Imperial College London
Department of Computing

Gaze Contingent Robotic Control in Minimally Invasive Surgery

Kenko Fujii

November 2014

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computing of Imperial College London
and the Diploma of Imperial College London

Declaration

I herewith certify that all material in this dissertation which is not my own work has been properly acknowledged.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Kenko Fujii

Abstract

Recent advances in minimally invasive surgery (MIS) have allowed patients to benefit from reduced trauma, faster recover times, and shorter hospitalisation. As a result, surgeons operate in a less-than-comfortable posture which is both physically and mentally challenging. This has led to a more demanding training scheme to acquire the relevant surgical skills. Navigating and operating with flexible instruments such as endoscopes can also induce spatial disorientation within the surgeon, where such instances are associated with increased pain for the patient and more critically, risk of perforating delicate patient tissue. Furthermore, the increased physical separation between the surgeon and the operative site and newly introduced surgical instruments have significantly changed the ergonomics and surgical workflow, which in turn increases the cognitive burden on the surgeon.

The perceptual and ergonomic challenges during flexible endoscope based MIS are investigated through using the wealth of perceptual information the gaze can provide. In particular, the visualisation, navigation and ergonomic issues during MIS procedures are studied. A gaze parameter based framework is introduced to assess the use of a new field-of-view expansion technique for improved visualisation and camera trajectory comprehension when disorientated. Flexible instruments such as endoscopes suffer from disorientation inducing perceptual-motor misalignment. A misalignment quantification approach and a classification method based on gaze to infer varying perceptual-motor misalignment are developed to enable ergonomic assessment during endoscopic procedures. A novel robotic assisted gaze controlled camera system is developed to improve camera navigation where the user's control intentions are conveyed to the robotic laparoscope via real-time gaze gestures. To further improve the ergonomics of the gaze contingent system, an online calibration algorithm is integrated into the system. Throughout the thesis, detailed validation and discussion of the results are conducted to demonstrate the potential clinical value of the work.

Acknowledgements

The work in this thesis was made possible with help from family, friends and colleagues. I am deeply grateful to my supervisor Professor Guang-Zhong Yang for continuously believing in my work, and for giving me this once in a life time opportunity to work in this exciting field.

I thank my colleagues with whom I have worked: Kumuthan Sriskandarajah, Johannes Totz, Antonino Salerno, Ka-Wai Kwok, George Mylonas, Kunal Shetty, and Colin Sugden. Without them, neither the chapters of this thesis nor the associated papers would have become a reality. I also thank my colleagues whom I have not worked with directly: Menglong for being an amazing friend and favourite coffee break buddy. Marco Visentini-Scarzanella for always injecting positivity. Matina Giannarou for always being helpful. Konrad Leibrandt and Gauthier Gras for always willing to brainstorm ideas. Petros Giataganas for being resourceful and the go to person for Kuka issues.

My deepest gratitude is towards my close friends and family. Bainee for always uplifting my spirit with positive energy and providing useful views to my new ideas, thank you! Pan for being a wonderful friend and attentive listener. Charlie for being the best graphic illustrator teacher. Kenei and Hitomi for being supportive all the time. And lastly to my parents who always love and supported me throughout my life.

Contents

1 Introduction **29**

2 Eye Tracking Technology and Applications to Minimally Invasive Surgery **37**

2.1 Anatomy of the Human Visual System 37

2.1.1 Foveal and Parafoveal Vision in the Human Visual System 39

2.1.2 Characteristics of Eye Movements 40

2.2 Overview of Eye Tracking Techniques 43

2.2.1 Feature-Based Eye Trackers 45

2.2.1.1 Detection of Eye Features 46

2.2.1.2 Calibration Procedure 49

2.2.1.3 2D Regression-Based Eye Trackers 52

2.2.1.4 3D Model-Based (Geometrical) Eye Trackers 53

2.2.2 Calibration Drift and Practical Issues 59

2.2.3 Calibration Free Eye Tracking 59

2.3 Surgical Eye Tracking Applications 60

2.3.1 Visual Perception and Disorientation 62

2.3.2 Gaze Contingent Perceptual Docking 63

2.4 Conclusions 65

3 Eye Tracking Applied to Endoluminal and Transluminal Surgery **67**

3.1 Endoscopy Procedure 68

3.2 Challenges in Endoscopy 69

3.2.1 Endoscope Visualisation and Orientation 70

3.2.2 Varying Perceptual-motor Misalignment during Navigation 71

3.3 Endoscopy Technological Advancements 76
3.3.1 Retroscopes 76
3.3.2 Narrow Band Imaging and Confocal Optical Microscopy 77
3.3.3 Scope Guide - External Endoscope Shape Visualisation 78
3.3.4 Overtubes and Variable Stiffness Endoscopes 78
3.3.5 Robotic Endoscopes 79
3.4 Case Study: Characteristics of Gaze Behaviour during Upper and Lower Endoscopic Navigation 80
3.4.1 Experimental Setup 81
3.4.2 Experimental Protocols 81
3.5 Results 82
3.5.1 Gaze Comparison during Upper and Lower Endoscopy 83
3.5.2 Poor Camera Visualisation 86
3.5.3 Retroflexion 88
3.5.4 Looping 91
3.6 Conclusions 94

4 Usability Study of Dynamic View Expansion for Surgical Visualisation 97
4.1 Methods 99
4.1.1 Dynamic View Expansion 100
4.1.2 Markov Chain and State Transition Graphs 102
4.2 Experimental Setup 104
4.3 Results 105
4.3.1 Response Correctness and Latency 105
4.3.2 Visual Search Behaviour Characterised by State Transition Diagrams 110
4.4 Conclusions 116

5 Perceptual-Motor Misalignment during Surgical Navigation 121
5.1 Methods 124
5.1.1 Perceptual-Motor Misalignment Angle 125
5.1.2 Task Performance Metrics and Gaze Parameter Analysis 128
5.1.3 Classification of Gaze Behaviour 130
5.2 Experimental Setup 133

5.3 Results . . . 136
5.3.1 Statistical Analysis . . . 137
5.3.2 Classification of Instances of Variable Perceptual-Motor Misalignment . . . 141
5.4 Conclusions . . . 143

6 Gaze Contingent Laparoscope for Surgical Navigation 147
6.1 Methods . . . 149
6.1.1 Gaze Gesture Recognition Process . . . 151
6.1.2 Robot Control Process . . . 155
6.2 Implementation . . . 162
6.2.1 Gaze Contingent Control Modes and UI . . . 162
6.2.2 Experimental Setup . . . 164
6.2.3 Experimental Protocols . . . 165
6.3 Results . . . 168
6.3.1 HMM Gaze Gesture Performance Assessment . . . 168
6.3.2 Comparative Analysis of the Three Control Modalities 169
6.4 Conclusions . . . 172

7 Online Calibrated Gaze Contingent Control 175
7.1 Gaze Contingent System Modifications . . . 177
7.1.1 Updated Gaze Gesture Recognition Process . . . 178
7.1.2 Updated Robot Control Process . . . 179
7.2 Online Calibration Algorithm . . . 180
7.3 Implementation . . . 188
7.3.1 Gaze Contingent Control Modes . . . 189
7.3.2 Experimental Setup . . . 192
7.3.3 Experimental Protocols . . . 193
7.3.3.1 Robotic Laparoscope System Usability Study . . . 193
7.3.3.2 Gaze Tracker Online Calibration Performance Study . . . 196
7.4 Results . . . 198
7.4.1 Robotic Laparoscope System Usability Studies . . . 199
7.4.1.1 Updated HMM Gaze Gesture Performance Assessment . . . 199

7.4.1.2 Comparative Analysis of the Three Control Modalities 199
7.4.1.3 Comparative Analysis of the First and Second Trial 205
7.4.1.4 Comparative Analysis against the Gaze Contingent System in Chapter 6 208
7.4.2 Performance of Online Gaze Calibration Algorithm . . 210
7.5 Conclusions 213

8 Summary and Conclusions **217**
8.1 Contributions of this Thesis 218
8.2 Ongoing Work and Future Perspectives 220

Bibliography **225**

List of Figures

2.1 A horizontal cross section of a human eye. Figure adapted from [26]. 38

2.2 θ represents the angular divergence (°). D represents the distance from the subject's eye to the gaze tracker screen (cm), which is outputted by the gaze tracker system. S represents the distance offset from a reference point (cm). 40

2.3 Density distributions of rod and cone receptors across the surface of the retina. Cone density is greatest at the fovea and falls sharply to a minimum by 10° from the eye fixation point. Beyond 10°, the retina contains a thin, uniform distribution of cones. Rods are not present in the fovea. They are most dense in the near periphery out to 18°. While they decline in number further into the periphery, they are still far denser than cones and dominate vision in the peripheral visual field. The blind spot, the region of the retina where the optic nerve leaves the eye, contains no photoreceptors. Graph adapted from [9]. 41

2.4 (a) An example of a scleral contact lens with a suction ring mechanism. Image courtesy of Skalar Medical, Delft Holland. The picture shows the invasiveness of this method of eye tracking. (b) An example of an EOG eye tracker. The electrodes are placed directly on the skin surface. It is considered an inexpensive technique, but is intrusive in nature. Image courtesy of Macrovision. 44

2.5 Categorical diagram of non-invasive video based eye trackers. 45

2.6 Feature-based commercial eye trackers. (a) The Tobii PCEye Go eye tracking device for laptops and desktop computers by Tobii Technology AB (b) The SMI Eye Tracking Glasses 2.0, an example of a head mounted eye tracker by SensoMotoric Instruments GmbH. 46

2.7 A general flow chart of feature-based eye trackers. Features of the eye are detected and tracked. A calibration procedure is then conducted to obtain a mapping correspondence between extracted features to screen coordinates, where subsequently the user's PoR can be estimated using the mapping obtained from the calibration procedure. 47

2.8 (a) Bright pupil is observed when the incident light ray is on-axis. (b) Dark pupil is observed when the incident ray is off-axis. (c) Thresholded difference of the two images gives a clear segmentation of the pupil. 48

2.9 Purkinje reflections are distinctive bright spots captured on video when a fixed infrared light source is directed on the eye. Four reflections are captured including; P1, reflection from the outer surface of the cornea also commonly referred to as the "glint"; P2, reflection from the inner surface of the cornea; P3, reflection from the outer (anterior) surface of the lens; P4, reflection from the inner (posterior) surface of the lens. 49

2.10 Calibration procedure for an eye tracker involves displaying a dot at nine regularly spaced coordinates, one at a time on the screen, where the user is required to fixate on them. During this procedure, the distinctive eye feature coordinates (pupil and glint locations) for each nine points are recorded with the corresponding screen coordinates. A polynomial function can be used to establish the correspondence relationship between the eye image and screen plane. 51

2.11 The general model of the eye used in 3D model-based eye trackers. The cornea is modelled as a perfectly spherical surface. By positioning the camera and light source in known geometrical locations, estimation of the 3D location of the cornea centre and pupil centre can be obtained. Through connecting the cornea centre with the pupil centre, the optical axis direction is obtained. A calibration procedure is used to obtain the angular offset between the optical and visual axes, allowing for estimation of the user's PoR. A regression-based calibration maps the point of intersection of the optical axis with the screen to the true gaze screen coordinates. 55

3.1 Illustration of oesophagogastroduodenoscopy (OGD) and the stomach anatomy. 69

3.2 Illustration of lower endoscopy (colonoscopy) and the human colon anatomy. 70

3.3 Illustration of how the sigmoid colon can move and stretch from (a) the point of entering the anus (b) to insertion into the sigmoid colon and (c) if pushed further can form a loop. 71

3.4 Illustration of a retroflexed endoscope with the respective visualisation. 72

3.5 Illustrations of the different sigmoid and transverse loops that occur during colonoscopy. (a) sigmoid spiral "N" loop (b) sigmoid alpha loop (c) sigmoid reverse alpha loop (d) sigmoid counter-clockwise spiral loop (e) transverse deep loop (f) transverse deep gamma loop. 75

3.6 Third Eye retroscope by Avantis Medical Systems, Inc. (a) image of the Third Eye retroscope, (b) illustrates how the Third Eye retroscope can easily detect lesions behind folds, (c) illustrates the endoscopist's view of the forward camera view on the left and the Third Eye retroscope view on the right. 77

3.7 Scope Guide system by Olympus Corp. (a) shows the endoscope stack and how the overall endoscope shape can be visualised, (b) the magnetic field emitter and tracker enables the endoscopist to visualise the endoscope shape without having to use harmful radiation such as in fluroscopy. 78

3.8 Viewing from top left to bottom right, the video sequence illustrates the poor camera visualisation that can occur during endoscopic procedures. The small FOV combined with the paradoxical endoscope tip movement during looping can make it difficult to comprehend the endoscope's position from the camera visualisation alone. This particular colonoscopy video sequence was 45 seconds in duration. 87

3.9 Viewing from top left to bottom right, the video sequence is the magnified camera visualisation of the video sequence shown in Fig. 3.8. Poor visualisation can be caused by the injected water during bowel cleansing causing multiple colour reflections. Even when the camera visualisation is clear, the small FOV makes it difficult to fully visualise the gastric wall immediately surrounding the endoscope tip. 89

3.10 Viewing from top left to bottom right, this video sequence shows an example visualisation of the endoscope retroflexed during an upper endoscopic procedure. The video sequence consists of examination of the fundus and gastroesophageal junction. 90

3.11 Viewing from top left to bottom right, this video sequence illustrates looping of the endoscope in the patient's lower gastrointestinal tract. During looping, camera visualisation is poor and the endoscopist frequently gazes between the Scope Guide monitor and the camera visualisation screen. The endoscopist uses a magnetic tracker and places this on the patient's abdomen to isolate the colonoscope tip. The position of the tracker shows up on the Scope Guide monitor which can help the endoscopist understand where the endoscope tip is within the patient's gastrointestinal tract. 92

3.12 Viewing from top left to bottom right, the video sequence is a continuation of the looped episode from Fig. 3.11. The endoscopist who is still looped, continues to move the endoscope whilst cross referencing between the Scope Guide and endoscope camera visualisation. The endoscopist adjusts the stiffness of the endoscope where he fixates at the endoscope's stiffness controls to try to get out of the loop. The endoscopist eventually overcomes the loop. 93

4.1 Experimental setup showing a study participant conducting the experiment in front of a Tobii 1750 gaze tracker used to collect the data. 99

4.2 (a) The fade-to-grey DVE texture processing flow chart. Video frames are processed with SLAM to estimate the surgical scene's geometry and create a non-planar panorama with older video frames. (b) An example of a fade-to-grey DVE image. (c) Same example images with conventional full-colour DVE technique. 100

4.3 Segmented regions on example fade-to-grey DVEs of the three sequences used. 102

4.4 (a) One fade-to-grey DVE image with (b) the corresponding trajectory options. 105

4.5 Comparison of response correctness and response latency between the two groups of images. (a) response correctness obtained when viewing full-colour and fade-to-grey DVE images. (b) response latency obtained when viewing full-colour and fade-to-grey DVE images. 106

4.6 Comparison of correctness between the two groups of images. (a) Response correctness for Group 1 images. The percentage of the correct responses when using the fade-to-grey DVE method is a higher than when using full-colour DVE. (b) Response correctness for Group 2 images. The correctness for Group 2 images are similar, albeit, the fade-to-grey method achieving slightly better performance. 109

4.7 Comparison of latency response measured in seconds between the two groups of images. (a) Response latency for Group 1 images. The response latency for Group 1 images is similar between the two methods, (b) Response latency Group 2 images. It took longer for the subjects to respond using the conventional full-colour DVE method, while with the fade-to-grey DVE images, the response latency is reduced. 109

4.8 Fade-to-grey DVE images 7-9 (a)-(c) with respective camera trajectories overlaid. The overlap of the live video frame with the trajectory introduces a discontinuity in the fade-to-grey expanded view, potentially making these images more challenging to estimate the trajectory. 111

4.9 Image 5 (a) full-colour DVE and (b) fade-to-grey DVE. Some of the limitations of this DVE visualisation technique is observable in this image. 112

4.10 Markov state transition graphs for Group 1. There are notable transitions in the fade-to-grey DVE state transition graphs between states 1 and 3 for (a) and (b), and state 1 and 2 for (c) and (d) which are faint in the corresponding full-colour DVE state transition graphs. 114

4.11 Comparison of the two DVE visualisation methods and the respective gaze heat-maps for Image 3 (a) Fade-to-grey DVE method. Fixations are more concentrated around the areas of interest and in particular in regions 1 and 4 (b) Conventional DVE method. Fixations are more scattered 115

4.12 Markov state transition graphs for Group 2. (a)-(c) State transition probabilities have similar characteristics of cross-referencing between key regions of the image. (d) For this particular image, participants were able to use the blur present in the full-colour version to infer the camera trajectory. . . . 117

4.13 Comparison of the two DVE visualisation methods and the respective gaze heat-maps for Image 2. The density of fixations across the regions of interest for both methods is similar. (a) Fade-to-grey DVE method. (b) Conventional DVE method.118

5.1 Illustration of the user experiment setup for this study. The user control arrangements and the simulated articulated robot inside the colon are shown. 125

5.2 Illustration of how the disparity between the camera image plane and the instrument end-effector position during motor coordination can lead to a misalignment angle between the surgeon's intended motor input and the resultant camera motion that is subsequently observed. (a) Perceptual-motor misalignment when controlling a flexible endoscope device. (b) Perceptual-motor misalignment when controlling the articulated robotic endoscope that is used in the study of this chapter. 127

5.3 Illustration of the colonoscopy task. (a) The camera view from the distal-tip shows the gastrointestinal tract task environment. (b) Illustration of the external view of how the articulated robotic endoscope progresses through the colon. (c) The polyp removal scene showing optimal positioning indicated by a large yellow circle. (d) Participants' eye tracking gaze overlaid as a heat-map during the polyp removal scene. Note the focus of participants' interest on the polyp and towards the lumen. 135

5.4 Performance improvement of having no perceptual-motor misalignment compared to having varying perceptual-motor misalignment during navigation for each performance metric. . . 138

5.5 Time normalised saccade count by control mode aggregating all participants. 139

5.6 Time normalised fixation count by control mode aggregating all participants. 139

5.7 Average fixation duration by control mode aggregating all participants. 141

5.8 Illustration of the separation hyperplane, defined from the SVM, between the instances of variable perceptual-motor misalignment (blue) and the instances of no misalignment (red). The support vectors on the boundary of the classification is depicted with green. It is observed that fewer saccades and fixations with longer fixation durations are linked to varying perceptual-motor misalignment. 142

6.1 A surgeon using the gaze contingent camera system. The operator is able to simultaneously navigate and perform a bi-manual surgical task. 150

6.2 Illustration of the (a) '*pan*' and (b) '*zoom*' gaze gestures for the gaze contingent laparoscope system. 150

6.3 The gaze contingent system's internal control mechanism can be broken down into two processes: the gaze gesture recognition process and the robot control process. 151

6.4 Flow chart of the gaze gesture recognition process. The stream of the user's raw PoR's are filtered from which the gaze gestures are segmented and then encoded. This encoded observation sequence is tested for inference of both gaze gesture HMMs. If one of the two possible gaze gestures is recognised, then the camera is activated and PoR information is sent to the robotic arm. 153

6.5 The recognition accuracy and false positive rate across different number of states for the (a) '*pan*' gaze gesture and (b) '*zoom*' gaze gesture. The inference values of unintended gaze gestures and the respective inference values of intended gaze gestures for (c) '*pan*' gestures and (d) '*zoom*' gestures respectively. 156

6.6 The position of the surgeon's PoR on the gaze tracker screen is used to update a minimum jerk trajectory planner providing the desired pose x_d to the robot controller based on a Cartesian impedance control scheme. The controller computes the command torque t_c for each joint according to the compensation of the whole robot dynamics and the error signal e estimated as the difference between the reference pose x_d and the actual pose x in Cartesian space retrieved by the robot's forward kinematics. 158

6.7 (a) Reference frames placement; (b) RCM constraint. 159

6.8 Gaze gesture controlled robotic laparoscope UI. (a) The UI waits for the user to perform a '*zoom*' or '*pan*' gaze gesture. If a '*zoom*' gaze gesture is detected then the user will be taken to (b), or if a '*pan*' gesture is detected it will take them to (c). While in (b) or (c) the camera can be stopped by looking into the white circle for $750ms$. After the camera stops, the user is taken back to (a). (d) Illustration of the camera velocity and respective regions under '*pan*' camera control mode. . . 163

6.9 The experimental setup includes a commercial eye tracker (Tobii 1750), a $10mm$ zero degree Stortz laparoscope and Tele Pack light box, a Kuka LWR and a upper gastrointestinal phantom with simulated white lesions. 165

6.10 Training task requesting participants to navigate the robotic laparoscope within a conventional box trainer to find numbers on a 4×5 grid in ascending magnitude. 166

6.11 Group task completion times by control modality. 170

6.12 Camera path for a given subject during (a) gaze gesture based gaze contingent control and (b) camera assistant control of the laparoscope. 170

6.13 Group camera path lengths (cm) by control modality. The camera path length obtained during the gaze gesture control mode showed significantly shorter path length against the camera assistant. 171

6.14 Group NASA-TLX scores by control modality. No significant difference in NASA-TLX scores were observed between the gaze gesture and pedal modality against the human camera assistant. 172

7.1 The gaze gesture sequence for the (a) '*activate camera*' (pan and zoom) and (b) '*tilt camera*' (rotation) gaze gestures enabling activation of the gaze contingent laparoscope system. 178

7.2 HMM training 10-fold cross validation for the (a) '*activate camera*' and (b) '*tilt camera*' HMM. The inference values of intended gaze gestures for (c) '*activate camera*' gestures and (d) '*tilt camera*' gestures. 179

7.3 The reference frames placement for the robotic platform (Kuka LWR). 181

7.4 Flow chart of the online calibration algorithm. 183

7.5 (a) Calibration point extrapolation process when a '*activate camera*' gaze gesture is performed on initialisation of the system. (b) Calibration point extrapolation process when a '*tilt camera*' gaze gesture is performed on initialisation of the system. (c) Resultant calibration mapping showing the extrapolated and original calibration points once both gaze gestures have been performed. (d) The five screen coordinate locations used for the TPS calibration mapping. 188

7.6 Illustration of the UI for the gaze contingent laparoscope system. (a) Panning during '*activate camera*' mode. (b) Zooming in during '*activate camera*' mode. (c) Zooming out during '*activate camera*' mode. (d) Tilting the camera during '*tilt camera*' mode. (e) Camera is stopped by gazing at the bottom left hand corner for $750ms$. 191

7.7 The experimental setup for the modified gaze contingent laparoscope system uses equipment from the previous chapter, but in addition an NDI Polaris infrared tracker together with reflectance markers attached onto both instrument handles to record the subject's instrument tip trajectory data. 193

7.8 (a) Panning and zooming training task where participants had to control the robotic laparoscope to navigate within a box trainer to find numbers on a 4×5 grid in ascending order. (b) The rotation task involved the participants tilting the camera view left or right to correctly orientate the camera. . 195
7.9 Task completion time by control mode aggregating all participants. 200
7.10 Camera path length by control mode aggregating all participants. 202
7.11 Camera workspace during each control modality. 203
7.12 Left and right instrument path length by control mode aggregating all participants data. 204
7.13 NASA-TLX graph scores for all participants comparing the new gaze contingent UIs with the camera assistant. 205
7.14 Comparison chart of the accuracy obtained when the gaze tracker has no calibration, the online calibration (after one, two, five and ten gaze gestures are performed) and after the five and nine point offline calibration. The online calibration accuracy performance is comparable to that of the offline calibration with as little as one gaze gesture being performed. 212
7.15 Comparison chart of the precision obtained when the gaze tracker has no calibration, the online calibration (after one, two, five and ten gaze gestures are performed) and after the five and nine point offline calibration. The precision of the gaze tracking is maintained longitudinally via the online calibration algorithm. 213

8.1 Illustration of the four most common patterns in the colon (a) normal, (b) adenocarcinoma, (c) adenoma, (d) hyperplastic. Images adapted from [157] 222

List of Tables

2.1 Summary of 2D regression-based and 3D model-based feature eye trackers with their respective key characteristics. 61

3.1 Gaze parameters of the surgeon averaged across the whole procedure. 84
3.2 Percentage of fixation time by the surgeon on each respective AOI during the upper and lower endoscopic procedures. . . . 85
3.3 Intubation time and NASA-TLX score of the surgeon during the upper and lower endoscopic procedures. 86
3.4 Gaze parameters of the surgeon during retroflexion in upper endoscopic procedures. 91
3.5 Gaze parameters of the surgeon during looping in lower endoscopic procedures. 94

4.1 Group-averaged percentage of correct responses and response latency over all images with respective standard deviations. On average, participants answered more correctly and quickly when observing fade-to-grey DVE images to determine the camera trajectory. 106
4.2 Percentage of correct responses and response latency over all images separated into surgeons and non-surgeons. 107
4.3 Group-averaged percentage of correctness and response latency per image, for each DVE method. 107

5.1 Average performance metrics of the group with respective standard deviation values and ANOVA test p-values. 137
5.2 Median and interquartile range values of gaze parameters of the nineteen participants with respective Mann-Whitney U test p-values. 140

6.1 Motion parameters for each robot movement. 161
6.2 Recognition accuracy and the false positive rate of the HMMs during the experimental trial. 169
6.3 Comparative analysis of the three control modalities. Median and interquartile range in parentheses. 169

7.1 Motion parameters for each region of the gaze tracker screen. Time (s) and lengths (mm). 181
7.2 The recognition accuracy and the false positive rate for both HMM gaze gestures. 199
7.3 User performance metrics of the assistant, gaze gesture activated and pedal activated system over both trials. 201
7.4 First and second trial user performance metrics of new laparoscope control scheme (gaze gesture and pedal) compared to the camera assistant. 206
7.5 User performance comparison of proposed system versus the previous gaze contingent system introduced in Chapter 6. . . 209
7.6 Comparative performance of online, offline and no calibration. 211
7.7 The recognition accuracy and the false positive rate for both HMM gaze gestures during use of the online gaze tracker calibration algorithm. 213

List of Symbols and Acronyms

AESOP	Automatic Endoscope Optimal Position
ANOVA	Analysis of Variance
AOI	Area of Interest
ASGE	American Society for Gastrointestinal Endoscopy
CGC	Collaborative Gaze Channelling
CR	Cross-Ratio
CT	Computer Tomography
DoF	Degrees of Freedom
DVE	Dynamic View Expansion
EOG	Electro-Oculography
ESD	Endoscopic Submucosal Dissection
FOV	Field of view
HMM	Hidden Markov Model
HVS	Human Visual System
IQR	Inter-Quartile Range
JAG	Joint Advisory Group
LWR	Light Weight Robot
MEMS	Microelectromechanical systems
MIS	Minimally Invasive Surgery

NASA-TLX	National Aeronautical Space Agency - Task Load Index
NOTES	Natural Orifice Transluminal Surgery
OA	Optical Axis
OGD	Oesophagogastroduodenoscopy
PCCR	Pupil Center minus Corneal Reflection
pCLE	probe-based Confocal Laser Endomicroscopy
PDA	Personal Digital Assistant
PGY	Post Graduate Year
PoR	Point of Regard
RBF	Radial Basis Function
RCM	Remote Centre of Motion
ROI	Region of Interest
SILS	Single Incision Laparoscopic Surgery
SLAM	Simultaneous Localisation and Mapping
SMI	SensoMotoric Instruments
SVM	Support Vector Machine
TPS	Thin Plate Spline
UI	User Interface
VA	Visual Axis
VF	Virtual Fixture

1 Introduction

Technical innovation in surgery has brought significant benefits for patients since the days of open surgery. Ever since the first laparoscopic cholecystectomy was performed in 1985 [166], minimally invasive surgery (MIS) has gained widespread adoption. By 1992, it is reported that over 70% of cholecystectomies were conducted laparoscopically throughout the United States, Europe and Japan [153]. In the era of MIS patients are offered better cosmetic results, less post-operative pain and shorter recovery times. The benefits of requiring less post-surgical medication and shorter hospitalisation have also had a positive impact for hospitals and healthcare providers. The advantages of performing surgery using less invasive techniques has induced surgeons to translate MIS to other surgical procedures including, cardiothoracic, neurosurgery, urology, and general surgery [33]. Although the advantages of minimally invasive techniques are clear, there are still challenges, many of which shift the burden to the surgeon carrying out the surgical procedure. The intrinsic setup of MIS requires specialised equipment, which introduces a number of changes in the way the surgeon accesses, observes and manipulates tissue and in turn increases operation complexity and affects the surgeon's performance. As a result, MIS procedures affect the operator ergonomics making it more taxing on the surgeon. Introduction of robotic equipment into the surgical theatre has improved some of the ergonomic issues associated with MIS, but has its own disadvantages such as the increased physical separation of the surgeon from the operative site and the associated high costs, which have hindered wide spread adoption of robotic surgical systems.

The aim of this thesis is to investigate and solve the challenges experienced by surgeons during MIS. Eye tracking can provide a wealth of perceptual information and has been previously adopted in a number of areas. Through using the information obtainable from human eye movements, we aim to

understand the perceptual and ergonomic challenges during MIS and to subsequently develop solutions to overcome the following issues:

- *Visualisation*: the key hole nature of laparoscopic instruments has introduced indirect visualisation of the operative site via cameras. The camera visualisation can suffer from a small Field of View (FOV) and low image quality which is inferior to direct visualisation of the operating scene with the naked eye during open surgery.

- *Navigation*: small incisions and miniaturised instruments make navigation to the operative site more difficult, requiring more surgical training and development of dexterous skills as well as the need to develop further complex instruments that can translate these movements.

- *Ergonomics*: the increased physical separation between the surgeon and the operative site, and newly introduced surgical instruments has significantly changed the surgeon's ergonomics and surgical workflow. This in turn increases the burden on the surgeon and requires the need for further training.

In this thesis, the gaze information from the human, and the accuracy of robotics are utilised to improve the camera navigation experience for the surgeon. The consistency and repeatability of the robotic components under long surgical procedures, combined with the cognitive decision making capabilities of the surgeon has the potential to enable a framework that offers the advantages of both whilst the surgeon maintains full control of the robotic assistance. The following paragraphs outline the chapter structure of this thesis.

Chapter 2 reviews the background, theoretical foundation and current state-of-the-art implementations of eye tracking. Specifically, the review focuses on non-invasive feature-based eye trackers. The challenge in eye feature extraction and tracking lies in the robustness of the performance during fast eye rotations, blinking, head motion and during changes of external illumination conditions. Technical details behind the calibration of feature-based eye trackers are also discussed. Subsequently, the current state-of-the-art eye tracking techniques are then reviewed along with their

respective advantages and disadvantages focusing on the issue of calibration and gaze estimation accuracy. Interestingly, to date, there is still no eye tracking method that does not require an explicit calibration procedure. From the review, it is concluded that the accuracy of a calibration-free eye tracking technique will depend on head motion range, camera calibration, number of calibration points, number of cameras and light sources. This chapter ends with a review of the previous eye tracking surgical applications for enhancing the surgeon's visual perception and ergonomics whilst performing a MIS procedure for better patient safety and surgical outcomes.

Chapter 3 reviews previous literature on endoscopic procedures to understand the main challenges when performing such procedures. The aim of the chapter is to investigate the difficulties faced by surgeons in the operating theatre, and to distinguish the main issues, in particular when performing endoscopy. A case study is conducted where gaze tracking is utilised to understand the surgeon's visual behaviour and ergonomics during real endoscopic procedures. Video data of the endoscopic suite and eye tracking data are recorded, where three main areas of improvement are identified. These include, the visualisation, endoscope visual-motor control, and ergonomics. Addressing these challenges are important for endoscopy, as well as for any MIS procedure. The results from the study illustrate the potential in utilising the surgeon's gaze information to understand the challenges of a surgical procedure, as the fixation duration and surgeon's fixation on particular areas of interest (AOI) can provide cues to their cognitive state. The subsequent chapters aim to address the three aspects which were identified in this chapter.

Chapter 4 aims to address the visualisation issues during MIS caused by the narrow FOV. In order to alleviate the problem of the narrow FOV, an image expansion technique called Dynamic View Expansion (DVE) can be used. Recently, a new mosaic based image view expansion technique called the fade-to-grey DVE was introduced by Totz *et al.* [154]. This image expansion technique differs from conventional full-colour DVE imaging methods as it provides the operator with temporal visual cues of the recent camera trajectory to enable the operator to infer the camera movement for reorientation. In this chapter, we propose a new framework using the gaze information to test the usability of the two DVE image visualisation

techniques. By modelling the visual attention strategy of both DVE image techniques with Markov chains, differences in participant search behaviour of both DVE methods were quantified. In addition, performance metrics including the trajectory estimate and response times were assessed. Results highlight the benefits of the new fade-to-grey DVE image visualisation technique during camera trajectory estimation with statistical significance in performance improvement versus conventional DVE methods. The results demonstrate that not only is the new fade-to-grey DVE technique able to extend the FOV for the surgeon, but there is potential clinical value of visualising these DVE images during identification of the camera trajectory for subsequent reorientation. The work in this chapter illustrates how eye tracking can be used for visual search behaviour assessment during a task.

Chapter 5 focuses on investigating the issues related to visual-motor control during endoscopic navigation. A quantification method of the perceptual-motor misalignment is introduced and gaze data recorded during the study was used to extract relevant features to train a Support Vector Machine (SVM) classifier. Through gaze data, application of this classifier allows inference of the instances with varying perceptual-motor misalignment during an endoscopy without the need to measure the misalignment angle. This is in contrast to conventional endoscopy where such measurement is difficult to determine due to lack of predefined kinematic structure. More importantly it provides a low cost means to infer challenging instances caused by varying perceptual-motor misalignment through tracking only the user's gaze information. This chapter also investigates how a variable perceptual-motor misalignment can affect task performance by creating an articulated robotic endoscope simulation environment. The results from the study demonstrate that varying perceptual-motor misalignment can have a negative impact on the task completion time, travel distance as well as finer endoscopic manoeuvres. The findings from this chapter, therefore, provides an important quantitative basis for assessing the ergonomics, as well as perceptual behaviour and potential cognitive loadings of future flexible devices, including robotic systems for emerging surgical techniques such as single port access or natural orifice MIS procedures. Furthermore, the obtained results highlight the potential usefulness of articulated robotic endoscopes, since robotic control has the ability to continuously re-align any perceptual-motor mis-

alignment.

In **Chapter 6**, a new gaze controlled camera system is introduced with the aim to improve navigation and ergonomics of MIS procedures by using eye tracking. A laparoscope robotic arm holder is developed taking into account previous results from Chapter 5. This gaze controlled camera system enables the operator to control the 6 Degrees of Freedom (DoF) laparoscope robotic arm through the surgeons' Point-of-Regard (PoR) as the control input. To overcome the fulcrum effect, attention was paid so that there is no misalignment between the resulting camera visualisation movement with the user input control, thus providing a natural camera control. To improve the surgeon-robotic interface ergonomics, real-time Hidden Markov Model (HMM) were used to recognise gaze gestures and activate the gaze contingent camera control. Through implementation of multiple gaze gestures, surgeons were able to convey their desired camera control mode via the use of pan and zoom gaze gestures, thus eliminating the need for external physical input devices or a camera assistant. User trials were conducted to demonstrate and assess the performance of the gaze gesture control modality. It was found that surgical navigation tasks were easily reproducible and complete. The results also supports the claims that the proposed gaze contingent robotic camera system with built-in safety mechanisms are ergonomic to use, inducing no significant cognitive burden when compared to using a camera assistant or when compared to using a foot pedal activated camera system. This means that the participants did not feel that the gaze gestures and gaze control were a complex user modality. More importantly, there is potential cost reduction opportunities through using such a system as it can reduce the need for a highly skilled camera assistant and allow a surgeon to operate by them self, providing cost and staffing efficiencies.

Chapter 7 further expands the system developed in the previous chapter with the addition of pan, tilt and zoom capabilities. The new system's camera motion is activated with gaze gestures but now has an efficient switching mechanism between panning and zooming in / out through a combination of gaze and head position tracking. Further improvements to the usability and ergonomics of the system are introduced by integrating an online gaze tracker calibration scheme, which allows the user to immediately start using the gaze contingent system. Notably, the online nature of the calibration

will enable maintaining an accurate and precise PoR estimation, avoiding the calibration drift problem as discussed in Chapter 2. A comprehensive usability study involving seventeen surgical residents is conducted to assess the modified control scheme of the gaze gesture activated robotic laparoscope system. The results demonstrate the system to be superior than instructing a camera assistant across various metrics. Our proposed system also removes the need for a foot pedal, which exists in many legacy systems, improving the ergonomics of the operation setting. Analysis of the camera workspace occupied during the user trials demonstrates that the gaze contingent laparoscope system is able to navigate a similar working volume to that of a human camera assistant. Moreover, the learning effect of using the system is investigated, and analysis showed that participants were able to learn to use the gaze contingent system quickly and outperform the camera assistant in many of the performance metrics by their second trial. A comparison of the gaze gesture activated system introduced in this chapter to that of the system presented in Chapter 6 is also conducted. The group of users who used the proposed gaze gesture system in this chapter produced significantly faster task completion times. The results from the various usability studies of the new gaze gesture activated robotic laparoscope system highlight the clinical value of the proposed system.

Finally, **Chapter 8** concludes the thesis by outlining future research directions and the grand challenges of integrating the proposed concepts for clinical application in robotic MIS.

The work presented in this thesis has resulted in a number of publications in peer reviewed international journals, and conference proceedings. It has also been presented at a number of international conferences and scientific symposiums. The main publications related to this thesis include:

- "The disparity between the neurobiology of classical neuropsychological vigilance and intra-operative vigilance in surgery: a functional Near Infrared Spectroscopy study.", Colin Sugden, Daniel Leff, Felipe Orihuela-Espina, James DRC, Kenko Fujii, Ara Darzi and Guang-Zhong Yang. Human Brain Mapping, 2011.
- "Stealth Calibration Eye Tracking Algorithm for Minimally Invasive Surgery.", Kenko Fujii, George Mylonas and Guang-Zhong Yang.

Hamlyn Symposium on Medical Robotics, 2011.

- "Enhanced visualisation for minimally invasive surgery.", Johannes Totz, Kenko Fujii, Peter Mountney, and Guang-Zhong Yang. International journal of computer assisted radiology and surgery 7, no. 3 (2012): 423-432.

- "Visual search behaviour and analysis of augmented visualisation for minimally invasive surgery.", Kenko Fujii, Johannes Totz, and Guang-Zhong Yang. Augmented Environments for Computer-Assisted Interventions, pp. 24-35. Springer Berlin Heidelberg, 2012.

- "Effects of Perceptual-Motor Misalignment: A Flexible Robotic Navigation Study.", Kenko Fujii, Ka-Wai Kwok and Guang-Zhong Yang. Modeling and Monitoring of Computer Assisted Interventions (M2CAI), 2012.

- "Gesture based Gaze Contingent Control of a Robotic Arm for Laparoscopic Applications.", Kenko Fujii, Antonino Salerno, Kumuthan Sriskandarajah, Ka-Wai Kwok, Guang-Zhong Yang. Hamlyn Symposium on Medical Robotics, 2013.

- "Gaze contingent cartesian control of a robotic arm for laparoscopic surgery.", Kenko Fujii, Antonino Salerno, Kumuthan Sriskandarajah, Ka-Wai Kwok, Kunal Shetty, and Guang-Zhong Yang. Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on. IEEE, 2013

- "Gaze Gesture based Human Robot Interaction for Laparoscopic Surgery.", Kenko Fujii, Kumuthan Sriskandarajah, Antonino Salerno, and Guang-Zhong Yang. Human–Computer Interaction, 2014. (under review)

2 Eye Tracking Technology and Applications to Minimally Invasive Surgery

Tracking of the human gaze has provided a means to understand human visual search behaviour and visual perceptual ergonomics [174, 52]. The ergonomic factors intrinsic to MIS mean that surgeons today experience greater intra-operative stress. Through the use of eye tracking it has been shown that the human gaze can be used to assess ergonomics and to obtain useful information to increase the surgeon's usability of robotic surgical tools. Such research can lead to improving visual perception and ergonomics for the surgeon to decrease their intra-operative stress levels, leading to improved situational awareness of the surgeon as well as increased patient safety. In order to understand how eye tracking technology works, this chapter will start with a description of the necessary details of the human visual system and the characteristic eye movements. An outline of eye tracking methods will then be given in Section 2.2 along with an overview of how video feature-based eye trackers work. In addition, a review of the state-of-the-art feature-based eye trackers along with a discussion on the current hurdles to enable widespread use of eye trackers in the surgical theatre are given. In Section 2.3, we discuss the eye tracking applications to date which assists and enhance the surgeon's visual perception and ergonomics in MIS. Section 2.4 covers the discussion and conclusions of this chapter.

2.1 Anatomy of the Human Visual System

Fig. 2.1 shows a horizontal section through a model of the human eye (approximately 23-25mm in diameter). When looking at a person's face, the parts of the eye that are visible are the sclera, which is the white part of the

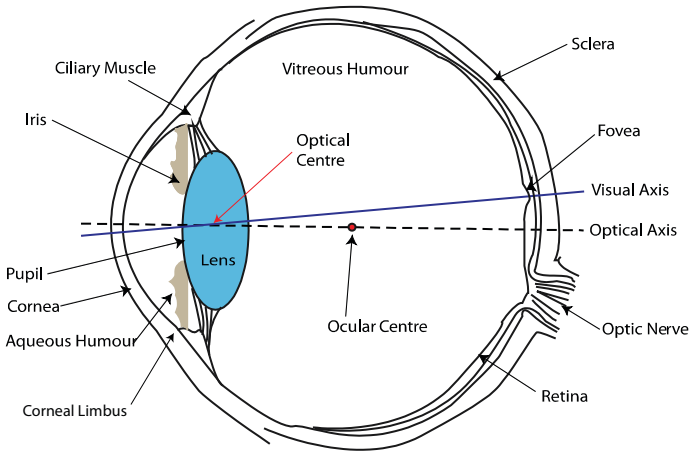


Figure 2.1 – A horizontal cross section of a human eye. Figure adapted from [26].

eye, the iris and the pupil. At the front, the sclera merges into a transparent bulged layer know as the cornea at the corneal limbus. The cornea covers the whole of the iris. The iris adjusts the amount of light entering the eye through the pupil (the hole of the iris), whose aperture size can vary from $2mm$ in bright light to about $8mm$ in diameter in darkness. Immediately behind the iris lies a bi-convex lens, whose shape is changed by the ciliary muscles to focus the incoming light onto the retina located at the rear of the eye. The retina effectively consists of a mosaic of light-sensitive photoreceptors (cones and rods) upon which the image is formed when the eye is in use. The retina contains roughly 7 million cones and 120 million rods.

The resolving power of the retina is the highest in the region of a small area known as the fovea centralis. When the eye fixates onto an object it effectively rotates until the image falls on the fovea. The fovea contains the highest concentration of cones and effectively no rods. Cones are sensitive to brighter, chromatic light. Rods are spread around the peripheral areas of the retina and are sensitive to dim or achromatic light. It is the cones and not rods that make the largest contribution to the fine-grained spatial resolvability of the human visual system (HVS). Nerve impulses from the retina receptors are communicated to the brain through the optic nerve [70].

When an eye is at rest, the ciliary muscles relax causing the lens to flatten, bringing parallel rays to a focus on the retina so that distant objects are

clearly seen. When having to view nearer objects the power of the optical system of the eye is increased, a process known as accommodation. The cililary muscles at such time contracts, causing the front surface of the lens to become round and subsequently allowing the person to focus on the close object.

The visual axis is described by the straight line that passes through both the optical centre (cornea centre) and the centre of the fovea. In contrast, the optical axis is defined as the line connecting the centre of the pupil with the optical centre. The angle between these two axes varies between individuals but is in general in the range of 4-8°. Studies also indicate that this can change even for the same person under different lighting conditions [177]. Of the two axes, the visual axis is more important in eye tracking as this represents the person's PoR.

2.1.1 Foveal and Parafoveal Vision in the Human Visual System

The intricacies involved in the distinction between the foveal and parafoveal vision of the HVS plays an important part in eye tracking. The dimensions of retinal features are conventionally described in terms of projected scene dimensions in units of degrees of visual angle θ, as described by Eq. 2.1.

$$\theta = \arctan\left(\frac{S}{2D}\right), \tag{2.1}$$

where D represents the distance in centimetres from the subject's eye to the gaze tracker screen, which is an output of the gaze tracker system. S represents the distance offset in centimetres from a reference point as illustrated in Fig. 2.2.

In general, the human visual field approximately corresponds to an elliptical visual field with a horizontal major axis of 180° of visual angle and a minor vertical axis of 130°. Within this field spatial acuity of the HVS varies causing the eye to move in various directions, when scanning across a scene. Highest acuity of the HVS is found in the *fovea*, a circular region subtended by 2° of visual angle. The fovea zone is better than the *parafovea* in terms of visual acuity, which extends to 4-5° and drops off sharply beyond.

The rest of the visual field, the *far peripheral* vision, has poor resolvable power and is mostly used for perception of ambient motion [35]. From

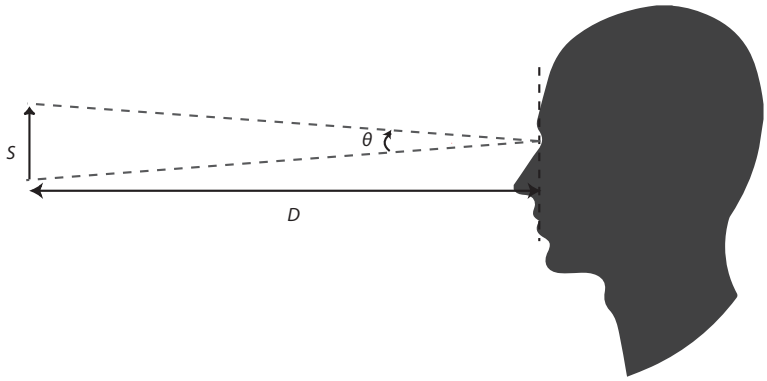


Figure 2.2 – θ represents the angular divergence (°). D represents the distance from the subject's eye to the gaze tracker screen (cm), which is outputted by the gaze tracker system. S represents the distance offset from a reference point (cm).

Fig. 2.3 it can be understood as to why there is higher visual acuity in the fovea over the parafovea and also in the parafovea over the far peripheral regions. There are more cone receptors per mm^2 in the fovea relative to the parafovea and similarly in the parafovea over the far peripheral region. This non-uniform distribution of cone receptors is what necessitates the various movements of the eye when a person conducts a visual search of their environment. The various characteristic eye movements are discussed in the following subsection.

2.1.2 Characteristics of Eye Movements

Given the differences between the foveal and parafoveal vision, it follows that in order to visualise an object clearly it is necessary to move the eye. This necessity gives rise to the different types of eye movements which are discussed here. On average humans use 30° of visual field before moving their head. When trying to observe beyond this visual angle, one will normally move their head [174].

- Saccades: Saccades are rapid eye movements used for repositioning the fovea to a new visual location. Saccades take approximately 30-120ms and traverse a wide visual angle of 1-40°, although typically the range is between 15-20°, giving is a fast acceleration at approximately 40,000°$s^{-2}$and a peak velocity of 600°s^{-1}[174]. Saccades are ballistic,

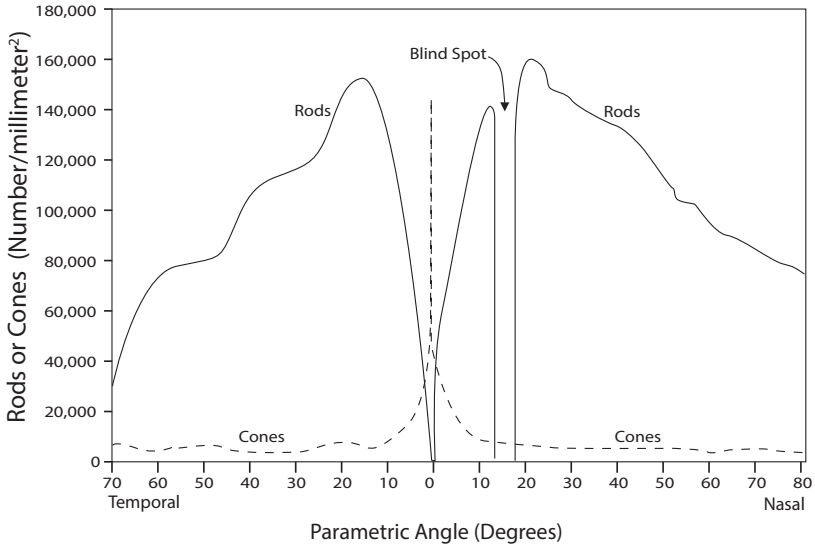


Figure 2.3 – Density distributions of rod and cone receptors across the surface of the retina. Cone density is greatest at the fovea and falls sharply to a minimum by 10° from the eye fixation point. Beyond 10°, the retina contains a thin, uniform distribution of cones. Rods are not present in the fovea. They are most dense in the near periphery out to 18°. While they decline in number further into the periphery, they are still far denser than cones and dominate vision in the peripheral visual field. The blind spot, the region of the retina where the optic nerve leaves the eye, contains no photoreceptors. Graph adapted from [9].

i.e., once initiated their trajectory and destination cannot be altered. Vision during a saccade is suppressed [41] and between the onset of a stimulus that may attract a saccade and until the saccade occurs, there is a 100-300ms delay. In addition, there is a 200ms refractory period after one saccade before it is possible to make another one.

- Miniature: Miniature is a group of involuntary eye movements including tremor, drift and microsaccades which form the characteristics of a fixation. When an individual looks at an object of interest the eye remains relatively stable (typically for longer than 150ms), but there are in fact several types of small, jittery motions occurring, generally less than 1° in size. There is a sequence of a slow drift followed by a sudden microsaccade, a tiny saccade-like jump to correct the effect of the drift. Additionally, a high frequency movement called tremor occurs to help hold a fixed position. The small movements of the eye impose a lower limit of accuracy (approximately 0.5° of visual angle) for eye tracking techniques. Nevertheless, it is these small eye movements that allow humans to continuously see, as research have shown that if an image is artificially stabilised on the retina, vision fades away within about a second and the scene will become blank [121].

- Smooth Pursuits: Pursuit movements occur when the human is visually tracking a moving object. It is much slower than a saccade and is generally synchronised with the velocity of the moving object up to velocities of $15°s^{-1}$. Beyond this speed the smooth pursuit movements are supplemented with saccadic movements, and above $100°s^{-1}$ pursuit is entirely saccadic [80].

- Compensatory: Compensatory movement of the eye occurs to enable fixation on to an object to be maintained during large head movements, by moving the eyes in the opposite direction to the rotation of the head.

- Nystagmus: Nystagmus movements describe the movement of the eye in response to motions of the head when viewing a moving or repetitive pattern. There are two phases, the first being one with smooth motion to follow an object as the head motion causes the eye to move

across the visual field, and the second one being a rapid motion in the opposite direction in order to come back to a more central position.

- Vergence: The movement of both eyes in opposite directions relative to the head to allow for adjustment of the angle between the eyes, ultimately allowing focus on objects at varying distances.

2.2 Overview of Eye Tracking Techniques

Although there are a variety of eye tracking techniques available, the goal of using these devices is the same, *i.e.,* to enable recording of the ocular movement during a visual search task. The eye movements resulting from visual search consist of a fixation scan path where the areas of interest become clear from the subject's eye movement recording. Eye tracking techniques can be divided into *intrusive* and *non-intrusive* methods. Intrusive eye trackers tend to have superior accuracy over non-intrusive methods but due to their invasive nature are unsuitable for long-term use. Intrusive eye trackers include scleral contact lens / search coil and electro-oculography (EOG) eye trackers. The non-intrusive eye trackers often use video recordings of the eye and computer vision techniques to estimate the user's PoR.

Intrusive Eye Trackers One of the most traditional eye tracking methods is the sclera contact lens [126]. A small coil (search coil) is embedded into a contact lens that fits tightly via suction over the sclera. The suction mechanism prevents slip of the contact lens during fast eye rotations. The users gaze is tracked from measuring an induced voltage in the search coil by an external electromagnetic field. Fig. 2.4(a) shows an illustration of a search coil. Although this method has the highest accuracy of around 0.08°, it is also the most intrusive method. Insertion of the lens requires care and practice. Wearing the lens also can cause discomfort.

Another intrusive eye tracking technique is EOG, which relies on the corneoretinal potential that creates an electric field in the front of the head. As the eye rotates, this electric field changes accordingly. To detect the change in movement of the eyes, electrodes are placed on the skin around the eyes. This eye tracking method is considered cost effective and easy to set up, but it is invasive due to the requirement of having to place electrodes

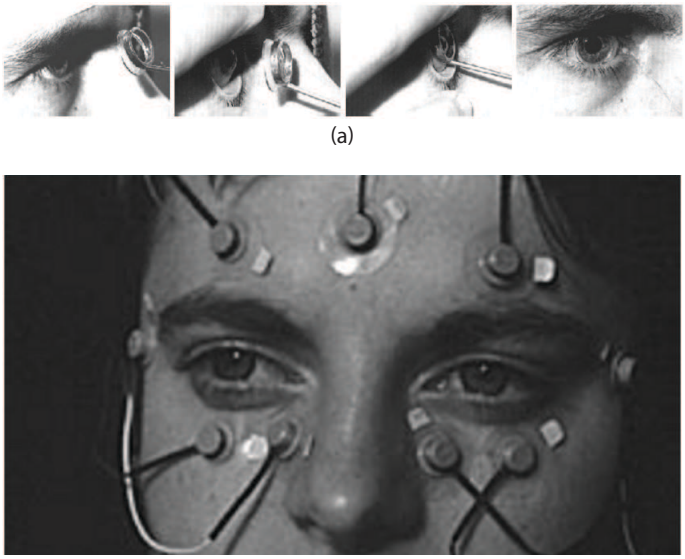


Figure 2.4 – (a) An example of a scleral contact lens with a suction ring mechanism. Image courtesy of Skalar Medical, Delft Holland. The picture shows the invasiveness of this method of eye tracking. (b) An example of an EOG eye tracker. The electrodes are placed directly on the skin surface. It is considered an inexpensive technique, but is intrusive in nature. Image courtesy of Macrovision.

around the eyes (see Fig. 2.4(b)) and the method's accuracy is limited to about 2° of visual angle. Importantly, because both The scleral contact lens and EOG only measures the position of the eye in relation to the head, in order to know the PoR, head tracking needs to be used in conjunction with both of these methods [69].

Non-Intrusive Eye Trackers The most commonly used eye trackers today are video camera based eye trackers. These systems are non-intrusive permitting usage during long sessions. Video based eye trackers can be categorised further into feature-based or appearance-based methods as illustrated in the Fig. 2.5.

Appearance-based eye trackers typically detect and track the eye directly using images of the whole eye, which are based on the photometric appearance as characterised by the colour distribution or filter responses of

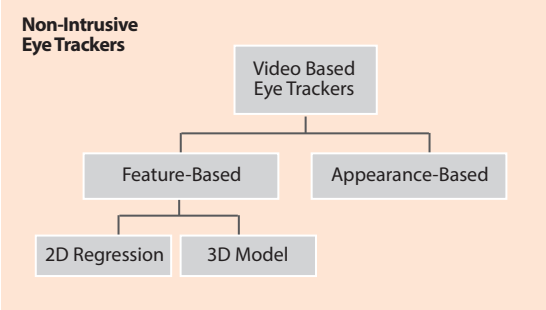


Figure 2.5 – Categorical diagram of non-invasive video based eye trackers.

the eye and its surroundings. Appearance-based methods can be image template-based, where both the spatial and intensity information of each of the image pixels are preserved, or holistic based, where the intensity distribution is characterised by ignoring the spatial information. The research presented in this thesis focuses on using feature based video eye trackers, thus appearance based eye trackers will not be discussed further.

2.2.1 Feature-Based Eye Trackers

Feature-based eye trackers detect and track either natural of artificially induced features. Natural features include the limbus (the boundary of the sclera and iris) or pupil. The centres of these features coincide with the optical axis. Assuming that the angle between the optical and visual axes remains constant, the feature-based eye trackers can provide the necessary information for tracking the user's PoR. Artificially induced features are reflections of light on the cornea (glints or Purkinje images) created by illumination with a light source such as infrared light. Examples of commercially available video feature-based eye trackers are shown in Fig. 2.6. Fig. 2.6(a) is an example of a feature-based video gaze tracker for a laptop / desktop computer use designed by Tobii Technology AB. Fig. 2.6(b) is a head mounted feature-based video eye tracker by SensoMotric Instruments GmbH.

Feature-based eye trackers can further be sub-categorised into 2D regression-based or 3D model-based eye trackers. 2D regression-based (interpolation-based) eye trackers use a mathematical mapping to create a correspondence between eye features to gaze screen pixel coordinates. This mapping is typ-



Figure 2.6 – Feature-based commercial eye trackers. (a) The Tobii PCEye Go eye tracking device for laptops and desktop computers by Tobii Technology AB (b) The SMI Eye Tracking Glasses 2.0, an example of a head mounted eye tracker by SensoMotoric Instruments GmbH.

ically a parametric equation such as a polynomial. In contrast, 3D model-based methods compute the gaze direction from the eye features based on a geometric model of the eye. The PoR is obtained by intersecting the gaze direction with the point being viewed on the screen. Many of today's feature-based eye trackers incorporate components similar to those shown in the block diagram in Fig. 2.7. Features of the eyes are extracted and tracked. An offline calibration procedure is then performed to obtain a correspondence mapping between the eye features and screen coordinates. Typically, this process involves presenting a moving dot on the screen where the user is required to gaze on the dot. Once the calibration procedure is completed, the obtained mapping can be used for gaze estimation. The following subsections will focus on explaining the feature-based eye tracker mechanism flow chart illustrated in Fig. 2.7 in greater detail.

2.2.1.1 Detection of Eye Features

The first step in feature-based eye tracking is the detection and tracking of naturally occurring and artificially generated features of the eye. The natural features include the limbus or the pupil. The artificial features are generated by illuminating the eye with infrared light, inducing corneal reflections, much like the "red-eye" effect captured on camera images when using a bright flash.

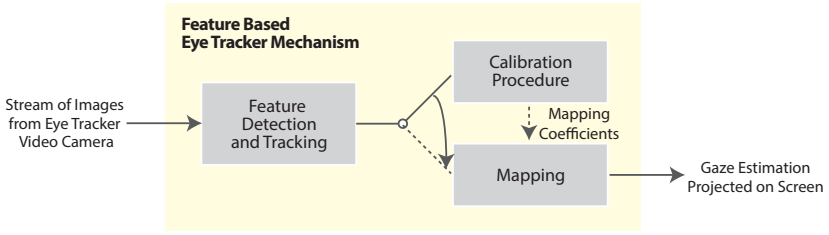


Figure 2.7 – A general flow chart of feature-based eye trackers. Features of the eye are detected and tracked. A calibration procedure is then conducted to obtain a mapping correspondence between extracted features to screen coordinates, where subsequently the user's PoR can be estimated using the mapping obtained from the calibration procedure.

Limbus / Pupil Centre Detection Limbus detection and tracking can be challenging as partial occlusion of the limbus by the eyelids or eyelashes often results in poor vertical tracking. In order to overcome the partial occlusion of limbus tracking Zhu *et al.* [185] created a subpixel iris tracker using the edges from the limbus, and by applying an ellipse-specific direct least-square fitting method by Pilu *et al.* [118]. However, it is far more common to see the pupil centre being used for eye tracking as the infrared light used to generate the artificial corneal reflections also increases the contrast between the iris and pupil therefore allowing for easier pupil detection. Depending on whether the camera optical axis is coaxial to the infrared illuminator or not, the pupil will result in either a "bright pupil" or a "dark pupil" as illustrated in Fig. 2.8(a)-(b) respectively. Through applying appropriate thresholds, the pupil centre can be segmented accurately in both cases. Ebisawa *et al.* [42] developed a pupil extraction technique by flashing two light sources, one in line with the camera axis and one off-axis alternatively effectively generating both bright and dark pupil images respectively. Assuming that the eye does not move much during the consecutive frames, through capturing these features with a video camera and taking a thresholded difference of the two consecutive frames, it is possible to extract a clearer pupil to increase tracking accuracy of the eye as illustrated in Fig. 2.8(c) This method, however, may not work under fast eye rotations *i.e.*, when there are large changes in the eye position between consecutive frames.

To increase robustness of pupil contour detection against outliers such as eyelashes and corneal reflections, Ohno *et al.* [114] suggested a dual ellipse

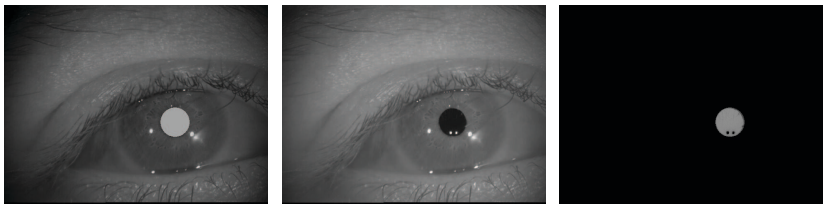


Figure 2.8 – (a) Bright pupil is observed when the incident light ray is on-axis. (b) Dark pupil is observed when the incident ray is off-axis. (c) Thresholded difference of the two images gives a clear segmentation of the pupil.

fitting mechanism. Initially, the pupil edges are used for ellipse fitting, but in a second fitting, only the pupil edges that are close to the computed ellipse edges are used. Overall, outliers are filtered out of the true pupil contour and thus increase the accuracy of the pupil center position estimation.

Corneal Reflections (Glints) The artificially induced corneal reflections, also known as Purkinje images (see Fig. 2.9), are distinctive spots created on the eye's cornea surface when infrared light is used to illuminate the eye. Typically, four reflections are captured including; P1, reflection from the outer surface of the cornea also commonly referred to as the "glint"; P2, reflection from the inner surface of the cornea; P3, reflection from the outer (anterior) surface of the lens; P4, reflection from the inner (posterior) surface of the lens. Today, tracking of the pupil and corneal reflections has become the most common approach in eye tracking. Tracking of these two types of ocular features enables the user's PoR estimation as the corneal reflections remain relatively stable under rotation of the eyeball. The eye corner has been suggested to be used instead of corneal reflections, since there would be an added benefit of not having to use infrared lights, however Sesma *et al.* [137] found that the corner of the eye moved with rotation of the eyeball introducing inaccuracies in PoR estimation.

For 2D regression-based eye trackers, the distance vectors between the pupil centre and corneal reflections are used to estimate the user's PoR. For 3D model-based eye trackers, the pupil centre, corneal reflections, the geometric location of infrared light source(s) and camera location(s) are used to calculate the cornea centre and subsequently, the optical axis vector

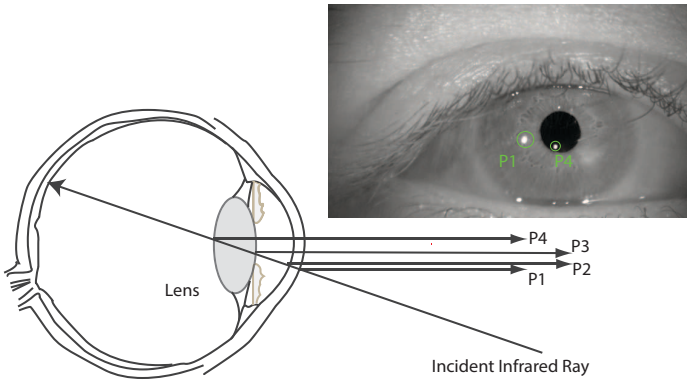


Figure 2.9 – Purkinje reflections are distinctive bright spots captured on video when a fixed infrared light source is directed on the eye. Four reflections are captured including; P1, reflection from the outer surface of the cornea also commonly referred to as the "glint"; P2, reflection from the inner surface of the cornea; P3, reflection from the outer (anterior) surface of the lens; P4, reflection from the inner (posterior) surface of the lens.

between the cornea centre and pupil centre to estimate the user's PoR. In 2D regression-based eye trackers, tracking the corneal reflections enables differentiating small head movements from eye rotation, which allows for measurement of the user's PoR under some head movement. Effectively, the artificially generated corneal reflections can be used as points of reference to distinguish how the eye is rotating. Tracking of both these features enables estimation and tracking of the user's optical axis. Thus, if the difference in the optical and visual axes is accounted for via a calibration procedure, it can enable estimation of the user's PoR.

2.2.1.2 Calibration Procedure

In order to establish correspondence between the ocular features in the eye plane and stimulus being viewed by the user in the screen plane, a subject-specific calibration procedure is required to enable estimation of the user's PoR. For both 2D regression-based eye trackers and 3D model-based eye trackers, this is done through calculation of a mapping function as illustrated in Fig. 2.10. Once the mapping function is obtained, it can be used to map the ocular feature coordinates from the eye images to the corresponding gaze pixel coordinates on a computer screen through interpolation. The

mapping function accounts for subject-specific ocular variations which may include:

- pupil size
- corneal radius and curvature
- cornea refractive index
- angular offset between the optical and visual axes
- interpupillary distance
- contact lenses or eyeglasses
- long eyelashes or heavy eye make-up

Without performing a subject-specific calibration procedure, the above factors can reduce the accuracy of the estimated user's PoR. Depending on the subject's ethnicity, sex or age, the above aspects can have a greater influence in the variation of the PoR accuracy.

The calibration procedure involves displaying a dot at a known set of coordinate points, typically nine points, one by one on the screen and the user is then required to fixate on them. During this procedure, the different ocular feature vectors V, are recorded with their corresponding fixated known screen coordinates S, where,

$$V = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{bmatrix} \text{ and } S = \begin{bmatrix} \boldsymbol{s^x} \\ \boldsymbol{s^y} \end{bmatrix}^T = \begin{bmatrix} s_1^x & s_2^x & \dots & s_n^x \\ s_1^y & s_2^y & \dots & s_n^y \end{bmatrix}^T.$$

Here n is the number of points displayed on the screen which in this example, $n = 9$.

In order to establish correspondence between the ocular feature vectors and the respective fixated screen coordinates, a set of mapping coefficients C needs to be solved for in order to provide this correspondence relationship *i.e.,* the following linear system needs to solved,

$$S = W^T C. \tag{2.2}$$

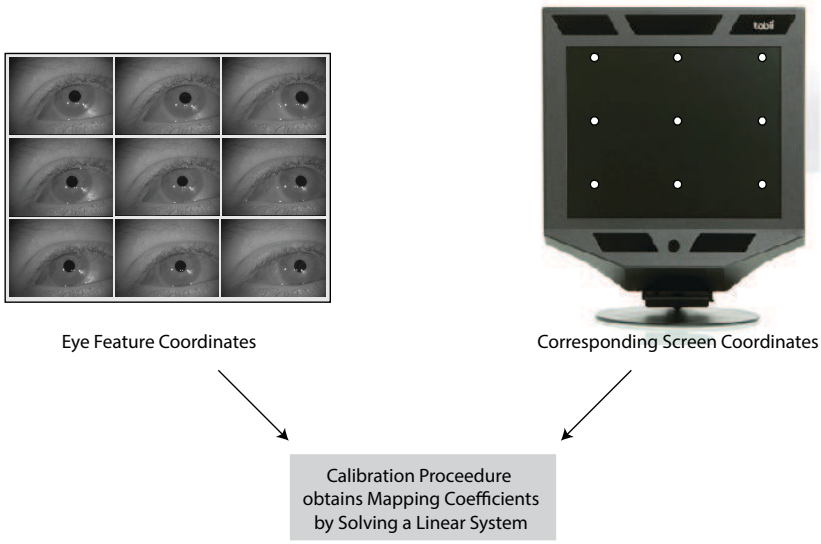


Figure 2.10 – Calibration procedure for an eye tracker involves displaying a dot at nine regularly spaced coordinates, one at a time on the screen, where the user is required to fixate on them. During this procedure, the distinctive eye feature coordinates (pupil and glint locations) for each nine points are recorded with the corresponding screen coordinates. A polynomial function can be used to establish the correspondence relationship between the eye image and screen plane.

Here W uses the ocular feature vector of V recorded during the calibration procedure to give,

$$W = \begin{bmatrix} 1 & x_1 & y_1 & x_1y_1 & x_1^2 & y_1^2 \\ 1 & x_2 & y_2 & x_2y_2 & x_2^2 & y_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_9 & y_9 & x_9y_9 & x_9^2 & y_9^2 \end{bmatrix}, \tag{2.3}$$

and the mapping coefficients C , is of the form,

$$C = \begin{bmatrix} C^x \\ C^y \end{bmatrix}^T = \begin{bmatrix} a_1 & a_2 & \dots & a_6 \\ b_1 & b_2 & \dots & b_6 \end{bmatrix}^T. \tag{2.4}$$

Using a linear least square method, we obtain,

$$C = \left(W^T W\right)^{-1} W^T S. \tag{2.5}$$

The obtained mapping coefficients C, are then used to form two parametric functions for x and y to create a $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ mapping [93] as follows,

$$E = \begin{bmatrix} E^x \\ E^y \end{bmatrix}^T = \begin{bmatrix} a_1 + a_2x + a_3y + a_4xy + a_5x^2 + a_6y^2 \\ b_1 + b_2x + b_3y + b_4xy + b_5x^2 + b_6y^2 \end{bmatrix}^T. \tag{2.6}$$

Eq. 2.6 enables estimation of the user's PoR E, when new incoming eye feature coordinates (x, y) are placed into this parametric function. In the following subsections, an overview of the state of the art 2D regression-based and 3D model-based eye trackers will be discussed.

2.2.1.3 2D Regression-Based Eye Trackers

Many 2D regression-based eye trackers use a single camera and one or more infrared light sources to track the "*Pupil Centre minus Corneal Reflection*" (PCCR) distance vectors of the user's eye [94, 89, 168]. Although these methods have some tolerance to head movements, they fail under larger head movements [93]. For techniques that rely on detecting and tracking only one natural feature such as the pupil or limbus, the user's head would have to be kept in a fixed position or a head mounted optical sensor would have to be worn to keep track of the head position for PoR estimation. Once

the coordinates of the pupil centre and corneal reflections are known, the PCCR distance vectors are then mapped to the computer screen through a vector transformation obtained via a subject-specific calibration procedure as previously discussed.

Merchant *et al.* [89] were one of the first to propose a real time video based eye tracker. They used a linear mapping to estimate the user's gaze where they noticed greater curvature with larger PCCR vectors. They compensated for these using a polynomial regression. Morimoto *et al.* [94] used a higher order polynomial function to map the user's PCCR vectors to the screen coordinates. Through using a second order parametric function, they report an accuracy of around 1°.

In order to account for larger head movements, White *et al.* [168] estimated four calibration mappings for four different head locations from a single static calibration using spatial symmetries. Head positions were located by using a second light source. By using two light sources as points of reference, the single static calibration was used to adjust for head movement. Sugioka *et al.* [150] used an ultrasonic distance meter to adapt for larger head movements. Distance information from the ultrasonic distance meter was used to drive a pan-tilt-zoom camera for tracking the gaze direction. They report an accuracy of 0.7°, but the tracking refresh rate is limited by the mechanical speed of the pan-tilt-zoom camera. Zhu *et al.* [186] introduced a 2D regression based method using two cameras to estimate the 3D head position. They use the 3D eye position to modify a regression function to compensate for head movements. Their system, however, requires a stereo camera calibration procedure and a long training process to form their calibration mapping. They report an accuracy of 1.5°. Cho *et al.* [22] used two wide-view cameras for stereo and a narrow-view camera mounted on a pan and tilt unit in a 2D regression-based calibration method to achieve a larger workspace of $100cm \times 40cm \times 100cm$. With a 7-point calibration method, they achieve an accuracy of 0.8°.

2.2.1.4 3D Model-Based (Geometrical) Eye Trackers

The general approach for model-based eye trackers is to estimate the 3D centre position of the cornea by assuming a spherical eyeball and cornea surface. After estimating the 3D location of the centre of the pupil, the

optical axis vector can be constructed by connecting the 3D location of the cornea centre with the pupil, as illustrated in Fig. 2.11. Both the optical and visual axes intersect at the cornea centre, thus making the cornea centre an important parameter to estimate in a 3D model-based eye tracker. In order to account for the angular offset between the optical and visual axes a calibration procedure is carried out. This can be a regression-based calibration where the point of intersection of the optical axis with the screen is mapped to the true gaze screen coordinates represented by the visual axis.

In reality, the spherical models of the cornea are not suitable for modelling the boundary area of the cornea and often lead to greater inaccuracies when the user moves their eye to the edges and or corners of the monitor (*e.g.,* glints move onto a non-spherical surface) [114, 159, 140, 50]. This is perhaps the main shortcoming of 3D model-based eye trackers; the use of a strong prior model of the eye leads to large performance variation between different subjects. The strong model assumptions can result in gaze estimation inaccuracies when estimating the user's PoR near the edges or corners of the display. In this sense, one of the important parameters to these models is the knowledge of the subject-specific cornea curvature. If the user's cornea curvature is known, it is possible to find its centre using only one camera and two light sources. Otherwise, estimation of the corneal centre requires at least two cameras and two light sources [140].

One of the advantages of 3D model-based eye trackers is the fact that they can use the 3D location estimation of the cornea centre as an indicator for the head location in 3D space to create head pose invariant models. The downside for many of the 3D model-based eye trackers [114, 159, 140, 50, 92, 55, 8, 142, 113, 102, 103, 45, 90] is that they rely on having accurate metric information and thus require a camera calibration and a global geometric model of light sources, cameras and monitor position. Thus, they are consequently inflexible when attempting to change the geometry of light sources, camera (*e.g.,* zoom) and screen position to particular needs. 3D model-based eye trackers that do not require camera calibration are the projective invariant cross-ratio (CR) method by Yoo *et al.* [180, 178, 179] and the different variants of the CR methods [29, 68, 53, 30, 31] and more recently, the work by Nakazawa *et al.* [104] which uses structured light reflected off the corneal surface to estimate the user's gaze under free head motion. The following part sub-categorises the 3D geometrical model-based

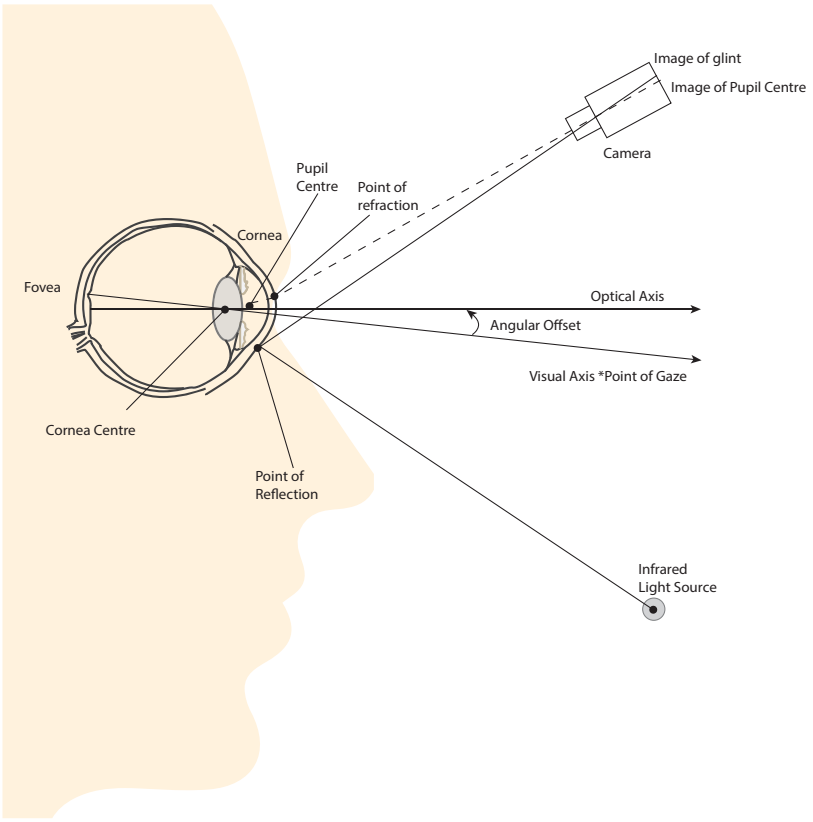


Figure 2.11 – The general model of the eye used in 3D model-based eye trackers. The cornea is modelled as a perfectly spherical surface. By positioning the camera and light source in known geometrical locations, estimation of the 3D location of the cornea centre and pupil centre can be obtained. Through connecting the cornea centre with the pupil centre, the optical axis direction is obtained. A calibration procedure is used to obtain the angular offset between the optical and visual axes, allowing for estimation of the user's PoR. A regression-based calibration maps the point of intersection of the optical axis with the screen to the true gaze screen coordinates.

eye tracking systems according to the number of cameras or light sources that each respective system uses.

- ***Single Camera and or Single Light Source:*** Ohno *et al.* [114] used a constant anthropometric average for the cornea curvature and refraction index to estimate the optical axis direction. Under a fixed head position, they use a two-point calibration method to calculate the angular offset between the optical and visual axes. They report an accuracy of approximately 1°. Villanueva *et al.* [159] used a single camera only to deduce the user's gaze direction and reported a similar accuracy. Shih *et al.* [140] proved that the use of a single light source and single camera cannot provide free head movement during gaze estimation. One downside with using only one camera is that subject specific eye parameters, such as the cornea curvature, cannot be estimated. Therefore, anthropometric averages need to be used in the models leading to larger performance variation between different subjects when compared to stereo camera arrangements.

- ***Single Camera and Multiple Light Sources:*** Morimoto *et al.* [92] used a single camera and two infrared light sources. Their method also requires a user specific calibration procedure to increase accuracy. Using simulations they report an accuracy of about 3°. A similar approach was taken by Hennessey *et al.* [55], where the authors use a single camera and two light sources and a four point calibration procedure to estimate the user's gaze under free head motion achieving an accuracy of 1°. Using a nine-point calibration procedure and one camera and two light sources, Guestrin *et al.* [50] reported a similar accuracy. One of the downsides of systems employing multiple light sources, such as the systems discussed here, is that they are faced with the increased chance of one or more of the glints disappearing, thus introducing errors into the gaze estimation.

- ***Multiple Cameras and Multiple Light Sources:*** To allow for free head motion, a larger field of view is needed. In contrast, in order to capture high-resolution eye images to provide accurate gaze estimates, a narrow field of view is desirable. Hence, single camera systems are faced with the trade-off problem of head motion workspace

against high-resolution eye images. Use of multiple cameras offers extended flexibility in this sense, albeit with an increased geometrical complexity and a camera calibration procedure required in the setup.

To enable a larger head motion workspace, Beymer *et al.* [8] used a pair of stereo cameras; one wide angle stereo camera is used to detect the face, from which information is used to steer the second narrow field of view stereo camera. The narrow field of view stereo pair is able to take large pupil images which are then used to fit projected model features from their 3D eye model for feature detection. This method estimates intrinsic parameters of the eye *i.e.,* radius of cornea, angular offset of the line of sight, through a calibration procedure. They reached an accuracy of 0.6°. Ohno *et al.* [113] utilised three cameras, a stereo pair wide-angle camera and a narrow angled camera mounted on a pan-tilt unit to enable eye tracking under free head movement. Their system uses a two-point calibration procedure to adjust for the angular offset between the optical and visual axes. They report an accuracy of approximately 1°, but systems using pan-tilt mechanical cameras are limited by the speed of the rotation mechanism.

Shih *et al.* [142] used two cameras and three light sources. With calibration they accomplished an accuracy of better than 1° but the head motion workspace is likely to be less than the techniques in [8, 113]. Nagamatsu *et al.* [102] introduced a one-point calibration gaze tracking method using stereo cameras. The one point calibration is required to correct for the offset between the visual axis and optical axis. They report an average system accuracy of around 1° with lower accuracy at the upper regions of the display.

In order to improve on their own work in [102], Nagamatsu *et al.* [101] modelled the eye with an aspherical shape through using a surface of revolution about the optical axis of the eye. They claim to have improved the accuracy but fail to report numerical figures. Eizenman *et al.* [45] used a stereo camera and multiple light sources and estimate the user's point of regard without the need for a calibration procedure. Through minimising the distance between the intersections of the visual axis of the left and right eyes with the surface of the display they are able to estimate the angular offset between the optical and

visual axes. They reported an accuracy of 1.3° based on simulations. More recently, Model *et al.* [90] built on [102] and demonstrated a similar accuracy through experiments instead of simulations.

Yoo *et al.* [180] devised the projective invariant CR method that closely estimates the optical axis using two cameras and five light sources. Their system uses a wide angle camera to direct a narrow angle camera mounted on a pan-tilt mechanism. Four light sources are placed around the corners of the computer screen to project these corners on the corneal surface. Through assuming a flat cornea and by considering the invariance property of CRs under perspective projection, they estimate the user's gaze. Overall, their method requires a four-point calibration procedure and they report an accuracy of around 1°. Later on, Yoo *et al.* [179] improved their CR method for improved robustness to head motion by introducing a scaling factor to compensate for the error introduced by assuming a flat cornea. They use the same hardware setup as their previous work and achieve an accuracy of 1° with a four-point calibration procedure. Coutinho *et al.* [29] reduced the number of parameters of Yoo's model and introduces a displacement vector correction for the optical and visual axes. Their method works best with a nine-point calibration procedure where they report an accuracy of 1°. Kang *et al.* [68] introduced a homography-based correction on top of the CR method by Coutinho *et al.* [29] and showed that via a four-point calibration, they were able to achieve an accuracy better than 1°. Their system consists of four infrared lights and two cameras (one wide angle and one narrow angle). Hansen *et al.* [53] also used a homography-based glint-normalisation on top of the CR method. Through using a Gaussian process interpolation they account for the CR method's model errors. They recreate the CR method by Yoo *et al.* [179] and [29] and demonstrate that their system has superior robustness to head movements. They use a four-point calibration procedure and report an accuracy of less than 1°. Coutinho *et al.* [30] improved on their work in [29] to increase the work space by dynamically adjusting the displacement vector by a scale factor to compensate for head motion. With a nine-point calibration, they report an 8% accuracy improvement in PoR estimation over a larger

head motion workspace. More recently, Coutinho *et al.* [31] improved on the CR method's errors by planarising the CR features, thus taking into account the curved cornea features. They recreate previous CR techniques [29, 30, 68, 53] and demonstrate that their new setup performs better under a larger head motion workspace. Unfortunately, they fail to disclose how many calibration points they used in their system. More recently, Nakazawa *et al.* [104] used an array of LED lights to project structured light reflected off the corneal surface and a geometric eye model to estimate the user's gaze under free head motion. They report an accuracy of approximately 1°.

2.2.2 Calibration Drift and Practical Issues

An additional challenge of an eye tracker is the "*Calibration Drift*", which can arise when the previously obtained subject-specific calibration needs to be updated after a given period. This drift may be induced by either internal or external factors. Internal factors may be a change in the individual's medical conditions or tiredness. External causes may include the room lighting conditions or the subject moving away from the original calibration position [19, 93, 29]. The internal and external factors can alter the ocular appearance of the same user, for example the pupil size [177]. Dry eyes can also alter the cornea surface causing the corneal reflections to be less visible. The drift can only be re-adjusted through re-calibrating the eye tracker. It has therefore been the case that a calibration procedure is performed before every session to account for any subject-specific ocular variations and calibration drift. The effects of the drift over time on the PoR estimation accuracy and precision have been documented [59, 112], which will require a recalibration procedure to adjust the accuracy and precision.

2.2.3 Calibration Free Eye Tracking

The research community's quest for developing a calibration-free, head movement robust eye tracker has so far been a trade-off between accuracy, head movement workspace, camera calibration, number of calibration points, eye tracking cameras and light sources. This can be understood from the summary shown in Table. 2.1. All feature based eye trackers (2D regression and 3D model-based) today still require an explicit offline calibration proced-

ure, especially if good eye tracking accuracy needs to be obtained. From a clinical perspective, the procedure can be repetitive and cumbersome as it requires explicit participation of the surgeon and can disturb the overall surgical workflow. Although there have been 3D model-based eye tracking methods that require as little as one-calibration point [102, 90] recent research highlights that the gaze estimation accuracy of these methods are still head position-dependent [17] and are still prone to calibration drift. Other calibration free methods have been suggested, which utilise the visual saliency information from images or videos to avoid active human participation of the user [149, 18, 5]. These techniques use the saliency information of the visual stimulus to estimate the likely regions which the human gaze would be attracted to. Through using machine learning techniques a mapping function or model parameters are incrementally estimated. Unfortunately, these methods also lack the necessary accuracy reporting 3° or worse.

For eye trackers to become readily usable and common place in the surgical theatre, there is a need to develop an eye tracking system that is user-friendly, *i.e.,* requires little calibration, robust in performance, accurate, and one which will not have to be recalibrated when used over long periods of time. To date, there is no eye tracking system that meets all these requirements. As part of this thesis, we introduce an implicit online calibration algorithm in Chapter. 7 which updates the calibration procedure in an ongoing manner to overcome the calibration drift. The implicit nature of the online calibration algorithm has the added benefit of seamlessly integrating into the surgical workflow, therefore, forgoing the need for the user to actively engage in an offline calibration procedure. The last subsection provides an overview of general and robotic MIS eye tracking applications.

2.3 Surgical Eye Tracking Applications

Eye tracking has been used extensively from behaviour analysis psychology [15, 34, 47] and cognitive science [74, 172] to usability and ergonomics assessment in aviation [41] and car driving simulators [136]. The widespread use of computers and the internet has also driven eye tracking to be applied for studies in web advertising and page layout [39, 116, 106]. In addition, eye tracking has also gained wide use in human-computer interaction systems for example as a pointing modality to replace the mouse [64] or for "*eye typ-*

Table 2.1 – Summary of 2D regression-based and 3D model-based feature eye trackers with their respective key characteristics.

Author	Type	Cameras	Lights	Free Head Pose	Camera Calibration	Personal Calibration	Accuracy (Degrees)	Comments
Morimoto *et al.* [94]	2D	1	2	No	No	Yes	1	Pan-tilt mechanism
White *et al.* [168]	2D	1	2	Yes	No	Yes	1.5	Ultrasonic distance meter
Sugioka *et al.* [150]	2D	1	1	Yes	No	Yes	0.7	Long training process
Zhu *et al.* [186]	2D	2	1	Yes	Yes	Yes	1.5	
Cho *et al.* [22]	2D	2+1	2	Yes	Yes	Yes	1	Pan-tilt mechanism
Ohno *et al.* [114]	3D	1	1	No	Yes	Yes (2-points)	>1	
Villanueva *et al.* [159]	3D	1	0	No	Yes	Yes (2-points)	1	
Morimoto *et al.* [92]	3D	1	2	Yes	Yes	Yes	3	Simulations Only
Hennessey *et al.* [55]	3D	1	2	Yes	Yes	Yes (4-points)	1	
Guestrin *et al.* [50]	3D	1	2	Yes	Yes	Yes	1	
Beymer *et al.* [8]	3D	2+2	2	Yes	Yes	Yes	0.6	3D head position + Stereo camera (Narrow FOV)
Ohno *et al.* [113]	3D	2+1	1	Yes	Yes	Yes (2-points)	1	3D head position + Stereo camera (Narrow FOV)
Shih *et al.* [142]	3D	2	2	Yes	Yes	Yes	<1	
Nagamatsu *et al.* [102]	3D	2	2	Yes	Yes	Yes (1-point)	>1	Greater error near the top of the screen
Eizenman *et al.* [45]	3D	2	2	Yes	Yes	No	1.3	Simulations only
Yoo *et al.* [180]	3D	1+1	5	Yes	No	Yes (4-points)	<2.5	
Yoo *et al.* [179]	3D	1+1	5	Yes	No	Yes (4-points)	1	
Coutinho *et al.* [29]	3D	1+1	5	Yes	No	Yes	1	Improved free head space vs. [179]
Kang *et al.* [68]	3D	1+1	4	Yes	No	Yes (4-points)	<1	Improved free head space vs. [29]
Hansen *et al.* [53]	3D	1	4	Yes	No	Yes (9-points)	1	Improved free head space vs. [29]
Coutinho *et al.* [30]	3D	1+1	5	Yes	No	Yes (9-points)	1	Improved free head space vs. [29]
Coutinho *et al.* [31]	3D	1+1	5	Yes	No	Yes	<1	Improved free head space vs. [30]
Nakagawa *et al.* [104]	3D	1+1	42+2	Yes	No	Yes	1	LED Array Projection

ing" [87]. There has also been interest in utilising eye tracking technology by the computer graphics community in gaze-contingent displays where the resolution and quality of an image or animation is adjusted according to the viewer's eye movements on single user displays to economise on bandwidth [85]. Since the human visual system can only capture high resolution images near the fovea boundary, the idea is to save resources that are wasted in producing high resolution images on the screen outside this visual boundary.

In medicine, eye tracking has been used for medical image understanding and analysis for training purposes. The knowledge of visual search patterns of experienced observers has been recognised to be useful for improving the learning process of trainees, since the search patterns provide an effective means to learn how experts assimilate different visual clues to reach a confident diagnosis [32, 107, 120]. Eye tracking technology offers the potential to overcome key problems in medical image analysis, including errors caused by overlooking important diagnostic features during observation of images and the variation of interpretation between different observers [73, 108, 127]. Eye tracking has also been widely used to improve the visual perception and ergonomics during a MIS procedure. Enhanced visual perception and enhanced ergonomics are important for the surgeons to supplement their capabilities and to reduce the burden experienced while performing surgery. The last part of this chapter provides an overview of the legacy eye tracking surgical applications, as it provides an important aspect in understanding the research conducted in this thesis.

2.3.1 Visual Perception and Disorientation

One of the fundamental challenges in visual perception in MIS, which is more apparent in natural orifice transluminal endoscopic surgery (NOTES), is the spatial disorientation. The Oxford dictionary defines spatial disorientation as, "*The loss of one's sense of direction, position, or relationship with one's surroundings.*" Spatial disorientation can be caused by a number of factors including off-axis instrument manipulation during retroflexion of the flexible endoscope, missing fixed anatomical references and a limited field-of-view during navigation [123]. Disorientation is also a common occurrence in aviation and there has been a study involving eye tracking to characterise eye movement behaviour during disorientation [21]. The authors report that

there is an increase in pupil dilation as well as evidence of nystagmus indicated by the number of involuntary saccades during disorientation. They also noted flight performance degradation during disorientation reflected by a significant deviation in maintaining airspeed.

In surgery, the behavioural characteristics of eye movements during spatial disorientation are still unknown. Sodergren *et al.* [144] conducted a study to understand whether there are common patterns of visual attention strategies employed by surgeons that are associated with a greater chance to successfully re-orientate under spatial disorientation during laparoscopic cholecystectomy. They presented a number of still images to a group of surgeons and asked them to interpret the orientation of the presented images. Through recording the gaze of the subjects, they found that there were quantifiable visual attention strategies employed by surgeons during laparoscopic cholecystectomy associated with successful re-orientation. They report that a good re-orientation strategy relies on identifying and focusing on a central object within the scene and judging position of its surrounding peripheral objects suggesting integration of both geometric and feature information in a systematic way. In contrast, the use of inconsistent feature cues for re-orientation was associated with less effective performance.

2.3.2 Gaze Contingent Perceptual Docking

The concept of gaze driven "*Gaze Contingent*" tools within the perceptual docking paradigm by Yang *et al.* [176] was introduced to enhance the surgeon's visual perception, instrument, and camera control modality where the robotic systems are able to acquire operator specific motor and perceptual behaviour during a surgical procedure through their gaze behaviour. Gaze contingent motion stabilisation of tissue during surgery by Mylonas *et al.* [99, 97] allow the surgeons to visualise moving organs such as the heart under a static frame of reference, enabling them to improve decision making when performing surgery on moving tissue. The surgeon's 3D fixation points are used for dynamic motion compensation targeted to the fixation point. Visentini-Scarzanella *et al.* [160] further proposed a technique for intra-operative 3D reconstruction and tissue motion compensation through using gaze information. Furthermore, gaze information can be used for path planning by utilising the gaze-focus to create a suitable ablation path for

focused energy delivery in mitral valve surgery. Stoyanov *et al.* [147] introduced a technique to use the 3D gaze fixation path of the operator to plan an accurate ablation path during surgery. Noonan *et al.* [109], used 3D fixation information from the surgeon's eye gaze to control an articulated robotic device. Their experimental results show potential value of their system, for a gaze contingent articulated robot to provide suction to remove blood and other fluids during surgery. More recently, Noonan *et al.* [110] introduced a gaze contingent controlled laparoscope that provides the operator with sufficient distal dexterity to achieve stable off-axis visualisation in an intuitive, hands-free way that will ultimately allow the operator to simultaneously control other instruments.

Eye tracking has also been used as a control modality for microbots [76]. A gaze-led controlling method of three microbots was compared against controlling the same microbots with a button pad. Superior task completion time with the gaze contingent selection method was achieved due to the improved ergonomics as the operator was able to maintain his hands on the instrument control device when using the gaze contingent method. Clancy *et al.* [24] introduced a gaze contingent autofocus camera system using an eye tracker and a liquid lens. Their system adjusts the camera focus according to the user's gaze position and provides an intuitive interface that removes the need for an external pedal for camera focus control. Kwok *et al.* [75] introduced the concept of Collaborative Gaze Channelling (CGC) where the gaze information is utilised to improve the effectiveness of collaborative surgery with multiple surgeons. Through relaying the gaze information, each operator's gaze is shared within the same surgical workspace allowing the surgeons to understand each other's intention. They demonstrated that the visual gaze can be an effective way of communicating during collaborative tasks for robotic surgery.

Safety boundaries in the form of virtual fixtures (VFs) introduced by Rosenberg [128] are used for path guidance of a surgical instrument and obstacle avoidance to prevent the instrument from inadvertently damaging sensitive tissue. This is particularly useful for robotic systems as the lack of haptic feedback to the surgeon's hands can result in accidental damage of sensitive tissue. Mylonas *et al.* [100] devised a method to dynamically introduce virtual fixture safety boundaries through the surgeon's 3D PoR. Initially, the eye gaze is used to track the moving tissue. Safety boundaries

are then dynamically prescribed based on the surgeon's fixation point intra-operatively. Since the fixation point conforms to the moving tissue, the safety boundaries that are centred on the fixation will also be adjusted, thus removing the need for prior knowledge of the soft-tissue morphology. Such a technique has the potential to enhance the surgeon's visuo-motor control capability whilst reducing the cognitive burden during an operation.

2.4 Conclusions

In this chapter, important background knowledge to understand eye tracking technology was described, including details on the anatomy of the eye, the human visual system, and the characteristics of eye movements. The dynamic nature of eye movements causes many challenges for robust tracking of the human eye and one of the fundamental technical limitations of present day eye tracking technology lies in the assumption that the gaze consists solely of the foveal vision *i.e.,* parafoveal vision plays an insignificant part. This is unfortunately not true and research has shown that humans are able to direct attention to objects outside of their foveal vision without moving their eyes.

An overview of eye tracking technology was provided with a particular focus on non-invasive feature-based eye trackers. Details of the working mechanism of video-based eye trackers were discussed including how features of the eyes are extracted and then mapped onto the computer screen plane to understand the user's PoR. The technical challenge in eye feature extraction and tracking lies in the robustness of the performance during fast eye rotations, blinking, head motion and during changes of external illumination conditions. Details of the calibration procedure for feature-based eye trackers were also discussed. Subsequently, the current state-of-the-art 2D regression-based and 3D model-based eye tracking techniques were reviewed along with their respective advantages and disadvantages focusing on the issue of calibration and gaze estimation accuracy. The accuracy of a calibration-free eye tracking technique will depend on head motion range, camera calibration, number of calibration points, number of cameras and light sources. Furthermore, the calibration can deteriorate and drift with time. To date, there is still no eye tracking method that does not require an explicit calibration procedure.

This chapter ends with a review of various eye tracking surgical applications for enhancing the surgeon's visual perception and ergonomics whilst performing a MIS procedure. Through the use of the gaze information, in the gaze contingent perceptual docking paradigm, there have been a number of applications that show potential to increase the surgeon's efficiency and enhance their capability which in turn can lead to improved patient safety and surgical outcomes. The remainder of this thesis focuses on how eye tracking may be applied in the surgical theatre to understand the surgeon's perception and improve the ergonomics during MIS. Over the next chapters, this thesis aims to provide the reader a better understanding of how the surgeon's gaze information may be used to further enhance the MIS experience and improve performance, workflow and overall ergonomics, which could ultimately lead to better surgical outcomes.

3 Eye Tracking Applied to Endoluminal and Transluminal Surgery

As discussed in Chapter 2, there has been interest in applying eye tracking in surgical environments to provide enhanced visualisation techniques as well as improved ergonomics. In order to develop clinically useful applications, it is of utmost importance to understand the challenges faced by clinicians in the operating theatre. During any surgical procedure, considerable time is spent on navigating to the operative site. Once arriving at the target anatomy, effort is spent on maintaining an optimal FOV for the operation to be performed. Therefore, methods that provide good camera visualisation during navigation and during the surgical operation are important aspects for the surgeon. Since this thesis aims to investigate the challenges in MIS, this chapter focuses on understanding the challenges in one of the most common diagnostic and therapeutic procedures performed today in hospitals, the endoscopic procedure [125]. The chapter begins with a description of the endoscopic procedure of both upper and lower parts of the gastrointestinal tract, referencing the key anatomical structures. In Section 3.2, the challenges during endoscopic procedures discussed in previous literature are reviewed to give the reader an in depth understanding of the difficulties commonly encountered when performing endoscopy. Section 3.3 presents a number of different techniques that have been previously implemented and applied to alleviate some of these problems. Subsequently, Section 3.4 describes an experiment based case study to help gain qualitative understanding of the important characteristics of gaze behaviour during upper and lower endoscopy. Section 3.5 analyses the results obtained during the case study, in particular, focusing on the differences in performance metrics, characteristic gaze parameters and visual attention strategies with respect

to key AOIs during the challenging instances of upper and lower endoscopy procedures. The chapter finishes in Section 3.6 by discussing the research issues that are to be further investigated in the rest of the thesis.

3.1 Endoscopy Procedure

Endoscopy is a general term to describe the visualisation of the inner lining of a tubular body structure such as the gastrointestinal tract, usually via a natural orifice, with an additional ability to carry out interventions. Mainly visualisation is accomplished by a flexible scope with a controllable tip. Endoscopes today employ advancements in fibre optics with a light source and lenses in order to illuminate the FOV. They contain shared internal channels for passing interchangeable flexible instruments which can be used for: a) biopsy, b) diathermy, c) incision, d) injection, e) snaring, f) clipping, g) banding and, h) retrieval. The shared channels can also be used for insufflation, suction, and irrigation of the lumen in order to create, maintain, and optimise the FOV. Although visualisation can also be achieved with rigid scopes (*e.g.,* rigid oesophagoscopy), they are limited to a maximum range of approximately $25cm$ from the point of entry.

Endoscopy is most commonly performed in the upper and lower parts of the gastrointestinal tract. When performed in the upper aspect, it is called an Oesophagogastroduodenoscopy (OGD) and when performed in the lower aspect it can be called either flexible sigmoidoscopy (partial colonic examination) or colonoscopy (complete colonic examination).

With reference to Fig. 3.1, during an OGD procedure, the endoscopist navigates from the patient's mouth over the tongue into the oropharynx, and down to the oesophagus. Gradually, the endoscope is passed down the oesophagus into the stomach and eventually through the pylorus to examine the duodenum. Once this is completed, the endoscope is withdrawn back into the stomach to conduct a thorough examination including retroflexing the tip (a 180° flexion of the endoscope tip) of the scope (J manoeuvre) to examine the fundus and gastroesophageal junction. Any additional procedures are performed at this stage including biopsies before the endoscope is fully withdrawn from the patient.

In the case of a colonoscopy procedure, the endoscopist navigates from the patient's anus up the rectum, the sigmoid colon, descending colon, past the

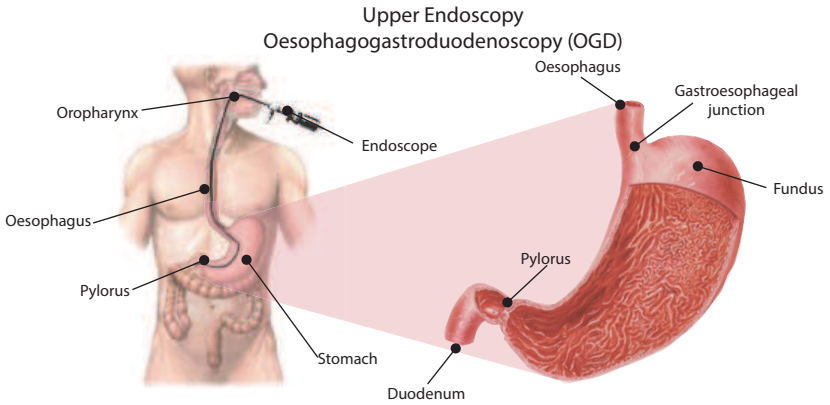


Figure 3.1 – Illustration of oesophagogastroduodenoscopy (OGD) and the stomach anatomy.

splenic flexure to go through the transverse, hepatic flexure and the ascending colon to finally reach the caecum and terminal ileum. The navigational path and anatomical features are illustrated in Fig. 3.2. Conventionally, once the endoscopist reaches the caecum, examination of the colon walls and folds follows during the withdrawal process in order to detect any cancerous lesions.

Advances in the management of medical conditions of the gastrointestinal tract can be largely attributed to the widespread uptake of endoscopy and its allied technologies. This chapter focuses on discussing the existing challenges in endoscopy, highlighting the visual and navigational difficulties which may even be challenging for the experts [82, 133, 170].

3.2 Challenges in Endoscopy

Flexible endoscopes have some unique design limitations that make their use problematic for novices or trainees. The American Society for Gastrointestinal Endoscopy (ASGE) recommends that a minimum of 140 colonoscopies are performed by a trainee endoscopist before they can be deemed to have achieved competency [62], whilst the Joint Advisory Group (JAG, UK) requires a minimum of 200 cases [65]. The main challenges in endoscopic procedures are induced by the visualisation, orientation and perceptual-motor misalignment factors. These are discussed in more detail in the following

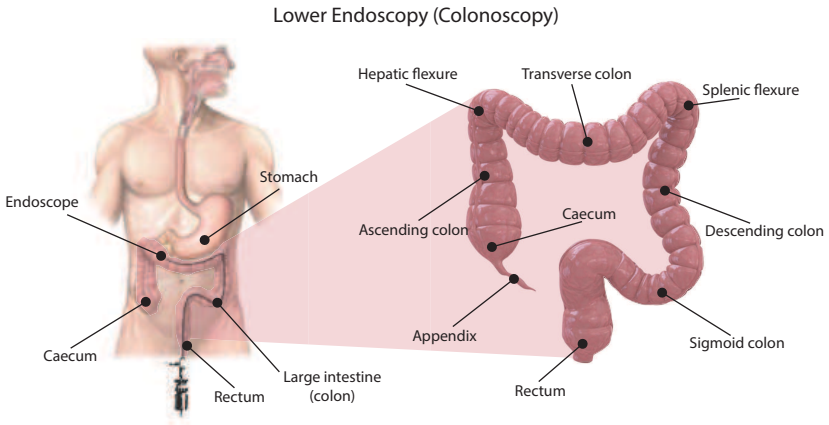


Figure 3.2 – Illustration of lower endoscopy (colonoscopy) and the human colon anatomy.

subsections.

3.2.1 Endoscope Visualisation and Orientation

The FOV with a conventional forward viewing endoscopes is 140°. This results in a significant proportion of the luminal wall not being adequately visualised [36]. This combined with the poor frequency of recognisable landmarks, results in frequent revisiting or missing of areas that need to be assessed. More critically, this can result in missed lesion detection, especially if the lesion is located on the proximal aspect of folds or flexures. Pickhardt *et al.* [117] mapped locations of non-rectal neoplasms that were detected by Computer Tomography (CT) colonography but were missed by colonoscopy and found that 67% of missed lesions were on the proximal aspects of haustral folds, rectal valves, flexures and the ileocaecal valve.

In addition, the design of the flexible endoscope with varying shaft stiffness makes it a 'floppy' and dependent device, *i.e.,* it is unable to support its own weight and is therefore guided by the shape of the colon, resulting in a large number of shape configurations. Coupled with the fact that the gastrointestinal tract changes shape, and can be highly mobile and elastic, particularly around the sigmoid colon and transverse colon as illustrated in Fig. 3.3, means that the endoscopist is constantly battling to maintain a satisfactory view and orientation in order to carry out surveillance and

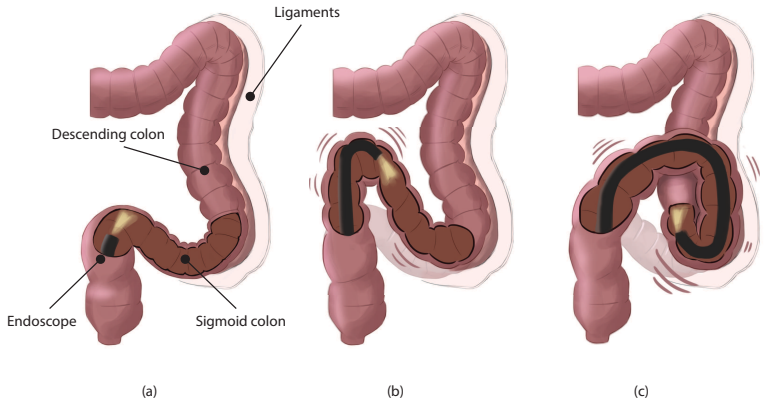


Figure 3.3 – Illustration of how the sigmoid colon can move and stretch from (a) the point of entering the anus (b) to insertion into the sigmoid colon and (c) if pushed further can form a loop.

interventional procedures. Poor visualisation can lead to a loss of one's sense of direction, position within the colon, and could therefore lead to disorientation.

3.2.2 Varying Perceptual-motor Misalignment during Navigation

Endoscope controls and views are designed to be fixed in alignment; despite this, there are certain scenarios that result in a paradoxical movement of the scope in relation to the view which can induce varying perceptual-motor misalignment. For the trainee this can be disorientating thereby prolonging the procedure time unnecessarily. There is also potential for increase in the discomfort that the patient experiences. The two scenarios that result in paradoxical movement are i) retroflexion (J manoeuvre) and ii) looping. These will be described further below.

Retroflexion Retroflexion or also referred as, 'J' manoeuvre, occurs when the endoscopist applies a 180° flexion to the tip in either the anteroposterior axis or the left-right axis. Fig. 3.4 shows the endoscope retroflexed and the corresponding camera view that it visualises. The illustration is that of a retroflexed endoscope within the human stomach, but is also performed during lower endoscopy. Retroflexion of the endoscope tip provides a rear-

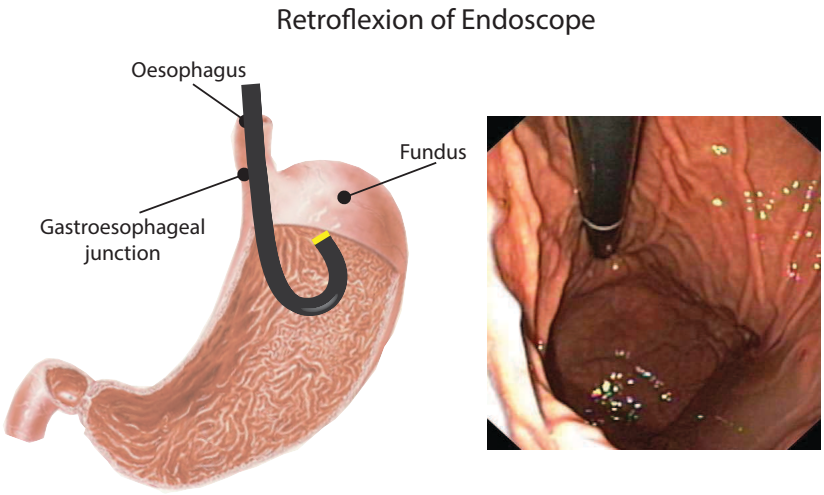


Figure 3.4 – Illustration of a retroflexed endoscope with the respective visualisation.

view of the endoscope which has been reported to provide important visual information that is not observable by the forward view alone [158]. Varadarajulu *et al.* [158] reports that during their study, 50% of patients with adenomatous polyps in the rectal vault were identified only by retroflexion during screening.

The consequences of performing a retroflexion is the mirroring in the movement of the scope tip to the controls. Despite this being an intentional manoeuvre (in the majority of occasions), perceptually the novice finds this phenomenon disorientating, and processing initially requires an adaptation period. Therefore, the ability to successfully retroflex the endoscope tip is a necessary skill for trainees to acquire during their endoscopic training [62].

Looping "Looping" is when the shaft of the flexible scope forms a loop configuration that results in paradoxical movement at the tip of the scope – thus when the endoscopist advances the instrument, the scope tip can move backwards and in an unintended direction [170]. This movement is not intended by the endoscopist and there is considerable literature [82, 133, 134, 138, 170] that states the frequent occurrence of looping as one of the largest factors that makes endoscopy challenging. When progressing to a loop formation, the long endoscope structure can also become twisted

along the length of the colon, due to the flexible tube structure, making the orientation of the camera visualisation to become misaligned from the endoscope controls. Any subsequent movement of the endoscope tip at this point can alter this mapping due the dependent and flexible nature of the endoscope, giving rise to a varying perceptual misalignment between the expected camera movement and actual resultant camera movement when the operator controls the endoscope tip from the proximal end of the endoscope. This results in a varying perceptual-motor misalignment. Looping can be one of the main causes of incomplete colonoscopic examinations, *i.e.,* failure to reach the caecum (refer to Fig. 3.2 for the colon anatomy), because when too much of the endoscope is left in the loop, the length of the endoscope becomes insufficient to reach the caecum. Saunders *et al.* [133] points out that when persistent looping occurs, the type of looping and location cannot be accurately assessed even for experienced endoscopists. Looping can occur due to patient-specific anatomical constraints including fixed (low elasticity) sigmoid colon [133] therefore making it hard to pass through the double bend of the sigmoid colon. In particular, the female pelvis anatomy is deeper than the male one, and furthermore, women tend to have longer colons (especially the transverse colon section) [131, 134]. These anatomical differences can cause the transverse colon to reach deep into the pelvis creating sharper colonic bends. Overall, this increases the difficulty in passing the endoscope, inducing loop formation and can result in greater discomfort during the procedure [61]. It has been widely reported that a colonoscopy can be more challenging in women [23, 61, 134, 6, 148]. Andersen *et al.* [6] reports that the following patient factors contribute to a more challenging endoscopy; a) older age b) a low body mass index (less than 25) c) history of constipation or reported laxative use. All of these factors contribute to more frequent loop formation which can make navigating inside the colon harder.

The U.S. Multisociety Task Force on Colorectal Cancer has set a target of 90% for caecal intubation, and encourages a 95% completion rate for screening circumstances [125]. Previous literature discussing completion rates from gastroenterologists at U.S. academic centres range from 88.0-97.2% with community gastroenterologists completion rates at similar levels of 83.6-97.0%. However, in a real-life study conducted by Cotton *et al.* [28], it is reported that of 69 participating endoscopists in seven major

U.S. centres, only 55% achieved a caecal intubation rate greater than 90%, with 9% actually achieving intubation rates of less than 80%. Furthermore, a study from the UK [11] reports gastroenterologists achieving caecal intubation rates of only 76.9% of the procedures. This figure becomes even lower at 56.9% when deducting the cases which didn't provide well documented evidence such as photographic evidence of reaching the ileocaecal valve (caecum). They report that reasons for failing to reach the caecum included patient discomfort (34.7%), looping (29.7%) and poor bowel preparation (19.6%). Shah *et al.* [138] found that looping occurred in 91% of the colonoscopic procedures in their study. A more recent study by Gavin *et al.* [49] reports an improved caecal intubation rate of 95.8% in the UK. Nevertheless, it is clear that looping during colonoscopy can be problematic for the endoscopist.

In order to develop a better understanding of looping, the following section describes the different types of loops that can occur in the colon. There have been reported a number of typical and atypical loop configurations and locations. These include:

1. Sigmoid spiral "N" loop (Fig. 3.5(a)) at the proximal descending colon.
2. Sigmoid alpha loop (Fig. 3.5(b)) at the splenic flexure.
3. Sigmoid reverse alpha loop (Fig. 3.5(c)) at the mid-sigmoid colon.
4. Sigmoid counter-clockwise spiral loop (Fig. 3.5(d)) at the distal sigmoid colon.
5. Transverse deep loop (Fig. 3.5(e)) at the start of the mid-transverse colon.
6. Transverse deep gamma loop (Fig. 3.5(f)) beyond the mid-transverse colon.

Shah *et al.* [138] reports that loops (1 & 5) are typical, whilst loops (2, 3, 4 & 6) are atypical. In order to navigate the instrument tip safely and comfortably for the patient and thereby achieve caecal intubation, it is critical to straighten out loops as they form. Thus, it becomes a critical skill for endoscopists to be able to detect loop formation and avoid forceful pushing. Overall, this can help reduce complications, such as perforation

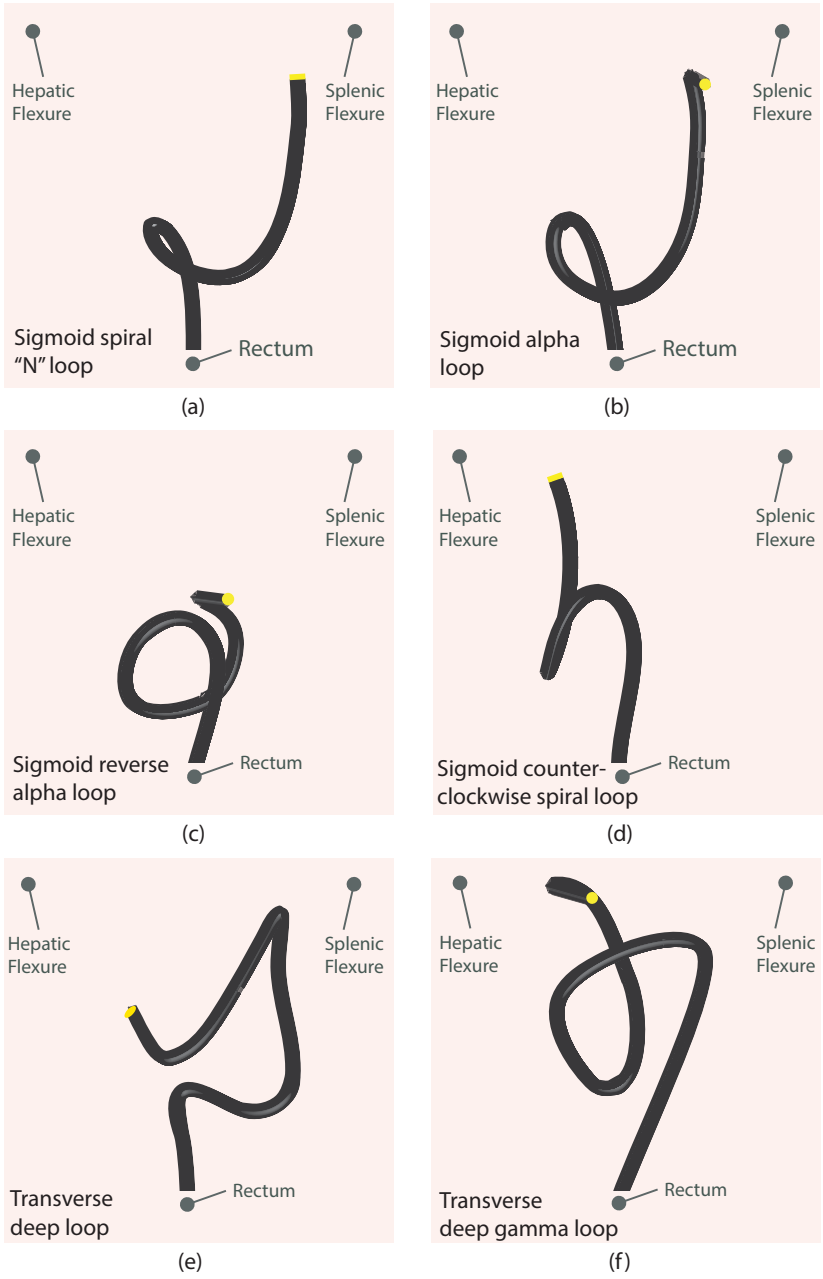


Figure 3.5 – Illustrations of the different sigmoid and transverse loops that occur during colonoscopy. (a) sigmoid spiral "N" loop (b) sigmoid alpha loop (c) sigmoid reverse alpha loop (d) sigmoid counter-clockwise spiral loop (e) transverse deep loop (f) transverse deep gamma loop.

of the colon. In the inevitable circumstances that a loop forms, the endoscopist needs to pull the endoscope backwards while applying a torque to the endoscope. Alternatively, endoscopists may ask the patient to change their body position and apply abdominal support from external hand pressure. This can often straighten the endoscope and allow the scope to move forward [167]. Shah *et al.* [138] report that although loops occur in men and women equally, atypical loops which can take longer to straighten out, occurred more frequently in women.

3.3 Endoscopy Technological Advancements

The technical challenges faced by endoscopists are a mixture of visualisation, orientation and perceptual-motor misalignment factors which are caused by physical limitations of the endoscope. These design limitations of the conventional endoscope and challenges related to the instrument interaction with the body are further magnified in advanced and hybrid techniques such as NOTES and endoscopic submucosal dissection (ESD). NOTES surgery requires the surgeon/endoscopist to navigate and manipulate the flexible scope outside the lumen of the gastrointestinal tract [84]. Without the constraints of the lumen wall to guide the scope, intraperitoneal navigation is extremely limited. In ESD, an electrocautery knife is used to resect the lesion in one piece, aiming to reduce the risk of recurrence and to allow a more accurate histopathological assessment. ESD requires the endoscopist to have dexterous control of the endoscope tip in order to excise early cancers with high precision and can be difficult to perform during a colonoscopy [151]. The balance of securing clear margins around the lesion versus perforating the viscus, makes this procedure time consuming and technically challenging. In order to overcome these problems various strategies have been employed, with many others still under research and development.

3.3.1 Retroscopes

The "Third Eye" retroscope system [156] is shown in Fig. 3.6(a) has been developed to providing a continuous retrograde view to existing forward-viewing endoscopes. The retroscope is passed down the working channel and bends 180° after the tip is outside the endoscope. Fig. 3.6(b) illustrates

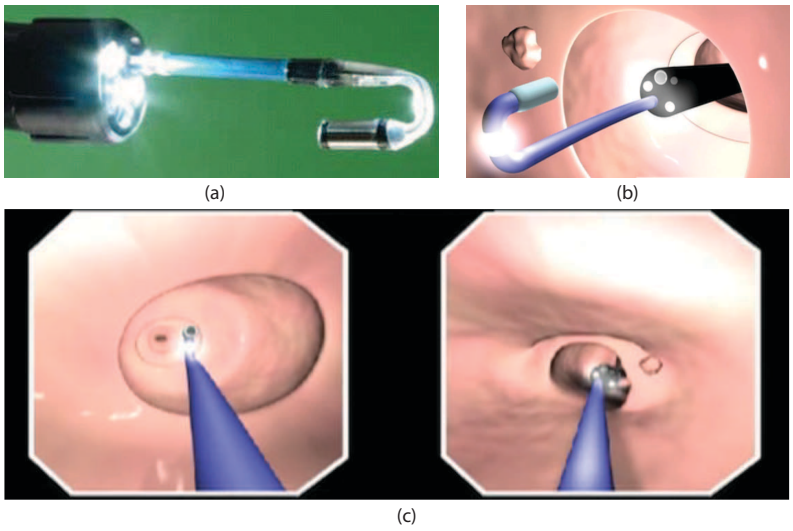
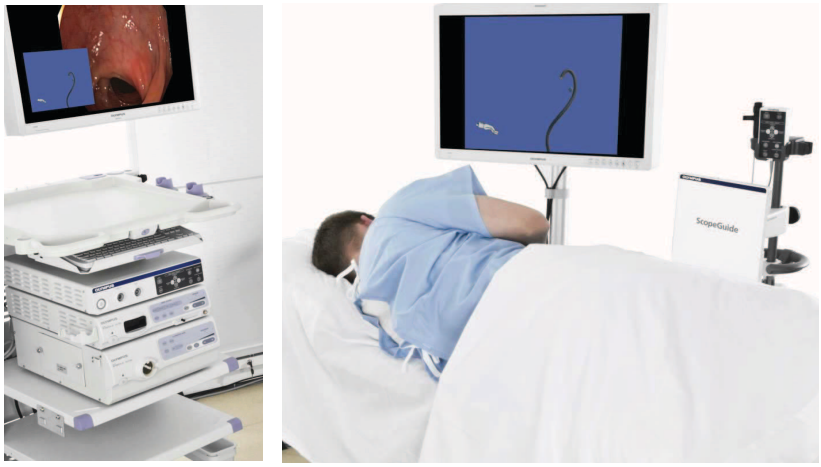


Figure 3.6 – Third Eye retroscope by Avantis Medical Systems, Inc. (a) image of the Third Eye retroscope, (b) illustrates how the Third Eye retroscope can easily detect lesions behind folds, (c) illustrates the endoscopist's view of the forward camera view on the left and the Third Eye retroscope view on the right.

how the Third Eye can help detect lesions behind folds. The instrument provides an additional visualisation to the forward conventional endoscope view and the endoscopist can visualise the front and rear-view simultaneously as shown in Fig. 3.6(c). This disposable endoscope can however add costs to the overall procedure and is likely to extend the length of endoscopy procedures as in circumstances that biopsies and polypectomies need to be taken, the retrograde scope needs to be withdrawn from the working channel [155]. A clinical study has reported improved adenoma detection rates [36] however, the clinical uptake of the Third Eye retroscope has yet to take off [181].

3.3.2 Narrow Band Imaging and Confocal Optical Microscopy

Narrow band imaging is recognised as a technique for improving the visualisation and therefore characterisation of lesions by the use an electronically activated filter for specific green and blue wavelengths of light [83]. Confocal microscopy is a probe based imaging technique used to increase optical



(a) (b)

Figure 3.7 – Scope Guide system by Olympus Corp. (a) shows the endoscope stack and how the overall endoscope shape can be visualised, (b) the magnetic field emitter and tracker enables the endoscopist to visualise the endoscope shape without having to use harmful radiation such as in fluroscopy.

resolution and contrast of a micrograph by using fluorescence point illumination and a spatial pinhole to eliminate out-of-focus light in specimens. It requires the use of a fluorescent tracer (*e.g.,* fluorescein) and is still under research [88, 83, 48]. Studies have so far failed to show that such techniques are clinically superior to standard definition white light endoscopy [181].

3.3.3 Scope Guide - External Endoscope Shape Visualisation

In order to help with loop detection and resolution during colonoscopy, a magnetic endoscope position detection system has been developed by Olympus Corp. This system shown in Fig. 3.7, is able to provide a real-time shape overview of the colonoscope position and configuration within the patient's colon. This method is preferable than fluroscopy, as there is no radiation exposure. However, the system is costly at $100,000.

3.3.4 Overtubes and Variable Stiffness Endoscopes

The Shapelock developed by USGI Medical USA is a metal overtube that is flexible when initially inserted into the gastrointestinal tract but its stiff-

ness can be adjusted to become rigid when desired. The product is used to enable passage of a colonoscope in a tortuous sigmoid colon through the rigid overtube without encountering sigmoid looping [170, 54]. The Endo-Ease Vista developed by Spirus Medical, West Bridgewater USA is another spiral overtube where it has been reported that using the overtube resulted in an increase in caecal intubation in patients whom standard colonoscopy has failed due to a redundant colon [135]. Olympus has developed variable stiffness colonoscopes to improve colonoscope navigation. The idea is that a stiffer colonoscope shaft can help reduce recurrent looping but at the same time can make passage through an angulated sigmoid colon more difficult causing further stretching and pain. Thus, by having an endoscope that can vary its stiffness as needed, it can potentially improve colonoscopic navigation [12]. Furthermore, in order to accommodate for the idiosyncrasies of the female colon, Olympus has also designed variable stiffness colonoscopes that tapers from the adult diameter to the paediatric diameter of $11.0mm$ at $25cm$. This device uses the features of both the adult colonoscope and the thinner paediatric colonoscope to improve colonoscopy in female colons [170]. However, clinical use of the product has found conflicting reports in clinical trials with the overtube colonoscope [20].

3.3.5 Robotic Endoscopes

There are a number of robotic endoscope platforms in development [162]. The NeoGuide [43] endoscope (acquired by Intuitive Surgical Inc. in 2009) consists of multiple articulated computer controlled segments along its length. Through the use of a sensor placed at the anus, the computer constructs a three-dimensional path map of the endoscope and uses this to make each segment of the endoscope follow the path traversed by the endoscope tip in a 'follow the leader' manner. Choset *et al.* [115] introduced a highly articulated robotic probe to provide a flexible assess route for epicardial surgery. The articulated section of their instrument comprises of two concentric tubes made of 50 links serially connected via spherical joints, where the instrument motion also achieves a 'follow the leader' technique. Yang *et al.* [139] pioneered a snake-like articulated hyper-redundant robot comprising of modular joint units. In their robotic system, each joint can be independently actuated offering the potential for greater instrument artic-

ulation. These systems offer the advantages of a self-supporting endoscope with a motorised tip for accurate navigation and improved stability. Such robotic endoscope platforms can be useful for more technical MIS procedures such as ESD and NOTES, on which the medical community is conducting a lot of research aiming to achieve better patient outcomes.

3.4 Case Study: Characteristics of Gaze Behaviour during Upper and Lower Endoscopic Navigation

Orientation, perception and visual identification of anatomical reference points are key elements of endoscopy. Disorientation may result in increased procedural errors and major complications. This is particularly true given the visually-driven constructs used in modern interventional endoscopy, such as natural orifice transluminal access methods which rely on a camera image relaying positional information back to the endoscopist.

To understand orientation, disorientation, and how this relates to an endoscopist's experience and skill, quantification is needed. Given that the visual system is responsible for the overwhelming majority of sensory input during endoscopy, a quantification of visual behaviour is likely to hold the key to understanding disorientation and other factors related to perception. Through using eye tracking technology, we aim to gain a preliminary understanding of important characteristics of gaze behaviour (eye movements and regions of attention) during upper and lower endoscopic procedures. Furthermore, characteristic differences in gaze behaviour of expert endoscopists between upper and lower endoscopies are investigated. Attention to specific anatomical and spatial reference points during an endoscopy procedure are also delineated.

Today's eye tracking devices, such as those discussed in Chapter 2 (Fig. 2.6(b)) can be worn unobtrusively, much like regular glasses. The tracked PoR data can be combined with image data from a forward-facing camera to determine the subject's areas of interest, deriving attention-related data such as gaze behaviour, or structures of interest.

Endoscopy is a skill which requires a significant period of training in order achieve competence [62]. The learning curve for acquiring basic com-

petency to perform a colonoscopy is approximately 200 procedures [163]. The acquisition of advanced interventional skills, such as ESD, is even more challenging.

Shortening of learning curves for skill acquisition have been demonstrated in laparoscopic surgery by the use of gaze training [169], a form of training to learn where to look, during the acquisition of laparoscopic technical skills. Laparoscopic surgery, similarly to endoscopy, is a skill which requires the practitioner to overcome perceptual-motor challenges associated with disorientation and navigation. As discussed in Chapter 2, eye tracking has been applied to the fields of usability and ergonomics where gaze analysis can help understand characteristics of specific perceptual features that result in making a procedure difficult.

3.4.1 Experimental Setup

In order to gain an understanding of the gaze behaviour and characteristics in the endoscopy suite, a pilot-study in a prospective, single-centre, observational, cohort study design was ethically approved by the local research ethics committee, reference 13/LO/0119. It was conducted at the Endoscopy unit in St Mary's Hospital, London, UK. OGDs, colonoscopies and sigmoidoscopies were sought. A senior consultant endoscopist with a minimum experience of 5,000 endoscopy procedures who was comfortable wearing eye tracking glasses whilst performing a procedure was enrolled. Two patients undergoing both upper and lower endoscopic procedures participated.

3.4.2 Experimental Protocols

Informed consent was obtained from the patients and the endoscopist with dedicated information leaflets provided. Anonymised demographic data about the procedure, patient and endoscopist was recorded and stored in accordance with data protection requirements. Details that were recorded included: task difficulty, experience, subject ID, procedure, interventions.

SensoMotoric Instruments (SMI) eye tracking glasses were used. The glasses have a spatial accuracy of 0.1° visual angle and 0.5° precision. It records at 30Hz, has a high definition scene camera recording the environment and two infrared cameras for binocular tracking. Data were collected onto

a personal digital assistant (PDA) attached by a belt clip to the endoscopist. The data were then downloaded and analysed on proprietary software (BeGaze, SensoMotoric Instruments, Berlin, Germany).

Eye tracking glasses were worn by the endoscopist before the procedure commenced and a one point calibration was carried out. Data were recorded throughout the endoscopy, from intubation until extubation, and procedural feedback about task difficulty was acquired from the endoscopist.

The surgeon was asked to rate the subjective difficulty of each respective upper and lower endoscopic procedure through the use of a National Aeronautical Space Agency - Task Load Index (NASA-TLX) questionnaire, a validated subjective questionnaire comprising of six variably weighted parameters that contribute to task workload [182]. The questionnaire is scored out of 100, and a higher score indicates a more challenging procedure. Basic patient demographic data were recorded, including age, gender, and complications. Video sequence of the surgeon's point of view of the procedure and the gaze data were recorded intra-operatively by the eye tracking device.

3.5 Results

In the result analysis, particular attention is paid first to discuss the gaze parameters during the whole procedure. Gaze parameters to be analysed include the surgeon's fixation duration and pupil diameter. Previous literature reports longer fixation duration [66, 67, 44] and pupil dilation [2, 7, 119] during challenging mental and cognitive tasks. The analysis discusses any differences between these gaze parameters during upper and lower endoscopy.

In order to associate the surgeon's PoR information with semantically meaningful context during the endoscope procedures, ten key AOI were identified, and each PoR on a frame-by-frame basis was assigned to one of the ten AOI whilst observing the video sequence. The ten AOI are later described in greater detail. The AOI PoR analysis along with the percentage dwell time on the ten AOI has potential to highlight the surgeon's visual search behaviour characteristics. Furthermore, metrics such as intubation time and the subjective difficulty of each procedure obtained from the NASA-TLX questionnaire are also discussed. This offers a way to highlight any differences in the difficulty in performing upper versus lower endoscopic

procedures.

Video and gaze data during challenging episodes of upper and lower endoscopy were subsequently segmented and analysed, including; i) sections with poor camera visualisation, ii) retroflexion of the endoscope, and iii) situations where the endoscope has formed a loop. Observation of such video sequences enables qualitative visualisation to help understand the challenges of the endoscopist in the endoscopy suite. The segmented three challenging episodes are shown in Fig. 3.8-Fig. 3.12. Along with these video sequences, the surgeon's gaze parameters during the three challenging episodes are compared to their overall averaged gaze parameters during each respective procedure. If we assume that the procedure average's values of these gaze parameters represent the control case, then comparison with the gaze parameters during the challenging episodes may provide insight to characteristic gaze parameters during the challenging episodes.

AOI Definitions Of the ten AOI, six were related to visualisation of objects inside the endoscopy suite; the patient, the nurse, instrumentation (for polypectomies), exterior view of the endoscope, the patient alarm monitor, and the Scope Guide monitor. The remaining four AOI were related to the endoscope camera visualisation and these include the gastrointestinal tract wall, the lumen, folds along the gastrointestinal tract, and the endoscope instrument visible from the endoscope camera view during a retroflexed position.

3.5.1 Gaze Comparison during Upper and Lower Endoscopy

The diagnostic/therapeutic procedures are divided into upper and lower endoscopy. In order to understand the nature of both procedures, the gaze parameters and performance metrics are analysed. The gaze parameters, including the fixation duration and pupil diameter measured over the whole experiment are summarised in Table. 3.1. The average fixation duration is slightly longer and the pupil diameter slightly smaller during colonoscopy when compared against OGD procedures. Given that the surgeon found colonoscopy to be more challenging than OGD procedures, as indicated by the higher NASA-TLX scores (36.5 vs 18.5 as shown in Table. 3.3), longer fixation durations, an indicator for greater cognitive focus during colono-

Table 3.1 – Gaze parameters of the surgeon averaged across the whole procedure.

	Average Fixation Duration (ms)		Average Pupil Diameter (mm)	
Patient Number	OGD	Colonoscopy	OGD	Colonoscopy
1	300.8	322.2	4.3	3.5
2	291.5	314.0	3.2	3.0
Average	**296.1**	**318.1**	**3.7**	**3.2**

scopy versus OGD (318.1ms vs 296.1ms) is consistent with the NASA-TLX scores. The smaller pupil diameter, which is an indicator of less cognitive demand during colonoscopy versus OGD (3.2mm vs 3.7mm), however contradicts the NASA-TLX scores. The inconsistency may be caused by the lighting condition, as the different lighting conditions during upper and lower endoscopic procedures could not be fully controlled for. Bright light causes the pupil to contract and become smaller. During colonoscopy, the surgeon uses a Scope Guide, fixating 9.5% of the time at the Scope Guide screen to help navigate the endoscope, as illustrated in Table. 3.2. Since the Scope Guide monitor is very bright (see Fig. 3.11), the pupil size would be biased towards becoming smaller than it would originally had if the scope guide was not used during the procedure.

When studying Table. 3.2, which summarises the percentage of fixation time on each AOI by the surgeon, it is clear that the surgeon fixated a greater proportion of time on the endoscope controls (Endoscope outside) during colonoscopy (4.69% vs 0.26%). This was to adjust or straighten the endoscope during looping, which was not necessary during an OGD procedure. An additional difference is the fixating on the Scope Guide monitor. However, this would not be a fair comparison, since during OGD procedures, the surgeon did not have the Scope Guide turned on. OGD procedures are considered easier and a Scope Guide is not usually used. Another observable difference is the greater proportion of time spent on fixating at folds during colonoscopy vs OGD procedures (10.37% vs 1.05%). This is likely due to the anatomical differences as there are more folds in the lower gastrointestinal tract.

When studying the intubation time (see Table. 3.3), colonoscopy procedures lasted a lot longer than OGD procedures (909s vs 266s) which is

Table 3.2 – Percentage of fixation time by the surgeon on each respective AOI during the upper and lower endoscopic procedures.

Fixation Time (%)	Wall	Fold	Lumen	Endoscope Outside	Endoscope Inside	Instrumentation	Scope Guide Monitor	Patient Alarm Monitor	Patient	Nurse
OGD	52.81	1.05	26.89	0.26	2.99	10.02	0.00	0.09	1.41	4.48
Colonoscopy	59.21	10.37	9.16	4.69	0.78	3.39	9.51	0.05	2.34	0.48

Table 3.3 – Intubation time and NASA-TLX score of the surgeon during the upper and lower endoscopic procedures.

	Intubation Time (s)		NASA-TLX Score	
Patient Number	OGD	Colonoscopy	OGD	Colonoscopy
1	280	973	22.0	30.7
2	252	846	15.0	42.3
Average	**266**	**909**	**18.5**	**36.5**

partially caused by the longer anatomical pathway, but also may indicate the comparatively more challenging nature of lower endoscopic procedures against upper endoscopic procedures.

One thing which should be noted here is that with such a small surgeon and patient sample size, these preliminary statistical comparison results should be interpreted with caution and used more as a way to qualitatively understand how the gaze parameters may change during upper and lower endoscopic procedures.

3.5.2 Poor Camera Visualisation

There were several episodes with poor camera visualisation, many of which can be caused by the combination of a small FOV, poor patient bowel preparation, and water injected during endoscope intubation. The video sequence in Fig. 3.8 with the respective magnified camera screen frames shown in Fig. 3.9 illustrates a typical example sequence of poor camera visualisation that can occur during a colonoscopy procedure. This video sequence lasted for 45 seconds and was partially caused by the endoscopist being looped in the colon. In fact, it was noticeable that the camera visualisation can become prohibitively worse during looping from observing additional procedures, as the camera visualisation at the endoscope tip moved in an unpredictable manner. During poor visualisation, the surgeon's gaze behaviour was frequently relying on the Scope Guide endoscope shape tracker to reorientate the endoscope and to grasp a better sense of the location of the endoscope within the patient. This may be due to the poor camera visualisation failing to provide any useful visual feedback to the surgeon.

The close up view of the camera visualisation monitor in Fig. 3.9 illustrates the challenges of endoscopy. Injected water for bowel cleansing can cause the camera visualisation to have multiple colour reflections. Further-

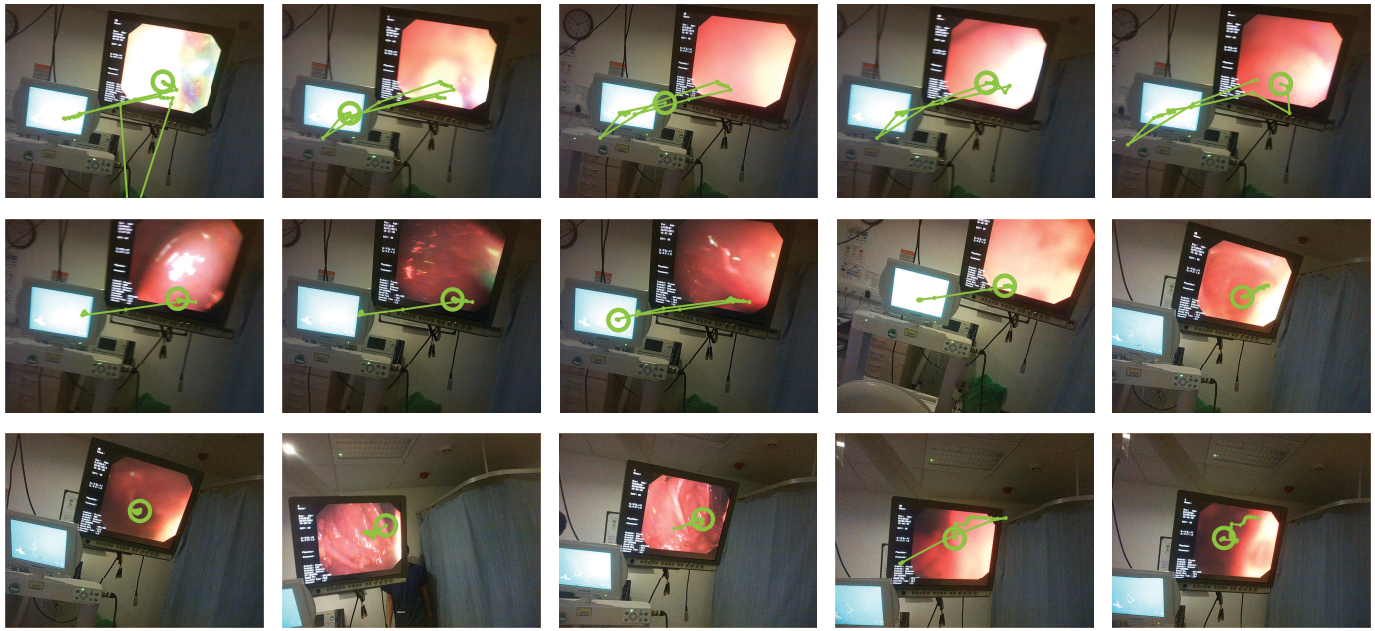


Figure 3.8 – Viewing from top left to bottom right, the video sequence illustrates the poor camera visualisation that can occur during endoscopic procedures. The small FOV combined with the paradoxical endoscope tip movement during looping can make it difficult to comprehend the endoscope's position from the camera visualisation alone. This particular colonoscopy video sequence was 45 seconds in duration.

more, even when the visualisation is clear, the small FOV makes it difficult to fully visualise the gastric wall immediately surrounding the endoscope tip without moving the endoscope tip up, down, left and right. Therefore it seemed that during intubation, navigating to the caecum was challenging. During extubation, effort is needed in order to detect lesions, especially those hidden behind gastrointestinal folds and bends. Overall, a high degree of dexterity and spatial awareness is required by the surgeons during the procedure.

3.5.3 Retroflexion

Earlier it was noted that retroflexion of the endoscope can be a challenging control technique, particularly for novice endoscopists. Since the recruited endoscopist for this pilot study was a consultant and thus very experienced, it is likely to differ to gaze behaviour or gaze parameters obtained from novices. However, in order to obtain a better understanding of the retroflexion process, it is useful to visualise the video sequence of then a retroflexion of the endoscope is performed. The video sequence shown in Fig. 3.10 is that of a retroflexion by the recruited endoscopist during an upper endoscopic procedure. The black tube-like endoscope can be observed within the endoscope camera monitor view. During the 10 second video sequence, the endoscopist flexes the scope by 180° to examine the fundus and gastroesophageal junction. From observing the video, the endoscopist performs this part of the procedure with little difficulty.

During both the upper and lower endoscopic procedures, the gaze parameters of the endoscopist during retroflexion were extracted and analysed. The data summary is observed in Table. 3.4. The surgeon did not perform any retroflexion manoeuvre during the lower endoscopy. When comparing the average fixation duration during OGD retroflexion to the average fixation duration during the whole OGD procedure (see Table. 3.1), ($336.6ms$ vs $296.1ms$), the fixation duration during retroflexion is higher. In addition, the average pupil diameter was also larger during retroflexion comparatively to the average pupil diameter during the whole OGD procedure ($4.0mm$ vs $3.7mm$). This result might be due to higher cognitive demand required for retroflexion manoeuvres. Although the surgeon was very experienced, retroflexion might be cognitively more demanding when compared to the rest

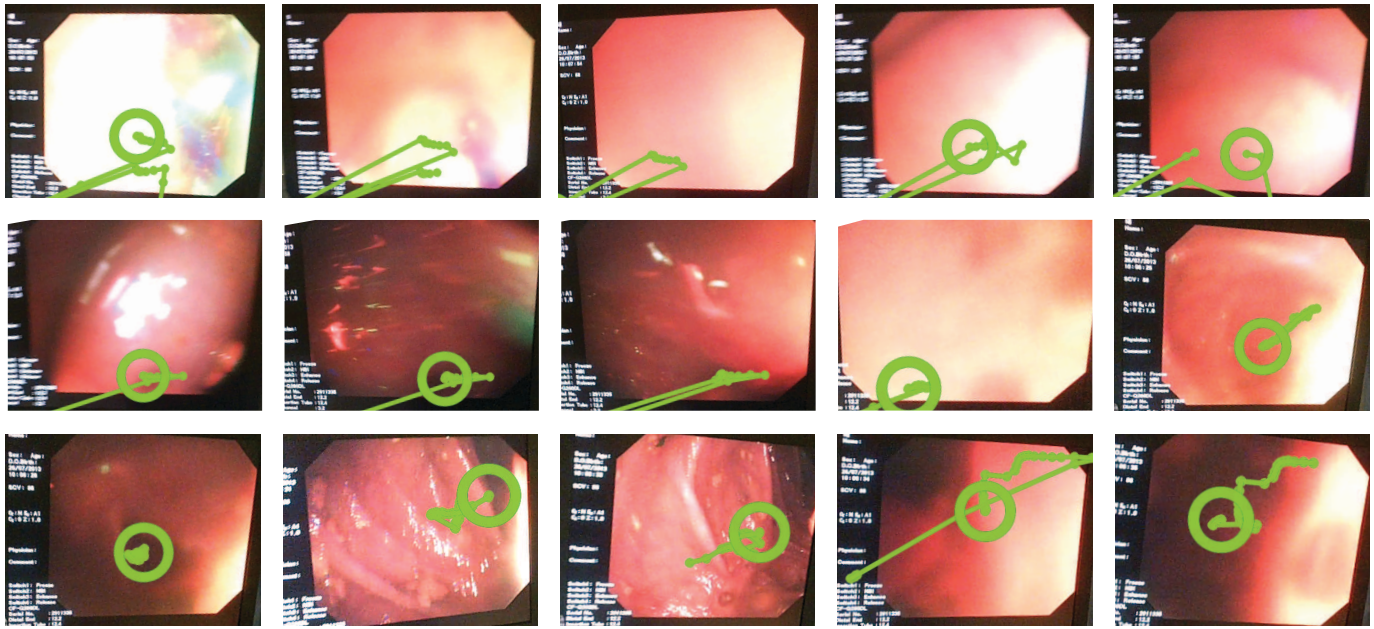


Figure 3.9 – Viewing from top left to bottom right, the video sequence is the magnified camera visualisation of the video sequence shown in Fig. 3.8. Poor visualisation can be caused by the injected water during bowel cleansing causing multiple colour reflections. Even when the camera visualisation is clear, the small FOV makes it difficult to fully visualise the gastric wall immediately surrounding the endoscope tip.

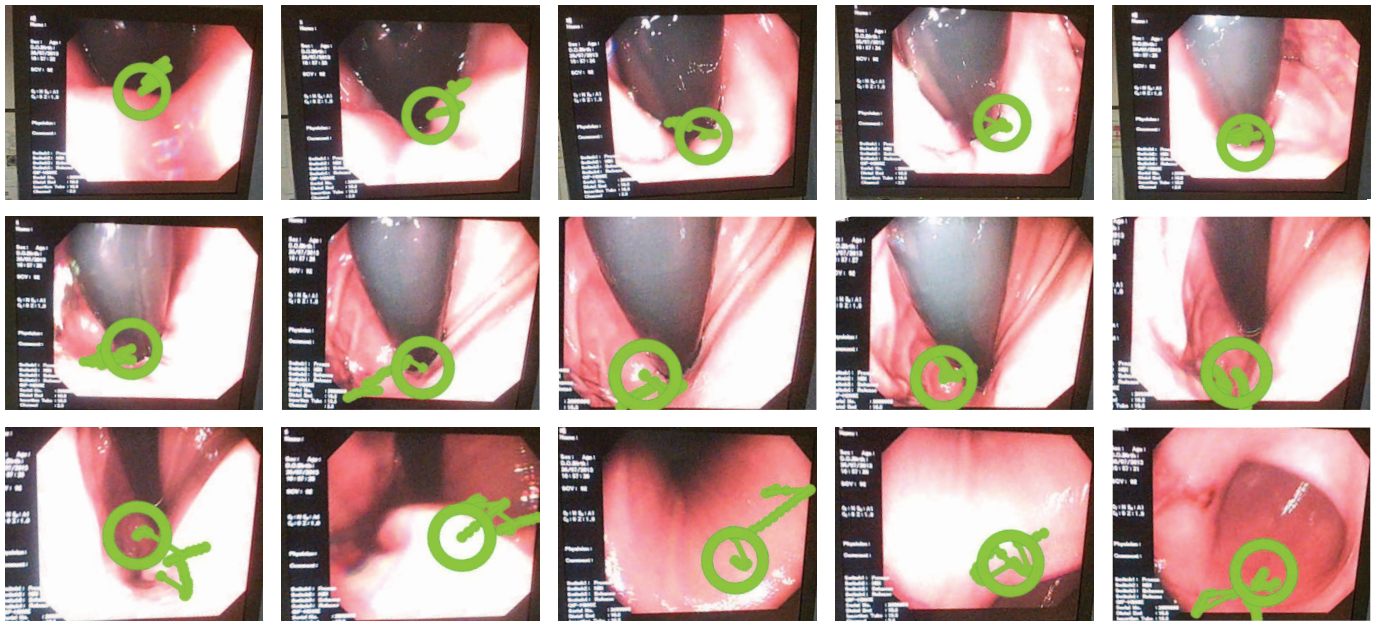


Figure 3.10 – Viewing from top left to bottom right, this video sequence shows an example visualisation of the endoscope retroflexed during an upper endoscopic procedure. The video sequence consists of examination of the fundus and gastroesophageal junction.

Table 3.4 – Gaze parameters of the surgeon during retroflexion in upper endoscopic procedures.

Patient Number	Average Fixation Duration (ms)	Average Pupil Diameter (mm)
1	293.3	4.2
2	380.0	3.8
Average	**336.6**	**4.0**

of the procedure.

3.5.4 Looping

Looping only occurred in the lower gastrointestinal tract and occurred in both patients. No looping was observed during the OGD procedures, presumably because the human gastric passage is linear and the upper gastrointestinal tract more fixed than the lower gastrointestinal tract. An example of a looping episode is illustrated in Fig. 3.11 and Fig. 3.12 which lasts 35 seconds and 40 seconds respectively. Referring to Fig. 3.11, in the looped episode video sequence the camera visualisation is poor and the endoscopist frequently gazes between the Scope Guide monitor and the camera visualisation screen. As the endoscopist struggles to understand the spatial location of the colonoscope tip, the endoscopist then asks the nurse for a magnetic tracker and places this on the patient's abdomen to isolate the colonoscope tip. The position of the tracker shows up on the Scope Guide monitor and helps to localise the colonoscope tip within the patient's gastrointestinal tract. Even still, the endoscopist continues to struggle to overcome the loop, and thus repeatedly manoeuvres the endoscope whilst cross referencing between the Scope Guide and endoscope camera visualisation as illustrated in Fig. 3.12. The endoscopist then adjusts the stiffness of the endoscope to help overcome the looping, as can be seen in Fig. 3.12 where there are fixations on the endoscope controls. The combined knowledge and endoscope adjustments allows the endoscopist to finally overcome the loop after one minute and fifteen seconds.

Looping can cause significant challenges even for consultant surgeons, and therefore study of the gaze parameters during such episodes are interesting.

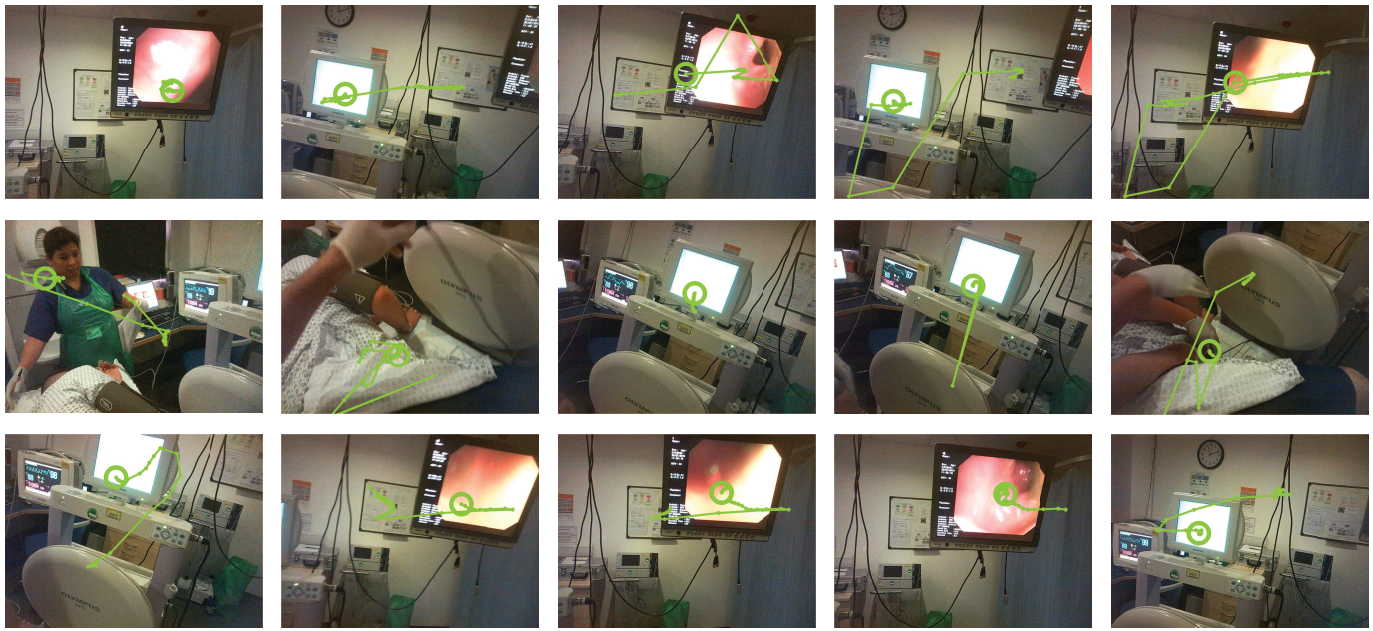


Figure 3.11 – Viewing from top left to bottom right, this video sequence illustrates looping of the endoscope in the patient's lower gastrointestinal tract. During looping, camera visualisation is poor and the endoscopist frequently gazes between the Scope Guide monitor and the camera visualisation screen. The endoscopist uses a magnetic tracker and places this on the patient's abdomen to isolate the colonoscope tip. The position of the tracker shows up on the Scope Guide monitor which can help the endoscopist understand where the endoscope tip is within the patient's gastrointestinal tract.

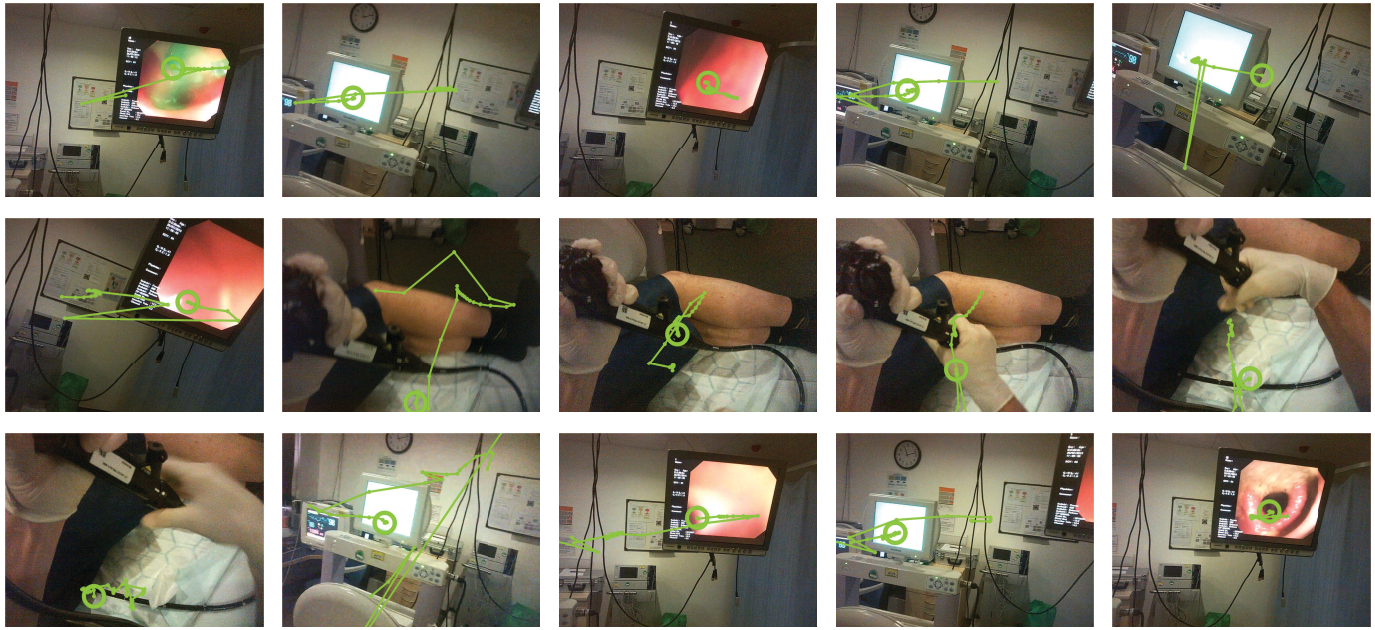


Figure 3.12 – Viewing from top left to bottom right, the video sequence is a continuation of the looped episode from Fig. 3.11. The endoscopist who is still looped, continues to move the endoscope whilst cross referencing between the Scope Guide and endoscope camera visualisation. The endoscopist adjusts the stiffness of the endoscope where he fixates at the endoscope's stiffness controls to try to get out of the loop. The endoscopist eventually overcomes the loop.

Table 3.5 – Gaze parameters of the surgeon during looping in lower endoscopic procedures.

Patient Number	Average Fixation Duration (ms)	Average Pupil Diameter (mm)
1	353.8	3.3
2	372.7	2.9
Average	**363.2**	**3.1**

Table. 3.5 summarises the gaze parameters extracted upon looping episodes during colonoscopy. Comparatively to the average gaze parameters during the whole colonoscopy procedure, the average fixation duration was slightly longer during the looping episode (363.2ms vs 318.1ms). For the pupil diameter, there was little difference between the loop episode parameter value versus the averaged value of the whole procedure (3.1mm vs 3.2mm), but this might be due to the endoscopist fixating more onto the Scope Guide bright monitor screen more frequently, cancelling out any dilation effects caused by increased cognitive load that might have been reflected in the pupil dilation otherwise.

3.6 Conclusions

This chapter focussed on investigating the challenges faced by surgeons in the operating theatre, specifically in the areas of perception, navigation and ergonomics. Particular attention was paid towards understanding the challenges faced by endoscopists. However, given the surgical community's interest towards less invasive methods for patients, overcoming the mentioned challenges are important for any surgical procedure. Significant research has already been conducted for the development of natural orifice access surgical instruments such as articulated multiple joint robotic systems to perform MIS via a single port or natural orifices where a review of such systems can be found in [162]. Gaze tracking provides a way to understand the surgeon's visual behaviour and ergonomics, and from the study in this chapter, it was noted that endoscope camera visualisation can become challenging at times, particularly when having to orientate the endoscope during looped circumstances. The results from the study highlight

the potential usefulness in the surgeon's gaze behaviour information to understand the difficulty of a particular surgical procedure, as the fixation duration and surgeon's fixation on particular AOI can provide cues to their cognitive state. Without additional guidance during navigation by external endoscope shape sensing devices such as the Scope Guide, the poor visualisation can hinder intubation progress during endoscopy. Furthermore, during looping, when perceptual-motor misalignment may occur, where the visual axis of the camera is no longer aligned to the surgeon's motor control of the endoscope, can pose challenges to the surgeon where de-looping becomes more of an ad-hoc trial and error of different manoeuvres such as twisting the scope, tightening the stiffness, applying abdominal pressure to the patient stomach and asking the patient to change their body position.

In the subsequent chapters, the following key research issues are further investigated:

- Improved surgical visualisation with larger FOV imagery
- The varying perceptual-motor misalignment during endoscopic navigation
- Improving the ergonomics in an already crowded surgical theatre environment with multiple devices and systems

The next chapter focuses on improving the camera visualisation by creating a larger FOV image visualisation technique that also encodes the camera trajectory information. A comprehensive user study is carried out to assess the clinical efficacy of this new technique.

4 Usability Study of Dynamic View Expansion for Surgical Visualisation

As discussed in Chapter 3, poor camera visualisation poses challenges during endoscopic navigation, potentially causing the endoscopist to become disorientated during navigation. Poor camera visualisation and disorientation causes problems in newer surgical techniques such as Single Incision Laparoscopic Surgery (SILS) and NOTES [123], which can hinder safe navigation in-vivo. Disorientation during endoscopy is induced by a number of factors including looping in the gastrointestinal tract and retroflexion, causing difficulty in interpreting anatomical references. These challenges were discussed in Chapter 3 and it was highlighted that a high degree of dexterity and spatial awareness is required by the surgeon since the endoscope provides only a limited FOV.

DVE is a panoramic image where a visualisation is progressively stitched using sequential images from the video feed to create an effectively enlarged FOV [96, 164]. Similar mosaicing techniques have been used to provide a larger FOV in order to navigate during surgery [81]. However, when observing an expanded FOV, it is not trivial to determine the motion trajectory of the endoscope. While most DVE techniques manage to achieve visually pleasing and artifact-free results, the challenge of distinguishing the current endoscope view from expanded still remains. A new DVE technique was introduced by Totz *et al.* [154], which not only increased the camera visualisation FOV, but also provide grey-scale visual cues to encode the camera motion trajectory. The grey-visual cues encode the motion trajectory by slowly fading the colour to grey-scale along its path aiming to facilitate the temporal evaluation of the scene. This provides a reference for changes in camera position and orientation so as to help the operator understand

how the camera has moved over the past few seconds. This fade-to-grey DVE method aims to combine spatial and temporal references to assist the surgeon to re-orientate in circumstances of disorientation.

The purpose of this chapter is to investigate the effectiveness of this fade-to-grey DVE visualisation scheme compared to conventional full-colour DVE in terms of usability and more importantly for visual search behaviour. There has been significant research conducted to understand visual search behaviour [174] and how it subsequently changes during re-orientation [144]. The conducted analysis aims to gain a deeper understanding of the visual search behaviour, when the two different visualisation methods are used. A detailed usability study measuring the response correctness and speed of participants estimating the camera trajectory for both DVE methods was conducted. Eye tracking data were recorded to study the visual search behaviour of participants whilst they estimate the camera trajectory. The dynamics of participants' gaze movement is characterised using Markov chains to understand the difference in visual search behaviour between the two DVE image stimuli. Resultant state transition graphs and heat-maps formed from the participants' eye movements were used to investigate the difference in visual search behaviour between grey and less grey regions. Characterising visual search behaviour using Markov modelling has been previously carried out by Dempere-Marco *et al.* [37] and Hacisalihzade *et al.* [51]. For this study, a novel state definition based on the greyness of the given Region of Interest (ROI) was used in the state transition graphs.

The study aims to understand the current advantages and limitations of the fade-to-grey DVE technique over conventional DVE to aid during disorientation. Section 4.1.1 gives the technical background on the DVE methodology used for the study and Section 4.1.2 explains how Markov Chains can be used to assess the visual behaviour of participants while performing a task. The user-study experiment is described in Section 4.2. In Section 4.3, the results regarding the response correctness and latency for both DVE imaging techniques are analysed. Furthermore, the difference in visual search behaviour by using the transition graphs formed from the participant's eye movements as well as the statistics of some key characteristics of the visual behaviour for both techniques are discussed. Conclusion and discussion regarding the study follows in Section 4.4.



Figure 4.1 – Experimental setup showing a study participant conducting the experiment in front of a Tobii 1750 gaze tracker used to collect the data.

4.1 Methods

In order to provide assessment of the new DVE scheme, a detailed user study was conducted by comparing the conventional DVE and the new fade-to-grey DVE method. Three sequences were used, two of which were taken from porcine in-vivo abdominal explorations [95] and one from silicone NOTES phantom recordings [25]. For any given viewpoint, two DVE images were generated in pairs; one fade-to-grey and one full-colour visualisation. During the experiment the subjects were shown one generated image at the time and were asked to identify the camera trajectory from the visualisations shown. Showing several pairs of images generated with both methods in a randomised order allowed direct comparison of performance for the two methods. An illustration of the experimental setup is shown in Fig. 4.1. In order to assess the visual behaviour of the subjects, eye tracking data were recorded with a Tobii 1750 gaze tracker throughout the experiment and in order to perform a detailed analysis of the data, Markov chains and state transition graphs were used. More details on the DVE method as well as of the techniques used to assess the visual behaviour are given in the next subsections.

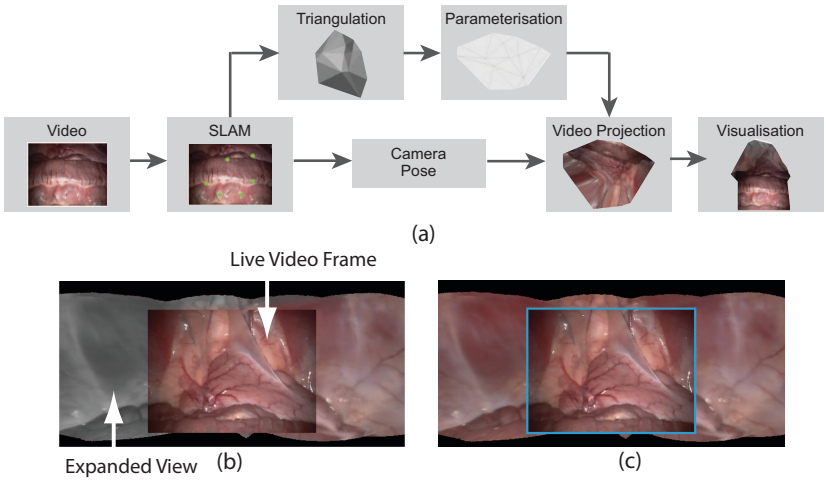


Figure 4.2 – (a) The fade-to-grey DVE texture processing flow chart. Video frames are processed with SLAM to estimate the surgical scene's geometry and create a non-planar panorama with older video frames. (b) An example of a fade-to-grey DVE image. (c) Same example images with conventional full-colour DVE technique.

4.1.1 Dynamic View Expansion

The basic concept of DVE is to augment live camera images with previously observed footage by creating a non-planar mosaic. This expands the camera view to a larger FOV based on the recent camera trajectory. Due to significant illumination changes in endoscopic views, noticeable seams appear in expanded views. Previous methods attempted to blend seams between video and expanded regions with Poisson Image Editing [96]. Blending techniques create aesthetically clean looking with minimum artefacts images but at the same time raise an issue of perception and recognition of what is live video and what is outdated information. To address this problem, a new fade-to-grey DVE visualisation method was proposed by Totz *et al.* [154] that renders historic information on the expansion by desaturating it to grey-scale, whilst retaining high fidelity of the current live-view.

The DVE image generation process is shown in (Fig. 4.2(a)). Every incoming video frame is processed using Simultaneous Localisation and Mapping (SLAM). In surgical scene reconstruction, SLAM is used to predict and estimate the relative position of the camera and a set of features in 3D [96], which usually are landmarks on the tissue surface. A sparse point cloud

is created and the points are used as features in order to match areas of interest, which incrementally form the SLAM map. By rendering the point cloud into a triangle mesh by using Delaunay-triangulation and by selecting parts of frames as texture maps and blending them, a panorama of the scene is built and a photorealistic view of the scene is generated. To simplify the video frame accumulation for panorama construction, a single planar texture domain is computed, covering every point of the mesh. Into this texture domain, video frames are projected over time. For visualisation, the triangulated SLAM map is texture-mapped with the texture domain image and rendered from the SLAM estimated camera position. Combined with the unaltered live video frame, this provides the perspective correct progressively constructed expanded view. This essentially expands the narrow FOV, as the historic view of the scene reconstructed by SLAM is combined with the current view. The challenge though is to be able to clearly distinguish the current view from the historic expansion view. In order to make the distinction between the live camera frame and the historic data in the expanded view, a slowly fading grey scale is applied on the historic data. The 'age' of every pixel in a texture image is set to zero in the beginning and is increased every time the texture map is updated, while it resets to zero every time new pixels are written in the texture image. Further details can be found in [154].

This DVE technique not only highlights the difference between live video and the expanded view, but it also implicitly encodes the camera's trajectory, providing a reference for changes in position and orientation in addition to anatomical features. An example of this visualisation scheme is illustrated in (Fig. 4.2(b)). The utility of this visualisation method was examined by having participants estimate the camera trajectory for the fade-to-grey and full-colour DVE visualisations. The fade-to-grey images were produced by using the fading method described above, while the conventional DVE images were presented with a rectangle highlighting the current live video frame. The fade-to-grey encoding is expected to enhance orientation, since the trajectory of the camera movement is implied by the fading grey scale. Moreover, this visualisation method is expected to be interpreted more easily, as it imitates the human vision system by keeping the most recent and important information as colour in the centre of the image, whilst the surrounding appears grey.

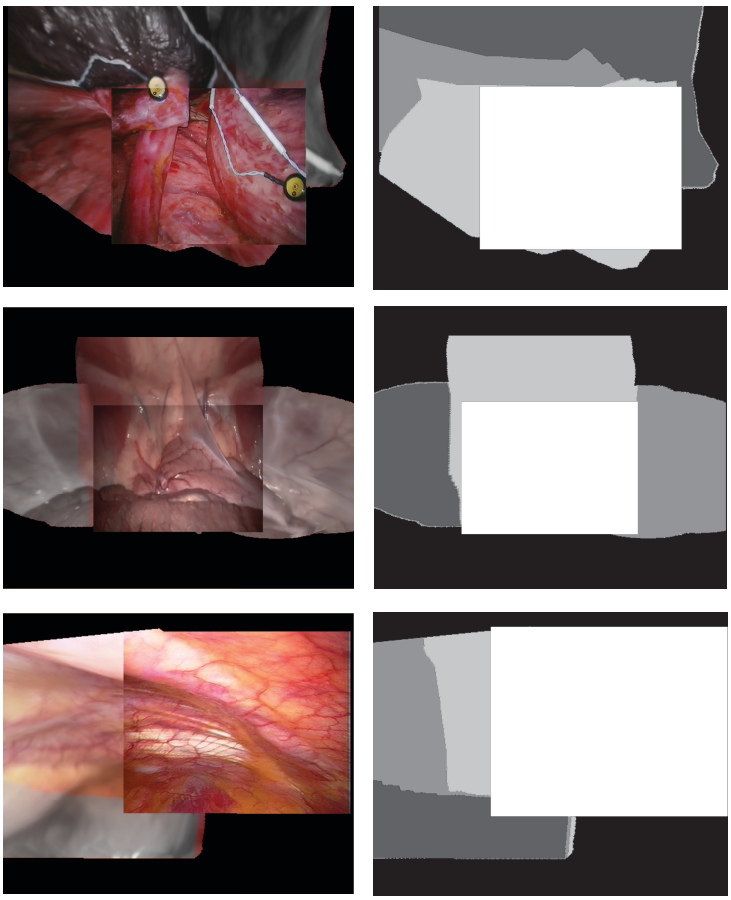


Figure 4.3 – Segmented regions on example fade-to-grey DVEs of the three sequences used.

4.1.2 Markov Chain and State Transition Graphs

In order to analyse participants' visual search behaviour, Markov chains were used to model their eye movements. Markov chains are commonly used to analyse systems with stochastic processes whose characteristic behaviour can be modelled by enumerating all the states it can enter. Such modelling has been previously used for characterisation of eye tracking visual search behaviour and scan path analysis [37, 51]. For this study, a first order Markov chain was used.

The recorded gaze data were first processed using a fixation filtering al-

gorithm by Salvucci *et al.* [132] to extract the participants' focus of interest for each DVE image. Here we assume that the participants' attention is only focused on the fixation points; therefore mechanisms for assimilating visual information such as factors relating to parallel search and peripheral vision were excluded. The fixation filtering algorithm is dispersion based, where the dispersion threshold and fixation duration were set to 1° of visual angle and $150ms$ respectively.

The states in the Markov chain were constructed according to how old the image region was. Thus, all ten fade-to-grey DVE images were segmented into four ROIs according to the temporal order of each region. Previous methods have used clustered eye tracking data as ROIs [37] or were arbitrarily defined manually by the author [51]. This novel definition of ROI effectively segments the fade-to-grey image into different ROIs according to their greyness. Fig. 4.3 illustrates the ROIs for three different DVEs: the darkest grey is the oldest temporal region of the DVE image and the white region is the live video frame. The other two shades of grey represent the temporal regions in-between. The segmented ROIs are subsequently used as independent states in the Markov chain analysis of the visual search behaviour.

The extracted fixation data were assumed to follow the Markov property and each fixation was then projected onto the ROI image. Since each fixation was assigned to only one of four regions on the ROI image, this enabled association of each fixation to one of the four states. Transition matrices were formed by summing the transitions between each respective state per image per subject. Note that intrastate-transitions were ignored, as only visual search behaviour between different regions of the image was of interest. Subsequently, transition probabilities P_{ij} between states i and j were obtained by normalising each interstate-transition by the total number of interstate-transitions t_{ij} from that state, giving rise to a 4×4 transition probability matrix. The same ROIs extracted from the fade-to-grey DVE images were also used for the corresponding full-colour DVE images during the gaze data analysis to directly compare the group's visual search behaviour during visualisation of each respective DVE image stimuli. Once a transition matrix was formed per participant per image, an equally weighted average of the 15 participants' transition matrices were calculated per image to create transition matrices that represent the group visual search beha-

viour for both DVE imaging techniques.

4.2 Experimental Setup

In order to provide detailed visual assessment of the new DVE scheme, the experimental setup described in Section 4.1 was used. Fifteen participants, consisting of seven surgeons and eight research engineers, were recruited and seated in front of a Tobii 1750 gaze tracker (Fig. 4.1). Participants were calibrated offline to the eye tracker first and their gaze was recorded in the background for the duration of the experiment.

The task given to participants was to identify the camera trajectory from the visualisations shown in full-screen. An example visualisation is shown in Fig. 4.4(a). While viewing the full-screen DVE visualisations, participants were asked to hypothesise the camera trajectory. No time restriction was imposed at this stage of viewing the full-screen visualisations. However, participants were encouraged to answer as promptly as possible and they were informed that the amount of time spent viewing the full-screen images, *i.e.,* the response latency, was being recorded. Following a keyboard press, they proceeded to the next stage to choose their identified trajectory from a list of three options, as illustrated in Fig. 4.4(b). Only one of the options was correct, with their order randomised. Initially, the DVE image was shown as background underneath the options but disappeared after three seconds. This short period of time was deemed long enough to relate the trajectories to the previous image, but also short enough to avoid participants cross-referencing the options and ignoring the previous full-screen image; *i.e.,* participants had to be confident about their estimated trajectory before switching to the options. This ensured that a sufficient amount of reliable eye tracking data capturing the participants' visual search behaviour could be collected.

The procedure was repeated with a total of twenty images; ten for each DVE method in a randomised order and each of the ten visualisation images was used only once to avoid potential learning effects. In addition, some images were flipped left-to-right in a randomised method to decrease the chance of learning from previous images. Participants were asked to work through the experiment as quickly, but as accurately as possible. The participants' selected option and response latency were recorded for offline

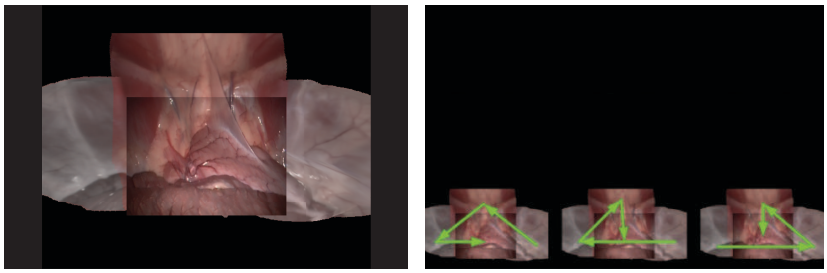


Figure 4.4 – (a) One fade-to-grey DVE image with (b) the corresponding trajectory options.

performance analysis. Prior to the experiment, participants received an introduction and a chance to familiarise themselves with the test procedure in a training session.

4.3 Results

In order to analyse the effectiveness of the proposed scheme, a comparison of the response correctness and latency was conducted. A statistical analysis was also carried out and involved conducting normality tests, followed by Analysis of variance (ANOVA) tests. A 'p-value' <0.05 was considered significant. The visual search assessment analysis was made by using Markov transition matrices, but also by comparing the density and trajectory of fixations analysed by the use of heat-maps. The detailed results are presented in the following sections.

4.3.1 Response Correctness and Latency

A comparison of the percentage of correct responses of the participants' answers and the response latency averaged over all images and participants is presented in Table. 4.1. The data for each DVE method is shown with its respective standard deviation values. As can be seen, participants identified the camera trajectory more correctly and consistently for fade-to-grey DVE images, indicated by the statistically significant p-value of 0.015. Participants also reached a decision faster on average with fade-to-grey DVE images. The average response latency was comparatively longer for fade-to-grey DVE images over the full-colour DVE images, also showing statistical

Table 4.1 – Group-averaged percentage of correct responses and response latency over all images with respective standard deviations. On average, participants answered more correctly and quickly when observing fade-to-grey DVE images to determine the camera trajectory.

Method	Percentage Correctness ± Stdev	Average Response Latency ± Stdev
Full-colour	45% ± 15.5%	7.79s ± 0.98s
Fade-to-grey	63% ± 8.5%	6.45s ± 1.55s

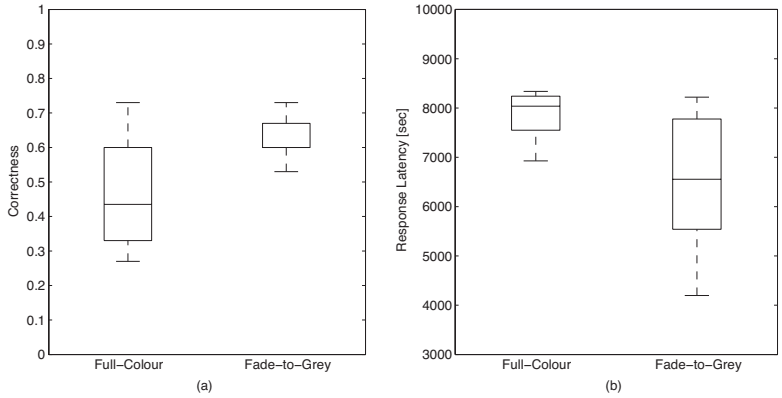


Figure 4.5 – Comparison of response correctness and response latency between the two groups of images. (a) response correctness obtained when viewing full-colour and fade-to-grey DVE images. (b) response latency obtained when viewing full-colour and fade-to-grey DVE images.

significance with a p-value of 0.029. The overall group's percentage of correct responses and response latency is illustrated in Fig. 4.5.

A comparison of the percentage of correct responses and response latency averaged over all images but the eight research engineers, and seven surgeons grouped separately, is shown in Table. 4.2 The behaviour of the two groups were similar and no statistically significant difference was observed between the two group's performance for each DVE method. The p-value for the percentage of correctness for the full-colour DVE images was p=0.34 between the two groups, while for the fade-to grey it was p=0.28. The average response time is also similar among the two groups, with a p-value of p=0.57 for the full-colour DVE method and a p-value of p=0.43 for the fade-

Table 4.2 – Percentage of correct responses and response latency over all images separated into surgeons and non-surgeons.

Method	Percentage Correctness		Average Response Latency	
	Non-Surgeons	Surgeons	Non-Surgeons	Surgeons
Full-colour	46%	43%	7.80s	7.77s
Fade-to-grey	64%	62%	6.20s	6.74s

Table 4.3 – Group-averaged percentage of correctness and response latency per image, for each DVE method.

Image Number	Percentage Correctness (%)		Average Response Latency (s)	
	Full-colour	Fade-to-grey	Full-colour	Fade-to-grey
1	60.0	73.3	8.01	5.58
2	53.3	60.0	8.24	5.60
3	26.7	60.0	8.12	7.78
4	26.7	80.0	8.34	4.37
5	73.3	60.0	8.06	5.54
6	46.7	60.0	9.42	7.52
7	40.0	53.3	6.93	7.66
8	33.3	66.7	7.57	8.22
9	40.0	53.3	7.55	8.04
10	60.0	66.7	5.69	4.20

to-grey DVE method. However, both groups answered more correctly by observing the fade-to-grey DVE images when having to determine the camera trajectory. Furthermore, even though there was no significant difference between research engineers and surgeons in terms of response latency, both groups were faster at providing their camera trajectory estimation when visualising the fade-to-grey DVE images.

In order to help understand the cause of the larger response latency standard deviation for fade-to-grey DVE images and do a further break-down of the results, the response latency and percentage of correct responses, averaged across all 15 participants for each of the 10 images for both DVE methods is shown Table. 4.3. It is observable that for some images the difference in the percentage of correct response is a lot higher for the fade-to-grey method, while for other images the difference between the two methods is not as noticeable. Similarly, when comparing the average response latency for the two methods, it is noted that for some images the subjects reach a decision much faster in the case of the fade-to-grey DVE method. However,

in some cases, like Image 8, it takes longer for the subject to respond for the fade-to-grey method.

Some of the images are more challenging than others, which on one hand makes the dataset a good sample of real scenarios, but on the other, makes the analysis more challenging. It was decided to separate the images in two groups according to their response accuracy and latency. The fade-to-grey DVE images with higher percentage of correct responses, but a similar average response latency between the two DVE methods, were categorised as Group 1 and included, Image 3, Image 4, Image 7, Image 8 and Image 9. The fade-to-grey DVE images' percentage of correct responses for the images in Group 1 turned out to be significantly better than the full-colour DVE counter part, p=0.017. The difference in the response time between the two methods for Group 1 images was not significant p=0.614. The rest of the images were assigned to Group 2, *i.e.,* Image 1, Image 2, Image 5, Image 6 and Image 10. For Group 2 images, there was no significant difference in percentage of correct response rate between the two DVE methods with p=0.338, but the fade-to-grey DVE images had significantly lower response latency with p=0.001. This shows that the fade-to-grey DVE images helped the subjects reach a decision much faster. An illustrative summary is observable in Fig. 4.6 and Fig. 4.7, where the comparison in percentage of correctness and the response latency between the two groups is shown respectively. The results indicate that in both cases, the fade-to-grey method is beneficial. However, it seems that some images were more challenging than others, namely Images 7-9. These images will be discussed further below.

From Table. 4.3 it is clear that images 7-9 were the only fade-to-grey DVE images of Group 1 that participants spent more time studying them in comparison to their respective full-colour images before making a decision. The key difference between these three images and the rest of the images was that these were the only images that had the live video frame overlapping with the camera trajectory. This is illustrated in (Fig. 4.8). This overlap introduces a discontinuity in the fade-to-grey expanded view and also creates a greater number of potential trajectory permutations. Thus, it seems plausible that participants found these three images' trajectory estimation more challenging and therefore spent a longer period viewing them. However, like the rest of the images in Group 1, although participants spent on

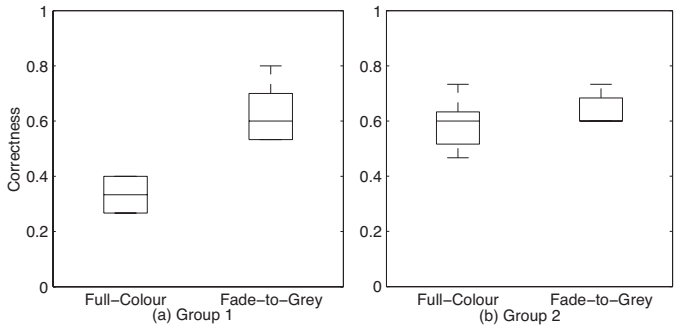


Figure 4.6 – Comparison of correctness between the two groups of images. (a) Response correctness for Group 1 images. The percentage of the correct responses when using the fade-to-grey DVE method is a higher than when using full-colour DVE. (b) Response correctness for Group 2 images. The correctness for Group 2 images are similar, albeit, the fade-to-grey method achieving slightly better performance.

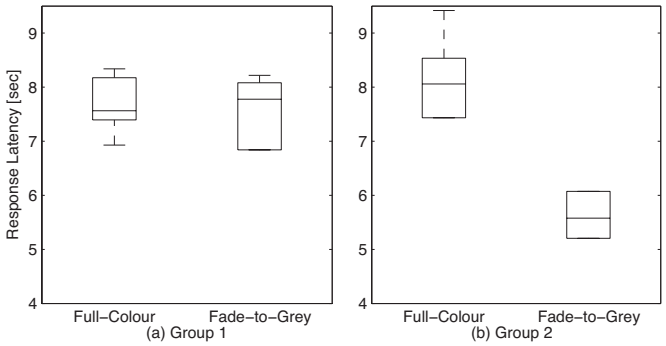


Figure 4.7 – Comparison of latency response measured in seconds between the two groups of images. (a) Response latency for Group 1 images. The response latency for Group 1 images is similar between the two methods, (b) Response latency Group 2 images. It took longer for the subjects to respond using the conventional full-colour DVE method, while with the fade-to-grey DVE images, the response latency is reduced.

average a longer time observing the fade-to-grey DVE images 7-9, the correct trajectory estimation rate was still higher for these images when comparing the percentage correctness to their corresponding full-colour versions. Therefore, the grey cues provided by the fade-to-grey technique were most likely to be helpful to participants. It could be speculated that with the full-colour DVE images 7-9, participants decided to guess the camera trajectory, as they couldn't find helpful visual cues to enable them to confidently estimate a trajectory, thus spending comparatively less time observing these images as can be seen from Table. 4.3.

When the group-averaged percentage of correct responses of the two DVE imaging techniques are compared, apart from image 5, participants answered more correctly for the fade-to-grey DVE stimuli. Image 5, which was the outlier for percentage of response correctness in the fade-to-grey DVE image stimuli, shows a limitation of the fade-to-grey DVE technique. Due to the nature of the grey-scale shading, originally darker regions that are actually newer temporally can appear greyer and therefore older than an originally brighter but actually older region. It was noticed that participants were confusing greyness with intensity: in (Fig. 4.9(b)) region 2 is perceived more grey than region 1 even though region 2 is simply darker due to its larger distance to the scope light source. In addition, the full-colour visualisation of image 5 (Fig. 4.9(a)) had noticeable image blur, which coincided with historical camera movement. This is caused by re-projections of the texture domain used for video data accumulation due to topology changes in the SLAM map after feature additions and deletions. Older regions of the expanded view appear more blurred because more re-projections have occurred since that respective area of the surgical scene had been observed. Hence, there is a chance that participants were using the blur cues to estimate camera trajectory, which help explain the higher percentage of response correctness for this particular image.

4.3.2 Visual Search Behaviour Characterised by State Transition Diagrams

In order to assess differences in the participants' visual search behaviour when viewing the two different DVE image stimuli, Markov transition probability matrices were formed from the filtered fixations obtained during the

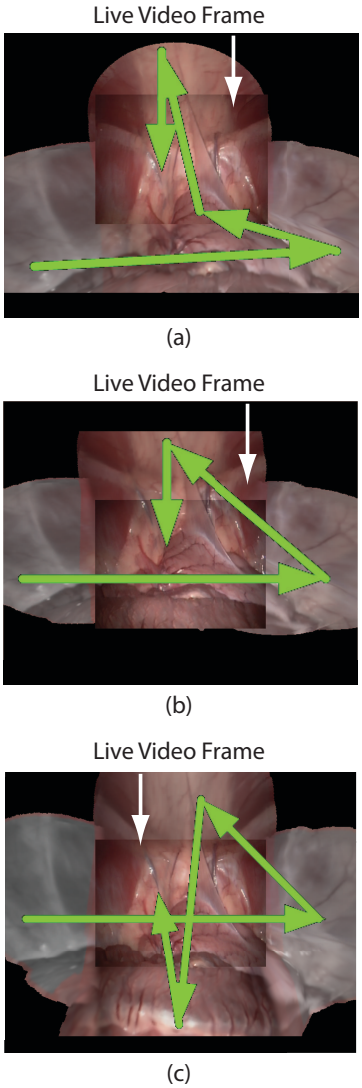


Figure 4.8 – Fade-to-grey DVE images 7-9 (a)-(c) with respective camera trajectories overlaid. The overlap of the live video frame with the trajectory introduces a discontinuity in the fade-to-grey expanded view, potentially making these images more challenging to estimate the trajectory.

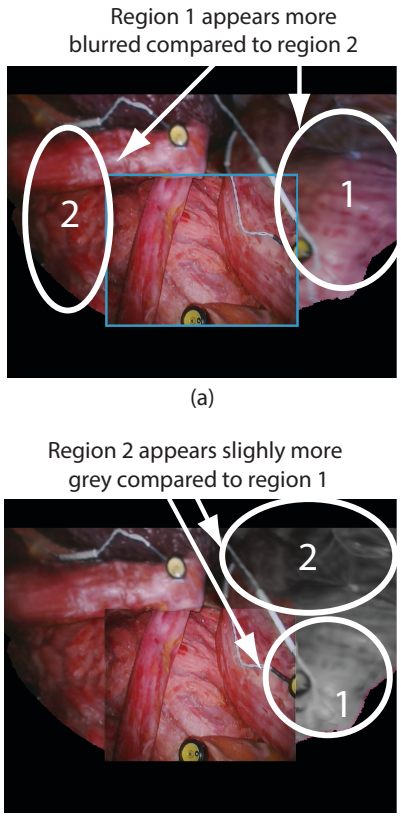


Figure 4.9 – Image 5 (a) full-colour DVE and (b) fade-to-grey DVE. Some of the limitations of this DVE visualisation technique is observable in this image.

user experiments. The dispersion based fixation filtering algorithm incorporated a dispersion threshold and fixation duration of 1° of visual angle and $150ms$ respectively. Further details of how the Markov transition probability matrices were formed can be found in Section 4.1.2. From the statistical analysis conducted in the previous section, it was found that different images cause different behaviours and they were divided in two groups, Group 1 and Group 2. It is expected that the images of the two groups cause a different visual behaviour. In order to assess the differences, state transitions graphs of four images from each group were plotted and analysed. The state transition graphs for the images' stimuli that had the largest difference in percentage of correct responses between the two DVE methods, *i.e.,* Group 1, are shown in Fig. 4.10 , while the ones for the images that resulted in a significant response latency difference, *i.e.,* Group 2, are shown in Fig. 4.12.

From Fig. 4.10(a)-(b) the observable key difference in participants' visual search behaviour when viewing the two DVE image stimuli lies in the comparatively higher probability of participants cross-referencing between the oldest region (state 1) and the almost live region (state 3), when looking at fade-to-grey DVE images. This implies a higher probability of saccading from the most-grey region to the least-grey region. Whilst for both DVE image stimuli participants showed frequent transitions between state 1 and 4, and 3 and 4, the transition between state 1 and 3 was much weaker in full-colour DVE image stimuli. The state transition graphs in Fig. 4.10(c)-(d) show a similar pattern, where the transition between states 1 and 4, 2 and 4, 3 and 4 are present in both graphs; a key transition between states 1 and 2 is lacking in the full-colour DVE transition graph. The presence of grey cues, which gives rise to cross-referencing between key transition states, suggests that the grey cues were influencing participants' visual search behaviour and a large proportion of the group was making use of these cues to help them during the trajectory estimation.

Fig 4.11 shows the stimuli used for the two methods and the corresponding gaze-hotspot images which represent their visual search behaviour. The heat-maps represent the density of the fixations of all subjects, which are overlaid on to the image that was visualised as the stimuli. It is observable from the fade-to-grey DVE image in Fig 4.11(a) that subjects concentrated mostly on region 4, which is the newest frame and regions 1 and 2, before deciding on the trajectory. In contrast, the gaze behaviour for the conven-

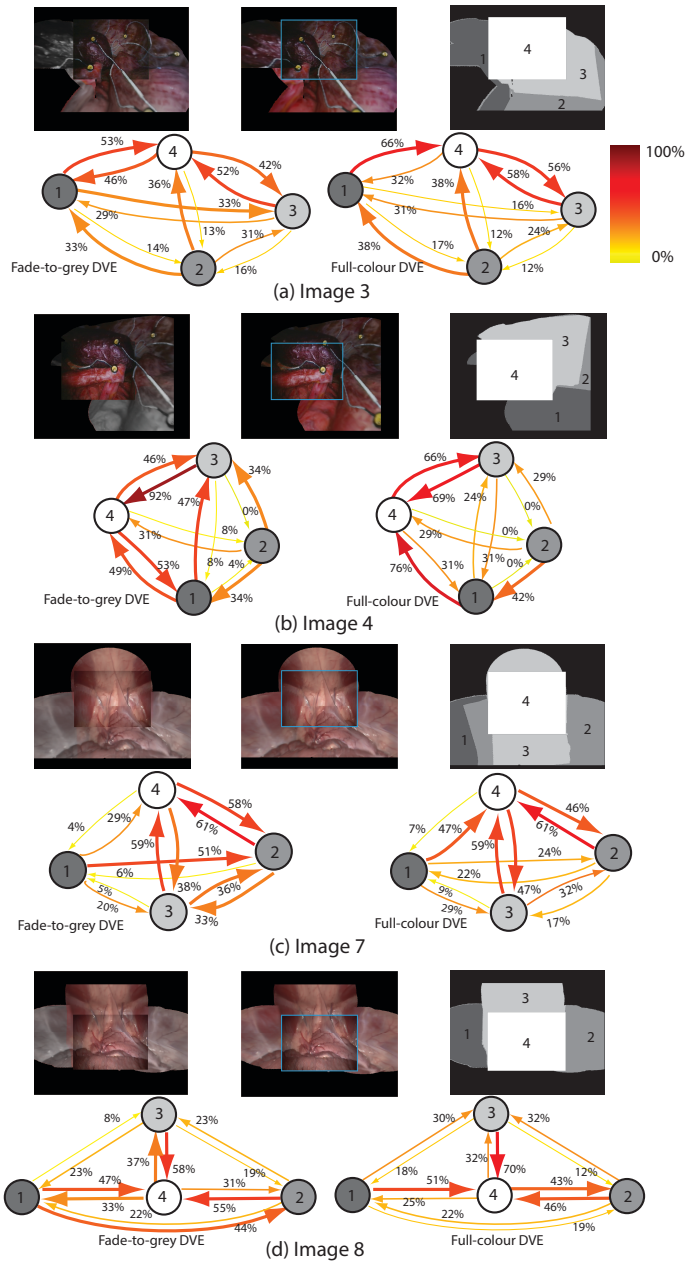


Figure 4.10 – Markov state transition graphs for Group 1. There are notable transitions in the fade-to-grey DVE state transition graphs between states 1 and 3 for (a) and (b), and state 1 and 2 for (c) and (d) which are faint in the corresponding full-colour DVE state transition graphs.

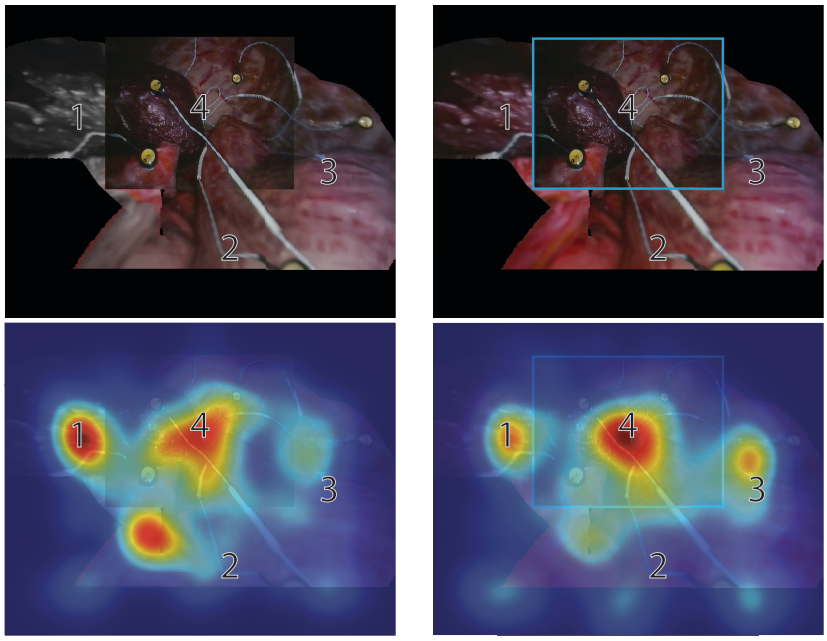


Figure 4.11 – Comparison of the two DVE visualisation methods and the respective gaze heat-maps for Image 3 (a) Fade-to-grey DVE method. Fixations are more concentrated around the areas of interest and in particular in regions 1 and 4 (b) Conventional DVE method. Fixations are more scattered

tional DVE, shown in Fig 4.11(b), is more random; subjects focused mainly on the central frame and the rest of the time around the other regions but without a distinct fixation focus. This is probably because there is little information to help the subject reach a decision. This is also reflected in the response correctness of the group's trajectory estimates where the group average shows significantly improved trajectory estimation for fade-to-grey DVE images over full-colour DVE images. In addition, feedback from a large number of participants, expressed that it was easier to estimate the camera trajectory from the fade-to-grey DVE images, whilst often they found themselves guessing or using the blur information as a cue for the full-colour DVE images.

A different visual behaviour was observed when studying the transition graphs of Group 2 images shown in Fig. 4.12(a)-(c). It is observable that

both types of DVE images exhibit similar graph probabilities across states 3 and 4, 1 and 4, and 1 and 3, and there is no marked difference in visual search behaviour. The colours of the arrows, which represent the probability between the transition states, are very similar for all images generated by both DVE methods. This means that both methods induce a similar subject gaze behaviour. The gaze heat-map also indicates similar gaze behaviour, where Fig. 4.13 shows the heat-map of one of the images *i.e.,* Image 2. The concentration of fixations for the fade-to-grey method, depicted in Fig. 4.13(a), is similar to the one captured for the conventional DVE method, as it can be seen in Fig. 4.13(b). This may explain the similarity in the percentage of correctness between both DVE methods, but the subjects were still able to decide on the trajectory faster for the fade-to-grey method.

The only image which is different to the rest of the images of Group 2 is Image 5 illustrated in Fig. 4.12(d). For this particular image the group's percentage of correct responses was slightly higher for the full-colour DVE image at 73.3% compared to 60% for the fade-to-grey DVE version. The higher response correctness for this full-colour DVE image could have been caused by the blur cues illustrated in Fig. 4.9(a) and by the limitation of the fade-to-grey DVE technique, shown in Fig. 4.9(b). The blur cues provide temporal trajectory information that was introduced by the limitations of the full-colour DVE technique thus potentially aiding in estimating a correct trajectory. The technical limitation of the fade-to-grey DVE technique caused darker regions of the image to appear greyer and therefore appear more aged than they actually were, misleading participants.

4.4 Conclusions

Prevention and overcoming of operator disorientation during colonoscopy, SILS or NOTES procedures is one of the critical challenges that need to be overcome to enable widespread adoption of these less invasive but more technical procedures. Dynamic view expansion provides a potential solution to the limited field of view and potentially alleviates the problem of disorientation during endoscopic interventions.

In this chapter, a systematic usability study of a new fade-to-grey DVE visualisation technique was conducted to assess its effectiveness in providing the operator visual cues regarding the recent camera trajectory for reorient-

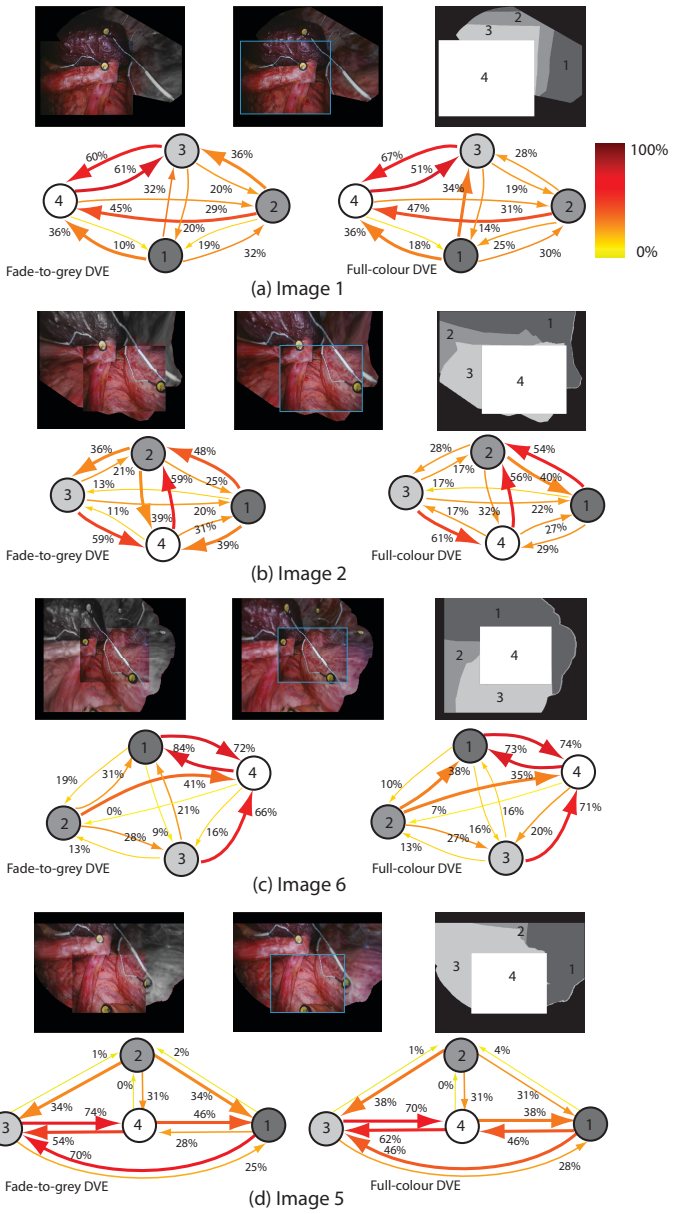


Figure 4.12 – Markov state transition graphs for Group 2. (a)-(c) State transition probabilities have similar characteristics of cross-referencing between key regions of the image. (d) For this particular image, participants were able to use the blur present in the full-colour version to infer the camera trajectory.

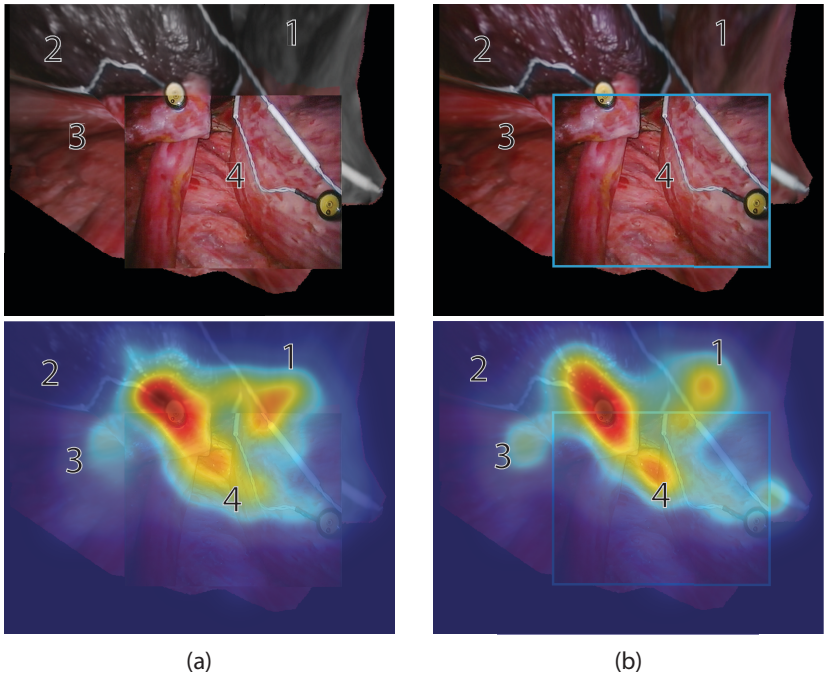


Figure 4.13 – Comparison of the two DVE visualisation methods and the respective gaze heat-maps for Image 2. The density of fixations across the regions of interest for both methods is similar. (a) Fade-to-grey DVE method. (b) Conventional DVE method.

ation. Depending on the level of complexity of the trajectory and the quality of the expanded view, the new DVE technique can be beneficial either in achieving higher percentage of correct responses or in reducing the response latency. Quantitative data were presented in terms of performance and response latency. The statistical analysis provides evidence for the benefits of the new method and facilitates the analysis by enabling formation of two groups of images. The two groups of images were causing different results and gaze behaviour.

Visual search behaviour was modelled using Markov chains and a new ROI formulation based on the greyness of various regions in the fade-to-grey DVE images. Modelling of visual attention using Markov chains combined with gaze heat-maps was effective in quantitatively highlighting the group's search behaviour differences between the two DVE methods and also between the two groups of images. The state transition graphs showed that subjects cross-referenced more between older and newer temporal regions when viewing the fade-to-grey DVE images and the density of fixations indicated by the heat-map showed concentration in the regions of interest where temporal cues were stronger. This may explain the superior performance in trajectory estimate during visualisation of Group 1 fade-to-grey DVE images. In contrast, the state transition graphs and the gaze heat-maps for Group 2 fade-to-grey DVE images showed little difference from their full-colour DVE images counterparts, which may help explain the similar rate of response correctness. In some of the Group 2 images, the blur cues in the conventional full-colour DVE image gave trajectory cue information. However, even then, the fade-to-grey DVE technique helped subjects achieve a smaller response latency. It should be noted that the fade-to-grey DVE imaging technique is far from perfect. Like its full-colour counterpart, the computational demands of generating the fade-to-grey DVE images hinders it from being implemented in real-time. Furthermore, the SLAM feature tracking capabilities can suffer under tissue deformation causing fixed anatomical features to move, therefore, affecting the resulting quality of both DVE image techniques. In addition, the fade-to-grey DVE technique has potential to cause confusion among users where darker regions can falsely appear more grey (more aged). However, the results from this study still demonstrate that there is some potential clinical value of visualising fade-to-grey DVE images. When computational demands are no longer an issue,

it might be useful to have it as an optional visualisation modality which can be switched on demand if the surgeon needs information to reorientate them self.

In summary, the detailed user study comparison of the two different visualisation techniques showed that the fade-to-grey DVE imaging method was helpful to users during identification of the camera trajectory in comparison to the conventional full-colour DVE imaging method. Furthermore, the work in this chapter demonstrates how eye tracking can be applied to assess the visual search behaviour of participants while performing a given task. The visual attention strategies correlate with the statistical analysis in terms of correctness and response latency results. In endoscopic procedures, where the cognitive burden during looping can be high, gaze data may be a valuable tool to help assess trainee performance. The limited FOV is one of the challenges apparent in endoscopic procedures and visualisation is only one aspect of improving the surgical capabilities. A system which can assess disorientation during navigation and compensate for it would be another way towards improved navigation.

In this chapter, gaze tracking was shown to be valuable in assessing the usability of newly developed techniques, in this case, for improving endoscope camera visualisation. In the next chapter, we discuss a way to quantify perceptual-motor misalignment. Through a study involving this misalignment, we quantify the gaze behaviour and investigate how varying perceptual-motor misalignment can affect task performance.

5 Perceptual-Motor Misalignment during Surgical Navigation

In Chapter 3 it was noted that another challenge during endoscopy is overcoming 'looping'. Looping is when the shaft of the flexible scope forms a loop configuration that results in paradoxical movement at the tip of the scope. Prevention or loop resolution can also be useful for more advanced surgical procedures like SILS and NOTES [84]. Currently, endoscopy is associated with a steep learning curve and even the most experienced gastroenterologists can become looped whilst navigating along the colon, inducing some degree of spatial disorientation [16, 14]. Previous studies have suggested that disorientation during MIS can be caused by a combination of the following factors: the limited manipulability of the endoscope due to the endoscope being a dependent device, *i.e.,* it is unable to support its own weight and therefore is guided by its environment [84], the lack of meaningful haptic feedback, the lack of salient tissue landmarks and limited FOV of the camera visualisation to help during navigation and orientation [13].

Holden and Flach [56] investigated the mismatch of spatial orientation between a laparoscope camera visualisation and the physical instrument workspace. They report that such a mismatch can lead to a decline in surgeon's performance which suggests that performance can depend on having a consistent mapping between the virtual eyes and hands. Their results also indicate that moving the camera during surgery can potentially disrupt coordinated action due to increased burden induced by proprioceptive adaptation [152]. To overcome this issue Holler *et al.* [58, 57] used a Microelectromechanical systems (MEMS) tri-axial sensor attached to the end of an endoscope to correct for this mismatch during NOTES surgery. Through measuring the impact of gravity on each of the three orthogonal accelerometer axes, they rotate the camera visualisation to correct for the horizon. In their usability task, participants were asked to touch locations in a por-

cine abdomen using separately inserted needle holders. When comparing the performance of participants during corrected horizon visualisation and non-corrected, they report that participants were able to grasp the target locations quicker and more efficiently when viewing the corrected horizon camera visualisation. Their study however does not investigate the effects of using the horizon correction during navigation, *i.e.,* the endoscope camera visualisation was kept fixed during their study and did not move. The navigational challenges induced by looping can only occur during when the camera is continuously moving.

This chapter focuses on the challenges during looping and on gaining an understanding of the gaze behaviour during such instances. Loop formation occurs when the leading end of the endoscope remains stuck even though the surgeon pushes the endoscope further. In this situation, the endoscope, due to its long flexible tubular structure, can get twisted along the length of the colon, causing the orientation of the camera visualisation to become misaligned from the endoscope controls. This gives rise to a perceptual misalignment between the expected camera movement and actual resultant camera movement since the operator controls the endoscope tip from the proximal end of the endoscope. Compounding the dependent nature of the endoscope instrument, any subsequent movement of a looped endoscope tip can yield bending of the endoscope in a paradoxical manner inducing a varying perceptual-motor misalignment. The effect of this misalignment makes searching for lesions challenging, but more critically, there is immense difficulty in manipulating the controls on the endoscope during looping [14]. This misalignment which was also discussed in Chapter 3, is the perceptual-motor misalignment.

The variation in perceptual-motor misalignment can occur during manoeuvring flexible surgical instruments guided by a camera visualisation at the distal tip. This problem differs from the problem investigated by Holden and Flach [56], and Holler *et al.* [58, 57] as their camera visualisation is placed on a different frame of reference to that of the instruments. In addition to that, the camera view remains static. In contrast, perceptual-motor misalignment describes the challenge that arises during navigation, *i.e.,* a moving camera view, caused by applying torque to an endoscope along the gastrointestinal tract, where a change in the mapping of the motor control at the proximal end of the endoscope and its camera orientation at the

distal tip occurs. In the varying perceptual-motor misalignment circumstance, the camera and the instrument share the same frame of reference, and the perceptual challenge for the endoscopist is caused by the change in the mapping of the motor input to the camera visualisation movement. When such a misalignment of the endoscope occurs, proprioceptive adaptation is required by the surgeon to re-establish veridical perceptual response [124]. In such a state, there is a risk of over stretching the bowel causing discomfort and pain to the patient or worse, perforating the patient's colon. Thus it is clinically of strong interest to be able to detect such instances and to warn and assist the surgeon. If challenging instances of varying perceptual-motor misalignment can be inferred from the gaze information, this can then be applied during clinical training programmes or even during live surgery, without having to use extra hardware to measure the perceptual-motor misalignment on conventional endoscopes.

In Chapter 4 we discussed the challenges arising from a narrow FOV and poor understanding of the camera trajectory during endoscopy and introduced a grey-scale DVE method to improve visualisation during endoscopy. The work presented in this chapter investigates how varying perceptual-motor misalignment during endoscopy can affect the surgeon's performance and to what extent the visual behaviour can indicate when the subjects are disoriented. The idea is that if we can classify the varying perceptual-motor misalignment from the gaze behaviour, then we can apply this classification method to infer when varying perceptual-motor misalignment occurs from the gaze data only and without having to quantify the actual angle, which is very difficult to measure in conventional endoscopes since they do not conform to a predefined kinematic model. Through tracking the gaze data, detecting instances of varying perceptual-motor misalignment can then lead to provision of warning or assistance during such challenging navigational periods. In order to achieve this, a surgical simulation of an articulated hyper-redundant robot simulating a flexible endoscope (colonoscopy) procedure is implemented. A method to quantify the perceptual-motor misalignment is introduced and the performance of the subjects during the colonoscopy task is assessed during two different perceptual-motor misaligned control modes. In Section 5.1 the methods used to quantify the perceptual-motor misalignment, the surgical performance assessment metrics, and the visual search behaviour classification during misaligned instances are discussed. Section

5.2 describes the experimental setup, while in Section 5.3 the statistical analysis and classification of the gaze data obtained from the experiments are presented. The chapter finishes with a conclusion and discussion on the study conducted.

5.1 Methods

In order to conduct the colonoscopy user study for this chapter, an articulated robot structure based on Shang *et al.* [139] was re-created virtually using OpenGL to have a controlled and repeatable user study environment. The lightweight universal-joint-based flexible access virtual platform compromises of a series of link units forming a snake-like robot structure and can provide wide exploration capability of the operative site including retroflexion [77]. Each joint in the articulated robot is independently actuated, offering a hyper-redundant kinematic structure, thus providing flexibility whilst maintaining accurate control of the joints during navigation. A simulated environment using this flexible robot platform has benefits of offering the ability to record the robot's joint information and thus the variation in the perceptual-motor misalignment that occurs during the task. The hyper-redundant kinematic robot comprises of forty universal joints linked together and has both a virtual camera and light source mounted at the distal tip to provide adequate tissue visualisation during the user study. The relative pose of the camera visualisation and lighting is changed simultaneously with the articulated robot movement. Controlling the hyper-redundant, multiple-joint robot is done through obtaining the 3D coordinate input from a SensAble Phantom Omni. An illustration of the experiment setup is shown in Fig. 5.1. A Tobii gaze tracker is used to record subject gaze data, while participants performed the task. Recording of such gaze data enables correlation analysis of performance with gaze parameters.

During the colonoscopy task, only the last two distal joints of the hyper-redundant robot are controlled, which facilitates both navigation and quantification of the perceptual-motor misalignment angle. However, only a single joint can be controlled at a given time. Switching between the two joints is done by pressing a button on the haptic controller. The kinematic structure of the flexible hyper-redundant robot [139, 161] together with only being able to control one joint at a time gives rise to a peak-to-peak vary-



Figure 5.1 – Illustration of the user experiment setup for this study. The user control arrangements and the simulated articulated robot inside the colon are shown.

ing perceptual-motor misalignment of up to 60°. The use of this hyper-redundant kinematic robot allows simulation of the varying perceptual-motor misalignment conditions which can occur during flexible endoscopy. Details of the two different perceptual-motor misalignment conditions are discussed in Section 5.2.

In the subsequent subsections the following are described, i) the quantification method of the perceptual-motor misalignment angle, ii) the statistical analysis methods of the performance metric and gaze data and, iii) a classification algorithm to differentiate the instances when the user is misaligned or when the user is not.

5.1.1 Perceptual-Motor Misalignment Angle

The perceptual-motor misalignment angle is the perceived misalignment angle between the expected camera movement and actual resultant camera movement as the endoscopist moves the endoscope tip. In order to understand the method of quantifying the perceptual-motor misalignment angle, it is helpful to refer to the illustration of an endoscope shown in Fig. 5.2(a). Assume $\delta\mathbf{q}$ is an incremental motion of the endoscope that the surgeon moves the end-effector. Here $[\delta\mathbf{q}^x; 0]$ and $[0; \delta\mathbf{q}^y]$ represent the motor directions along the x- and y-axes in the instrument frame of reference plane.

The 3D displacement vector of the endoscope $\delta\mathbf{x}_{eff}$, caused by a motor input in the x-direction $\delta\mathbf{q}^x$, can be expressed as,

$$\delta\mathbf{x}_{eff}(\mathbf{q}, \delta\mathbf{q}^x) = \mathbf{x}_{eff}(\mathbf{q}) - \mathbf{x}_{eff}(\mathbf{q} + \delta\mathbf{q}^x). \tag{5.1}$$

By projecting this 3D vector to the camera visualisation plane, the incremental camera motion direction $\delta\mathbf{x}_{cam}$, can be expressed as,

$$\delta\mathbf{x}_{cam}(\mathbf{q}, \delta\mathbf{q}^x) = \mathbf{C}_z \times (\delta\mathbf{x}_{eff}(\mathbf{q}, \delta\mathbf{q}^x) \times \mathbf{C}_z) \tag{5.2}$$

where $\mathbf{C}_z \in \Re^3$ is a unit vector indicating the direction of the camera principal axis that is always perpendicular to the $\mathbf{C}_x \in \Re^3$ and $\mathbf{C}_y \in \Re^3$ unit vectors along the camera visualisation plane. To calculate the perceptual-motor misalignment introduced by the instantaneous configuration $\mathbf{q}$, the angle θ_x, relative to the x-coordinate of the camera imaging plane can be defined as:

$$\theta_x = cos^{-1}\left(\frac{\delta\mathbf{x}_{cam} \cdot \mathbf{C}_x}{\|\delta\mathbf{x}_{cam}\| \, \|\mathbf{C}_x\|}\right) \tag{5.3}$$

To determine the misalignment θ_y along the y-axis, $\delta\mathbf{q}^y$ and $\mathbf{C}_y$ become the input to Eq. (5.1) and Eq. (5.3) respectively, instead of $\delta\mathbf{q}^x$ and $\mathbf{C}_x$. For a looped endoscope, the nature of the endoscopic configuration $\mathbf{q}$, can alter the mapping between the expected motor control input $\delta\mathbf{q}$ by the operator and the resultant end-effector camera visualisation movement due to a misalignment angle θ_x and θ_y. This resulting alteration to the mapping measured by θ_x and θ_y is the perceptual-motor misalignment. During control of the endoscope, the overall camera pose is manipulated by the operator such that these perceptual-motor misalignment angles, θ_x and θ_y, will be time-varying functions based on the configuration $\mathbf{q}$.

The hyper-redundant articulated robot used in this study is illustrated in Fig. 5.2(b), where only the the last two joints (Joint 1 and Joint 2) are used during the navigation task of the study. For the articulated endoscope robot, the kinematic mapping from the n joint variables $\mathbf{q} \in \Re^n$ to the end-effector position $\mathbf{x}_{eff}$ relative to the world frame can be generalised as $\mathbf{x}_{eff}(\mathbf{q}) \in \Re^3$. To understand how misalignment can be introduced during a single-joint control mode of such an articulated endoscopic robotic instrument, let us assume that joint 1 is perfectly aligned to the camera plane x

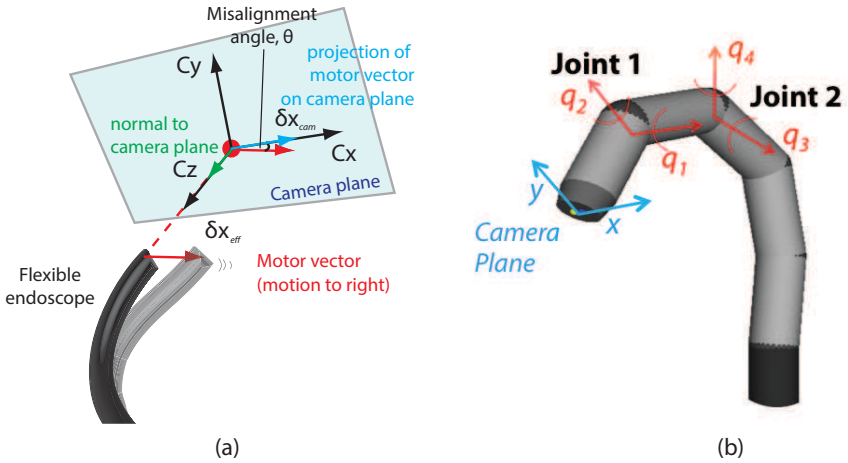


Figure 5.2 – Illustration of how the disparity between the camera image plane and the instrument end-effector position during motor coordination can lead to a misalignment angle between the surgeon's intended motor input and the resultant camera motion that is subsequently observed. (a) Perceptual-motor misalignment when controlling a flexible endoscope device. (b) Perceptual-motor misalignment when controlling the articulated robotic endoscope that is used in the study of this chapter.

and y. In this configuration, rotation of q_1 will move the camera in the y direction, or up and down. Rotation of q_2 will translate into movement in the x direction, or left and right. This shows that when controlling only Joint 1, no perceptual-motor misalignment is introduced. However, once Joint 1 has been moved, and the operator starts to control joint 2, *e.g.,* rotating q_3 via a motor input in the x direction, this no longer yields a movement of the camera plane in a purely x direction. Instead it will result in a misalignment between the operator's intended camera movement and the resulting camera movement at the end-effector when joint 2 is moved. In Chapter 3 it was discussed that looping occurs frequently at the sigmoid colon and the splenic flexure, where there are sharp bends in the colonic anatomy. At such instances of navigation, the participant would have to adjust both joints in the articulated robot, introducing a varying perceptual-motor misalignment. Hence, the simulation was deemed to be a close resemblance of the varying perceptual-motor misalignment challenges, which occur during looped instances in flexible endoscopic navigation.

5.1.2 Task Performance Metrics and Gaze Parameter Analysis

In order to understand how the perceptual-motor misalignment can affect task outcome, the following data were recorded:

- Time to complete the colonoscopy task
- Total endoscope tip travel distance
- The distance from the target lesion position relative to the camera visualisation position
- The angle from the target lesion position relative to the camera visualisation position

From the subject's task completion time, an indication of the ease or difficulty of each control mode can be assessed. Previous literature has shown that shorter instrument travel distance allows for more natural movement and less of a need for correction during motor control, and has been associated with better surgical performance [1]. By calculating both the robot

configuration and the polyp position with respect to the endoscopic camera view, performance related to polyp targeting accuracy can be measured. Performance measurements including the average distance from the polyp and the average angle of approach to the polyp, in the form of an error measurement, can indicate the intricate movements during each control scheme [86]. Navigation during colonoscopy requires dexterous movements and steady control of the endoscope tip. Participants are informed of their expectation to position the robotic endoscope in an optimal angle and distance from the polyps. As shown in Fig. 5.3(c), feedback of the distance and angle error are given through a yellow circle, whose radius and colour changes according to the distance and angle of approach error. Optimal positioning is indicated by a bright, large yellow circle.

Gaze data were recorded during the experiment to compute the participant's gaze parameters. By comparing a range of gaze parameters during the two different perceptual-motor misalignment control modes, *i.e.,* no perceptual-motor misalignment and varying perceptual-motor misalignment, the aim was to characterise gaze behaviour during normal navigation and looped conditions respectively. The gaze parameters extracted for gaze analysis were as follows:

- Gaze path length
- Time normalised gaze path length (cm per minute)
- Saccade amplitude
- Number of saccades
- Time normalised saccade count (saccades per minute)
- Number of fixations
- Time normalised fixation count (fixations per minute)
- Mean fixation duration

Statistical analysis was conducted on the above task performance and gaze parameter data. This involved conducting normality tests. Normality tests on the task performance data showed that the data followed a normal distribution, therefore ANOVA tests were applied during analysis. In contrast,

Mann-Whitney U test were applied to the gaze parameter data due to the non-normal distribution. For both types of statistical analysis tests, a 'p-value' <0.05 was considered significant.

5.1.3 Classification of Gaze Behaviour

Characterisation of the gaze behaviour associated with the time periods of varying perceptual-motor misalignment can potentially provide insight into the performance and the gaze signature associated with looping (disorientation). Here, variation of perceptual-motor misalignment greater than 30° will be identified and labelled using the method described in Section 5.1.1. The hypothesis is that relatively large changes of perceptual-motor misalignment angles translate to the perceptual and motor challenges that are experienced during looping and can alter the resultant gaze behaviour. In varying perceptual-motor misalignment, subjects are in a more challenging situation to navigate forward, and may reflect the disorientation experienced by the subject. During little variability in perceptual-motor misalignment, *i.e.,* less than 30°, it is assumed that navigation resembles the normal and less challenging periods.

In order to find relevant gaze parameter features that can help distinguish varying and no perceptual-motor misalignment, Mann-Whitney U statistical analysis tests comparing the gaze parameters obtained during varying and no perceptual-motor misalignment tasks are conducted as described previously in Section 5.1.2. The gaze parameters which show significant difference during varying and no perceptual-motor misalignment trials are used as features for a classification algorithm to identify instances of potential disorientation.

In this study, three gaze parameters were identified as relevant features in order to distinguish instances of varying perceptual-motor misalignment from those where there was no misalignment. These included, i) the time normalised saccade count, ii) the time normalised fixation count and iii) the mean fixation duration. All gaze parameters showed significant difference when comparing the gaze parameter data obtained during the varying and no misalignment control tasks (see Table. 5.2). Previous research reports that the fixation duration is correlated to the cognitive processing where longer fixations indicates greater processing [66, 67, 44]. It is expected that

subjects experienced difficulty when varying perceptual-motor misalignment occurred, needing greater focus on the procedure and therefore inducing longer fixations. The ability to identify looping and the subsequent associated disorientation instances are clinically useful for training as well as for providing in-vivo assistance during the procedure. During the colonoscopy task there are periods when the subject is performing polyp removal subtasks. Gaze parameter data during these periods were removed from the classification data as they are not related to navigational gaze behaviour.

In order to identify periods of varying misalignment, the data need to be divided along the temporal dimension using a time window. Depending on the length of the time window, the amount of available classification data are altered. The length of the time window is a trade off between including enough information to obtain good classification accuracy, whilst making it small enough so that the delay in detecting of such instances is not too long (semi-online) to make it useful for the potential application. Having a semi-online system will enable the provision of assistance at times of challenging misalignment conditions during the procedure, therefore is an important feature.

Since the aim of the classification is to separate instances of varying misalignment from the instances of no misalignment (normal gaze behaviour), a binary classification problem is formulated by separating the gaze data associated with the varying and no perceptual-motor misalignment into two classes. Each class's dataset has three dimensions according to the three gaze parameter features, *i.e.,* the time normalised fixations, the time normalised saccades and the mean fixation duration. In the field of machine learning, extensive research has been conducted on supervised learning techniques, *i.e.,* developing models from a training set of correctly identified observations. The choice of algorithm, depends mainly on the amount of available data and on the nature of the data, such as the dimensionality. SVM, a machine learning technique introduced by Cortes and Vapnik [27], is suitable to binary classification problems and is applied here to classify the instances of varying perceptual-motor misalignment from the normal gaze behaviour.

In principal, a SVM defines a classification function which corresponds to a separating hyperplane $f(x)$ that provides maximal margin between the data vectors of the two classes. Let the set of labelled training data be

$(y_1, x_1) \ldots (y_n, x_n)$, $y_i \{-1, 1\}$, with x_i being the vector of the features and $y_i = 1$ being a positive example, *i.e.,* varying misalignment instances, and $y_i = -1$ a negative example, *i.e.,* normal gaze behaviour. This set is linearly separable if there exist a vector $\bar{w}$ and a scalar b that satisfies the following formula: $y_i(\bar{w}x_i + b) > 1$, $i = 1 \ldots l$,which means that when $f(x) = \bar{w}\bar{x_i} + b$ is positive, $y_i = 1$ while when $f(x) = \bar{w}\bar{x_i} + b$ is negative $y_i = -1$. The optimal hyperplane is then defined as,

$$f(x) \quad = \quad \bar{w}\bar{x_i} + b = 0. \tag{5.4}$$

It separates the data with a maximal margin, as it determines the direction $\frac{\bar{w}}{\|\bar{w}\|}$ where the distance between the projections of the training vectors x_i of two different classes is maximal. In order to construct such optimal hyperplanes one only has to take into account a small amount of the training data, the so called support vectors, for which $y_i(\bar{w}x_i + b) = 1$. These vectors are the data which fall on the margin of the separating hyperplane and they essentially are enough to define the margin between the two classes.

In case the classes are not perfectly separable, a set of slack variables ξ_k is introduced to allow for class overlap, which results in $y_i(\bar{w}x_i + b) > 1 - k$, $i = 1 \ldots l$. This means that a sample can be on the wrong side of the hyperplane by the small value ξ_k. A new optimisation function which takes into account the slack variables is formed and separates the training set with a minimal number of errors.

The above formulations can solve the simple linear case in which the data are separated by a hyperplane. In order to allow for the SVMs to form a non-linear boundary between the two classes, a kernel function is used. A kernel function, when applied on the feature data, projects the features in a high dimensional space. In this new space, a linear function can be found to make the two classes separable. This means that the problem is solved by running the same hyperplane search algorithm on the newly projected data points rather than on the original data sets, which were not linearly separable. The choice of the kernel depends on the dataset and it is normally chosen by using a validation technique and choosing the function that gives better accuracy. A commonly used kernel is the Gaussian kernel, which is

defined as,

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right). \tag{5.5}$$

In the case of the Gaussian kernel, the parameter σ needs to be defined. If σ is large, then the Gaussian tends to fall relatively slowly, which results in higher bias and low variance, while with a small σ, the reverse happens. In our analysis, a 10-cross-fold validation is used to define which kernel performs better and the respective parameters of the kernel. The same validation method is used to define the resulting classification accuracy of the SVMs. The analysis of the classification performance of the SVMs and the details of the final used parameters are described in Section 5.3.2.

5.2 Experimental Setup

The simulation colonoscopy task using the hyper-redundant, multiple-joint robot is controlled by obtaining the 3D coordinate input from a SensAble Phantom Omni, as described in Section 5.1. Subjects were asked to perform a colonoscopy which consisted of navigating along the colon, searching for polyps and removing them until reaching the end of the colon. Such simulated tasks provide sufficient information to analyse the efficiency and the performance of the subjects conducting the task. In order to assess whether varying perceptual-motor misalignment increases burden to the operator, two control modes were used. These two control modes accommodated for different conditions during navigation and enabled the analysis of the respective performance metrics and gaze parameters obtained during these control modes. These two control modes are:

1. No perceptual-motor misalignment control mode

   Navigation during no perceptual motor misalignment involves using a control scheme capable of using only a 3-DoF control input. During this control scheme, any misalignment between the visual and motor axes is compensated from the articulated robotic control scheme by re-aligning the motor control with the visual coordinates. When using this control scheme, a left, right, up and down motor input would translate into a left, right, up and down endoscope camera movement

respectively.

2. Varying perceptual-motor misalignment control mode

   During the varying misalignment control scheme, the last two distal joints of the hyper-redundant robot are controlled. However, only a single joint can be controlled at a given time. Switching between the two joints can be conducted by pressing one of the buttons on the haptic controller. The kinematic structure of the flexible hyper-redundant robot [139, 161], together with only being able to control one joint at a time gives rise to a peak-to-peak varying misalignment of up to 60°. Hence, a right motor input can effectively translate into an endoscope camera movement with varying misalignment angle of up to 60° with respect to the motor input. At sharp colonic bends, which is where looping commonly occurs in commonly used flexible endoscopes, the participant would have to make control adjustments to both of the articulated robot joints, introducing a varying perceptual-motor misalignment. Hence, the control scheme was deemed to be transferable to the varying perceptual-motor misalignment problems which occur during looped instances of current flexible endoscope navigation.

Each participant was asked to traverse the colon using the above control modes in a randomised order; the no perceptual-motor misalignment control mode, and varying perceptual-motor misalignment angle control mode.

The Phantom Omni workspace is not sufficient to allow navigation of the entire surgical workspace. Therefore, in order to ensure that the entire workspace in the surgical simulation was navigational with the Phantom Omni, a clutch button that disables the camera screen movement with respect to any Phantom Omni motor input is introduced. The clutch mechanism allows the user to reposition the Phantom Omni stick while not affecting the location of the endoscope tip. The mapping from the planar input motion of Omni to the robot motion in joint space can be expressed as $H(\delta x, \delta y) = \delta \mathbf{q}$. Provided with proper alignment between the visual and motor axes, *e.g.,* when the user moves the Omni purely along the x-axis, the misalignment θ_x in Eq. (5.3) will always be minimised to zero such that $\delta x_{cam}(q, H(\delta x, 0)) = \alpha C_x : \alpha \in \Re^+$.

A Tobii eye-tracker was used to capture gaze information which was then analysed offline. Based on this platform, the participants were instructed to

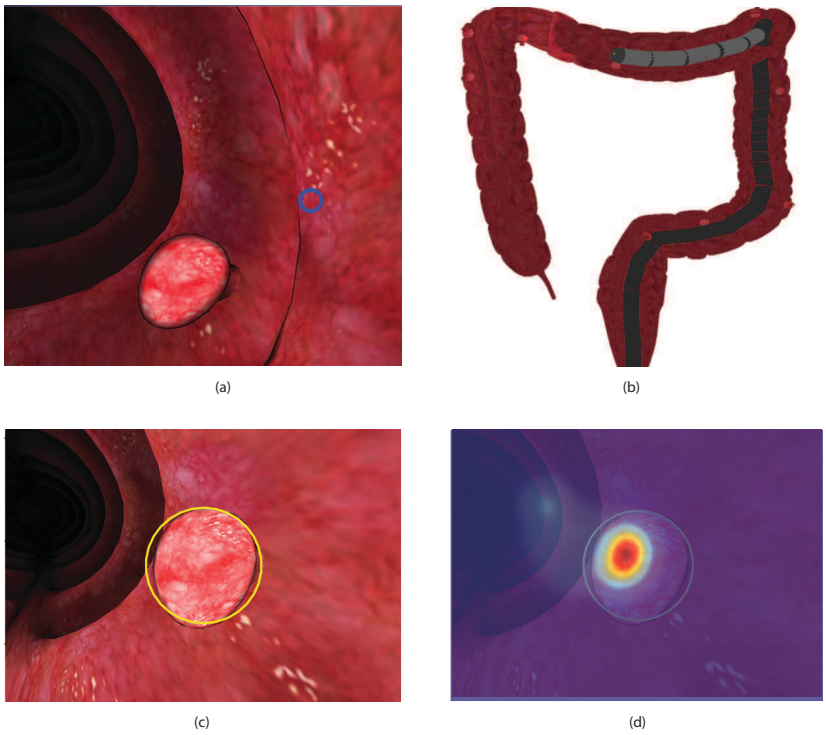


Figure 5.3 – Illustration of the colonoscopy task. (a) The camera view from the distal-tip shows the gastrointestinal tract task environment. (b) Illustration of the external view of how the articulated robotic endoscope progresses through the colon. (c) The polyp removal scene showing optimal positioning indicated by a large yellow circle. (d) Participants' eye tracking gaze overlaid as a heat-map during the polyp removal scene. Note the focus of participants' interest on the polyp and towards the lumen.

locate 10 different polyps placed throughout the colon. An illustration of the experiment environment is shown in Fig. 5.3. Fig. 5.3(a) shows the camera visualisation of the colon during the task while Fig. 5.3(b) illustrates the external view of how the articular robotic endoscope progresses through the colon. Fig. 5.3(c) shows a polyp removal scene. Here subject participants are given feedback of distance and angle through a yellow circle, where optimal positioning is indicated by a bright, large yellow circle. Fig. 5.3(d) illustrates the fixation density heat-map as the subject removes a polyp.

Instruction was given to participants that accuracy and safety should be given priority over speed, just as with real surgery. Training was given

to each participant for both control modalities before gaze data were recorded. Coaching was provided during training to help them manoeuvre through the sigmoid colon and past the splenic flexure. During training, participants could visualise both the forward camera view of the endoscope and an external orientation view of the robot as illustrated in Fig. 5.3(a)-(b) respectively. However, during the actual trial, only the endoscope forward camera view could be seen.

A total of twenty-four subjects volunteered as participants. Of the participants who took part in the experiment, the age ranged from 18-36 with the background of the participants categorised as surgical or non-surgical. Due to the novel nature of the robotic articulated device, level of surgical experience was deemed irrelevant for the experiment. Data could not be used from five participants due to incompleteness/failure to track gaze data, leaving nineteen subjects for subsequent analysis. In order to perform a detailed analysis, several task performance metrics and gaze parameters were recorded as discussed in Section 5.1.2. The results of statistical and classification analysis are presented in the following section.

5.3 Results

A summary of the group's task performance metrics and gaze parameters together with statistical analysis for the two control modes are summarised in Table. 5.1 and Table. 5.2 respectively. Results with significant differences are indicated with an asterisk '*' mark in all tables. Results are represented as mean and standard deviation for the ANOVA statistical analysis tests on the performance metrics and in median and Inter-Quartile Ranges (IQR) in parentheses for the Mann-Whitney U statistical analysis tests on gaze parameters. All statistical analysis was performed using MATLAB.

From the statistical tests, the features which showed significant difference in the gaze parameter data comparison tests during the two different perceptual-motor control modes were used as the gaze features for SVM classification purposes. The training and test data to be then used for the SVM classification was the gaze data collected during the varying perceptual-motor misalignment control trials only. The aim of the classification was to distinguish relatively large varying perceptual-motor misalignment in a given time window from small variations in perceptual-motor misalignment

Table 5.1 – Average performance metrics of the group with respective standard deviation values and ANOVA test p-values.

	Task Efficiency		Targeting Error	
	Task Time (s)	Travel Distance (mm)	Distance Error (mm)	Angle Error (°)
(1) No Misalignment	187.8 ± 39.1	69.1 ± 41.6	1.7 ± 0.8	4.7 ± 3.2
(2) Varying Misalignment	273.4 ± 61.5	210.6 ± 101.3	2.1 ± 0.8	6.5 ± 4.0
p-value (1) vs (2)	**<0.001***	**<0.001***	**0.011***	**0.043***

throughout the control scheme.

5.3.1 Statistical Analysis

Detailed analysis was carried out to compute instrument manipulation and gaze parameters as described in Section 5.1.2. Statistical tests were conducted on subject group data for each of the parameters previously discussed. Results are summarised in Table. 5.1 and Table. 5.2 respectively. As can be observed from Table. 5.1, the overall task time and travel distance was significantly longer during the trials using the varying perceptual-motor misalignment control scheme (273.4s vs 187.8s; p<0.001) and (210.6mm vs 69.1mm; p<0.001) respectively. The group also struggled to control the endoscope tip during finer manipulations during the use of the varying perceptual-motor misalignment control scheme, reflected in significantly larger target-polyp distance (2.1mm vs 1.7mm; p=0.011) and angle approach error values (6.5° vs 4.7°; p=0.043). Fig. 5.4 demonstrates the performance difficulty of the varying perceptual-motor misalignment control over the no misalignment control. The no perceptual-motor misalignment control shows improvements over all performance indices, especially for the endoscope travel distance. It is clear that the group found using the varying perceptual-motor control scheme more challenging, as proprioceptive adaptation would be required when having large varying misalignment [124], overall increasing the workload which is reflected in the participant's longer instrument travel distance.

The gaze path length was significantly longer during varying perceptual-motor misalignment when compared against the no perceptual-motor misalignment control scheme (660.5cm vs 401.7cm; p=0.007), however once normalised for the duration of the trial, (normalised gaze path length) this

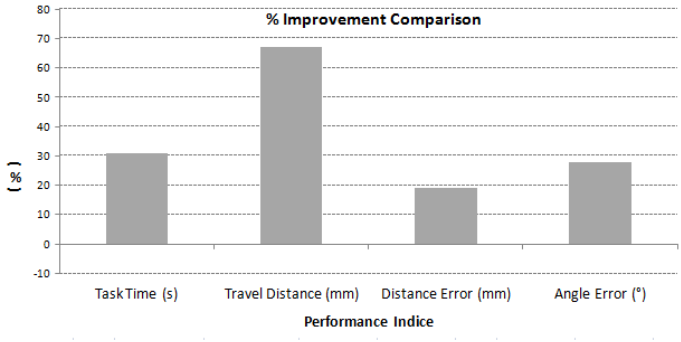


Figure 5.4 – Performance improvement of having no perceptual-motor misalignment compared to having varying perceptual-motor misalignment during navigation for each performance metric.

metric no longer showed any significant difference between the two control modalities (183.2cm per min vs 191.7cm per min; p=0.414). This indicates that the gaze path length is a reflection of the task time which resulted to be significantly different in Table. 5.2. The saccade amplitude did not result in any significant difference between the two control modalities. Interestingly, the number of saccades and the time normalised saccade count showed significant difference between the two control modalities (116.0 vs 69.0; p=0.002) and (30.0 saccades per min vs 38.1 saccades per min; p=0.047). The results suggest that participants performed fewer saccades during the varying perceptual-motor misalignment control mode. The number of fixations and the time normalised fixation count also showed significant difference between the two control modes (238.0 vs 129.0; p<0.001) and (61.9 fixations per min vs 71.7 fixations per min; p=0.001). In addition, the mean fixation duration was also significantly longer during the varying perceptual-motor misalignment control scheme (490.6ms vs 455.4ms; p=0.004) showing that participants performed fewer fixations of longer duration during the varying perceptual-motor controls scheme. This observation is consistent with previous literature, where it has been shown that increased cognitive load results in longer fixation durations [66, 67, 44]. When there is varying perceptual-motor misalignment, the results suggest that the group had a more focused visual attention strategy with fewer but longer fixations, resulting in performing fewer saccades. This could also suggest that during varying perceptual-motor misalignment, the group of participants were less

Figure 5.5 – Time normalised saccade count by control mode aggregating all participants.

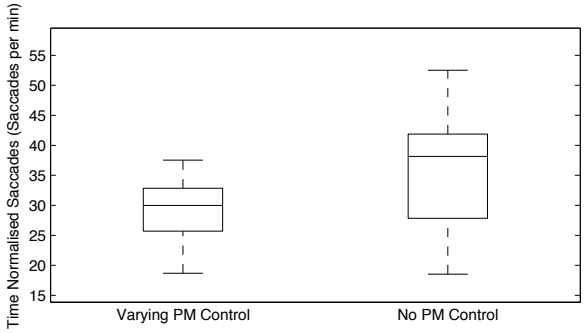
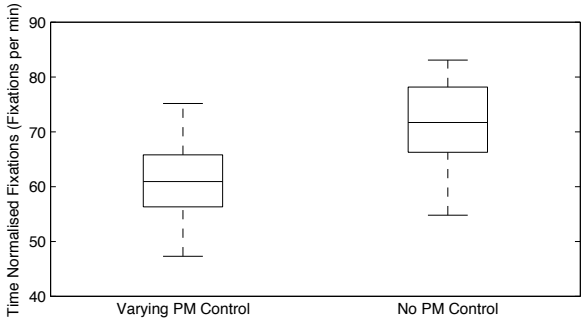


Figure 5.6 – Time normalised fixation count by control mode aggregating all participants.



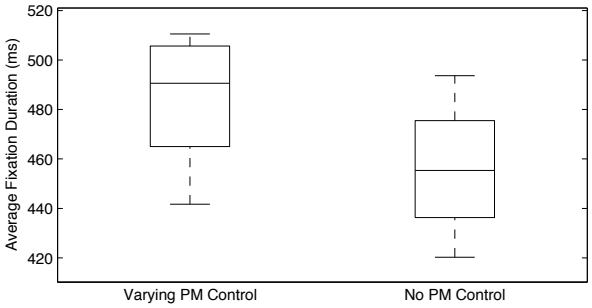
capable of being attentive to their surroundings, potentially affecting their situational awareness. Overall, the resulting gaze parameters are likely to be a reflection on the more challenging nature of the navigational task.

Aggregated data of the gaze parameters which resulted in significant differences when comparing the gaze data recorded during varying and no perceptual-motor misalignment control trials are summarised in Fig. 5.5 (time normalised saccade count), Fig. 5.6 (time normalised fixation count) and Fig. 5.7 (the mean fixation duration). These three gaze parameters are to be used as features to classify the SVM to help distinguish varying and no perceptual-motor misalignment, which may help to identify instances of potential disorientation.

Table 5.2 – Median and interquartile range values of gaze parameters of the nineteen participants with respective Mann-Whitney U test p-values.

	Gaze Path Length (cm)	Normalised Gaze Path Length (cm per min)	Saccade Amplitude (°)	Number of Saccades	Time Normalised Saccade Count (Sacc. per min)	Number of Fixations	Time Normalised Fixation Count (Fix. per min)	Mean Fixation Duration (ms)
(1) No Misalignment	401.7 [235.0]	191.7 [92.6]	5.7 [2.2]	69.0 [31.5]	38.1 [13.2]	129.0 [44.5]	71.7 [11.2]	455.4 [37.5]
(2) Varying Misalignment	660.5 [627.4]	183.2 [104.8]	6.2 [1.5]	116.0 [49.5]	30.0 [7.0]	238.0 [81.0]	61.9 [9.0]	490.6 [39.6]
p-value (1) vs (2)	**0.007***	0.414	0.759	**0.002***	**0.047***	**<0.001***	**0.001***	**0.004***

Figure 5.7 – Average fixation duration by control mode aggregating all participants.



5.3.2 Classification of Instances of Variable Perceptual-Motor Misalignment

The statistical analysis in the previous section showed significant difference in the time normalised saccade count, the time normalised fixation count and the mean fixation duration. These three features were used for the SVM classification to classify the instances of misalignment versus the instances of normal navigation. The data separation time window used was 30 seconds, which was chosen to be able to potentially use the trained SVM to infer with reasonable temporal resolution whether the surgeon is experiencing a varying perceptual-motor misalignment and therefore is potentially disorientated. SVMs are not computationally demanding and can therefore classify a data set in real time.

The 3D-dataset of the saccade count, fixation count and duration of fixations, is not linearly separable (as observable in Fig. 5.8), and therefore a kernel needed to be used. In order to find an appropriate kernel, three kernels were tested for by conducting a 10-cross fold validation. The kernels include the quadratic, polynomial (third order) and Gaussian kernel. During the cross-fold validation phase, each kernel was tested and the SVM parameters were fine-tuned. The boundary defines by how much a data point can lie in the wrong class. From the 10-cross-fold validation, the Gaussian kernel with a σ value equal to 2, gave the best recognition accuracy of 89.8%. The result of the Gaussian kernel SVM classification is illustrated in Fig. 5.8. The data instances of varying perceptual-motor misalignment

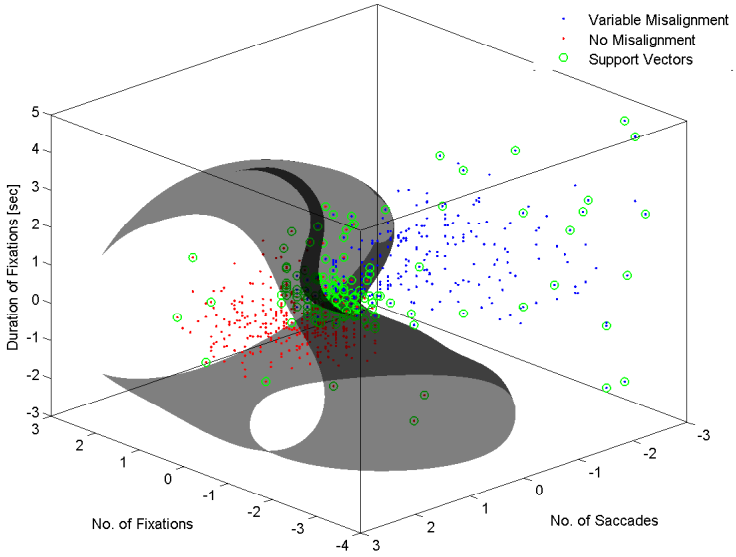


Figure 5.8 – Illustration of the separation hyperplane, defined from the SVM, between the instances of variable perceptual-motor misalignment (blue) and the instances of no misalignment (red). The support vectors on the boundary of the classification is depicted with green. It is observed that fewer saccades and fixations with longer fixation durations are linked to varying perceptual-motor misalignment.

are illustrated with blue and the red points indicate the data during little amount of perceptual-motor misalignment. The support vectors, which are shown by a green circle, are the data that define the margin between the two classes. It can be observed by the distribution of the data that this is a challenging classification problem and that it is only possible to get a good classification accuracy through the use of a kernel.

The results of the classification analysis, indicate that during varying perceptual-motor misalignment, fewer fixations of longer durations combined with fewer saccades indicate that the subjects are experiencing difficulty during navigation. The coinciding of the three gaze features in such a manner provide a strong indication of varying perceptual-motor misalignment, struggle and potential disorientation which can be used as a way to

assess trainee surgeons or to distinguish and provide surgeons with some in-vivo assistance. The SVM classifier could be applied during conventional flexible endoscopy or during robotic procedures without having to measure the perceptual-motor misalignment, as instances of varying perceptual-motor misalignment can be inferred from the tracked gaze feature data. Having said the above, there are a number of limitations to the analysis, first and foremost is the assumption that perceptual-motor misalignment leads to disorientation. Although the analysis of performance metrics show that varying perceptual-motor misalignment control of an endoscope is more challenging, it is still unknown how long it takes to become disorientated and how to repeatedly induce a person to become disoriented. Furthermore, the SVM classification method would only be able to infer varying perceptual-motor every half a minute. When the time window was reduced further, the classification failed to distinguish the varying perceptual-motor instances due to the lack of available data within the time window. The semi-online classification method cannot provide instant inference of a varying perceptual-motor misalignment from gaze data.

5.4 Conclusions

In this chapter, through re-creating an articulated robotic endoscope in a simulated environment we have investigated how a variable perceptual-motor misalignment may affect task performance. Participants were asked to traverse the colon and perform a colonoscopy task using two control modes, one which is capable of producing varying perceptual-motor misalignment and one with no misalignment. The articulated robotic kinematics were used as an approximation to the perceptual-motor misalignment that can be induced during looping of flexible endoscopes. Task efficiency and instrument targeting accuracy metrics were used to assess how performance differed when using a varying-perceptual motor misalignment scheme to when no perceptual-motor misalignment was present during navigation. ANOVA tests were carried out on the performance metrics which showed that during the varying perceptual-motor misalignment control scheme, the task time and travel distance was significantly longer and the group of participants also found it more challenging to control the endoscope tip during finer manipulations resulting in significantly larger target-polyp dis-

tance and angle approach error values. The Mann-Whitney U statistical tests when comparing gaze parameter data obtained during the trials using the varying perceptual-motor misalignment control scheme against the no perceptual-motor misalignment control scheme resulted in a significant difference between the time normalised saccade and fixation count as well as the mean fixation duration. The gaze features characteristic during varying perceptual-motor misalignment resulted in fewer but longer fixations with less saccades.

Classifying the visual search behaviour during varying perceptual-motor misalignment was also performed using SVMs, a binary classification method, to differentiate instances of large variation in perceptual-motor misalignment from those with little to no perceptual-motor misalignment. Using a Gaussian kernel adjustment, a recognition accuracy of 89.2% was achieved, enabling inference of varying perceptual-motor misalignment from the gaze data alone. Application of the classification scheme can allow inference of the instances with varying perceptual-motor misalignment during an endoscopy without having to measure the perceptual-motor misalignment angle of conventional endoscopes. This can be useful as it is difficult to measure the perceptual-motor misalignment angle in conventional flexible endoscopes as the flexible nature of endoscopes gives it no predefined kinematic structure, and more importantly it provides a low cost means to infer challenging instances caused by varying perceptual-motor misalignment through the user gaze behaviour only. A short fall of the gaze behaviour classification is that it may only indicate the cognitive and motor challenge experienced by the participants, rather than capture the gaze behaviour during perceptual-motor misalignment. Nevertheless, it can be expected that a perceptual-motor misalignment would increase the cognitive burden to the participant, which should be captured in the classification conducted in this chapter. The implications of our findings provide an important quantitative basis for assessing the ergonomics, as well as perceptual behaviour and potential cognitive loadings of future flexible devices, including those with robotic assistance for emerging surgical techniques such as single port access or natural orifice MIS procedures. It also highlights the potential usefulness of articulated robotic endoscopes, since robotic control has the ability to continuously re-align any perceptual-motor misalignment. The use of an articulated endoscopic robot with hyper kinematic redundancy facilitates

adequate surgical exploration [78, 139] and can allow for continuous alignment of the motor controls to the visualisation axis through realignment of the physical motor control with the camera visualisation. Therefore future use of fully articulated flexible robots to replace the traditional colonoscopy or any endoscopic procedure represents a unique opportunity of automating certain aspects of the navigation process with improved consistency and safety as there is potential to address the navigational challenges during looping. In the next chapter, a gaze contingent camera interface is introduced where the camera movement is continuously aligned to the user input control, allowing for improved camera navigation and ergonomics.

6 Gaze Contingent Laparoscope for Surgical Navigation

Surgical navigation can pose a number of challenges to the surgeon. The need to manoeuvre whilst maintaining orientation requires dexterous hand skills, which combined with the small camera workspace and narrow FOV can make the surgical procedure more difficult. This is universally true whether using endoscopes or laparoscopes. During laparoscopic surgery, the surgeon requires the bi-manual manipulation of surgical instruments, therefore a camera assistant is usually present to hold and navigate the laparoscope for the operating surgeon. This arrangement is not without its challenges. The narrow FOV of laparoscopic cameras requires the assistant to frequently move the laparoscope in order to visualize the full surgical workspace during the MIS procedure. Understanding the surgeon's desired camera orientation and visualisation via verbal communication is not always straightforward and failure to provide good visualisation can not only induce greater mental workload on the surgeon [183], but also lead to unrecognised collateral injuries [105]. The need for good camera handling has been recognised as an important step in the training curricula for surgical residents [141]. In spite of such training, issues related to assistant fatigue, hand tremor and the confined shared workspace for the surgeon and camera assistant persist. There is also man-power and cost implications of requiring a highly skilled camera assistant, which if addressed, could allow a surgeon to operate solo, thereby providing cost and staffing efficiencies.

In order to address the above deficiencies, a number of commercial robotic assisted camera systems have been developed. These include the EndoAssist [71] which is controlled by the user's head-mounted infrared emitter, the verbally controlled Automatic Endoscope Optimal Position (AESOP) system [72], and the finger stick controlled SoloAssist from AktorMed [145]. Additionally, there have been a number of gaze contingent controlled ro-

botic assisted camera systems [110, 146]. Existing gaze contingent control systems move the camera by following the user's gaze position on the visualisation screen. However, they have only had a panning control of the laparoscope, failing to provide the navigational functionality during surgery. Furthermore, these systems are activated by a foot-pedal and controlled by dwell-time on fixed regions of the screen to convey the panning direction [110]. Unfortunately, introducing more hardware such as foot-pedals can lead to instrument clutter in an already complex environment and existing dwell-time methods to control the camera require use of large amounts of screen coverage, where further screen area would be needed to introduce additional functionality such as zoom control. An alternative method to detect user intention during camera control is blink detection, which was previously used in gaze typing [46]. However, requiring the surgeon to close their eyes for a fixed amount of time could be dangerous for the patient.

In order to overcome these problems, this chapter introduces a gaze contingent robotic camera control system, where the camera is activated via real-time gaze gestures rather than an external switch. Through the use of multiple gestures, which map to specific camera control commands, it becomes possible to initialise different camera control modes such as panning and zooming of the camera. The work presented here builds on the 'Perceptual Docking' paradigm introduced by Yang *et al.* [175]. The main challenge in using gaze contingent control is the difficulty to design an intuitive control interface that enables the user to convey their intention without being affected by aberrant or idiosyncratic saccadic eye movements.

Gaze gestures can be single-stroke [38, 91] or multi-stroke based [130]. The use of gaze gestures have previously been applied to eye typing [171], Human Computer Interaction, mobile phone interaction [129] and gaming [63]. Here, HMM based gaze gesture recognition is applied to learn multiple input commands of a surgeon in order to convey their desired camera control mode. This hands-free activation and camera control method empowers the surgeon to use both their hands to operate whilst eliminating the need for a camera assistant. With laparoscopy, the surgeon's visual control and feedback is removed. In the proposed system, the important role of the surgeon's eyes, (*i.e.*, localising, tracking and orienting) are restored back into the 'control loop' of the camera. In the following section, the HMM based gaze gesture training and recognition algorithm are explained in Section

6.1.1, followed up by the robotic control in Section 6.1.2. In Section 6.2, the experimental setup, the system user interface (UI) and the experimental protocol are delineated. Section 6.3 discusses the results and the chapter finishes with a discussion and conclusion.

6.1 Methods

In order to enable hands-free camera activation for gaze contingent camera control, as illustrated in Fig. 6.1, one solution is to design gaze-based gestures. Through the use of gaze gestures, the surgeon is able to convey their intention without the need of activating a physical switch. This allows them to use both hands to control their instruments with improved workspace.

Prior to using the gaze contingent camera system, a subject-specific eye calibration is required to accurately estimate the user's PoR. This is typically done offline. Afterwards, the remote centre of motion (RCM) for the laparoscope is set, which is then inserted into the phantom, and the surgeon can promptly start using the robotic laparoscopic system. In order to move the camera, one of the two possible gaze gestures is performed: '*pan*' or '*zoom*'. The '*pan*' gaze gesture which is illustrated in Fig. 6.2(a) is defined by the following three-stroke sequence of eye movements: gaze at the centre of the screen, then to the bottom right corner, back to the centre, and finally back to the bottom right corner. Conversely, the '*zoom*' gaze gesture, shown in Fig. 6.2(b), is similar to the '*pan*' gaze gesture but instead, the user is required to look at the bottom left corner of the screen. Since the human eye in its natural form is used as an information gathering function [174] rather than to convey intention to control external devices, the challenge is distinguishing natural gaze patterns from intentional gaze gestures. To this end, pattern recognition methods are necessary to learn these gaze gestures. Furthermore, gaze gestures towards the corner of the screen are chosen to prevent obstruction of the camera view, minimise occupying screen real-estate and to also reduce detection of involuntary gaze gestures. Once a gaze gesture is captured, the camera will activate and the user can control the robotic arm as the surgeon's PoR is closed in the feedback loop as described later in Section 6.1.2. In order to stop the robot arm from moving, the user looks into the white circle centre of the screen for $750ms$.

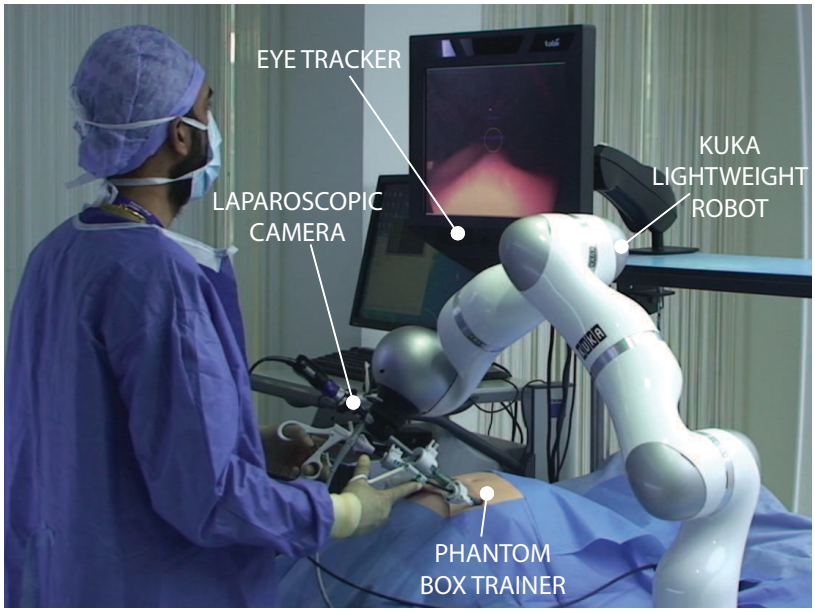


Figure 6.1 – A surgeon using the gaze contingent camera system. The operator is able to simultaneously navigate and perform a bi-manual surgical task.

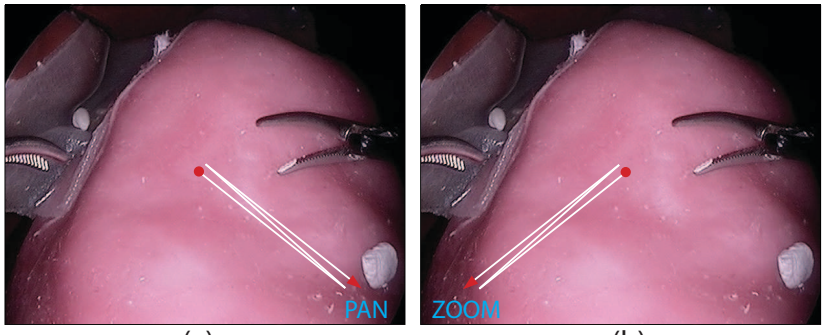


Figure 6.2 – Illustration of the (a) '*pan*' and (b) '*zoom*' gaze gestures for the gaze contingent laparoscope system.

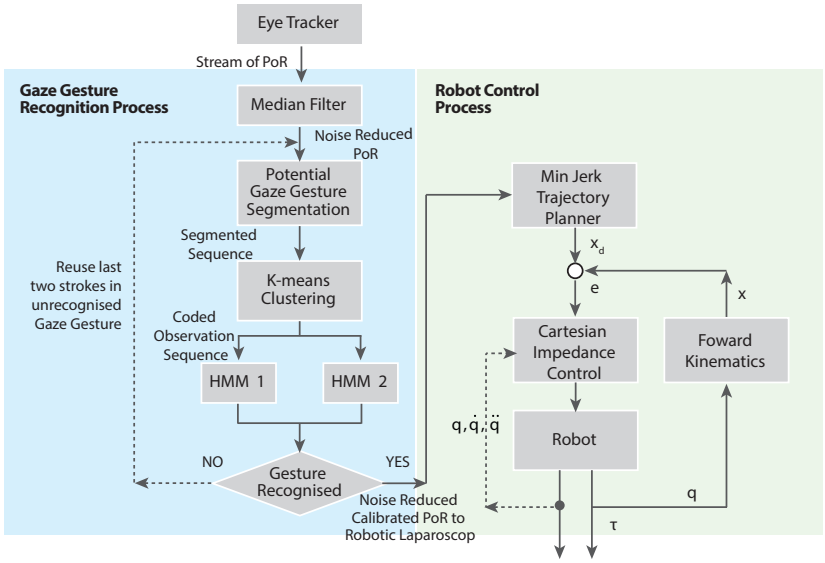


Figure 6.3 – The gaze contingent system's internal control mechanism can be broken down into two processes: the gaze gesture recognition process and the robot control process.

The gaze contingent system's internal control mechanism is illustrated in Fig. 6.3. The system can be separated into two processes: the gaze gesture recognition process and the robot control process. The gaze gesture recognition process is the user control interface to activate controlling of the robot arm. The robot control process enables moving of the laparoscope mounted on the robot arm. The following subsections will go into further detail explaining the two processes.

6.1.1 Gaze Gesture Recognition Process

The gaze gesture recognition process designed for the gaze contingent laparoscope system is illustrated in Fig. 6.4. The PoR gaze output is first passed through a median filter with a $150ms$ time window followed by gaze gesture segmentation. Since the nature of the gaze gestures used in the system requires the user to produce three sequential fast eye movements, *i.e.*, saccades, the gesture segmentation algorithm was designed to detect three saccades. These saccades are detected by searching for two sequential gaze data samples which exceed a $300°s^{-1}$ velocity threshold. A ve-

locity threshold saccade detection technique [132] was implemented with additional temporal constraints to ensure that the gaze gestures were not excessively long (each gaze stroke was required to be less than $750ms$) overall allowing segmentation of continuous and user-intended gaze gestures. Furthermore, any segmented gaze gesture could potentially contain part of an intended gaze gesture (*i.e.,* one third or two thirds of the intended three-stroke gaze gesture). Therefore, the last two strokes of any segmented potential gaze gesture are stored in memory where any subsequent incoming gaze stroke would be added onto the sequence. Note that if a gaze gesture is recognised, this memory would be cleared and the algorithm would start its search again for a three-stroke gaze gesture sequence.

Relevant features from the segmented trajectories of a gaze gesture are then extracted via a k-means clustering algorithm. Each cluster's symbol number, centroid coordinates and radius are used collectively to create a discrete codebook that captures the relevant features of the gaze gestures. The codebook was designed offline from 300 gaze gesture training data sequences and five clusters were chosen for the codebook. Each potential gaze gesture sequence is encoded using the codebook where symbol numbers are assigned to each observation by using the distance between the observation and the centroid of each cluster provided that it is within the defined radius. If an observation is outside the feature space, it is discarded.

In order to recognise the segmented potential gaze gestures, two '*left-to-right*' HMMs [122] were used for each camera control activation mode. The '*pan*' gaze gesture is modelled by HMM1 and enables panning control of the laparoscope (user interface detailed more in Section 6.2.1). The '*zoom*' gaze gesture is modelled by HMM2 and enables zooming in and out. A single HMM represents a single gesture and as many HMMs can be added as necessary depending on the number of gestures that need to be incorporated into a given system. For the system design presented here, two gestures need to be utilised.

HMMs are stochastic sequences where the hidden states are associated with a probability [122] and is typically represented by a set of interconnected N states $S = [S_1, ..., S_N]$, and a number of k discrete symbols $V = [v_1, ..., v_k]$. The model is described by an observation sequence $O = [O_1, ..., O_M]$, a state transition matrix $E = [e_{ij}]$, with transition probabilities from state i to state j as $e_{ij} = P(q_{t+1} = S_j | q_t = S_t)$, where

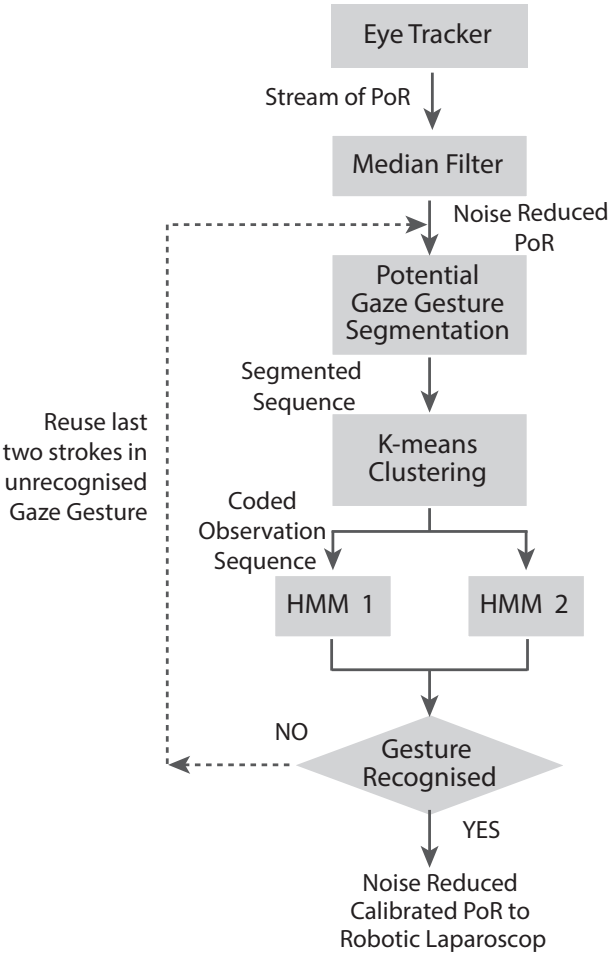


Figure 6.4 – Flow chart of the gaze gesture recognition process. The stream of the user's raw PoR's are filtered from which the gaze gestures are segmented and then encoded. This encoded observation sequence is tested for inference of both gaze gesture HMMs. If one of the two possible gaze gestures is recognised, then the camera is activated and PoR information is sent to the robotic arm.

$1 \leq i,\ j \leq N$, and q_t is the state at time t, and the emission probability matrix $F = [f_{jk}]$, where f_{jk} is the probability of generating the symbol v_k from the state q_j, with $f_{jk} = P(v_k \text{ at } t | q_t = S_j)$, $1 \leq j \leq N$ and $1 \leq k \leq M$. The initial state probability distribution is denoted by $P_i = p_j$, where $j = 1, 2, \ldots, N$.

Prior to being able to recognise the segmented gaze gestures, each of the HMMs model parameters were trained offline using a set of gaze gesture training data. After obtaining the model parameters, the forward-backward algorithm is used to obtain the probability of the encoded gaze gesture sequence given the respective trained HMM. The recognised gesture is the one with the maximum inference value from the HMM, given that it is above the inference value threshold defined during the training. Once one of the gaze gestures are recognised, the noise reduced PoR is sent to the robotic arm in order to control it (described further in Chapter 6.1.2), otherwise no input is given to the robotic arm.

HMM Training To enable the gaze contingent system to recognise the two gaze gestures, these two HMMs' model parameters needed to be obtained offline using training data. One of the challenges in a gaze gesture recognition system is to enable the system to recognize the different intended gestures accurately (*i.e.,* high recognition accuracy) and discriminate against natural visual search activity (*i.e.,* low false positive rate). Hence, both intentional and unintentional gaze gesture data were used to define the model parameters. 300 intentional gaze gesture sequences for the two gaze gestures were collected from ten participants. From the same ten subjects, unintentional gaze gesture data were collected during a five minute web browsing task. Each of the 300 training sequences was encoded using the formulated k-means clustering codebook. Within this set of training data observations O, an initial state probability P, are defined and then the optimal state transition and emission probabilities that describe the set of training observations are iteratively obtained by using the Baum-Welch algorithm [122]. To obtain a good trade-off between recognition accuracy, a low false positive rate, and overall complexity of the system, a 10-fold cross validation was run across the HMM with different number of states (between 2 and 8 states, total of seven runs). 90% of the encoded training sequences are used with the Baum-Welch algorithm to iteratively obtain the

HMMs parameters. With each of the seven training instances, a detection probability threshold was set at a 95% confidence limit of the training data sequences' inference values obtained from the trained HMM. The recognition accuracy of the HMM is then defined by using this threshold on the remaining 10%, the validation data set. Fig. 6.5(a)-(b) illustrates the recognition accuracy and false positive rate of the system across the increasing number of states used in the HMM for the '*pan*' and '*zoom*' gaze gesture respectively. It was decided that 6 states were to be used for the HMMs as this showed a good trade-off between accuracy, robustness and complexity of the system.

Fig. 6.5(c)-(d) show the inference value histograms obtained from the gaze gesture testing data using 6 state HMMs along with unintended gaze gesture data (natural browsing data and the alternative gesture data) for the '*pan*' and '*zoom*' gaze gesture respectively. The white bars in the histogram show the inference values of the unintended gaze gesture data, whilst the dark grey values identifies the overlap of inference values between the intended and unintended gaze gesture data. From Fig. 6.5(c)-(d), a cut-off threshold for the inference values is observable, making it possible to differentiate the intentional and unintentional gaze gestures. An inference threshold of 0.65 was used for the '*pan*' gaze gesture detection and an inference threshold of 0.7 was used for the '*zoom*' gaze gesture detection.

6.1.2 Robot Control Process

The robot control process is illustrated in Fig. 6.6. Once a gesture is recognised, the position of the surgeon's gaze coordinate $PoR = [x_{eye}\ y_{eye}]^T$ on the gaze tracker screen is employed to update a minimum jerk trajectory planner providing the desired pose x_d to the robot controller based on a Cartesian impedance control scheme. The controller computes the command torque τ_c for each joint according to the compensation of the whole robot dynamics and the error signal e estimated as the difference between the reference pose x_d and the actual pose x in Cartesian space retrieved by the robot's forward kinematics. The Cartesian impedance control has been chosen to guarantee both a safe human robot interaction and an intuitive camera positioning during surgery. The control law is described by,

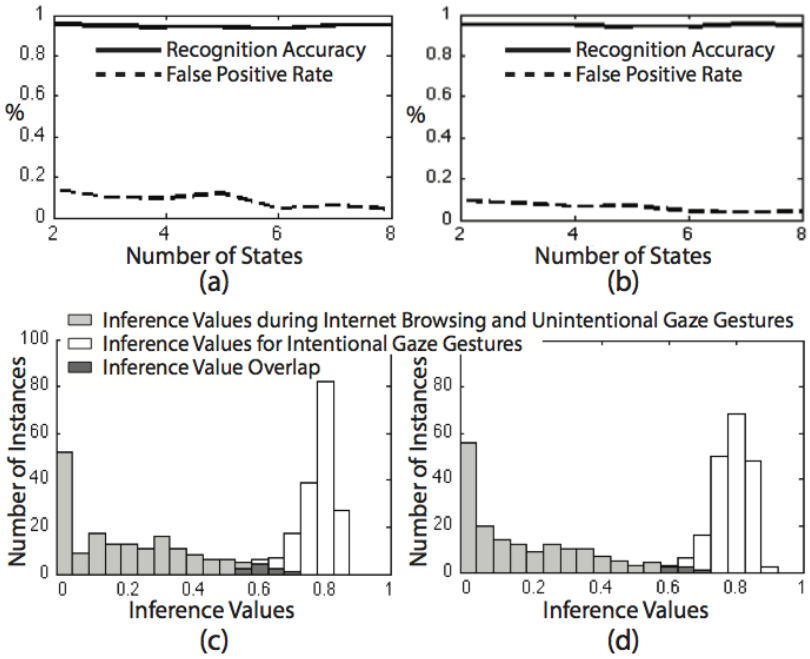


Figure 6.5 – The recognition accuracy and false positive rate across different number of states for the (a) '*pan*' gaze gesture and (b) '*zoom*' gaze gesture. The inference values of unintended gaze gestures and the respective inference values of intended gaze gestures for (c) '*pan*' gestures and (d) '*zoom*' gestures respectively.

$$\tau_c = J^T(q)K\left(x_d - x(q)\right) + D(d_c) + F_d(q, \dot{q}, \ddot{q}) \tag{6.1}$$

where, $J(q) \in \Re^{6\times n}$ is the Jacobian matrix depending on robot configuration q for a robot with n joints, $K \in \Re^{6\times n}$ is a diagonal stiffness matrix, $x = [p\ \varphi]^T$ is the actual pose in terms of position p and orientation φ. $D(d_c)$ is a damping vector depending on the damping coefficients d_c and $F_d(q, \dot{q}, \ddot{q})$ is the robot dynamics compensation including the terms of inertia matrix, centrifugal and Coriolis torques, friction and gravitational torque vector. The vectors of joint position, velocity and acceleration are represented by $q, \dot{q}, \ddot{q}$ respectively. The main benefit in using this control approach is that the robot's visco-elastic properties during interaction with patient tissue can be regulated by tuning robot stiffness K and damping D matrices according to the robot behaviour like a passive mass-spring-damper system. In the '*Camera Assistant*' mode mentioned in Section 6.2.1, the control law is given by $\tau_c = F_d(q, \dot{q}, \ddot{q})$. Further details about the impedance control law can be found in [3, 4].

With reference to Fig. 6.7, the pose vector x of the camera frame C with respect to the base frame B can be retrieved by the forward kinematics in terms of the following homogeneous transformation matrix:

$$T_c^b(q) = \begin{bmatrix} R_c^b & p_c^b \\ O^T & 1 \end{bmatrix} = T_e^b(q)T_c^e, \tag{6.2}$$

where $R_c^b \in SO(3)$ and p_c^b are respectively the rotation matrix and origin position of tool frame C with respect to the base frame B. $T_e^b(q)$ is the homogeneous transformation matrix of the robot end-effector frame E with respect to the base frame B based on robot forward kinematics and $T_c^e(q)$ is the fixed homogeneous transformation between the camera frame C and the end-effector frame E.

In order to convert the surgeon's PoR on a screen into a robot movement in terms of desired pose x_d in Eq. (6.2), a minimum jerk trajectory planner has been implemented for each discrete motion of the frame C with respect to the base frame B. In particular, each movement is broken down into a series of finer point-to-point motions to achieve a safer camera movement in the Cartesian space with a start and stop motion typology. Therefore, the overall movement of the robotic arm is composed of a sequence of point-to-

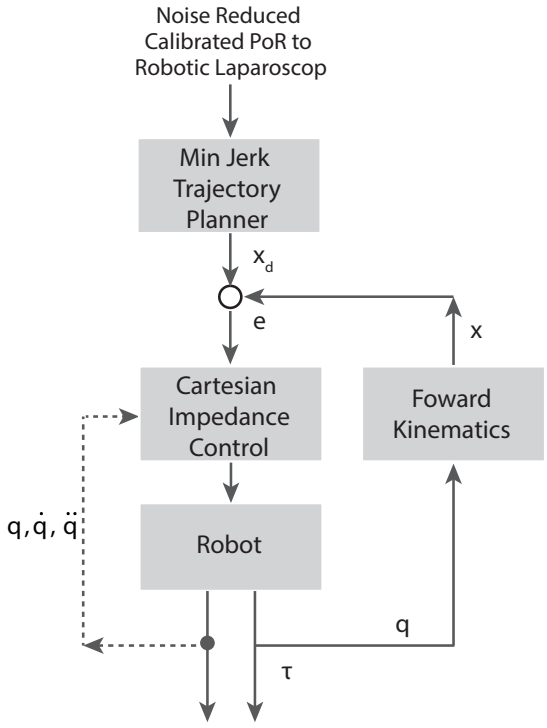


Figure 6.6 – The position of the surgeon's PoR on the gaze tracker screen is used to update a minimum jerk trajectory planner providing the desired pose x_d to the robot controller based on a Cartesian impedance control scheme. The controller computes the command torque τ_c for each joint according to the compensation of the whole robot dynamics and the error signal e estimated as the difference between the reference pose x_d and the actual pose x in Cartesian space retrieved by the robot's forward kinematics.

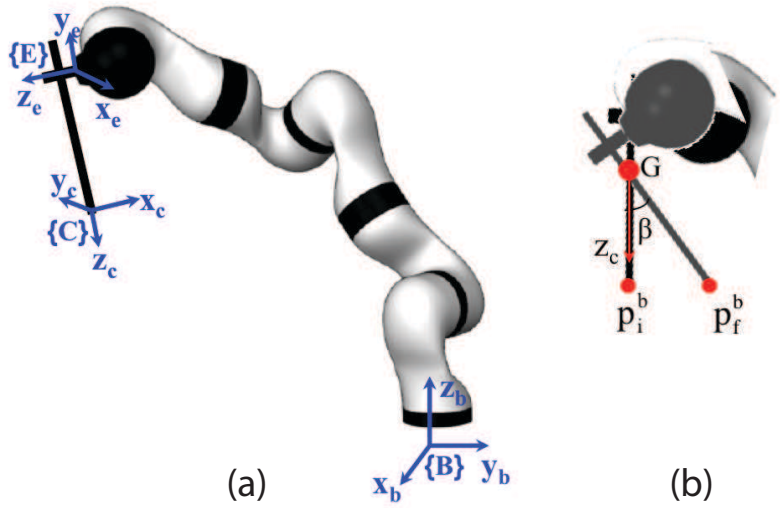


Figure 6.7 – (a) Reference frames placement; (b) RCM constraint.

point motions, each of them consisting of a segment in the Cartesian space with the assumption of small movements. At the end of the current point-to-point motion, the trajectory planner is updated with the new data on the surgeon's PoR if the robot motion has been enabled by a gaze gesture. A generic segment connecting (in base frame B) the initial point p_i^b to the end point p_f^b in a time interval t_f is described by the parametric function,

$$p(s(t)) = p_i^b + \frac{s(t)}{\left\|p_f^b - p_i^b\right\|}(p_f^b - p_i^b), \tag{6.3}$$

where the arc length $s(t)$ is the fifth order polynomial function of time:

$$s(t) = \sum_{i=0}^{5} c_i t^i. \tag{6.4}$$

The point p_i^b represents the co-ordinates of the point p_c^b when a new motion planning begins, whereas the computation of final point p_f^b has been based on the PoR coordinates. Assume that, after an initial calibration, the coordinates $PoR = [x_{eye}\ y_{eye}]^T$ on the eye tracker screen correspond to the xy plane of the camera frame C in Fig. 6.7, then the direction, α of the next camera motion in the same plane can be reconstructed as

$$\alpha = \arctan\left(\frac{y_{eye}}{x_{eye}}\right). \tag{6.5}$$

The proposed gaze contingent system computes the final position p_f^c with respect to the camera frame can be expressed as:

$$p_f^c = L_d \cdot [\cos(\alpha) \;\; \sin(\alpha) \;\; 0]^T \tag{6.6}$$

where L_d is the constant displacement in the xy plane and the corresponding point p_f^b of the motion planning in base frame B is given by

$$[p_f^b \; 1]^T = T_c^b [p_f^c \; 1]^T. \tag{6.7}$$

Eq. (6.3)–(6.7) provide a smooth point-to-point motion [79] in the xy plane of the reference frame C in Fig. 6.7(a) with respect to base frame B. Thus far, only the control of the laparoscope position has been discussed. As regards to the control of the orientation, an additional constraint needs to be taken into account; *i.e.,* the RCM. In fact, during a minimally invasive laparoscopic surgery, the access trocar provides a fixed insertion point that acts as a fulcrum point as shown in Fig. 6.7(b). In order to implement a RCM passing through the camera tool, the following rotation matrix has been computed to constrain the tool orientation by means of angle/axis technique [143]

$$R(\beta, \hat{a}) = \begin{bmatrix} \hat{a}_x^2(1-c_\beta)+c_\beta & \hat{a}_x\hat{a}_y(1-c_\beta)-\hat{a}_z s_\beta & \hat{a}_x\hat{a}_z(1-c_\beta)+\hat{a}_y s_\beta \\ \hat{a}_x\hat{a}_y(1-c_\beta)+\hat{a}_z s_\beta & \hat{a}_y^2(1-c_\beta)+c_\beta & \hat{a}_y\hat{a}_z(1-c_\beta)-\hat{a}_x s_\beta \\ \hat{a}_x\hat{a}_z(1-c_\beta)-\hat{a}_y s_\beta & \hat{a}_y\hat{a}_z(1-c_\beta)+\hat{a}_x s_\beta & \hat{a}_z^2(1-c_\beta)+c_\beta \end{bmatrix} \tag{6.8}$$

where β is the rotation angle around the axis $\hat{a} = [\hat{a}_x \; \hat{a}_y \; \hat{a}_z]^T$ in the base frame.

Let the point G be the position in frame B of the RCM constraint in Fig. 6.7. The rotation matrix $R(\beta, \hat{a})$ constrains the orientation of the laparoscope to pass through G during the motion execution from point p_i^b to point p_f^b. Given the vector $v = p_f^b - p_G^b$ and the z-axis unit vector z_c of the reference frame C, the rotation angle β and rotation axis $\hat{a}$ are respectively

Table 6.1 – Motion parameters for each robot movement.

	Planner Parameters		Avg. Speed	Condition
Static	L_d=0.0 mm	t_f=0.010 s	0.00 mms^{-1}	$r < r_1 = 0.18$
Medium	L_d=5.5 mm	t_f=0.325 s	16.9 mms^{-1}	$r_1 \leq r < r_2 = 0.38$
Fast	L_d=7.0 mm	t_f=0.300 s	23.3 mms^{-1}	$r > r_2$

$$\beta_f = atan2\left(\frac{\|z_c \times \nu\|}{\|\nu\|}, \frac{z_c^T \cdot \nu}{\|\nu\|}\right),\ \hat{a} = \frac{z_c \times \nu}{\|z_c \times \nu\|}. \tag{6.9}$$

Therefore, the homogeneous transformation matrix at instants $t = 0$ and $t = t_f$ are

$${}^{i}T_c^b(q) = \begin{bmatrix} {}^{i}R_c^b(q) & p_i^b \\ 0^T & 1 \end{bmatrix},\ {}^{f}T_c^b(q) = \begin{bmatrix} R(\beta, \hat{a}){}^{i}R_c^b(q) & p_f^b \\ 0^T & 1 \end{bmatrix}, \tag{6.10}$$

where ${}^{i}R_c^b$ is the rotation matrix of the frame C at the instant time $t = 0$. The angle β varies between $\beta|_{t=0} = 0$ and $\beta|_{t=t_f} = \beta_f$ with the timing law expressed in Eq. (6.4). In summary, the motion of the camera tool can be decomposed in terms of simultaneous rotation around the RCM and translation of the frame C, where Eq. (6.3)–(6.7) regulate the translation and Eq. (6.8)–(6.10) are in charge of the rotation. The procedure described so far addresses the motion modality of camera panning in Section 6.2.1.

In Table. 6.1, the control parameters for three different camera velocities based on the normalised Euclidean distance r, of the gaze from the centre of the screen have been chosen for the '*pan*' according to clinical requirements and the aforementioned assumption of small movements of each segment to avoid unexpected motion in the joint space.

Without a loss of generality, the same motion generation approach was applied to '*zoom*' mode by considering

$$p_f^b = p_f^b \pm z_c \cdot L_d,\ {}^{f}R_c^b(q) = {}^{i}R_c^b(q),$$

where the ±was used to zoom in and out respectively. The variables L_d and t_f have been fixed to $5.0mm$ and $0.500s$ respectively for zooming action.

6.2 Implementation

In this section, the UI of the following three control modes are described:

1. Gaze gesture control; gesture based activation and camera control with their PoR.

2. Pedal activated control; where the user activates the control mode of the camera with a dual-switch foot-pedal and directs the camera with their PoR.

3. Camera assistant control; a camera assistant follows verbal instructions of the experimental participant and navigates the camera.

Additionally, the experimental setup and the protocols are described in detail.

6.2.1 Gaze Contingent Control Modes and UI

The implemented gaze gesture control UI for the gaze contingent robot system is illustrated in Fig. 6.8. The UI is designed to enable the user to intuitively switch between the '*pan*', '*zoom*' and stopping of the camera. At the outset, the camera system is stationary and is waiting for a gesture input from the user. The user will be observing a camera view as in Fig. 6.8(a) where there are guidance '*zoom*' and '*pan*' text in the left and right bottom corner respectively. Whilst using the system, the user is able to see their PoR on the screen in the form of a white dot. They also have the option to turn this off. If a '*zoom*' gaze gesture is triggered then the UI switches from Fig. 6.8(a) to Fig. 6.8(b) and the user will be able to zoom in by looking anywhere above the horizon and zoom out by looking below. Guidance text is also overlaid on the camera view and the camera can be stopped by looking into the white circle in the centre for $750ms$. In contrast, if a '*pan*' gaze gesture is inputted by the user, the UI switches from Fig. 6.8(a) to Fig. 6.8(c) where the camera pans in the vector direction of the gaze from the screen centre.

If the user's PoR is within a region of $r \leq 0.1$ from the screen centre, then the camera will stop after $750ms$. In '*pan*' mode, the screen area is separated in three radial regions where gazing within each region would prompt the camera to move at a different velocity. This region separation is illustrated

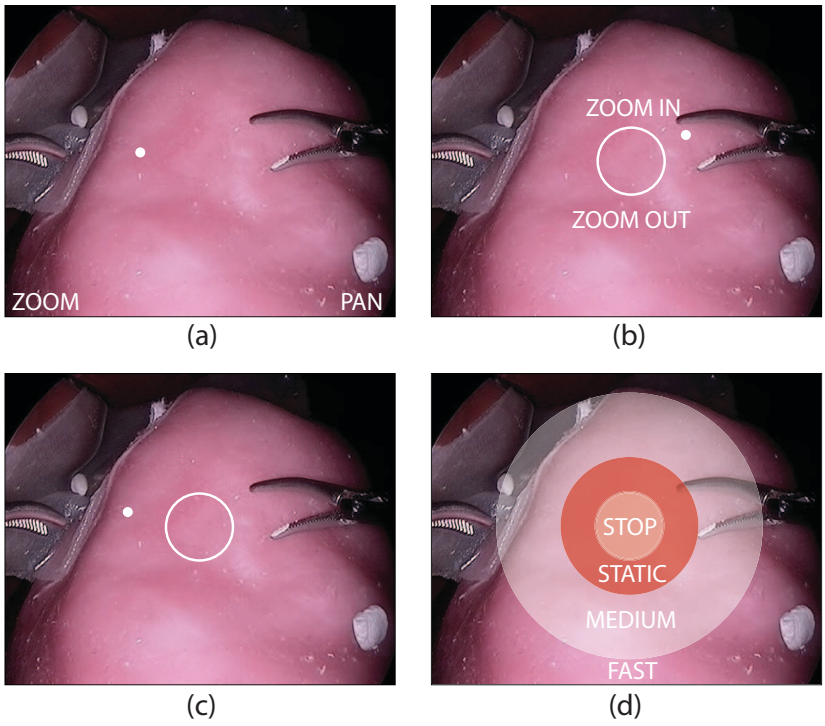


Figure 6.8 – Gaze gesture controlled robotic laparoscope UI. (a) The UI waits for the user to perform a '*zoom*' or '*pan*' gaze gesture. If a '*zoom*' gaze gesture is detected then the user will be taken to (b), or if a '*pan*' gesture is detected it will take them to (c). While in (b) or (c) the camera can be stopped by looking into the white circle for $750ms$. After the camera stops, the user is taken back to (a). (d) Illustration of the camera velocity and respective regions under '*pan*' camera control mode.

in Fig. 6.8(d) with respective radial region values described as in Table. 6.1. If the PoR is within the static region, the camera will remain static. This region was introduced to enable the surgeon to have a stable working area to operate whilst maintaining an active '*pan*' mode. If the PoR falls within medium or fast region, the camera pans in a medium or fast velocity set at $16.9mms^{-1}$ and $23.3mms^{-1}$ respectively. The different speed regions were introduced to enable an intuitive control of the camera whilst maintaining a known maximum velocity that would be safe for the patient.

In order to address the potential safety issue when the eye tracker looses tracking of the user's eyes, an effective safety mechanism was introduced where the robotic laparoscopic holder would immediately stop under lost gaze tracking conditions. If the user's gaze is re-detected, then the robotic laparoscope would resume in the same user control mode.

The pedal activated control mode is similar to the UI of the gaze gesture control except that the onset camera screen view shown in Fig. 6.8(a) would not have the '*zoom*' and '*pan*' text in the corner as a foot pedal is being used to activate the camera arm. During pedal activation control mode, a dual-switch foot pedal is used. To activate the '*zoom*' or '*pan*' mode the user is required to step on the left or right pedal respectively. Navigation of the camera is conducted using gaze with the same UI as in Fig. 6.8(b) and Fig. 6.8(c) for '*zoom*' and '*pan*' mode respectively. During camera assistant mode, an operator manually controls the laparoscopic camera using the method explained in Section 6.1.2.

6.2.2 Experimental Setup

The experimental platform which is illustrated in Fig. 6.9 includes a Tobii 1750 eye tracker, a $10mm$ zero degree Stortz laparoscope, a Storz Tele Pack light box, a Kuka Light Weight Robot (LWR) [4] and an upper gastrointestinal phantom with simulated white lesions. Both the robot and the HMM gaze gesture recognition processes have been implemented in C++ and run at 200Hz and 33.3Hz respectively. All experimental data including the HMM gaze gestures, the camera-view video, the PoR and camera position in Cartesian space obtained from the robot forward kinematics, were recorded at 33.3Hz with respective time stamps. The bi-directional communication and synchronization between the devices was achieved using the

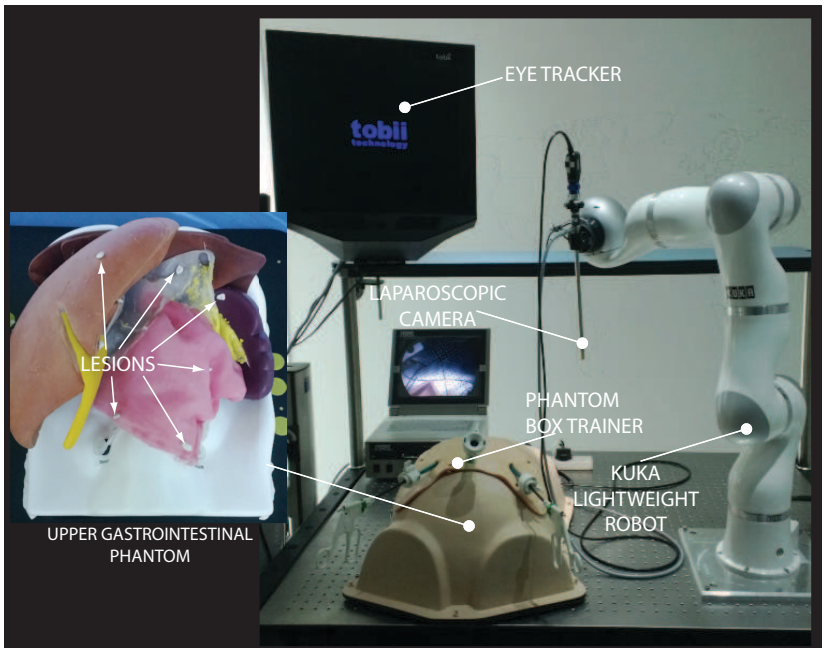


Figure 6.9 – The experimental setup includes a commercial eye tracker (Tobii 1750), a 10*mm* zero degree Stortz laparoscope and Tele Pack light box, a Kuka LWR and a upper gastrointestinal phantom with simulated white lesions.

UDP protocol via an Ethernet cable.

6.2.3 Experimental Protocols

Eleven surgical residents with a postgraduate year between 3-7 (PGY3-7, male=10, female=1) were consented to participate in this study. The laparoscopic experience was 536 (+/- 315) cases. All subjects were initially trained to use both the gaze gesture and pedal activated systems on an abstract navigation task. This was done to mitigate the potential confounding effects of learning when performing the subsequent study task.

The abstract training task involved navigating the robotic laparoscope within a conventional box trainer to find numbers on a 4×5 grid (illustrated in Fig. 6.10) in ascending magnitude. The numerals were of differing font sizes, necessitating zooming and panning to successfully complete the task. Training was completed when they had met a minimum baseline pro-

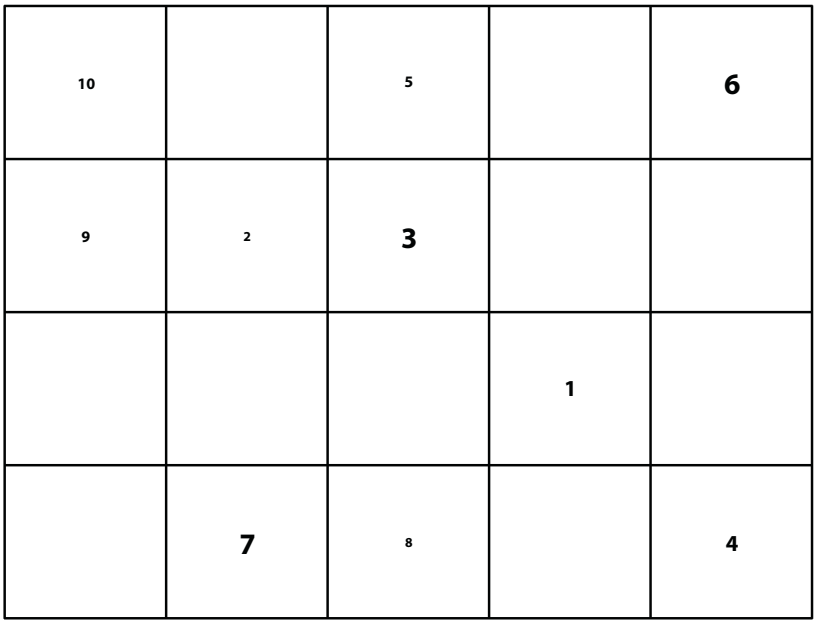


Figure 6.10 – Training task requesting participants to navigate the robotic laparoscope within a conventional box trainer to find numbers on a 4×5 grid in ascending magnitude.

ficiency task completion time, when there was no further improvement in completion time and they could reproduce their best time on three consecutive occasions.

A modified upper gastrointestinal staging laparoscopy phantom was used in an immersive laparoscopic box trainer. The subjects were tasked with identifying and "biopsying" (*i.e.,* removing) a set number of randomly placed lesions on the phantom. Importantly, this simulated task required subjects to use a bi-manual technique, with one instrument either manipulating or retracting tissue, whilst the other removes the lesion. The surgeons were allowed and encouraged to physically look at the phantom model (before lesions were placed) to familiarise themselves with it. This was introduced to minimise the potential confounding factor of learning the phantom model. Participants were asked to perform the lesion removing task twice, for each of the three camera control modes mentioned previously, namely via i) gaze gesture control, ii) pedal activated control, and iii) camera assistant control. Thus, overall each participant performed the lesion removal task six times

in total. The sequence in which subjects performed the three task modes was randomized to mitigate the learning effects on the phantom. The same human camera assistant was used throughout the study for all subjects. The assistant was experienced in controlling the robotic laparoscope arm.

Eye tracking data for each participant was recorded during the gaze gesture control mode to assess gaze gesture usability. A high performance gaze gesture recognition algorithm plays a critical role in the usability and ergonomics of the gaze contingent system. In order to assess the gaze gesture recognition process, post-hoc observation of the recorded camera-view videos was conducted and the occurrences of true positive (*i.e.,* correctly identified), false positive (*i.e.,* incorrectly identified) and false negative (*i.e.,* incorrectly rejected) gaze gestures were recorded. The number of true negative (*i.e.,* correctly rejected) gaze gestures were obtained by filtering the recorded gaze data and counting the rejected potential gaze gestures during the trials. The recognition accuracy is obtained by,

$$\frac{\sum True\ Positive\ Gestures}{\sum True\ Positive\ Gestures + \sum False\ Negative\ Gestures}, \tag{6.11}$$

and the false positive rate is obtained by,

$$\frac{\sum False\ Positive\ Gestures}{\sum False\ Positive\ Gestures + \sum True\ Negative\ Gestures}. \tag{6.12}$$

If the trained HMM gaze gestures provide low recognition accuracy, this can induce frustration in the surgeon as their intention to control the camera is rejected, delaying the use of the laparoscope camera. On the contrary, a high false positive rate would lead to involuntary activation and subsequent movement of the camera, again resulting in a poor user experience, and overall could lead to a low uptake of the system.

In addition, the system usability during each laparoscope control mode is assessed quantitatively through the use of the following performance metrics.

- Task completion time measured in seconds.
- Camera path length measured over a single trial in centimetres - to assess camera control efficiency and usability of the system.

- NASA-TLX questionnaire - subjects completed a questionnaire after each lesion removal task. This subjective questionnaire is validated and comprises of six variably weighted parameters that contribute to task workload [182].

At the end of the trial, subjects completed a questionnaire providing their laparoscopic experience and rated how difficult it was to learn to use a) the gaze gesture system and b) the foot pedal activation system on a visual analogue scale.

6.3 Results

In order to investigate the usability performance of the new gaze contingent laparoscope system design, the following statistical analysis studies were performed.

1. Recognition accuracy and false positive rate assessment of the HMM gaze gestures.
2. Comparative analysis of the three control modalities.

The statistical comparative analysis of the three control modalities of the system were carried out on the performance metrics mentioned in Section 6.2.3. All statistical analysis involved first conducting normality tests, followed by Mann-Whitney U test for non-parametric continuous variables between modalities. A 'p-value' <0.05 was considered significant. Results with significant differences are indicated with an asterisk '*' mark in all tables. Results are represented as medians with IQRs in parentheses, unless otherwise stated. All statistical analysis was performed using MATLAB.

6.3.1 HMM Gaze Gesture Performance Assessment

The recognition accuracy and false positive rate results are shown in Table. 6.2. The overall average recognition accuracy for the HMM based gaze gestures is 97.0% with an average false positive rate of 1.4%. These results demonstrate that gaze gestures provide high recognition accuracy and a low false positive rate that makes the use of HMM based gaze gestures both a user-friendly and safe method for the surgeon to convey their intention to control the robotic system.

Table 6.2 – Recognition accuracy and the false positive rate of the HMMs during the experimental trial.

	Recognition Accuracy	False Positive Rate
HMM1 ('*Pan*')	95.6%	2.2 %
HMM2 ('*Zoom*')	98.3%	0.6 %

Table 6.3 – Comparative analysis of the three control modalities. Median and interquartile range in parentheses.

	Task time (s)	Cam Path Length (cm)	NASA-TLX Score	Sub. Skill Acquisition Difficulty†
(1) Assistant	290.0 [181.0]	171.5 [126.3]	43.3 [25.0]	-
(2) Gaze Gesture	266.5 [182.0]	89.3 [57.6]	42 [39.1]	0.853 [0.15]
(3) Pedal	288.0 [158.8]	109.8 [87.5]	27 [28.6]	0.884 [0.14]
p-value (1) vs (2)	0.536	**0.028***	0.876	-
p-value (1) vs (3)	0.697	0.078	0.262	-
p-value (2) vs (3)	0.780	0.388	0.273	0.178

*$p<0.05$, †Visual analogue scale score: 1=very easy, 0=very difficult

6.3.2 Comparative Analysis of the Three Control Modalities

The comparative analysis of the three control modalities are summarised in Table. 6.3. All eleven subjects met the baseline proficiency and training requirements to be included in the user performance based quantitative analysis. Two of the subjects wore glasses and three wore contact lenses.

The task completion times showed no significant difference (p=0.536 and p=0.697) between the three control modalities, indicating that the proposed gaze contingent system is as competent as using a camera assistant, which is the current gold standard. The group task completion times are observable in Fig. 6.11.

The camera path length was measured in order to assess how efficient and user-friendly our system is compared to the other two modes. Fig. 6.12(a)-(b) illustrates an example camera path of a single subject during the gaze gesture based control of the camera and the camera assistant respectively. It is clear that the camera assistant shows a more volatile camera trajectory. In contrast, the camera path during the gaze gesture based gaze contingent control mode is smooth and still offers the same navigational workspace of an assistant.

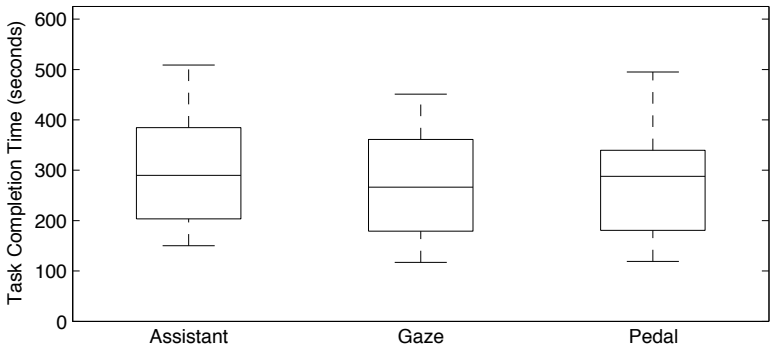


Figure 6.11 – Group task completion times by control modality.

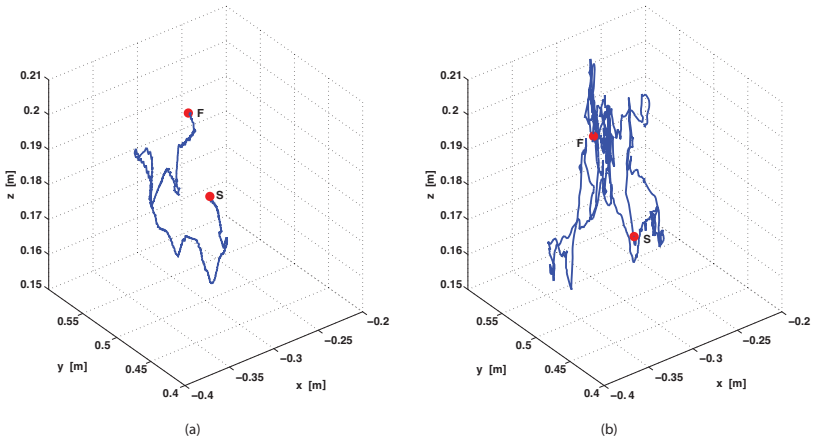


Figure 6.12 – Camera path for a given subject during (a) gaze gesture based gaze contingent control and (b) camera assistant control of the laparoscope.



Figure 6.13 – Group camera path lengths (cm) by control modality. The camera path length obtained during the gaze gesture control mode showed significantly shorter path length against the camera assistant.

Fig. 6.13 shows the group's total camera path lengths during each control modality. A significantly shorter camera path length for the gaze gesture modality compared to the assistant was observed with median and interquartile range values of (89.3 [57.6]cm vs 171.5 [126.3]cm; p=0.028). The pedal activation control mode showed a shorter path length against the camera assistant, but did not result in statistical significance (109.8 [87.5]cm vs 171.5 [126.3]cm; p=0.078).

The gaze contingent system was also assessed for its contribution to the workload of the surgeon with the NASA-TLX questionnaire. It is desirable that any surgical innovation does not add to the cognitive burden of the surgeon. Importantly, there was no statistically significant difference in NASA-TLX scores for the gaze gesture modality versus the human camera assistant (p=0.876) or the foot pedal (p=0.262) as shown in Fig. 6.14.

Visual analogue scores for rating subjective difficulty in skill acquisition between the gaze gesture and the pedal activation control modalities were not significantly different (0.853 [0.15] vs 0.884 [0.14]; p=0.178). 100% of surgeons rated both systems as 'easy' to learn. Pedals are considered commonplace in the operating theatre. Thus, for gaze gestures to be as easy to learn as the foot pedal is a measure of the intrinsic ergonomics of gaze gestures.

In the questionnaire, the subjects were asked to state their preferred mode

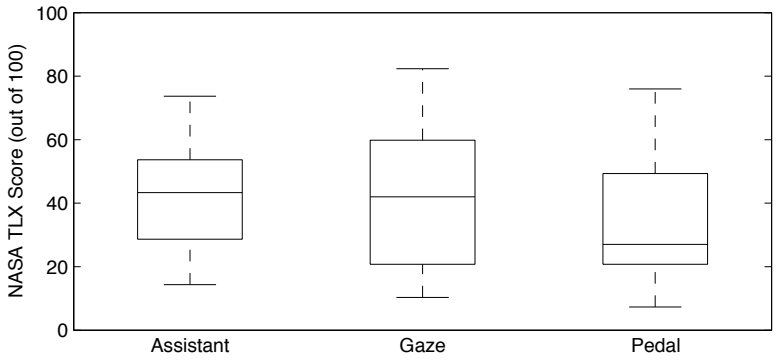


Figure 6.14 – Group NASA-TLX scores by control modality. No significant difference in NASA-TLX scores were observed between the gaze gesture and pedal modality against the human camera assistant.

and 91% of surgeons (10/11) documented they would use the gaze gesture system in clinical practice and also recommend it to their surgical colleagues. Participants raised a number of points for development with regards to the system. These include enabling panning during zooming in, a different stop mechanism (the current system requires to fixate into a circular region in the centre, which overlaps with the region of interest for the surgeon to perform the task), and a camera tilt control to enable rotation around the laparoscope's longitudinal axis. The current stop mechanism sometimes lead to surgeons expressing frustration that the camera stops accidentally.

6.4 Conclusions

In this chapter, an approach to ergonomically control a 6 DoF camera through the surgeons' PoR (*i.e.,* 2 DoF) as the control input was introduced. The use of real-time HMM gaze gestures to convey the user's camera control intention of pan and zoom was proposed. Controlling of the laparoscope robot arm holder was implemented to ensure that there was no misalignment between the resulting camera visualisation movement with the user input control, resulting in a natural camera movement due to overcoming of the fulcrum effect. The ability to reproducibly complete a surgical navigation task was demonstrated using the gaze gesture control modality. The

efficiency of the system was demonstrated by the significantly shorter camera path length compared to the current gold standard (camera assistant). Thus, surgeons obtained the desired FOV more accurately and required less fine-tuning. The longer camera path length during the use of a camera assistant was most likely a reflection of the difficulty in communicating how the camera should be moved to another human though verbal commands. Validation of the HMMs showed that HMMs were effective in recognising gaze gestures with mean experimental recognition accuracy and false positive rate of 97.0% and 1.4% respectively.

The results obtained from the user trials support the claim that the proposed gaze contingent robotic camera system with built-in safety mechanisms was ergonomic to use, inducing no significant cognitive burden when compared to using a camera assistant or when compared to using a foot pedal activated camera system. This means that the participants did not feel that the gaze gestures and gaze control as a complex user modality. This was also reflected in the subjective visual analogue scores. Additionally, the system has the desirable feature of enabling the surgeon to achieve comparable performance without the need of additional foot pedals. More importantly, there is potential cost reduction opportunities through using such a system as it can reduce the need to use a highly skilled camera assistant and allow a surgeon to operate by them self, providing cost and staffing efficiencies.

Through the user trials, there were a number of improvements suggested by participating surgeons and these included, i) a tilt control capability of the camera to enable rotation of the camera view around the laparoscope's longitudinal axis, ii) mechanism of simultaneous panning whilst zooming in, offering a more natural camera movement experience, and iii) a different camera stopping mechanism. The tilt control was suggested based on the group of surgeons' previous experiences as well as literature suggesting that having a tilted horizon at certain angles can increase the risk of laparoscopic procedures [184]. Therefore integrating the capability to correct for tilt angle rotation was seen important. The stopping region in the UI illustrated in this chapter is currently in the centre of the screen. This can overlap with the region of interest for the surgeon during the operation causing the camera to stop unintentionally. In the next chapter, adjustments are made to the gaze contingent laparoscope system to include the above suggested

points as well as to introduce further improvements to the usability and ergonomics of the system by introducing an online calibration scheme of the eye tracker. This online calibration scheme will allow the user to start using the gaze contingent camera immediately on the fly, improving usability and the overall surgical workflow.

7 Online Calibrated Gaze Contingent Control

In Chapter 6, a robotic assisted laparoscope holder was introduced to provide surgeons with the ability to simultaneously navigate and perform a surgical procedure without having assistance from a human camera assistant. The usability studies revealed that a gaze gesture activated gaze contingent robotic laparoscope holder enabled the surgeons to have a more efficient camera manoeuvrability when compared to using a camera assistant. At the end of Chapter 6, a number of discussion points for improvement were raised. These included:

- **Introducing tilt control of the camera**. This would enable rotation around the laparoscope's longitudinal axis, therefore enabling horizon correction which is useful for reorientation [184].

- **Enabling simultaneous zooming in and panning.** The previous system enables zooming in or out when fixating above or below the horizon. This means that the zooming is along the laparoscope depth axis only. Surgeons expressed a need to enable panning whilst zooming inwards.

- **A different camera stopping method.** The previous system required the user to fixate into a circular region in the centre of the screen to deactivate the camera movement. This overlaps with the region of interest for the surgeon during the operation. The arrangement sometimes led to surgeons expressing frustration that the camera stops without them wanting it to. Surgeons conveyed their desire for a camera mechanism that would remain stationary while the camera is activated but only deactivates through an explicit signal to the system.

This chapter focuses on incorporating the above clinical suggestions and further improving the ergonomics and usability of the gaze contingent system. The novelties of the work presented in this chapter include; i) the introduction of tilt control (rotation around the laparoscope's longitudinal axis), with which the user is able to pan, tilt and zoom the camera, ii) enabling simultaneous panning whilst zooming the camera inwards, iii) modification of the switching between panning and zooming control by the user moving their head forward to zoom in and backward to zoom out, offering seamless conveying of their control intention, iv) an online calibration method for the gaze tracker, which overcomes the need of an explicit offline gaze calibration before using the gaze contingent system, offering a fast, frustration-free user experience of the system. Comprehensive usability studies to assess the modified system, including analysis of the user's instrument trajectories during use of the system for better objective assessment are used in this chapter.

The hands-free gaze gesture activated gaze-directed camera empowers the surgeon to use both their hands to operate, whilst eliminating the need for a camera assistant. The online nature of the calibration algorithm introduced in this chapter enables continuous updating of the gaze tracker calibration, thus helping to overcome the calibration drift that can occur over longer periods of using gaze trackers [59, 112]. The proposed online calibration algorithm can be used for any remote gaze tracker system in condition that there is a user-interactive aspect to the system application, for example a menu navigation application involving the user's PoR, gaze typing or an automatic scrolling mechanism during reading. The online calibration works by obtaining calibration points from the user interactive aspects and use these to remap the user's PoR. A study to demonstrate the performance of the online calibration algorithm is presented and is compared against a conventional offline gaze tracker calibration.

This chapter is presented as follows; Section 7.1 highlights the changes made to the gaze gesture recognition process, the training of the gaze gestures and the robot arm control algorithms. Section 7.2 describes the online calibration algorithm. Section 7.3, gives an overview of the system user control interface, the experimental setting and experimental protocols. In Section 7.4 the results are discussed and this is followed by the conclusion in Section 7.5.

7.1 Gaze Contingent System Modifications

The gaze gesture activated robotic laparoscope system's fundamental internal control mechanism, (gaze gesture recognition and robot control processes) is kept the same as in Chapter 6, shown in Fig. 6.3. The possible gaze gestures the user can perform are two: '*activate camera*' (for pan and zoom) or '*tilt camera*' (for rotation). These are different to the ones presented in Chapter 6.

The '*activate camera*' gaze gesture sequence illustrated in Fig 7.1(a) is defined by a three-stroke sequence of eye movements: gazing at the centre, then bottom right corner, back to the centre, and finally back to the bottom right corner of the screen. The '*activate camera*' gaze gesture activates a camera control mode that enables the user to perform pan and zoom control in one control mode. On initialisation, the camera records the distance of the user's eyes from the gaze tracker screen. If the user maintains their head position within a distance threshold, the camera pans according to the vector direction and distance from the screen centre. If the user wants to zoom the camera inwards, they simply are required to lean forward by the distance threshold and the camera will zoom in. During zooming inwards, the user is able to simultaneously pan. This setup enables a smooth simultaneous zooming and panning control, as suggested by the group of clinical participants in the study of Chapter 6. Additionally, in order to zoom the camera view outwards, the user is requested to lean back by the distance threshold. More detailed UI controls are described in Section 7.3.1.

The '*tilt camera*' gaze gesture sequence is shown in Fig. 7.1(b), is similar to the '*activate camera*' gaze gesture sequence but instead, the user is asked to gaze at the bottom left corner of the screen. The '*tilt camera*' gaze gesture activates the camera control that enables rotation of the camera view along the longitudinal axis of the laparoscope. Once a gaze gesture sequence is performed and detected, camera motion is enabled and the user can manoeuvre the laparoscope robotic arm holder with their PoR.

In order to implement a different stopping mechanism to that described in Chapter 6.2.1, the user now is asked to fixate at the bottom left corner for $750ms$ instead of the centre of the screen to stop the camera. This was introduced by request of a number of surgeons who participated in the previous user study in Chapter 6 who found the screen centre stopping

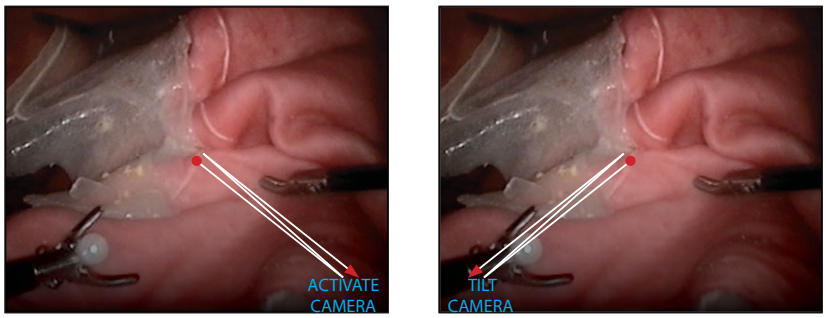


Figure 7.1 – The gaze gesture sequence for the (a) '*activate camera*' (pan and zoom) and (b) '*tilt camera*' (rotation) gaze gestures enabling activation of the gaze contingent laparoscope system.

region unintuitive due to the camera accidentally stopping if they gazed within the centre of the screen.

7.1.1 Updated Gaze Gesture Recognition Process

The gaze gesture recognition process is the same to that presented in Chapter 6, with the only difference being that more training data were collected from an additional ten participants. The codebook was designed offline from twenty participants providing 600 gaze gesture training data sequences. Unintentional gaze gesture data were also collected from the additional participants during a five-minute web browsing task. Each of the 600 training sequences was encoded using the formulated k-means clustering codebook. The Baum-Welch algorithm is used to iteratively obtain the optimal state transition and emission probabilities that describe the training data set observations O. A 10-fold cross validation was performed across a different number of states (between 2 and 8 states, total of seven runs as shown in Fig. 7.2(a)-(b)) for each of the two HMMs which describe the two gaze gestures to find a good trade-off between recognition accuracy, a low false positive rate, and overall complexity of the system. With each of the seven training instances, a detection probability threshold was set at a 95% confidence limit of the training data sequences' inference values obtained from the trained HMM. The recognition accuracy of the HMM is obtained by using this threshold on the remaining 10%, the validation data set. Inference value histograms obtained from testing the 6 state HMMs with the

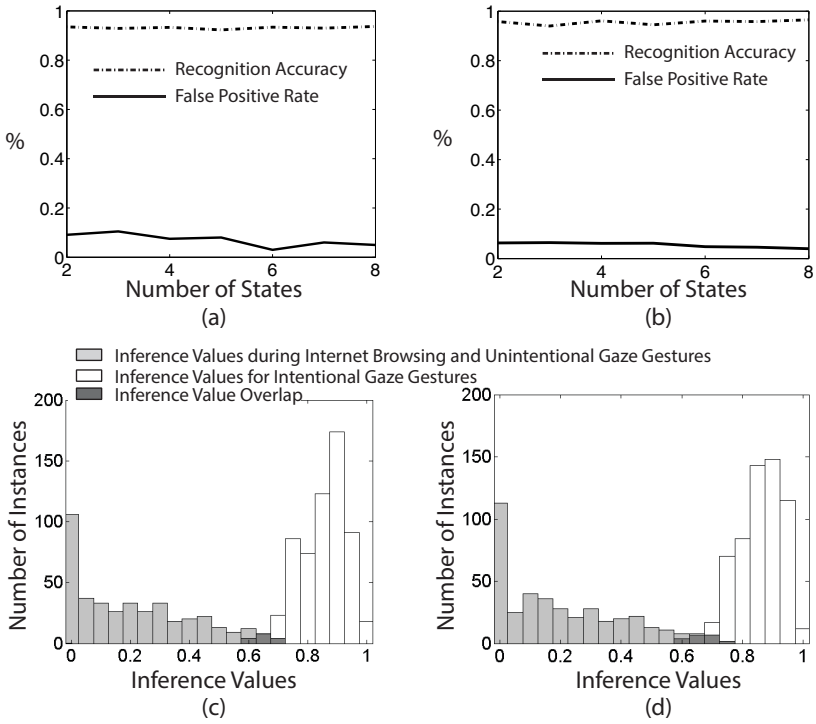


Figure 7.2 – HMM training 10-fold cross validation for the (a) '*activate camera*' and (b) '*tilt camera*' HMM. The inference values of intended gaze gestures for (c) '*activate camera*' gestures and (d) '*tilt camera*' gestures.

unintended gaze gesture data are shown for the '*activate camera*' and '*tilt camera*' gestures in Fig. 7.2(c)-(d). From this figure, it was found that a 6 state HMM with an inference threshold of 0.7 would provide the best overall performance for both the '*activate camera*' and '*tilt camera*' gaze gesture detection.

7.1.2 Updated Robot Control Process

In addition to the above changes to the gaze gesture recognition process and gesture training, a number of modifications were also made to the robot control process. The majority of the robot control process remains the same as Section 6.1.2. The following description describes only the changes made to the robot control process to enable the list of modifications mentioned in

the beginning of this chapter. Eq. (6.6) which describes the final position p_f^c with respect to the camera frame $\{C\}$ (illustrated in Fig. 7.3) is now expressed as

$$p_f^c = \begin{bmatrix} L_d \cos(\alpha) & L_d \sin(\alpha) & L_z \end{bmatrix}^T, \tag{7.1}$$

to enable simultaneous motion of pan and zoom. L_d and L_z are respectively the displacement in the plane xy (*i.e.,* pan motion) and in z-axis (*i.e.,* zoom motion) of the frame $\{C\}$. If only the pan mode is active $L_z = 0$, whereas in case of zoom out motion, the equation becomes

$$p_f^c = \begin{bmatrix} 0 & 0 & -L_z \end{bmatrix}^T. \tag{7.2}$$

In Table. 7.1, the control parameters for three different camera velocities based on the normalised Euclidean distance r, of the gaze from the centre of the screen have been chosen for the panning and zooming in during the '*activate camera*' control mode according to clinical requirements and the previously mentioned assumption of small movements of each segment to avoid unexpected motion in the joint space. The action of zooming out has constant parameters $L_d = 0$, $L_z = -0.005m$ and $t_f = 0.5s$.

In order to perform the '*tilt camera*' mode, *i.e.*, the rotation γ around the longitudinal axis of the laparoscope (*i.e.,* unitary vector z_c in Fig. 7.3), the parameters $L_d = 0$; $L_d = 0$ (*i.e.,* $p_f^b = p_i^b$) and ${}^fT_c^b$ in Eq. (6.10) becomes

$${}^fT_c^b(q) = \begin{bmatrix} {}^iR_c^b(q)\, R_z(\gamma) & p_i^b \\ 0^T & 1 \end{bmatrix},$$

where $R_z(\gamma)$ represents the elementary rotation matrix of the frame $\{C\}$ about the unitary vector z_c shown in Fig. 7.3. In addition, mechanical joint limits have been taken into account in the motion planning by keeping the joint variables as close as possible to the centre of their ranges.

7.2 Online Calibration Algorithm

As mentioned in the introduction of this chapter, an online calibration method for the gaze tracker is introduced. Conventional remote gaze trackers require an offline calibration procedure to map the optical axis (OA) (*i.e.*, the straight line that passes through both the pupil and cornea centre) of

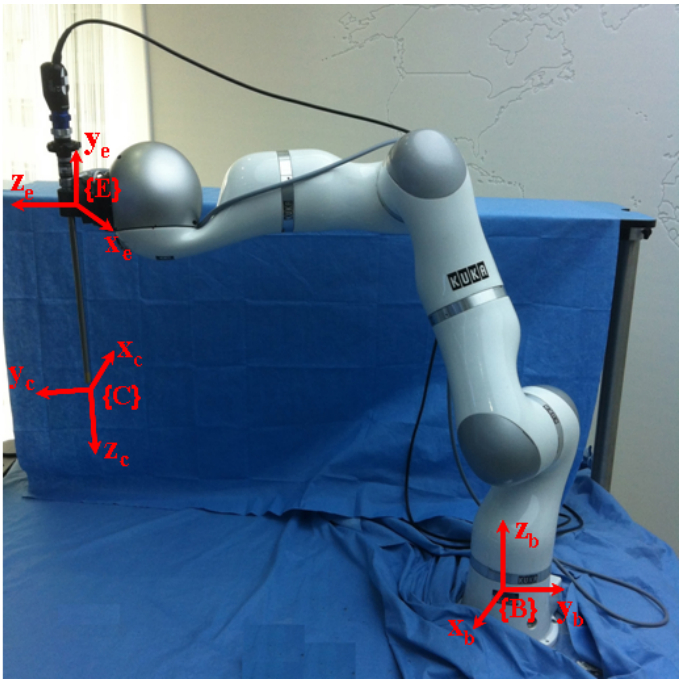


Figure 7.3 – The reference frames placement for the robotic platform (Kuka LWR).

Table 7.1 – Motion parameters for each region of the gaze tracker screen. Time (s) and lengths (mm).

	Pan Only Planner Parameters	Pan and Zoom in Simultaneously Planner Parameters	Condition
Static	$L_d = 0.0$ $L_z = 0.0$ $t_f = 0.010$	$L_d = 0.0$ $L_z = 5.0$ $t_f = 0.500$	$r < r_1 = 0.18$
Medium	$L_d = 5.5 L_z = 0.0$ $t_f = 0.325$	$L_d = 6.0 L_z = 5.0$ $t_f = 0.600$	$r_1 \leq r < r_2 = 0.38$
Fast	$L_d = 7.0 L_z = 0.0$ $t_f = 0.300$	$L_d = 8.0 L_d = 5.0$ $t_f = 0.500$	$r \geq r_2$

the user's eye to account for their visual axis (VA) (*i.e.,* the PoR, or the line joining the cornea and the fovea centre) [52]. This process typically requires the user to gaze at a moving spot presented on the screen. During this procedure a set of the user's pupil coordinates at a number of predetermined locations, also known as calibration points, are recorded. The PoR of the user is then estimated through forming a mapping, which relates the captured pupil coordinates to their respective screen coordinates. It has been found in numerous cases that such calibration methods drift over time [59, 112], thus affecting the accuracy and precision of the PoR, potentially requiring the surgeon to recalibrate during surgery. This would lead not only to poor user-experience but also, more critically, would raise safety related reservations in using the gaze contingent robotic laparoscope system in the operating theatre.

To overcome this problem, an online calibration process that seamlessly and progressively adapts to the changing user's gaze is introduced. Since the proposed online calibration process replaces the conventional offline calibration, the surgeon will be able to use the robotic laparoscope system immediately. Furthermore, the adaptive nature of the algorithm overcomes the calibration drift as it updates with continued use thus, allowing the surgeon to use the gaze contingent system for longer periods without recalibrating during an operation. The online calibration algorithm takes advantage of the pre-learnt gaze gesture information to extract calibration points in an ongoing manner to form and update the mapping function.

The proposed online calibration algorithm illustrated in Fig. 7.4 shares many similarities with the gaze gesture recognition process introduced in Fig. 6.4 as it integrates with the original gaze gesture recognition process. Since the nature of the gaze gestures requires the user to look at the centre and at one of the bottom corners, the online calibration algorithm uses these coordinates to populate the subject-specific calibration mapping on the fly by extracting the pupil locations at these instances.

Referring to Fig 7.4, the online calibration algorithm works by first taking the stream of unmapped optical axis based PoR coordinates from the gaze tracker and median filtering them to reduce noise. Following this, potential gaze gestures are extracted in the same manner as the gaze gesture recognition process shown in Fig 6.4. It is important to emphasise that, at this stage, the gaze gesture sequence consists of a series of inaccurate uncalib-

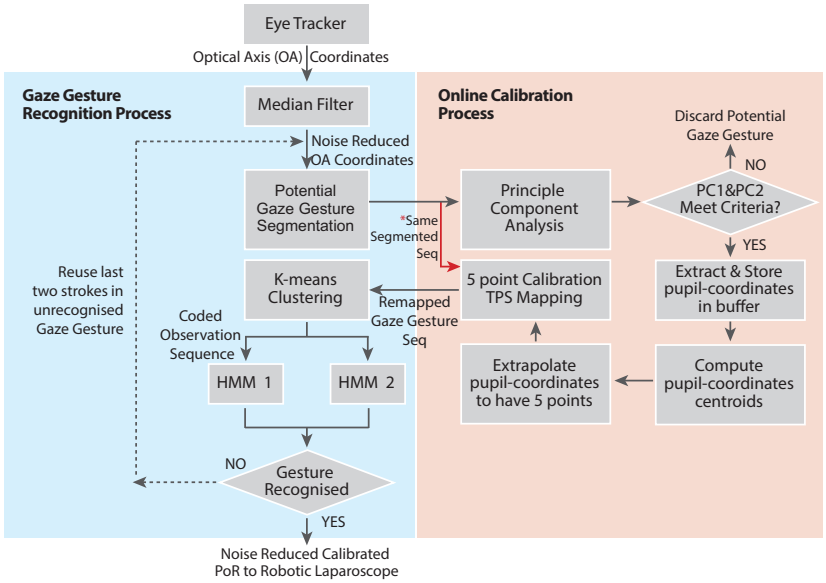


Figure 7.4 – Flow chart of the online calibration algorithm.

rated PoR coordinates. Hence, a technique is needed to distinguish if a potential gaze gesture is of a plausible trajectory and orientation. Principal components analysis is applied to the segmented potential gaze gesture. Since the majority of the trajectory's information is contained in the first and second principal component (PC1 and PC2 respectively), and with the additional knowledge that the gaze gestures here are of an elongated diagonal, when the PC1/PC2 ratio is above a threshold of 5, the false gaze gestures are filtered out. The angle of PC1 indicates the quadrant location of the potential gaze gesture, and can accordingly distinguish whether the segmented gaze gesture is associated to HMM1 or HMM2. If PC1 lies on the forth quadrant, then it is related to HMM1 ('*activate camera*' mode), while if it lies on the third quadrant then it is associated to HMM2 ('*tilt camera*' mode).

If the principal component criteria are met, then a series of pupil coordinates during two fixation points, *i.e.,* the coordinates associated with the calibration point at the screen centre and the coordinates associated with one of the calibration points at the bottom corner, are extracted from the gaze gesture. Here, only the pupil coordinates during the period when the velocity is below $300^{\circ}s^{-1}$ are stored into an array buffer. There are three

array buffers, one for each stream of calibration point coordinates associated with gazing at the centre, bottom left and bottom right corner of the screen. To enable the online calibration algorithm to adapt to potential calibration drifts, the array buffers store only the most recent five streams of calibration coordinates for each of the three calibration point locations. If the array buffer already has five streams of calibration coordinates, the first stream is displaced and the previous four steams are shifted along where the new incoming stream of calibration coordinates are added at the end of the array buffer. Therefore buffers are also indexed to distinguish the start and end of each respective calibration point coordinate stream in each respective array buffer.

The contents of the array buffers are filtered subsequently by computing the centroid and standard deviation of each calibration point. The final pupil centroid *(i.e.,* calibration points) to be used in the mapping function is recomputed excluding any coordinates that fall outside one standard deviation from the initial computed centroid. This is done to avoid potential noise in the gaze gesture. These centroids are to be used in the calibration mapping to map the OA to the VA.

The calibration mapping incorporated in the algorithm is a thin plate spline (TPS) based radial basis function (RBF) mapping. TPS, a special polyharmonic spline, is chosen for the gaze calibration mapping due to its elegant characteristics to interpolate surfaces over scattered data. The TPS was first introduced by [40], and has previously been used for various computer vision applications such as in image registration [10]. Given a vector $\boldsymbol{x} \in \mathbb{R}^m$, a RBF is a real-valued function whose value depends only on the distance from the origin, so that $\phi(\boldsymbol{x}) = \phi(\|\boldsymbol{x}\|)$, where $\|.\|$ denotes the Euclidean norm. Any function that satisfies this property is a radial function. Sums of RBFs are commonly used to approximate functions. They take the form,

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i \phi(\|\boldsymbol{x} - \boldsymbol{c}_i\|). \tag{7.3}$$

In Eq. (7.3), the approximating function $f(\boldsymbol{x})$ is represented as a sum of n radial basis functions, each associated with a different centre $\boldsymbol{c}_i$, and weighted by an appropriate coefficient ω_i. This function defines a spatial mapping that maps any location x in space to a new location $f(\boldsymbol{x})$. In order

to use the TPS as the basis function, the TPS takes the form $\phi(\boldsymbol{r}) = \boldsymbol{r}^2 ln(\boldsymbol{r})$, where $\boldsymbol{r} = \|\boldsymbol{x} - \boldsymbol{c}_i\|$.

In order to guarantee the unique existence of interpolants with a small variation to Eq. (7.3) lower order polynomials are added to $f(\boldsymbol{x})$ along with some extra conditions. Specifically, for the TPS, we obtain,

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i \phi(\|\boldsymbol{x} - \boldsymbol{c}_i\|) + a + \boldsymbol{b}^T x, \tag{7.4}$$

where, $\boldsymbol{x} \in \mathbb{R}^m$, a is a constant and $\boldsymbol{b} \in \mathbb{R}^m$. This gives rise to unique existence of interpolating $f(\boldsymbol{x})$ using the TPS basis function $\phi(\boldsymbol{r}) = \boldsymbol{r}^2 ln(\boldsymbol{r})$ with its value at origin declared to be zero. In addition, the uniqueness can be guaranteed provided that $\boldsymbol{c}_i$ are not collinear and the following conditions are fulfilled:

$$\sum_{i=1}^{n} \omega_i = 0 \text{ and } \sum_{i=1}^{n} \omega_i \boldsymbol{c}_i = (0, 0, ..., 0)^T. \tag{7.5}$$

Here, $(0, 0, ..., 0)^T$ denotes the zero vector in m dimensions. Under these conditions, the scalars ω_i, a and the vector $\boldsymbol{b}$ can be uniquely solved for.

To apply the TPS basis function as the gaze calibration mapping, the pupil coordinate centroids, $\boldsymbol{P} = (x, y)$ is defined as the input, thus $m = 2$. The mapping to the screen coordinates $f(\boldsymbol{P})$ is represented as,

$$f(\boldsymbol{P}) = \sum_{i=1}^{n} \omega_i \phi(\|\boldsymbol{P} - \boldsymbol{c}_i\|) + a + \boldsymbol{b}\boldsymbol{P}, \tag{7.6}$$

where $\boldsymbol{c}_i = (\check{\boldsymbol{x}}_i, \check{\boldsymbol{y}}_i)$ is the n eye feature vectors obtained during calibration, and $\boldsymbol{b} = [a_x, a_y]$. The conditions for the scalars ω_i, a, and the vector $\boldsymbol{b}$ to be solved uniquely become,

$$\sum_{i=1}^{n} \omega_i = 0 \text{ and } \sum_{i=1}^{n} \omega_i \check{\boldsymbol{x}}_i = 0 \text{ and } \sum_{i=1}^{n} \omega_i \check{\boldsymbol{y}}_i = 0. \tag{7.7}$$

Since the mapping in Eq. (7.4) specifies an approximation function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, in order to describe an eye tracking calibration mapping in $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, two TPS functions are used; one for the x-axis of the screen $f^x(\boldsymbol{P})$ and one for the y-axis of the screen $f^y(\boldsymbol{P})$. These mapping functions share their calibration points, $\boldsymbol{c}_i = (\check{\boldsymbol{x}}_i, \check{\boldsymbol{y}}_i)$ to give,

$$M(\boldsymbol{P}) = \begin{bmatrix} f^x(\boldsymbol{P}) \\ f^y(\boldsymbol{P}) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} \omega_i^x \\ \omega_i^y \end{bmatrix} \phi(\|\boldsymbol{P} - \boldsymbol{c}_i\|) + \begin{bmatrix} a^x \\ a^y \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}^x \\ \boldsymbol{b}^y \end{bmatrix} \boldsymbol{P} \quad (7.8)$$

Here ω_i^x, a^x and $\boldsymbol{b}^x$ represent the scalar coefficients and vector for the x-axis, and ω_i^y, a^y and $\boldsymbol{b}^y$ represent the scalar coefficients and vector for the y-axis. The resultant mapping function $M(\boldsymbol{P})$ now describes the mapping of the pupil coordinates in the eye image plane to the screen pixel coordinate plane. The overall TPS calibration mapping is divided into two parts; the sum of the RBF weighted by the TPS coefficients ω_i^x and ω_i^y, which are bounded and asymptotically flat, and an affine part described by the last three terms.

Three coefficients are required to express an affine transform, which means a minimum of three calibration points are needed to compute the TPS mapping. However, the gaze contingent laparoscope system utilises two gaze gestures, giving room to obtain between two and three calibration points only. Therefore, prior to using the calibration points for the online mapping, the final calibration points associated with the centre, bottom left and right corners of the screen are extrapolated to increase the number of calibration points to five. Note that on initialisation of the system, there can only be two calibration points obtained from the first gaze gesture received from the user, *i.e.,* either a '*activate camera*' or '*tilt camera*' gaze gesture. Therefore, symmetry along both the vertical and horizontal eye rotation is assumed and the two calibration points are extrapolated to five calibration points, when the first gaze gesture has been successfully performed. This scenario is illustrated in Fig. 7.5(a)-(b). When both gaze gestures have been performed, then three calibration points are used to extrapolate to five calibration points as shown in Fig. 7.5(c). A conventional offline calibration procedure uses anything between five to nine calibration points to establish a gaze mapping function for accurate PoR estimation.

Once five calibration points have been extrapolated, the parameters including ω_i^x, ω_i^y, a^x, a^y, $\boldsymbol{b}^x$ and $\boldsymbol{b}^y$ can be obtained by solving the following linear system,

$$\begin{bmatrix} \boldsymbol{S}^T \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L} & \boldsymbol{C} \\ \boldsymbol{C}^T & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{t}^x & \boldsymbol{t}^y \end{bmatrix}, \quad (7.9)$$

where $\boldsymbol{S}$ is the five calibration point's respective screen coordinates,

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}^x \\ \boldsymbol{s}^y \end{bmatrix}^T = \begin{bmatrix} s_1^x & s_2^x & \cdots & s_5^x \\ s_1^y & s_2^y & \cdots & s_5^y \end{bmatrix}^T, \tag{7.10}$$

which are also illustrated in Fig. 7.5(d). The calibration points' respective pupil feature points represented by $\boldsymbol{c}_i = (\check{\boldsymbol{x}}_i, \check{\boldsymbol{y}}_i)$ form the array $\boldsymbol{C} = (1, \check{\boldsymbol{x}}_i, \check{\boldsymbol{y}}_i)$ which is a 5×3 matrix, $\boldsymbol{L} = \phi(\|\boldsymbol{c}_j - \boldsymbol{c}_i\|)$ is a 5×5 matrix, $\boldsymbol{O}$ and $\boldsymbol{0}$ are 3×3 and 3×2 zero matrices respectively. $\boldsymbol{t}^x$ and $\boldsymbol{t}^y$ are arranged as,

$$\boldsymbol{t}^x = \begin{bmatrix} \omega_1^x & \cdots & \omega_5^x & a^x & \boldsymbol{b}^x \end{bmatrix}^T \text{ and } \boldsymbol{t}^y = \begin{bmatrix} \omega_1^y & \cdots & \omega_5^y & a^y & \boldsymbol{b}^y \end{bmatrix}^T. \tag{7.11}$$

Both $\boldsymbol{t}^x$ and $\boldsymbol{t}^y$ are 8×1 vectors. By further denoting,

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{L} & \boldsymbol{C} \\ \boldsymbol{C}^T & \boldsymbol{O} \end{bmatrix}, \tag{7.12}$$

we can invert the system in Eq.(7.9) to solve for the scalar and vector parameters $\boldsymbol{t}^x$ and $\boldsymbol{t}^y$ as follows,

$$\begin{bmatrix} \boldsymbol{t}^x & \boldsymbol{t}^y \end{bmatrix} = \boldsymbol{K}^{-1} \begin{bmatrix} \boldsymbol{S}^T \\ \boldsymbol{0} \end{bmatrix}. \tag{7.13}$$

Once a calibration mapping is formed, this is used to remap the previously segmented gaze gesture. The same segmented gaze gesture is remapped via the calibration mapping and tested against the two gaze gesture HMM models. If the gaze gesture returns an inference value above either of the two HMM thresholds, the extracted pupil coordinates are deemed accurate to be used in the calibration mapping and are kept in the respective buffer arrays as calibration points, otherwise they are discarded. As the user continues to use the gaze contingent laparoscope system and inputs gaze gestures, the stored running buffer of five streams of pupil coordinate points in each three respective calibration point locations are used to build a more robust gaze tracking calibration mapping, whilst also accounting for any calibration drift as the user moves their head. The online algorithm therefore can potentially lead to an improved and sustained accurate PoR estimation. The overall online calibration algorithm integrates closely with the gaze

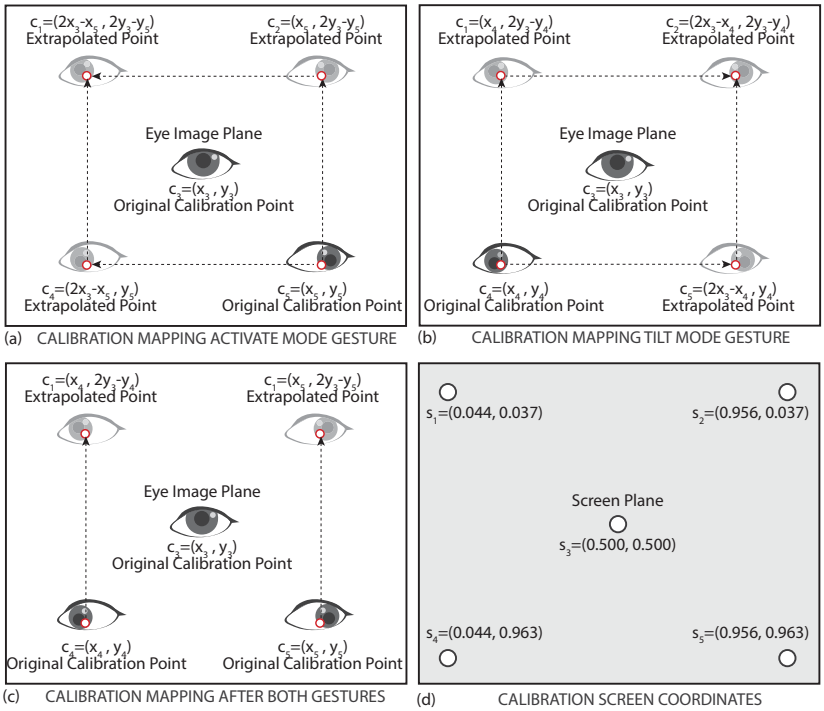


Figure 7.5 – (a) Calibration point extrapolation process when a '*activate camera*' gaze gesture is performed on initialisation of the system. (b) Calibration point extrapolation process when a '*tilt camera*' gaze gesture is performed on initialisation of the system. (c) Resultant calibration mapping showing the extrapolated and original calibration points once both gaze gestures have been performed. (d) The five screen coordinate locations used for the TPS calibration mapping.

gesture recognition process, thus enabling the surgeon to seamlessly start using the robotic laparoscope system without having to perform an offline gaze tracker calibration. The next section will explain the implementation of the system, including the gaze contingent control user interface in further detail.

7.3 Implementation

In this section, the UI of the three control modes are described:

1. Gaze gesture control; gesture based activation and camera control

with their PoR and head position.

2. Pedal activated control, where the user activates the control mode of the camera with a dual-switch foot-pedal and directs the camera with their PoR and head position.

3. Camera assistant control, a camera assistant follows verbal instructions of the experimental participant and navigates the camera.

Following this, the experimental setup and protocols are described in detail. The experimental protocol consists of two subsections. The first explains the usability experiments of the robotic laparoscope system whereas the second covers the performance study of the online gaze tracker online calibration.

7.3.1 Gaze Contingent Control Modes

In this chapter, a modified gaze gesture control UI is implemented to introduce tilt control of the camera, to allow simultaneous zooming in and panning, and to have a camera stopping method that is not at the centre of the screen. The modified UI is illustrated in Fig. 7.6. The system UI enables pan, zoom and tilt of the laparoscope camera. The pan and zoom control is combined into one control mode, *i.e.*, '*activate camera*' mode.

On system initialisation, the camera is stationary and the system waits for a gaze gesture input from the user. The user has the option to control the camera via '*activate camera*' or '*tilt camera*' modes. '*Activate camera*' mode enables panning or zooming and switching between the panning and zooming is enabled with a movement of the head forward or backward. On initialising the '*activate camera*' control mode, the distance from screen to the user's eyes is calculated and stored as the '*original distance*'. During this control mode, the camera will pan in the vector direction of the PoR from the screen centre with a speed in accordance to Table. 7.1 as long as the user's head is within a region of ±5cm from the *'original distance*'. More specifically, in the panning mode, the screen area is separated into three radial speed regions as shown in Fig. 7.6(a). Gazing within each region would move the camera at a different velocity accordingly. If the surgeon's PoR is within the central screen region, the camera maintains stationary allowing for a stable view whilst performing tasks. If the surgeon's gaze falls within the medium and fast regions, the camera moves at a velocity of

$16.9mms^{-1}$ and $23.3mms^{-1}$ respectively. These three speed regions were introduced to enable user-friendly control of the camera whilst maintaining a predefined maximum velocity that would be safe for the patient. Thus, even if the surgeon performs a glance towards the edge of the screen, the camera still follows the predefined safe camera speed. In order to zoom in, the user is required to simply lean their head forward by more than $5cm$ from the '*original distance*' position while the '*activate camera*' mode is active. This control scheme is illustrated in Fig. 7.6(b). Whilst zooming in, the gaze can also direct the camera to enable simultaneous zooming and panning, again with the same speeds as in Table. 7.1. Conversely, if the user desires to zoom out during the '*activate camera*' mode, the user is asked to lean their head back by $5cm$ from the '*original distance*' position as shown in Fig. 7.6(c). During zooming outwards, panning via gaze direction is disabled.

In the '*tilt camera*' mode, the system allows the camera to rotate the view around the laparoscope's longitudinal axis. You can activate this mode by performing the '*tilt camera*' gaze gesture. Once this mode is activated, the original positions of the left and right eye are recorded and the horizon going through the left and right eye is calculated and stored as the '*base line*'. Through tracking the position of the left and right eye and calculating the relative angle they make with this '*base line*', the camera tilt can be controlled. In order to tilt the camera left or right, the user is instructed to tilt their head by more than $\pm 15°$ to the left or right respectively. The robot performs incremental rotation steps with an angular velocity of $5°s^{-1}$ in the respective tilt direction. This can be understood from Fig. 7.6(d). In order to maintain a tilt in one direction, the user simply maintains their head position *i.e.,* a tilt of the head greater than 15° will maintain the camera tilt motion in that given direction. While using the system, the user can see their PoR represented on the screen as a white moving dot, which can optionally be deactivated. During each control mode, guidance text is overlaid on to the camera view and the camera can also be stopped by looking at the '*stop camera*' text at the bottom left hand corner of the screen for $750ms$ as shown in Fig. 7.6(e).

The pedal activated control mode has a similar UI to that of the gaze gesture control mode except that the '*activate camera*' and '*tilt camera*' modes are activated via a pedal, thus the guidance text is not observable on the

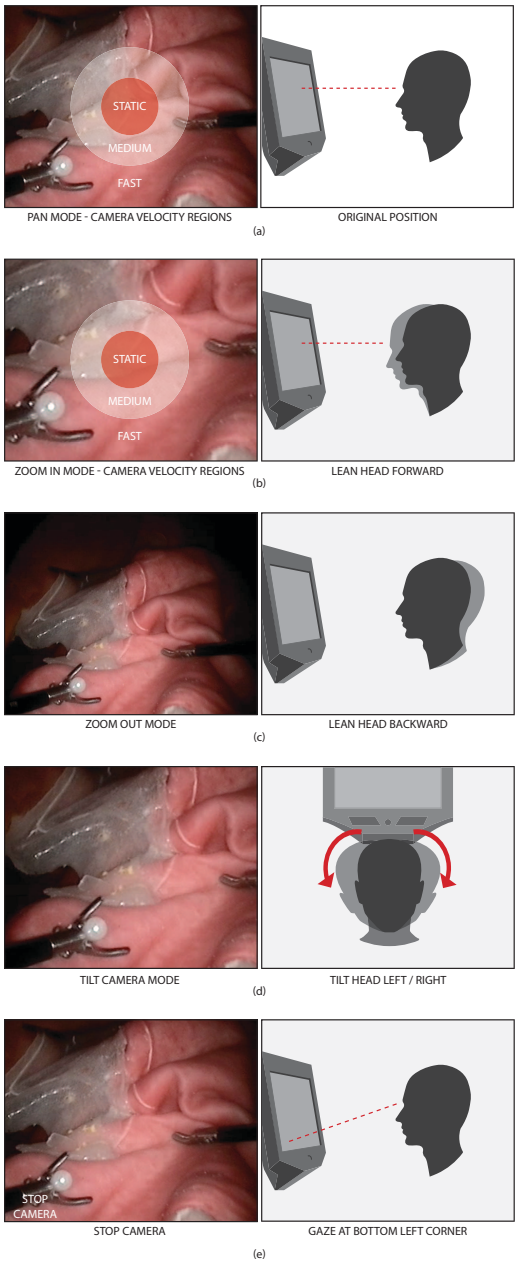


Figure 7.6 – Illustration of the UI for the gaze contingent laparoscope system. (a) Panning during '*activate camera*' mode. (b) Zooming in during '*activate camera*' mode. (c) Zooming out during '*activate camera*' mode. (d) Tilting the camera during '*tilt camera*' mode. (e) Camera is stopped by gazing at the bottom left hand corner for $750ms$.

screen. The user is asked to use a dual-switch foot-pedal to activate each respective camera control mode. The left and right foot-pedal activates the '*tilt camera*' and '*activate camera*' mode respectively. Once a camera mode is activated by depressing the pedal, the pedal needs to be kept depressed to maintain operating in the chosen control mode. The laparoscope is navigated in exactly the same manner as shown in Fig 7.6(a)-(d). Note that during pedal activation mode, there is no '*stop camera*' text displayed at the bottom left corner of the screen as the camera stops by releasing the foot-pedal.

Lastly, during the camera assistant mode, an assistant operator manually controls the laparoscope camera according to the surgeon's verbal guidance.

7.3.2 Experimental Setup

The experimental setup is illustrated in Fig. 7.7. It uses the following same equipment: a Tobii 1750 remote gaze tracker, a LWR [3], a $10mm$ zero degree Karl Storz laparoscope, a Storz Tele Pack light box and camera, two Storz Matkowitz grasping forceps, a dual-switch foot-pedal and an upper gastrointestinal phantom with simulated white lesions. Additionally to the experiment setup in Chapter 6, surgical instrument tip trajectory data in Cartesian space was recorded using an NDI Polaris infrared tracker together with reflectance markers attached onto both instrument handles.

The HMM gaze gesture recognition process, the online gaze tracker calibration algorithm and the robot control process were implemented in C++. The gaze based algorithms runs at 33.3Hz and the robot control algorithm updates at 200Hz. Experimental data, which consisted of subject PoR, gaze gestures, camera-view feed, the laparoscope camera tip position in Cartesian space, were recorded at a rate of 33.3Hz. Here the camera tip position was obtained from the robot's forward kinematics. The NDI Polaris tracker which logged data at 17Hz, the laparoscope camera and eye tracker were synchronised on one PC. This PC was also synchronised via the UDP protocol with the LWR PC. Instrument trajectory tracking was undertaken as peer-reviewed literature has shown that instrument trajectory information including, path length and speed correlate to the level of surgical performance [60]. Thus, the use of the same group of surgeons during the studies, will reflect the usability of the system. (During the camera assistant control

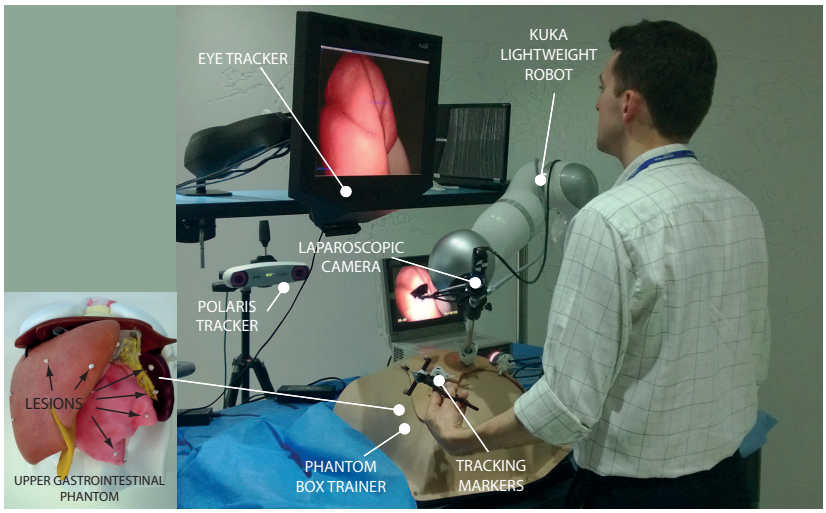


Figure 7.7 – The experimental setup for the modified gaze contingent laparoscope system uses equipment from the previous chapter, but in addition an NDI Polaris infrared tracker together with reflectance markers attached onto both instrument handles to record the subject's instrument tip trajectory data.

mode, a third infrared reflectance marker was mounted onto the laparoscope and tracked via the NDI Polaris tracker.)

7.3.3 Experimental Protocols

The experimental protocol consists of two subsections. The first describes the usability experiments of the robotic laparoscope system whereas the second explains the performance study of the online gaze tracker online calibration. Each subsection covers the quantitative performance metrics used to analyse the collected results from each respective study.

7.3.3.1 Robotic Laparoscope System Usability Study

In this usability study, seventeen surgical residents with a postgraduate year between 3-7 (PGY3-7, male=16, female=1) were recruited. The mean laparoscopic experience was 676 (±293) cases. All participants were trained to use the gaze gesture and pedal activated systems on an abstract navigation task before starting the study. This was performed to mitigate against the

potential learning effects when performing the subsequent phantom based task. The abstract training task participants performed is illustrated in Fig. 7.8(a). It required the subject to navigate the laparoscope system inside a conventional box trainer to locate numbers in ascending order. The numbers of varying font sizes were placed randomly on a 4×5 grid in order to require the user to both pan and zoom during the training. Subject training was halted when: a minimum baseline proficiency task completion time was met; when they showed no further improvement in completion time; and, when they could reproduce a similar completion time consecutively on three occasions. A second training task was used for the novel tilt control modality of the camera. Subjects were asked to re-align three operative scenes to a conventional anatomical orientation. Each of the three operative scenes shown in Fig. 7.8(b) were displayed one at a time, and were laid out so that the scene is shown across the whole camera FOV. In order to re-align the scene, the user tilted the camera left, right, then left by 15°, 65° and 35° respectively. Once a scene was correctly re-aligned, the next scene was presented to the participant.

The validation task involved subjects identifying and biopsying a set number of randomly placed lesions on an upper gastrointestinal phantom. The task was a simulated upper gastrointestinal staging laparoscopy (the same task as that conducted in Chapter 6) and the phantom was placed in a laparoscopic box trainer. The task required subjects to use a bi-manual technique, typically, one instrument manipulating and/or retracting tissue, and the other grasping and removing the lesion. The surgeons were encouraged to physically look at the phantom model, before lesions were placed, to familiarise themselves with it. This was introduced to minimise the potential confounding factor of learning the phantom used during the task. Participants were asked to conduct the lesion removing task twice, for each of the three camera control modes: i) gaze gesture activation, ii) foot-pedal activation, and iii) verbal communication with a human camera assistant. Each participant overall performed the lesion removal task six times in total. To mitigate learning effects on performing the lesion removing task, the sequence in which subjects performed the task during the three control modes were randomised. The same human camera assistant was recruited throughout the study for the camera assistant mode.

Subject gaze data were recorded during the gaze gesture control mode to

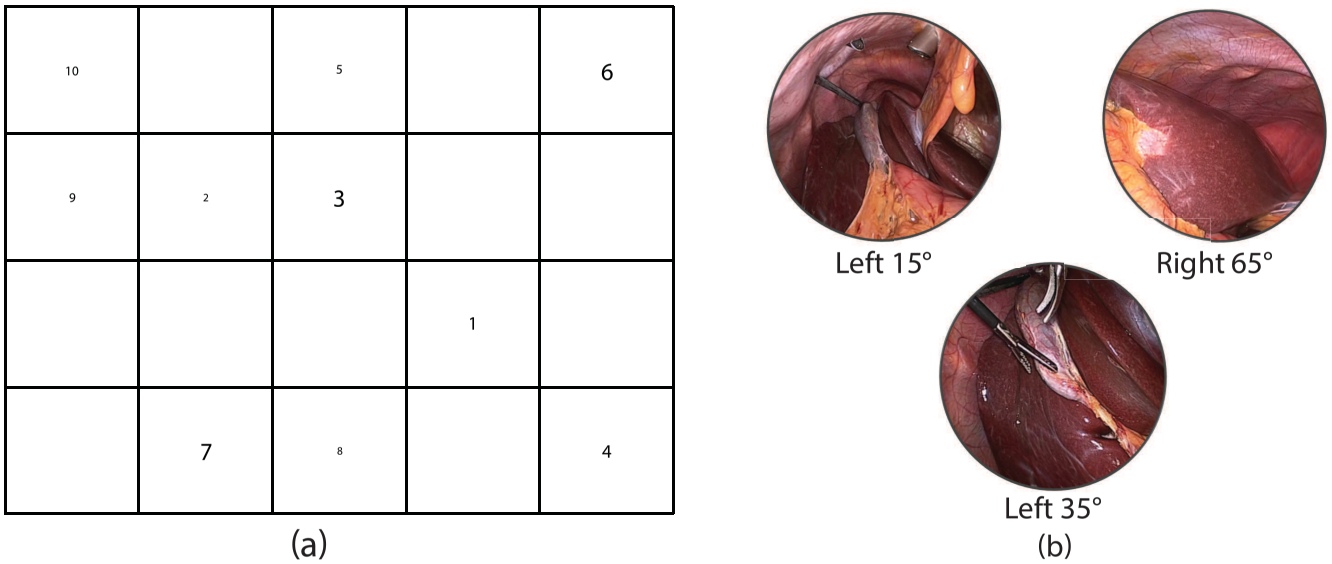


Figure 7.8 – (a) Panning and zooming training task where participants had to control the robotic laparoscope to navigate within a box trainer to find numbers on a 4×5 grid in ascending order. (b) The rotation task involved the participants tilting the camera view left or right to correctly orientate the camera.

assess the usability of the gaze gesture algorithm. Usability was assessed from post-hoc observation of the recorded camera-view videos where the occurrences of true positive, false positive and false negative gaze gestures were counted. The number of true negative gaze gestures were obtain by filtering the recorded gaze data and counting the rejected potential gaze gestures during the trials. The recognition accuracy is obtained by Eq. (6.11) and the false positive rate is obtained by Eq. (6.12).

The system usability during each laparoscope control mode is assessed via the following quantitative performance metrics:

- Task completion time measured in seconds.
- Camera path length measured over a single trial in centimetres.
- Camera workspace measured in centimetre cube - an empirical assessment of whether the group of participants were able to move the laparoscope system over a comparable workspace to that of a camera assistant.
- Instrument path length measured over a single trial in centimetres - to assess the ergonomics of the system.
- Instrument speed measured in centimetres per second - to measure the ergonomics of the system. This metric is obtained by dividing the camera path length measured over a single trial by the respective task completion time.
- NASA-TLX questionnaire - subjects completed a questionnaire after each lesion removal task.

7.3.3.2 Gaze Tracker Online Calibration Performance Study

Twenty-five subjects participated in the study to assess the performance of the online calibration algorithm independently (male=20, female=5). The aim of this study was to assess whether the online calibration algorithm could calibrate '*on the fly*' with a range of different subjects and maintain a high level of accuracy and precision over time. Furthermore, the study compares the online calibration algorithm performance to when the gaze tracker is not calibrated, and when an offline 5 and 9 point calibration procedure is conducted. All gaze gestures performed were also being recorded

for offline analysis to quantify the recognition accuracy and false positive rate, ultimately to check whether if the online calibration adversely affects the performance of the gaze gesture recognition process.

All participants were trained to use the gaze gesture activated system before starting the study. Each participant was required to successfully perform both gaze gestures ten times. Post training, each participant was asked to perform a calibration performance task under the following eye tracker calibration conditions; i) no calibration, ii) five point offline calibration, iii) nine point offline calibration, iv) after one gaze gesture (online calibration), v) after two gaze gestures (both '*activate camera*' and '*tilt camera*' gesture (online calibration), vi) after five gaze gestures (online calibration), vii) after ten gaze gestures (online calibration). The performance task comprised of the participant observing one by one, nine evenly distributed white dots displayed on a screen. During this, the participants' gaze was recorded for offline analysis.

In order to measure the accuracy of each calibration method, the angular distance of the user's fixation point, θ , from the nine reference points which they were asked to observe is computed in degrees of visual angle via Eq. (2.1). The overall accuracy of the calibration is defined as,

$$\theta_{Avg.Accuracy} = \frac{1}{n}\sum_{i=1}^{n}\theta_i, \tag{7.14}$$

where θ_i is an angular distance of the recorded fixations from each of the reference targets $i = 1, 2, \cdots, n$, where n is the number of reference targets. This is measured in degrees of visual angle. In addition, precision of the calibration is defined as,

$$\theta_{precision} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\theta_j^2}, \tag{7.15}$$

where θ_j, is the angular distance of m successive PoRs within a fixation point *i.e.*, $j = 1, 2, \cdots, m$, and the $\theta_{precision}$ value is computed when the user is observing one of the nine reference point targets. The overall precision is obtained by taking the average between the nine precision measurements *i.e.*,

$$\theta_{Avg.Precision} = \frac{1}{n}\sum_{k=1}^{n}\theta_{precision}^{(k)}, \tag{7.16}$$

where $k = 1, 2, \ldots, n$ and n is the number of reference targets. Precision is also measured in degrees of visual angle.

During the performance study for the gaze gesture based online calibration algorithm, each trial was carried out over twenty minutes. The participant performed one gaze gesture, then the performance task, then was asked to take a five minute break by moving away from the desk. Subsequently, the same procedure was repeated after performing two, five and ten gaze gestures in the same session. The gaze gesture count is accumulated to emulate the subject performing the calibration on the go. The study was executed in this manner to enable understanding of the online calibration's longitudinal performance.

7.4 Results

The results section is divided into two sections. The first covers the assessment of the usability of the new robotic laparoscope system's design. The second analyses the online gaze tracker calibration algorithm's performance.

In order to investigate the usability performance of the new gaze contingent laparoscope system design, the following statistical analysis studies were performed.

1. HMM gaze gesture recognition accuracy and false positive rate assessment.

2. Comparative analysis of the three control modalities.

3. Comparative analysis of the first and second trials assessing learning effects.

4. Comparative analysis of results from study 2. against the system in Chapter 6.

The statistical analysis of the usability tests of the system *i.e.,* (studies 2 - 4), were carried out on the performance metrics mentioned in Section 7.3.3.1.

Table 7.2 – The recognition accuracy and the false positive rate for both HMM gaze gestures.

	Recognition Accuracy	False Positive Rate
HMM1 (‘*Activate Camera*’)	94.88%	1.91%
HMM2 (‘*Tilt Camera*’)	98.08%	0.41%

Assessment of the online gaze tracker calibration algorithm performance involved calculating the accuracy and precision of the gaze tracker online calibration algorithm, as discussed in Section 7.3.3.2. The longitudinal performance of the online calibration was statistically compared to the five point and nine point offline calibration's performances. The gaze gesture recognition accuracy and false positive rate during the trials were also assessed to check that the online calibration was not adversely affecting the performance of the gaze gesture recognition process.

For all statistical analysis, normality tests were initially performed, followed by Mann-Whitney U test for non-parametric continuous variables between modalities. A ‘p-value’ <0.05 was considered significant. Results with significant differences are indicated with an asterisk ‘*’ mark in all tables. Results are represented as medians with IQRs in parentheses, unless otherwise stated. All statistical analysis was performed using MATLAB.

7.4.1 Robotic Laparoscope System Usability Studies

7.4.1.1 Updated HMM Gaze Gesture Performance Assessment

The recognition accuracy and false positive rate results are shown in Table. 7.2. The overall average recognition accuracy for the HMM based gaze gestures is 96.48% and an average false positive rate of 1.17%. These results demonstrates that gaze gestures provide high recognition accuracy and a low false positive rate that makes the use of HMM based gaze gestures both a user-friendly and safe method for the surgeon to convey their intention to control the robotic system.

7.4.1.2 Comparative Analysis of the Three Control Modalities

The comparative analysis of the new gaze contingent control modalities versus with the camera assistant are summarised in Table. 7.3. All seven-



Figure 7.9 – Task completion time by control mode aggregating all participants.

teen subjects met the baseline proficiency and training requirements to be included in the user performance based quantitative analysis. Three of the subjects wore glasses and four wore contact lenses.

The task completion time over the trials were significantly shorter for the gaze gesture activated system compared to both the camera assistant and the pedal activated control scheme with median and IQR values of (190.50 [67.50]s vs 240.50 [118.75]s; p=0.015) and (190.50 [67.50]s vs 246.00 [159.50]s; p=0.014) respectively. In contrast to this, the pedal activated control modality showed no statistical difference in the task completion time versus the camera assistant (246.00 [159.50]s vs 240.50 [118.75]s; p=0.864). The comparative task completion times are shown in the box plot in Fig. 7.9. This result indicates that the overall subject group could complete the same bi-manual task faster with the gaze gesture activated control modality over both the pedal activated system and the camera assistant. This might be attributed to the seamless nature of the gaze gesture activated system as the surgeon can maintain their visual attention on the laparoscope camera view, without for example, having to look down to search for the foot-pedal or to verbally communicate to another human how to move the camera, both of which can induce delay in the task completion times.

Significantly shorter camera path lengths were observed for the gaze gesture and pedal activation modalities compared to the camera assistant with median and IQR values of (97.58 [46.02]cm vs 449.46 [291.12]cm; p=1.83×10^{-12}) and (104.81 [61.13]cm vs 449.46 [291.12]cm; p=2.39×10^{-12})

Table 7.3 – User performance metrics of the assistant, gaze gesture activated and pedal activated system over both trials.

	Task Time (s)	Cam Path Length (cm)	Left Instr. Path Length (cm)	Right Instr. Path Length (cm)	Left Instr. Speed (cms^{-1})	Right Instr. Speed (cms^{-1})	NASA-TLX Score
(1) Assistant	240.50 [118.75]	449.46 [291.12]	1155.71 [634.70]	1250.16 [699.20]	4.75 [1.00]	6.02 [1.89]	32.67 [20.58]
(2) Gaze Gesture	190.50 [67.50]	97.58 [46.02]	730.81 [328.70]	725.93 [336.29]	4.45 [1.67]	5.16 [1.89]	40.00 [28.83]
(3) Pedal	246.00 [159.50]	104.81 [61.13]	868.26 [513.18]	896.05 [449.48]	4.15 [1.59]	4.71 [1.29]	45.34 [27.76]
p-value (1) vs (2)	**0.015***	**1.83×10^{-12}***	**3.30×10^{-5}***	**5.33×10^{-8}***	0.076	**0.009***	0.524
p-value (1) vs (3)	0.864	**2.39×10^{-12}***	**0.034***	**2.90×10^{-4}***	**0.016***	**1.83×10^{-12}***	0.051
p-value (2) vs (3)	**0.014***	0.552	**0.030***	**0.022***	0.727	0.213	0.275



Figure 7.10 – Camera path length by control mode aggregating all participants.

respectively. Note that the camera path length showed no statistical difference between the gaze gesture activated and pedal activated control modes (97.58 [46.02]cm vs 104.81 [61.13]cm; p=0.552). Illustration of the groups' comparative camera path lengths is shown in the box plot in Fig. 7.10. The fact that the gaze gesture activated control mode resulted in a significantly shorter task completion time but not in a significant shorter camera path length, could be attributed to a more ergonomic user experience of not having to depress an external pedal device whilst performing a bi-manual instrument task. The pressing of a pedal can change the surgeon's posture and balance, thus adversely affecting the ergonomics during the task.

In order to assess whether the gaze contingent system is able to cover a similar workspace volume comparable to that of a human camera assistant, the camera tip trajectories from the group of surgeons were combined into one point cloud for each control modality. The point clouds were used to obtain a surface mesh via Delaunay triangulation and the overall volume occupied by the camera tip workspace was then computed using the convex hull algorithm. As can be seen from the illustration in Fig. 7.11, all three camera control methods show a similar workspace volume of $2558.47cm^3$, $2556.40cm^3$ and $2159.60cm^3$ for the camera assistant, gaze gesture activated and pedal activated control scheme respectively. Since the camera workspace occupied by the gaze contingent control system is similar to that of the human camera assistant, together with the significantly shorter cam-

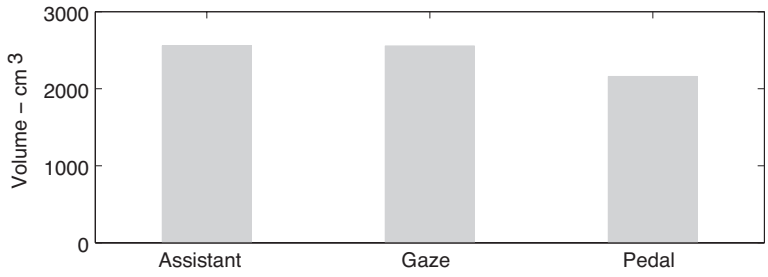


Figure 7.11 – Camera workspace during each control modality.

era path length in the earlier analysis, it indicates that the robotic system is able to provide a more efficient way of navigating the workspace.

The instrument path lengths (both left and right) were significantly shorter during the use of the gaze gesture activated control scheme compared to the camera assistant with (730.81 [328.70]cm vs 1155.71 [634.70]cm; p=3.30×10^{-5}) and (725.93 [336.29]cm vs 1250.16 [699.20]cm; p=5.33×10^{-8}) for the left and right instruments respectively. Similarly, the instrument path lengths were also significantly shorter when using the pedal activated control scheme when compared against the camera assistant with left and right path lengths of (868.26 [513.18]cm vs 1155.71 [634.70]cm; p=0.034) and (896.05 [449.48]cm vs 1250.16 [699.20]cm; p=2.90×10^{-4}) respectively. Importantly, note that the instrument path lengths recorded during the gaze gesture activated control scheme resulted in statistically shorter when compared against the pedal activated camera control scheme with (730.81 [328.70]cm vs 868.26 [513.18]cm; p=0.030) and (725.93 [336.29]cm vs 896.05 [449.48]cm; p=0.022) for left and right instruments respectively. The group's instrument path lengths during each control scheme is shown in the box plot of Fig. 7.12. A shorter instrument path length, which is a reflection of efficient instrument movements, have previously been associated with better surgical performance in the clinical setting [1]. Since the group of surgeons participating in the study remained the same, any changes in the instrument path length can be inferred as an indirect measure of usability.

The instrument speed during the gaze gesture activated control scheme was significantly slower for the right instrument when comparing it against those obtained when having used a camera assistant (5.16 [1.89] cms^{-1} vs

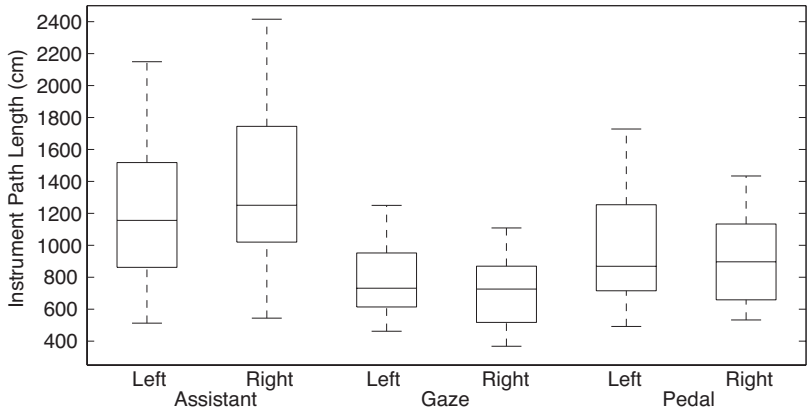


Figure 7.12 – Left and right instrument path length by control mode aggregating all participants data.

6.02 [1.89] cms^{-1}; p=0.009). Interestingly, the left instrument did not result in any statistical difference in speed (4.45 [1.67]cms^{-1} vs 4.75 [1.00]cms^{-1}; p=0.076). The slower right instrument is likely to be attributed to the learning of using the new gaze contingent laparoscope system. Previous literature has shown that slower instrument speeds are associated with lower levels of surgical experience [173]. Since the same group of surgeons performed the lesion removal task for each control modality, the relatively slower right instrument speed during the use of the gaze gesture activated control mode is caused by the usability of the system. Instrument speeds during the pedal activation control scheme were also significantly slower for both left and right instruments compared to the camera assistant trials (4.15 [1.59]cms^{-1} vs 4.75 [1.00]cms^{-1}; p=0.016) and (4.71 [1.29]cms^{-1} vs 6.02 [1.89]cms^{-1}; p=2.66×10^{-5}) respectively. There was no statistical difference in the instrument speeds between the gaze gesture activated and pedal activated control modality for either instrument. Since the slower instrument speeds could be attributed to the learning effect of using the system, further statistical analysis is performed in the next subsection where instrument speeds are analysed longitudinally.

Each control scheme was also assessed for its contribution to the cognitive workload of the participant through the NASA-TLX questionnaire. When introducing new technology into the operating theatre, it is desirable that

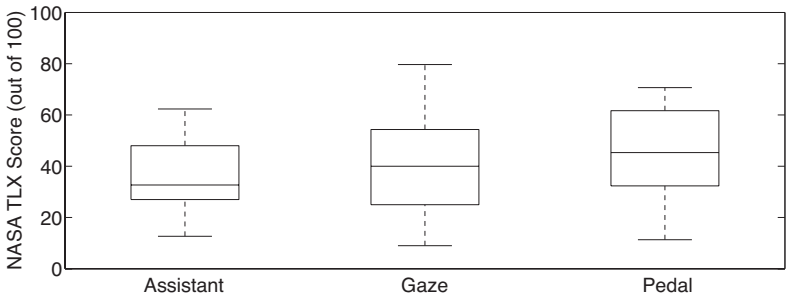


Figure 7.13 – NASA-TLX graph scores for all participants comparing the new gaze contingent UIs with the camera assistant.

the new technology does not add to the cognitive burden of the surgeon. From the results we observe, there was no statistically significant difference in the NASA-TLX score outcome for the new gaze gesture activated control scheme when compared to using a camera assistant (p=0.524) or the foot-pedal activation method (p=0.275). The overall group NASA-TLX scores are shown in the box plot in Fig. 7.13.

7.4.1.3 Comparative Analysis of the First and Second Trial

During the experiment, each participant completed two trials using each control modality. The first trial is separated from the second and a comparison analysis between the first and second trial for every control modality is performed. This comparison was conducted to investigate the learning effect of using the system. Results are summarised in Table. 7.4.

The task completion time showed significant improvement during the second trial in comparison to the first during the use of the gaze gesture activated (p=0.003) and pedal activated (p=0.013) laparoscope control schemes. This indicates that there is a learning effect in using the control schemes. Importantly also, a significant improvement in task completion time is observed when comparing the task completion time during the second trial of the gaze gesture activated control scheme over the camera assistant (p=0.005). However, no significant improvement was observed for the task completion time during the second trial of the pedal activated system over the second trial camera assistant (p=0.361). Moreover, there was no statistical significant change in task completion times between the

Table 7.4 – First and second trial user performance metrics of new laparoscope control scheme (gaze gesture and pedal) compared to the camera assistant.

	Task Time (s)	Cam Path Length (cm)	Left Instr. Path Length (cm)	Right Instr. Path Length (cm)	Left Instr. Speed (cms^{-1})	Right Instr. Speed (cms^{-1})	NASA-TLX Score
(1) Assistant Trial 1	244.00 [112.00]	452.83 [299.33]	1319.55 [870.70]	1571.32 [1001.31]	5.15 [1.18]	6.14 [1.85]	34.33 [27.67]
(2) Assistant Trial 2	237.00 [100.00]	418.08 [200.73]	980.93 [453.92]	1219.14 [485.69]	4.68 [0.68]	5.68 [1.88]	32.00 [18.67]
(3) Gaze Gesture Trial 1	223.00 [76.00]	111.68 [37.93]	882.68 [323.62]	807.08 [363.22]	4.32 [1.64]	4.80 [2.03]	50.67 [36.33]
(4) Gaze Gesture Trial 2	163.00 [48.00]	80.01 [27.19]	683.79 [207.53]	706.00 [305.61]	4.65 [1.58]	5.76 [1.83]	31.67 [22.33]
(5) Pedal Trial 1	300.00 [150.00]	133.57 [72.38]	952.54 [656.80]	971.78 [380.28]	3.66 [0.67]	4.53 [1.39]	52.00 [27.66]
(6) Pedal Trial 2	184.00 [110.00]	94.46 [45.10]	786.99 [202.07]	852.98 [384.30]	4.86 [1.11]	4.91 [1.13]	39.67 [22.00]
p-value (1) vs (2)	0.558	0 .731	0.228	0.535	0.228	0.535	0.642
p-value (3) vs (4)	**0.003***	**0.002***	0.0538	0.302	0.130	**0.039***	0.130
p-value (5) vs (6)	**0.013***	**0.030***	0.2856	0.228	**0.009***	**0.033***	0.113
p-value (1) vs (3)	0.361	**1.00×10^{-6}***	**0.007***	**4.43×10^{-4}***	**0.007***	**0.001***	0.310
p-value (1) vs (5)	0.179	**1.69×10^{-6}***	0.191	0.058	**2.61×10^{-4}***	**4.82×10^{-5}***	0.076
p-value (3) vs (5)	**0.036***	0.705	0.215	0.079	0.535	0.449	0.547
p-value (2) vs (4)	**0.005***	**7.05×10^{-7}***	**0.001***	**3.08×10^{-5}***	0.731	0.705	0.945
p-value (2) vs (6)	0.361	**7.05×10^{-7}***	**0.042***	**0.002***	0.535	0.085	0.309
p-value (4) vs (6)	0.054	0.513	**0.023***	0.1131	0.757	0.228	0.335

first and second camera assisted trials, indicating that the improvement in task completion times were not due to the subjects learning the task. The results indicates that the group was able to quickly learn how to use the gaze gesture activated control scheme and was able to obtain a better task completion time with this new control technique.

The camera path lengths support the hypothesis that a robotic controlled laparoscope is able to provide a more stable and efficient control scheme for the surgeon. This is reflected in the results where the camera path lengths are shorter during both the first and second trials over the camera assistant for the gaze gesture and pedal activated systems as illustrated in Table. 7.4. Interestingly, there is a statistically significant shortening of the camera path obtained during the second trial of using the gaze gesture and pedal activated systems in comparison to their first trials (p=0.002) and (p=0.030) respectively. This is consistent with the improvement in the resulting task completion times during the second trial for both control schemes which shows that participants were able to quickly learn how to use the laparoscope system.

The instrument path lengths resulting from the first and second trials show no significant difference between trials for the gaze gesture activated or pedal activated systems. However, significantly shorter left and right

instrument path lengths were obtained during the gaze gesture activated control scheme when comparing to those obtained during the camera assistant control scheme during both first and second trials as shown in Table. 7.4. Comparatively significantly shorter instrument path lengths were obtained during the second trial pedal activated control scheme versus the camera assistant (p=0.042) and (p=0.002) for left and right respectively. This illustrates the learning effect taking place during the two trials of the gaze contingent control schemes. The fact that the cohort of surgeons were able to obtain a more efficient instrument economy in their second trial potentially indicates the ergonomic benefit of the gaze controlled robotic laparoscope over having a camera assistant to move the laparoscope on their behalf.

The left and right instruments' speed recorded during the gaze gesture and pedal activated control methods in the analysis in Section 7.4.1.2 resulted in statistically slower speeds as discussed earlier (see Table. 7.3). However, when the gaze contingent control scheme's instrument speeds obtained during the first and second trials are compared, a learning effect is observed. For instance, during the first trial of the gaze gesture and pedal activated control schemes, significantly slower instrument speeds are observed compared to the camera assistant mode. By the second trial no significant difference is observed in the instrument speeds during the gaze gesture and pedal activated control methods when compared to the camera assistant. Therefore once the surgical participants learnt the new gaze contingent control schemes, their instrument economy was reaching their normal behaviour (as if they were using a camera assistant). It should be noted also that during the second trial of using the gaze gesture activated control scheme, the cohort of surgeons showed statistically shorter instrument path lengths compared to the camera assistant mode. The combination of significantly shorter instrument path lengths and indifferent instrument speeds during this control scheme (compared to having a camera assistant), highlights the improved instrument efficiency that can be attained.

The NASA-TLX scores showed no significant changes in the scores when conducting intra trial and inter comparisons of the three control modalities. This highlights that the gaze gesture activated control scheme did not add any significant burden on to the group of surgeons.

7.4.1.4 Comparative Analysis against the Gaze Contingent System in Chapter 6

The final usability analysis is a comparison of the system presented in this chapter to that in Chapter 6. The two systems share some similarities: both systems use the same Kuka LWR arm and two gaze gestures to control the camera. Furthermore, the same panning and zooming speeds are used. The main differences between the systems is that the system presented in Chapter 6 separates the pan from the zoom control. In order to switch from panning the camera to zoom, the user would have to stop the camera and then perform a gaze gesture to switch to '*zoom*' control. In contrast, the system UI presented in this chapter combines the pan and zoom control into one 'activate camera' control mode, where the user can switch between the panning and zooming by moving their head forward or backward. In addition, the new system enables an extra '*tilt camera*' control to rotate the camera view along the laparoscope's longitudinal axis.

The study conducted in Chapter 6 had a subject group size of eleven participants with laparoscopic experience of 536 ($\pm$315) cases. This is comparable to the subject group from the user trials presented in this chapter which is seventeen subjects with laparoscopic experience of 676 ($\pm$293) cases. The task completed by subjects was identical in both studies. Since the comparative analysis study of the first and second trial in Section 7.4.1.3 indicated learning effects of using the system taking place, data from the second trials were only used in the analysis here. Statistical p-value tests were performed comparing the two systems and a summary of these results are presented in Table. 7.5.

When comparing the gaze gesture activated system presented in this chapter to the previous one in Chapter 6, a significantly shorter task completion time is observed for the system introduced in this chapter. The pedal activated control method in contrast showed no significant difference in task completion times. This result potentially indicates that the new UI, which enables quick switching between panning and zooming, is more ergonomic when the gaze gestures are used to activate the camera, but is not necessarily the case when the user is required to depress a foot-pedal. The subtle change in the user's balance and posture when having to use another limb to depress the pedal and activate the camera might be the cause of the

Table 7.5 – User performance comparison of proposed system versus the previous gaze contingent system introduced in Chapter 6.

	Task Time (s)	Cam Path Length (cm)	NASA-TLX Score
(1) Current Gaze Gesture	190.50 [67.50]	97.58 [46.02]	40.00 [28.83]
(1) Previous Gaze Gesture	281.00 [172.00]	89.61 [87.05]	41.33 [39.08]
(3) Current Pedal	246.00 [159.50]	104.81 [61.13]	45.34 [27.76]
(4) Previous Pedal	265.00 [160.00]	107.61 [88.42]	27.00 [28.32]
p-value (1) vs (2)	**0.032***	0.913	0.910
p-value (3) vs (4)	0.499	0.411	**0.017***

disparity.

The camera path lengths resulted in no statistical difference for both the gaze gesture and the pedal activated control schemes. This is not surprising as the system's parameters for panning and zooming were kept identical. Therefore, the significantly shortened task time resulting from subject trials during the use of our proposed gaze gesture activated control scheme is likely due to the improved system ergonomics.

The NASA-TLX showed no significant change between the previous and current gaze gesture activated control schemes. Interestingly, the pedal activated control scheme that is proposed in this chapter shows a significant increase in the NASA-TLX score when compared to the pedal system in Chapter 6. This increase in cognitive burden for the cohort of surgeons might have been caused by the restricted posture during zooming in or out. Since the foot-pedal needs to maintain depressed during camera control, the limb used for the foot-pedal is essentially in a fixed position. During zooming in or out since the new system requires moving their head forwards or backwards, the combined UI may have been awkward at times for the surgeon.

7.4.2 Performance of Online Gaze Calibration Algorithm

Although previous gaze tracker systems have achieved a reduced number of calibration points as little as one point, and a good accuracy of 1.0° of visual angle, these systems utilise multiple cameras and or multiple light sources [52], and they still require an offline calibration that is susceptible to calibration drift which was discussed in Chapter 2. The proposed online calibration algorithm can be applied to any remote gaze tracker system as long as there is a user-interactive element to the system application, such as menu navigation with the user's PoR, eye typing or another example might be an automatic scroll mechanism during reading. Once a user interaction is recognised, calibration points can be captured and used to remap the user's PoR.

The accuracy and precision performance of the online gaze tracker calibration algorithm compared to having no calibration and an offline calibration is summarised in Table. 7.6.

From the accuracy performance chart in Fig. 7.14, it is observable that the online calibration has consistent accuracy throughout the trial (after performing one, two, five and ten gaze gestures). Offline calibration methods have previously shown to deteriorate in accuracy performance [112], which would in turn hinder gaze tracking techniques to be applied in the surgical theatre. During the fifteen to twenty minute duration of the trial, the gaze tracker's calibration accuracy is maintained which is a positive attribute. In addition, the accuracy of the online calibration after one gaze gesture is comparable to that of the five point offline calibration (1.048° vs 0.914°; p=0.245) or nine point offline calibration (1.048° vs 0.886°; p=0.085) respectively. The significance test's p-values confirm this, as no significant difference was observed between the online calibration's accuracy after one gaze gesture to that obtained from a five point offline calibration or a nine point offline calibration. Importantly, the statistical comparison tests show that the accuracy of the online calibration becomes better with more gaze gestures; the online calibration accuracy after five gaze gestures and ten gaze gestures shows a statistically significant improvement in accuracy over that obtained after one gaze gesture with accuracy of (0.914° vs 1.048°; p=0.024) and (0.801° vs 1.048°; p=2.37×10^{-5}) respectively. Furthermore, the accuracy after ten gaze gestures shows statistical improvement over that

Table 7.6 – Comparative performance of online, offline and no calibration.

	Avg. Accuracy ± Stdev (Deg Visual Angle)	Avg. Precision ± Stdev (Deg Visual Angle)
(1) No calibration	3.884 ± 2.464	0.255 ± 0.192
(2) Online 1 time	1.048 ± 0.730	0.255 ± 0.146
(3) Online 2 times	0.931 ± 0.574	0.228 ± 0.143
(4) Online 5 times	0.919 ± 0.665	0.224 ± 0.156
(5) Online 10 times	0.801 ± 0.574	0.207 ± 0.126
(6) 9pt offline	0.886 ± 0.512	0.207 ± 0.147
(7) 5pt offline	0.914 ± 0.493	0.214 ± 0.155
p-value (1) vs (2)	**1.26 ×10-48***	0.785
p-value (1) vs (3)	**8.48×10-53***	0.591
p-value (1) vs (4)	**2.38×10-52***	0.939
p-value (1) vs (5)	**3.61×10-56***	0.959
p-value (1) vs (6)	**7.89×10-55***	0.208
p-value (1) vs (7)	**4.21×10-54***	0.878
p-value (2) vs (3)	0.174	0.743
p-value (2) vs (4)	**0.024***	0.840
p-value (2) vs (5)	**2.37×10-5***	0.767
p-value (2) vs (6)	0.085	0.088
p-value (2) vs (7)	0.245	0.710
p-value (3) vs (4)	0.311	0.579
p-value (3) vs (5)	**0.022***	0.549
p-value (3) vs (6)	0.673	0.056
p-value (3) vs (7)	0.882	0.512
p-value (4) vs (5)	0.074	0.945
p-value (4) vs (6)	0.518	0.155
p-value (4) vs (7)	0.198	0.903
p-value (5) vs (6)	**0.006***	0.129
p-value (5) vs (7)	**0.001***	0.878
p-value (6) vs (7)	0.533	0.259

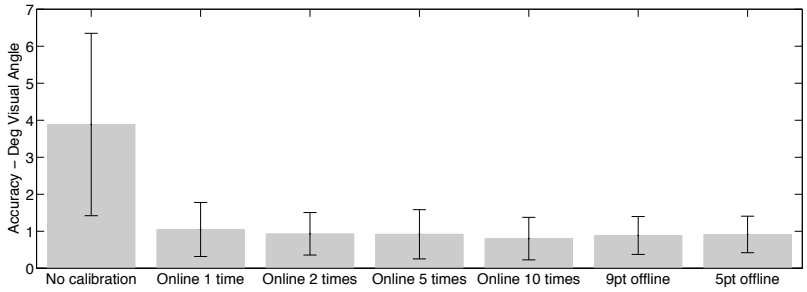


Figure 7.14 – Comparison chart of the accuracy obtained when the gaze tracker has no calibration, the online calibration (after one, two, five and ten gaze gestures are performed) and after the five and nine point offline calibration. The online calibration accuracy performance is comparable to that of the offline calibration with as little as one gaze gesture being performed.

obtained after two gaze gestures (0.801° vs 0.914°; p=0.002) as well as the five point offline (0.801° vs 0.914°; p=0.001) and nine point offline (0.801° vs 0.886°; p=0.006) calibrations. Referring back to Fig. 7.14, as expected, the calibration accuracy with no calibration has a poor accuracy (3.884° ± 2.464°). This is also reflected in the statistical comparison tests against the online and offline calibration methods (see Table. 7.6), where the accuracy improves significantly after the gaze tracker has been calibrated with any offline or online calibration method.

In terms of precision analysis, it is observable from both Table. 7.6 and Fig. 7.15 that the online calibration algorithm maintains the precision throughout the study, which is desirable. In previous literature, it was raised that the precision of offline calibration methods also drifted with time [112]. Here we are able to show from the statistical comparison tests that the online calibration algorithm is able to consistently achieve statistically indifferent precision to those of offline calibration techniques.

The last performance analysis of the online gaze tracker calibration algorithm is investigating the recognition accuracy and false positive rates attained during use of the algorithm. The results are summarised in Table. 7.7. It would be undesirable if the use of the online gaze tracker algorithm lead to a reduction in the recognition accuracy and the false positive rate of the gaze gestures, as this would hinder the user from controlling the

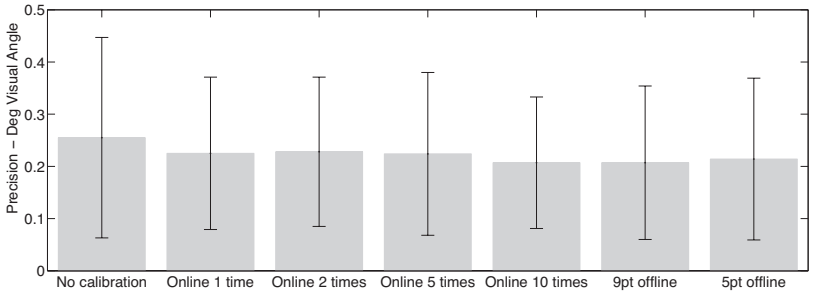


Figure 7.15 – Comparison chart of the precision obtained when the gaze tracker has no calibration, the online calibration (after one, two, five and ten gaze gestures are performed) and after the five and nine point offline calibration. The precision of the gaze tracking is maintained longitudinally via the online calibration algorithm.

Table 7.7 – The recognition accuracy and the false positive rate for both HMM gaze gestures during use of the online gaze tracker calibration algorithm.

	Recognition Accuracy	False Positive Rate
HMM1 ('Activate Camera')	97.64%	0.82%
HMM2 ('Tilt Camera')	95.97%	0.37%

gaze contingent laparoscope. The overall average recognition accuracy for the HMM based gaze gestures is 96.81% and an average false positive rate of 0.60%. These results are comparable to those obtained when the gaze tracker is calibrated offline (shown in Table. 7.2), thus demonstrating that the online calibration algorithm does not affect the usability of the gaze gesture activated laparoscope system.

7.5 Conclusions

In this chapter, we introduce a modified gaze contingent robotic laparoscope which allows for pan, tilt and zoom motion capabilities in Cartesian space. Real-time HMM gaze gestures were used to convey the camera controls, and head motion enabled switching between panning and zooming in or out was implemented.

A novel implicit online gaze tracker calibration algorithm is also intro-

duced and results showed that the algorithm is able to obtain a comparable accuracy and precision to that of a conventional explicit offline gaze tracker calibration after one gaze gesture is performed. The introduction of this online gaze tracker calibration would enable forgoing the need to perform an explicit offline calibration before being able to use the gaze tracker, making the surgical workflow more seamless and efficient. The online nature of the calibration will enable maintaining an accurate and precise PoR estimation, avoiding the calibration drift problem.

In addition, a comprehensive usability study involving seventeen surgical residents was conducted to assess the new gaze gesture activated robotic laparoscope system. During the usability study, instrument trajectories were also tracked along with camera trajectories to enable an improved objective usability assessment. Results demonstrated that once the group of surgeons learnt how to use the system, they were able to perform a surgical navigation task quicker, with a superior camera and instrument efficiency when compared to instructing a camera assistant or when using a pedal activated control scheme. The gaze gestures provide an effective means to convey the surgeon's desired camera control method and the seamless switching between panning and zooming in and out by leaning forward or backward are likely to have contributed to the improved user performance. Although pedals are commonly used in the operating theatre today, having to depress a foot-pedal can change the ergonomics of the operation. Analysis of the camera workspace occupied during the user trials demonstrated that the gaze contingent laparoscope system is able to navigate across a similar working volume to that of a human camera assistant.

The learning effect of using the system was also investigated, and analysis showed that participants were able to learn to use the gaze contingent system quickly and outperform the camera assistant in many of the performance metrics by their second trial. During the first trial, instrument trajectory speeds were slower during the gaze gesture and pedal activated control schemes, showing the group was learning to simultaneously control the camera and perform the bi-manual lesion removing task. However, by the second trial, the instrument speeds were similar to those observed during the camera assistant control scheme, indicating that participants were able to reach their normal instrument economy quickly. Task completion times also improved significantly in the second trial for the gaze gesture activated

control scheme. NASA-TLX scores indicated no signs of cognitive burden to the user when compared to using the conventional camera assistant or pedal activated control. This shows that the group of participants did not feel the gaze gesture activated system to be a complex control scheme.

A comparison of the gaze gesture activated system introduced in this chapter to that of the system presented in Chapter 6 was also conducted. The group of users who used the proposed gaze gesture system in this chapter produced faster task completion times. The pedal activated control schemes did not show any difference in terms of task completion times, but the pedal activated control scheme introduced in this chapter resulted in significantly higher NASA-TLX scores, indicating that the surgical residents felt a higher cognitive burden whilst using it. This is perhaps due to the introduction of using head motion to zoom in and out, which could have caused awkward posture when combined with depressing a pedal to control the camera.

The comprehensive analysis of the results from the various usability studies of the new gaze gesture activated robotic laparoscope system highlights the potential clinical value of the proposed system. The system modifications introduced in this chapter, together with the ergonomic improvements should provide a translatable control scheme for improved camera navigation whether it integrates into a laparoscope or an articulated hyper-redundant robot system.

8 Summary and Conclusions

The surgical community is continuously pushing the technological boundary to improve patient outcomes. Whilst MIS can offer reduced patient trauma and faster recovery times, it has also required longer surgical training for skill acquisition. Subsequently, hospitals and healthcare systems have had to absorb these costs. The rational behind the development of robotic assisted surgical instrumentation is to improve the user ergonomics for the surgeon whilst offering patients all the benefits of conventional MIS. Furthermore, creating a scalable platform to enable faster skill acquisition that improves surgical safety, usability, workflow and reduction of surgical procedure times, will enable cost reduction for hospitals in the longer term. With the future trend pointing towards further miniaturised surgical instruments and perform natural orifice access MIS, there is a need to develop more sophisticated articulated instruments. This is likely to, again, change the safety and ergonomics in the operating theatre raising the need for development of better human-robot interaction methods.

To this end, utilisation of the human visual perception in the form of *Perceptual Docking* has been introduced [175] in order to improve the human-robot interaction technique, where the information of the surgeon's gaze obtained via an eye tracker is used to improve the safety and user ergonomics. A number of applications have been introduced that utilise the human visual perception including, gaze contingent motion stabilisation of beating heart tissue for in-situ tissue visualisation [99, 97] and gaze contingent trajectory planning where the surgeon's gaze fixation path is used during surgery for laser ablation planning [147], which enhances the surgeon's capability during surgery. Gaze contingent motor channelling, a method to dynamically introduce virtual fixture like safety boundaries through the surgeon's 3D PoR was introduced to increase safety during surgery [100]. Gaze contingent instrument placement using the three-dimensional fixation information of the surgeon's gaze to position instruments or control an articulated camera

[109, 111] were introduced to improve the ergonomics and workflow during the operation by utilising the information obtained from the human visual perception. However, there are still limitations and improved ergonomics are desired.

In order to improve the ergonomics in the operating theatre, it is important first to recognise the main issues of MIS procedures. From the study conducted in this thesis, the challenges during endoscopic MIS procedures were identified as the endoscope visualisation's narrow FOV, awkward navigation instances during looping caused by perceptual-motor misalignment, and the generally poor ergonomics. This thesis focuses on finding effective solutions to provide better ergonomics in the operating theatre by minimising these identified problems.

8.1 Contributions of this Thesis

A number of novel ideas, which have been explored to assess and tackle the previously stated problems, were outlined in this thesis. Chapter 4 focused on developing a gaze information based framework to assess a new endoscope camera visualisation, the fade-to-grey DVE, for the purpose of improving the camera visualisation. It was shown that visual attention strategies could be a valuable tool to assess the usability of this new endoscope camera visualization technique. The modelling of eye movements with Markov chains combined with the use of gaze heat-maps, was proven to be effective in comparing different DVE techniques and in highlighting the differences.

Chapter 5 aimed to understand the visual-motor navigation challenges. A method to quantify perceptual-motor misalignment that can occur during looping of a flexible endoscope was introduced and was measured during an articulated robot simulator colonoscopy task. It was identified that during challenging episodes of the colonoscopy, where large varying perceptual-motor misalignment was prevalent, such instances were associated with fewer fixations of longer duration and less saccades. These difficult instances during a task were then inferable from the subject gaze parameters by using a trained SVM classification algorithm. This gaze parameter classifier provides an important quantitative basis for assessing the ergonomics, as well as perceptual behaviour and potential cognitive loadings of devices

used in MIS procedures.

In Chapter 6, a gaze contingent laparoscopic robot system was developed to improve the surgical navigation. Real-time HMM gaze gesture activated camera control was introduced to improve the usability and human-robot interfacing, where the camera control was activated via commands from the operator's eye movements. The proposed hands-free system enabled the surgeon to be part of the robot control feedback loop, allowing user-friendly camera navigation and providing the surgeon the ability to simultaneously navigate and perform a surgical procedure without having assistance from a human camera assistant.

Chapter 7 further expanded the system developed in Chapter 6 with the addition of pan, tilt and zoom capabilities. Further improvements to the usability and ergonomics of the system were introduced by developing a novel online calibration algorithm that was able to obtain a comparable accuracy and precision to that of a conventional explicit gaze tracker calibration. The online calibration scheme allowed the user to immediately start using the gaze contingent system, improving the user experience and the overall surgical workflow. Notably, the online nature of the calibration will enable maintaining an accurate and precise PoR estimation, avoiding the calibration drift problem as discussed in Chapter 2. From the comprehensive user trials involving a cohort of surgeons, it has been shown that the proposed gaze gesture activated and gaze directed laparoscope system could be as effective as a camera assistant with the benefit of faster task completion times and shorter corresponding instrument path lengths with and no significant cognitive burden introduced within the cohort of participating surgeons. The gaze contingent camera control system introduced should provide a translatable control scheme for improved camera navigation whether it is for a laparoscope or an articulated hyper-redundant robotic system. It must be mentioned that although gaze contingent control of instrumentation has been proposed for a number of applications, the main contribution of this thesis is the seamless incorporation of gaze information, which is robust to aberrant eye behaviour and has been proven to be easy to use in a clinical setting.

8.2 Ongoing Work and Future Perspectives

The work presented in this thesis has made attempts to help solve a number of identified problems during MIS. There are however other areas which the work can be further applied to for improvement of ergonomics and usability. Here, we will discuss some future directions to extend the work in this thesis.

The gaze contingent camera control framework can be utilised to control hyper-redundant robots. With the medical engineering research community trending towards less invasive, better cosmetic results and patient outcomes, it is likely that natural orifice access MIS will become the standard and thus more articulated instruments, such as hyper-redundant robots will be required to enable such procedures [162]. The dilemma in using these more complex hyper-redundant robots is the difficulty in controlling their multiple joints, making the elongated robotic structure problematic to manipulate. The gaze contingent control scheme could be applied to control such multi-DoF robots. In Chapter 6 it was demonstrated that the surgeons were capable of controlling the 6 DoF robotic laparoscope through the use of their PoR, thus simplifying the camera control in a user-friendly manner. The next logical step would be to integrate this control strategy to improve the control ergonomics of these emerging articulated robots.

Another area of development would be to extend the work of this thesis and utilise the three-dimensional gaze information obtainable from binocular eye trackers. Previous work by Mylonas *et al.* [98] integrated a binocular eye tracker into a da Vinci surgical console by Intuitive Surgical, Inc. where it was demonstrated how the three-dimensional fixation information of the surgeon could be obtained. The challenge with binocular eye trackers is the longer explicit calibration procedure due to the added depth dimension. There is potential to integrate the implicit online calibration procedure in Chapter 7 to improve the usability of binocular eye trackers as in majority of cases the surgeon would focus their vergence on the tissue surface. Thus, if we can obtain the three-dimensional geometry of the surface tissue, there is potential to use this information as an intrinsic constraint to calibrate the binocular eye tracker. Furthermore, once the binocular eye tracker is implicitly calibrated, the gaze fixation information can be used for Cartesian space positioning of surgical instruments for example, focussed energy delivery. Stoyanov *et al.* [147] previously demonstrated the potential of using the

gaze information to plan ablation trajectories with the surgeon's PoR. The gaze contingent control interface in Chapter 6 could be used to control the energy delivery device with the surgeon in the control feedback loop. This way the surgeon will be able to visualise the progress of the tissue debulking while performing the focused energy delivery. Aberrant eye movements can be removed by median filtering and by also pre-defining how much the robotically controlled energy delivery device can move in the control loop, thus maintaining patient safety.

The work from Chapter 5 can be extended to help detect and classify hesitation and introduce inline decision support. Detection of surgeon hesitation is an unexplored area and an algorithm that can detect hesitation of the surgeon can play an important part in providing decision support for error reduction during training as well as surgical procedures. Given the general trend of the extended longevity projection of the human population, healthcare providers are expected to be increasingly burdened with the amount of services they need to provide. Thus, provision of decision support to surgeons who may be overworked, fatigued and cognitively overloaded will become a key feature in the surgical theatre. Hesitation could originate during surgery due to inexperience during uncertain and unexpected events. For example, it may arise from sudden bleeding or something more common as uncertainty and fear of mistakenly cutting the common bile duct, a common mistake that can lead to complications, during laparoscopic cholecystectomies. The challenge in detection of hesitation is extraction of relevant information from a combination of visual, perceptual and motor information sources for example, gaze, instrument trajectory, heart rate information, to then filter out the essence of hesitation behaviour. Part of the challenge in hesitation classification lies in the potential subject specific variability and uncertainty of whether a classification method is obtainable for a large sample of humans. Furthermore, it is likely that hesitation detection will be a surgical procedure or even task specific, hence the context of the task will needed to be taken into account. The research opportunity for hesitation detection lies in the need to reduce major surgical errors that can lead to surgical complications.

The eye tracking based work in this thesis can also be branched out to develop scene association and clinical decision support particularly in probe-based Confocal Laser Endomicroscopy (pCLE), a relatively new technique

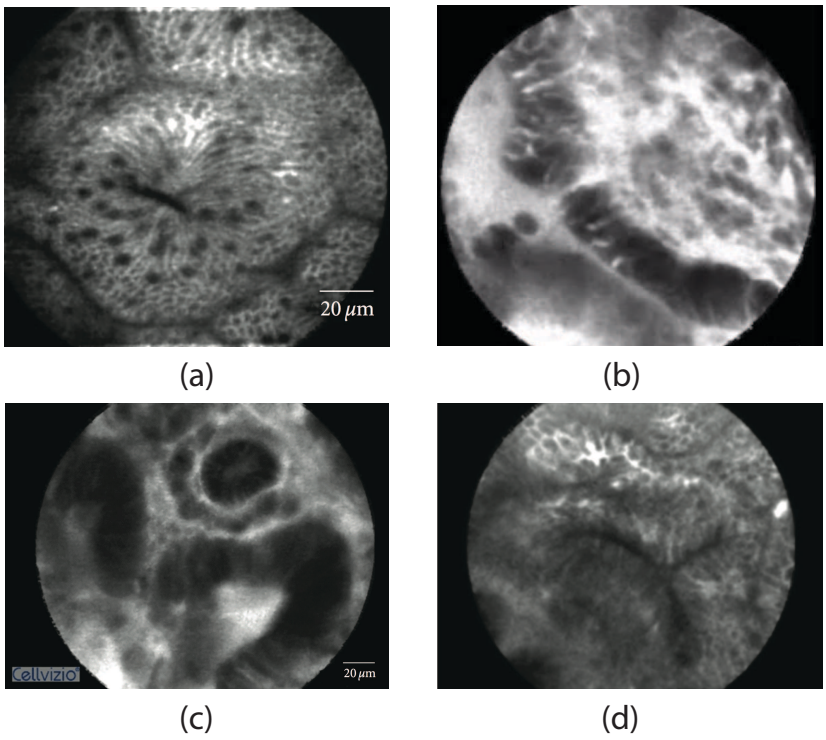


Figure 8.1 – Illustration of the four most common patterns in the colon (a) normal, (b) adenocarcinoma, (c) adenoma, (d) hyperplastic. Images adapted from [157]

in endoluminal imaging. The introduction of pCLE enables the in-vivo visualisation of tissue, which is in contrast to conventional biopsy, which prolongs the treatment process as tissue samples need to be examined after tissue extraction. The Cellvisio system developed by Mauna Kea Technologies offers the possibility to visualise the cellular structure of tissue by inserting a miniaturised confocal laser into the instrument channel of the endoscope. In Fig. 8.1, example images recorded with a Cellvisio system of typical common patterns in the colon namely, normal, adenocarcinoma, adenoma and hyperplastic are illustrated in Fig. 8.1(a)-(d) respectively.

Depending on the structure of the tissue the produced image has different characteristics, which can enable an 'optical biopsy' and allow the endoscopist to make real-time decisions about the nature of the examined tissue. However, good visualisation of tissue, which is paramount to enable

the endoscopist to diagnose with confidence whether an area is malignant or not, is challenging. A content-based image retrieval system, a system that can retrieve images similar to the query, *i.e.,* the image of interest, by comparing that query to a large number of images stored in a database, can prove helpful in this instance. A large number confocal images taken from colon tissue labelled by clinicians as normal, hyperplastic, adenoma and adenocarcinoma are stored in the database. A content-based image retrieval system designed for confocal images and using such a database can be useful to aid endoscopists in establishing the final diagnosis based on similar images retrieved from the system. Most retrieval systems have sophisticated algorithms to encode the content of each image and enable a quick search through a large database. The first step is to find the informative areas, called keypoints, on an image and then create a signature from the surrounding area. Identifying keypoints on an image is often critical for the accuracy of the system and it is a challenging problem, especially when the images are not very descriptive like the images from confocal endoscopy. However, because of the nature of the confocal images no algorithm has been found effective enough in selecting keypoints [165] and the accuracy of the correct retrieved images is maximised when the user selects the keypoints manually. This complicates and delays the procedure and can also be tiresome for the endoscopist. Using eye tracking instead to identify the areas where the subject is focusing, can provide a promising solution to the problem. In a preliminary study conducted, deciding the saliency of the areas of the examined image is based on the density of the fixations *i.e.*, a gaze heat-map, akin to what was previously illustrated in Chapter 5, and its associated areas of interest. It was found that there is a correlation between deliberately selected keypoints and areas of focus, as defined by the density of the fixations. The information obtained via the human gaze holds promising potential and is worth exploring further to enable a gaze based automatic image retrieval system, which can be a useful tool for diagnostic examinations using confocal endomicroscopy and can lead to more wide-spread use of the pCLE technology.

In this thesis, a number of contributions to improve visualisation, navigation and ergonomics during MIS were introduced with detailed methodology and validation to justify the potential clinical value of the work. The findings from the work presented in this thesis can provide ideas to develop

robotic systems which incorporate better human-robotic interfacing protocols that can have a positive impact on both the surgical workflow as well as on patient outcomes. The work presented in this thesis could not have been accomplished without the dynamic and collaborative nature involving practising clinical surgeons, engineers and computer scientists from a multitude of disciplines that can come together under one roof at the Hamlyn Centre for Robotic Surgery at Imperial College London.

Bibliography

[1] R. Aggarwal, T. Grantcharov, K. Moorthy, T. Milland, P. Papasavas, A. Dosis, F. Bello, and A. Darzi. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. *Annals of Surgery*, 245(6):992–9, 2007.

[2] Sylvia Ahern and Jackson Beatty. Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, 1979.

[3] A Albu-Schaffer, S. Haddadin, Ch. Ott, A. Stemmer, T. Wimbock, and G. Hirzinger. The dlr lightweight robot: design and control concepts for robots in human environments. *Industrial robot*, 34(5):376, 2007.

[4] A Albu-Schaffer, C Ott, and G Hirzinger. A unified passivity-based control framework for position, torque and impedance control of flexible joint robots. *The international journal of robotics research*, 26(1):23, 2007.

[5] F. Alnajar, T. Gevers, R. Valenti, and S. Ghebreab. Calibration-free gaze estimation using human gaze patterns. *International Conference on Computer Vision*, pages 137 – 144, 2013.

[6] Joseph C Anderson, Catherine R Messina, William Cohn, Edward Gottfried, Scott Ingber, Gary Bernstein, Eugene Coman, and Joseph Polito. Factors predictive of difficult colonoscopy. *Gastrointestinal endoscopy*, 54(5):558–562, 2001.

[7] Jackson Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2):276–292, 1982.

[8] D. Beymer and M. Flickner. Eye gaze tracking using an active stereo head. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–451–8 vol.2, 2003.

[9] Kenneth R. Boff and Janet E. Lincoln. Engineering data compendium. human perception and performance. 1988.

[10] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:567–585, 1989.

[11] CJA Bowles, R Leicester, C Romaya, E Swarbrick, CB Williams, and O Epstein. A prospective study of colonoscopy practice in the uk today: are we adequately prepared for national colorectal cancer screening tomorrow? *Gut*, 53(2):277–283, 2004.

[12] JC Brooker, BP Saunders, SG Shah, and CB Williams. A new variable stiffness colonoscope makes colonoscopy easier: a randomised controlled trial. *Gut*, 46(6):801–805, 2000.

[13] Caroline G. L. Cao and Paul Milgram. Direction and location are not sufficient for navigating in nonrigid environments: An empirical study in augmented reality. *Presence: Teleoperators and Virtual Environments*, 16(6):584–602, 2007.

[14] Caroline G.L. Cao and Paul Milgram. Disorientation in minimal access surgery: A case study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(26):169–172, 2000.

[15] P. A. Carpenter. Eye fixations and cognitive processes. *Cognitive Psychology*, 8:441, 1976.

[16] O. W. Cass. Objective evaluation of endoscopy skills during training. *Annals of Internal Medicine*, 118(1):73–74, 1993.

[17] Juan J. Cerrolaza, Arantxa Villanueva, and Rafael Cabeza. Study of polynomial mapping functions in video-oculography eye trackers. *ACM Trans. Comput.-Hum. Interact.*, 19(2):1–25, 2012.

[18] Jixu Chen and Qiang Ji. Probabilistic gaze estimation without active personal calibration. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 609–616, 2011.

[19] Daniel Cheng and Roel Vertegaal. An eye for an eye: a performance evaluation comparison of the lc technologies and tobii eye trackers. In *Eye Tracking Research & Application: Proceedings of the 2004 symposium on Eye tracking research & applications*, volume 22, pages 61–61, 2004.

[20] Wu Bin Cheng, Michael AJ Moser, Sivaruban Kanagaratnam, and Wen Jun Zhang. Overview of upcoming advances in colonoscopy. *Digestive endoscopy*, 24(1):1–6, 2012.

[21] B. Cheung and K. Hofer. Eye tracking, point of gaze, and performance degradation during disorientation. *Aviation Space and Environmental Medicine*, 74(1):11–20, 2003.

[22] Dong-Chan Cho and Whoi-Yul Kim. Long-range gaze tracking system for large movements. *Biomedical Engineering, IEEE Transactions on*, 60(12):3432–3440, 2013.

[23] James M Church. Complete colonoscopy: how often? and if not, why not? *American Journal of Gastroenterology*, 89(4), 1994.

[24] N.T. Clancy, G.P. Mylonas, Guang-Zhong Yang, and D.S. Elson. Gaze-contingent autofocus system for robotic-assisted minimally invasive surgery. In *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pages 5396–5399, 2011.

[25] J. Clark, M. H. Sodergren, D. P. Noonan, A. Darzi, and G-Z. Yang. The natural orifice simulated surgical environment (nosse): Exploring the challenges of notes without the animal model. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 19(2):211, 2009.

[26] TN Cornsweet. Visual perception, 1970. *Academic, New York*.

[27] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[28] Peter B. Cotton, Patrick Connor, Daniel McGee, Paul Jowell, Nick Nickl, Steve Schutz, Joseph Leung, John Lee, and Eric Libby. Colonoscopy: Practice variation among 69 hospital-based endoscopists. *Gastrointestinal Endoscopy*, 57(3):352–357, 2003.

[29] F.L. Coutinho and C.H. Morimoto. Free head motion eye gaze tracking using a single camera and multiple light sources. In *Computer Graphics and Image Processing, 2006. SIBGRAPI '06. 19th Brazilian Symposium on*, pages 171–178, 2006.

[30] Flavio L Coutinho and Carlos H Morimoto. A depth compensation method for cross-ratio based eye tracking. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 137–140. ACM, 2010.

[31] Flavio Lvio Luiz Coutinho and Carlos H Morimoto. Augmenting the robustness of cross-ratio gaze tracking methods to head movement. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 59–66. ACM, 2012.

[32] Helen C. Cowley, Alastair G. Gale, and A. R. M. Wilson. Mammographic training sets for improving breast cancer detection. volume 2712, pages 102–112. SPIE, 1996.

[33] S. A. Darzi and Y. Munz. The impact of minimally invasive surgical techniques. *Annu Rev Med*, 55:223–37, 2004. Darzi, Sir Ara Munz, Yaron Review United States Annual review of medicine Annu Rev Med. 2004;55:223-37.

[34] Maria De Luca, Enrico Di Pace, Anna Judica, Donatella Spinelli, and Pierluigi Zoccolotti. Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia*, 37(12):1407–1420, 1999.

[35] RL De Valois and KK De Valois. Spatial vision. 1988. *Oxford: Oxford University Press*, 10:145–175.

[36] Daniel C DeMarco, Elizabeth Odstrcil, Luis F Lara, David Bass, Chase Herdman, Timothy Kinney, Kapil Gupta, Leon Wolf, Thomas

Dewar, and Thomas M Deas. Impact of experience with a retrograde-viewing device on adenoma detection rates and withdrawal times during colonoscopy: the third eye retroscope study group. *Gastrointestinal endoscopy*, 71(3):542–550, 2010.

[37] L. Dempere-Marco, Hu Xiao-Peng, S. L. S. MacDonald, S. M. Ellis, D. M. Hansell, and Yang Guang-Zhong. The use of visual search for knowledge gathering in image decision support. *IEEE Transactions on Medical Imaging*, 21(7):741–754, 2002.

[38] Heiko Drewes, Alexander De Luca, and Albrecht Schmidt. Eye-gaze interaction for mobile phones. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, pages 364–371. ACM, 2007.

[39] Xavier Dreze and Fred Zufryden. Testing web site design and promotional content. *Journal of Advertising Research*, 37:77–91, 1997.

[40] Jean Duchon. Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive theory of functions of several variables*, pages 85–100. Springer, 1977.

[41] Andrew Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer, 2007.

[42] Y. Ebisawa and S. i Satoh. Effectiveness of pupil area detection technique using two light sources and image difference method. In *Engineering in Medicine and Biology Society, 1993. Proceedings of the 15th Annual International Conference of the IEEE*, pages 1268–1269, 1993.

[43] Axel Eickhoff, Jacques Van Dam, Ralf Jakobs, Valerie Kudis, Dirk Hartmann, Ulrich Damian, Uwe Weickert, Dieter Schilling, and Jürgen F Riemann. Computer-assisted colonoscopy (the neoguide endoscopy system): results of the first human clinical trial ("pace study"). *The American journal of gastroenterology*, 102(2):261–266, 2007.

[44] Shahram Eivazi, Roman Bednarik, Markku Tukiainen, Mikael von und zu Fraunberg, Ville Leinonen, and Juha E.s Jääskeläinen. Gaze

behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 377–380. ACM, 2012.

[45] M. Eizenman, D. Model, and E. D. Guestrin. Covert monitoring of the point-of-gaze. In *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, pages 551–556, 2009.

[46] Marcela Fejtová, Jan Fejt, and Lenka Lhotská. Controlling a pc by eye movements: The memrec project. In *Computers Helping People with Special Needs*, pages 770–773. Springer, 2004.

[47] B Fischer, K Hartnegg, and A Mokler. Dynamic visual perception of dyslexic children. *Perception*, 29(5):523–530, 2000.

[48] Sebastian Foersch, Axel Heimann, Ali Ayyad, Gilles A Spoden, Luise Florin, Konstantin Mpoukouvalas, Ralf Kiesslich, Oliver Kempski, Martin Goetz, and Patra Charalampaki. Confocal laser endomicroscopy for diagnosis and histomorphologic imaging of brain tumors in vivo. *PloS one*, 7(7):e41760–, 2012.

[49] Daniel R Gavin, Roland M Valori, John T Anderson, Mark T Donnelly, J Graham Williams, and Edwin T Swarbrick. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the uk. *Gut*, 62(2):242–249, 2013.

[50] E. D. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *Biomedical Engineering, IEEE Transactions on*, 53(6):1124–1133, 2006.

[51] S. S. Hacisalihzade, L. W. Stark, and J. S. Allen. Visual perception and sequences of eye movement fixations: a stochastic modeling approach. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):474–481, 1992.

[52] D. W. Hansen and Ji Qiang. In the eye of the beholder: A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):478–500, 2010.

[53] Dan Witzner Hansen, Javier San Agustin, and Arantxa Villanueva. Homography normalization for robust gaze estimation in uncalibrated setups. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 13–20. ACM, 2010.

[54] Rami Hawari and Pankaj Jay Pasricha. Going for the loop: a unique overtube for the difficult colonoscopy. *Journal of clinical gastroenterology*, 41(2):138–140, 2007.

[55] Craig Hennessey, Borna Noureddin, and Peter Lawrence. A single camera eye-gaze tracking system with free head motion. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, pages 87–94. ACM, 2006.

[56] J. G. Holden. Perceptual-motor coordination in an endoscopic surgery simulation. *Surgical endoscopy*, 13(2):127–132, 1999.

[57] K Holler, Jochen Penne, Joachim Hornegger, Armin Schneider, Sonja Gillen, Hubertus Feussner, Jasper Jahn, Javier Gutierrez, and Thomas Wittenberg. Clinical evaluation of endorientation: Gravity related rectification for endoscopic images. In *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*, pages 713–717. IEEE, 2009.

[58] Kurt Holler, Jochen Penne, Armin Schneider, Jasper Jahn, Javier Gutierrez Boronat, Thomas Wittenberg, Hubertus Feussner, and Joachim Hornegger. Endoscopic orientation correction. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2009*, pages 459–466. Springer, 2009.

[59] Anthony J Hornof and Tim Halverson. Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers*, 34(4):592–604, 2002.

[60] P. D. van Hove, G. J. M. Tuijthof, E. G. G. Verdaasdonk, L. P. S. Stassen, and J. Dankelman. Objective assessment of technical surgical skills. *British Journal of Surgery*, 97(7):972–987, 2010.

[61] T Hull and JM Church. Colonoscopy: how difficult, how painful? *Surgical endoscopy*, 8(7):784–787, 1994.

[62] PRINCIPLES OF TRAINING IN. Principles of training in gastrointestinal endoscopy. *Gastrointestinal Endoscopy*, 49(6), 1999.

[63] Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. Designing gaze gestures for gaming: an investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 323–330. ACM, 2010.

[64] Robert JK Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 11–18. ACM, 1990.

[65] JAG. ——— certification criteria. http://www.thejag.org.uk/ training-forendoscopists/endoscopytrainees/traineecertification/ certification-criteria.aspx. accessed 24 April 2014.

[66] Marcel Adam Just and Patricia A Carpenter. Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4):441–480, 1976.

[67] Marcel Adam Just and Patricia A Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87:329–354, 1980.

[68] JJ Kang, ED Guestrin, WJ MacLean, and M Eizenman. Simplifying the cross-ratios method of point-of-gaze estimation. In *30th Canadian medical and biological engineering conference (CMBEC30)*, 2007.

[69] A. E. Kaufman, A. Bandopadhay, and B. D. Shaviv. An eye tracking computer user interface. In *Virtual Reality, 1993. Proceedings., IEEE 1993 Symposium on Research Frontiers in*, pages 120–121, 1993.

[70] M. P. Keating. *Geometric, Physical, and Visual Optics (Second Edition)*. 2002.

[71] Sashi S Kommu, Peter Rimington, Christopher Anderson, and Abhay Rané. Initial experience with the endoassist camera-holding robot in laparoscopic urological surgery. *Journal of Robotic Surgery*, 1(2):133–137, 2007.

[72] BM Kraft, C Jäger, K Kraft, BJ Leibl, and R Bittner. The aesop robot system in laparoscopic surgery: Increased risk or advantage for surgeon and patient? *Surgical Endoscopy And Other Interventional Techniques*, 18(8):1216–1223, 2004.

[73] H. L. Kundel. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative radiology*, 13(3):175, 1978.

[74] Marta Kutas and Kara D Federmeier. Minding the body. *Psychophysiology*, 35(2):135–150, 1998.

[75] Ka-Wai Kwok, Loi-Wah Sun, George P Mylonas, David RC James, Felipe Orihuela-Espina, and Guang-Zhong Yang. Collaborative gaze channelling for improved cooperation during robotic assisted surgery. *Annals of biomedical engineering*, 40(10):2156–2167, 2012.

[76] Ka Wai Kwok, Loi Wah Sun, Valentina Vitiello, D James, George P Mylonas, Ara Darzi, and Guang-Zhong Yang. Perceptually docked control environment for multiple microbots: application to the gastric wall biopsy. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 2783–2788. IEEE, 2009.

[77] Ka-Wai Kwok, Kuen Hung Tsoi, V. Vitiello, J. Clark, G.C.T. Chow, W. Luk, and Guang-Zhong Yang. Dimensionality reduction in controlling articulated snake robot for endoscopy under dynamic active constraints. *Robotics, IEEE Transactions on*, 29(1):15–31, 2013.

[78] Ka-Wai Kwok, Valentina Vitiello, and Guang-Zhong Yang. Control of articulated snake robot under dynamic active constraints. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 229–236. Springer, 2010.

[79] K. J. Kyriakopoulos and G. N. Saridis. Minimum jerk path generation. In *IEEE International Conference on Robotics and Automation, 1988. Proceedings., 1988*, pages 364–369 vol.1, 1988.

[80] Michael F. Land. Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, 25(3):296–324, 2006.

[81] Mirna Lerotic, Adrian J Chung, James Clark, Salman Valibeik, and Guang-Zhong Yang. Dynamic view expansion for enhanced navigation in natural orifice transluminal endoscopic surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 467–475. Springer, 2008.

[82] D C Lindsay, J G Freeman, I Cobden, and C O Record. Should colonoscopy be the first investigation for colonic disease? *BMJ*, 296(6616):167–169, 1988.

[83] Jen-Jane Liu, Michael J Droller, and Joseph C Liao. New optical imaging technologies for bladder cancer: considerations and perspectives. *The Journal of urology*, 188(2):361–368, 2012.

[84] A. Loeve, P. Breedveld, and J. Dankelman. Scopes too flexible...and too stiff. *Pulse, IEEE*, 1(3):26–41, 2010.

[85] Lester C Loschky and George W McConkie. User performance with gaze contingent multiresolutional displays. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 97–103. ACM, 2000.

[86] I. Scott MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7(1):91–139, 1992.

[87] Päivi Majaranta and Kari-Jouko Räihä. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 15–22. ACM, 2002.

[88] Nikolay L Martirosyan, Daniel D Cavalcanti, Jennifer M Eschbacher, Peter M Delaney, Adrienne C Scheck, Mohammed G Abdelwahab, Peter Nakaji, Robert F Spetzler, and Mark C Preul. Use of in vivo near-infrared laser confocal endomicroscopy with indocyanine green to detect the boundary of infiltrative tumor: Laboratory investigation. *Journal of neurosurgery*, 115(6):1131–1138, 2011.

[89] John Merchant, Richard Morrissette, and James L. Porterfield. Remote measurement of eye direction allowing subject motion over one cubic foot of space. *Biomedical Engineering, IEEE Transactions on*, BME-21(4):309–317, 1974.

[90] Dmitri Model and Moshe Eizenman. User-calibration-free remote gaze estimation system. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 29–36. ACM, 2010.

[91] Emilie Mollenbach, John Paulin Hansen, Martin Lillholm, and Alastair G Gale. Single stroke gaze gestures. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 4555–4560. ACM, 2009.

[92] Carlos H. Morimoto. Detecting eye position and gaze from a single camera and 2 light sources. volume 4, pages 40314–40314, 2002.

[93] Carlos H. Morimoto and Marcio R. M. Mimica. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.*, 98(1):4–24, 2005.

[94] Carlos Hitoshi Morimoto, David Koons, A Amit, Myron Flickner, and Shumin Zhai. Keeping an eye for hci. In *Computer Graphics and Image Processing, 1999. Proceedings. XII Brazilian Symposium on*, pages 171–176. IEEE, 1999.

[95] P. Mountney, D. Stoyanov, and G.Z. Yang. Three-dimensional tissue deformation recovery and tracking. *Signal Processing Magazine, IEEE*, 27(4):14–24, 2010.

[96] Peter Mountney and Guang-Zhong Yang. Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 1184–1187. IEEE, 2009.

[97] George Mylonas, Danail Stoyanov, Ara Darzi, and Guang-Zhong Yang. Assessment of perceptual quality for gaze-contingent motion stabilization in robotic assisted minimally invasive surgery. pages 660–667. 2007.

[98] George P Mylonas. *Gaze-Contingent Perceptual Docking for Minimally Invasive Robotic Surgery.* PhD thesis, Imperial College London, 2009.

[99] George P. Mylonas, Ara Darzi, and Guang-Zhong Yang. Gaze contingent depth recovery and motion stabilisation for minimally invasive robotic surgery. pages 311–319. 2004.

[100] George P Mylonas, Ka-Wai Kwok, Ara Darzi, and Guang-Zhong Yang. Gaze-contingent motor channelling and haptic constraints for minimally invasive robotic surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 676–683. Springer, 2008.

[101] Takashi Nagamatsu, Yukina Iwamoto, Junzo Kamahara, Naoki Tanaka, and Michiya Yamamoto. Gaze estimation method based on an aspherical model of the cornea: surface of revolution about the optical axis of the eye. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 255–258. ACM, 2010.

[102] Takashi Nagamatsu, Junzo Kamahara, Takumi Iko, and Naoki Tanaka. One-point calibration gaze tracking based on eyeball kinematics using stereo cameras. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 95–98. ACM, 2008.

[103] Takashi Nagamatsu, Ryuichi Sugano, Yukina Iwamoto, Junzo Kamahara, and Naoki Tanaka. User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 251–254. ACM, 2010.

[104] Atsushi Nakazawa and Christian Nitschke. Point of gaze estimation through corneal surface reflection in an active illumination environment. In *Computer Vision–ECCV 2012*, pages 159–172. Springer, 2012.

[105] Charles C Nduka, Paul A Super, JR Monson, and Ara W Darzi. Cause and prevention of electrosurgical injuries in laparoscopy. *Journal of the American College of Surgeons*, 179(2):161–170, 1994.

[106] J. Nielson. F-shaped pattern for reading web content. 2006.

[107] Rie Niimi. Eye-tracking device comparisons of three methods of magnetic resonance image series displays. *Journal of Digital Imaging*, 10(4):147–151, 1997.

[108] CF Nodine and HL Kundel. A visual dwell algorithm can aid search and recognition of missed lung nodules in chest radiographs. *Visual Search*, 2:399–406, 1990.

[109] D. P. Noonan, G. P. Mylonas, A. Darzi, and Yang Guang-Zhong. Gaze contingent articulated robot control for robot assisted minimally invasive surgery. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 1186–1191, 2008.

[110] David P Noonan, George P Mylonas, Jianzhong Shang, Christopher J Payne, Ara Darzi, and Guang-Zhong Yang. Gaze contingent control for an articulated mechatronic laparoscope. In *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*, pages 759–764. IEEE, 2010.

[111] David P Noonan, George P Mylonas, Jianzhong Shang, Christopher J Payne, Ara Darzi, and Guang-Zhong Yang. Gaze contingent control for an articulated mechatronic laparoscope. In *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*, pages 759–764. IEEE, 2010.

[112] M. Nystrm, R. Andersson, K. Holmqvist, and J. Weijer. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods, Instruments, & Computers*, 45(1):272–288, 2013.

[113] Takehiko Ohno and Naoki Mukawa. A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 115–122. ACM, 2004.

[114] Takehiko Ohno, Naoki Mukawa, and Atsushi Yoshikawa. Freegaze: a gaze tracking system for everyday gaze interaction. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 125–132. ACM, 2002.

[115] Takeyoshi Ota, Amir Degani, David Schwartzman, Brett Zubiate, Jeremy McGarvey, Howie Choset, and Marco A Zenati. A highly articulated robotic surgical system for minimally invasive surgery. *The Annals of thoracic surgery*, 87(4):1253–1256, 2009.

[116] Bing Pan, Helene A Hembrooke, Geri K Gay, Laura A Granka, Matthew K Feusner, and Jill K Newman. The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 147–154. ACM, 2004.

[117] Perry J Pickhardt, Pamela A Nugent, Pauline A Mysliwiec, J Richard Choi, and William R Schindler. Location of adenomas missed by optical colonoscopy. *Annals of internal medicine*, 141(5):352–359, 2004.

[118] M. Pilu, A. W. Fitzgibbon, and R. B. Fisher. Ellipse-specific direct least-square fitting. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 599–602 vol.3, 1996.

[119] Marc Pomplun and Sindhura Sunkara. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI*, 2003.

[120] K. Preston White, T. L. Hutson, and T. E. Hutchinson. Modeling human eye behavior during mammographic scanning: preliminary results. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 27(4):494–505, 1997.

[121] R. M. Pritchard. Stabilized images on the retina. *Scientific American*, 204(6):72, 1961.

[122] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[123] D. Rattner. Asge/sages working group on natural orifice translumenal endoscopic surgery. *Surgical endoscopy*, 20(2):329, 2006.

[124] G. M. Redding and B. Wallace. Adaptive mechanisms in perceptual-motor coordination: Components of prism adaptation. *Journal of Motor Behavior*, 20(3):242–254, 1988.

[125] Douglas K Rex, John H Bond, Sidney Winawer, Theodore R Levin, Randall W Burt, David A Johnson, Lynne M Kirk, Scott Litlin, David A Lieberman, Jerome D Waye, James Church, John B Marshall, and Robert H Riddell. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the u.s. multi-society task force on colorectal cancer. *Am J Gastroenterol*, 97(6):1296–1308, June 2002.

[126] David A. Robinson. A method of measuring eye movemnent using a scieral search coil in a magnetic field. *Bio-Medical Electronics, IRE Transactions on*, 10(4):137–145, 1963.

[127] P. J. Robinson. Radiology's achilles' heel: error and variation in the interpretation of the rontgen image. *The British journal of radiology*, 70(839):1085, 1997.

[128] Louis B Rosenberg. Virtual fixtures: Perceptual tools for telerobotic manipulation. In *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*, pages 76–82. IEEE, 1993.

[129] D Rozado, T Moreno, J San Agustin, FB Rodriguez, and P Varona. Controlling a smartphone using gaze gestures as the input mechanism. *Human–Computer Interaction*, (just-accepted), 2013.

[130] David Rozado, Francisco B. Rodriguez, and Pablo Varona. Low cost remote gaze gesture recognition in real time. *Applied Soft Computing*, 12(8):2072–2084, 2012.

[131] S Sadahiro, T Ohmura, Y Yamada, T Saito, and Y Taki. Analysis of length and surface area of each segment of the large intestine according to age, sex and physique. *Surgical and Radiologic Anatomy*, 14(3):251–257, 1992.

[132] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.

[133] BP Saunders, FA Macrae, and CB Williams. What makes colonoscopy difficult. *Gut*, 34(suppl 1):T179, 1993.

[134] Brian P Saunders, Manabu Fukumoto, Steve Halligan, Craig Jobling, Mohammed E Moussa, Clive I Bartram, and Christopher B Williams. Why is colonoscopy more difficult in women? *Gastrointestinal endoscopy*, 43(2):124–126, 1996.

[135] Drew B Schembre, Andrew S Ross, Michael N Gluck, John J Brandabur, Susan E McCormick, and Otto S Lin. Spiral overtube-assisted colonoscopy after incomplete colonoscopy in the redundant colon. *Gastrointestinal endoscopy*, 73(3):515–519, 2011.

[136] Charles T Scialfa, Lisa McPhee, and Geoffrey Ho. The effects of a simulated cellular phone conversation on search for traffic signs in an elderly sample. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 45–50. ACM, 2000.

[137] Laura Sesma, Arantxa Villanueva, and Rafael Cabeza. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 217–220. ACM, 2012.

[138] Syed G Shah, Brian P Saunders, Jim C Brooker, and Christopher B Williams. Magnetic imaging of colonoscopy: an audit of looping, accuracy and ancillary maneuvers. *Gastrointestinal endoscopy*, 52(1):1–8, 2000.

[139] J. Shang, D. P. Noonan, C. Payne, J. Clark, M. H. Sodergren, A. Darzi, and G. Z.s Yang. An articulated universal joint based flexible access robot for minimally invasive surgery. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1147–1152, 2011.

[140] Shih Sheng-Wen, Wu Yu-Te, and Liu Jin. A calibration-free gaze tracking technique. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 4, pages 201–204 vol.4, 2000.

[141] Shohan Shetty, Lucian Panait, Jacob Baranoski, Stanley J Dudrick, Robert L Bell, Kurt E Roberts, and Andrew J Duffy. Construct and face validity of a virtual reality–based camera navigation curriculum. *Journal of Surgical Research*, 177(2):191–195, 2012.

[142] Sheng-Wen Shih and Jin Liu. A novel approach to 3-d gaze tracking using stereo cameras. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 34(1):234–245, 2004.

[143] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo. Robotics: modelling, planning and control. *Springer*, 2009.

[144] Mikael H Sodergren, Felipe Orihuela-Espina, James Clark, Julian Teare, Guang-Zhong Yang, and Ara Darzi. Evaluation of orientation strategies in laparoscopic cholecystectomy. *Annals of surgery*, 252(6):1027–1036, 2010.

[145] K. Sriskandarajah, S. Gillen, A. Di Marco, M. Sodergren, D.R.C. James, J. Clark, H.D. Feussner, A. Darzi, and G.-Z. Yang. Robotic tele-manipulating devices for laparoscopy improve surgical performance in simulated porcine laparoscopic cholecystectomies on the elite simulator. *Proceedings of the 2012 Hamlyn Symposium on Medical Robotics*, pages 36–37, 2012.

[146] C. Staub, S. Can, B. Jensen, A. Knoll, and S. Kohlbecher. Human-computer interfaces for interaction with surgical tools in robotic surgery. In *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on*, pages 81–86, 2012.

[147] Danail Stoyanov, George P Mylonas, and Guang-Zhong Yang. Gaze-contingent 3d control for focused energy ablation in robotic assisted surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 347–355. Springer, 2008.

[148] Sarah E Streett. Endoscopic colorectal cancer screening in women: can we do better? *Gastrointestinal endoscopy*, 65(7):1047–1049, 2007.

[149] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. Calibration-free gaze sensing using saliency maps. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2667–2674. IEEE, 2010.

[150] A. Sugioka, Y. Ebisawa, and M. Ohtani. Noncontact video-based eye-gaze detection method allowing large head displacements. In *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines*

for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE, volume 2, pages 526–528 vol.2, 1996.

[151] Shinji Tanaka, Shiro Oka, Iwao Kaneko, Mayuko Hirata, Ritsuo Mouri, Hiroyuki Kanao, Shigeto Yoshida, and Kazuaki Chayama. Endoscopic submucosal dissection for colorectal neoplasia: possibility of standardization. *Gastrointestinal endoscopy*, 66(1):100–107, 2007.

[152] M. J. Tarr and S Pinker. Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2):233–282, 1989.

[153] Russell H Taylor. *Computer-integrated surgery: technology and clinical applications*. Mit Press, 1996.

[154] Johannes Totz, Kenko Fujii, Peter Mountney, and Guang.Zhong Yang. Enhanced visualisation for minimally invasive surgery. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–10, 2011.

[155] G Triadafilopoulos and J Li. A pilot study to assess the safety and efficacy of thethird eye retrograde auxiliary imaging system during colonoscopy. *Endoscopy*, 40(06):478–482, 2008.

[156] George Triadafilopoulos, H. David Watts, Jack Higgins, and Jacques Van Dam. A novel retrograde-viewing auxiliary imaging device (third eye retroscope) improves the detection of simulated polyps in anatomic models of the colon. *Gastrointestinal Endoscopy*, 65(1):139 – 144, 2007.

[157] Vivian M Ussui and Michael B Wallace. Confocal endomicroscopy of colorectal polyps. *Gastroenterology research and practice*, 2012, 2012.

[158] Shyam Varadarajulu and William H Ramsey. Utility of retroflexion in lower gastrointestinal endoscopy. *Journal of clinical gastroenterology*, 32(3):235–237, 2001.

[159] Arantxa Villanueva, Rafael Cabeza, and Sonia Porta. Eye tracking: Pupil orientation geometrical modeling. *Image and Vision Computing*, 24(7):663–679, 2006.

[160] Marco Visentini-Scarzanella, George Mylonas, Danail Stoyanov, and Guang-Zhong Yang. i-brush: A gaze-contingent virtual paintbrush for dense 3d reconstruction in robotic assisted surgery. pages 353–360. 2009.

[161] Kwok K.W. Payne-C. Yang G.Z. Vitiello, V. Dof minimization for optimized shape control under active constraints for a hyper-redundant flexible robot. *Information Processing in Computer-Assisted Interventions*, pages 67–78, 2011.

[162] Valentina Vitiello, Su-Lin Lee, Thomas P Cundy, and Guang-Zhong Yang. Emerging robotic platforms for minimally invasive surgery. *Biomedical Engineering, IEEE Reviews in*, 6:111–126, 2013.

[163] ST Ward, MA Mohammed, R Walt, R Valori, T Ismail, and P Dunckley. An analysis of the learning curve to achieve competency at colonoscopy using the jets database. *Gut*, 2014.

[164] Alexander Warren, Peter Mountney, David Noonan, and Guang-Zhong Yang. Horizon stabilized-dynamic view expansion for robotic assisted surgery (hs-dve). *International journal of computer assisted radiology and surgery*, 7(2):281–288, 2012.

[165] Pijika Watcharapichat. Image processing and classification algorithm to detect cancerous cells morphology when using in-vivo probe-based confocal laser endomicroscopy for the lower gastrointestinal tract. Master's thesis, Imperial College London, 2012.

[166] Wolfgang Wayand. The history of minimally invasive surgery. *Global Surgery 2004*, pages 37–38, 2004.

[167] Jerome D Waye, Siroun A Yessayan, Blair S Lewis, and Thomas L Fabry. The technique of abdominal pressure in total colonoscopy. *Gastrointestinal endoscopy*, 37(2):147–151, 1991.

[168] Jr. White, K. P., T. E. Hutchinson, and J. M. Carley. Spatially dynamic calibration of an eye-tracking system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(4):1162–1168, 1993.

[169] Mark R Wilson, Samuel J Vine, Elizabeth Bright, Rich SW Masters, David Defriend, and John S McGrath. Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. *Surgical endoscopy*, 25(12):3731–3739, 2011.

[170] Todd N Witte and Robert Enns. The difficult colonoscopy. *Canadian Journal of Gastroenterology*, 21(8):487, 2007.

[171] Jacob O Wobbrock, James Rubinstein, Michael W Sawyer, and Andrew T Duchowski. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 11–18. ACM, 2008.

[172] David S Wooding. Eye movements of large populations: Ii. deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments, & Computers*, 34(4):518–528, 2002.

[173] S. Yamaguchi, K. Konishi, T. Yasunaga, D. Yoshida, N. Kinjo, K. Kobayashi, S. Ieiri, K. Okazaki, H. Nakashima, K. Tanoue, Y. Maehara, and M. Hashizume. Construct validity for eye-hand coordination skill on a virtual reality laparoscopic surgical simulator. *Surgical Endoscopy And Other Interventional Techniques*, 21(12):2253–7, 2007.

[174] G. Z. Yang, L. Dempere-Marco, X. P. Hu, and A. Rowe. Visual search: psychophysical models and practical applications. *Image and Vision Computing*, 20(4):291, 2002.

[175] Guang-Zhong Yang, George P Mylonas, Ka-Wai Kwok, and Adrian Chung. Perceptual docking for robotic control. In *Medical Imaging and Augmented Reality*, pages 21–30. Springer, 2008.

[176] Guang-Zhong Yang, George P Mylonas, Ka-Wai Kwok, and Adrian Chung. Perceptual docking for robotic control. In *Medical Imaging and Augmented Reality*, pages 21–30. Springer, 2008.

[177] Y. Yang. Pupil location under mesopic, photopic, and pharmacologically dilated conditions. *Investigative ophthalmology & visual science*, 43(7):2508, 2002.

[178] Dong Hyun Yoo and Myung Jin Chung. Non-intrusive eye gaze estimation without knowledge of eye pose. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 785–790. IEEE, 2004.

[179] Dong Hyun Yoo and Myung Jin Chung. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Comput. Vis. Image Underst.*, 98(1):25–51, 2005.

[180] Dong Hyun Yoo, Jae Heon Kim, Bang Rae Lee, and Myoung Jin Chung. Non-contact eye gaze tracking system by mapping of corneal reflections. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 94–99, 2002.

[181] Patrick E Young and Craig M Womeldorph. Colonoscopy for colorectal cancer screening. *Journal of Cancer*, 4(3):217, 2013.

[182] Bin Zheng, Xianta Jiang, Geoffrey Tien, Adam Meneghetti, O. NeelyM Panton, and M. Stella Atkins. Workload assessment of surgeons: correlation between nasa tlx and blinks. *Surgical endoscopy*, 26(10):2746–2750, 2012.

[183] Bin Zheng, Erwin Rieder, Maria A Cassera, Danny V Martinec, Gyusung Lee, O Neely M Panton, Adrian Park, and Lee L Swanström. Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (notes) procedures. *Surgical endoscopy*, 26(5):1352–1358, 2012.

[184] Anlong Zhu, Chao Yuan, Daxun Piao, Tao Jiang, and Hongchi Jiang. Gravity line strategy may reduce risks of intraoperative injury during laparoscopic surgery. *Surgical endoscopy*, 27(12):4478–4484, 2013.

[185] Jie Zhu and Jie Yang. Subpixel eye gaze tracking. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 124–129, 2002.

[186] Zhiwei Zhu, Qiang Ji, and Kristin P Bennett. Nonlinear eye gaze mapping function estimation via support vector regression. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 1132–1135. IEEE, 2006.