D points onto pixel addresses is performed using the camera matrix (Equations 7 and 8 in Section 3.4.1). Given a set of endoscope coordinates $C_{test} = (x_t,\quad y_t,\quad z_t,\quad \theta_{xt},\quad \theta_{yt},\quad \theta_{zt})$, the camera matrix $\boldsymbol{P_{test}} = \boldsymbol{K}\ [\ \boldsymbol{R_{test}} \mid \boldsymbol{t_{test}}\ ]$ is composed using

$$\boldsymbol{R_{test}} = \boldsymbol{R_x}(\theta_{xt})\boldsymbol{R_y}(\theta_{yt})\boldsymbol{R_z}(\theta_{zt}) \tag{16}$$

$$\boldsymbol{t_{test}} = -\boldsymbol{R}\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} \tag{17}$$

In Equation 16, $\boldsymbol{R_x}$, $\boldsymbol{R_y}$, and $\boldsymbol{R_z}$ are the standard coordinate axis rotation matrices. Equation 17 is necessary because the endoscope's position is specified in CT-space, but

the translation part of the camera matrix is the position of the origin in the camera's reference frame. The projection of the CT-space coordinates $(X_i',\ Y_i',\ Z_i')$ creates a third set of pixel addresses:

$$\boldsymbol{P_{test}} \begin{bmatrix} X_i' \\ Y_i' \\ Z_i' \\ 1 \end{bmatrix} = \begin{bmatrix} u_{ti}' \\ v_{ti}' \\ 1 \end{bmatrix} \tag{18}$$

The distances between $(u_i,\ v_i)$ and $(u_t',\ v_{ti}')$ create a set of residuals that are used as the objective of a least-squares minimization that provides the desired endoscope coordinates C_0. Each correspondence provides two residuals (one for u and one for v), so a minimum of three correspondences between F_0 and $V(C_{guess})$ must be selected to determine the six endoscope coordinates. The determination of endoscope coordinates in this manner will be referred to as resectioning.

4.3 The novel method: path-based volumetric search

4.3.1 Algorithm description

In any endoscopic video there are likely to be many scenarios that may cause frame-to-frame tracking to fail, including lighting changes from the endoscope's dynamic gain, transient muscle motion causing large structural changes, erratic camera motion, and blurry frames. But frame-to-frame tracking requires determination of the

coordinates of the endoscope for all frames from F_0 up to F_{reg}, even though only the coordinates C_{reg} are of interest, and the longer this sequence of frames, the greater the chance of encountering an impasse. A more robust method would be to search directly for the desired frame's coordinates without considering any frames before it, and that is the goal of the path-based volumetric search.

The approach for this method is motivated by the observation that at any given location in the airways of the head and neck, only a small subset of the endoscope's orientation space $(\theta_x,\ \theta_y,\ \theta_z)$ is realistically possible. For example, the endoscope will never be positioned near the epiglottis, but looking in the superior direction, and the roll angle of the camera is constrained by the operator's hand on the endoscope's control section. These physical constraints can be used to initialize the virtual endoscope's view direction close to the correct direction, and a large majority of the orientation space can be excluded from the search. The endoscope's position space $(x,\ y,\ z)$ must be searched as well. This is accomplished by generating a sparse set of seed points to perform a coarse search that places the virtual endoscope near the correct location. This result is refined with a local search to obtain the final coordinates C_{reg}. The path-based volumetric search algorithm consists of the following steps, which are also illustrated in Figure 6:

1. Select a registration frame F_{reg}
2. Create a possible path through the volume for the virtual endoscope.
    a. Manually select a small set of points covering the length that the endoscope can travel.

   b. Interpolate between these points at an interval δ.
   c. Assign view directions to each point such that the virtual endoscope looks at the next point in the path.
3. At each path point, create a set of seed points that samples the cross-sectional area of the surface mesh.
   a. Slice the surface mesh in the plane perpendicular to the virtual endoscope's view direction.
   b. Calculate the desired number of seed points η based on the area of the slice A_s.
   c. Use k-means clustering[71] to generate seed points within the slice.
   d. Assign to each seed point the same view direction as that of the path point from which the slice was created.
4. Perform the coarse search by starting from each seed point in each slice and searching for the virtual endoscope coordinates that maximize the similarity between F_{reg} and the virtual image $V(C)$. In this step, the virtual endoscope's position $(x,\ y,\ z)$ is fixed at each seed point, and the view direction $(\theta_x,\ \theta_y,\ \theta_z)$ is optimized.
5. Create a 3 x 3 x 3 grid of points with spacing γ centered on the best overall result from step 4. Assign each grid point the same view direction as this result.
6. Perform the fine search by starting from each grid point and searching for the virtual endoscope coordinates that maximize the similarity measure between F_{reg} and the virtual image $V(C)$. In this step, all six coordinates $(x,\ \ y,\ \ z,\ \ \theta_x,\ \ \theta_y,\ \ \theta_z)$ are optimized.

The path-based volumetric search is more complex than frame-to-frame tracking, particularly for the slicing of the mesh and clustering of points in step 3, which are described in greater detail in Section 4.3.2. This algorithm has the advantage that the results for a given frame do not depend on any frames before or after, and manual input is required only for the selection of the initial path points in step 2. This selection is a simple process because it does not matter if the selected path is the same as the actual path the endoscope takes in the recorded video. The path is only used to initialize the seed point view directions to be reasonably close to what can be expected at any given location in the anatomy. In general, the path was created by manually selecting ~10 points on a single CT slice, and more details are provided in Sections 5.2.5 and 6.2.6.

There are several variables in the algorithm that can be adjusted. The first is δ, the interpolation interval along the path in step 2b. If this value is too large, there may not be a seed point close enough to the correct location to produce a virtual image that is a good match for the registration frame. If it is too small, an excessive amount of computation time will be required. The determination of this value is discussed in Sections 5.2.5, 5.3.1.1, and 5.3.2.1. The second variable is η, the number of seed points in step 3b. This number must vary from slice to slice to avoid under-sampling of large slices and over-sampling of small ones. The considerations for the number of seed points are similar to those for the interpolation spacing, so it is reasonable to specify this number such that the distance between the seed points is comparable to the interpolation spacing. The calculation of the number of seed points is discussed in

Section 4.3.2. The third variable is γ, the spacing of the grid for the fine search in step 5. It is set to half the interpolation spacing. These variables are summarized in Table 6.

Table 6: Description of the variables in the path-based volumetric search algorithm.

Symbol	**Name**	**Comment**
δ	Path interpolation interval	See step 2b. Specified by user. Evaluation of different values discussed in Sections 5.2.5, 5.3.1.1, and 5.3.2.1.
η	Number of seed points in a slice	See step 3b. Calculation described in Section 4.3.2.
γ	Grid spacing for fine search	See step 5 and 6. Set to equal $\delta/2$.

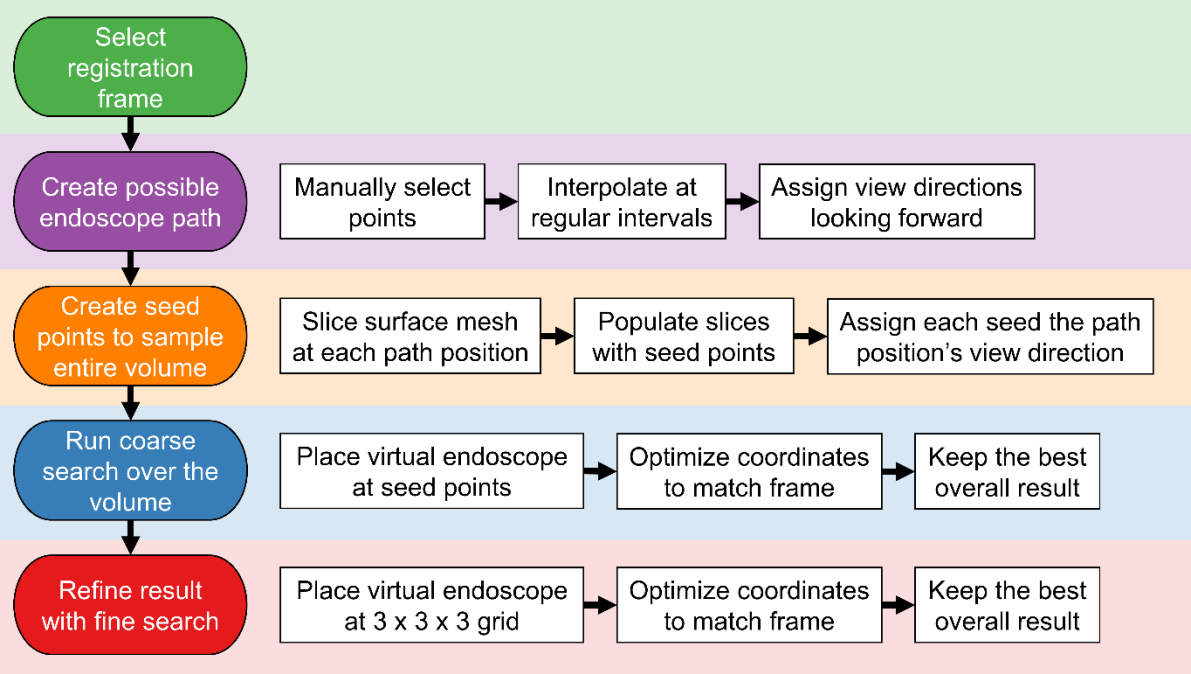

Figure 6: Flowchart showing the steps of the path-based volumetric search registration algorithm.

This figure has been reproduced and modified from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

4.3.2 Slicing the surface mesh and clustering seed points

Step 3 of the path-based volumetric search involves slicing the surface mesh to create a cross-section perpendicular to the virtual endoscope's view direction. This is accomplished using VTK methods to render an image of the cross section, and an example is shown in Figure 7. A `vtkCutter` object is created with its input set as the surface mesh and its cut function is set as a `vtkPlane` defined with its origin at the

virtual camera's position and its normal along the virtual camera's direction of projection. The output of the `vtkCutter` is a list of CT-space points defining the intersection between the plane and the surface mesh. These points are transformed to camera-centered coordinates using Equation 6, and their X and Y extents are used to calculate the distance behind the virtual endoscope from which the image must be rendered in order to contain the entire slice. The rendered image is an outline of the slice, which is flood-filled from the virtual endoscope's position. Any non-filled regions are discarded to eliminate unnecessary seed points in unconnected regions, such as the contralateral nasal cavity.

Next, the pixel addresses in the filled slice are treated as a set of individual observations and partitioned with k-means clustering. The number of clusters η is calculated based on the area of the filled slice A_s. The distance from which the slice image was rendered and the intrinsic parameters of the virtual camera are known, so it is possible to express this area in cm^2. The end goal is to create the seed points such that their spacing is approximately equal to the path interpolation interval δ in step 2b. It was observed that k-means clustering tends to produce patterns similar to a hexagonal lattice when applied to large, uniform areas, so η is calculated as the area of the filled slice divided by the area of a regular hexagon whose size is such that the spacing between hexagon centers in a lattice is equal to δ:

$$\eta = \frac{slice\ area}{hexagon\ area} = \frac{2A_s}{\sqrt{3}\delta^2} \tag{19}$$

This is illustrated in Figure 8. After running k-means clustering with η clusters, the centroid of each cluster is converted to a 3D point in CT-space using the known distance from which the slice image was rendered and the intrinsic parameters of the virtual camera. This produces the seed points for the slice, and each is assigned the same view direction as the path point from which the slice was generated. This is repeated for each path point, producing the desired set of seed points that samples the entire volume.

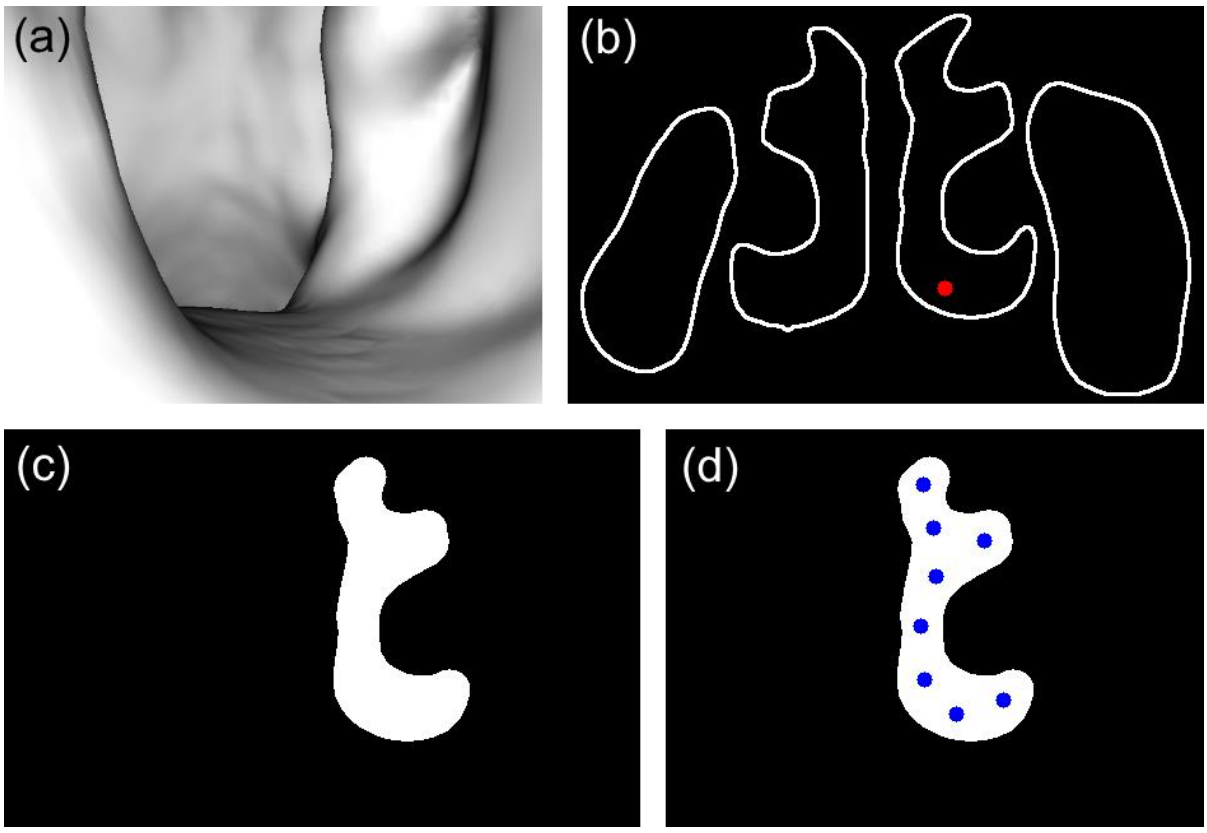

Figure 7: Example of surface slicing and seed point clustering in the path-based volumetric search registration algorithm. (a) A virtual endoscopic image in the left posterior nasal cavity. The medial wall and floor of the nasal cavity are visible on the left and bottom of the image, and the camera is looking towards to posterior wall of the pharynx. (b) The slice outline created in step 3a. The both sides of the nasal cavity and the maxillary sinuses are included. The virtual endoscope's position is shown as a red dot. (c) The outline has been flood-filled from the virtual endoscope's position and unconnected regions have been discarded. (d) The centroids of the k-means clusters are shown as blue dots. The number of clusters η was calculated using Equation 19 with δ = 5 mm. The actual distances between adjacent centroids were 4.3-5.0 mm. This image was cropped slightly to fit the template.

This figure has been reproduced and modified from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

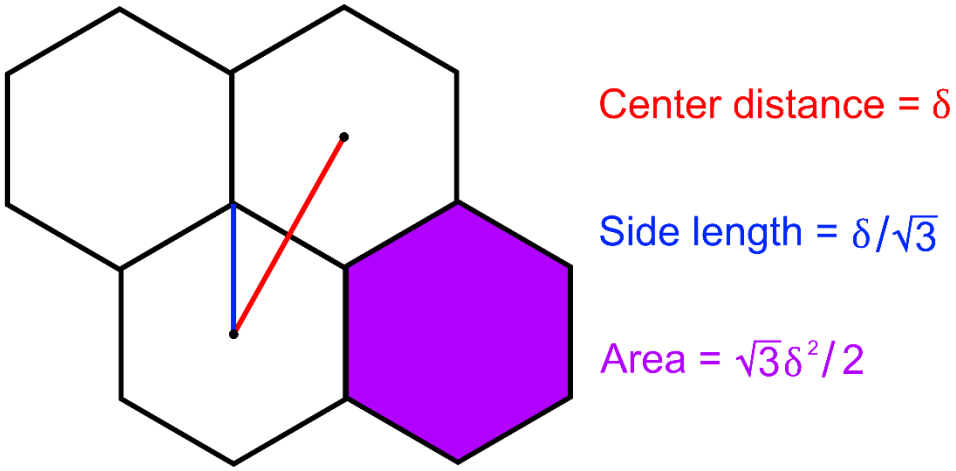

Figure 8: Illustration of the hexagon lattice used to calculate the number of seed points in Equation 19. The desired seed point distance is equal to the path interpolation interval δ. This gives a side length of $\delta/\sqrt{3}$ and area of $\sqrt{3}\delta^2/2$.

4.4 Methods for projective measurements

4.4.1 General concepts

The two registration methods, frame-to-frame tracking and path-based volumetric search, provide the registered endoscope coordinates C_{reg} that are used to place the virtual endoscope at the correct position and orientation for a given video frame. These coordinates are sufficient if the goal is to localize the endoscope for guidance during a procedure, or to transfer information such as a radiation dose distribution from CT to endoscopic video. However, establishing spatial correspondence from endoscopic video to CT requires some additional steps. With virtual endoscopy, depth information is still

available because the surface mesh provides a 3D model of the anatomy, and measuring CT-space coordinates from video-frame pixel addresses can be thought of conceptually as projecting a ray from the virtual camera's position through the pixel in the image plane and returning the coordinates of the point where it intersects the surface mesh. These 2D-to-3D measurements will be referred to as projective measurements.

Projective measurements can be taken simply using the VTK classes `vtkPointPicker` and `vtkCellPicker`. Both have methods to take a pixel address, project a ray through it into the scene, and return information about what it hits. `vtkPointPicker` can only return the nearest vertex in the mesh. `vtkCellPicker` calculates the exact intersection point, and it can also return the vector normal to the surface at the intersection point, which may be useful for identifying high-uncertainty regions (see Section 4.4.3). However, it takes ~30 ms to do this operation. This does not sound long, but even for a region as small as a circle with a 50-pixel radius it would take nearly 4 minutes to project every pixel. Though it is likely that a sparse set of pixels within this region would be sufficient for most applications, it is more convenient to have a faster method to compute the 3D coordinates for every pixel in the virtual image and downsample the data as needed. A method to accomplish this, and two derivatives that are useful for characterizing projective measurements, are presented in Sections 4.4.2-4.4.4.

4.4.2 Fast projective measurements via the world transform

The VTK rendering pipeline makes use of the Z-buffer, which is an array that stores information about the depth of objects in the scene relative to the camera. The Z-buffer may be implemented in hardware or software, and it is used to determine which objects are visible in a rendered scene. If the depth of the 3D point corresponding to a pixel address is known, the full projective measurement for that pixel can be calculated by multiplying the pixel address by the inverse of the camera matrix in Equation 8 and scaling the result by the known depth.

An image of the Z-buffer can be created using the class `vtkWindowToImageFilter` with the Z-buffer set as the input. The values in the depth buffer image are in the range [0, 1]. They are normalized to the range between the near and far clipping planes, which is an attribute of the `vtkCamera` that determines the range of depth values included in the rendered images. There are two important considerations for using the Z-buffer image to calculate 3D coordinates. The first is that the Z-buffer is not linear. The conversion from Z-buffer value Z_b to depth Z is given by

$$Z = \frac{2Z_{far}Z_{near}}{Z_{far} + Z_{near} - \left(Z_{far} - Z_{near}\right)(2Z_b - 1)} \tag{20}$$

Where Z_{far} and Z_{near} are the values of the clipping range. Note that the output Z is the Z-component of the camera-centered coordinates described in Section 3.4.1 and Equation 2. The second consideration for using the Z-buffer image to calculate 3D

coordinates is that the Z-buffer has a finite precision, and due to the non-linearity of Equation 20, the accuracy of calculated 3D coordinates will suffer if the clipping range is too large. This is especially true if Z_{near} is close to zero.

Taking these into account, the following procedure was used to calculate the CT-space coordinates corresponding to each pixel in a virtual image.

1. Place the virtual endoscope at the desired coordinates.
2. Build the depth image by extracting the Z-buffer image in 1-cm strips to preserve resolution.
    a. Set the clipping range to [0.1, 1] extract the Z-buffer image, and convert to camera-centered depth using Equation 20.
    b. Set the clipping range to [1, 2] and repeat. Pixels already assigned to the depth image in the previous iteration are given priority to avoid including regions of the outside of the surface that are made visible as Z_{near} is moved away from the camera.
    c. Repeat this process with the clipping range set to [2, 3], [3, 4], and so on until a desired maximum depth is reached.
3. Create a list of all pixel addresses $(u, v, 1)$ in the virtual image and convert to image plane coordinates by multiplying them by the inverse of the virtual endoscope's calibration matrix (see Equations 4 and 5 and Table 5).
4. Convert the image-plane coordinates to camera-centered coordinates by scaling them by their corresponding values in the depth image.

5. Convert the camera-centered coordinates to CT-space coordinates by multiplying them by the inverse of the virtual endoscope's camera matrix (see Equations 7 and 8).

The calculation CT-space coordinates for each pixel in the virtual image, and the resulting m x n x 3 coordinate array, will be referred to as the world transform of the virtual image. The calculations in steps 2d-2f are vectorized, so the world transform is quite fast. The speed depends on the size of the surface mesh, and the maximum depth used in step 2c, but the computation time is generally less than 5 seconds. The same computation would take over 2.5 hours using the `vtkCellPicker` method. The world transform is accurate as well, differing from `vtkCellPicker` by only by 0.13 ± 0.82 mm. The large standard deviation is due to the presence of a small number of outliers where the two methods disagree. This can happen at occluding edges and in distant regions of the image, but in those areas, a shift of a few pixels results in a large change in the projected point anyway. Excluding just the largest 1% of errors from the comparison brings the difference down to 0.08 ± 0.09 mm.

4.4.3 Computation of measurement angle from the world transform

Assuming that there is an inherent uncertainty in the solid angle through which a pixel is projected, it is reasonable to expect that the apparent measurement uncertainty will be affected by scene geometry, including the distance between the endoscope's camera and the measured position and the angle at which the camera views the surface.

This is illustrated in Figure 9. The camera-to-surface distance is easy to calculate from the world transform using the virtual endoscope's coordinates, but calculating the angle at which the camera views the surface is more complication.

Throughout this dissertation, the term measurement angle will be used to refer to the angle between a vector connecting a point on the surface mesh to the virtual endoscope and the vector normal to the surface mesh at that point. The measurement angle is calculated by fitting a 3D plane to each point in the world transform using its surrounding 8-neighbors. The equation for a plane with the normal vector $(a,\ b,\ c)$ containing the point $(X_0,\ Y_0,\ Z_0)$ can be written as

$$a(X - X_0) + b(Y - Y_0) + c(Z - Z_0) = 0 \tag{21}$$

By designating a point in the world transform as $(X_0,\ Y_0,\ Z_0)$, its surrounding 8-neighbors $(X_{ni},\ Y_{ni},\ Z_{ni})$ can be used to form an over-determined system of equations:

$$\begin{bmatrix} X_{n1} - X_0 & Y_{n1} - Y_0 & Z_{n1} - Z_0 \\ \vdots & \vdots & \vdots \\ X_{n8} - X_0 & Y_{n8} - Y_0 & Z_{n8} - Z_0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{22}$$

The non-trivial solution for $(a,\ b,\ c)$ is found by taking the singular value decomposition of the matrix[72]. The last right-singular vector of the decomposition is the desired surface normal. Finally, the measurement angle is calculated from the dot

product of the surface normal and the vector connecting the surface point to the virtual endoscope.

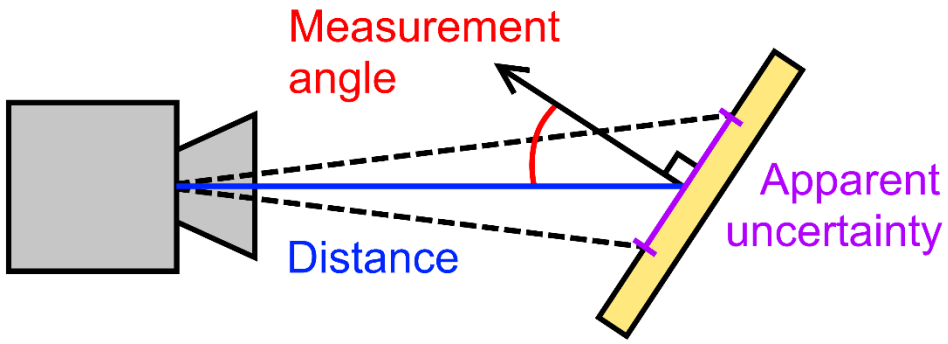

Figure 9: Illustration of the expected impact of scene geometry on projective uncertainty. The dashed lines represent the inherent uncertainty in the solid angle through which a pixel is projected. The apparent measurement uncertainty is increased with the camera-to-surface distance or the measurement angle increase.

This figure has been reproduced and modified from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

4.4.4 Computation of the edge mask from the world transform

In a given virtual endoscopic image, there are likely to be occluding edges and regions with measurement angles close to 90 degrees. In these areas, very large projective measurement errors are likely, because a shift of only a few pixels will result in a large change in the projected coordinates. It may be desirable to avoid making

projective measurements in these areas, and the world transform provides a convenient method of identifying them. For each point in the world transform, the distances to each of its 8-neighbors are calculated. If at least one of these distances is larger than a threshold, the point is marked as an edge point. This process creates a binary array showing edge regions in the virtual image to exclude from projective measurements. This array it will be referred to as the edge mask, and an example is shown in Figure 10.

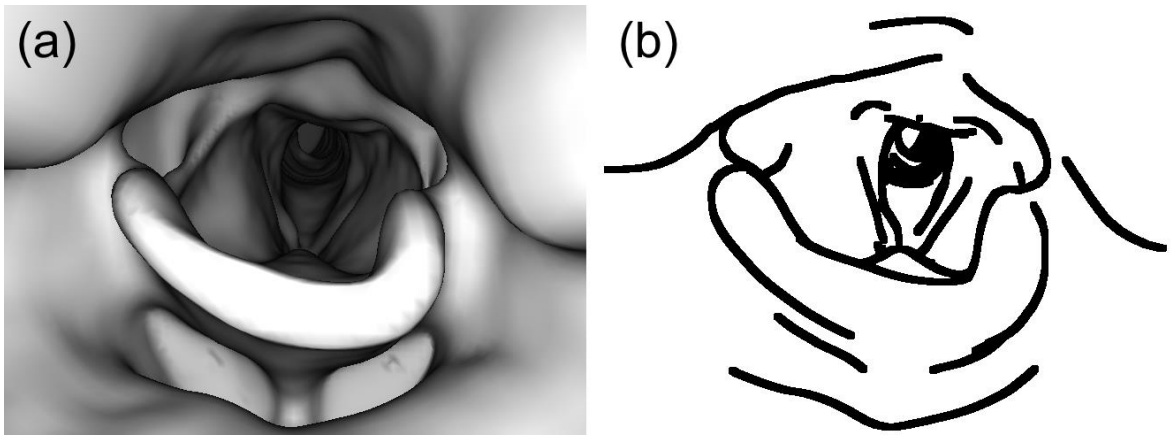

Figure 10: Example of the edge mask created from the world transform. (a) A virtual endoscopic image with the epiglottis in the foreground and the glottis in the background. (b) The edge mask for this image. It was created with a 2-mm threshold, and morphological erosion was applied to provide a buffer around the edge points.

5

Image registration in phantoms

Parts of this chapter are based on the following publication[64]:

W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

5.1 Introduction

The previous chapter described the methods developed to register endoscopic video to CT, and this chapter presents the testing of those methods in rigid phantoms. The creation of these phantoms, the acquisition of endoscopic video and virtual endoscopic images within them, and the details of the registration tests are discussed in Section 5.2. The results of the registration tests are presented in Section 5.3, and a discussion is given is Section 5.4.

5.2 Methods

5.2.1 Phantom design

Two clay phantoms were created to assess the registration methods described in Sections 4.2 and 4.3. Clay was used so that the phantoms would have irregular shapes and low-contrast surface textures. Both phantoms include fiducials that allowed for projective measurements to be compared to a ground truth. The first phantom (Figure 11) contains twelve 2-mm-diameter radiopaque markers embedded in the luminal surface, arranged in three rings along the phantom's length. These markers were used to make point measurements of registration accuracy. The second phantom (Figure 12) contains a 10 x 10 x 5 mm^3 piece of Superflab bolus material (Mick Radio-Nuclear Instruments, Mount Vernon, NY). The bolus protrudes into the lumen of the phantom. It was used to test registration accuracy by mapping an object contour from endoscopic video to CT.

There are a couple of notable differences in the design of the two phantoms. The first is that the internal dimensions of the marker phantom are larger than those of the bolus phantom. The internal diameter of the bolus phantom is ~5 cm, whereas that of bolus phantom is ~2 cm. The second is that the luminal surface of the marker phantom is more irregular. The internal dimensions of the bolus phantom are a better representation of what can be seen in patients via virtual endoscopy, and its smoother surfaces, which lack characteristic edges, provide a more challenging test for endoscopy-CT registration.

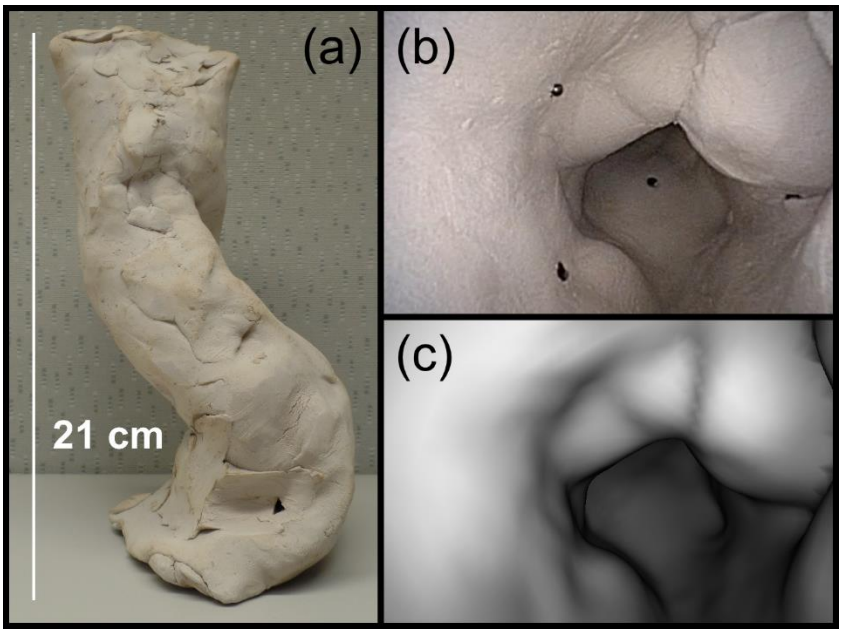

Figure 11: The marker phantom. (a) A photograph of the marker phantom showing its overall dimensions. The opening at the top for the endoscope is not visible. (b) An endoscopic video frame from inside the phantom, positioned near the top looking down. Four of the 2-mm radiopaque markers are visible. (c) The corresponding virtual endoscopic image.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

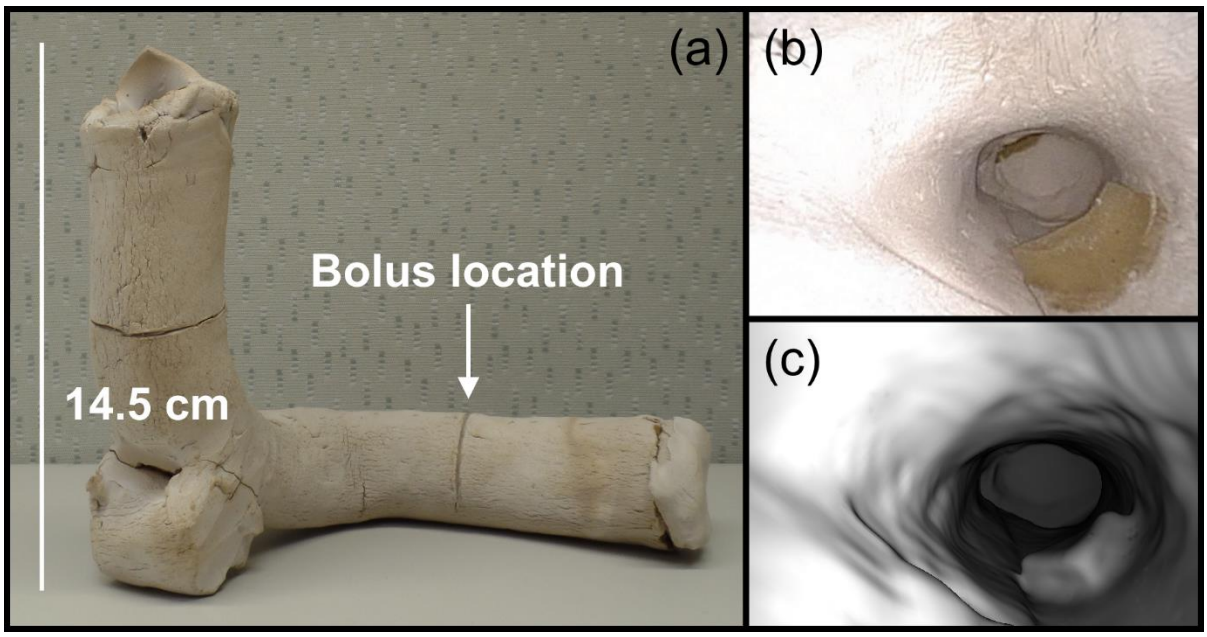

Figure 12: The bolus phantom. (a) A photograph of the bolus phantom showing its overall dimensions and the approximate location of the bolus material. (b) An endoscopic video frame from inside the phantom, positioned directly in front of the bolus, which is visible on the lower right. (c) The corresponding virtual endoscopic image.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

5.2.2 CT acquisition and virtual endoscopy

CT scans of the phantoms were acquired using a Lightspeed RT (GE Healthcare) with a 30-cm field of view and 1-mm slices. The luminal surfaces were segmented using the Pinnacle^3 treatment planning software (Philips Healthcare). The segmentation was performed semi-automatically by selecting a density threshold of 0.9 g/cm^3 and making a single mouse click inside the lumen on each axial slice. This threshold was chosen

because the density of the clay is ~1.8 g/cm^3, 0.9 g/cm^3 should draw the boundary at voxels that contain half air and half clay. The .roi files containing the segmented contours were used to create virtual endoscopy surface meshes as described in Section 3.3.1. The virtual endoscopy lighting parameters (Section 3.3.2 and Table 1) were chosen by visual inspection to match the overall appearance of recorded videos. The constant, linear, and quadratic attenuation values were set to 1, 1.4, and 2, respectively. The cosine exponent was set to 5. In the marker phantom, the intensity was set to 30. In the bolus phantom, the smaller dimensions made this setting too bright, so the intensity was reduced to 15. All virtual endoscopic images were smoothed with a 3 x 3 Gaussian kernel with σ = 1 pixel.

5.2.3 Endoscopic video datasets

Two endoscopic video sequences were recorded in each phantom with the endoscope moving in through the length of the phantom and back out. The lengths of the video sequences were 27 and 17 seconds in the marker phantom and 53 and 75 seconds in the bolus phantom, with the bolus visible for the last 42 and 64 seconds, respectively. The bolus phantom videos were longer due to the increased difficulty of navigating the endoscope through the smaller space.

A set of registration frames was selected for each phantom by sampling the videos at regular intervals and identifying the least blurry frame out of the sample as well as the five previous and subsequent frames. The sampling intervals were 1 and 2 seconds for the marker and bolus phantoms, respectively. The least blurry frame was identified

as the one with the largest variance after filtering with a 3 x 3 Laplacian. The endoscopic videos contained many frames that were unsuitable for registration, either due to the markers or bolus not being visible or due to under- or overexposure as the image sensor adjusted its gain. To avoid these scenarios, each set of frames was reviewed and unsuitable frames were rejected. This resulted in a total of 36 and 37 registration frames for the marker and bolus phantoms, respectively. The preprocessing for all registration frames included deinterlacing by replacing every other row with bilinear interpolation, distortion removal as described in Section 3.4, conversion to grayscale, and smoothing with a 3 x 3 Gaussian kernel with σ = 1 pixel.

The ground-truth CT-space marker positions were obtained manually. For each marker, the CT slice on which it appeared brightest was selected, and the coordinates of the voxel closest to the center of the lumen were recorded. The ground-truth bolus contour was created semi-manually. First, the luminal contours on all slices containing the bolus were copied. Then, the underside of the bolus was contoured manually, and all extraneous parts of the copied contour were removed. This process ensured that the luminal voxels of the ground truth contours, which were the only ones visible for measurement via virtual endoscopy, were identical to those from which the virtual endoscopy surface mesh was created.

5.2.4 Frame-to-frame tracking

Frame-to-frame tracking was performed as described in Section 4.2.1. Initial frames to start the tracking were selected near the entrances of the phantoms prior to the first

registration frame for each sequence, and the initial virtual endoscope coordinates were determined using the resectioning process described in Section 4.2.2. At each subsequent frame, the virtual endoscope's coordinates ($x, \quad y, \quad z, \quad \theta_x, \quad \theta_y, \quad \theta_z$) were optimized to match the next frame in the video sequence using the simplex method described in Section 4.1. The scale of the search space was set using

$$\Delta_{simplex} = (2\ mm, \quad 2\ mm, \quad 2\ mm, \quad 5\ deg, \quad 5\ deg, \quad 5\ deg) \tag{23}$$

5.2.5 Path-based volumetric search

The path-based volumetric search was performed as described in Section 4.3.1. In the marker phantom, the possible virtual endoscope path was created by manually selecting seven points on a coronal slice of the CT. This is illustrated in Figure 13. The bolus phantom does not have bilateral symmetry, so instead of manual selection, the path was created by calculating the centroid of the segmented surface for every fifth CT slice. To evaluate the impact of seed point spacing on registration accuracy, the search was performed using four sets of seed points that were created using path interpolation intervals of δ = 2.5, 5.0, 7.5, and 10.0 mm (see Section 4.3 for details). The corresponding numbers of seed points for each slice were calculated using Equation 19.

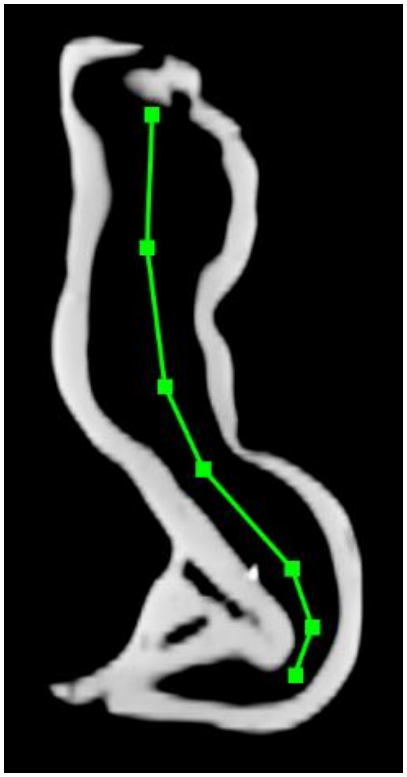

Figure 13: Virtual endoscope path creation in the marker phantom. A coronal CT slice shows the manually-selected points that comprise the virtual endoscope path used for the path-based volumetric search registration algorithm.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

At each seed point, the virtual endoscope's view direction $(\theta_x,\ \theta_y,\ \theta_z)$ was optimized to match the registration frame using the simplex method described in Section 4.1. The scale of the search space was set using

$$\Delta_{simplex} = (20\ deg,\quad 20\ deg,\quad 20\ deg) \tag{24}$$

This is larger than the 5-degree scale used for frame-to-frame tracking. A smaller value is appropriate in that scenario because the endoscope does not move much between adjacent frames, so the starting point for the optimization can be assumed to be very close to the optimum.

The result of the coarse search was the seed point with optimized view direction that provided the virtual endoscopic image that was most similar to the registration frame. For the fine search, the 3 x 3 x 3 grid was created with its center on this optimized seed point and the spacing between points equal to half the path interpolation interval, $\delta/2$. At each grid point, the virtual endoscope's coordinates $(x, \quad y, \quad z, \quad \theta_x, \quad \theta_y, \quad \theta_z)$ were optimized to match the registration frame using the simplex method. The scale of the search space was set using

$$\Delta_{simplex} = \left(\frac{\delta}{4}, \quad \frac{\delta}{4}, \quad \frac{\delta}{4}, \quad 2.5\ deg, \quad 2.5\ deg, \quad 2.5\ deg\right) \tag{25}$$

The result of the fine search was the optimized grid point that provided the virtual endoscopic image that was most similar to the registration frame.

5.2.6 Measurements of registration accuracy

The outputs of frame-to-frame tracking and the path-based volumetric search are sets of registered endoscope coordinates C_{reg} corresponding to each registration frame F_{reg} in the datasets. These coordinates were used to render virtual endoscopic images and take projective measurements as described in Section 4.4. In the marker phantom, a single pixel address was manually selected for each marker visible in each registration frame. Registration accuracy was quantified by calculating the 3D distance errors between the projective measurements for these pixel addresses and the ground-truth CT-space marker positions. This is illustrated in Figure 14. One to six markers were visible in each frame, for a total of 128 measurements. To investigate the impact of scene geometry, the measurement angles at the ground-truth marker positions and the measurement distances between the registered endoscope coordinates and the ground-truth marker positions were calculated (see Figure 9). Measurement angles were averaged over a circular ROI. The pixel radius of this ROI was varied such that it would correspond to a 2-mm radius at the measurement distance in the center of the image.

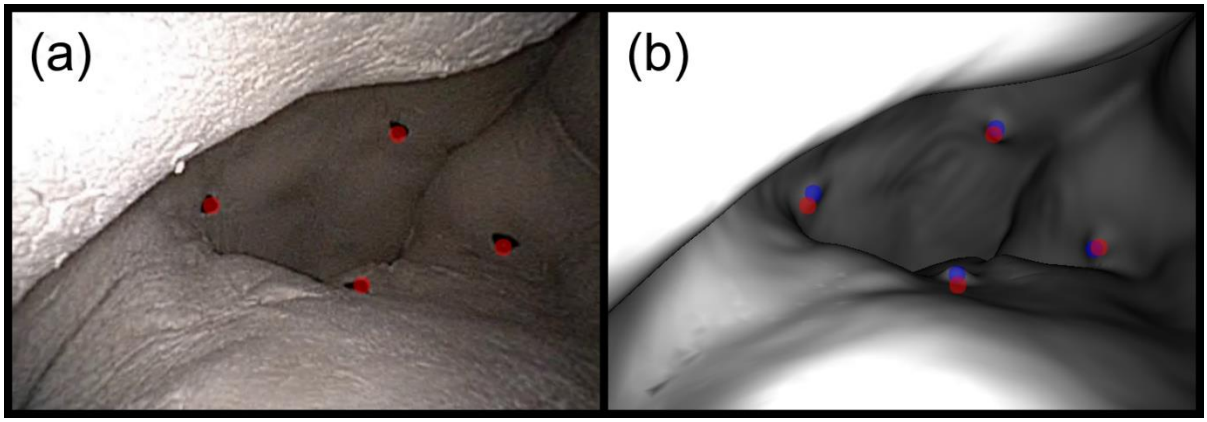

Figure 14: Projective measurement of CT-space marker positions. (a) A video frame with four markers visible. Their manually-selected pixel locations are overlaid in red. (b) The corresponding virtual endoscopic image rendered at the registered coordinates. The manually-selected pixel locations are overlaid in red, and the image projections of the ground-truth CT-space coordinates are overlaid in blue. The average error for these four measurements was 2.9 mm.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

In the bolus phantom, a video-frame contour was created by manually selecting a set of vertices outlining the bolus and filling the polygon to create a binary mask. The mask was used to sample the projective measurements in the world transform (see Section 4.4.2). This is illustrated in Figure 15. Registration accuracy was quantified by calculating the symmetric mean absolute surface distance (SMAD) between the projected mask and the luminal voxels of the ground-truth CT-space bolus contour:

$$SMAD = \frac{1}{(n_M + n_{CT})}\left(\sum_{i=1}^{n_M} d_i^{M\to CT} + \sum_{j=1}^{n_{CT}} d_j^{CT\to M}\right) \quad (26)$$

In this equation, n_M is the number of pixels in the mask and n_{CT} is the number of luminal voxels in the CT contour. The term $d_i^{M\to CT}$ is the minimum distance from the ith projected mask pixel to any luminal voxel, and $d_j^{CT\to M}$ is the minimum distance from the jth luminal voxel to any projected mask pixel. There were 612 luminal voxels in the CT contour, but the video-frame mask can contain tens of thousands of pixels. To prevent this disparity from skewing the calculation of SMAD, the mask and the world transform were downsampled using nearest-neighbor interpolation such that the number of projective measurements was approximately equal to 612. To investigate the impact of scene geometry, the measurement distances between the registered endoscope coordinates and the centroid of the ground-truth bolus contour were calculated. Measurement angles were not calculated because the larger size of the bolus compared to the markers, and the fact that it protrudes into the lumen, mean that there is no way to calculate a meaningful characteristic value.

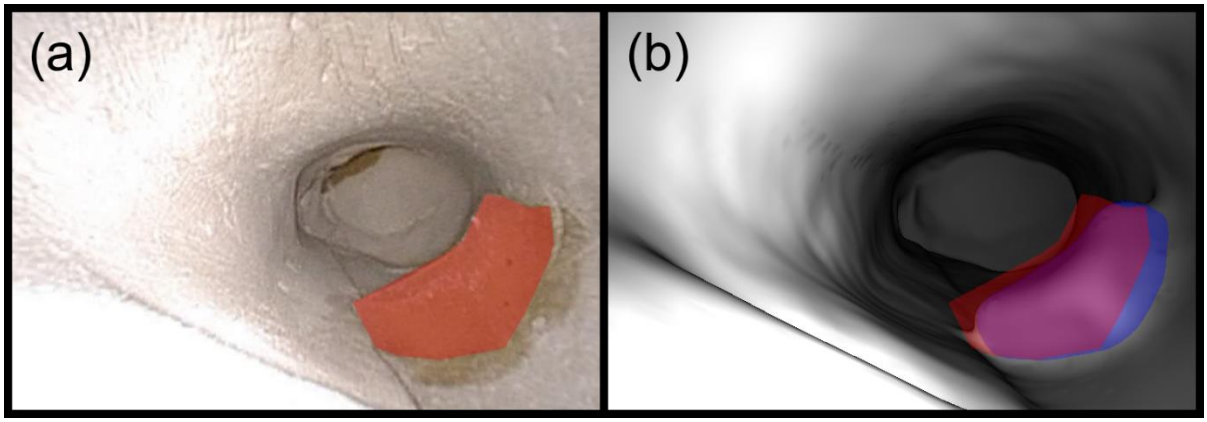

Figure 15: Projective measurement of CT-space bolus contour. (a) A video frame showing the bolus material with the manually-drawn contour mask overlaid in red. (b) The corresponding virtual endoscopic image rendered at the registered coordinates. The manually-drawn contour mask is overlaid in red, and the image projection of the CT-space bolus contour is overlaid in blue. The SMAD for this frame was 2.3 mm.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

5.3 Results

5.3.1 Marker phantom

5.3.1.1 Evaluation of seed point spacing for path-based volumetric search

The results of the path-based volumetric search using the four sets of seed points are summarized in Table 7. For three registration frames in sequence 2, the registered coordinates for all seed point sets were a very poor match to the registration frame,

such that none of the correct marker locations were visible anywhere in the virtual image. These frames, from which 7 out of the 128 marker position measurements were taken, were considered failures and excluded from quantitative analysis.

For the rest of the measurements, the median errors ranged from 3.0 to 3.5 mm, and no significant difference was found between the four sets ($p > 0.05$ using the Kruskal-Wallis H-test). Median errors were used for this comparison because the measurements were characterized by the presence of a small number of very large outliers. This is demonstrated by comparison of the 80^{th} percentile errors, which ranged from 6.1 to 6.4 mm, to the maximum errors, which ranged from 54.3 to 56.5 mm. The registration accuracy with each set of seed points was very similar for every frame. This is illustrated by a plot of the average measurement errors for each frame in the two video sequences, shown in Figure 16. The path-based volumetric search performed using the seed point set created with δ = 7.5 mm resulted in the smallest median point measurement error, so it was selected for comparison with the frame-to-frame tracking results in Section 5.3.1.2.

Table 7: Comparison of seed point spacing in the marker phantom for the path-based volumetric search. The first column gives the specified path interpolation interval used to create the set of seed points (see Table 6 and Equation 19). The second column gives the actual average distance between adjacent points in the set. The third column gives the median point measurement error after running the search using each set of seed points. The fourth and fifth columns give the 80th percentile and maximum measurement errors. All values are in mm.

Specification	**Actual spacing**	**Median error**	**80th percentile**	**Maximum**
$\delta = 2.5$	2.2 ± 0.1	3.5	6.2	54.9
$\delta = 5.0$	4.5 ± 0.2	3.4	6.2	56.3
$\delta = 7.5$	6.8 ± 0.4	3.0	6.1	54.3
$\delta = 10.0$	9.0 ± 0.6	3.3	6.4	56.5

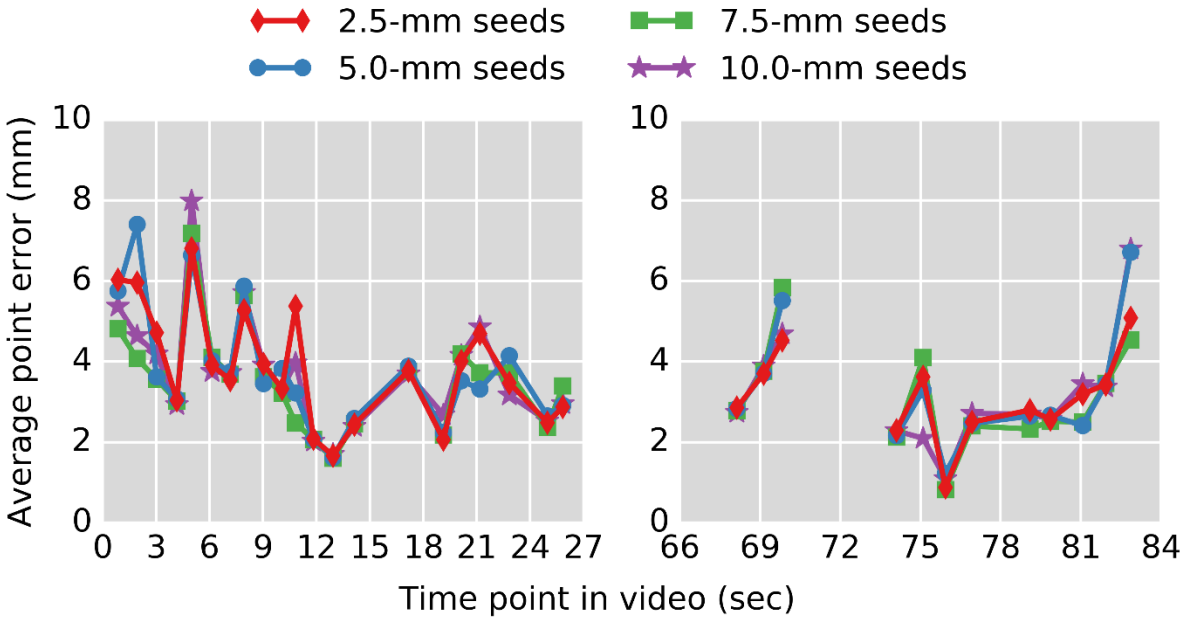

Figure 16: Comparison of seed point spacing in the marker phantom for the path based-volumetric search. These plots show the average point measurement error for each frame in video sequence 1 (left) and video sequence 2 (right) after running the search using the four sets of seed points. There is no observable trend between the sets, and there was no statistically significant difference between them either. Outliers larger than the 90th percentile were excluded from the averages. The gap in the plot for sequence 2 corresponds to the three failed frames.

5.3.1.2 Comparison of the two registration methods

The results of registration using the frame-to-frame tracking method and the path-based volumetric search method with 7.5-mm seed points are summarized in Table 8, and a plot of the average measurement errors for each frame in the two video sequences is shown in Figure 17. The two methods had very similar results, with median point measurement errors of 2.9 and 3.0 mm for frame-to-frame tracking and the path-based volumetric search, respectively. There was no significant difference between the two methods ($p > 0.05$ using the Wilcoxon signed-rank test). Both methods had a small number of very large outliers, and nearly all of these occurred for the same marker measurements in both cases. The most notable difference between the two is that frame-to-frame tracking successfully registered the three frames on which path-based volumetric search failed.

Table 8: Comparison of the two registration methods in the marker phantom. The most notable difference is that frame-to-fracking tracking successfully registered the three frames on which the path-based volumetric search failed.

Metric	**Frame-to-frame tracking**	**Path-based volumetric search**
# of failed frames	0	3
Median error	2.9	3.0
80^{th} percentile	5.6	6.1
Maximum	53.9	54.3

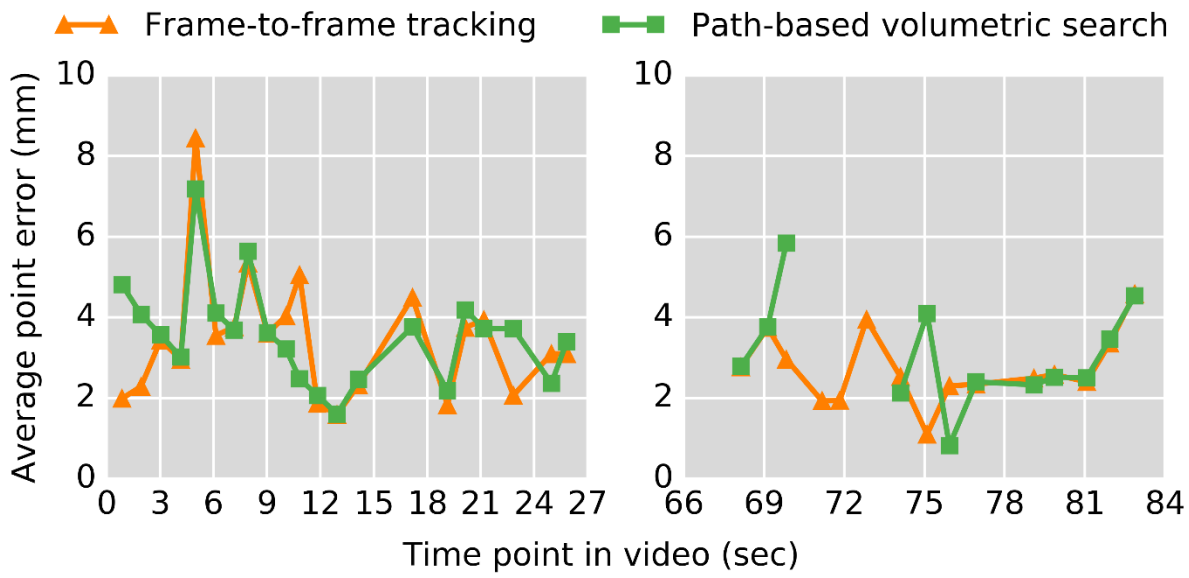

Figure 17: Comparison of the two registration methods in the marker phantom. These plots show the average point measurement error for each frame in video sequence 1 (left) and video sequence 2 (right) for frame-to-frame tracking and path-based volumetric search. The gap in the plot of path-based volumetric search in sequence 2 corresponds to the three failed frames, and the most notable difference between the two methods is that frame-to-frame tracking successfully registered these frames. There was no statistically significant difference between the two methods. Outliers larger than the 90th percentile were excluded from the averages.

5.3.2 Bolus phantom

5.3.2.1 Evaluation of seed point spacing for path-based volumetric search

The results of the path-based volumetric search using the four sets of seed points are summarized in Table 9. There were no failed frames in this phantom, meaning that the bolus was visible in all registered virtual endoscopic images. The median symmetric

mean absolute distance (SMAD) between the measured and ground-truth bolus contours ranged from 3.5 to 5.7 mm. As in the marker phantom, no significant difference was found between the four sets ($p > 0.05$ using the Kruskal-Wallis H-test). Unlike the marker phantom, the registration accuracy for each frame was quite variable between the four sets of seed points, with an average range between the largest and smallest SMADs of 3.6 ± 2.0 mm. This is illustrated by a plot of the SMAD for each frame in the two video sequences, shown in Figure 18. The path-based volumetric searches performed using the seed point sets created with $\delta = 5.0$ and $\delta = 10.0$ mm resulted in the smallest median SMADs, so $\delta = 5.0$ was selected for comparison with the frame-to-frame tracking results in Section 5.3.2.2.

Table 9: Comparison of seed point spacing in the bolus phantom for the path-based volumetric search. The first column gives the specified path interpolation interval used to create the set of seed points (see Table 6 and Equation 19). The second column gives the actual average distance between adjacent points in the set. The third column gives the SMAD between measured and ground-truth bolus contours error after running the search using each set of seed points. The fourth and fifth columns give the 80th percentile and maximum SMADs. All values are in mm.

Specification	**Actual spacing**	**Median SMAD**	**80th percentile**	**Maximum**
$\delta = 2.5$	2.3 ± 0.1	5.7	6.5	8.0
$\delta = 5.0$	4.5 ± 0.3	3.5	6.6	7.9
$\delta = 7.5$	6.7 ± 0.5	5.0	6.5	10.8
$\delta = 10.0$	8.7 ± 0.5	3.5	5.3	8.4

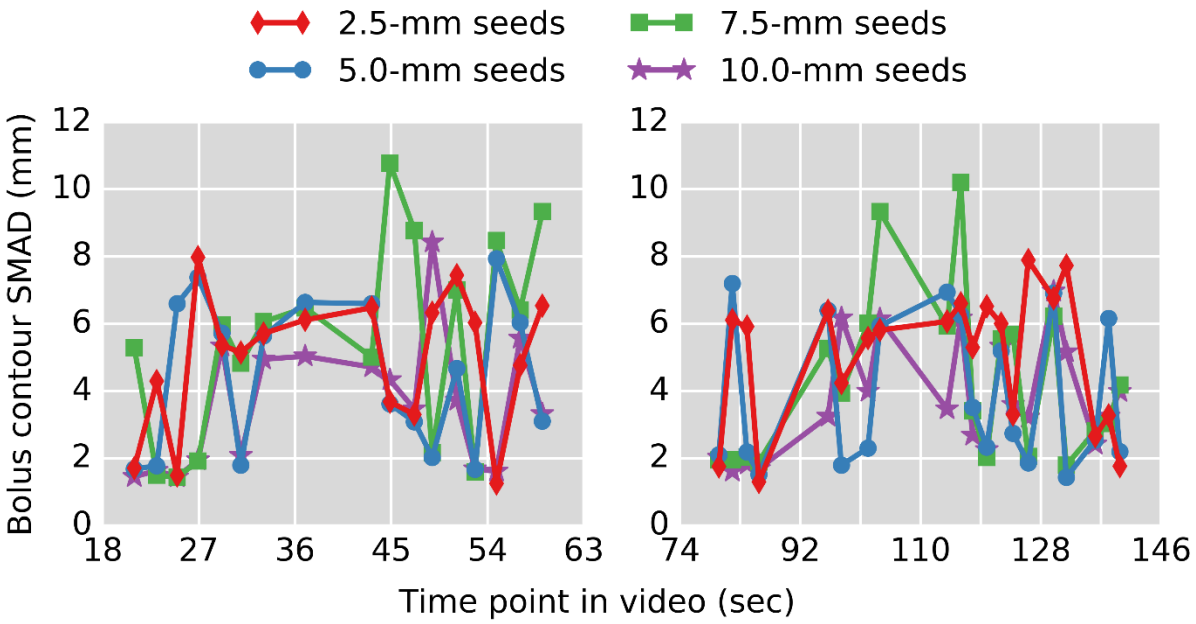

Figure 18: Comparison of seed point spacing in the bolus phantom for the path based-volumetric search. These plots show the SMAD between the measured and ground-truth bolus contour for each frame in video sequence 1 (left) and video sequence 2 (right) after running the search using the four sets of seed points. There is no observable trend between the sets, and there was no statistically significant difference between them either. The results vary more widely between seed point sets than in the marker phantom.

5.3.2.2 Comparison of the two registration methods

The performance of the two registration methods in the bolus phantom differed from that in the marker phantom. As mentioned in Section 5.3.2.1, there were no failed frames with path-based volumetric search. However, frame-to-frame tracking was unable to reach any of the registration frames. This is due to the fact that that there is very little characteristic structure in the phantom other than the piece of bolus. The

initial frames used to start the process were near the entrance to the phantom where the bolus is not visible, and the virtual endoscope became lost in both video sequences before it could reach the other end of the phantom. To get a better characterization of the potential performance of frame-to-frame tracking in this phantom, a second initial frame was chosen for each video sequence in which the bolus was already visible, and tracking was run again from there.

The registration results using the second frame-to-frame tracking run and the path-based volumetric search with 5.0-mm seed points are summarized in Table 10, and a plot of the average measurement errors for each frame in the two video sequences is shown in Figure 19. Frame-to-frame tracking failed for two frames in sequence 2, meaning that the bolus was not visible in the registered virtual endoscopic image. There were numerous frames for which the registered virtual image contained the bolus, but the image was a very poor match for the frame, resulting in very large registration errors. Path-based volumetric search successfully registered all frames, with a median SMAD of 3.5 mm. The median SMAD for frame-to-frame tracking was twice as large at 7.0 mm, and the difference between the two methods was statistically significant ($p < 0.001$ using the Wilcoxon signed-rank test).

Table 10: Comparison of the two registration methods in the bolus phantom. Path-based volumetric searched had better accuracy in this phantom. Frame-to-frame tracking failed for two frames, and had numerous frames for which the registered virtual endoscopic image was a very poor match, resulting in very large SMADs. All SMAD values are in mm.

Metric	**Frame-to-frame tracking**	**Path-based volumetric search**
# of failed frames	2	0
Median SMAD	7.0	3.5
80th percentile	13.6	6.6
Maximum	17.9	7.9

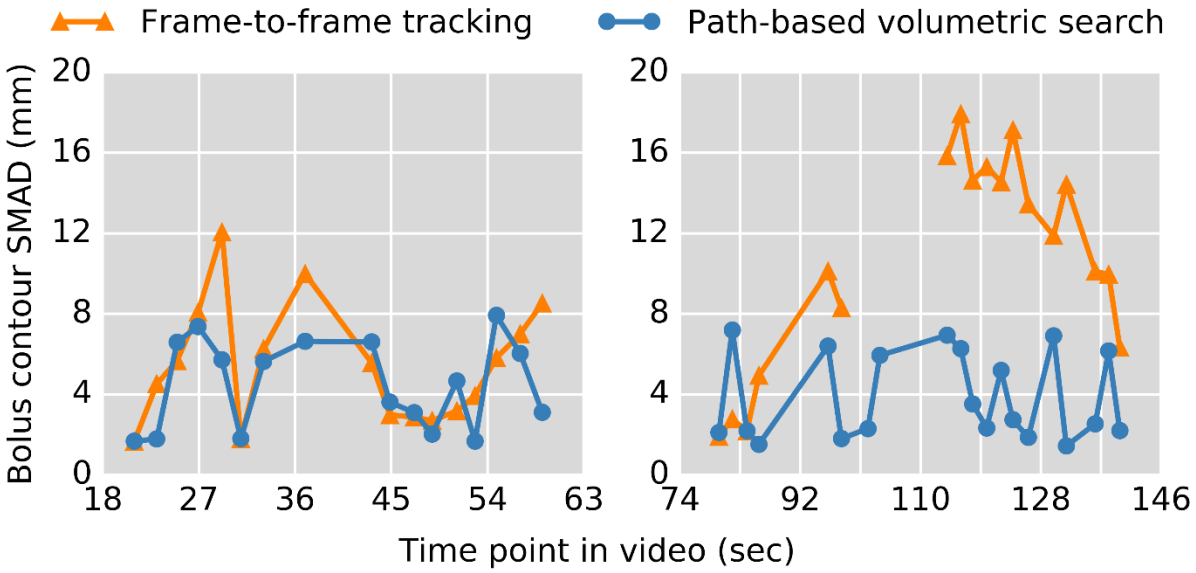

Figure 19: Comparison of the two registration methods in the bolus phantom. These plots show the SMAD between the measured and ground-truth bolus contour for each frame in video sequence 1 (left) and video sequence 2 (right) for frame-to-frame tracking and path-based volumetric search. Path-based volumetric search performed significantly better than frame-to-frame tracking in this phantom. The gap in the plot of frame-to-frame tracking in sequence 2 corresponds to the two failed frames, and the very large errors second half of the sequence are due to frames for which the registered virtual image contained the bolus, but was a very poor match to the frame.

5.3.3 The impact of scene geometry

Scatter plots showing point measurement errors in the marker phantom as a function of measurement angle and measurement distance are shown in Figure 20. The correlation between measurement angle and point error was very weak (Spearman's rho = 0.25, $p < 0.05$), and there was no correlation between measurement distance and point error (Spearman's rho = 0.05, $p > 0.05$). A scatter plot showing bolus contour SMADs as a function of the virtual endoscope's distance from the bolus centroid is shown in Figure 21. There was a moderate correlation between centroid distance and SMAD (Spearman's rho = 0.46, $p < 0.05$).

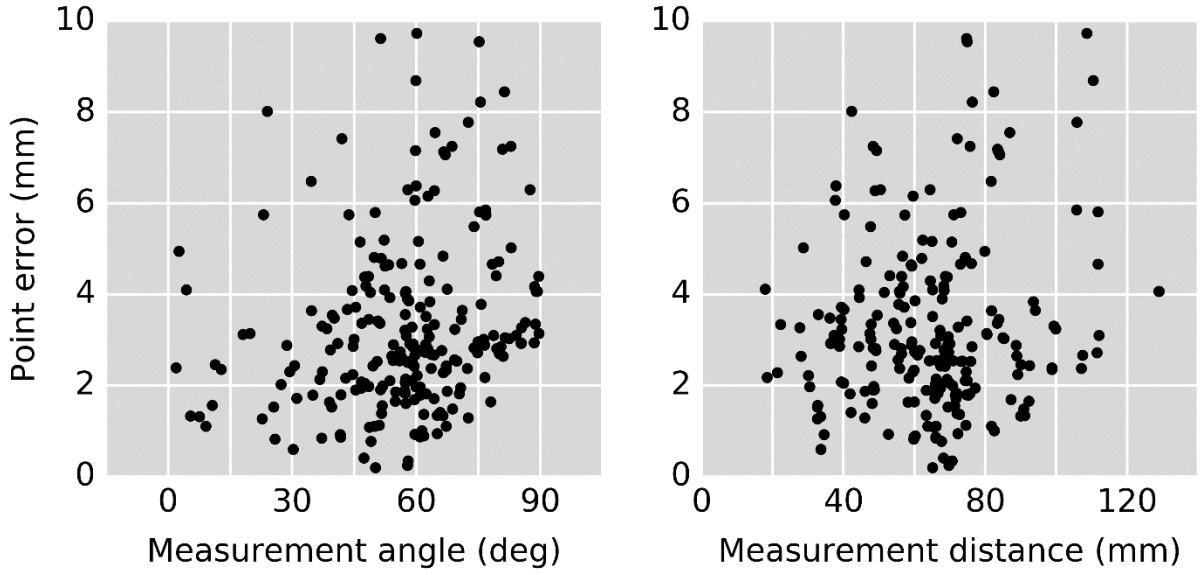

Figure 20: Dependence of point measurement errors on surface angle and distance. These plots show the point measurement errors in the marker phantom presented in Section 5.3.1.2 vs. their measurement angles (left) and distances (right). These parameters are discussed in Section 4.4.3. The correlation between measurement angle and point error was very weak, and there was no correlation between measurement distance and point error.

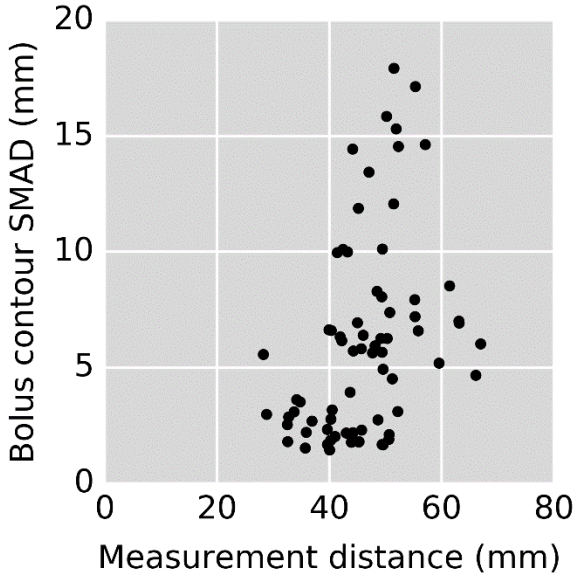

Figure 21: Dependence of bolus contour measurement on surface distance. This plot shows the bolus contour SMADS presented in Section 5.3.2.2 vs. the distance between the virtual endoscope and the bolus centroid. There was a moderate correlation between distance and SMAD.

5.4 Discussion

5.4.1 Summary

Two clay phantoms were created to evaluate the performance of the two endoscopy-CT registration methods. The first phantom contains small radiopaque markers embedded in the luminal surface that were used to map point measurements from endoscopic video to CT. The second phantom contains a piece of bolus material

protruding into the lumen that was used to map object contours from endoscopic video to CT. Two endoscopic video sequences were recorded in each phantom, and sets of registration frames were selected by sampling these videos at regular intervals (n = 36 and 37 for the marker and bolus phantoms, respectively).

In both phantoms, four sets of seed points for the path-based volumetric search were created using variable path interpolation intervals. However, no significant differences were found in registration accuracy between the four sets. In the marker phantom, frame-to-frame tracking successfully registered all frames, but path-based volumetric search failed to find acceptable virtual endoscope coordinates for three frames. On the rest of the frames, there was no significant difference between the performances of the two methods. In the bolus phantom, frame-to-frame tracking failed to reach any registration frames using the starting point near the entrance to the phantom, as the virtual endoscope became lost before reaching the bolus. Tracking was restarted after placing the virtual endoscope in view of the bolus. Frame-to-frame tracking still failed for two frames, and had very large contour mapping errors in many others due to the virtual image being a very poor match to the registration frame. Path-based volumetric search successfully registered all frames in the bolus phantom, and its registration accuracy was significantly better than that of frame-to-frame tracking.

The impact of scene geometry was investigated for both marker and bolus measurements. There was a weak correlation between surface angle and point error in the marker phantom, and a moderate correlation between centroid distance and contour error in the bolus phantom.

5.4.2 Seed point spacing for path-based volumetric search

Seed points for the path-based volumetric search are created by specifying a path interpolation interval δ, and then for each slice the number of seed points is calculated based on this interval such that the 3D spacing will be approximately equal to the interval. No difference in registration accuracy was found in phantoms using interpolation intervals of 2.5, 5.0, 7.5, and 10.0 mm. This is not surprising in the marker phantom. Its internal dimensions are relatively large, which made it easy to control the endoscope's motion as the video sequences were recorded. This meant that the registration frames were generally not very close to the walls, so for any seed point spacing, at least one of them was close enough that the view direction optimization in the coarse search found a reasonable match, and the fine search was able achieve an accurate registration. It is more surprising that seed point spacing had no significant effect in the bolus phantom, which has much smaller dimensions. The results in the bolus phantom were more variable than those in the marker phantom, so it is possible that the greater difficulty of registering video frames in that phantom washes out any effect of seed point spacing. This difficulty is discussed in Section 5.4.3.

5.4.3 Challenges in the bolus phantom

In the marker phantom, the registered virtual images were generally a very good match to the registration frames, both visually and quantitatively. This was not always the case in the bolus phantom, even for frames where the bolus contour SMAD was only

a few mm. One reason for this is the smaller dimensions and difficulty navigating the endoscope in the bolus phantom, which caused the registration frames to be closer to the walls. The bright areas on the walls when the endoscope's lights are very close, and the effect of the endoscope's image sensor adjusting its gain to account for this, are not always reproduced accurately in the virtual images. Another reason that the registered virtual images did not match the registration frames as well in the bolus phantom is the appearance of the bolus itself. Superflab is a translucent material, and its reflectance in the virtual images did not match that in the endoscopic videos at all. This prevented accurate registration in some cases. Finally, the segmentation density used to create the surface meshes may have played a role in reducing the registration accuracy. 0.9 g/cm^3 is a good threshold for the clay walls with a density of ~1.8 g/cm^3, but the Superflab bolus is less dense at ~1 g/cm^3. This threshold may have removed voxels from the bolus that should have been included in the mesh, and the bolus does appear to protrude further into the lumen in the video frames than it does in the virtual images. This difference in shape, combined with the visual effects already described, likely played a role in the highly-variable registration results in the bolus phantom.

5.4.4 The impact of scene geometry

Scene geometry had a smaller impact than expected on registration accuracy. Measurement angle had only a weak correlation in the marker phantom, and measurement distance had a moderate correlation in the bolus phantom. The lack of an effect for measurement angle is likely due to the structure of the surface meshes being

very irregular, both due to the actual structure of the phantoms and the discrete nature of the CT segmentation used to create the meshes. This suggests that the measurement angle describes only a very small local neighborhood. The measurement angles were averaged over several mm^2 using an adaptive ROI based on the measurement distance in an effort to account for this, but no strong effect was found. It is difficult to surmise why so little effect was seen with measurement distance either. Many of the largest errors occurred where the ground-truth is close to an occluding edge, and the projective measurement landed on the other side of that edge. In this scenario, it is plausible that the local structure of the phantom is more important than the measurement distance. Whatever the source of these weak effects, the phantom registration results suggest that the simple surface geometry illustrated in Figure 9 is not sufficient to model uncertainties in projective measurements.

6

Image registration in patients

Parts of this chapter are based on the following publication[64]:

W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

6.1 Introduction

In Chapter 4, the methods developed to register endoscopic video to CT were described. This chapter presents the testing of those methods in patients. The patient cohorts, the acquisition of endoscopic video and virtual endoscopic images, and the details of the registration tests are discussed in Section 6.2. The results of these tests are presented in Section 6.3, and a discussion is given in Section 6.4.

6.2 Methods

6.2.1 Patient cohorts

Two patient cohorts were used to investigate the feasibility of endoscopy-CT image registration. The first included three patients with head and neck cancer undergoing radiotherapy at The University of Texas MD Anderson Cancer Center in Houston, Texas. After approval from the Institutional Review Board, they were enrolled on a protocol allowing their routinely-obtained endoscopic examinations to be archived for this study. Informed consent was obtained in writing prior to enrolling each patient. These patients will be referred to as MDA1, MDA2, and MDA3. For MDA1, two endoscopic examinations were recorded during and at the end of the course of radiotherapy, 28 and 49 days after the planning CT was acquired. For MDA2 and MDA3, a single endoscopic examination was recorded one day after the planning CT was acquired.

The second patient cohort included two patients with head and neck cancer that received radiotherapy at Princess Margaret Hospital in Toronto, Ontario. These patients were part of the cohort used to evaluate endoscopy-CT registration via electromagnetic endoscope tracking by Weersink et al[41, 42], so they were not enrolled prospectively for the work presented in this dissertation. These patients will be referred to as PMH1 and PMH2. The timing of their endoscopic examinations and the endoscope model used are unknown.

There are a few notable differences between the two sets of videos. One is that the MDA videos were acquired with the patient seated, and the PMH videos were acquired

with the patients in the same position as in the planning CTs (supine, with the head and neck positioned with a molded thermoplastic mask). Another is that the PMH videos were recorded with distortion already removed. This meant that calibration parameters were not need to use these videos, but it also had the effect of "locking in" the interlacing artifacts. Both videos were recorded at nearly the same resolution, but the PMH videos have a large border, reducing the effective size of the image. The PMH videos were recorded at a reduced frame rate (6.2 frames per second vs. 30 frames per second) to account for the computational overhead of the electromagnetic tracking system. The electromagnetic sensors were fixed to the outside of the endoscope, so its working length, including the camera lens, was covered by a plastic sheath during the examinations. This introduced blurring artifacts. The visual differences between the videos are illustrated in Figure 22.

Figure 22: Differences between patient endoscopic video characteristics. On the left is a video frame from the MDA cohort, and on the right is a video frame from the PMH cohort. Distortion and interlacing have been removed from the MDA frame. The PMH videos were recorded with distortion already removed. Interlacing artifacts are present throughout the PMH videos, but they are not apparent in this frame, for which the camera was relatively stationary. The PMH videos were acquired with the endoscope covered by a plastic sheath, which introduced blurring artifacts. These can be seen as two faint circles on the right side of the frame.

6.2.2 CT acquisition and virtual endoscopy

The planning CTs for the MDA cohort were acquired using a Brilliance 65 (Philips Healthcare, Andover, MA) with a 50-cm field of view and 1-mm slice thickness. They were interpolated to a 30-cm field of view prior to segmenting the luminal surface in order to increase the number of triangles in the surface meshes, which can improve structural resolution. The planning CTs for the PMH cohort were acquired using an Aquilion ONE (Toshiba America Medical Systems, Tustin, CA) with a 50-cm field of view

and 2-mm slice thickness. The interpolation software was not available during the site visit over which these data were acquired, so the fields of view were not changed.

The luminal surfaces of the airways were segmented using Pinnacle^3 treatment planning software (Philips Healthcare, Andover, MA). The segmentation was performed semi-automatically by selecting a density threshold of 0.6 g/cm^3 and making a small number of mouse clicks inside the lumen on each axial slice. If the contours were to be drawn at the boundary of voxels containing half air and half tissue, the density threshold would be ~0.5 g/cm^3. 0.6 g/cm^3 was chosen based on the observation that thresholds between 0.5 and 1 can provide better virtual endoscopic detail for smaller structures like the epiglottis. The segmentation started in the nasal cavity approximately at the bottom of the eyes and extended inferiorly to the carina of the trachea. The .roi files containing the segmented contours were used to create virtual endoscopy surface meshes as described in Section 3.3.1. The lighting model used to render the virtual images is discussed in Section 6.2.4, and all virtual images were smoothed with a 3 x 3 Gaussian kernel with $\sigma = 1$ pixel.

6.2.3 Endoscopic video datasets

At this point in the study, MDA2 was excluded from further analysis. This choice was made because the patient had a large base-of-tongue tumor that obstructed much of the airway in the pharynx. This obstruction caused the surface mesh to lose most of the characteristic anatomical structure that allows the similarity measure to match virtual endoscopic images to video frames. The large size of the tumor also caused most

of the video to be close-up views of the surface, for which virtual endoscopy can reproduce neither the texture nor the specular reflections from fine structural details that are not present in the surface mesh. An example frame and virtual image are presented in Figure 23.

Figure 23: Examples of an endoscopic video frame and a virtual endoscopic image in patient MDA2. The frame on the left shows the large tumor that obstructed much of the airways. The left edge of the epiglottis is just visible near the top center of the frame. Most of the video consisted of close-up views of the surface. The virtual image on the right was rendered at a similar position. Most of the characteristic anatomical structure was not present due to the large tumor, and the airways superior to this position were cut off entirely by the CT segmentation. MDA2 was excluded from quantitative analysis due to these unfavorable characteristics.

The video sequences were 66 and 48 seconds long for MDA1, 64 seconds for MDA3, 49 seconds for PMH1, and 85 seconds for PMH2. Registration frames were selected for

the MDA patients by sampling the videos every 2 seconds and identifying the least blurry frame out of the sample and the five previous and subsequent frames. The least blurry frame was identified as the one with the largest variance after filtering with a 3 x 3 Laplacian. The selected frames were reviewed to reject those with characteristics unsuitable for registration, such as heavy motion blur, over- or under-exposure, and close-up views of the surface. The automatic sampling method was not used for the PMH patients, because the presence of interlacing artifacts caused the Laplacian filter to select unsuitable frames. Instead, registration frames were selected manually with an effort to sample a variety of distinct views throughout the anatomy. This was challenging for PMH2. The video appeared quite dark with heavy blurring in many frames due to the sheath on the endoscope, and much of it contained duplicate views of the anatomy.

After sampling all of the videos, the set of registration frames consisted of 15 and 12 frames from MDA1, 21 frames from MDA3, 14 frames from PMH1, and 5 frames from PMH2. The preprocessing for all registration frames included conversion to grayscale and smoothing with a 3 x 3 Gaussian kernel with σ = 1 pixel. For frames from MDA1, it also included deinterlacing by replacing every other row with bilinear interpolation and distortion removal as described in Section 3.4.

There were no fiducial markers in the patients that could be used to test registration accuracy, so the ground truth for each registration frame was a set of virtual endoscope coordinates. These were obtained in a two-step process. The first step was camera resectioning by alignment of 2D-3D point correspondences, as described in Section 4.2.2. These coordinates were refined by overlaying the virtual

image on the registration frame and using keyboard input to make fine adjustments to the virtual endoscope's coordinates to get the best visual alignment between anatomical structures in the two images.

At this point in the study, MDA3 was excluded from further analysis. This choice was made because every attempt to perform the camera resectioning resulted in the virtual endoscope being placed outside of the mesh, so ground-truth virtual endoscope coordinates could not be obtained. A likely source of this problem is differences between the anatomical configurations seen the endoscopic video and virtual images. The clearest evidence of this is an anterior offset of the glottis relative to the epiglottis in the virtual images, as shown in Figure 24. This example also shows that MDA3's surface mesh contained relatively little anatomical detail, and the epiglottis was not segmented cleanly from the walls of the pharynx. These factors may have also played a role in the failed resectioning. After exluding this patient, the final dataset used to test endoscopy-CT registration in patient consisted of four endoscopic examinations recorded in three patients, for a total of 46 registration frames with corresponding ground-truth virtual endoscope coordinates.

Figure 24: Examples of an endoscopic video frame and virtual endoscopic image in patient MDA3. A video frame viewing the epiglottis and glottis is shown on the left, and a virtual endoscopic image rendered at a similar position on the right. This patient's surface mesh contained relatively little anatomical detail, and the epiglottis was not segmented cleanly from the walls of the pharynx. The opening visible near the center of the virtual image is directly posterior to the epiglottis. Unlike the video frame, the glottis is not visible through this opening because it was displaced anteriorly in the virtual images (towards the bottom of the image). This difference in anatomical configuration probably contributed to the failed camera resectioning that prevented ground-truth virtual endoscope coordinates from being obtained.

6.2.4 Optimization of virtual endoscopy lighting parameters

In phantom tests, the virtual endoscopic lighting parameters were chosen by inspection. Unlike human tissue, the phantom surfaces were rigid, did not contain variable textures, and did not have saliva and other fluids affecting the reflectance. In that scenario, it is unlikely that the registration accuracy would be highly sensitive to changes in the lighting parameters. But given the increased complexity of registration with patient images, it is better to choose the lighting parameters objectively. This was

accomplished by searching parameter space to find the values that maximize the similarity between video frame and virtual image intensity histograms over the set of patient images described in Section 6.2.3.

Calculating a single histogram for the entire image would not provide the most meaningful results, because it removes any spatial information about the lighting effects. Instead, multiple histograms were calculated for each image using several circular ROIs. Placing these ROIs in the same locations for all frames would not provide meaningful results either, because there are generally structural differences between the registration frames and the ground-truth virtual images. If one of these differences is within an ROI, the histogram would be comparing the lighting for different structures in the two images. Structural differences were avoided by creating masks for each frame, which is described in Section 6.2.4.1. The placement of ROIs within these masks is explained in Section 6.2.4.2, and the metrics used to compare the histograms are given in Section 6.2.4.3. The optimization techniques used to determine the lighting parameters are discussed in Section 6.2.4.4. Because the determination of lighting parameters was a preliminary step in testing the registration methods in patients, the results of the optimization will be presented in this section as well.

6.2.4.1 Creation of structure masks

The airways of the head and neck are not rigid, so there are generally some differences between the structures seen in the endoscopic videos and those seen via virtual endoscopy. These differences are slight in some frames and very large in others,

and can change in a given location due to muscle motion. This means that the registration frames and the registered virtual images have some areas where structures do not match up. To avoid these areas when calculating ROI histograms, the registered virtual images were overlaid on the registration frames and binary masks were manually drawn using paint.net (dotPDN, LLC), a simple photo editing software. These masks remove 0-60% of the image area, with an average of 26 ± 18%. An example is given in Figure 25.

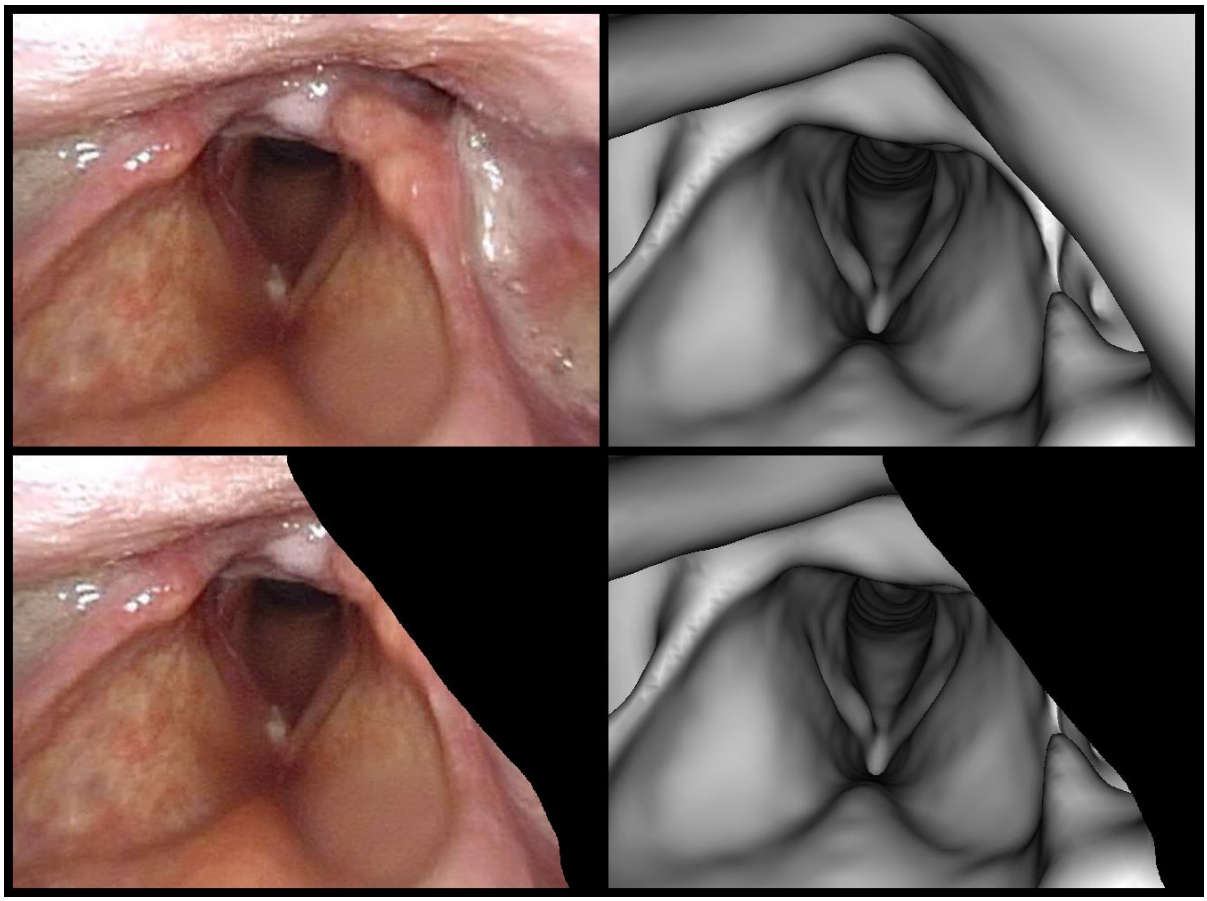

Figure 25: Example of a structural mask used for lighting optimization. The top row shows a registration frame and the corresponding ground-truth virtual image. There is a large edge on the right side of the virtual image that is not present in the frame. The bottom row shows the same images with the manually-draws structure mask applied, which removed the edge. This mask removes 28% of the image area.

6.2.4.2 Selection of ROI locations

After creating the structure masks, circular ROIs with 50-pixel radii were created in which to calculate histograms from the registration frames and virtual images. The number of ROIs was varied based on the percentage of the image that remained after

masking out structural differences, and ranged from 3 to 6. The ROI locations were selected manually by displaying the mask and using mouse input to choose their centers, with the goal to distribute them approximately evenly throughout the region included by the mask. Only the mask was displayed for this step, rather than the frame or virtual image, to avoid inadvertently biasing the ROI selection in favor of any particular anatomical features. An example is given in Figure 26.

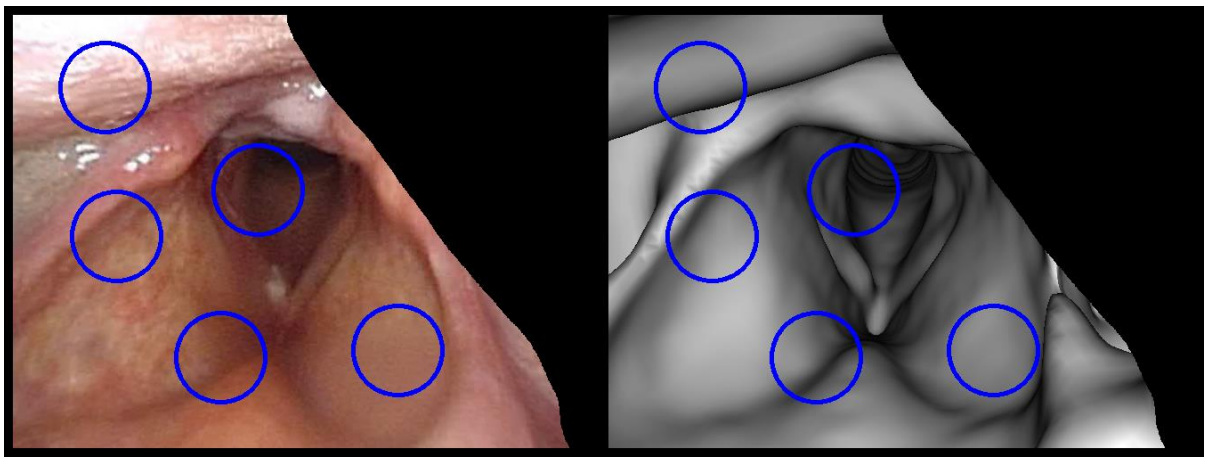

Figure 26: Example of the ROIs used to calculate image histograms. This frame and virtual image are the same ones presented in Figure 25. The ROI radii are 50 pixels, and their centers were selected manually on the binary structure mask.

6.2.4.3 Histogram comparison metrics

The optimization of virtual endoscopy lighting parameters is not a well-characterized problem, so a variety of metrics were used to compare the ROI histograms in the registration frames and ground-truth virtual images. They are given in the following equations, in which H_f and H_v are the histograms in the registration frame and ground-truth virtual images, $\bar{H}_f$ and $\bar{H}_v$ are their averages, and N is the number of histogram bins.

1. Euclidean distance

$$\sqrt{\sum_i \left(H_{fi} - H_{vi}\right)^2} \tag{27}$$

2. Correlation

$$\frac{\sum_i \left(H_{fi} - \bar{H}_f\right)\left(H_{vi} - \bar{H}_v\right)}{\sqrt{\sum_i \left(H_{fi} - \bar{H}_f\right)^2 \sum_i \left(H_{vi} - \bar{H}_v\right)^2}} \tag{28}$$

3. Intersection

$$\sum_i \min\left(H_{fi}, H_{vi}\right) \tag{29}$$

4. Chi-squared distance

$$\sum_i \frac{\left(H_{fi} - H_{vi}\right)^2}{H_{fi}} \tag{30}$$

5. Alternate chi-squared distance

$$2 \cdot \sum_i \frac{\left(H_{fi} - H_{vi}\right)^2}{H_{fi} + H_{vi}} \tag{31}$$

6. Hellinger distance

$$\sqrt{1 - \frac{1}{\sqrt{\bar{H}_f \bar{H}_v N^2}} \sum_i \sqrt{H_{fi} H_{vi}}} \tag{32}$$

7. Kullback-Leibler divergence

$$\sum_i H_{fi} \log\left(\frac{H_{fi}}{H_{vi}}\right) \tag{33}$$

6.2.4.4 Optimization methods and results

An objective function was defined by calculating the sum of a given histogram comparison metric over each ROI in each registration frame/ground-truth virtual image pair. All histograms were computed with 64 bins. The correlation and the intersection are maximized when the histograms are most similar, so their values were negated when they were used in the objective function. The goal of the optimization was to find the set of lighting parameters that minimized this objective function. Recall that the attenuation factor in the virtual endoscopy lighting model (Section 3.3.2) is given by

$$attenuation\ factor = \frac{\cos^e(\phi)}{a_c + a_l d + a_q d^2} \tag{34}$$

The parameters considered in this optimization were the linear and quadratic attenuation coefficients a_l and a_q, the cosine exponent e, and the intensity i, which sets the brightness. The angle ϕ is a property of the surface. The constant attenuation coefficient a_c was held equal to 1 because an increase in all the attenuation coefficients can be compensated for by an increase in intensity, so one of these parameters can be treated as a constant for the optimization.

The optimization was performed in two steps. The first step was a coarse brute-force calculation over the ranges in parameter space given in Table 11. These ranges were chosen based on experience to cover the meaningful range of lighting characteristics. The brute-force calculation was performed using each histogram

comparison metric in the objective function, and the best set of parameters for each metric is given in Table 12. All metrics had consistent results except for chi-squared distance, so it was excluded from further analysis.

The brute-force calculations showed that the best lighting parameters were near a_l = 1.5-2, a_q = 0, e = 0.5-1.5, and i = 3-4. The second step of the lighting optimization was to refine the results of these calculations by running a series of Nelder-Mead simplex optimizations starting from each point in a grid based on these approximate parameter values. The grid coordinates are given in Table 13. The series of optimizations was run using each histogram comparison metric in the objective function except for chi-squared distance, and the best result for each metric is given in Table 14.

Table 11: Lighting parameter ranges for the brute-force optimization. The ranges of values are given in the notation $start\ value : interval : stop\ value$.

Name	**Symbol**	**Range of values**
Linear attenuation	a_l	0 : 0.1 : 2
Quadratic attenuation	a_q	0 : 0.1 : 2
Cosine exponent	e	0 : 0.5 : 10
Intensity	i	1 : 1 : 30

Table 12: Results of the brute-force optimization of the lighting parameters. Each row gives the best set of parameter values when the specified histogram comparison metric was used in the objective function.

Comparison metric	a_l	a_q	e	i
Euclidean distance	1.8	0	0	3
Correlation	2	0	1.5	4
Intersection	1.5	0	1	3
Chi-squared distance	1.8	2	10	1
Alternate chi-squared distance	1.5	0	1	3
Hellinger distance	2	0	1.5	4
Kullback-Leibler divergence	1.6	0	0.5	3

Table 13: Lighting parameter values used to start the simplex optimizations. The optimizations were started from each point in the 5 x 2 x 4 x 4 grid defined by the values given here. The simplex scale vector $\Delta_{simplex}$ (Section 4.1) was set at half the spacing between these values.

Name	**Symbol**	**Grid coordinates**
Linear attenuation	a_l	1.25, 1.5, 1.75, 2, 2.25
Quadratic attenuation	a_q	0, 0.25
Cosine exponent	e	0, 0.75, 1.5, 2.25
Intensity	i	2, 3, 4, 5

Table 14: Results of the simplex optimizations of the lighting parameters. Each row gives the best simplex result when the specified histogram comparison metric was used in the objective function.

Comparison metric	a_l	a_q	e	i
Euclidean distance	1.83	0	0.01	3.01
Correlation	2.56	0	1.59	4.81
Intersection	1.71	0	0.99	3.29
Alternate chi-squared distance	1.75	0	0.99	3.34
Hellinger distance	1.74	0	1.10	3.39
Kullback-Leibler divergence	1.33	0	0.60	2.71

Each histogram comparison metric found a different set of lighting parameters to be best, but the results of the simplex optimizations were fairly consistent across all metrics, particularly for quadratic attenuation. The intersection, alternate chi-squared distance, and Hellinger distance found very consistent values for all parameters, so the final set of lighting parameters was chosen based on their results. This set of parameters is given in Table 15, and it was used to render all patient virtual endoscopic images used in the remainder of this dissertation.

Table 15: Final results of the lighting parameter optimization. These values were used to render all patient virtual endoscopic images. They were chosen based on the results of the histogram comparison metrics intersection, alternate chi-squared distance, and Hellinger distance given in Table 14.

Name	Symbol	Final value
Linear attenuation	a_l	1.75
Quadratic attenuation	a_q	0
Cosine exponent	e	1.05
Intensity	i	3.35

6.2.5 Frame-to-frame tracking

Frame-to-frame tracking was performed as described in Section 4.2.1. Initial frames to start the tracking were selected in the nasal cavity near the start of the video sequences. The initial virtual endoscope coordinates were determined manually. The long, narrow passages make resectioning difficult (see Section 4.2.2 for details), because the least-squares optimization to align the 2D-3D point correspondences is usually poorly-conditioned and moves the virtual endoscope outside the mesh. Fortunately, the space for the endoscope in the inferior nasal cavity is much smaller than that in the phantoms, so it is not challenging to place the virtual endoscope near the correct location. At each subsequent frame, the virtual endoscope's coordinates $(x, \ y, \ z, \ \theta_x, \ \theta_y, \ \theta_z)$ were optimized to match the next frame in the video sequence using the simplex method described in Section 4.1. As in the phantoms, the scale of the search space was set using

$$\Delta_{simplex} = (2\ mm,\quad 2\ mm,\quad 2\ mm,\quad 5\ deg,\quad 5\ deg,\quad 5\ deg) \qquad (35)$$

6.2.6 Path-based volumetric search

The path-based volumetric search was performed as described in Section 4.3.1. The possible endoscope path was created by manually selecting a small set of points on a single sagittal slice of the planning CT. An example is given in Figure 27. Seed points were created using a path interpolation interval of δ = 5 mm (see Section 4.3 for details). The corresponding numbers of seed points for each slice were calculated using Equation 19. In both phantoms, the registration accuracy did not depend on the interpolation interval, so this choice for δ was somewhat arbitrary. However, given the increased complexity of registration with patient images, and the fact that the only downside to decreasing the interval is increased computation time, choosing a smaller interval was prudent.

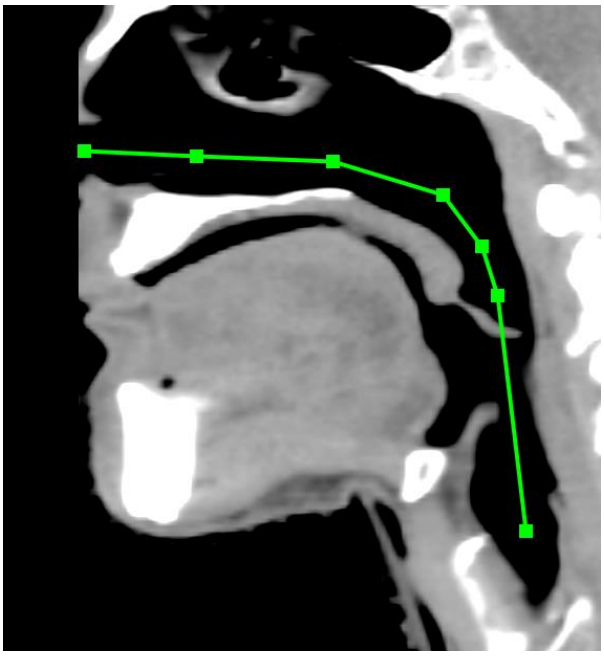

Figure 27: Virtual endoscope path creation in a patient. A sagittal CT slice shows the manually-selected possible virtual endoscope path used for the path-based volumetric search registration method. The nose is cropped out due to the field-of-view reduction described in Section 6.2.2.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, B. M. Beadle, R. Wendt III, A. Rao, X. A. Wang, and L. E. Court. "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking." *PLoS One* 12(5), 1-23 (2017).

At each seed point, the virtual endoscope's view direction $(\theta_x,\ \theta_y,\ \theta_z)$ was optimized to match the registration frame using the simplex method described in Section 4.1. As in the phantoms, the scale of the search space was set using

$$\Delta_{simplex} = (20\ deg,\quad 20\ deg,\quad 20\ deg) \tag{36}$$

This is larger than the 5-degree scale used for frame-to-frame tracking. A smaller value is appropriate in that scenario because the endoscope does not move much between adjacent frames, so the starting point for the optimization can be assumed to be very close to the optimum.

The result of the coarse search was the seed point with the optimized view direction that provided the virtual endoscopic image most similar to the registration frame. For the fine search, the 3 x 3 x 3 grid was created centered on this optimized seed point with 2.5-mm spacing. At each grid point, the virtual endoscope's coordinates $(x, \quad y, \quad z, \quad \theta_x, \quad \theta_y, \quad \theta_z)$ were optimized to match the registration frame using the simplex method. The scale of the search space was set using

$$\begin{aligned} &\Delta_{simplex} \\ &= (1.25\,mm, \quad 1.25\,mm, \quad 1.25\,mm, \quad 2.5\,deg, \quad 2.5\,deg, \quad 2.5\,deg) \end{aligned} \tag{37}$$

The result of the fine search was the optimized grid point that provided the virtual endoscopic image that was most similar to the registration frame.

6.2.7 Measurements of registration accuracy

The outputs of frame-to-frame tracking and the path-based volumetric search are sets of registered endoscope coordinates C_{reg} corresponding to each registration frame F_{reg} in the datasets. Unlike the phantoms, there were no fiducial markers in the patients that could be used to compare projective measurements to a ground truth, so another

method was needed to quantify registration accuracy. A straightforward option would be to take the registered and ground-truth endoscope coordinate vectors $(x,\ \ y,\ \ z,\ \ \theta_x,\ \ \theta_y,\ \ \theta_z)$ and calculate some distance metric. However, this is not particularly meaningful. The true test of endoscopy-CT registration is its ability to map points from 2D to 3D, and in that scenario a change in the virtual endoscope's position can be compensated for by an opposite change in its orientation. For example, if the virtual endoscope is moved to the right, its view direction can be turned back to the left, and the projective measurements may be largely unchanged. To account for this, the world transforms were taken for the ground truth and registered virtual endoscopic images, and the registration accuracy was quantified by the median projective distance error between these sets of CT-space points.

Scene geometry did not have a strong impact on registration accuracy in the phantoms. To investigate its role in patient images, the world transforms for the ground-truth virtual images were used to compute measurement angles and distances for each point. In addition to these quantities, edge masks were created for both the ground-truth and registered virtual images using a distance threshold of 2 mm and morphological erosion with a 3 x 3 structuring element for five iterations. These masks were used to investigate the potential identification of regions of high uncertainty to avoid when making projective measurements. See Section 4.4 for more details on the world transform, measurement angles, and edge masks.

6.3 Results

6.3.1 Comparison of the two registration methods

Frame-to-frame tracking was unable to reach any of the registration frames in any of the video sequences due to the virtual endoscope becoming lost in the nasal cavity. To get a better characterization of the potential performance of frame-to-frame tracking in patients, the virtual endoscope was placed at the ground-truth coordinates for the first registration frame and tracking was restarted from there. The registration results using the second frame-to-frame tracking run and the path-based volumetric search are summarized in Table 16, and plots of the registration accuracy for each frame in the four video sequences are shown in Figure 28-31. Note that in all of these figures, the first frame is omitted for frame-to-frame tracking due to the restart.

Frame-to-frame tracking failed for two frames in MDA1 video sequence 1, seven frames in MDA1 video sequence 2, and two frames in PMH2. Failed frames were identified as those for which the registered virtual image contained none of the anatomy recognizable in the registration frame, generally meaning that the virtual endoscope was directly in front of a wall. Using this criterion, path-based volumetric search successfully registered all frames in all video sequences. However, frames were labelled as successful on a very inclusive basis, and for many successful frames, the registered virtual image was still a poor match to the registration frame. This led to larger errors overall than those seen in the phantoms. Excluding failed frames, the average registration accuracy with frame-to-frame tracking was 21.7 ± 10.4 mm. Path-

based volumetric search performed significantly better, with an average registration accuracy of 12.5 ± 9.9 mm ($p < 0.001$ using the Wilcoxon signed-rank test). Its performance varied between video sequences. The best results were obtained for patient MDA1, video sequence 1, for which path-based volumetric search had a registration accuracy within 5 mm for 10 out of 15 frames.

Table 16: Comparison of the two registration methods in patients. Path-based volumetric search had significantly better registration accuracy ($p < 0.001$ using the Wilcoxon signed-rank test). Frame-to-frame tracking failed for 11 frames, and both methods had numerous frames for which the registered virtual image was not a great match to the registration frame, leading to larger errors overall than in the phantoms. All accuracy values are in mm.

Metric	**Frame-to-frame tracking**	**Path-based volumetric search**
# of failed frames	11	0
Median accuracy	23.4	9.9
80^{th} percentile	30.5	20.6
Maximum	38.0	41.0

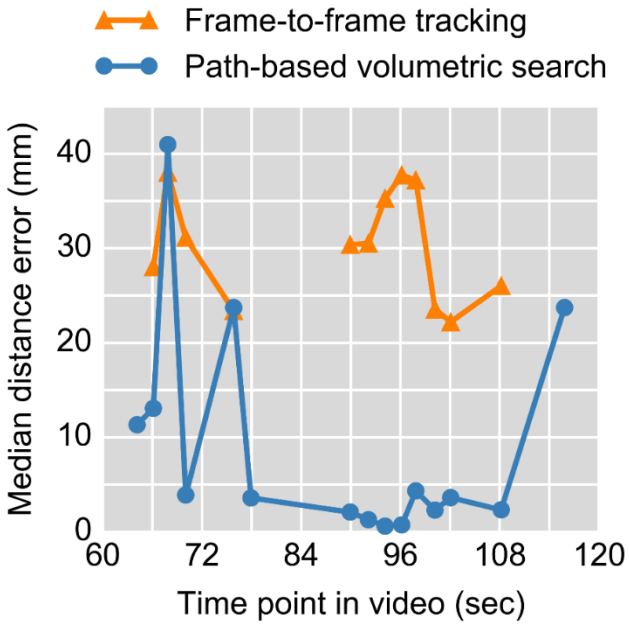

Figure 28: Comparison of the two registration methods for patient MDA1, video sequence 1. This plot shows the median distance error between the world transforms of the ground-truth and registered virtual images for each registration frame. Frame-to-frame tracking failed for two frames, indicated by gaps in the plot. Path-based volumetric search performed very well on this sequence, with a registration accuracy within 5 mm for 10 out of 15 frames.

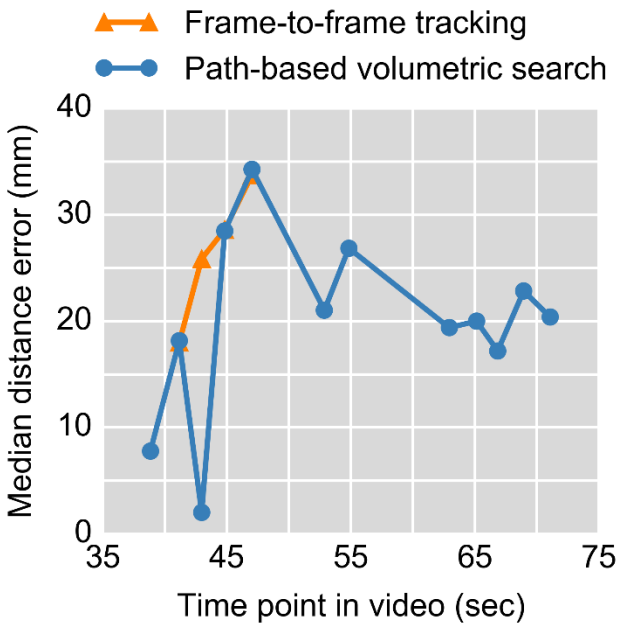

Figure 29: Comparison of the two registration methods for patient MDA1, video sequence 2. This plot shows the median distance error between the world transforms of the ground-truth and registered virtual images for each registration frame. Frame-to-frame tracking failed for seven frames after the virtual endoscope became stuck in front of a wall. Path-based volumetric search had worse performance here than in video sequence 1.

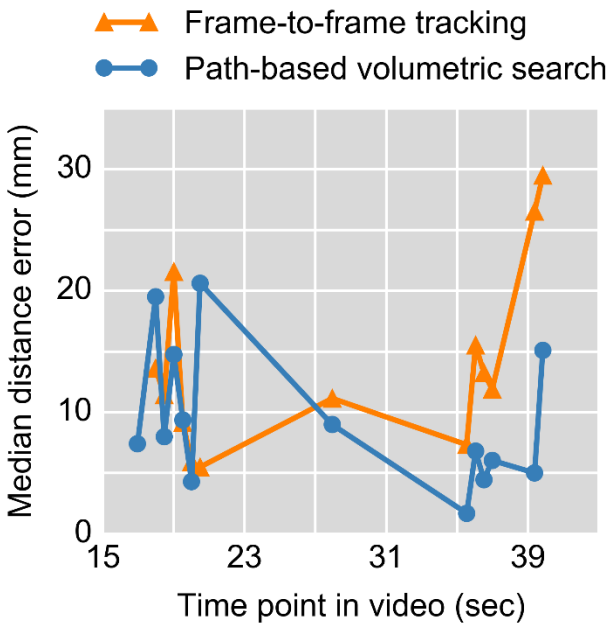

Figure 30: Comparison of the two registration methods in patient PMH1. This plot shows the median distance error between the world transforms of the ground-truth and registered virtual images for each registration frame. Frame-to-frame tracking did not fail for any frames, but its registration accuracy was worse towards the end of the sequence than that of path-based volumetric search. This shows that with frame-to-frame tracking, error can accumulate when the virtual endoscope fails to keep up, even when the results are not easily identified as failures.

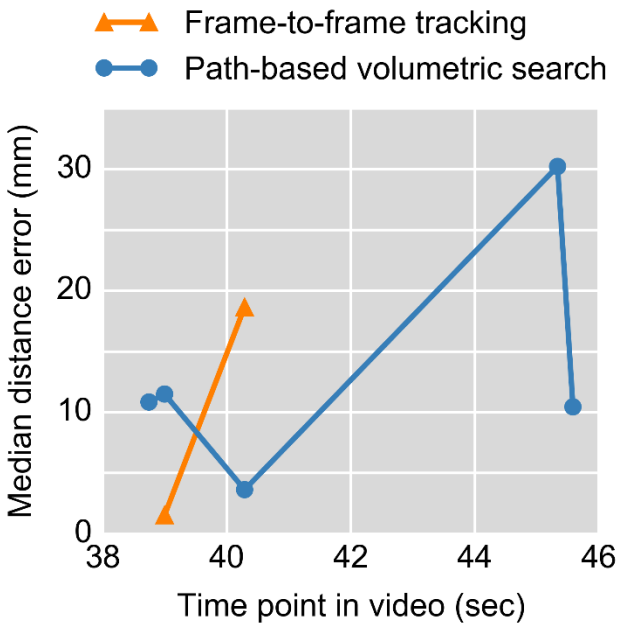

Figure 31: Comparison of the two registration methods for patient PMH2. This plot shows the median distance error between the world transforms of the ground-truth and registered virtual images for each registration frame. Frame-to-frame tracking failed for two frames. This video sequence contained many unfavorable characteristics, making it challenging to select registration frames.

6.3.2 The impact of scene geometry

For each frame, Spearman's rho was calculated between the distance errors for the virtual image registered with path-based volumetric search and the measurement angles or distances for the ground-truth virtual image. These correlations were variable between frames and weak overall, with an average of 0.17 ± 0.23 for measurement angle and 0.28 ± 0.31 for measurement distance (all $p < 0.001$). The weak correlation for measurement angle echoes the results in the marker phantom. Despite the weak

correlation for measurement distance, this value does play a role in projective measurement error. This is illustrated in Figure 32, which shows the median distance errors for all registration frames with progressive distance cutoffs applied. For example, with a cutoff of 60 mm, all points in the world transform with measurement distances beyond 60 mm were excluded. The errors were reduced by each cutoff, and the difference between every adjacent cutoff was significant ($p < 0.05$ using the Wilcoxon signed-rank test). This shows that excluding points with large measurement distance can reduce projective measurement errors.

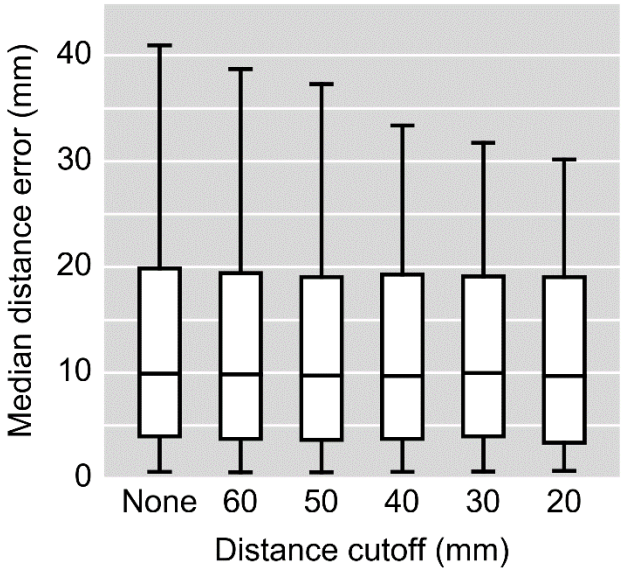

Figure 32: Reduction in registration error when distant points are excluded. These box plots show the median distance error with path-based volumetric search for all patient registration frames with progressive measurement distance cutoffs applied. For example, in the 60-mm plot, all points with a measurement distance larger than 60 mm have been excluded. The boxes show the median and quartiles, and the whiskers show the minimum and maximum. Though the cutoffs do not appear to have much effect on the first three quartiles, the reduction in error is significant between all adjacent cutoffs ($p < 0.05$ using the Wilcoxon signed-rank test).

6.4 Discussion

6.4.1 Summary

Two patient cohorts were used to evaluate the performance of the two endoscopy-CT registration methods. The first cohort included three head-and-neck radiotherapy patients prospectively enrolled for this study, for one of whom two endoscopic

examinations were recorded. The second cohort included two patients from previous studies of registration via electromagnetic endoscope tracking by Weersink et al[41, 42]. One patient in the first cohort was deemed unsuitable for registration due to a large base-of-tongue tumor that obstructed much of the airway, preventing clear views of the anatomy in the endoscopic video and limiting the structural detail in the virtual endoscopic images. Another patient from the first cohort was excluded from quantitative analysis because ground-truth virtual endoscope coordinates could not be obtained, possibly due to a difference between the anatomical configuration seen in the endoscopic video and that seen in the virtual endoscopic images.

A set of 46 registration frames was selected from the remaining patients, and ground-truth virtual endoscope coordinates were obtained manually for each. Each frame was registered to CT using the frame-to-frame tracking and path-based volumetric search methods. In all videos, frame-to-frame tracking failed to reach any registration frames from the starting point in the nasal cavity. Tracking was restarted after placing the virtual endoscope at the ground-truth coordinates for the first registration frame. After restarting, frame-to-frame tracking still failed for many frames, generally when the virtual endoscope became stuck directly in front of a wall. Path-based volumetric search successfully registered all frames, but registration errors were larger than those seen in phantoms.

The impact of scene geometry on registration accuracy was investigated for the path-based volumetric search results. There were weak correlations between registration accuracy and both measurement angle and measurement distance, but these correlations were highly variable from frame to frame. However, excluding

distant points from projective measurements did reduce overall errors, indicating that measurement distance may be an important consideration for applications of endoscopy-CT registration.

6.4.2 Exclusion of patients from the first cohort

Two patients were excluded from the first cohort. One was excluded due to a large tumor that prevented adequate visualization on both the endoscopic video and the virtual endoscopic images. The rest of the patients in both cohorts had early-stage disease, so there was no obstruction of the airways. This suggests that there is a subset of patients for whom endoscopy-CT registration may not be possible, at least not with the image-based methods used in this study.

The second patient was excluded due to an inability to obtain ground-truth virtual endoscope coordinates, possibly due to a difference between the anatomical configuration seen in the endoscopic video and that seen in the virtual endoscopic images. This difference may be the result of positioning difference between the seated endoscopic examination and the supine CT. This patient's virtual endoscopic images also had less detail overall and poor segmentation of the epiglottis, so the exact source of the difficulties is unclear. These characteristics suggest that patient positioning and CT acquisition may be important considerations for the robustness of endoscopy-CT registration.

6.4.3 Anatomical differences between real and virtual endoscopy

One feature of endoscopy-CT registration that becomes apparent with experience is that the configuration of the anatomy in the endoscopic video does not exactly match that in the virtual endoscopic images. A major source of this difference is muscle motion in the pharynx and larynx. This motion can be dramatic, such as when the patient swallows. Transient motion does not pose a problem for endoscopy-CT registration, because frames in those sequences would not be useful for registration in a clinical setting due to very poor visualization of the anatomy. However, the motion can also be subtle, such as gradual changes in the opening of the glottis and the positions of the walls. This type of motion is difficult to avoid by frame selection, and it can cause systematic differences between the anatomical configuration in the endoscopic videos and that in the virtual endoscopic images.

The most notable difference seen in this study was a displacement of the posterior wall of the pharynx. The pharynx appeared wider in virtual endoscopic images, and it was particularly problematic for registration because the posterior wall was much closer to the epiglottis in the videos. This is illustrated in Figure 33. The wall positioning caused a large bright area at the top of several registration frames that could not be reproduced in the virtual images near the ground truth coordinates, leading to large registration errors.

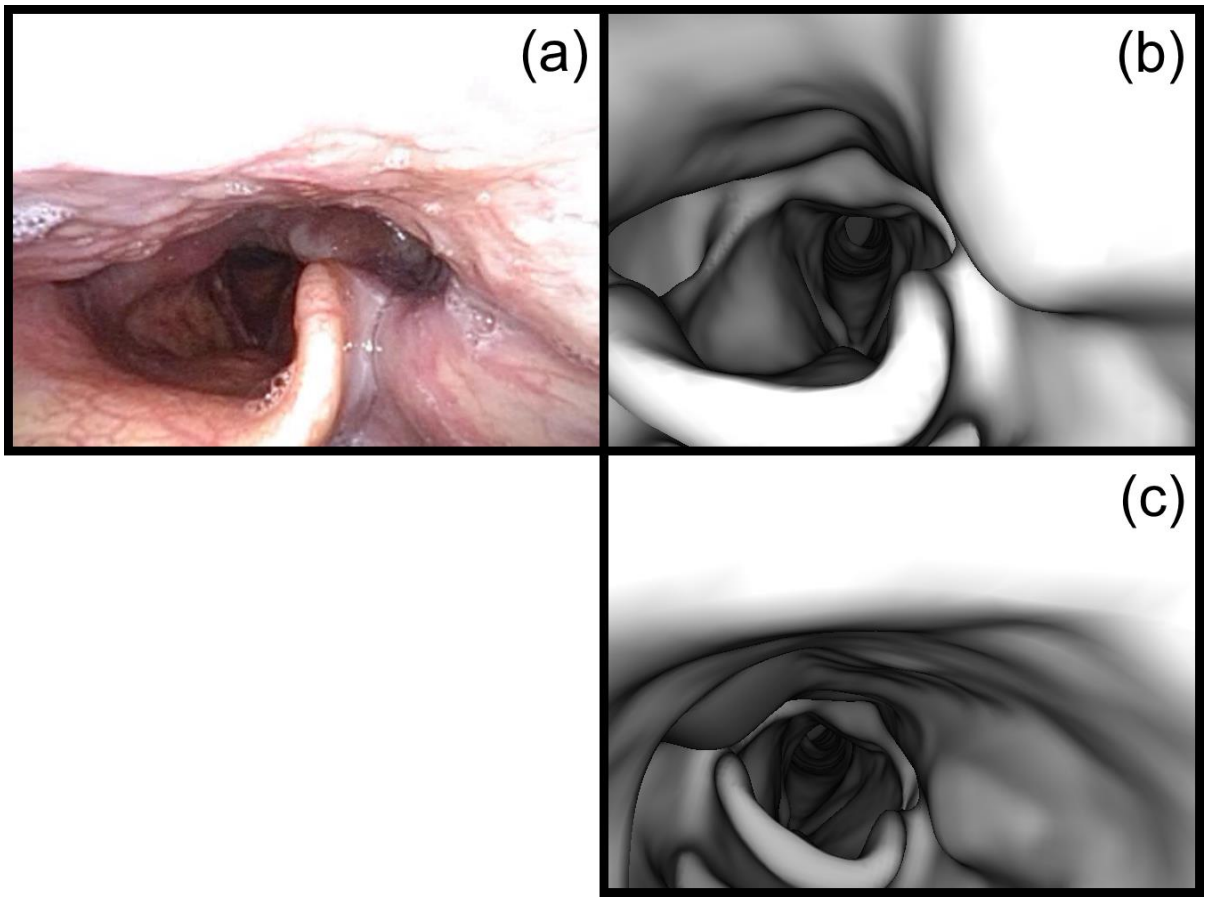

Figure 33: Example of anatomical differences between endoscopic video and virtual endoscopy. (a) A registration frame from patient MDA1. The posterior wall of the pharynx is quite close to the epiglottis, creating a large bright spot at the top of the frame. (b) The ground-truth virtual image for this frame. The pharynx appears much wider, and the bright region cannot be reproduced at the ground-truth endoscope coordinates. (c) The registered virtual image obtained with path-based volumetric search. The bright region at the top of the virtual image caused a false match to the registration frame. The median distance error for this frame was 28.5 mm.

6.4.4 The impact of manual inputs

The path-based volumetric search method requires manual input to create the possible endoscope path through the volume. This was done by selecting a small set of points from the nasal cavity past the glottis on a sagittal CT slice. This will introduce some variability between users, but it is important to note that path-based volumetric search does not depend on the possible path exactly matching the actual path taken in the recorded video. No effort was made to replicate the actual path when creating the paths used in this study.

The path is used to assign virtual endoscope view directions when slicing the surface and when initializing the simplex optimizations. It is possible that a highly-convoluted path drawn by a user could cause inadequate sampling of the volume if the slices are at odd angles, or could initialize the virtual endoscope with a view direction that causes the simplex to find a false optimum. However, the possible endoscope path is meant to be a simplified approximation like the one illustrated in Figure 27. If the path is drawn in this manner, it is unlikely that inter-user variability will impact the registration.

Manual input was also used to identify point correspondences to obtain ground-truth virtual endoscope coordinates via camera resectioning for patient images. An effort was made to select these points covering a range of distances from the camera, which improves the conditioning of the resectioning and should reduce the impact of inter-user variability. However, the impact of inter-user variability was not quantified for this or the path creation. The purpose of this study was to explore the feasibility of

endoscopy-CT image registration, but if path-based volumetric search is to be used clinically, these manual inputs must be standardized, or automated if possible.

6.4.5 Computation times

One drawback of the path-based volumetric search method is its computational complexity. The creation of the virtual endoscope path and the creation of the set of seed points by slicing the surface and k-means clustering (steps 2 and 3 in Section 4.3.1) need to be done only once for a given phantom or patient, but the coarse search over all the seed points and the subsequent fine search (steps 4-6 in section 4.3.1) must be done for each frame. This process required about 20 minutes per frame in the current implementation, which is written in the Python programming language. This is certainly a practical limitation, but there are many aspects of this algorithm that can be optimized to reduce the computation time. The most salient target is parallel computing for the coarse search, as the optimization at each seed point is independent of that at all other seed points.

6.4.6 Local optima in the Nelder-Mead simplex

Both frame-to-frame tracking and path-based volumetric search use the Nelder-Mead simplex algorithm to optimize virtual endoscope coordinates to find the best match for the registration frame. This method was chosen because it does not require any gradient computation, which is not possible in this scenario. However, it is a local

optimization technique that can find only the nearest optimum. This is not a problem with frame-to-frame tracking, because in a 30-frames-per-second video, the next frame is always going to be close enough to the previous solution that local optimization is sufficient. But when trying to register a frame with no nearby initial guess for the virtual endoscope coordinates, a global search is required.

Path-based volumetric search accomplishes this by treating the position and orientation components of virtual endoscope coordinate space separately. The slicing and clustering of seed points samples the position space of the entire volume. The manually-drawn possible endoscope path eliminates the need for a global search of orientation space, because it is used to initialize the virtual endoscope's view direction sufficiently close to the optimum. So even though a local optimization algorithm is used in both techniques, path-based volumetric search effectively performs a global optimization by using many start points that sample the entire volume. Its performance depends on the seed point sampling being dense enough that the optimum is not missed. However, it may be worthwhile to investigate performing the initial coarse search using a global optimization algorithm such as simulated annealing.

6.4.7 The role of patient positioning

Virtual endoscopy surface meshes were created from planning CTs, which are acquired in the supine position with the patient's head, neck, and shoulders positioned with a molded thermoplastic mask. The patient's head is typically tilted back to keep it out of the treatment beam. For the patients in the PMH cohort, the endoscopic

examinations were performed in the same position, with the thermoplastic mask. This was not the case for the MDA cohort, and would not be the case for any endoscopic examination performed in routine clinical practice. These examinations are performed with the patient seated in a chair, and the patient is not positioned in any particular way.

One weakness of the methods presented in this study is that it is likely that the different positions introduce anatomical differences between the endoscopic video and the virtual images. These differences could prevent successful registration when the virtual image cannot match the appearance of the video frame at the correct coordinates, and could introduce errors when mapping video frames to CT even when registration is successful. It is likely that positioning differences played a role in the exclusion of patient MDA3 from the study, and it is possible they influenced the anatomical differences discussed in Section 6.4.3. Though the PMH cohort did not have positioning differences, the sample size in this study is too small to draw any conclusions about the impact of positioning on registration accuracy.

Without a patient set including supine and upright CTs, it will be difficult to fully understand the impact of patient positioning. A possible solution to this problem is to allow the surface mesh to deform to match the appearance of the anatomy in the endoscopic video, but validation will remain a challenge. Another option that would reduce positioning differences is to perform the endoscopic examination with the patient in the supine position using the thermoplastic mask. This would not cause undue burden on the patient, but it is a change to standard clinical practice. The thermoplastic masks are affixed to mounts on the couch of a CT scanner or linear

accelerator, and scheduling time to use this equipment could pose a logistical challenge in busy clinics. Furthermore, development of image registration for seated-position endoscopy is appealing because it could be used in scenarios where supine endoscopy is not available, such as retrospective analysis of archived video or endoscopic examinations performed at outside institutions.

7

The influence of patient positioning and non-rigid anatomy

This chapter is based on the following publication[73]:

W. S. Ingram, J. Yang, R. Wendt III, B. M. Beadle, A. Rao, X. A. Wang, and L. E. Court. "The influence of non-rigid anatomy and patient positioning on endoscopy-CT image registration in the head and neck." *Medical Physics* (in press).

7.1 Introduction

Though virtual endoscopy has been used for registration of endoscopic video to CT scans in a variety of anatomical regions and clinical settings, very little attention has been paid to the sources of uncertainty in projective measurements to map video frames to CT space. Perhaps the most important source of uncertainty is differences between the airway surface structure that is seen on CT and that which is seen in the endoscopic video. These differences have several causes. The most apparent is muscle motion during the endoscopic examination when the patient swallows or speaks, which

can change the size of the lumen and the relative positions of anatomical structures. This effect can largely be eliminated by judicious selection of the video frames to be registered to CT.

Another cause of airway surface differences is patient positioning. Endoscopic examinations for head and neck cancer patients are performed with the patient seated upright in a chair, while the CT scans used for radiotherapy treatment planning are acquired in the supine position with the patient's head tilted back and secured in a molded thermoplastic mask. This changes the relative positions of different regions of the respiratory tract in the two data sets. Finally, the upper respiratory tract is not rigid, and there may be day-to-day changes in the positions of tissues that influence registration to the planning CT, which is acquired at a single time point. Anatomic variability in the head and neck has been studied previously[74, 75], but not for the surfaces of the respiratory tract, and how this variability affects endoscopy-CT registration remains unknown.

This chapter presents an analysis of the impacts of patient positioning and daily anatomical variations on the uncertainty of virtual endoscopic projective measurements in the head and neck. Unlike the previous chapters, no endoscopic videos were acquired for this analysis. The general approach was to take sets of projective measurements in the planning CT and compare them to projective measurements taken in a different CT when the virtual endoscope was placed at the same position. Measurements were taken throughout the airways of the head and neck in order to characterize the uncertainty in different anatomical regions. As with the phantom and patient analyses in Chapters 5 and 6, the dependence of projective

measurements on surface geometry was investigated as well. The unique patient cohort that enabled this analysis, and the virtual endoscopic methods used to make projective measurements, are described in Section 7.2. The results of the analysis are presented in Section 7.3, and a discussion of their interpretation and the strengths and weaknesses of this study is presented in Section 7.4.

7.2 Methods

7.2.1 Patient cohort

Nineteen head and neck cancer patients who received radiotherapy at MD Anderson Cancer Center were retrospectively selected for this study, which was approved by the Institutional Review Board. A simulation CT had been acquired for each patient for treatment planning. These scans were acquired using a 48 or 50 cm field of view and 0.25 or 0.3 cm slice thickness. The treatments had been delivered in 31-35 daily fractions, and on each day another CT scan had been acquired prior to irradiation using a CT-on-rails in the treatment room. These scans were acquired using the same parameters as the planning CTs. In the daily CTs, the head, neck, and shoulders were secured in molded thermoplastic masks, so the patient positioning was as close as possible to that in the simulation CT. These scans were used to evaluate the influence of daily variations in non-rigid anatomy on projective measurements.

In addition to the daily scans, pre-treatment diagnostic CT scans were acquired for thirteen of the patients. Six patients were omitted due to the large 5-mm slice thickness

of their diagnostic scans, which would have a negative effect on the resolution of virtual endoscopic images. These scans were acquired using a 25 or 26 cm field of view and 0.1 or 0.25 cm slice thickness. In the diagnostic scans, the thermoplastic masks were not used, and the patients' head, neck, and shoulders were not positioned in any particular way. These scans were used to investigate the impact of patient positioning. This scenario does not exactly model the differences between supine and seated positions, but it can provide some insight into the importance of reproducible positioning for projective measurements.

7.2.2 Virtual endoscopy

Virtual endoscopic images were rendered using the optimized lighting model discussed in Section 6.2.4. Due to the large number of CT scans used for this analysis, a fully-automated method was used to segment the CTs and create the virtual endoscopy surface meshes. First, each slice was converted to a binary image using a density threshold of 0.8 g/cm^3. This is higher than the 0.6 g/cm^3 threshold used in the patient analysis of Chapter 6. This choice was made based on the observation that lower thresholds closed off narrow passages in the nasal cavity and reduced anatomical detail for fine structures such as the epiglottis. No endoscopic videos were used in this analysis, so the choice of threshold was somewhat arbitrary.

After converting the slice to a binary image, morphological analysis was used to find the outlines of each object in the image. The coordinates of these outlines were used to create the surface mesh with the method described in Section 3.3.1. This

segmentation method included additional structures in the surface mesh from outside the airways. The additional structures did not affect the virtual endoscopic images, as the segmentation of the airway surface itself is the same as what would be obtained with the semi-automatic method used in Sections 5.2.2 and 6.2.2. The optimized lighting model presented in Section 6.2.4 was used to render virtual endoscopic images, all of which were smoothed with a 3 x 3 Gaussian kernel with $\sigma = 1$ pixel.

7.2.3 Virtual endoscope paths

The anatomy of the head and neck varies considerably in terms of the distances, structures, and muscle motion seen via endoscopy. In order to characterize the different anatomical regions, a virtual endoscope path from the nasal cavity to the glottis was created in the planning CT surface mesh for each patient. The camera positions were spaced 5 mm apart along the path, and the number of positions per patient ranged from 13 to 28. The large range is due to the omission of regions that were inaccessible to the virtual endoscope in certain patients. This occurred when the patients had large tumors that blocked part of the pharynx, and when the narrow passages of the anterior nasal cavity were closed off by the smoothing applied to the CT-space threshold contours when creating the surface meshes.

Due to differences in patient anatomy and positioning of the head and neck in the CT, each patient's virtual endoscope path is unique and it is difficult to make inter-patient comparisons of projective measurements at a given point on the path. To allow for some degree of inter-patient comparison, the points in the paths were classified into

three anatomical regions: the nasal cavity, the nasopharynx and oropharynx, and the hypopharynx and larynx. The coordinate ranges used for these classifications were chosen manually for each patient, and an example is shown in Figure 34. An example of the virtual endoscopic images in one patient's path is shown in Figure 35.

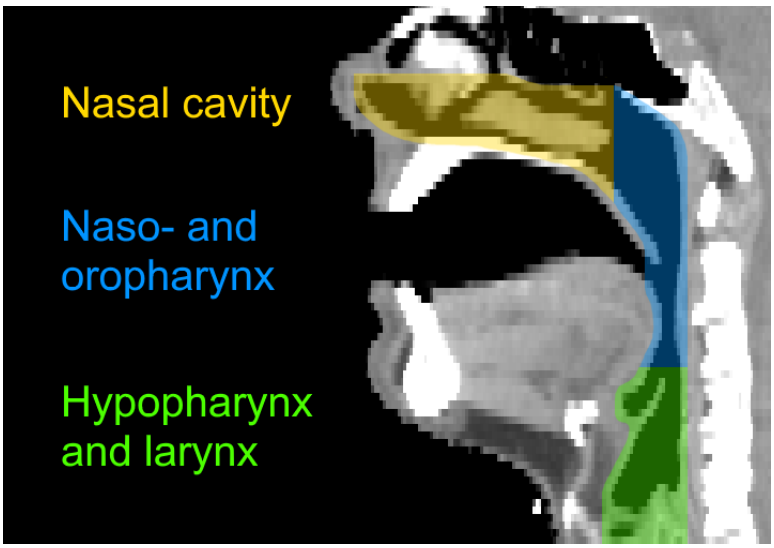

Figure 34: An example of the anatomical regions used to in the virtual endoscope path.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, R. Wendt III, B. M. Beadle, A. Rao, X. A. Wang, and L. E. Court. "The influence of non-rigid anatomy and patient positioning on endoscopy-CT image registration in the head and neck." *Medical Physics* (2017).

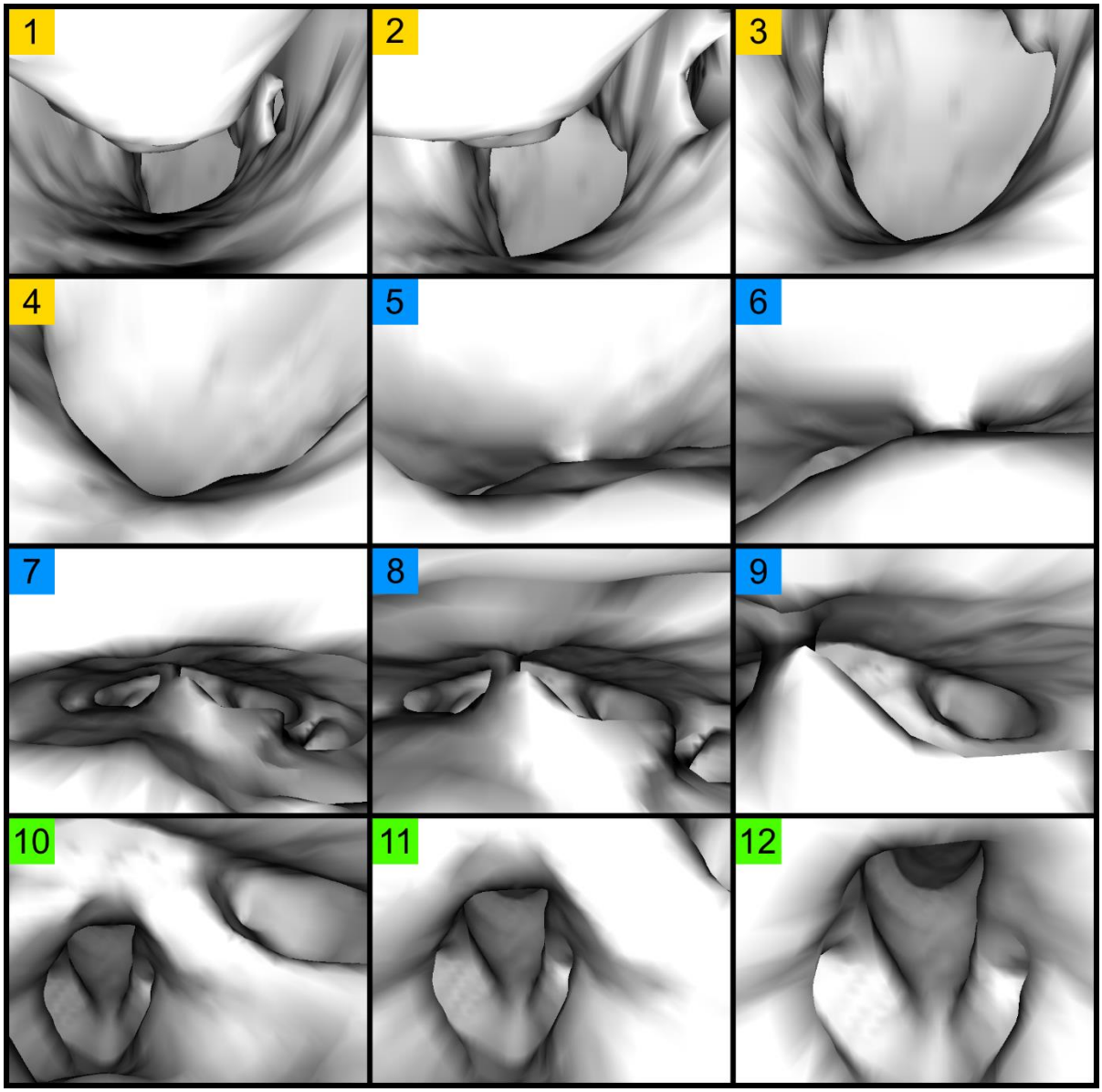

Figure 35: An example of the images in the virtual endoscope path. The colors of the numbered squares indicate the anatomical regions: nasal cavity (yellow), naso- and oropharynx (blue), and hypopharynx and larynx (green). For conciseness, every other image has been omitted, so adjacent virtual endoscope positions are 10 mm apart. In 1-4, the posterior wall of the pharynx and the floor of the nasal cavity are visible. In 5 and 6, the virtual endoscope is in the superior pharynx, and in 7-9 the epiglottis is visible. In 10-12, the virtual endoscope is approaching the glottis.

This figure has been reproduced and modified from the following publication:
W. S. Ingram, J. Yang, R. Wendt III, B. M. Beadle, A. Rao, X. A. Wang, and L. E. Court. "The influence of non-rigid anatomy and patient positioning on endoscopy-CT image registration in the head and neck." *Medical Physics* (2017).

7.2.4 Measurements of projective errors

The influences of non-rigid anatomy and patient positioning differences on the uncertainty of endoscopy-CT image registration were quantified by the range of 3D distance errors between reference projective measurements taken on the planning CTs and test projective measurements taken on the daily and diagnostic CTs. These measurements were taken using the world transform described in Section 4.4.2. The daily CTs were used to evaluate the impact of non-rigid anatomy, and the diagnostic CTs were used to evaluate the impact of patient positioning. The following steps, which are summarized in Figure 36, describe the calculation of the two sets of distance errors. "Reference" will be used to refer to the planning CT, and "test" will be used to refer to the daily or diagnostic CT:

1. Take the world transform at each position on the virtual endoscope path in the reference mesh. These sets of 3D points comprise the reference measurements.
2. Calculate the corresponding virtual endoscope path in the test mesh.
    a. Deformably register the reference CT to the test CT using validated in-house software[76].
    b. Assign each path position the average deformation vector of all surface voxels within 1 cm. This is done because the registration may be less reliable within the air cavity, so the nearby surface voxels may provide a more realistic deformation for the virtual endoscope.

3. Optimize the view direction at each position in the corresponding virtual endoscope path in the test mesh. This is done to reduce errors that are caused by the deformation of the path rather than by the effects of non-rigid anatomy or patient positioning.
    a. Calculate the similarity between the virtual endoscopic image at a given path position in the reference mesh and the virtual endoscopic image at the corresponding path position calculated in step 2. The similarity metric MI_{grad}, defined in Section 4.1, was used here as well.
    b. Search for the view direction in the test mesh that maximizes this similarity using the Nelder-Mead simplex method.
4. Take the world transform in the test mesh at each position on the virtual endoscope path calculated in steps 2 and 3. These sets of 3D points comprise the test measurements.
5. Using the deformation field obtained in step 2a, transform the reference measurements taken in step 1 into the coordinate space of the test measurements.
6. Calculate the distance errors between the transformed reference measurements obtained in step 5 and the test measurements taken in step 4.

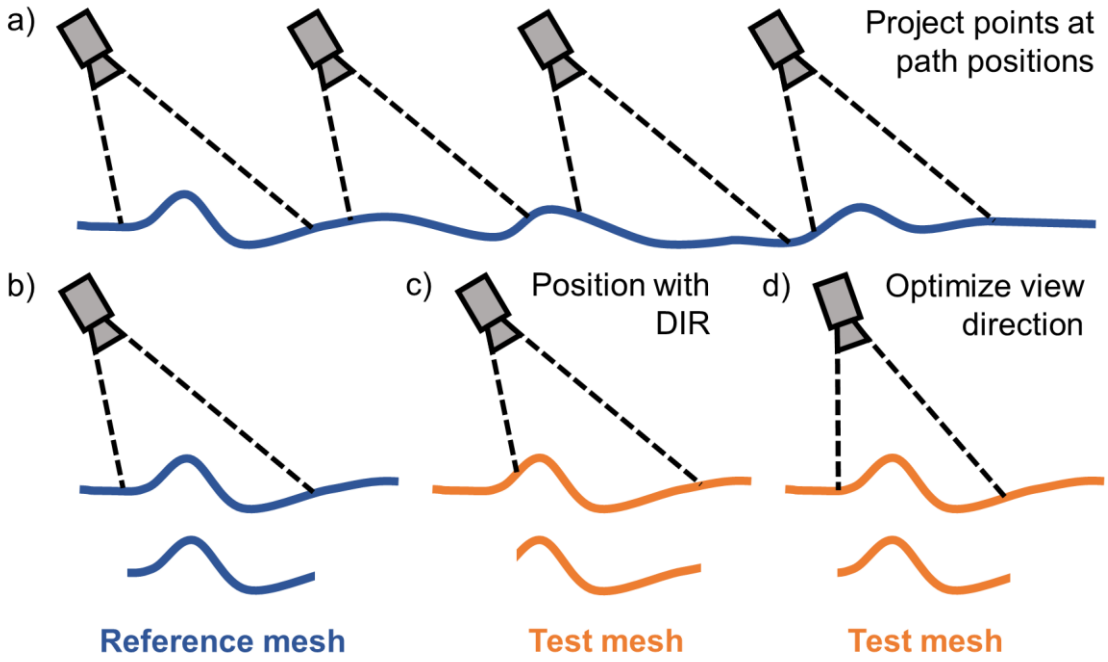

Figure 36: A schematic illustrating the measurement of projective errors. a) The virtual endoscope is placed at each position on its path through the reference mesh, which is represented by the blue line. b) At each position on the path, projective measurements are made, generating a set of CT-space reference points. c) The virtual endoscope is placed at the corresponding position in the test mesh, which is represented by the orange line. This position is calculated using deformable registration between the planning CT and the daily or diagnostic CT. d) The virtual endoscope's view direction is optimized to match the appearance of the reference virtual image. Then, projective measurements are made, generating a set of CT-space test points. Finally, distance errors are calculated between the reference and test points.

This figure has been reproduced from the following publication:
W. S. Ingram, J. Yang, R. Wendt III, B. M. Beadle, A. Rao, X. A. Wang, and L. E. Court. "The influence of non-rigid anatomy and patient positioning on endoscopy-CT image registration in the head and neck." *Medical Physics* (2017).

7.2.5 The impact of scene geometry

A simple geometric analysis shows that projective measurements should be influenced by the distance between the camera and the surface, and the angle at which it views the surface. This is illustrated in Figure 9. The results presented in Sections 5.3.3 and 6.3.2 did not find any strong correlations between these quantities and projective error, although excluding points with large measurement distances did reduce projective errors in patients. To further characterize geometric effects, the world transforms for the reference virtual images were used to compute measurement angles and distances for each point. The correlation between these values and projective measurement errors were calculated for each path position in the daily CTs. In addition to these quantities, edge masks were created for both the reference and test virtual images using a distance threshold of 2 mm and morphological erosion with a 3 x 3 structuring element for five iterations. The masks were used to investigate the exclusion of regions of high uncertainty when making projective measurements. See Section 4.4 for more details on the world transform, measurement angles, and edge masks.

7.2.6 The impact of the interval between CT acquisitions

In an ideal application of projective mapping for endoscopy-CT registration, the endoscopic examination and the CT would be acquired as close in time as possible to minimize anatomical differences between the two modalities. This is especially true for

patients with large tumors that shrink during the course of radiotherapy. One application of endoscopy-CT registration would be to assess the dose to various anatomical structures seen via endoscopy at follow-up examinations after the course of radiotherapy. In this context, it is important to know if anatomical changes over the course of radiotherapy can affect registration to the planning CT. To investigate this, the projective errors for the treatment-room CT on day 1 were compared to those for day 30.

7.3 Results

7.3.1 The influence of non-rigid anatomy

For each patient, the median projective distance error between the reference and test world transforms was calculated for each path position in each of the CTs from days 1-5. These measurements were grouped according to their anatomical region, and the results are presented in Table 17 and Figure 37. The nasal cavity had the smallest median projective errors overall, with the average ranging from 1.6 ± 1.1 mm to 1.9 ± 2.2 mm over the five CTs. The naso- and oro-pharynx had the largest errors, with averages ranging from 2.8 ± 2.1 mm to 3.2 ± 2.9 mm. The hypopharynx and larynx were in between, with averages ranging from 1.9 ± 1.0 mm to 2.3 ± 2.0 mm.

There were no significant differences between the five CTs ($p > 0.05$ in each region using the Kruskal-Wallis H-test). The differences between regions were all significant ($p < 0.0001$ for each pair of regions using the Mann-Whitney U-test). The trend between

regions was observed in most patients, but there was a high degree of variability between them. This is illustrated in Figure 38, which shows the projective errors at each path position for three patients.

Table 17: Averages of median projective errors in the three anatomical regions for each daily CT.

CT	Nasal cavity	Naso- and oropharynx	Hypopharynx and larynx
Day 1	1.6 ± 1.1	2.9 ± 2.3	1.9 ± 1.0
Day 2	1.6 ± 1.2	2.9 ± 2.6	2.1 ± 1.1
Day 3	1.9 ± 2.2	2.8 ± 2.1	2.1 ± 1.5
Day 4	1.9 ± 1.6	3.2 ± 2.9	2.3 ± 2.0
Day 5	1.9 ± 1.3	3.1 ± 2.3	2.1 ± 1.2

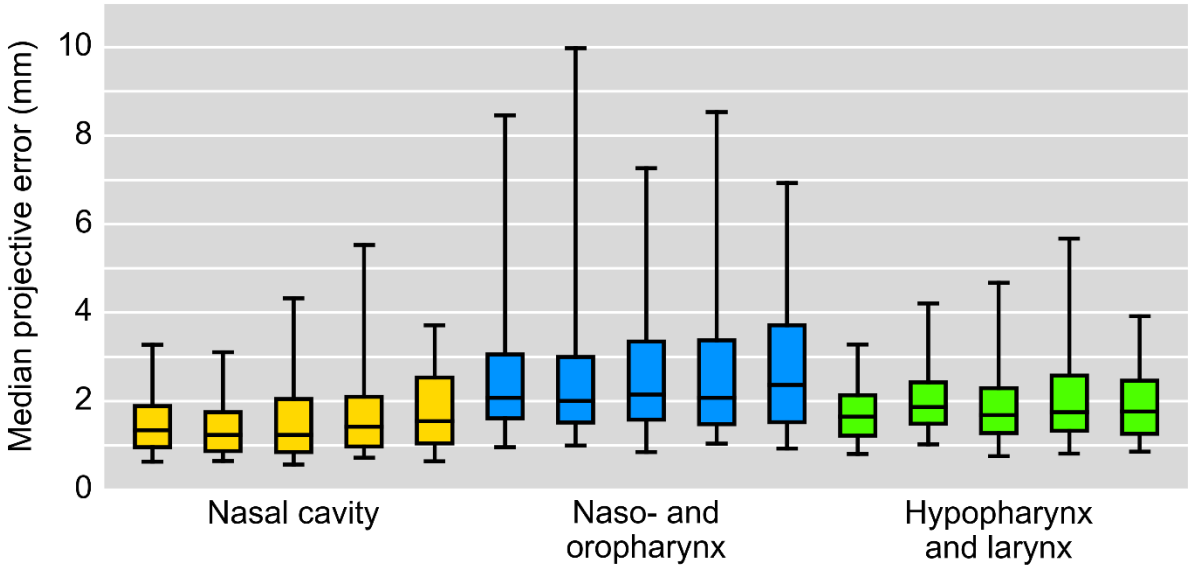

Figure 37: Median projective errors in the three anatomical regions for each daily CT. The boxes show the median and the quartiles, and the whiskers show the 5th and 95th percentiles. The smallest errors were found in the nasal cavity, and the largest were found in the naso- and oropharynx. The differences between regions were all statistically significant.

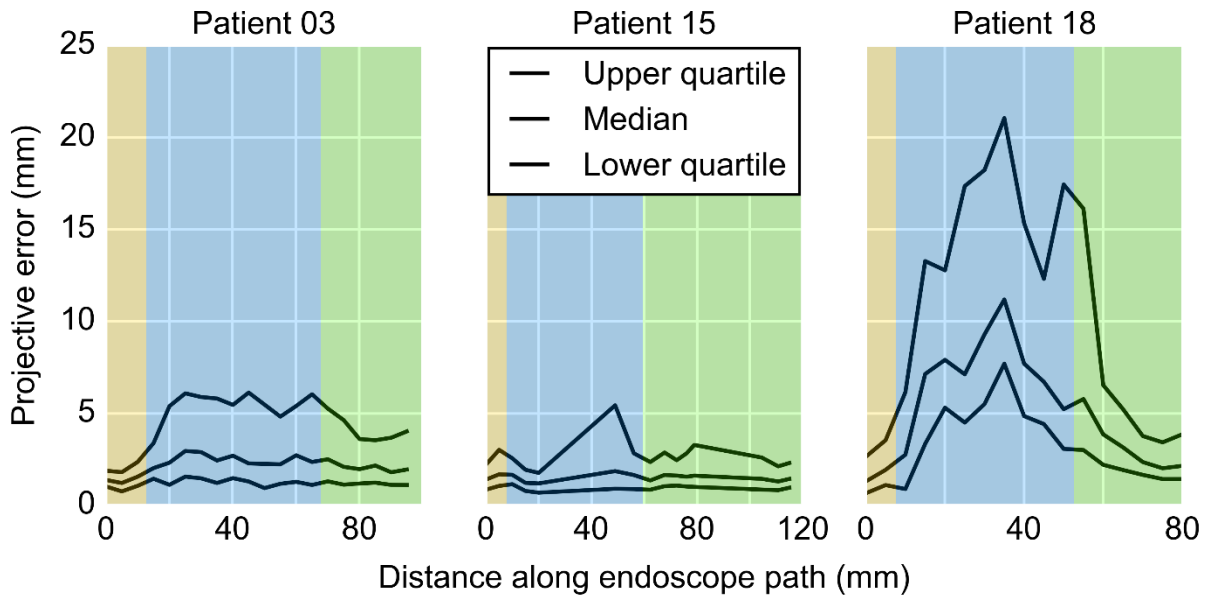

Figure 38: Projective errors at each virtual endoscope path position for three patients. These plots illustrate the inter-patient variability in trends between the different anatomical regions, which are indicated by the color shading. In Patient 03, the trend between regions is clear, with very small errors in the nasal cavity that become larger in the naso- and oropharynx, and then diminish in the hypopharynx and larynx. In Patient 15, the errors remain small throughout the path, but there is no clear trend. Patient 18 shows some of the largest overall errors, particularly in the naso- and oropharynx.

7.3.2 The influence of patient positioning

For each patient, the median projective distance error between the reference and test world transforms was calculated for each path position in the diagnostic CT. These measurements were grouped according to their anatomical region, and the results are presented in Figure 39. The nasal cavity had the smallest median projective errors overall, with an average of 3.5 ± 2.6 mm. The naso- and oropharynx had the largest

errors, with an average of 4.3 ± 2.9 mm. The hypopharynx and larynx were in between with an average of 4.2 ± 3.8 mm, but the results in this region were highly variable with very large outliers.

The projective errors in the diagnostic CTs were larger than those in the daily CTs in each anatomical region, and each of these differences was statistically significant ($p < 0.0001$ using the Mann-Whitney U-test). When comparing the diagnostic CT errors between regions, only the difference between the nasal cavity and the naso- and oropharynx was significant ($p < 0.01$ using the Mann-Whitney U-test).

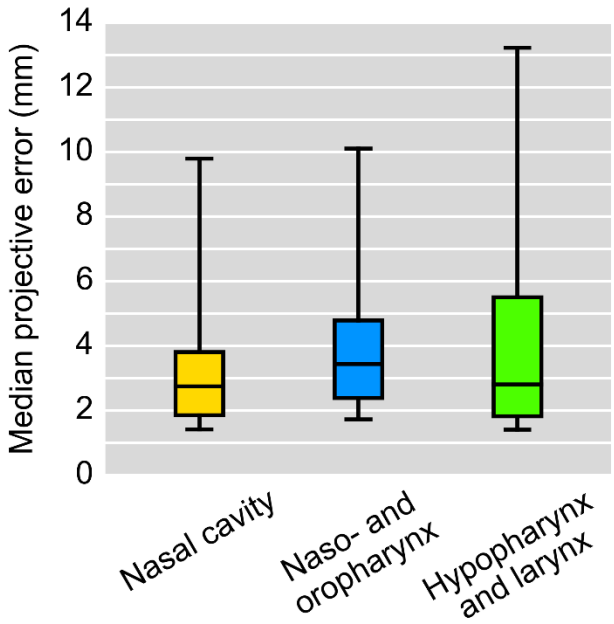

Figure 39: Median projective errors in the three anatomical regions for the diagnostic CTs. The boxes show the median and the quartiles, and the whiskers show the 5th and 95th percentiles. The smallest errors were found in the nasal cavity, and the largest were found in the naso- and oropharynx. The errors in the hypopharynx and larynx were highly variable with large outliers. Only the difference between the nasal cavity and the naso- and oropharynx was statistically significant.

7.3.3 The influence of scene geometry

The dependence of projective error on measurement angle and measurement distance was similar to that found in the analyses of Chapters 5 and 6. Spearman's rho was calculated for both of these quantities at each path position in the Day 1-5 CTs. The results were highly variable from position to position, and a summary is presented in Table 18. On average, a very weak correlation was found between angle and error, and a moderate correlation was found between distance and error. A better depiction of the influence of measurement distance is presented in Figure 40, which shows the median projective errors in the Day 1-5 CTs for all patients and all path positions with progressive distance cutoffs applied. The errors were reduced by each cutoff, and the difference between every adjacent cutoff is significant ($p < 0.0001$ using the Wilcoxon signed-rank test).

Edge masks were also successful in reducing projective errors, and the results are given in Table 19. In general, the edge masks removed approximately 10% of points from the world transforms, but reduced the median error for 99.5% of the virtual images tested. The magnitude of this reduction was only 0.2 mm on average. This is likely due to the fact that the edge mask excludes large outliers where the reference and test measurements fall on either side of an occluding edge. The median error for a virtual image would not be very sensitive to exclusion of these outliers.

Table 18: Correlation values between projective error and measurement angle and measurement distance. Spearman's rho was calculated at each path position in the Day 1-5 CTs for all patients, and the averages are presented here. All $p < 0.0001$ except for a few where the correlation value was very close to 0.

CT	Measurement angle correlation	Measurement distance correlation
Day 1	0.16 ± 0.25	0.41 ± 0.32
Day 2	0.17 ± 0.26	0.40 ± 0.32
Day 3	0.17 ± 0.25	0.42 ± 0.32
Day 4	0.13 ± 0.25	0.43 ± 0.32
Day 5	0.18 ± 0.24	0.41 ± 0.31

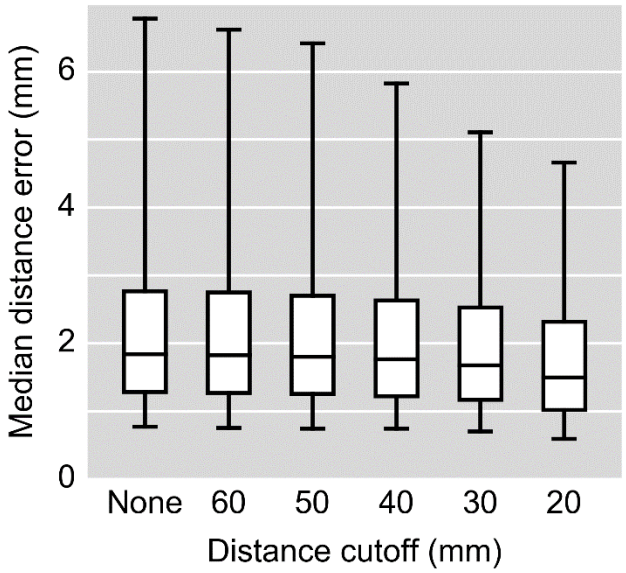

Figure 40: Reduction is projective errors when distant points are excluded. These box plots show the median projective errors in the Day 1-5 CTs for all patients and path positions with progressive measurement distance cutoffs applied. For example, in the 60-mm plot, all points with a measurement distance larger than 60 mm have been excluded. The boxes show the median and quartiles, and the whiskers for the 5th and 95th percentiles. The reduction in error is significant between all adjacent cutoffs ($p < 0.0001$ using the Wilcoxon signed-rank test).

Table 19: Projective errors with and without edge masks applied. All values were calculated at each path position in the Day 1-5 CTs for all patients, and the averages are presented here. Column 2 gives the percentage of pixels removed by the edge mask. Column 3 gives the original median projective errors, and Column 4 gives the errors after the masks were applied. The median error was reduced for every frame by applying the edge mask, and the magnitude of the reduction is given in the fifth column. All error values are in mm. The reduction in error was statistically significant for all CTs ($p < 0.0001$ using the Wilcoxon signed-rank test).

CT	Masked percentage	Unmasked error	Masked error	Error reduction
Day 1	10.4 ± 4.3	2.3 ± 1.8	2.1 ± 1.7	0.2 ± 0.2
Day 2	10.4 ± 4.3	2.3 ± 2.0	2.2 ± 1.8	0.2 ± 0.3
Day 3	10.4 ± 4.6	2.4 ± 2.0	2.2 ± 1.7	0.2 ± 0.5
Day 4	10.4 ± 4.1	2.6 ± 2.4	2.5 ± 2.3	0.2 ± 0.3
Day 5	10.2 ± 4.3	2.5 ± 1.9	2.3 ± 1.7	0.2 ± 0.3

7.3.4 The impact of the interval between CT acquisitions

The median projective errors for all patients in the Day 1 CT were compared to those in the Day 30 CT in each anatomical region. The errors in the Day 30 CT were larger than those in the Day 1 CT. The average increase was 0.5 ± 2.0 mm in the nasal cavity, 1.5 ± 3.4 mm in the naso- and oropharynx, and 2.3 ± 3.6 mm in the hypopharynx and larynx. The increase in projective errors was statistically significant in the naso- and oropharynx and in the hypopharynx and larynx ($p < 0.0001$ using the Wilcoxon signed-rank test).

7.4 Discussion

7.4.1 Summary

A patient cohort was used to investigate the influences of non-rigid anatomy and patient positioning on the uncertainty of virtual endoscopic projective measurements. The patients in this cohort had daily CT imaging in the same position as the planning CT during the course of radiotherapy. With these scans, projective measurements were compared to those in the planning CT to investigate the influence of non-rigid anatomy. Diagnostic scans were acquired as well, in which the patients were not positioned in any particular way. With these scans, projective measurements were compared to those in the planning CT to investigate the influence of patient positioning.

Projective measurements were taken along a virtual endoscope path from the nasal cavity to the glottis. To facilitate inter-patient comparison, the path points for each patient were classified into three anatomical regions: the nasal cavity, the naso- and oropharynx, and the hypopharynx and larynx. Reference measurements were taken in the planning CT surface mesh. Test measurements were taken in the daily or diagnostic CT surface meshes after placing the virtual endoscope at the corresponding path position with deformable CT-CT registration. At each path position, the median distance error between the reference and test measurements was calculated to quantify projective uncertainty.

In the daily CTs, the median projective errors were smallest in the nasal cavity, and largest in the naso- and oropharynx. This trend between regions was observed in most

patients, but the results were highly variable. In the diagnostic CTs, median projective errors were larger in all regions than those in the daily CTs, and the trends between regions were not as pronounced. An investigation of the dependence of projective errors on surface geometry found weak correlations between measurement angle and error, and moderate correlations between measurement distance and error. However, projective error was shown to decrease when points with large measurement distances were excluded. The edge masks were also demonstrated to reduce projective error, showing that proximity to occluding edges may be an important consideration when making projective measurements with virtual endoscopy.

7.4.2 Projective measurements without real endoscopic images

No real endoscopic images were used in this study; all comparisons were made between virtual endoscopic images rendered from different CTs. This choice was made in part because endoscopic examinations are not routinely recorded and saved as part of the patient's medical record at MD Anderson Cancer Center, so real endoscopic images were not available for the patient set used in this study. One advantage of this approach is that two images of the same modality are used when image similarity is calculated to optimize the virtual endoscope's view direction (step 3 in Section 7.2.4). If endoscopic examinations were available, another approach would be to use a registration method such as path-based volumetric search to find the registered endoscope coordinates in the reference and test CTs. However, this would include the additional uncertainty introduced by calculating image similarity between two different

modalities. Given that endoscopic videos were not available for the patients in this dataset, there was no way to quantify this additional uncertainty. By using only virtual endoscopic images, the methods in this study did not quantify the full uncertainty that would be expected in a clinical application of endoscopy-CT registration, but they have the advantage of isolating the influence of anatomical differences between CT scans.

7.4.3 The role of deformable CT registration

In this study, the daily and diagnostic CTs were deformably registered to the planning CT in order to place the virtual endoscope at the path positions in the daily and diagnostic CT surface meshes (step 2 in Section 7.2.4). This step is necessary because placing the virtual endoscope at the same location in the two meshes means that any differences between the projective measurements were the result of anatomical differences between the CTs, rather than differences in virtual endoscope placement. It is important to note that the airway segmentation used to create the surface meshes themselves was performed on each CT prior to applying the deformable registration (see Section 7.2.2). This means that any anatomical differences between the CTs were retained in the corresponding virtual endoscopic images.

One challenge is that the airway surfaces are not rigid, so given a position in one CT, the same position in another CT is not exactly defined in the context of projective measurements via virtual endoscopy. Rigid registration would maintain the same position with respect to the overall anatomy of the head and neck, but if the airway surfaces deform, the virtual endoscope could be placed several mm closer to or further

from the surface, which can cause a very noticeable change in a virtual endoscopic image. For this reason, deformable registration was chosen and the virtual endoscope was assigned the average deformation vector of all surface points within 1 cm, which should maintain the same position relative to the surface of the airway. This method of placing the virtual endoscope, and the subsequent optimization of its view direction to match the planning CT virtual endoscopic image (step 3 in Section 7.2.4), reduce errors in the projective measurements caused by any source other than the anatomical variations that this study sought to investigate.

7.4.4 Trends in projective errors

The effects of surface geometry were similar in this study to those observed in the analyses of Chapters 5 and 6. The influence of measurement distance illustrated in Figure 9 was demonstrated by the data, but the influence of measurement angle was not. There is no clear reason for this, but it is possible that it is due to the surface meshes not being completely smooth, both due to the anatomical structure and the density threshold segmentation. This means that the measurement angle may only describe a very small local neighborhood of the surface, so the simple geometrical assumptions made in Figure 9 do not hold. Another possibility is that the influence of measurement distance is much stronger for points far from the camera, preventing observation of a correlation between projective error and measurement angle.

This study also found that a longer interval of time between the planning CT and the daily CT increases the projective error. There were no significant differences

between the measurements from the Day 1-5 CTs, but the Day 30 CT had larger projective errors than those from the Day 1 CT. This suggests that minimizing the length of time between the endoscopic examination and the CT scan will be important for clinical application of projective mapping for endoscopy-CT registration. It also suggests that uncertainties would be larger for registration of follow-up endoscopic examinations performed after the course of radiotherapy. However, the data in this study do not support any particular recommendation for an acceptable interval of time.

One of the main goals of this study was to investigate the difference between the influences of daily anatomical variations and patient setup on projective uncertainty. The results show that patient setup has a larger impact, but not by much. One caveat to this difference is that the daily CT data set was larger, with five daily CTs for each of nineteen patients, whereas there was only one diagnostic CT for each of thirteen patients. Another caveat is that the patients' positions in the diagnostic scans are not the same as those in endoscopic examinations, so this data set is an imperfect representation of the influence of patient positioning. No attempt was made to model the effect of a seated position in the virtual endoscopic images. It would be challenging to do this in a meaningful way without seated-position CT scans, from which surface meshes could be created and compared to those from supine CT scans. However, because the diagnostic scans used in this study were acquired without the thermoplastic masks, they do provide some insight into the importance of positioning the patient's head and neck in exactly the same way for acquisition of the CT and the endoscopic video.

7.4.5 The clinical context of projective errors

The data presented in this study are not particularly meaningful without the context of clinically acceptable levels of uncertainty for image registration. There is no definitive acceptable level of uncertainty, but a multi-institution study of 21 deformable registration algorithms found average point errors between the phases of 4DCTs up to 3.0 mm in the lungs and 6.7 mm in the abdomen[77]. Another study of four deformable registration algorithms in the head and neck found overall surface distance errors of 4.6 mm in the best case[78]. A large majority of the median projective errors presented in this study are comparable to or less than these published values.

Another important consideration is the impact of these projective errors on the potential use of endoscopy-CT registration for tumor delineation during the treatment planning process. This is a difficult topic to evaluate because endoscopic images have never been used to do this clinically, and projective mapping via virtual endoscopy is fundamentally different than the volumetric registration that is commonly used in radiotherapy. Inter-observer variability in manual target delineation is not negligible in the head and neck, with one study finding a standard deviation of GTV delineation for nasopharyngeal cancer greater than 4 mm[79]. This is comparable to the median projective errors found in this study. However, this does not address the large outliers that were present in the data in this study and the analyses of Chapters 5 and 6. It is likely that many of these will be easy to identify and reject in the context of target definition by inspecting for large discontinuities between nearby projected points. The reduction in error when edge masks were applied shows that surface geometry may be

useful for avoiding some of these outliers as well. Further insight could be gained by studying the projective mapping of object contours that have CT-space ground truths.

7.4.6 Caveats to the projective errors presented in this study

There are several aspects of the methods used in this study that may have increased the reported errors beyond what could be expected in a fully-developed clinical application of projective mapping for endoscopy-CT registration. The first is that the deformed virtual camera path positions were held fixed and only their view directions were optimized. This may have the effect of placing the camera closer to walls or obstacles than it is in the reference image, which would lead to larger errors. In a clinical application, the position of the virtual camera would be optimized along with the view direction.

The second aspect is that the meshes and their corresponding virtual images are held rigid for the projection of pixel locations. This causes misalignments between the two images, resulting in large projective errors where the points fall on either side of an occluding edge. One strategy to reduce these errors is to allow the meshes to deform during the optimization process, or to warp the test virtual image onto the reference virtual image before projecting the pixel locations. However, no method has been investigated to accomplish this.

It is important to note that projective mapping is only a technique to transfer spatial information between an endoscopic video frame and a CT scan, and this study did not consider the uncertainty of the registration method used to find the CT-space

endoscope coordinates needed to render the virtual image. However, the results of this study are not dependent on the registration method. Even with the additional sources of error described in the previous paragraph, the uncertainty of projective mapping in the head and neck introduced by anatomical variations and patient positioning was found to be comparable to acceptable levels of uncertainty in other forms of medical image registration. This demonstrates that projective mapping is a promising technique for endoscopy-CT image registration in the head and neck.

8

Image processing parameters

8.1 Introduction

In the preceding chapters, the performance of two different endoscopy-CT registration algorithms has been evaluation, and the influences of scene geometry and anatomical variations on the registration uncertainty have been investigated. In these analyses, little attention has been given to the image processing parameters that can modify the endoscopic video frames and the virtual endoscopic images. These parameters may have significant impacts on registration accuracy, and if endoscopy-CT is to be developed into a robust clinical tool, it will be important to characterize them.

This chapter presents two separate but related analyses. The first is an evaluation of a variety of image processing parameters that may influence endoscopy-CT registration, including similarity measures, Gaussian and edge-preserving smoothing, downsampling, and masking to avoid the structural disparities described in Section 6.4.3. The second analysis is investigation of the importance of having an exact camera calibration model for the endoscope, with the goal of determining if endoscopy-CT registration would be feasible in settings where the calibration model is unknown. This situation could arise for retrospective analyses of archived video, and for prospective

patient evaluation if the endoscopic examination was performed at an outside institution. The calibration analysis is included in this chapter because calibration parameters are image processing parameters, as they determine the removal of distortion from endoscopic video frames and the view angle used to render virtual endoscopic images.

The main goal of these analyses was to determine how the image processing parameters affect the accuracy of registered endoscope coordinates. Rather than using the path-based volumetric search registration method (see Section 4.3), a slightly different approach was used to better characterize how the parameters affect the ability of a simplex-based optimization routine to finely distinguish virtual endoscope coordinates near the ground truth. With the calibration analysis, a secondary goal was to investigate the effect that the parameters have on projective measurement error, independent of the accuracy of registered endoscope coordinates. These methods are explained in detail in Section 8.2. The results of the analyses are presented in Section 8.3, followed by a discussion in Section 8.4.

8.2 Methods

8.2.1 Patient dataset

The patient dataset described in Section 6.2.3 was used for these analyses as well. It consisted of 46 registration frames from four endoscopic examinations of three patients. As in the analyses of Chapters 5, 6, and 7, the preprocessing for all registration

frames included conversion to grayscale and smoothing with a 3 x 3 Gaussian kernel with σ = 1 pixel. For frames from patient MDA1, it also included deinterlacing by replacing every other row with bilinear interpolation and distortion removal as described in Section 3.4. Frames from patients PMH1 and PMH2 were recorded with distortion already removed. Ground-truth virtual endoscope coordinates were obtained for each frame by camera resectioning (see Section 4.2.2) followed by manual refinement to visually align anatomical structures. The optimized lighting model presented in Section 6.2.4 was used to render virtual endoscopic images, all of which were smoothed with a 3 x 3 Gaussian kernel with σ = 1 pixel.

8.2.2 Volumetric grid search near the ground truth

Registered endoscope coordinates were computed for each registration frame using a method similar to the path-based volumetric search (see Section 4.3), but designed to sample the neighborhood of the ground-truth coordinates more densely. The general approach was to place the virtual endoscope at each point in a grid, optimize its view direction to match the registration frame, keep the best overall result, and refine it to get the best possible registered endoscope coordinates. The steps of this volumetric grid search are detailed below:

1. Select a registration frame F_{reg}.
2. Create a 9 x 9 x 9 grid of points centered on the ground-truth position with ±10-mm extent and 2.5-mm spacing. Discard any points outside the surface mesh.

3. Initialize the virtual endoscope's view direction at each grid point.
    a. Place the virtual endoscope at the ground-truth coordinates.
    b. Take the projective measurement of the pixel at the center of the image to get the grid focal point $(x_f,\ y_f,\ z_f)$.
    c. Place the virtual endoscope at the grid point with the view direction of the ground-truth coordinates.
    d. Adjust the virtual endoscope's view direction so the grid focal point is in the center of the image.
4. Starting from each grid point, search for the virtual endoscope coordinates that maximize the similarity between F_{reg} and the virtual image. In this step, the virtual endoscope's position $(x,\ y,\ z)$ is fixed at each grid point, and the view direction $(\theta_x,\ \theta_y,\ \theta_z)$ is optimized.
5. Keep the best overall result from the grid search and refine it by searching again for the virtual endoscope coordinates that maximize the similarity between F_{reg} and the virtual image. In this step, all six endoscope coordinates are optimized.

In steps 4 and 5, the virtual endoscope's coordinates were optimized to maximize the similarity between the registration frame and virtual endoscopic images. The similarity measures used for this process are discussed in Section 8.2.3. The optimizations were performed using the Nelder-Mead simplex algorithm. In step 4, the scale of the search space was set using

$$\Delta_{simplex} = (5\ deg, \quad 5\ deg, \quad 5\ deg) \tag{38}$$

A small search space was used for this step because the view initialization in step 3 should put the virtual endoscope's view direction near the optimum. For the refinement in step 5, an even smaller scale was set for the search space using

$$\begin{aligned} &\Delta_{simplex} \\ &= (1.25\ mm, \quad 1.25\ mm, \quad 1.25\ mm, \quad 2.5\ deg, \quad 2.5\ deg, \quad 2.5\ deg) \end{aligned} \tag{39}$$

See Section 4.1 for more details on the Nelder-Mead algorithm and the scale vector $\Delta_{simplex}$.

The output of step 5 is the desired registered endoscope coordinates. The accuracy of the registered coordinates was quantified by taking the world transforms of the ground-truth and registered virtual images and computing the median distance error between the two sets of CT-space points. See Section 6.2.7 for the rationale behind using the median projective error to quantify registration accuracy, and see Section 4.4 for details on the world transform. The volumetric grid search was used to evaluate the image processing parameters described in the following sections. First, the search was run for a variety of image similarity measures to identify the one that resulted in the best registration accuracy. Then, the search was run using that similarity measure with additional preprocessing parameters to identify those that could further improve registration accuracy.

8.2.3 Similarity measures

The two most fundamental image processing components of endoscopy-CT registration are the virtual endoscopy lighting model and the similarity measure used to compare video frames and virtual images. The lighting model was optimized by histogram comparison between the two modalities (see Section 6.2.4), so the next step is to determine the best similarity measure. This was accomplished by running the volumetric grid search with a variety of similarity measures and choosing the one that resulted in the smallest overall projective distance errors relative to the ground-truth endoscope coordinates.

In the phantom and patient analyses of Chapters 5, 6, and 7, the similarity measure MI_{grad} was chosen based on the characteristics of mutual information and the presence of structural edges in both video frames and virtual images. Nine more similarity measures were chosen for this analysis. Eight were selected based on their performance registering images with changes in illumination and images of different modalities[80], and one was designed specifically for virtual endoscopy[32]. Many of these are similarity measures rather than dissimilarity measures, so their values were negated for use as the objective function for simplex optimization. The similarity measures are defined below. In the following equations, f_i and v_i are pixel intensity values in the video frame and virtual image, $\bar{f}$ and $\bar{v}$ are the mean intensities, σ_f and σ_v are the standard deviations of intensities, and n is the number of pixels in one of the images.

1. Pearson correlation coefficient

$$= \frac{\sum_i (f_i - \bar{f})(v_i - \bar{v})}{\sqrt{\sum_i (f_i - \bar{f})^2}\sqrt{\sum_i (v_i - \bar{v})^2}} \tag{40}$$

2. Correlation ratio

$$= \sqrt{1 - \frac{1}{n}\sum_{i=0}^{255} n_i \sigma_i^2} \tag{41}$$

Here n_i is the number of pixels in the video frame with intensity i, and σ_i^2 is the variance of intensities in the virtual image corresponding to the intensity i in the video frame.

3. Spearman's rho

$$= 1 - \frac{6\sum_i [R(f_i) - R(v_i)]^2}{n(n^2)} \tag{42}$$

Here $R(f_i)$ and $R(v_i)$ are the ranks of f_i and v_i in the video frame and virtual image, respectively. When using this metric, the images were converted to floating point values and smoothed with a 3 x 3 Gaussian kernel with σ = 1 pixel to prevent ties in the ranks.

4. Material similarity

   This is a complex metric based on identification of peaks in the joint probability distribution of the video frame and virtual image. Its definition can be found elsewhere[80], and will not be included here for conciseness.

5. Normalized square L_2 norm

$$= \sum_i \left(\frac{f_i - \bar{f}}{\sigma_f} - \frac{v_i - \bar{v}}{\sigma_v} \right)^2 \tag{43}$$

6. Incremental sign distance

   To compute this metric, a vector is created for each image containing the signs of the differences between adjacent pixel intensities. Incremental sign distance is the Hamming distance between these two vectors.

7. Mutual information

$$= E(F) + E(V) - E(F \mid V) \tag{44}$$

Here $E(F)$, $E(V)$, and $E(F \mid V)$ are the entropy of the video frame, the entropy of the virtual image, and the joint entropy of the two images, respectively. Entropy is calculated as

$$E(F) = -\sum_i p_F(f_i) \log p_F(f_i) \tag{45}$$

where p_F is the probability distribution of the video frame, calculated by normalizing the image histogram. The virtual image entropy and the joint entropy are defined similarly.

8. Gradient-weighted mutual information

$$= MI(F,V) \cdot GW(F,V) \tag{46}$$

Here $MI(F,V)$ is the mutual information of the video frame and virtual image, and $GW(F,V)$ is a gradient-weighting factor calculated by

$$GW(F,V) = \sum_i \frac{\cos(\phi_i) + 1}{2} \cdot \min(|\nabla F(f_i)|, |\nabla V(v_i)|) \tag{47}$$

In this equation, $|\nabla F(f_i)|$ and $|\nabla V(v_i)|$ are the magnitudes of the intensity gradients in the video frame and virtual image, and ϕ_i is the angle between the

gradients. This factor favors alignment where both images have strong edges in the same location and the same direction.

9. Mutual information of gradient magnitudes

   This similarity measure is calculated in the same way as mutual information, but gradient magnitudes of the video frame and virtual image are used rather than pixel intensities.

10. Discriminative structural similarity measure

    This similarity measure was designed specifically for endoscopic video and virtual endoscopic images. Its calculation is complex, and involves breaking the image into many small sub-regions, rejecting those with characteristics unfavorable for matching the two images, and keeping those likely to contain meaningful structural information. Its definition can be found elsewhere[32], and will not be included here for conciseness.

8.2.4 Gaussian smoothing for virtual images

Virtual endoscopic images contain no noise, and the edges are very sharply defined. This is not the case for endoscopic video frames. It could be advantageous to apply Gaussian smoothing to the virtual images, which would smear the edges and make their

appearance more similar to that in video frames. To test this hypothesis, the volumetric grid search was run with three different Gaussian kernels applied to the virtual image: 5 x 5 with σ = 1 pixel, 9 x 9 with σ =2.6 pixels, and 13 x 13 with sigma = 5.2 pixels. The standard deviations were chosen such that the corner elements of the kernel were ~0.0025 for each size. Examples of a virtual image smoothed with the three kernels are shown in Figure 41.

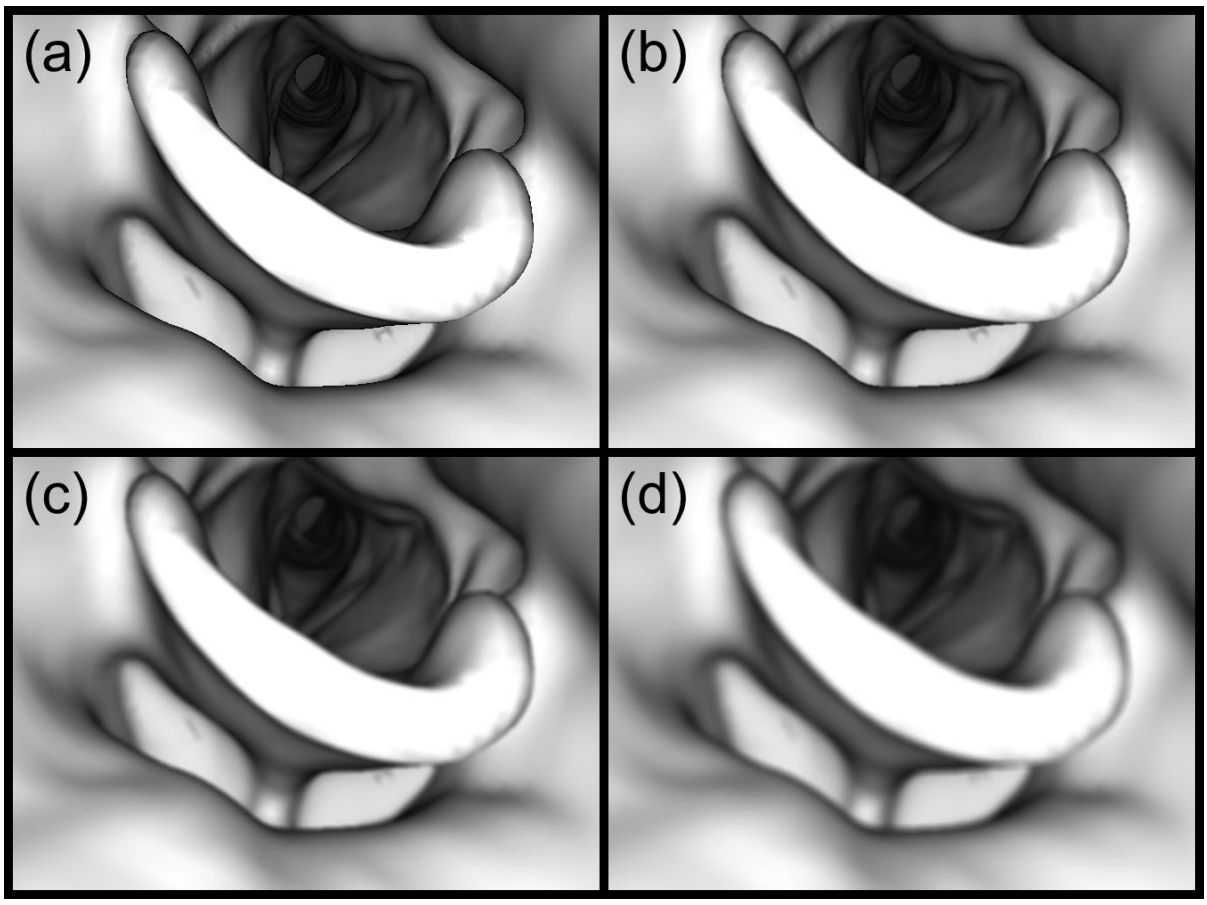

Figure 41: Examples of the Gaussian smoothing kernels applied to virtual endoscopic images. (a) No smoothing. (b) 5 x 5 kernel, σ = 1 pixel. (c) 9 x 9 kernel, σ = 2.6 pixels. (d) 13 x 13 kernel, σ = 2.6 pixels.

8.2.5 Edge-preserving smoothing for video frames

The major features present in both video frames and virtual images are structural edges and the overall changes lighting with distance from the camera. The virtual image surfaces are smooth, but the video frames contain variable textures in the epithelial tissue. It could be advantageous to smooth out these surface textures while preserving structural edges. To test this hypothesis, the volumetric search was run using the bilateral filter[81] to smooth the registration frames. This filter was applied to color images prior to grayscale conversion. It has two parameters σ_{color} and σ_{space} that determine its extent in color and coordinate space, respectively. Three combinations were chosen: $\sigma_{color} = \sigma_{space} = 15$, $\sigma_{color} = \sigma_{space} = 30$, and $\sigma_{color} = \sigma_{space} = 45$. Examples of a video frame smoothed with these filters are shown in Figure 42.

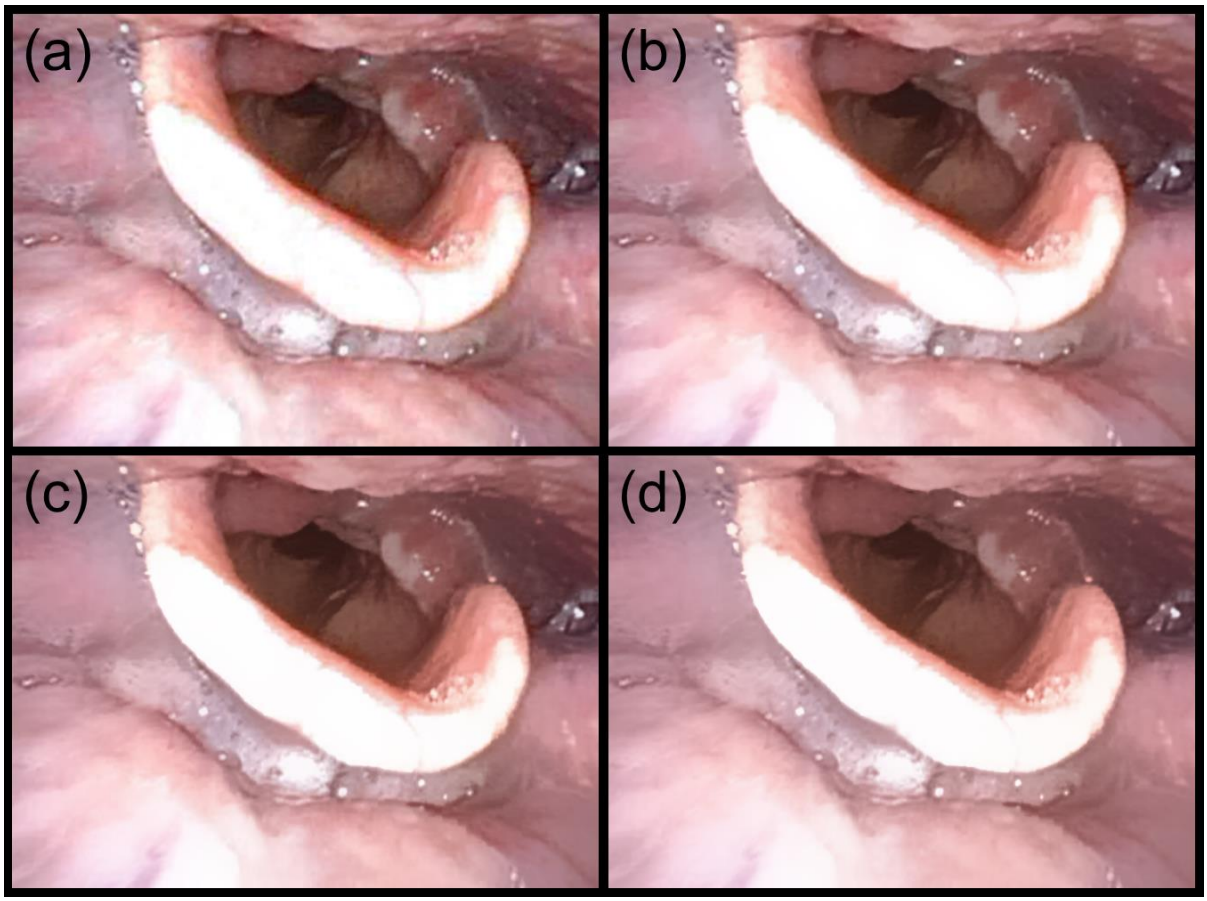

Figure 42: Examples of the bilateral filters applied to endoscopic video frames. (a) No filter. (b) $\sigma_{color} = \sigma_{space} = 15$. (c) $\sigma_{color} = \sigma_{space} = 30$. $\sigma_{color} = \sigma_{space} = 45$.

8.2.6 Downsampling

Downsampling images to a lower resolution is a common approach in image registration algorithms. It is generally used in a pyramid implementation, where the images are filtered and downsampled one or more times, and the registration results at lower resolutions are used to initialize the registration at higher resolutions. The

concept of a pyramid implementation does not extend as naturally to endoscopy-CT registration because the video frame and virtual image do not move on top of each other. However, it could be advantageous to downsample the video frames and virtual images for similarity calculations. To test this hypothesis, the volumetric grid search was run with the images downsampled to 0.5 and 0.25 times their original sizes. Downsampling was performed by smoothing the images with a 5 x 5 Gaussian kernel and discarding every other row and column. Examples of the downsampled images are shown in Figure 43. To test the potential use of a pyramid implementation, the registered coordinates from lower resolutions were refined with a simplex optimization over a small search space set by

$$\Delta_{simplex} = (0.5\ mm,\quad 0.5\ mm,\quad 0.5\ mm,\quad 1\ deg,\quad 1\ deg,\quad 1\ deg) \qquad (48)$$

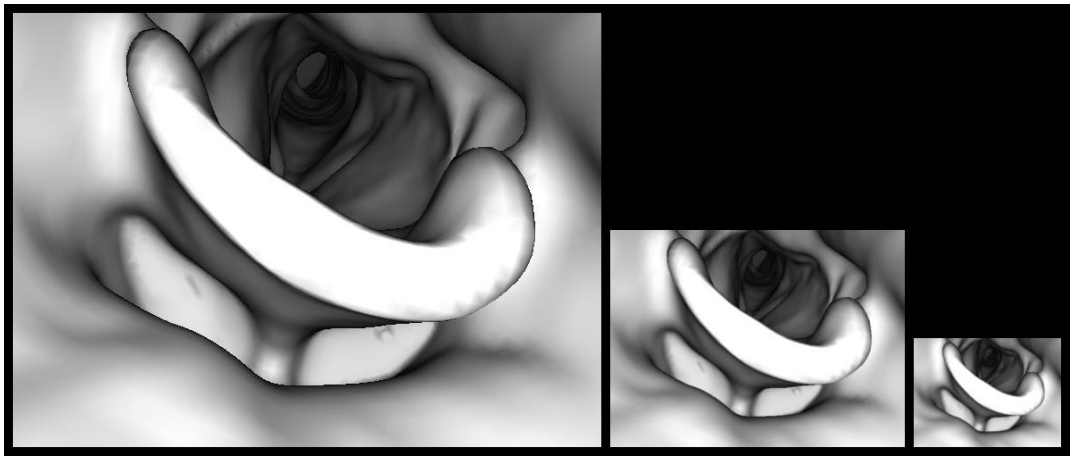

Figure 43: Examples of 2x and 4x downsampled virtual endoscopic images.

8.2.7 Masking

There are generally differences between the anatomical structures seen in the endoscopic video frames and those seen in the virtual images. When these differences are large, they can prevent the similarity measure from finding a good match near the correct endoscope coordinates. This is discussed in more detail in Section 6.4.3, and an example of the structural differences is given in Figure 33. It could be advantageous to mask out the areas in the registration frames that contain any structural differences. To test this hypothesis, the volumetric grid search was run using the structure masks created for the virtual endoscopy lighting optimization, which are discussed in Section 6.2.4.1. An example of these masks is given in Figure 25.

8.2.8 Analyses of calibration parameters

Calibration parameters were not available for the PMH cohort patient (see Section 6.2.1). In the analyses described in Sections 8.2.8.1 and 8.2.8.2, only the two endoscopic examinations of patient MDA1 were used, with a total of 27 registration frames.

8.2.8.1 View angle

Camera calibration provides the focal length of the endoscope's camera, which is used to set the view angle of the virtual camera with Equation 9. The calibrated view angle for the endoscope used in these analyses was 49.55 degrees. To investigate the

sensitivity of registration accuracy to the view angle, the volumetric grid search was run with the virtual camera's view angle set to 40, 45, 55, and 60 degrees. Examples of virtual images rendered with these view angles are shown in Figure 44.

The goal of the grid searches was to determine if an approximate view angle could be used in place of a calibrated view angle for endoscopy-CT registration. However, this does not consider the effect that the view angle has on projective measurements, even if accurately-registered endoscope coordinates can be obtained. To investigate this effect, the virtual endoscope was placed at the ground-truth coordinates, and the world transform was taken with the camera's view angle set to 40, 45, 55, and 60 degrees. The changes to the virtual image when the view angle is modified are radially symmetric, and their magnitude is larger farther from the center of the image. It may be the case that projective errors remain acceptably small within some central region of the image. To test this hypothesis, the world transforms were sampled using a series of five circular ROIs with evenly-spaced radii out to the image edge. These are shown in Figure 45.

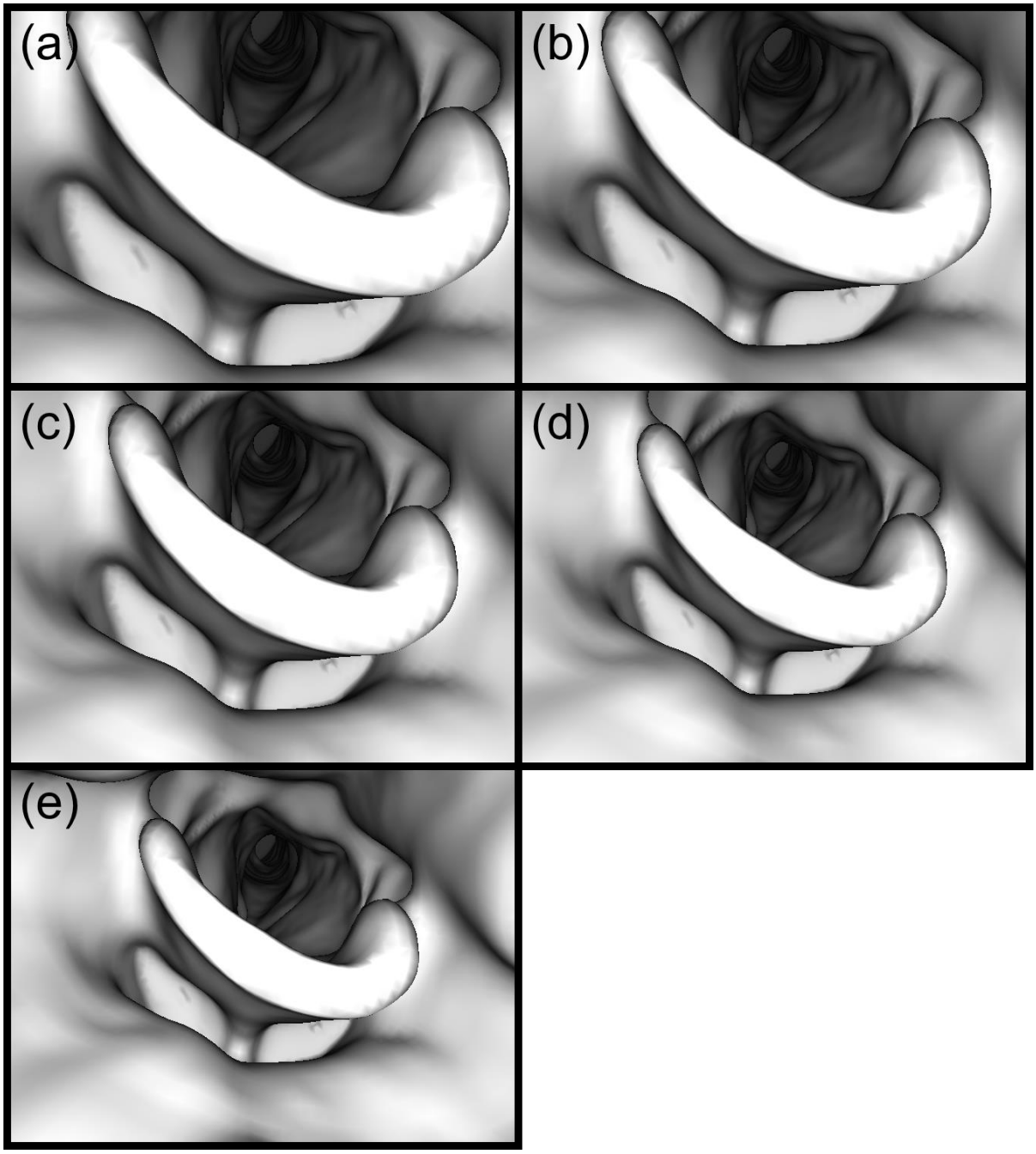

Figure 44: Examples of virtual endoscopic images rendered with variable view angles for the virtual camera. (a) 40 degrees. (b) 45 degrees. (c) 49.55 degrees, the view angle obtained from the camera calibration described in Section 3.4. (d) 55 degrees. (e) 60 degrees.

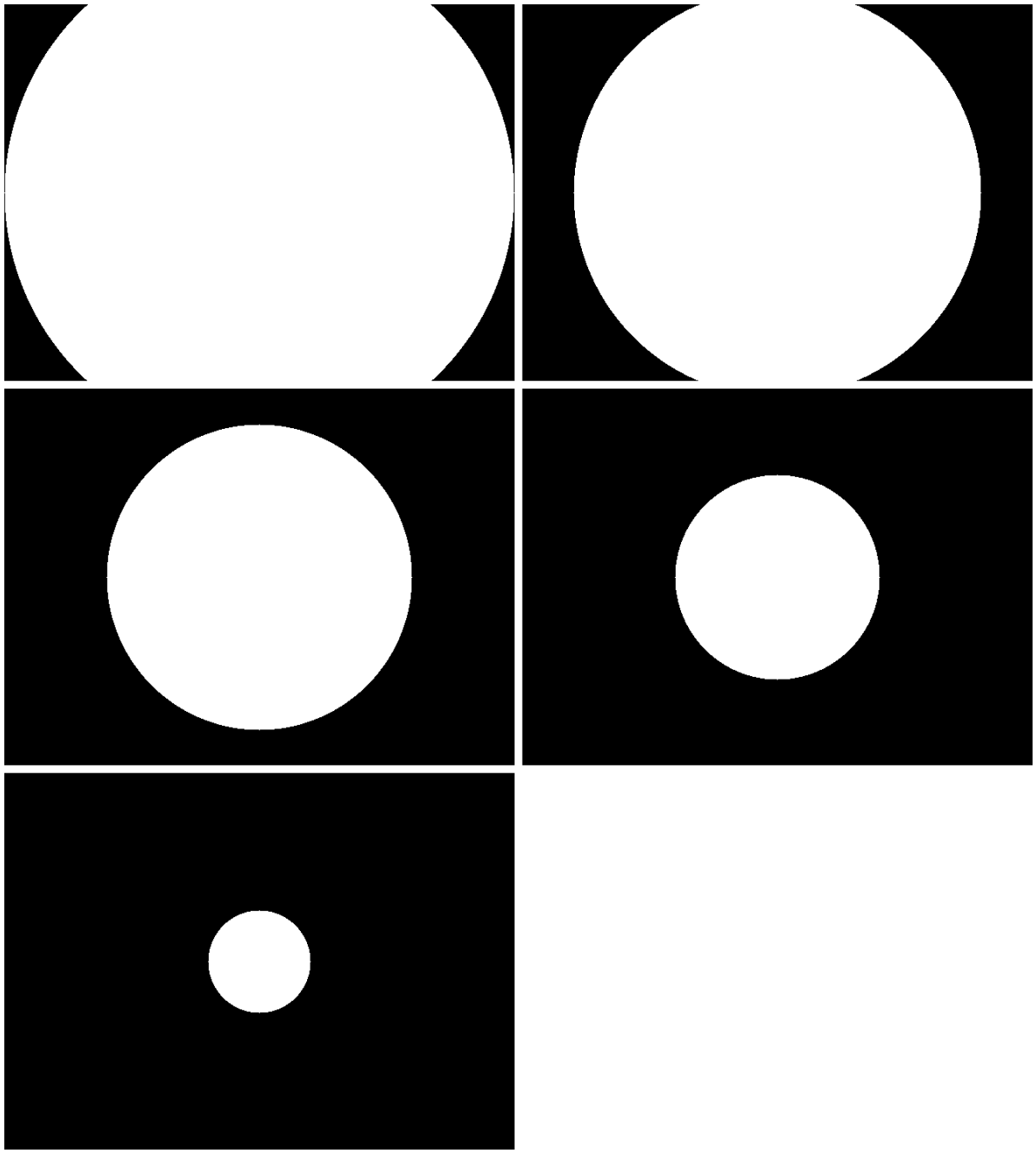

Figure 45: The circular ROIs used to sample projective errors for the analysis of calibration parameters. Their radii are 329, 263, 197, 132, and 66 pixels. They cover 90%, 66%, 38%, 17%, and 4% of the total image area, respectively.

8.2.8.2 Distortion parameters

Camera calibration provides a set of five distortion parameters, which are used to remove distortion from endoscopic video frames. A detailed description of the distortion model is given in Section 3.4.1. Three of the parameters describe radial distortion and two describe tangential distortion, and the calibrated values are (-0.3858, 0.212, -0.00115, 0.00083, -0.075). It was observed that for the endoscope used in these analyses, the distortion could be adequately modeled using only the first radial distortion parameter. To investigate the sensitivity of registration accuracy to the model used to remove distortion from the registration frames, the volumetric grid search was run with the first radial distortion parameter set to -0.27, -0.18, -0.09, and 0. This range of values covers full distortion removal with -0.27 to no distortion removal with 0. The other four parameters were set to 0. Examples of video frames processed with these reduced models are shown in Figure 46.

The goal of the grid searches was to determine if an approximate distortion model could be used in place of a calibrated model for endoscopy-CT registration. However, this does not consider the effect that the distortion model has on projective measurements, even if accurate registered endoscope coordinates can be obtained. To investigate this effect, the virtual endoscope was placed at the ground-truth coordinates, and the world transform was resampled with bilinear interpolation after distorting the virtual image pixel grid with the first radial coefficient set to -0.27, -0.18, -0.09, and 0. The other four parameters were set to 0. As with the view angle, the changes to the virtual image when the first radial distortion parameter is modified are

radially symmetric, and their magnitude is larger further from the center of the image. It may be the case that projective errors remain acceptably small within some central region of the image. To test this hypothesis, the world transforms were sampled using the five circular ROIs shown in Figure 45.

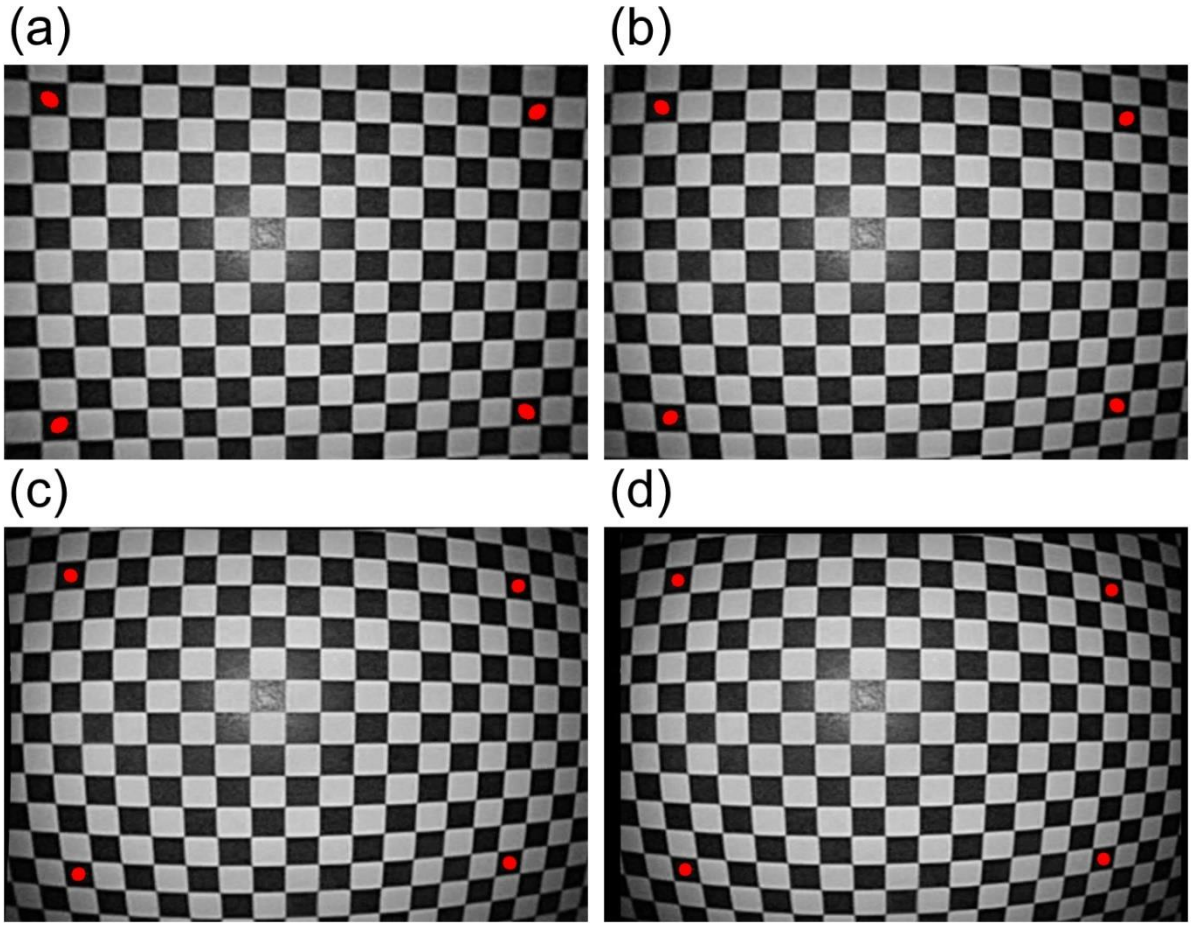

Figure 46: Examples of distortion removal with reduced distortion models. These endoscopic video frames are from a recording of the calibration rig described in Section 3.4.2. The red dots were added digitally prior to distortion removal to provide a reference point. (a) Distortion was removed with the parameters set to (-0.27, 0, 0, 0, 0). The result is very similar to distortion removal with the full model, which can be seen in Figure 5. (b) Distortion was removed with the parameters set to (-0.18, 0, 0, 0, 0). (c) Distortion was removed with the parameters set to (-0.09, 0, 0, 0, 0). (d) No distortion removal. The parameters were set to (0, 0, 0, 0, 0), so the only change in the image was the principal point shift described in Section 3.4.3. This shift accounts for the uneven black borders around the image.

8.3 Results

8.3.2 Similarity measures

The registration accuracy for each similarity measure was quantified by the median projective measurement error between the world transforms of the ground-truth virtual image and the registered virtual image obtained with the volumetric grid search. The results for each similarity measure are summarized in Table 20. Many of the similarity measures had a large number of failed frames, which were identified subjectively as frames for which the registered virtual image did not contain any recognizable structures. Most failed frames were false matches where the virtual endoscope was directly in front of a wall.

Only incremental sign distance and gradient-weighted mutual information had no failed frames. Gradient-weighted mutual information had the best performance overall, even when failed frames were excluded from the analysis. Its median projective errors were significantly smaller than those of all other similarity measures ($p < 0.001$ using the Wilcoxon signed-rank test). However, registration errors were large overall, with an average of 9.1 ± 5.7 mm for gradient-weighted mutual information. Based on these results, gradient-weighted mutual information was used as the similarity measure for all subsequent analyses discussed in this chapter.

Table 20: Results of the volumetric grid searches for all similarity measures. The second column gives the number of failed frames, which were identified subjectively. The third column gives the average median projective measurement error between the ground-truth and registered world transforms. The third column gives this value when failed frames are excluded from the average. Gradient-weighted mutual information had the best performance overall.

Similarity measure*	**# of failed frames (% of total)**	**Median error overall (mm)**	**Median error excluding failed frames (mm)**
PC	14 (30.4)	12.9 ± 7.3	11.0 ± 7.0
CR	36 (78.3)	22.8 ± 9.3	14.1 ± 8.2
SR	14 (30.4)	13.3 ± 7.8	9.6 ± 6.0
MS	46 (100.0)	31.1 ± 26.3	
L2N	14 (30.4)	13.0 ± 7.4	11.1 ± 7.2
ISD	0 (0.0)	15.7 ± 6.4	15.7 ± 6.4
MI	37 (80.4)	22.7 ± 9.1	12.9 ± 8.5
GWMI	0 (0.0)	9.1 ± 5.7	9.1 ± 5.7
MIGM	16 (34.8)	18.6 ± 8.5	14.6 ± 7.2
DSSM	23 (50.0)	18.5 ± 7.9	14.8 ± 7.2

*PC = Pearson correlation, CR = correlation ratio, SR = Spearman's rho, MS = material similarity, L2N = normalized square L_2 norm, ISD = incremental sign distance, MI = mutual information, GWMI = gradient-weighted mutual information, MIGM = mutual information of gradient magnitudes, DSSM = discriminative structural similarity measure.

8.3.2.1 Histogram bins

Mutual information is calculated using the joint histogram of the two images. The video frames and virtual images are 8-bit, so in all analyses presented so far in this dissertation, the joint histogram was computed with 256 bins. It could be advantageous to reduce the number of bins. To test this hypothesis, an additional set of volumetric grid searches was performed with the number of histogram bins set to 128, 64, and 32. The results are given in Table 21. 128 bins had the best performance, with an average median projective measurement error of 8.7 ± 5.6 mm, compared to 9.1 ± 5.7 mm with 256 bins. This difference was statistically significant ($p < 0.05$ using the Wilcoxon signed-rank test), and the median error was improved for 32 out of 46 frames.

Table 21: Results of the volumetric grid searches with different numbers of histogram bins. The second column gives the average median projective measurement error between the ground-truth and registered world transforms, and the third column gives the number of frames with a smaller median error than that with 256 bins. 128 bins had the best performance.

Histogram bins	Average median error (mm)	# of frames improved (% of total)
256	9.1 ± 5.7	
128	8.7 ± 5.6	32 (70)
64	8.9 ± 6.2	30 (65)
32	9.4 ± 6.1	25 (54)

8.3.3 Gaussian smoothing for virtual images

The results of the volumetric grid searches with different levels of Gaussian smoothing applied to the virtual endoscopic images are given in Table 22. All kernels resulted in smaller median projective errors relative to those for the 3 x 3 kernel, which was the base level of smoothing applied to all images throughout this dissertation. However, the improvement was statistically significant only for the 5 x 5 kernel ($p < 0.05$ using the Wilcoxon signed-rank test).

Table 22: Results of the volumetric grid searches with Gaussian smoothing of the virtual endoscopic images. The second column gives the average median projective measurement error between the ground-truth and registered world transforms, and the third column gives the number of frames with a smaller median error than that with the 3 x 3 kernel. All kernels resulted in reduced errors, but the difference was only statistically significant for the 5 x 5 kernel.

Gaussian kernel	**Average median error (mm)**	**# of frames improved (% of total)**
3 x 3, $\sigma = 1.0$	9.1 ± 5.7	
5 x 5, $\sigma = 1.0$	8.7 ± 5.5	29 (63)
9 x 9, $\sigma = 2.6$	8.7 ± 5.6	28 (61)
13 x 13, $\sigma = 5.2$	8.5 ± 5.5	29 (63)

8.3.4 Edge-preserving smoothing for video frames

The results of the volumetric grid searches with different bilateral filters applied to the endoscopic video frames are given in Table 23. There was no significant improvement in the median projective errors with any of the filters ($p > 0.05$ using the Wilcoxon signed-rank test).

Table 23: Results of the volumetric grid searches with edge-preserving smoothing of the endoscopic video frames. The second column gives the average median projective measurement error between the ground-truth and registered world transforms, and the third column gives the number of frames with a smaller median error than that with no smoothing. There was no significant improvement with any level of edge-preserving smoothing.

Filter	**Average median error (mm)**	**# of frames improved (% of total)**
None	9.1 ± 5.7	
$\sigma_{color} = \sigma_{space} = 15$	9.2 ± 6.2	21 (45.7)
$\sigma_{color} = \sigma_{space} = 30$	9.0 ± 5.9	28 (60.9)
$\sigma_{color} = \sigma_{space} = 45$	10.9 ± 7.0	20 (43.5)

8.3.5 Downsampling

The results of the volumetric grid searches with 2x and 4x downsampling are given in Table 24. Both levels of downsampling, and the subsequent refinement with a pyramid implementation, resulted in smaller median projective errors relative to those for the full-resolution images. However, the improvement was statistically significant only for 4x downsampling ($p < 0.05$ using the Wilcoxon signed-rank test). The refinement of the registered coordinates with a pyramid implementation made the results worse, and this difference was also statistically significant ($p < 0.01$ using the Wilcoxon signed-rank test).

Table 24: Results of the volumetric grid searches with downsampled images. The second column gives the average median projective measurement error between the ground-truth and registered world transforms, and the third column gives the number of frames with a smaller median error than that with the full-resolution images. 4x downsampling had the best results, but the pyramid refinement made them worse.

Downsampling	**Average median error (mm)**	**# of frames improved (% of total)**
None	9.1 ± 5.7	
2x	9.0 ± 6.2	26 (57)
2x pyramid	9.0 ± 6.2	28 (61)
4x	7.8 ± 5.1	30 (65)
4x pyramid	8.1 ± 5.4	29 (63)

8.3.6 Masking

The average median projective measurement error with the structure masks applied to the images was 5.5 ± 3.9 mm. The median error found with the unmasked images was 9.1 ± 5.7 mm. This difference was statistically significant ($p < 0.001$ using the Wilcoxon signed-rank test), and it was the largest improvement found with any of the image processing parameters. The median projective error was improved for 33 frames, which is 72% of the total number.

8.3.7 Analyses of calibration parameters

8.3.7.1 View angle

The results of the volumetric grid searches with different view angles used for the virtual camera are given in Table 25. The world transforms of the ground-truth and registered virtual images were taken with the calibrated view angles, so these values provide the same measure of registration accuracy as the previous tables in this chapter. The results with the reduced view angles of 40 and 45 degrees were worse than those with the calibrated view angle, and these differences were statistically significant ($p < 0.05$ using the Wilcoxon signed-rank test). However, the results with the increased view angles of 55 and 60 degrees were not significantly different from those with the calibrated view angle.

The results of the projective error analysis using the five circular ROIs are shown in Figure 47. When the virtual camera view angle was set to 45 degrees or 55 degrees, the majority of projective errors remained within 2 mm for all ROIs. With view angles of 40 and 60 degrees, the average median projective errors in each ROI were at least twice as large as their counterparts with view angles of 45 and 55 degrees. Interestingly, moving from the largest ROI to the intermediate ROIs tended to increase the spread of the projective errors, with smaller lower quartiles, larger upper quartiles, and often larger medians. For all view angles, the two smallest ROIs had smaller average median projective errors than the three largest ROIs.

Table 25: Results of the volumetric grid searches with different view angles for the virtual camera. The second column gives the average median projective measurement error between the ground-truth and registered world transforms. The third row contains the reference value obtained with the calibrated view angle. View angles larger than the calibrated value did not result in increased error.

View angle (deg)	**Average median error (mm)**
40	12.2 ± 7.0
45	10.6 ± 7.0
49.55	9.9 ± 6.8
55	10.1 ± 7.1
60	9.9 ± 7.7

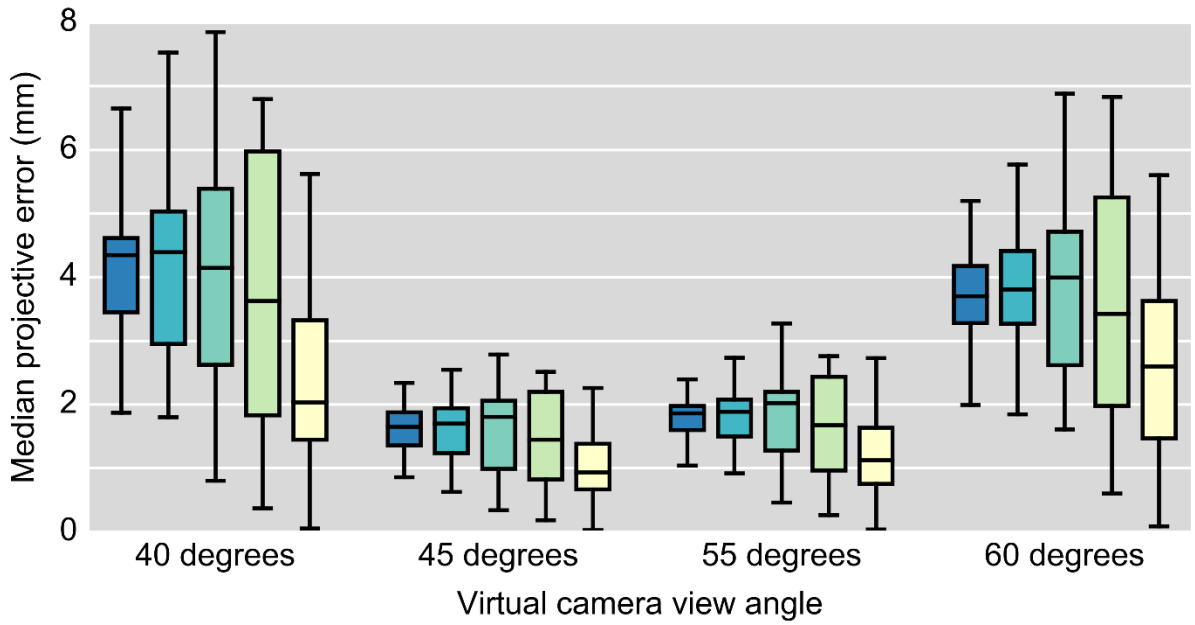

Figure 47: Projective measurement errors induced by changes to the view angle of the virtual camera. These plots show the median projective measurement error for all registration frames when the virtual camera's view angle is changed from the calibrated value of 49.55 degrees. For each view angle, the five plots correspond to the five circular ROIs shown in Figure 45, with the largest on the left and the smallest on the right. The boxes show the median and the quartiles, and the whiskers show the minimum and maximum. With view angles of 45 and 55 degrees, the majority of projective errors are less than 2 mm. Interestingly, sampling with the mid-sized ROIs tended to increase the median error and the spread of the errors.

8.3.7.2 Distortion parameters

The results of the volumetric grid searches with the four reduced models used to remove distortion from the registration frames are given in Table 26. The world transforms of the ground-truth and registered virtual images were not affected by the distortion model, so these values provide the same measure of registration accuracy as

the previous tables in this chapter. The average median projective measurement errors were nearly the same for every model, and there were no significant differences between them ($p > 0.05$ using the Kruskal-Wallis H-test).

The results of the projective error analysis using the five circular ROIs are shown in Figure 48. Median projective measurement errors remained small for all distortion models and all ROIs. Even with no distortion removal and the largest ROI, the maximum error was less than 1 mm. The exclusion of peripheral points with the smaller circular ROIs consistently reduced errors for all distortion models. When the distortion parameters were set to (-0.27, 0, 0, 0, 0), which approximates the calibrated values as shown in Figure 46, all errors in all ROIs were less than 0.02 mm.

Table 26: Results of the volumetric grid searches with different models used to remove distortion from the registration frames. The second column gives the average median projective measurement error between the ground-truth and registered world transforms. The first row contains the reference value obtained with the calibrated distortion model. Equivalent results were obtained with all models.

Distortion parameters	**Average median error (mm)**
(-0.3858, 0.212, -0.00115, 0.00083, -0.075)	9.9 ± 6.8
(-0.27, 0, 0, 0, 0)	9.9 ± 6.7
(-0.18, 0, 0, 0, 0)	9.9 ± 6.6
(-0.09, 0, 0, 0, 0)	10.1 ± 6.6
(0, 0, 0, 0, 0)	9.8 ± 6.4

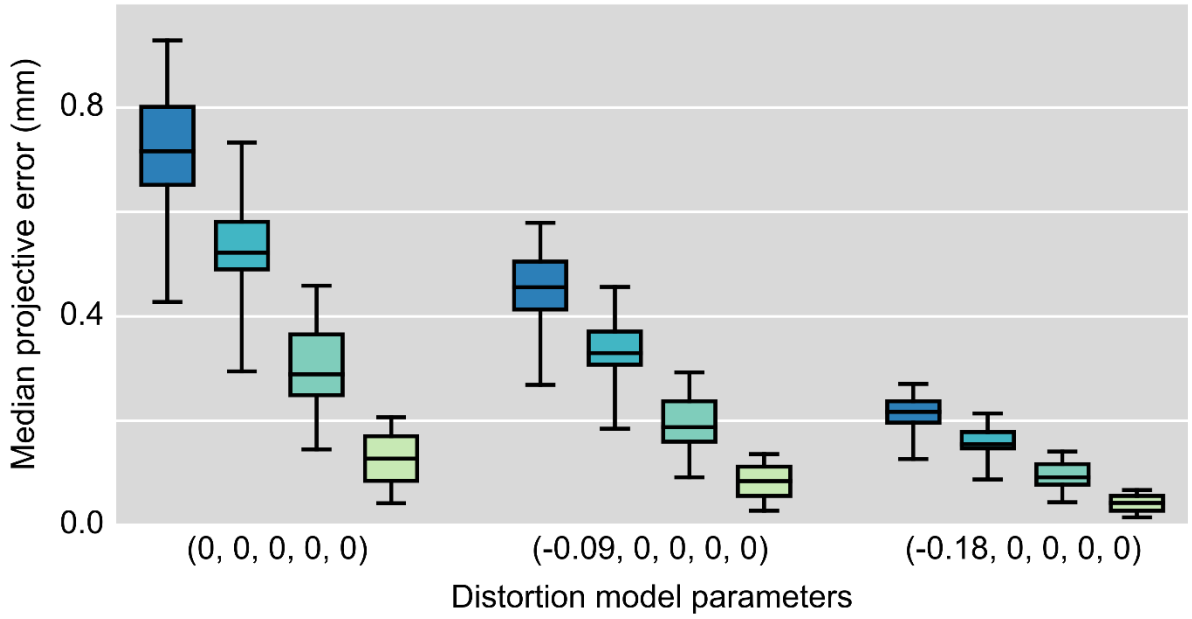

Figure 48: Projective measurement errors induced by changes to the distortion model. These plots show the median projective errors for all registration frames when the distortion model is changed from the calibrated values. For each model, the four plots correspond to the four largest circular ROIs shown in Figure 45. The smallest ROI was omitted because the maximum error for all frames and all models was less than 0.05 mm. The errors for (-0.27, 0, 0, 0, 0) are not shown because they were less than 0.02 mm for all frames and all ROIs. Even with no distortion removal and the largest ROI, the maximum projective error was less than 1 mm.

8.4 Discussion

8.4.1 Similarity measures and preprocessing

Many similarity measures had a large number of failed frames, where the registered virtual endoscope image was a meaningless view in front of the wall. In these

views, the virtual image contains no recognizable structure, and consists of smooth intensity gradients. It is unclear why this occurred so often for a variety of similarity measures. It is also surprising that the discriminative structural similarity measure, which was designed specifically for endoscopic video and virtual endoscopy, had weak performance. However, it was developed on bronchoscopic images, and there are several parameters that influence its calculation. For this analysis, these parameters were set to the values used in the publication[32], so it is possible that better performance could be obtained with different values. Gradient-weighted mutual information was the best similarity measure, and calculating it with 128 histogram bins rather than 256 further improved the registration accuracy.

Among the preprocessing parameters, Gaussian smoothing of virtual images, downsampling of both images, and masking both images to avoid structural disparities all improved the registration accuracy. With Gaussian smoothing, a 5 x 5 kernel with σ = 2.6 provided the best results. With downsampling, reducing the resolution by a factor of 4 from (659, 486) to (165, 122) provided the best results. Interestingly, refining these results at higher resolutions with a pyramid implementation resulted in worse registration accuracy. It is not clear why this occurred, but perhaps the sharp edges and other irregularities of virtual images are smoothed out by downsampling, allowing more stable matching to video frames. The concept of a pyramid implementation does not extend naturally to endoscopic video and virtual endoscopy, because the images do not move on top of each other during the optimization process.

By far, the largest reduction in projective measurement error was obtained by using the structure masks described in Section 6.2.4.1. This shows that one of the major

sources of error in endoscopy-CT registration is the structural disparities between endoscopic video and virtual endoscopy discussed in Section 6.4.3. This analysis shows that masking certain regions out of the similarity calculation can reduce the impact of these disparities. Future studies should investigate automatic generation of such masks, as the manually-drawn masks used for this analysis are time-consuming, and the results may suffer from inter-user variability.

8.4.2 Calibration parameters

These analyses show that endoscopy-CT registration is more sensitive to accurate determination of the endoscope camera's view angle than its distortion model. The registration accuracy was reduced when the virtual camera's view angle was set to be smaller than the calibrated value, but not when it was set larger. Projective measurement errors were also found to be more sensitive to the view angle than the distortion parameters. When the view angle was set to the calibrated value ± 5 degrees, the majority of projective measurement errors remained within 2 mm, which may be an acceptable uncertainty. Interestingly, taking projective measurements within a central region of the image did not have the expected reduction in error for the various view angles. This is illustrated in Figure 47 by the increased spread of the quartiles when moving from the largest circular ROI to the mid-sized ROIs. The likely cause for this effect is that the peripheral regions of a virtual endoscopic image generally contain the areas of the surface that are closest to the camera. This measurement distance sets a soft upper limit on the magnitude of projective errors, so when these points were

excluded by a smaller ROI, the apparent spread of measurement errors in the box plot increased.

Registration accuracy was not affected by changes to the distortion model, even when distortion was not removed at all from the registration frames. The projective measurement errors introduced by changes to the distortion model were smaller than those introduced by changes to the view angle, and the maximum error with no distortion removal and the largest ROI was still less than 1 mm. Projective measurements within central regions of the images had even smaller errors, which is illustrated in Figure 48. This was expected, because the magnitude of the distortion increases with radial distance from the center of the image.

The motivation for the analyses of view angle and distortion model was to determine if it would be feasible to register endoscopic video to CT if the endoscope's calibration parameters were not available. The results show that an exact distortion model is not necessary. A reduced model with a single radial distortion parameter is sufficient for accurate registration and introduces minimal error to the projective measurements. However, the view angle must be known with some accuracy. The results show that with a view angle as much as 5 degrees larger than the calibrated value, registration accuracy is not affected, and the errors introduced to the projective measurements are on the order of 2 mm. If an approximate view angle were to be used, the user might start from the manufacturer's specification, which is 90 degrees for the ENF-VQ endoscope used in this analysis. If this is assumed to be the diagonal view angle, it corresponds to a vertical view angle of about 61 degrees. The calibrated view

angle for the endoscope was 49.55 degrees, so this approximation would likely introduce unacceptable registration errors.

9

Discussion

9.1 Specific aim 1

Specific aim 1: Develop, test, and optimize a method to register endoscopic video of the head and neck to CT

Hypothesis: Endoscopic video frames can be registered to CT with an accuracy of 5 mm in rigid phantoms and 10 mm in patients.

The development of the path-based volumetric search, a novel registration algorithm for endoscopic video presented in Section 4.3, was essential to the completion of this aim. The structure of this dissertation presents that algorithm and the established frame-to-frame tracking algorithm as though they were selected from the outset and then tested on phantoms and patients. However, path-based volumetric search was devised and implemented early in the development of this body of research specifically to overcome the limitations of frame-to-frame tracking. One major drawback of frame-to-frame tracking is that an initial set of virtual endoscope coordinates must be established, which is not a trivial task. The other major drawback

is that the virtual endoscope can become lost. This was apparent in the bolus phantom, in which frame-to-frame tracking failed to reach any of the registration frames until the virtual endoscope was manually placed near the bolus and tracking was restarted. It was even more apparent in patients tests, in which frame-to-frame tracking failed for many frames even after being restarted, and had very large registration errors overall. It may be possible to automatically identify when the virtual endoscope is starting to drift off track and correct, but preliminary tests trying to identify failure points based on changes in the similarity measure did not provide meaningful results.

The strengths of the path-based volumetric search are its ability to search the entire volume in a reasonably efficient manner, and that the virtual endoscope view directions initialized using the path will always be reasonably close to the correct view direction except in the most extreme cases. One weakness is the requirement of manual input to create the path, which would introduce some inter-user variability. Another weakness is long computation times of ~15 minutes to register a single frame, which is certainly too long for routine clinical application. However, this project was approached with the goal of exploring feasibility rather than developing efficient software, and there are many ways that this algorithm could be made more computationally efficient. Two of the most salient targets to reduce computation times are parallel computing when performing the coarse search, and restriction of the search space to a smaller volume near the endoscope's position, rather than the entirety of the airways in the head and neck.

Phantom tests of image registration are presented in Chapter 5. In these tests, path-based volumetric search achieved a median point measurement error of 3.0 mm and a

median symmetric mean absolute distance (SMAD) between measured and ground-truth bolus contours of 3.5 mm. This confirms the hypothesis that endoscopic video frames can be registered to CT with an accuracy within 5 mm. Patient tests of image registration are presented in Chapter 6. There were no fiducial markers with ground truth coordinates in patient tests, so registration accuracy was quantified by the median projective measurement error relative to a ground-truth virtual endoscopic image. The median registration accuracy with path-based volumetric search was 9.9 mm. Registration accuracy was further improved by several of the image processing parameters incorporated in Chapter 8, particularly when the structure masks were used to compute similarity only in certain regions of the images. This confirms the hypothesis that endoscopic video frames can be registered to CT with an accuracy within 10 mm. However, there were many frames with larger errors, and registration tests were not possible on two of the patients. These limitations suggest that endoscopy-CT registration may not be robust enough for clinical application, even if the registration accuracy can be further improved to clinically-acceptable levels.

One feature common to all sets of measurements taken in phantoms and patients was the presence of very large outliers. This is an inherent characteristic of projective measurements with virtual endoscopy, and it generally occurs when a projected point is near an occluding edge in the scene with a large distance behind it. A small change in the virtual endoscope's position or orientation can cause the point to miss the edge and intersect the surface behind it instead, leading to errors as large as 100 mm or more. Errors such as these would generally be apparent in the context of the object that is being mapped from endoscopy to CT space, but they do pose a challenge for developing

a robust endoscopy-CT mapping tool that could be incorporated into the radiotherapy workflow. Both the phantom and patient studies found that the geometry of the surface mesh can provide information about the expected measurement uncertainty in different areas of the virtual endoscopic image. A dependence of measurement error on the distance between the endoscope and the surface was found consistently, and the edge masks described in Section 4.4.4 were found to be useful in excluding large-error points from sets of projective measurements.

9.2 Specific aim 2

Specific aim 2: Investigate the sources of uncertainty in projective mapping via virtual endoscopy and determine their impact on endoscopy-CT registration.

Hypothesis: Patient positioning will have the largest impact on registration errors.

Two of the major sources of uncertainty in endoscopy-CT image registration are the non-rigid anatomy of the airway surfaces in the head and neck, and the differences in patient positioning between seated endoscopic examinations and supine CT scans. These were investigated by taking repeated projective measurements in different CT scans with the virtual endoscope placed at the same position. The influence on non-rigid anatomy was investigated using a set of radiotherapy patients who had daily treatment-room CT imaging in the same position as the simulation CT, and the influence of patient positioning was investigated using diagnostic CTs of the same patients. In

diagnostic CTs, the patients' head, neck, and shoulders are not positioned in any particular way, so the differences in projective measurements provide some insight into the importance of reproducing the simulation-CT position for projective measurements with virtual endoscopy. This study has been described in more detail in Chapter 7.

In the daily CT scans, the projective measurement errors were on the order of 1.5-3.0 mm, and their magnitudes were dependent on anatomical region. The errors were larger in the diagnostic scans, on the order of 3.5-4.5 mm. This confirms the hypothesis that patient positioning has a larger impact on registration uncertainty than daily anatomical variations. These analyses suggest that daily variations and patient positioning difference impose a lower limit of 3-5 mm on the uncertainty of endoscopy-CT image registration, which is not insignificant in the context of applications to radiotherapy. Furthermore, the results of this study account only for the virtual endoscopic manifestations of these sources of uncertainty. It is likely that patient positioning differences also affect the anatomical configuration as it appears in endoscopic videos, so further studies will be necessary to fully understand their influence on registration uncertainty.

The third component of this specific aim was the analysis of the impact of variations in the camera calibration parameters on registration accuracy and projective measurement errors. This analysis has been presented in Sections 8.2.8 and 8.3.7. The calibration is not a source of uncertainty in the same sense as are anatomical variations. Instead, this analysis sought to determine whether or not endoscopy-CT registration was feasible in a scenario where the endoscope's calibration parameters were not available. Unlike CT, for which a DICOM file contains all the information necessary for

registration to another CT, an endoscopic video file does not contain the information necessary to remove distortion or set the view angle of the virtual endoscope.

The distortion model was found to have no impact on registration accuracy, and changes to the model introduced projective measurement errors of less than 1 mm in the worst case. Additionally, a reduced distortion model was found to provide equivalent results to the calibrated model. These results show that knowing the distortion component of the camera calibration may not be necessary for accurate endoscopy-CT registration. However, changes to the virtual endoscope's view angle had a larger impact, both in terms of registration accuracy and projective measurement errors. These results show that without a way to set the virtual endoscope's view angle with ~5 degrees of the calibrated value, accurate registration will not be possible.

9.3 Principal hypothesis

Principal hypothesis: Endoscopic video in the head and neck can be registered to CT without prospective physical endoscope tracking through the use of virtual endoscopy.

The results that are presented in Chapters 5 and 6 confirm the principal hypothesis that registration of endoscopic video in the head and neck to CT is possible. Though there were outliers with large registration errors, many endoscopic video frames were registered successfully, and the magnitudes of the anatomical uncertainties may prove to be manageable with further development of registration algorithms. However, the real problem with endoscopy-CT registration as presented in this dissertation is its

robustness. Out of three patients enrolled in the research protocol, registration was not possible for two. It may be that endoscopy-CT registration is only possible for a subset of head and neck radiotherapy patients, or it may be that greater consideration must be given to patient positioning for the endoscopic examination and to the segmentation methods that are used to create the virtual endoscopy surface mesh. Further studies with larger patient sets will be necessary to understand these problems.

9.3 Future directions

One of the major limitations of the methods used in this dissertation is that they treat the virtual endoscopy surface mesh as a rigid structure, which is at odds with the non-rigidity of the airways of the head and neck. This is likely to be the sources of many of the large patient registration errors in which structural differences prevented the virtual endoscopic images from reproducing the appearance of the registration frame near the correct coordinates. An example of this is given in Figure 33. The importance of accounting for these structural differences is further demonstrated by the large increase in the registration accuracy when the structure masks are applied to remove these disparate regions from the similarity calculation (see Sections 8.2.7 and 8.3.6).

Future work on this subject could incorporate a deformation model that allows the structure of the surface mesh to change. This could be done prior to registration based on the anatomical appearance of the endoscopic video, or as a part of the optimization process. An even simpler approach that could prove effective is to warp the registration frame onto the registered virtual image prior to making projective measurements. This

would help align the edges and prevent some of the large projective measurement errors that occur in those regions.

A related problem is segmenting the airways of the head and neck with sufficient detail to provide meaningful similarity calculations between endoscopic video frames and virtual endoscopic images. An example of this is given in Figure 24, in which the segmentation retained very little detail for the epiglottis. Rather than simply segmenting each patient's CT with a density threshold, a better approach may be to create a finely-detailed generic model of the airway surfaces. This could be registered to individual patients' CTs with an atlas-based approach, and could improve virtual endoscopic details for fine structures such as the epiglottis.

Bibliography

[1] N. Howlader, A.M. Noone, M. Krapcho, D. Miller, K. Bishop, S.F. Altekruse, C.L. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D.R. Lewis, H.S. Chen, E.J. Feuer, and K.A. Cronin, *SEER Cancer Statistics Review, 1975-2013* (National Cancer Institute, Bethesda, MD, 2016).

[2] P. Vineis, M. Alavanja, P. Buffler, E. Fontham, S. Franceschi, Y.T. Gao, P.C. Gupta, A. Hackshaw, E. Matos, J. Samet, F. Sitas, J. Smith, L. Stayner, K. Straif, M.J. Thun, H.E. Wichmann, A.H. Wu, D. Zaridze, R. Peto, and R. Doll, "Tobacco and cancer: recent epidemiological evidence," J. Natl. Cancer Inst. **96**(2), 99–106 (2004).

[3] M. Hashibe, P. Brennan, S. Chuang, S. Boccia, X. Castellsague, C. Chen, M.P. Curado, L. Dal Maso, A.W. Daudt, E. Fabianova, L. Fernandez, V. Wünsch-Filho, S. Franceschi, R.B. Hayes, R. Herrero, K. Kelsey, S. Koifman, C. La Vecchia, P. Lazarus, F. Levi, J.J. Lence, D. Mates, E. Matos, A. Menezes, M.D. McClean, J. Muscat, J. Eluf-Neto, A.F. Olshan, M. Purdue, P. Rudnai, S.M. Schwartz, E. Smith, E.M. Sturgis, N. Szeszenia-Dabrowska, R. Talamini, Q. Wei, D.M. Winn, O. Shangina, A. Pilarska, Z.-F. Zhang, G. Ferro, J. Berthiller, and P. Bofetta, "Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the INHANCE consortium," Cancer Epidemiol. Biomarkers, Prev. **18**(2), 541–550 (2009).

[4] S. Marur, G. D'Souza, W.H. Westra, and A.A. Forastiere, "HPV-associated head and neck cancer: a virus-related cancer epidemic," Lancet Oncol. **11**(8), 781–789 (2010).

[5] C. Suárez, J.P. Rodrigo, A. Ferlito, R. Cabanillas, A.R. Shaha, and A. Rinaldo,

"Tumours of familial origin in the head and neck," Oral Oncol. **42**(10), 965–978 (2006).

6 A. Argiris, M. V Karamouzis, D. Raben, and R.L. Ferris, "Head and neck cancer," Lancet **371**(9625), 1695–1709 (2008).

7 J. Bourhis, J. Overgaard, H. Audry, K.K. Ang, M. Saunders, J. Bernier, J.-C. Horiot, A. Le Maître, T.F. Pajak, M.G. Poulsen, B. O'Sullivan, W. Dobrowsky, A. Hliniak, K. Skladowski, J.H. Hay, L.H.J. Pinto, C. Fallai, K.K. Fu, R. Sylvester, and J.-P. Pignon, "Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis," Lancet **368**(9538), 843–854 (2006).

8 G. Starkschall, L. Dong, P.A. Balter, A.S. Shiu, F. Mourtada, M. Gillin, and R. Mohan, "Clinical Radiation Oncology Physics," in *Radiat. Oncol. Ration. Tech. Results*, 9th ed., edited by J.D. Cox and K.K. Ang (Mosby, Inc, Philadelphia, PA, 2010), pp. 50–91.

9 J. Kim and C. V Dang, "Cancer's molecular sweet tooth and the Warburg effect," Cancer Res. **66**(18), 8927–8930 (2006).

10 D. De Ruysscher, U. Nestle, R. Jeraj, and M. MacManus, "PET scans in radiotherapy planning of lung cancer," Lung Cancer **75**(2), 141–145 (2012).

11 E.G.C. Troost, D.A.X. Schinagl, J. Bussink, W.J.G. Oyen, and J.H.A.M. Kaanders, "Clinical evidence on PET-CT for radiation therapy planning in head and neck tumours," Radiother. Oncol. **96**(3), 328–334 (2010).

12 M. MacManus, U. Nestle, K.E. Rosenzweig, I. Carrio, C. Messa, O. Belohlavek, M. Danna, T. Inoue, E. Deniaud-Alexandre, S. Schipani, N. Watanabe, M. Dondi, and B. Jeremic, "Use of PET and PET/CT for radiation therapy planning: IAEA expert

report 2006-2007," Radiother. Oncol. **91**(1), 85–94 (2009).

13 D. Thorwarth, X. Geets, and M. Paiusco, "Physical radiotherapy treatment planning based on functional PET/CT data," Radiother. Oncol. **96**(3), 317–324 (2010).

14 B. De Bari, L. Deantonio, J. Bourhis, J.O. Prior, and M. Ozsahin, "Should we include SPECT lung perfusion in radiotherapy treatment plans of thoracic targets? Evidences from the literature," Crit. Rev. Oncol. Hematol. **102**, 111–117 (2016).

15 V.S. Khoo and D.L. Joon, "New developments in MRI for target volume delineation in radiotherapy," Br. J. Radiol. **79**(Special issue), S2–S15 (2006).

16 F.P.M. Oliveira and J.M.R.S. Tavares, "Medical image registration: a review," Comput. Methods Biomech. Biomed. Engin. **17**(2), 73–93 (2014).

17 S. Varadarajulu, S. Banerjee, B.A. Barth, D.J. Desilets, V. Kaul, S.R. Kethu, M.C. Pedrosa, P.R. Pfau, J.L. Tokar, A. Wang, L.-M.W.K. Song, and S.A. Rodriguez, "GI endoscopes," Gastrointest. Endosc. **74**(1), 1–6 (2011).

18 M.P. Fried, J. Kleefield, H. Gopal, E. Reardon, B.T. Ho, and F.A. Kuhn, "Image-guided endoscopic surgery: results of accuracy and performance in a multicenter clinical study using an electromagnetic tracking system," Laryngoscope **107**(5), 594–601 (1997).

19 P. Reittner, M. Tillich, W. Luxenberger, R. Weinke, K. Preidler, W. Köle, H. Stammberger, and D. Szolar, "Multislice CT-image-guided endoscopic sinus surgery using an electromagnetic tracking system.," Eur. Radiol. **12**(3), 592–596 (2002).

20 I. Bricault, G. Ferretti, and P. Cinquin, "Registration of real and CT-derived virtual

bronchoscopic images to assist transbronchial biopsy," IEEE Trans. Med. Imaging **17**(5), 703–714 (1998).

21 J.P. Helferty and W.E. Higgins, "Combined endoscopic video tracking and virtual 3D CT registration for surgical guidance," in *Proc. IEEE Conf. Image Process.*(2002), pp. 961–964.

22 K. Mori, D. Deguchi, J. Sugiyama, Y. Suenaga, J. Toriwaki, C.R. Maurer, H. Takabatake, and H. Natori, "Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images," Med. Image Anal. **6**(3), 321–336 (2002).

23 F. Deligianni, A. Chung, and G.-Z. Yang, "pq-space based 2D/3D registration for endoscope tracking," in *Proc. Int. Conf. Med. Image Comput. Comput. Interv.*(2003), pp. 311–318.

24 F. Deligianni, A. Chung, and G.-Z. Yang, "Patient-specific bronchoscope simulation with pq-space-based 2D/3D registration," Comput. Aided Surg. **9**(5), 215–226 (2004).

25 F. Deligianni, A.J. Chung, and G.Z. Yang, "Nonrigid 2D/3D registration for patient specific bronchoscopy simulation with statistical shape modeling: phantom validation," IEEE Trans. Med. Imaging **25**(11), 1462–1471 (2006).

26 W.E. Higgins, J.P. Helferty, K. Lu, S.A. Merritt, L. Rai, and K.-C. Yu, "3D CT-video fusion for image-guided bronchoscopy," Comput. Med. Imaging Graph. **32**(3), 159–173 (2008).

27 L. Rai, J.P. Helferty, and W.E. Higgins, "Combined video tracking and image-video registration for continuous bronchoscopic guidance," Int. J. Comput. Assist.

Radiol. Surg. **3**(3), 315–329 (2008).

[28] D. Deguchi, K. Mori, M. Feuerstein, T. Kitasaka, C.R. Maurer, Y. Suenaga, H. Takabatake, M. Mori, and H. Natori, "Selective image similarity measure for bronchoscope tracking based on image registration," Med. Image Anal. **13**(4), 621–633 (2009).

[29] T. Reichl, X. Luo, M. Menzel, H. Hautmann, K. Mori, and N. Navab, "Deformable registration of bronchoscopic video sequences to CT volumes with guaranteed smooth output," in *Proc. Int. Conf. Med. Image Comput. Comput. Interv.*(2011), pp. 17–24.

[30] X. Luó, M. Feuerstein, D. Deguchi, T. Kitasaka, H. Takabatake, and K. Mori, "Development and comparison of new hybrid motion tracking for bronchoscopic navigation," Med. Image Anal. **16**(3), 577–596 (2012).

[31] S.A. Merritt, R. Khare, R. Bascom, and W.E. Higgins, "Interactive CT-video registration for the continuous guidance of bronchoscopy," IEEE Trans. Med. Imaging **32**(8), 1376–1396 (2013).

[32] X. Luo and K. Mori, "A discriminative structural similarity measure and its application to video-volume registration for endoscope three-dimensional motion tracking," IEEE Trans. Med. Imaging **33**(6), 1248–1261 (2014).

[33] X. Luo, Y. Wan, X. He, J. Yang, and K. Mori, "Diversity-enhanced condensation algorithm and its application for robust and accurate endoscope three-dimensional motion tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*(2014), pp. 1250–1257.

[34] T. Anayama, J. Qiu, H. Chan, T. Nakajima, R. Weersink, M. Daly, J. McConnell, T.

Waddell, S. Keshavjee, D. Jaffray, J.C. Irish, K. Hirohashi, H. Wada, K. Orihashi, and K. Yasufuku, "Localization of pulmonary nodules using navigation bronchoscope and a near-infrared fluorescence thoracoscope," Ann. Thorac. Surg. **99**(1), 224–230 (2015).

35 D.J. Mirota, A. Uneri, S. Schafer, S. Nithiananthan, D.D. Reh, M. Ishii, G.L. Gallia, R.H. Taylor, G.D. Hager, and J.H. Siewerdsen, "Evaluation of a system for high-accuracy 3D image-based registration of endoscopic video to C-arm cone-beam CT for image-guided skull base surgery," IEEE Trans. Med. Imaging **32**(7), 1215–1226 (2013).

36 M.P. Fried, S.R. Parikh, and B. Sadoughi, "Image-guidance for endoscopic sinus surgery," Laryngoscope **118**(7), 1287–1292 (2008).

37 D.J. Mirota, M. Ishii, and G.D. Hager, "Vision-based navigation in image-guided interventions," Annu. Rev. Biomed. Eng. **13**(1), 297–319 (2011).

38 D.I. Rosenthal, J.A. Asper, J.L. Barker, A.S. Garden, K.S.C. Chao, W.H. Morrison, R.S. Weber, and K.K. Ang, "Importance of patient examination to clinical quality assurance in head and neck radiation oncology," Head Neck **28**(11), 967–973 (2006).

39 A. Trotti, L.A. Bellm, J.B. Epstein, D. Frame, H.J. Fuchs, C.K. Gwede, E. Komaroff, L. Nalysnyk, and M.D. Zilberberg, "Mucositis incidence, severity and associated outcomes in patients with head and neck cancer receiving radiotherapy with or without chemotherapy: a systematic literature review," Radiother. Oncol. **66**(3), 253–262 (2003).

40 T. Rancati, M. Schwarz, A.M. Allen, F. Feng, A. Popovtzer, B. Mittal, and A. Eisbruch,

"Radiation dose-volume effects in the larynx and pharynx," Int J Radiat Oncol Biol Phys **76**(3 Suppl), S64-9 (2010).

41 R.A. Weersink, J. Qiu, A.J. Hope, M.J. Daly, B.C.J. Cho, R.S. DaCosta, M.B. Sharpe, S.L. Breen, H. Chan, and D.A. Jaffray, "Improving superficial target delineation in radiation therapy with endoscopic tracking and registration," Med. Phys. **38**(12), 6458–6468 (2011).

42 J. Qiu, A.J. Hope, B.C.J. Cho, M.B. Sharpe, C.I. Dickie, R.S. DaCosta, D.A. Jaffray, and R.A. Weersink, "Displaying 3D radiation dose on endoscopic video for therapeutic assessment and surgical guidance," Phys. Med. Biol. **57**(20), 6601–6614 (2012).

43 Q. Zhao, S. Pizer, M. Niethammer, and J. Rosenman, "Geometric-feature-based spectral graph matching in pharyngeal surface registration," Med. Image Comput. Comput. Interv. **17**(1), 259–266 (2014).

44 Q. Zhao, T. Price, S. Pizer, M. Niethammer, R. Alterovitz, and J. Rosenman, "Surface registration in the presence of missing patches and topology change," in *Proc. Med. Image Underst. Anal.*(2015), pp. 8–13.

45 Q. Zhao, T. Price, S. Pizer, M. Niethammer, R. Alterovitz, and J. Rosenman, "The endoscopogram: a 3D model reconstructed from endoscopic video frames," in *Proc. Int. Conf. Med. Image Comput. Comput. Interv.*(2016), pp. 439–447.

46 L.M. Auer and D.P. Auer, "Virtual endoscopy for planning and simulation of minimally invasive neurosurgery," Neurosurgery **43**(3), 529–537 (1998).

47 A. Ferlitsch, P. Glauninger, A. Gupper, M. Schillinger, M. Haefner, A. Gangl, and R. Schoefl, "Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy," Endoscopy **34**(9), 698–702 (2002).

[48] C.P. Davis, M.E. Ladd, B.J. Romanowski, S. Wildermuth, J.F. Knoplioch, and J.F. Debatin, "Human aorta: preliminary results with virtual endoscopy based on three-dimensional MR imaging data sets," Radiology **199**(1), 37–40 (1996).

[49] C.L. Kay, D. Kulling, R.H. Hawes, J.W.R. Young, and P.B. Cotton, "Virtual endoscopy - comparison with colonoscopy in the detection of space-occupying lesions of the colon," Endoscopy **32**(3), 226–232 (2000).

[50] S. Gilani, A.M. Norbash, H. Ringl, G.D. Rubin, S. Napel, and D.J. Terris, "Virtual endoscopy of the paranasal sinuses using perspective volume rendered helical sinus computed tomography," Laryngoscope **107**(1), 25–29 (1997).

[51] R.P. Gallivan, T.H. Nguyen, and W.B. Armstrong, "Head and neck computed tomography virtual endoscopy: evaluation of a new imaging technique," Laryngoscope **109**(10), 1570–1579 (1999).

[52] A.J. Burke, D.J. Vining, W.F. McGuirt, G. Postma, and J.D. Browne, "Evaluation of airway obstruction using virtual endoscopy," Laryngoscope **110**(1), 23–29 (2000).

[53] H. Scharsach, M. Hadwiger, A. Neubauer, S. Wolfsberger, and K. Bühler, "Perspective isosurface and direct volume rendering for virtual endoscopy applications," in *Eurographics/IEEE-VGTC Symp. Vis.*(2006), pp. 315–322.

[54] W.E. Lorenson and H.E. Cline, "Marching cubes: a high resolution 3D surface reconstruction algorithm," ACM SIGGRAPH Comput. Graph. **21**(4), 163–169 (1987).

[55] W. Schroeder, K. Martin, and B. Lorenson, *The Visualization Toolkit*, 4th ed. (Kitware, Inc., 2006).

[56] D. Schreiner, *OpenGL Programming Guide: The Official Guide to Learning OpenGL, Versions 3.0 and 3.1*, 7th ed. (Pearson Education, Inc., Boston, MA, 2009).

[57] D.C. Brown, "Close-range camera calibration," Photogramm. Eng. **37**(8), 855–866 (1971).

[58] R.Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TVcameras and lenses," IEEE J. Robot. Autom. **3**(4), 323–344 (1987).

[59] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*(1997), pp. 1106–1112.

[60] T.A. Clarke and J.G. Fryer, "The development of camera calibration methods and models," Photogramm. Rec. **16**(91), 51–66 (1998).

[61] P.F. Sturm and S.J. Maybank, "On plane-based camera calibration: a general algorithm, singularities, applications," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*(1999), pp. 1432–1437.

[62] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. IEEE Conf. Comput. Vis.*(1999), pp. 666–673.

[63] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. (O'Reilly Media, Inc., Sebastopol, CA, 2008).

[64] W.S. Ingram, J. Yang, B.M. Beadle, R. Wendt III, A. Rao, X.A. Wang, and L.E. Court, "The feasibility of endoscopy-CT image registration in the head and neck without prospective endoscope tracking," PLoS One **12**(5), 1–23 (2017).

[65] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, "Image registration by

maximization of combined mutual information and gradient information," IEEE Trans. Med. Imaging **19**(8), 809–814 (2000).

66 F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," IEEE Trans. Med. Imaging **16**(2), 187–198 (1997).

67 C. Studholme, D.L.G. Hill, and D.J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," Pattern Recognit. **32**(1), 71–86 (1999).

68 R.C. Gozalez and R.E. Woods, "Chapter 3: Intensity Transformations and Spatial Filtering," in *Digit. Image Process.*, 3rd ed.(Pearson Education, Inc., Upper Saddle River, NJ, 2008), pp. 104–198.

69 F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," in *Proc. IEEE*(2003), pp. 1699–1721.

70 J.A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J. **7**(4), 308–313 (1964).

71 J. Macqueen, "Some methods for classification and analysis of multivariate observations," Proc. Fifth Berkeley Symp. Math. Stat. Probab. **1**(14), 281–297 (1967).

72 G.H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," Numer. Math. **14**(5), 403–420 (1970).

73 W.S. Ingram, J. Yang, R. Wendt III, B.M. Beadle, X.A. Wang, and L.E. Court, "The influence of non-rigid anatomy and patient positioning on endoscopy-CT image registration in the head and neck," Med. Phys. (2017).

74 S. van Kranen, S. van Beek, C. Rasch, M. van Herk, and J.-J. Sonke, "Setup

uncertainties of anatomical sub-regions in head-and-neck cancer patients after offline CBCT guidance," Int. J. Radiat. Oncol. **73**(5), 1566–1573 (2009).

75 L. Zhang, A.S. Garden, J. Lo, K.K. Ang, A. Ahamad, W.H. Morrison, D.I. Rosenthal, M.S. Chambers, X.R. Zhu, R. Mohan, and L. Dong, "Multiple regions-of-interest analysis of setup uncertainties for head-and-neck cancer radiotherapy," Int. J. Radiat. Oncol. **64**(5), 1559–1569 (2006).

76 H. Wang, L. Dong, J. O'Daniel, R. Mohan, A.S. Garden, K.K. Ang, D.A. Kuban, M. Bonnen, J.Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," Phys. Med. Biol. **50**(12), 2887–2905 (2005).

77 K.K. Brock, "Results of a multi-institution deformable registration accuracy study (MIDRAS)," Int. J. Radiat. Oncol. **76**(2), 583–596 (2010).

78 A.S.R. Mohamed, M.-N. Ruangskul, M.J. Awan, C.A. Baron, J. Kalpathy-Cramer, R. Castillo, E. Castillo, T.M. Guerrero, E. Kocak-Uzel, J. Yang, L.E. Court, M.E. Kantor, G.B. Gunn, R.R. Colen, S.J. Frank, A.S. Garden, D.I. Rosenthal, and C.D. Fuller, "Quality assurance assessment of diagnostic and radiation therapy–simulation CT image registration for head and neck radiation therapy: anatomic region of interest–based comparison of rigid and deformable algorithms," Radiology **274**(3), 752–763 (2015).

79 C. Rasch, R. Steenbakkers, and M. Van Herk, "Target definition in prostate, head, and neck," Semin. Radiat. Oncol. **15**(3), 136–145 (2005).

80 A.A. Goshtasby, "Chapter 2: Similarity and Dissimilarity Measures," in *Image Regist. Princ. Tools, Methods*(Springer-Verlag, London, 2012), pp. 7–66.

81 C. Tomasi and R. Manduchi, “Bilateral filtering for fray and color images,” in *Proc. IEEE Conf. Comput. Vis.*(1998), pp. 839–846.

Vita

W. Scott Ingram was born in Greensboro, North Carolina on May 16, 1988. He is the youngest of three children to Cathy and Haywood Ingram. He graduated from Walter Hines Page High School in Greensboro, North Carolina in June 2006, and he attended Johns Hopkins University in Baltimore, Maryland from September 2006 to May 2010. There he received the degree of Bachelor of Science with a major in Physics and a minor in Mathematics. After becoming interested in the field of medical physics, he completed an internship in the Department of Radiation Oncology at the Johns Hopkins Hospital in Baltimore, Maryland in the summer of 2010, and subsequently shadowed physicists at the Cone Health Cancer Center in Greensboro, North Carolina. He gained a broader clinical, biological, and personal perspective of cancer therapy through the Howard Hughes Medical Institute Med-Into-Grad program at The University of Texas MD Anderson Cancer Center in Houston, Texas in the summer of 2011. In September 2011, he entered The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences in Houston, Texas. He conducted his dissertation research under the supervision of Laurence Court, Ph.D., and his research interests include image processing, image registration, and computer vision.