HAND EYE COORDINATION IN SURGERY

Jit Hung Julian John Leong MA(Oxon) MBBS MRCS(Eng)

Department of Biosurgery and Surgical Technology
Royal Society/Wolfson Foundation Medical Image Computing Laboratory
Imperial College London
University of London

January 2009

This thesis is submitted to the University of London in partial fulfilment of the requirements for the degree of Doctor of Philosophy. Except where otherwise indicated, all the work described in this thesis is original and contains results of the authors own research.

ABSTRACT

The coordination of the hand in response to visual target selection has always been regarded as an essential quality in a range of professional activities. This quality has thus far been elusive to objective scientific measurements, and is usually engulfed in the overall performance of the individuals. Parallels can be drawn to surgery, especially Minimally Invasive Surgery (MIS), where the physical constraints imposed by the arrangements of the instruments and visualisation methods require certain coordination skills that are unprecedented. With the current paradigm shift towards early specialisation in surgical training and shortened focused training time, selection process should identify trainees with the highest potentials in certain specific skills. Although significant effort has been made in objective assessment of surgical skills, it is only currently possible to measure surgeons' abilities at the time of assessment. It has been particularly difficult to quantify specific details of hand-eye coordination and assess innate ability of future skills development. The purpose of this thesis is to examine hand-eye coordination in laboratory-based simulations, with a particular emphasis on details that are important to MIS.

In order to understand the challenges of visuomotor coordination, movement trajectory errors have been used to provide an insight into the innate coordinate mapping of the brain. In MIS, novel spatial transformations, due to a combination of distorted endoscopic image projections and the "fulcrum" effect of the instruments, accentuate movement generation errors. Obvious differences in the quality of movement trajectories have been observed between novices and experts in MIS, however, this is difficult to measure quantitatively. A Hidden Markov Model (HMM) is used in this thesis to reveal the underlying characteristic movement details of a particular MIS manoeuvre and how such features are exaggerated by the introduction of rotation in the endoscopic camera. The proposed method has demonstrated the feasibility of measuring movement trajectory quality by machine learning techniques without prior arbitrary classification of expertise. Experimental results have highlighted these changes in novice laparoscopic surgeons, even after a short period of training.

The intricate relationship between the hands and the eyes changes when learning a skilled visuomotor task has been previously studied. Reactive eye movement, when visual input is used primarily as a feedback mechanism for error correction, implies difficulties in hand-eye coordination. As the brain learns to adapt to this new coordinate map, eye

movements then become predictive of the action generated. The concept of measuring this spatiotemporal relationship is introduced as a measure of hand-eye coordination in MIS, by comparing the Target Distance Function (TDF) between the eye fixation and the instrument tip position on the laparoscopic screen.

Further validation of this concept using high fidelity experimental tasks is presented, where higher cognitive influence and multiple target selection increase the complexity of the data analysis. To this end, Granger-causality is presented as a measure of the predictability of the instrument movement with the eye fixation pattern. Partial Directed Coherence (PDC), a frequency-domain variation of Granger-causality, is used for the first time to measure hand-eye coordination. Experimental results are used to establish the strengths and potential pitfalls of the technique. To further enhance the accuracy of this measurement, a modified Jensen-Shannon Divergence (JSD) measure has been developed for enhancing the signal matching algorithm and trajectory segmentations. The proposed framework incorporates high frequency noise filtering, which represents non-purposeful hand and eye movements. The accuracy of the technique has been demonstrated by quantitative measurement of multiple laparoscopic tasks by expert and novice surgeons.

Experimental results supporting visual search behavioural theory are presented, as this underpins the target selection process immediately prior to visual motor action generation. The effects of specialisation and experience on visual search patterns are also examined. Finally, pilot results from functional brain imaging are presented, where the Posterior Parietal Cortical (PPC) activation is measured using optical spectroscopy techniques. PPC has been demonstrated to involve in the calculation of the coordinate transformations between the visual and motor systems, which establishes the possibilities of exciting future studies in hand-eye coordination.

ACKNOWLEDGEMENTS

孫子曰：凡用兵之法，馳車千駟，革車千乘，帶甲十萬，千里饋糧
"Sun Tzu said: In the operations of war, where there are in the field a thousand swift chariots, as many heavy chariots, and a hundred thousand mail-clad soldiers, with provisions enough to carry them a thousand LI" The art of war – Sun Tzu (translated by Lionel Giles)

These were words of Sun Tzu in the 6th Century BC, discussing the importance of the preparation and cost of battle. Likewise, the preparation of this PhD thesis has felt like travelling a thousand LI (2.78 LI = 1 mile), some of which treacherous, whilst most others enlightening. I would like to express my gratitude to my supervisors, for equipping me with all the provisions to prepare for this long and rewarding journey.

I am indebted to Professor Ara Darzi for giving me the opportunity to complete this thesis, and his guidance to steer my journey towards the final goal. His unwavering determinations and clear directions set a fine example of an academic surgeon. I would like to thank Professor Guang-Zhong Yang for his time, his patience, his wisdom, and most of all his kindness, for guiding me throughout this thesis. I would also like to thank Mr Roger Emery for being my mentor, his vision and advice have helped me develop my research and clinical interests.

I would not be able to complete this journey without the help of these friends and colleagues: Louis Atallah, Marios Nicolaou and George Mylonas for the time they sacrificed to help design the experiment and analysis framework. Daniel Leff for all our discussions and brainstorms. Felipe Orihuela-Espina for his passionate explanations of mathematics. Fani Deligianni and Adam James for their help in some of the software and hardware implementations. Mr Thanos Athanasiou for involving me in his exciting work.

I would also like to thank Sanjay Purkayastha, for being a true friend throughout some of the difficult periods. Ee Ling for believing that this thesis can be finished.

Most of all, I would like to thank my parents and my brother for their patience throughout this journey.

JL, Jan 09.

To my parents.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

2D	*Two-dimensions*
3D	*Three-dimensions*
A&E	*Accident and Emergency Department Senior House Officers*
ADEPT	*Advanced Dundee Endoscopic Psychomotor Tester*
API	*Application Programming Interfaces*
AR	*Autoregressive model*
CCD	*Charge-coupled device*
CDF	*Centroid distance function*
Con	*Consultant Orthopaedic Surgeon*
DICOM	*Digital Imaging and Communications in Medicine*
EM	*Expectation Maximisation Algorithm*
E-T	*Eye-Target distance*
EWTD	*European Working Time Directive*
FN	*False Negatives*
fNIRS	*Functional Near Infrared Spectroscopy*
FR	*First Rotation Task*
F/T	*Forces and Torques*
GMC	*General Medical Council*
GMM	*Gaussian Mixture Model*
HbO_2	*Oxyhaemoglobin*
HHb	*Deoxyhaemoglobin*
HMM	*Hidden Markov Model*
ICSAD	*Imperial College Surgical Assessment Device*
IR	*Internal Representation*
IRED	*Infra-Red*
IR-LED	*Infra-Red Light Emitting Diode*
JSD	*Jensen-Shannon divergence*
KL	*Kullback-Leibler distance*
LR	*Left Rotation Task*
MIS	*Minimally Invasive Surgery*
MMC	*Modernising Medical Careers*
MRCS	*Membership of the Royal College of Surgeons*
MSE	*Mean Squared Error*
MTAS	*Medical Training Application Service*
NDI	*Northern Digital Inc*
NOSsE™	*Natural Orifice Ssimulated sSurgical Environment*
NOTES	*Natural Orifice Translumenal Endoscopic Surgery*
NR	*Normal Rotation Task*

OSATS	*Objective Structured Assessment of Technical Skill*
PDC	*Partial Directed Coherence*
POST	*Post-Training Task*
PPC	*Posterior Parietal Cortex*
PRE	*Pre-Training Task*
Rad	*Consultant Radiologist*
RMS	*Root Mean Square*
RMSE	*Root Mean Squared Error*
RR	*Right Rotation Task*
RT	*Reaction Time*
SAGES	*Society of American Gastrointestinal Surgeons*
SHO	*Orthopaedic Senior House Officer*
SIFT	*Scale-Invariant Feature Transform*
SpR	*Orthopaedic Specialist Registrar*
SR	*Second Rotation Task*
TCP/IP	*Transmission Control Protocol and the Internet Protocol*
TDF	*Target Distance Function*
TFT	*Thin-Film Transistor*
TIFF	*Tagged Image File Format*
TP	*True Positives*
T-T	*Tip-Target distance*
USMLE	*United States Medical Licensing Examination*
VAR	*Vector Autoregressive model*
VOG	*Video-Oculography*

Chapter 1

Introduction

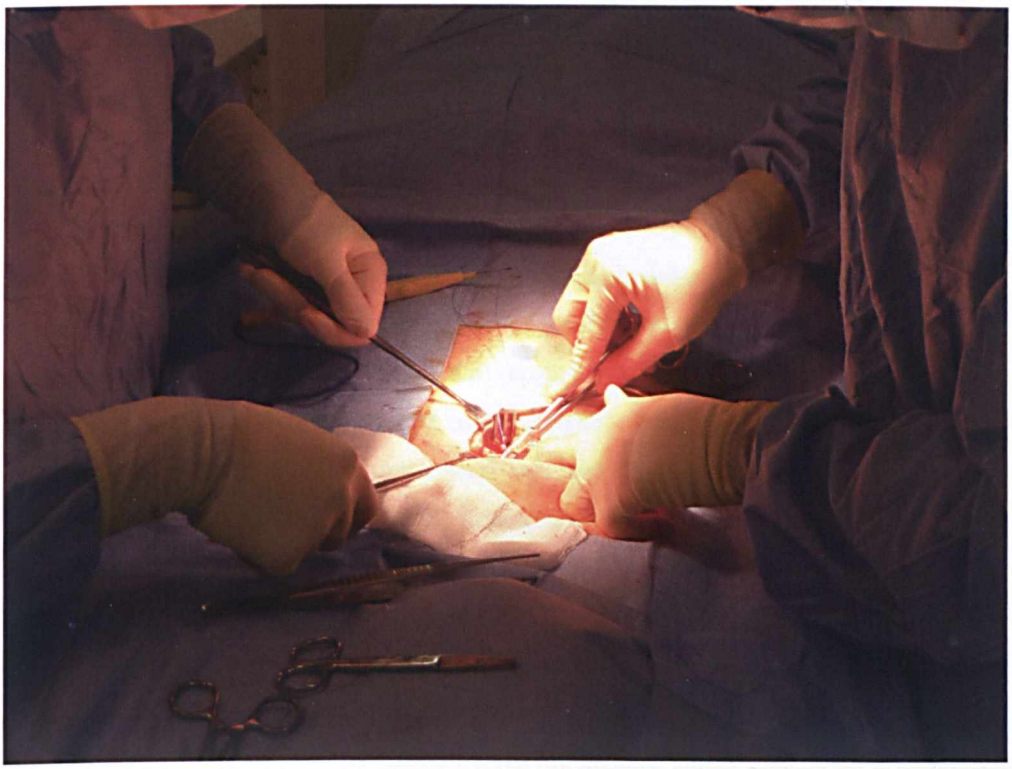

Figure 1.1 Photograph of a typical surgical scene.

Major challenges are faced by practising and training surgeons, with the continuous improvements in surgical technologies, the shortened training period and work hours, and the increased public awareness. One of the most important technological advances was the introduction of Minimally Invasive Surgery (MIS), which conveyed the advantages of quicker recovery time, smaller incisions and reduced trauma to the patients. However, the initial introduction of MIS carried higher complication rate and longer operation time, due to the significant learning curve involved in acclimatisation for the MIS environment

(1,2). The intrinsic setup of MIS requires specialised equipment, which restricts the visual information available, dampens haptic feedback, and introduces novel sensorimotor transformation to the surgeons. Newer technologies are now becoming available to improve surgical outcomes and patients' morbidity from surgery, these include Natural Orifice Translumenal Endoscopic Surgery (NOTES) and robotic assisted surgery, and undoubtedly a significant initial learning period would be required. This has renewed the debate on the delicate relationship between surgical training, assessment and patient safety.

With the introduction of more focused surgical training curricula, trainee selection and assessment need to advance with current technologies and become more specific to the surgical sub-speciality. Existing research has shown qualitative observation score and dexterity measurements can be used to monitor the progress of trainees, although predictions of the surgeons' future abilities remain a challenge. Observation score continues to be the 'gold standard' of objective skills assessment, as it utilises expert judgement in formulating the scores. Despite its reliability and validity, it remains too expensive and labour intensive for large scale deployment (3). On the other hand, dexterity measurements can be automated using motion tracking devices. These devices measure the distance each hand travels, and scoring is based on the principles of economy of movement, where shorter distance equates higher dexterity. Dexterity measurement using this technique is highly influenced by variations in the tasks, which renders assessment of live operations difficult. Cognitive influences are also neglected in these measurements, which may explain their lack of predictive validity.

One of the greatest initial challenges in MIS is the coordination of the instruments with the visual feedback available from the laparoscopic camera The combination of the "fulcrum" effect and distortions resulted from the camera lenses creates an unusual and challenging visuomotor coupling. This is akin to the measurement of hand-eye coordination in daily living tasks, although in adults this attribute is well developed and can only be distinguished when skilled tasks are performed, as in professional sportsmen. The measurement of hand-eye coordination in surgery remains the focus of this thesis.

Quantification of hand-eye coordination, however, is difficult. The combinatorial problem includes multiple inputs from the visual, tactile and proprioceptive system; whilst coordinating motor output through movements across joints of the axial and peripheral skeleton. Furthermore, interactions between different brain regions can

influence motor execution by integrating previously stored memories of the attributes of specific motor tasks, hence affecting the final quality of movement trajectories. The study of each individual component is complex in itself, hence rendering the task of their combined effect near impossible.

Traditional measures of hand-eye coordination adopted a goal directed approach, where artificially created complex coordination tasks were used to test the individual's capabilities. Quantification relied on their abilities to complete such tasks, and the speed in achieving them. These principles underpin the current methods of surgical skills assessments, where time measurements and outcomes of the operations performed by trainees are assessed. Although this may be useful in assessing performance of specific tasks, its generalised application as a measure of hand-eye coordination remains unaccepted.

More recent attempts have focused on measuring the quality of movement trajectories in laboratory-based tasks. In these studies, simple reaching tasks are performed whilst visual input is either distorted or impaired, and the variability and the error of the final hand movement output are measured. The extent of the error measured implies the reliance on the visual input for motor generation in these tasks. This indirect measure mirrors dexterity measurement in surgical skills assessment, where movement economy, or the reduction in motion error, equates dexterity. Again, the outcome is highly influenced by any variations of the experimental tasks, and tests outside the laboratory environment are difficult to standardise.

The spatiotemporal relationship between the hands and the eyes has been introduced as a concept for measuring their coordination. In learning a novel task, visual feedback is crucial for motor error reduction, especially in the terminal phase of the movement trajectory. Reactive eye movements, where hand motion precedes the saccades of the eyes, are used for visual feedback and final adjustments of the hand movement towards the target. On the other hand, predictive eye movements are displayed in familiar tasks, where the initial mental estimation of the target position is accurate enough for reaching the target. This relationship remains the most promising method to quantify hand-eye coordination, as this measure is direct and cognitively influenced.

Thus far, the measurement of this spatiotemporal relationship remains qualitative. Complex motion trajectories are generated in skilled motor tasks, and this is reflected by

intricate eye movements which are easily influenced by cognitive factors, such as distractions, attention and concentration. In order to measure these highly variable data streams, simple statistical calculations are not sufficient. A robust signal matching algorithm is needed to locate similarities between the two (or three in bimanual tasks) time series, then a causality measurement is deployed to identify signal precedence. A multidisciplinary approach is necessary, as signal matching and causality are problems not confined to movement data, but regularly utilised in economical, neurological and computer vision data.

The purpose of this thesis is to provide a method of measurement for analysing the hand-eye coordination in surgery. MIS provides a perfect platform to introduce a novel visuomotor environment for assessment of hand-eye coordination. It is hypothesised that the quality of the movement trajectories produced by novices in MIS changes through learning of the MIS environment, and this reflects on the continuing adaption and development of the internal mapping of the task. Validation is sought by including data from practised MIS surgeons, where the quality improves; and camera rotation tasks, where the disruption in hand-eye coordination increases the difficulties of the tasks.

This thesis further aims to quantify the spatiotemporal relationship of the instruments and the eyes, as a direct measurement of hand-eye coordination in surgery. Robust signal matching and causality algorithms are adapted and developed, in order to provide a generalised measurement. The effects of training and coordination disruption are examined in detail to provide further validation of the method.

Chapter 2 reviews the historical evolution of MIS; its conception dates back centuries ago, although its safe and popular practice only started in the last two decades. Human factors involved and the difficulties of MIS are further addressed, specifically related to visual and motor learning. In this chapter, the concept of sensorimotor transformation is introduced and its importance in relation to MIS explored. In MIS, this is further complicated by the coordinate changes presented by the laparoscopic camera setup. This chapter concludes by summarising the current status in research into the measurement of hand-eye coordination, and its potential application in MIS.

To further understand the human visual behaviour, purposeful eye movements are listed in Chapter 3. The selections of the hardware setup for eye tracking are described, along with discussions of the relative merits and disadvantages of the different techniques

available. In terms of the instrument tips, the goal is to track the tip position on the laparoscopic screen; due to technical constraints described in the chapter, an indirect method of tracking is used. In order to calculate the Two-Dimensional (2D) instrument tip projection from its Three-Dimensional (3D) position, intrinsic and extrinsic properties of the camera need to be calculated and its position known at all times. Options of hardware for tracking the instruments and camera positions are discussed further, and camera calibration and robotic hand-eye coordination algorithms are used for transforming 3D instrument tip positions. Issues of synchronisation of the data streams are raised and bespoke solutions are outlined.

Further to the review in Chapter 2, the sensorimotor transformation involved MIS is too complex to model precisely, Chapter 4 introduces the use of a stochastic framework for analysing instrument movement trajectories. Hidden Markov Models (HMM) are able to predict the state of the model by its observable outcomes and have been extensively used for pattern recognition tasks such as speech and handwriting. In this chapter, instrument tip trajectories are first transformed to Centroid Distance Function (CDF) to reduce the dimensionality of the data and render it rotational and translational invariant. These trajectories are then used to train the HMM, using the leave-one-out technique, and in turn the trained HMM is used to calculate the quality of each trajectory. To validate this method of classification, a simple laparoscopic task is used to test novices and experienced surgeons. Screen rotation is introduced to accentuate the difficulty of sensorimotor transformation, and the effects of training and screen rotation are measured and discussed in details.

Eye movement data is included in the analysis for the first time in Chapter 5. The concept of spatiotemporal relationship between the instrument and the eye is introduced, where the eye movement lags behind the instruments (reactive) when novel sensorimotor tasks are first attempted. As the tasks become familiar, the eyes appear to predict the movement of the instruments. Detailed qualitative analysis of the eye and instrument data streams provided obvious differences between experts and novices, and also gradual improvements during training. Interestingly, initial screen rotation tasks exert the most influence on the quality of eye and instrument movements. In terms of spatiotemporal relationship, quantitative proof of this observation is provided in simple Cartesian distance calculations, as the experiment is designed to avoid complex cognitive interactions to influence the movement generated. This method of analysis is possible as

the target location retrieval and trajectory segmentation are done using bespoke hardware modifications, with the compromise of the experimental design.

In order to improve the usability of the measurement of spatiotemporal relationship between the hand and instruments based on the findings in Chapter 5, signal matching algorithms are explored for data analysis. Partial Directed Coherence (PDC) is used to analyse the instrument and eye data for the first time, and is discussed in details in Chapter 6. PDC is a frequency-based variation of Granger-causality, which is an econometric technique to calculate the relationship between two time series, and whether one can be used to forecast another. This technique eliminates the requirements of the target coordinates, and analysed the whole data streams in the frequency domain. Data from two experiments are collected for validation of this method, where the latter experiment resembles more closely to real laparoscopic surgery with complex interactions between the two hands and the eyes are analysed.

Chapter 7 is an extension on the analysis in Chapter 6, and models the instrument and eye movements into probability distributions using principles based on the Jensen-Shannon Divergence (JSD). Measuring the statistical difference between the two distributions provides a level of signal matching between them, and an adjustable threshold can be set compared to PDC in Chapter 6. Modifications of the JSD are necessary to analyse data in different frequencies, which allows for filtering of high frequency noise to improve the signal-to-noise ratio. Actual quantification of the temporal difference between the hand and the eye is calculated, and larger data sets using both simple and complex laparoscopic tasks are collected, and expert data are included to validate the results.

Chapter 8 summarises the findings of the thesis, and systematically discusses each chapter presented in terms of strengths and weaknesses. It also highlights the potential improvements in the experimental setup for its implementation as an assessment and selection tool. In this chapter, a pilot study including preliminary functional brain data is outlined as possible future work.

In a skilled motor task, visual search is an essential component primarily used for target selection and location, prior to movement generation. Appendix A presents a study on the quantification of visual search pattern, based on the Global-focal model. Static medical images are used for the study, and X-Ray radiographs are chosen as they are familiar to physicians across specialities. The effects of training and specialisation are compared,

using Kullback-Leibler distance (similar to JSD) and double Gaussian curve fitting algorithms.

The original contributions of the thesis are:

- Proposed a novel concept for measuring hand-eye coordination in MIS;
- Modification and development of the hardware essential for synchronous tracking of instruments, eyes and camera. Implementation of software algorithm to calculate the projection of the instrument tip positions on the laparoscopic video;
- Calculation of the quality of movement trajectories without prior classification of the measurements;
- Measurement of the spatiotemporal relationship between the hand and eye movement in simulated laparoscopic tasks and validation and modification of this measurement for its diversify implementations;
- Quantification of the subtle differences in visual search pattern, as a function of experience and training.

Quantification of hand-eye coordination has not been achieved before, even outside the field of surgical assessment. This is the first time where the combination of different disciplines allows the actual measure of the spatiotemporal difference between the hand and eye, compared to previous observational measures. The implementation of this method can extend to all aspects of surgery, although MIS is an obvious exemplar with disrupted spatial environment presented for the first time.

Publications in peer reviewed journals as a result of the development of this thesis are:

- Leong JJH, Atallah L, Mylonas G, Leff DR, Emery RJ, Darzi AW, Yang GZ. Investigation of Partial Directed Coherence for Hand-Eye Coordination in Laparoscopic Training. In: T.Dohi, I. Sakuma, and H.Liao (Eds). Medical Imaging and Augmented Reality (MIAR) 2008, Lecture Notes in Computer Science 5128, pp. 270-278, 2008.

- Leff DR, Orihuela-Espina F, Leong J, Darzi A, Yang GZ. Modelling dynamic fronto-parietal behaviour during minimally invasive surgery--a Markovian trip distribution approach. Medical Image Computing Computer Assisted Intervention (MICCAI) 2008, Lecture Notes in Computer Science, 11(Pt 2), pp. 595-602, 2008.

- Aggarwal R, Leong J, Leff D, Warren O, Yang GZ, Darzi A. New Technologies for the Surgical Curriculum. World Journal of Surgery 32(2), pp. 213-216, 2008.

- Leong JJH, Nicolaou M, Emery RJ, Darzi AW, Yang GZ. Visual Search Behaviour in Skeletal Radiographs: a cross speciality study. Clinical Radiology 62(11), pp.1069 – 1077, 2007.

- Leong JJ, Nicolaou M, Atallah L, Mylonas GP, Darzi AW, Yang GZ. HMM assessment of quality of movement trajectory in laparoscopic surgery. Computer Aided Surgery 12(6), pp. 335 – 346, 2007.

- Leong JJH, Nicolaou M, Atallah L, Mylonas GP, Darzi AW, Yang GZ. HMM Assessment of Quality of Movement Trajectory in Laparoscopic Surgery. In: Larsen R, Nielsen M, and Sporring J (Eds.). Medical Image Computing and Computer-Aided Intervention (MICCAI) 2006, Lecture Notes in Computer Science 4190, pp. 752 – 759, 2006.

International conference presentations:

- Leong JJH, Atallah L, Mylonas G, Leff DR, Emery RJ, Darzi AW, Yang GZ. Investigation of Partial Directed Coherence for Hand-Eye Coordination in Laparoscopic Training. Medical Imaging and Augmented Reality (MIAR) 2008, Tokyo. (podium)

- Leff DR, Orihuela-Espina F, Leong JJH, Darzi AW, Yang GZ. Modelling Dynamic Fronto-Parietal Behaviour during Minimally Invasive Surgery - a Markovian Trip Distribution Approach. Medical Imaging Computing and Computer Aided Intervention (MICCAI) 2008, New York. (poster)

- Leong JJH, Nicolaou M, Atallah L, Mylonas GP, Darzi AW, Yang GZ. HMM Assessment of Quality of Movement Trajectory in Laparoscopic Surgery. 9th Medical Image Computing and Computer-Assisted Intervention (MICCAI), Copenhagen, Denmark Oct 2006 (poster)

- JJH Leong, M Nicolaou, X-P Hu, R Emery, A Darzi, GZ Yang. Visual Search Behaviour in Fracture Detection and Localisation in Plain Radiographs. 13th European Conference on Eye Movement (ECEM). Bern, Switzerland. Aug 2005 (Poster)

- J Leong, M Nicolaou, R Emery, GZ Yang, A Darzi. Visual Search Behaviour in Fracture Detection and Localisation in Plain Radiographs. XXIII SICOT/SIROT Triennial World Congress, Istanbul, Turkey. Sept 2005 (Podium)

- Julian Leong, M Nicolaou, R Emery, A Darzi, GZ Yang. Visual Search Behaviour in Fracture Detection and Localisation in Plain Radiographs. Medical Image Perception Conference XI (MIPS), Windermere, UK. Sept 2005 (Podium)

Chapter 2

Hand-eye Coordination in Minimally Invasive Surgery

"Seeing comes before words. The child looks and recognized before it can speak" – John Berger

2.1 Introduction

Hand-eye coordination is the ability to utilise the information received through the eyes to control, guide, and direct the hands to complete a given task. This also involves the integration of body and joint position sense (proprioception), as well as continuous visual feedback in order to correct any errors made. However, the most important component in this feedback loop is the visual system (4).

Measurement of hand-eye coordination is difficult, the traditional method involves using tasks that are known to be complex hand-eye coordination problems to measure outcomes such as accuracy and reaction time. This type of measurement is indirect and has to be applied in context. For example, good hand-eye coordination in one activity does not always apply to another (5).

Others have used hand movement trajectory as a surrogate measurement of hand-eye coordination. This indirectly implies the point of reference the brain is using to perform an action, this point of reference (otherwise known as the internal representation) must transform from the input system (eyes) to the output system (hands) to ensure efficient movement executions. This transformation is clearly learned at an early age and further refined or adapted when highly skilled tasks are performed.

Hand-eye coordination is developed through infancy. By the age of 2 months, the eyes can focus well enough to follow a moving object with their gaze. By 3 months, most infants bring their hands into their field of vision, establishing the connection between the information received through the eyes as a consequence of the actions of the hands. Subsequent refinement develops through the early teenage years and in context of specific skilled tasks (6).

Minimally Invasive Surgery (MIS) poses a unique environment where the visuomotor axis is disrupted and the hand-eye (in this case, the instrument tips) coordination in the normal world no longer applies. Figure 2.1 demonstrates some of this disruption in further details, note the surgeon's action follows the blue arrow, whereas the monitor displaying the surgical field follows the red arrow. Furthermore, the assistant standing opposite the surgeon is looking at the monitor directly behind him, this would introduce complex coordinate transformation for each action to be translated into the surgical field, thus impeding the normal hand-eye coordination.

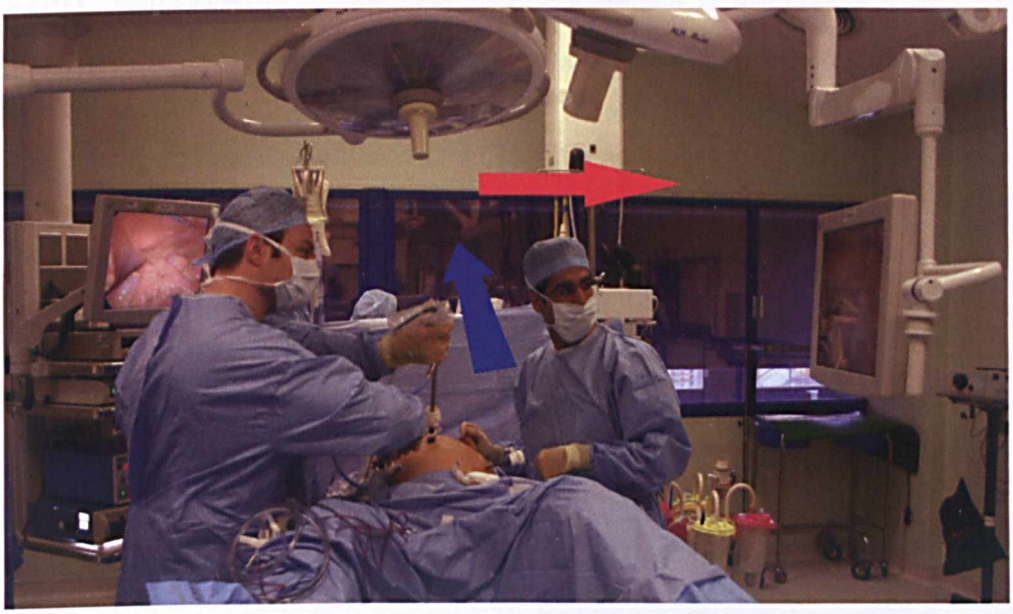

Figure 2.1 A typical laparoscopic operation scene demonstrating the motor axis (blue) and visual axis (red) pointing at different directions.

This chapter will give a background of MIS and its development, as well as important recent developments to improve the ergonomics of such surgery to improve patient safety and shorten the learning curve for trainee surgeons. Arguments will be presented that the measure of hand-eye coordination is important in trainee selection and assessment, as it

represents a direct measurement of the intrinsic skills of the surgeon. Previous work that has inspired the development of the hypothesis of this thesis will also be discussed.

2.2 Minimally Invasive Surgery

2.2.1 History

MIS simply describes the use of small incisions in surgery. In order to perform MIS, special instruments are needed. Visualisation is achieved through the use of a rigid endoscope and the image is captured in the camera attached and projected onto a 2D screen. The body cavity is insufflated using specific inert media to expand the working space. The surgical instruments are small devices attached to one end of rigid rods and controlled by handles attached to the other. These rigid instruments are then inserted into the patient's body cavity through small 5-10mm incisions.

2.2.1.1 Endoscopy

Before the 19th century, the interior of the living body could only be faintly observed through natural body orifices, using crude specula and candlelight. The first ideas of endoscopy came from Phillip Bozzini of Frankfurt, who published his theories in 1806 on its use for examination of natural openings and the organs inside body cavities. He constructed an instrument, which he called the "lichtleiter" or light conductor, which consisted of a speculum and reflectors using candle light (7).

When endoscopy permitted enough vision inside the urethra, Antonin Desormeaux was able to perform the first true endoscopic operation in 1853, removing a papilloma from the urethra. He improved the illumination by replacing candle with a lamp powered by turpentine and alcohol. In 1865, Adolph Kussmaul of Freiberg pioneered direct oesophagoscopy using a tube-shape speculum to which he had attached the endoscope of Desormeaux for illumination. As shown in Figure 2.2, for the safety of the patients, he enlisted the help of a sword swallower for the development of a safe technique. The modern endoscope was invented by Maximilian Carl-Friedrich Nitze, who tried to overcome the problem of illumination by putting the light source at the distal tip of the instrument in 1877. The main problem with this approach was the heat generated by the platinum wire inside the body cavity, which was overcome by a water cooling system, and he became the first person to insert the light source to illuminate the bladder directly (8). Between 1891 and 1894, he constructed an operating cystoscope, where the hot light

bulb was first used to coagulate a bladder papilloma, and subsequently constructed cold and hot wire loops for galvanocautery as an endoscope accessory instrument (9).

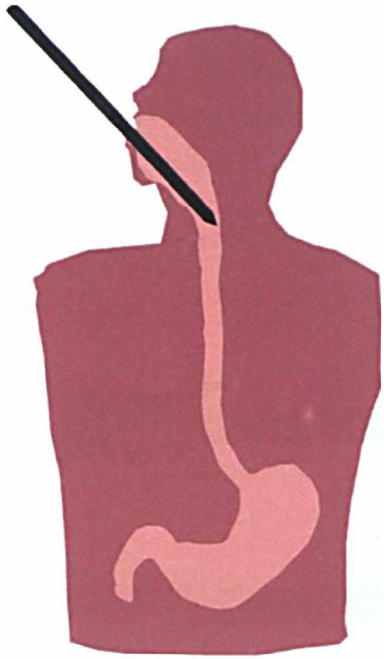

Figure 2.2 A schematic illustration of Kussmaul performing direct oesophagoscopy with the help of a sword swallower.

2.2.1.2 Laparoscopy

The first documented laparoscopy was undertaken in 1901 by Dimitri Oskarovich Ott of St Petersburg, Russia, using a gynaecologic head mirror, an external light source, and a speculum to perform the procedure (7). He described the intervention as "ventroscopy." By the end of the decade, he published his first endoscopic inspection of the abdominal cavity via a mini-laparotomy in 1909. In 1902, Georg Kelling examined the peritoneal cavity of a dog using the Nitze cystoscope. His interest was in the problem of gastrointestinal bleeding and a treatment technique called a *lufttamponade*. By insufflating the abdominal cavity with 50mmHg of air in animals, he noted that the organs had shrunk and become colourless, also halting the bleeding. He adopted the term *Kölioskopie* to describe the method for safely inducing a pneumoperitoneum as well as the location and appropriate technique for port placement (10,11).

However, it was Hans Christian Jacobaeus of Stockholm, Sweden, who coined the term *laparoscopy* and the first clinical applications of the method. In 1910, he performed

laparoscopy on 17 patients with ascites for diagnostic purposes (12). In 1912, he differentiated between laparoscopic operations on patients with and without ascites, where the operations on patients with ascites were easy and problem free. On average, the patients had between 8 and 10 litres of fluid, with a maximum of 23 litres. The operations on ascites-free patients were much more complicated, however, and the risks of intestinal injuries were significantly higher. This is of no surprise, as the design of the trocar for air insufflation was rather rudimentary, as shown in Figure 2.3 (10).

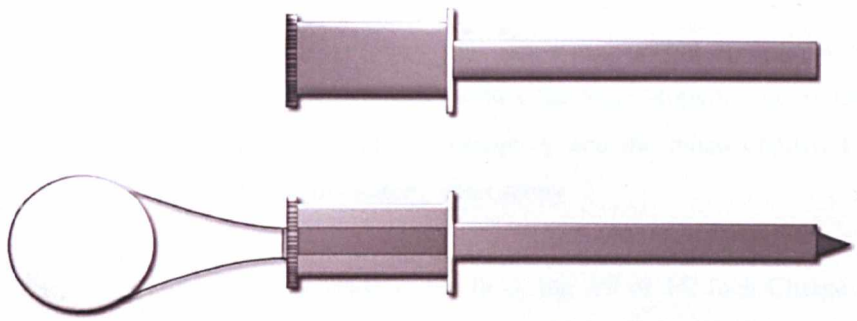

Figure 2.3 Design of the Jacobaeus' trocar.

The next advancement in endoscopy was the development of fibreoptics for flexible endoscope. John Logie Baird, most famous for demonstrating the first working television, patented the idea of transmitting images through a flexible glass cable in 1928 (British patent 285,738) (13). These ideas influenced Harold H. Hopkins and Narinder Kapany from Imperial College London who published a letter in Nature in 1954, which laid out the principles of coherent image transmission for sending images along an aligned bundle of flexible glass fibres (14). They developed the endoscope with these flexible glass fibres (fibreoptics) illuminated by a simple light bulb at the proximal end, and called this system a flexible fibrescope. In the same issue of Nature, Van Heel also presented a similar concept (15). Making use of these principles, Hirschowitz, a gastroenterologist at the University of Michigan, and his group created the first prototype of the fibreoptic instrument 3 years later for examining the stomach and duodenum. Karl Storz (1911-1996) realised that besides transmitting visual information, the system of glass fibres could be used for the purpose of light transmission, and he licensed the idea of a fibreoptic external cold light transmission (8,16).

This also coincided with Hopkins' other invention which revolutionised modern endoscopy, the rod-lens system in 1965. The traditional system consisted of a tube of air with thin lenses of glass. By contrast, the rod lens system consisted of a glass tube with thin lenses of air. The three advantages are: first, the total light transmitted is increased by factor 2.72; second, mounting the rod-lenses permitted a greater diameter to be used for lenses for a given outer diameter of the telescope, these two factors improved the optic efficiency by nine fold (8); lastly, the use of multilayer antireflection coating improved the brightness and contrast of the image. Therefore, Hopkins' rod lenses had clear advantages over the Nitze system.

Karl Storz was informed of this revolutionary idea and arranged an appointment with Hopkins, who then agreed to cooperate and produce the Storz-Hopkins endoscopes, using the cold external light source carried by fibreoptics, and the much improved rod-lens system. This system is still used in modern laparoscopy.

Many other improvements were made in the field, the 2/3 or 1/2 inch Charge-Coupled Device (CCD) camera introduced in 1969 by Bell Laboratories in the US was one of the most important. They are lightweight, low-powered, extremely sensitive image sensors, and are approximately 15 times more sensitive to light than standard regular photographic film. Video technology also permitted transitions of the laparoscopic images to one or more television sets, or even remote video displays using digital video recording technology. Clinicians were then able to document their endoscopic findings effectively with cameras and video systems.

The first laparoscopic appendicectomy was performed by Kurt Semm in Kiel on the 12th September 1980 (17). His technique greatly influenced the first laparoscopic cholecystectomy performed by Erich Mühe in 1985 (18). However, in the Society of American Gastrointestinal Surgeons (SAGES) convention in 1990, the French surgeon, Phillipe Mouret of Lyon was given credit for performing the first laparoscopic cholecystectomy. Mühe's lack of recognition was partly due to the rejection of the German Surgical Society in 1986, and also of his lack of publication in English. In 1999, he was finally recognised by SAGES for having performed the first laparoscopic cholecystectomy and was invited to present the Storz lecture (18,19). Laparoscopic cholecystectomy now remains one of the commonest laparoscopic operations performed (20) and has overtaken the open procedure as shown in Figure 2.4. Despite the obvious

advantages of MIS, the uptake of complex laparoscopic surgery is still confined to super-specialist centres, which is largely due to the technical demand of MIS.

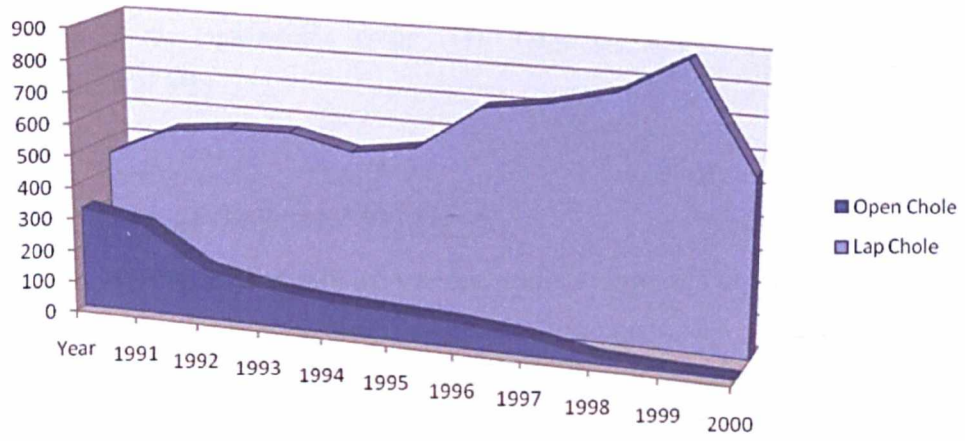

Figure 2.4 Performance of primary open and laparoscopic cholecystectomy at a teaching hospital in the United States from 1991 to 2000. [adapted from (21)]

2.2.2 Advantages of MIS

The advantages of using a series of small incisions are obvious: increased speed of recovery, reduced pain and better cosmetic results. In a recent Cochrane review comparing laparoscopic versus open cholecystectomy, it was found that there were no significant differences in mortality, intra-operative complications, minor complications, and bile duct injuries in 38 randomised trials. However, the laparoscopic group had shorter incisional wounds, shorter hospital stay and earlier return to work (22).

Advantages	Disadvantages
Smaller incision (less trauma)	Reduced depth perception
Faster recovery	Poor visuomotor axis
Less pain	Reduced tactile feedback
Lower incidence of wound dehiscence	Fulcrum effect
Lower incidence of adhesions	Reduced degrees of freedom
Better access to some areas, *e.g.* pelvis	Steep learning curve
Better cosmetic results	Longer procedure
Lower incidence of incisional herniae	Higher equipment costs

Table 2.1 Advantages and disadvantages of MIS. [adapted from (23)]

In a recent prospective cohort study, 360 patients were followed up after either open or laparoscopic hernia repair for 30-36 months. There were no differences in recurrence rates, but there were lower rates of major morbidity and post operative inpatient admission in the laparoscopic group (24). Table 2.1 summarises advantages and disadvantages of MIS:

2.2.3 Technical challenges in MIS

The technical difficulties of MIS are intrinsic to the arrangement of the instruments, the insertion of elongated instruments through small incisions, and visualisation of the operating screen on a 2D screen removed from the actual operating field. Despite the introduction of MIS over 20 years ago, many surgeons are still restricted to relatively simple procedures, *e.g.*, laparoscopic cholecystectomy and diagnostic arthroscopy, whereas complex procedures such as laparoscopic colectomy and arthroscopic soft tissue repair are still reserved for highly specialised surgeons (25).

2.2.3.1 Human factors in MIS

First, the alignment of the **visuomotor axis** can affect the performance of MIS, as demonstrated in Figure 2.1 previously, where the surgical field is transmitted remotely in a video screen often placed in a different axis than the instruments (26). The OR1™ (Karl Storz, Tuttlingen, Germany) has been developed partly to improve this alignment, where the video screens are suspended from the ceiling and can be moved freely inside the operating theatre.

Second, the instruments are inserted through small incisions which act as fulcrums. This **'fulcrum effect'** is demonstrated in Figure 2.5, showing the movement of the instruments are opposite to the movement of the hands of the surgeon. This introduces a further change in coordinate system for the surgeons to cope with. In a series of experiments, it was found that novices' performance and learning are facilitated if this "fulcrum" effect is abolished, using y-axis inversion of the laparoscopic screen image (27,28).

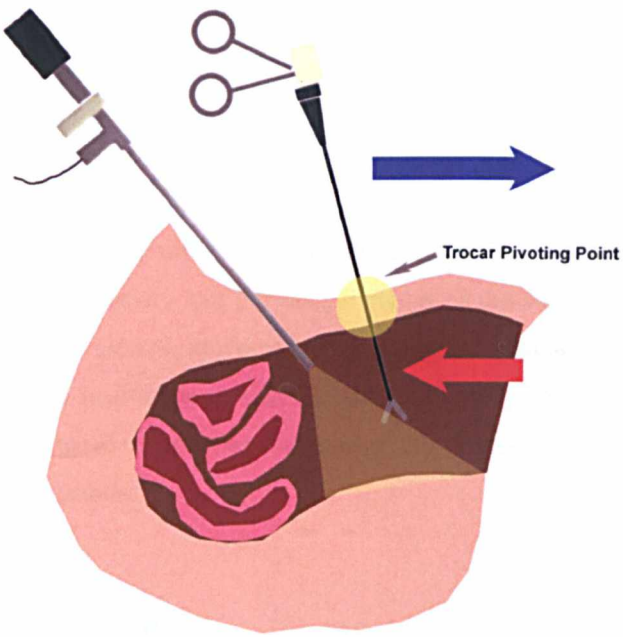

Figure 2.5 Illustrating the "fulcrum" effect – blue arrow represents the direction of movement of the instrument handle, red arrow represents the instrument tip movement, yellow circle shows the trocar pivoting point.

Third, **the lack of 3D vision** greatly impairs depth perception (29-31), surgeons often have to adopt other subtle visual cues to compensate on the 2D view. Monocular cues, such as relative size, interposition, familiar size, texture gradient, linear perspective and colour have been described by Nicolaou previously (23). Furthermore, by analysing the visual fixations of surgeons during MIS, they found that the 'invisible' shadow can be a useful cue for depth perception (31). Although earlier reports have shown no difference between the performances of laparoscopic surgery using a 2D or 3D camera, these earlier devices were more prone to induce ocular discomfort and headache (32). More recent reports, using newer technologies including the da Vinci Surgical System (Intuitive Surgical, California, USA), showed that 3D cameras can significantly improve the performance of laparoscopic novices and experts (33,34).

Fourth, the **reduced number of degrees of freedom** in the relatively primitive instrument designs. In open operations, surgeons use short instruments that are afforded a wide range of movement through the flexibility of their hands, wrists, shoulders and hips. The hand has about 27 degrees of freedom (35). In contrast, laparoscopic instruments are long and straight so that they can fit through trocars and reach remote anatomic structures within the abdomen. These trocars confine the instruments to motion within a cone,

reducing the degrees of freedom to 4. The apex of this cone is fixed in space by the trocar.

Other factors such as **reduced haptic and tactile feedback** not only increase the burden of the visual system to detect instrument collisions, but also to assess the texture of the tissue. The **scaling of the instrument movement** is determined both by the length of instrument inside the body cavity and the magnification of the laparoscopic camera; as more instrument length is inserted, movement of the surgeons' hands would translate into longer path length of the instrument tips. These all contribute to the complexity of MIS, and highlight the importance of selection and assessment of trainee MIS surgeons which are described in further details.

2.2.4 Assessment in Surgery

Surgical skills assessment has been the highlight of our Department's research for over a decade. The need for assessment is continuously motivated by high profile negligent cases, the increased public awareness, and changes in the service and training structures. Since the introduction of MIS, experienced surgeons found that they had to relearn different skills sets required, which provided further drive into the need for assessment. However, the existing assessment methodologies rely on indirect measures of technical skills, and motor skills evaluation still remains in the research domain.

2.2.4.1 Changes in Service and Training Structure

William Halstead at Johns Hopkins in 1889 developed what is now known as the "Halsteadian" model of training, which was based on the pyramidal German training system. In those days, of the 8 residents admitted in the first year, only 1 would have some prospect of obtaining full surgical training (36). The model was adapted to produce one outstanding individual. The competition for training has not necessarily improved since, although several major changes in the United Kingdom have been made in the last 15 years.

First, Kenneth Calman in 1993 introduced the concept of the National Training Number and the Specialist Registrar grade, in order to "produce a shorter, more structured and organised training pathway, so that independent clinical competence as a consultant can be achieved much earlier than in the past in many disciplines" (37). The recommendations included combining the Registrar and Senior Registrar grades, and

defining training curricula and minimum requirements for each speciality. Formal educational agreements were designed to emphasise structured learning as well as apprenticeship (38).

The European Working Time Directive (EWTD) is a directive from the Council of Europe to protect the health and safety of workers in the European Union. It lays down minimum requirements in relation to working hours, rest periods, annual leave and working arrangements for night workers. The Directive was enacted into UK law as the Working Time Regulations, which took effect from the 1st October 1998. The timeline for implementation proposed a 56 hour maximum working week by August 2007 and 48 hour by 2009, compared to the often quoted 100+ hour working week by junior doctors (39).

Modernising Medical Careers (MMC) was launched in February 2003, with the aim to establish a more streamlined, competency based training. The Senior House Officer grade needed to be modernised, as there was no limit to time spent in the grade, coupled with the lack of educational and career pathways. The new changes introduced aims to accelerate the entry into specialty training, which would lead to a Certificate of Completion of Training seamlessly (40).

Surgeons who become consultants 15 years ago would have trained for an average of 13 years and over longer hours, which would have amounted to more than 30,000 hours in training. The combined effect of the EWTD and MMC will reduce the number of hours to around 15,000 hours, also to be taken into account the decrease in case load due to the increased number of trainees (41). With these changes in place, training has to be more focused and target driven; but more importantly, progress needs to be assessed and competencies need to be achieved in an objective manner.

2.2.4.2 Technical Skills Assessment

It has long been acknowledged that technical skills acquisition is difficult to measure objectively. It is important to stress that technical skills form only a part of the complete assessment of a surgeon (42). **Knowledge, decision-making, communication** and **leadership** skills are also important qualities that should be taken into account. For centuries, the method of surgical assessment and training is based on the apprenticeship model and augmented by an examination by the official certifying body. More recently, technology has allowed further investigation into the motor and cognitive behaviour of

the surgeons, and various methodologies have been developed, albeit most of which still remain in the research domain.

Knowledge Examinations and Direct Observations

During the 16th and 17th Century, the criteria for admission to membership of the Incorporation of Barbers and Surgeons (later became the Royal College of Surgeons of the City of Edinburgh in 1778) were as follows: six years of training by master surgeons whilst giving satisfactory service, payment of a statutory fee, and most importantly, passing an examination conducted by the senior members (43). Not much has changed since then, except the content and format of the examination have adapted with time, and the fees have gone up with inflation.

Up to the time of writing of this chapter, the format for the Membership of the Royal College of Surgeons (MRCS) examination is undergoing further changes. However for the last decade, the examination consists of 2 multiple choice papers (Applied Basic Sciences and Principles of Surgery-in-General), a clinical examination, six *viva voce* examinations (Principles of Surgery, Anatomy, Pathology, Physiology, Critical Care, Operative Surgery) and a communication skills test. This is usually accompanied by an examination of the procedure logbook, which is self recorded and verified by the supervising trainers.

Viva voce examination and multiple choice questions, like the MRCS examinations and the American Board of Surgery In-Training examinations, have not been shown to correlate with technical skills level (44). Procedural logs merely represent quantity of operations performed but lack content validity. The reliability of direct observations depends on the subjectivity of the observer, however this can be made valid by addition of a criteria based scoring system (45).

Objective Structured Assessment of Technical Skill (OSATS)

The use of checklist and global performance score has shown promise in the last decade. Kopta in the 1970's demonstrated that using checklist to assess performance of orthopaedic residents showed high inter-rater reliability (46). However, it was the Objective Structured Assessment of Technical Skill (OSATS) (45) by a Toronto group in 1997, where a *global assessment score* combined with *specific operation checklists* gained the widest acceptance in skills research. In this study, the reliability and validity of these scoring scales were established in both live and bench model surgical tasks. It was

also found that the *global scoring system* had the highest reliability, which consisted of 7 categories assessing aspects of operative skill on a five-point Likert scale (47), combined with behavioural descriptors anchoring points 1, 3 and 5. A modified version of OSATS is used in Chapter 4.

OSATS is generally regarded as the 'gold standard' of skills assessment (48). Its strength lies in its simplicity, and hence the repeatability amongst different surgical specialities, even in different institutions across the world. However, it relies on the judgement of the "expert OSATS raters", so cannot be completely devoid of subjectivity. Each procedure also needs to be observed by two such raters, whether the performance is live or recorded, still adds to a significant amount of time and labour.

Motion Analysis

More automated methods of assessing surgical skills led to the development of motion tracking of the surgeons' hands. The idea developed from the observation that expert surgeons often demonstrate higher dexterity skills which relates to the concept of 'economy of movement'. The Imperial College Surgical Assessment Device (ICSAD) is based on an electromagnetic motion tracking system, using an electromagnetic field generator and two sensors that are attached to the dorsum of the surgeons' hands. Using bespoke computer software developed at the Imperial College London, kinematic data is recorded and filtered of any fine tremor. The number of hand movements made and path length travelled in standardised bench models (48), laparoscopic box trainers (49), and live laparoscopic surgery (50) were found to decrease with the experience of the surgeon.

Other dexterity measuring devices, like the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT), using a dual gimbal mechanism, equipped with potentiometers, which accepts regular laparoscopic instruments for surgical assessment, has also shown to be valid and reliable (51).

Virtual Reality

Virtual reality computer simulators have been shown to be effective both as assessment and training tools. The Minimally Invasive Surgical Trainer – Virtual Reality (Mentice, Gothenburg, Sweden), a low fidelity laparoscopic simulator, has been shown to be valid in assessment in MIS, and studies show that the skills acquired from the simulator transfers to the operating theatre (52-54). The higher fidelity Procedicus Virtual Arthroscopy simulator (Mentice, Gothenburg, Sweden) has been design to replicate the

critical steps of shoulder arthroscopy with modules testing knowledge of anatomy, navigational skills in the shoulder joint, manipulation of virtual objects and some therapeutic manoeuvres. In some of the modules, the scoring systems (based on time taken and instrument path length) have been shown to correlate with the experience in arthroscopic surgery (55,56).

However, all these skills modalities are useful to evaluate a surgeon's technical skills at the time of assessment, they lack the measure of intrinsic components which may predict superior performance in the future, which would be paramount for selecting the future laparoscopic surgeons.

2.2.5 Trainee Selection

The selection of trainee has been reliant on undergraduate and postgraduate examination results, academic excellence, interview skills and references from supervising consultants; this, however, can be subjective and lack predictive validity in further development of surgical skills. The introduction of the MMC and the accompanying selection process called Medical Training Application Service (MTAS) in 2007, has renewed further concerns about the validity of the criteria used for admission into specialist training.

Previous to 2007, candidates are short-listed for interviews based on some essential and desirable criteria, and ranked accordingly. In a 2002 – 2003 study of the candidates for a London regional interview for higher training in orthopaedics, the chances of being selected for interviews are significantly influenced by the self recorded operative experience, the specific regions and hospitals the candidates worked in, and the number of years in service since graduation. Furthermore, academic records such as publications, presentations and higher degrees did not influence the outcome (57).

In an American study performed in the same time, the top 10 selection criteria ranked by Programme Directors in Orthopaedics included: having worked at the director's institution, the United States Medical Licensing Examination (USMLE) Part 1 score, letter of recommendation by an orthopaedic surgeon, and membership in an honour society called the Alpha Omega Alpha. In terms of the letter of recommendation, the most important aspect is when the letter is written by someone known to the Programme Director (58).

These criteria are clearly not validated, and lack subjectivity. Indeed, the USMLE Part 1 score is shown not to correlate with the Orthopaedic In-Training Examination score, which is a yearly residency examination (59). Furthermore, in a recent prospective study with a 20 year follow up, A-level grades at high school have long term predictive validity for undergraduate and postgraduate careers in medicine, measuring time to achieve postgraduate membership qualifications and drop out rate (60). However, this is not specific to surgical specialities, and it would be hard to justify its use for surgical trainee selection.

The introduction of the MTAS selection process sparked many controversies and was the subject of the first junior doctors demonstration in recent years. Professor Sir John Tooke was commissioned to perform an inquiry into MTAS, and concluded that "the process used for selection into specialty training whilst promoted as theoretically sound was lacking in face validity, was rushed in implementation and was technically deficient (61)."

The General Medical Council (GMC) released a document called Tomorrow's doctors, emphasising that standards should be set for knowledge, skills, attitude and behaviour for doctors (62). Professor Tooke also highlighted the importance of a more comprehensive assessment of knowledge, skills and behaviour, as a basis for critical selection process. At present, knowledge is assessed by examinations and interview stations; behaviour is generally judged by the candidates' presentations at the interview, and also the feedback from their trainers and peers. Skills, on the other hand, are only indirectly measured using operative log and trainer assessment.

Recently, the assessment of technical skills has been introduced at the surgical selection process. Although at a survey of surgical training programmes in London, testing of visuospatial and technical abilities was piloted at selection in only one surgical department. Practical skills were tested in 3 out of 9 specialities at selection, however, only one speciality continued this assessment during higher training (63). In terms of predictive validity of these visuospatial abilities, Madan *et al* studied undergraduate medical students concerning the use of non-surgical dexterity skills to predict MIS skills. The survey inquired about typing, computer gaming, sewing, musical instruments, chopsticks, and operating tools skills. They were then tested in their performances of surgical tasks. It was found that only chopsticks skills were associated with better

performance in one of the surgical tasks, however the authors concluded that it was difficult to predict baseline laparoscopic skills (5).

2.3 Visuomotor learning

The major flaw in the assessment tools developed thus far is the complete separation of motor and cognitive components of motor skills evaluation. MIS is not just simply a motor task, it involves heavy cognitive influences that rely on visual information processing and subsequent skilled, coordinated movement. However, traditional disciplines in studying motor learning concentrate on characteristics of movement parameters rather than in-depth understanding of visual processing skills. In order to understand the cognitive aspect of technical skills development, this section attempts to summarise the body of literature on visual and motor learning separately.

2.3.1 Visual learning

In a visuomotor task, vision is mainly used for two functions: first, to search for the target for the motor system to action upon; second, to provide sensory feedback for any error in the action produced.

2.3.1.1 Visual search

During a visual search, a saccade moves the gaze of the eyes to look at the current area of interest. The time order of the fixation points represents the actual visual search that takes place, further explanations of eye movements are listed in Chapter 3.

The Feature Integration Theory was introduced by Triesman and Gelade (64), where a visual search model consisted of two different stages. In ***parallel*** visual search, the object is considered to be significantly different to the distractors, and all objects can be processed in parallel to identify the target. The parallel description implies that all of the objects are processed concurrently, and the time taken is largely **independent** of the set size. Because of this, it produces a flat Reaction Time (RT)/set size slope that is indicative of parallel searches. In **serial** visual search, the target object is not significantly different from the distractors. Each object in the search space has to be attended to separately and a decision made as to whether it is the target or not before moving onto the

next object. A steeper RT/set size slope is indicative of serial searches. This is illustrated in Figure 2.6 (65).

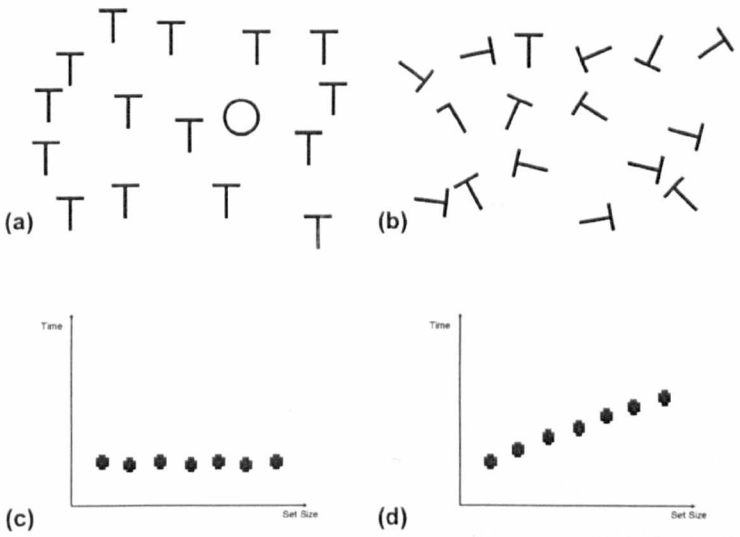

Figure 2.6 (a), (b) showing an example of a parallel and serial search respectively and their corresponding RT/set size slope (c), (d). [adapted from (65)]

Wolfe *et al* later proposed the Guided Search Model (66), which aimed to explain the continuum, rather than the strict serial/parallel dichotomy, that was observed. It envisages that there exists an early parallel stage, which closely collaborates with later serial mechanisms. The model suggests that an activation map is initially pre-attentively created to direct attention to the subsequent locations of interest in the visual field either through bottom-up (stimulus-driven) or top-down (user-driven processes). In the Guided Search Model, a ranking of stimuli is calculated by combining the information derived from the bottom-up and top-down processes. Thus, attention is thought to be directed by the visual system to highly prioritised regions.

Expert search behaviour

A more systematic framework has been proposed by Nodine and Kundel, modelling skilled search behaviour in image understanding. The global-focal model suggested four stages of search that include (i) *global impression* which is defined as the initial search using mainly peripheral vision guidance and lasts for less than 200ms; (ii) *discovery search* which utilises the information from step one, and involves a detailed inspection of

the target; (iii) *reflective search* which involves gathering evidence from cross referencing other potential targets; and (iv) *post search recall* which describes the period when the image is no longer available, and is recalled from memory (67). Appendix A studies the effect of training and specialisation on the development of the global focal search model, the study includes subjects (including surgeons and radiologists) at different experience level, and provides mathematical proof of this model.

There is evidence that it is the global impression phase that improves in experts. In a classic study, Chase and Simon found that after viewing chess position for only a few seconds, chess masters were able to reproduce these positions much more accurately than less-skilled players (68). However, when random board configurations were used, there was little difference in performance between experts and novices. They hypothesised advantage lies in the early perceptual organisation and internal representation of the chess position (68).

Using a combination of the gaze-contingent window paradigm and the change blindness flicker paradigm, Reingold *et al* further extended the theory that chess masters had dramatically larger visual spans whilst processing structured, but not random, chess positions. In addition, in a check-detection task, a minimised 3×3 chessboard containing a King and potentially checking pieces was displayed (69). In this task, experts made fewer fixations and had a greater proportion of fixations between individual pieces, rather than on pieces, as shown in Figure 2.7.

Their results provide strong evidence for a perceptual encoding advantage for experts attributable to chess experience, rather than to a general perceptual or memory superiority (69). This has indeed been found with other visual context effects (*e.g.* word, letter, object, face and scene superiority effects), a coherent and familiar context enhances the perception of constituent elements (70).

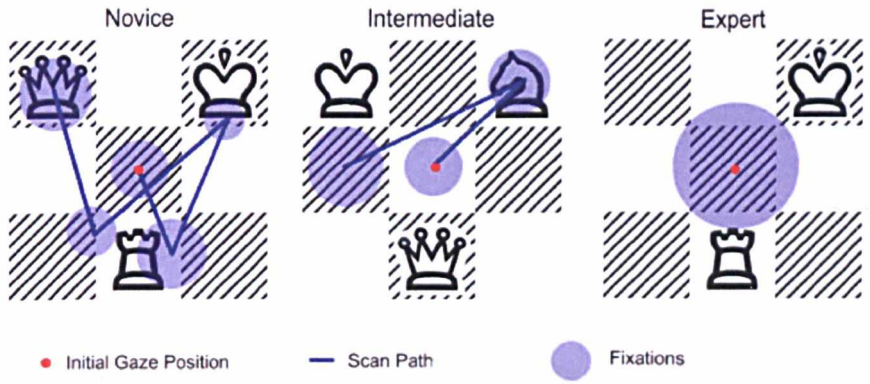

Figure 2.7 Showing a check detection task with the fixation and saccades superimposed, comparing novice, intermediate and expert in chess. [adapted from (69)]

This kind of perceptual learning theory has been extensively documented before. When first encountering a complex perceptual task, humans are typically uncertain about which are the relevant cues that will allow them to best perform the visual task. With practice, they learn to attend to visual cues that contain information and ignore those cues that are not informative. Eckstein *et al* modelled this learning using an optimal Bayesian learner, where the model learns by using the image data in the present trial to modify the weights given to a nonlinear transformation of the responses of each sensory unit in future trials. This compared similarly with actual human performance, which implied that human visual learning relies on fine tuning attention to specific visual features through previous experiences (71).

2.3.1.2 Vision used as sensory feedback

Sensory prediction errors occur when an initial motor command is generated but the predicted sensory consequences do not match the observed values. When subjects reach to a stationary target, it has been reported that gaze is anchored on the target and may anticipate the hand motion to the target. Evidence for predictive, saccadic eye movements was found in a task in which subjects had to learn coordinated hand movements to control a cursor motion on the screen. These authors found that, initially, subjects tracked the cursor with their eyes, but as they learned the task, eye movements changed to a predictive mode (72). This implies that vision is used initially as sensory feedback to correct motor errors, but as these errors decrease through learning, this feedback mechanism is less relied on.

By comparing predicted sensory events with the actual sensory events, the visual system can monitor task progression and adjust subsequent motor commands if errors are detected. This sensory feedback is not visual alone, but includes tactile, proprioceptive and audio feedback. However, there is good evidence that visual feedback is most important in the initial learning period (73).

2.3.2 Motor learning

Motor learning is a multifaceted set of internal processes through which there is a relatively permanent change in human performance through practice, provided the change cannot be attributed to a human's maturation, temporary state, or instinct. Though motor learning is usually implied by improvement in performances, it is not directly observable (74).

2.3.2.1 Fitts and Posner model

Paul Fitts and Michael Posner presented the classic learning stages model in 1967 (75). The first stage, the **cognitive stage** of learning, the learner focuses on cognitively oriented problems. Performance is characterised by a large number of errors, and the errors are not consistent. There is generally a lack of awareness of the ways to improve performance.

The second stage of learning is called the **associative stage** of learning. There is generally a variable amount of practice for transition into this stage. The learner associates specific environmental cues with the movements required to achieve the goal of the skill. They make fewer and less gross errors since the fundamentals or mechanics of the skill have been acquired, although they need to be improved. This is also known as the refining stage, in which the individual focuses on performing the skill successfully and being more consistent from one attempt to the next.

The last stage of learning is the **autonomous stage** of learning. Here the skill has become almost automatic, or habitual. People in this stage do not consciously think about the specific movement characteristics of what they are doing whilst performing the skill. They can often perform another task at the same time. Performance variability during this stage is very small, additionally these skilled performers can detect their own errors and make the proper adjustments to correct them.

2.3.2.2 Gentile's two-stage model

This model was proposed by Gentile (76), in contrast to Fitts and Posner, motor skill learning is progressing through at least two stages and they are presented from the perspective of the goal of the learner, as illustrated in Figure 2.8.

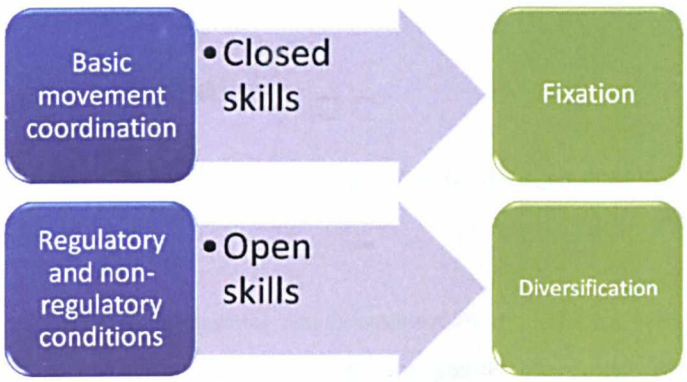

Figure 2.8 Gentile's two-stage model showing basic movement coordination, regulatory and non-regulatory conditions used in the model.

In the **initial stage**, the learner has two important goals to achieve. One is to acquire basic movement coordination to achieve the goal, these movements must also match the regulatory conditions of the environment context in which the skill is performed. Regulatory conditions refer to those characteristics that determine the movement characteristics to achieve an action goal, for example, the size and shape of the cup are important as one reaches and grasps it.

The second goal of the learner is to discriminate between regulatory and non regulatory conditions in the environmental context. Non regulatory conditions are those characteristics of the performance environment that do not influence the movement characteristics required to achieve an action goal, such as the colour of the cup.

In the **later stages** of learning, the learner needs to acquire three general characteristics. First, the person must be adaptable to perform in different situations. Second, the person must increase their consistency in achieving the goal of the skill. Third, the person must learn to perform the skill with an economy of effort.

This is characterised by the concept of fixation and diversification, which depends on the type of skills learnt. Closed skills require fixation of the basic movement coordination pattern acquired during the first stage of learning. This means that the learner must refine this pattern to achieve consistency. On the other hand, open skills require diversification of the basic movement pattern. An important requirement is to adapt quickly to the continuously changing spatial and temporal regulatory conditions of the skill.

2.4 Hand-eye coordination

"Coordination is an act of managing interdependencies between activities" (Malone and Crowston 1991).

Hand-eye coordination is the mental transformation of the effector's coordinates, as a feedback to the visual paradigm, which in turn guides movement generations. This fundamental skill is dependent on the integration of multiple sensory (visual, tactile and proprioceptive) and motor (hands, shoulders and body) systems, and the complexity of this combinatory problem lies in the numerous possible interactions between them. Although it is crucial in everyday tasks, individuals such as professional sportsmen are renowned to excel in hand-eye coordination, it is the precision of their actions when reacting to a visual or tactile stimulus that make them successful (77-79).

In order to reach a visually perceived target, the brain must transform visual information into coordinates appropriate for movement (80). Therefore, in order to study hand-eye coordination, the relationship between these coordinate systems should be discussed in further details.

2.4.1 Coordination systems

Coordinate systems are different ways of representing a point in the 3D space. The most commonly used system is the Cartesian coordinate system, named after René Descartes (1596-1650), where the position of a point is represented by three orthogonal coordinates. Another way of representing a position is the Polar coordinate system (or the spherical coordinate system in 3D), where the coordinates are expressed as the distance from the origin and the angles from the baseline.

2.4.2 Basic model of hand-eye coordination

It is believed that the coordinates of a target is perceived visually as direction and distance from the eye, hence in the polar system where the hand would be $(\theta_{hand}, r_{hand})$ and the target would be $(\theta_{target}, r_{target})$ as shown in Figure 2.9. There is substantial evidence that this eye-centred representation of the world is used directly for motor control, hence influencing the trajectory and error of the final hand movements (81-83).

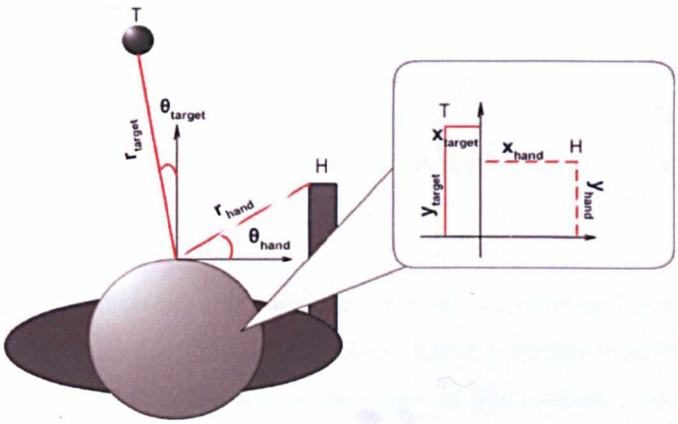

Figure 2.9 Illustrating the coordinate transformation from an eye-centred reference in Polar Coordinates to an Internal Representation in Cartesian Coordinates H – hand, T – target.

However, others have suggested that the eye-centred representation is transformed into a master map of coordinates, the Internal Representation (IR) (84), which is believed to be the mental estimation of the distance between two targets, probably in the Cartesian 'world' coordinate (*x*, *y*) (85). For simplicity, it is assumed that the origin of the IR and the visual system are the same.

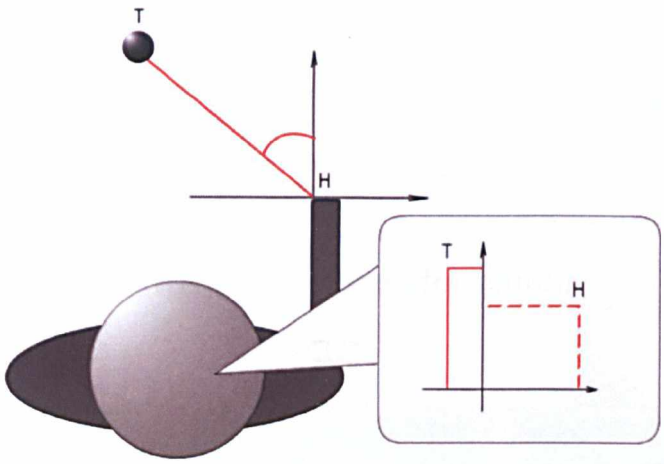

Figure 2.10 Illustrating the transformation of the Internal Representation to hand-centred Polar Coordinates.

In order to move the hand towards the target, movements are generated in a hand-centred (or shoulder-centred) polar coordinate system (83), hence a further transformation from the IR (Cartesian coordinate). However, as the origin of this rotation is at the hand, not the head (the origin of the visual world), a translation of the origin is needed to be considered, as illustrated in Figure 2.10.

This is a much simplified model of the transformation from a head origin visual system to an internal representation, then back to a hand (or shoulder) centred motor coordinate, as summarised in Figure 2.11. The assumptions are that: first, the origin of the visual system is at one point, however the eyes' movement in relation to Listing's plane and the binocular systems introduce further complications; second, there is no movement of the head or the body introduced in this model; third, no error is introduced in the reaching task; forth, joint position sense has not been accounted for, although recent report showed that this is of limited influence (4). The term eye-centred representation is also misleading. In binocular vision, a virtual cyclopean eye located midway between the left and right eyes is often used as the reference point (86).

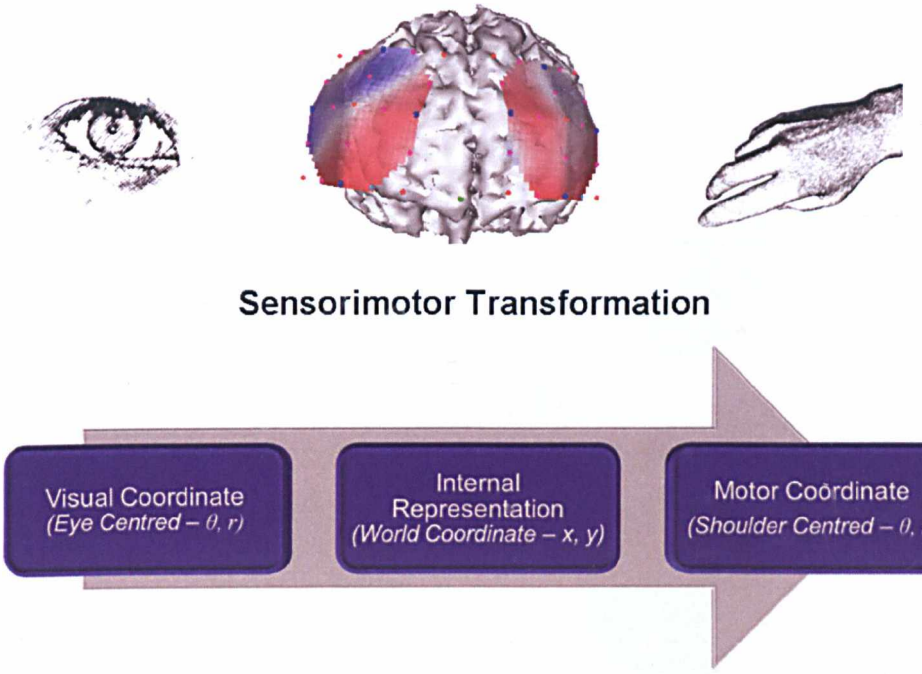

Figure 2.11 Simplified summary of coordinate transformation from visual to motor coordinates.

2.4.3 Listing's law

The eyes rotate with three degrees of freedom, this further complicates any coordinate transformation as described above. This means that the eyes can rotate about the following axes: 1) a vertical axis to generate horizontal eye movements (abduction and adduction), 2) a horizontal axis to generate vertical eye movements (elevation and depression), and 3) the line of sight to generate torsional eye movements (excyclotorsion and incyclotorsion). Figure 2.12 illustrates these movements. In theory, the eye could assume an infinite number of torsional positions for any gaze direction (87).

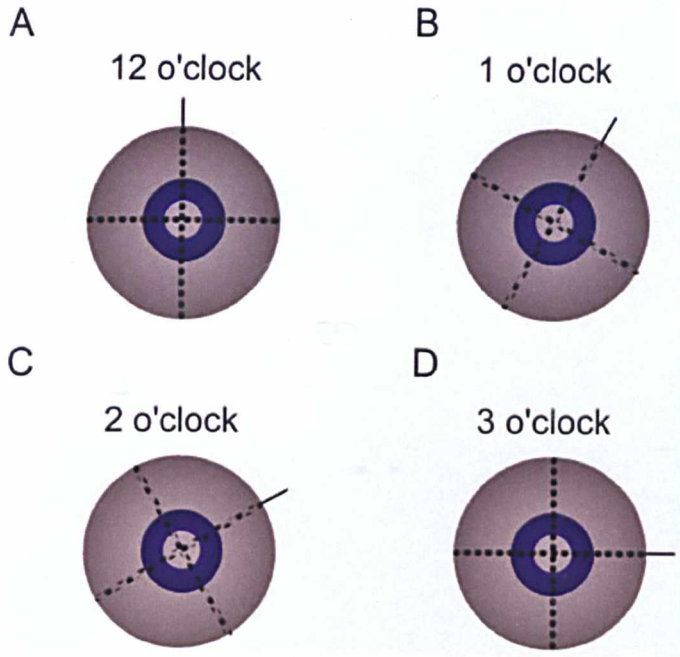

Figure 2.12 (A) A schematic of an eye directed straight ahead at the reader, with the thick black vertical (solid) line represents its superior pole, which is at 12 o'clock. There are many different torsional positions that the eye can adopt when it looks straight ahead: (B) at 1 o'clock, (C) 2 o'clock, (D) 3 o'clock, and so on.

Listing's law states that, when the head is fixed, there is an eye position called the primary position, such that the eye assumes only those orientations that can be reached from primary position by a single rotation about an axis in a plane called Listing's plane. This plane is orthogonal to the line of sight when the eye is in primary position. In other words, one can visualize any given eye movement as caused by rotation about an axis, and the collection of these axes constitutes Listing's plane, as shown in green in Figure 2.13 (87,88).

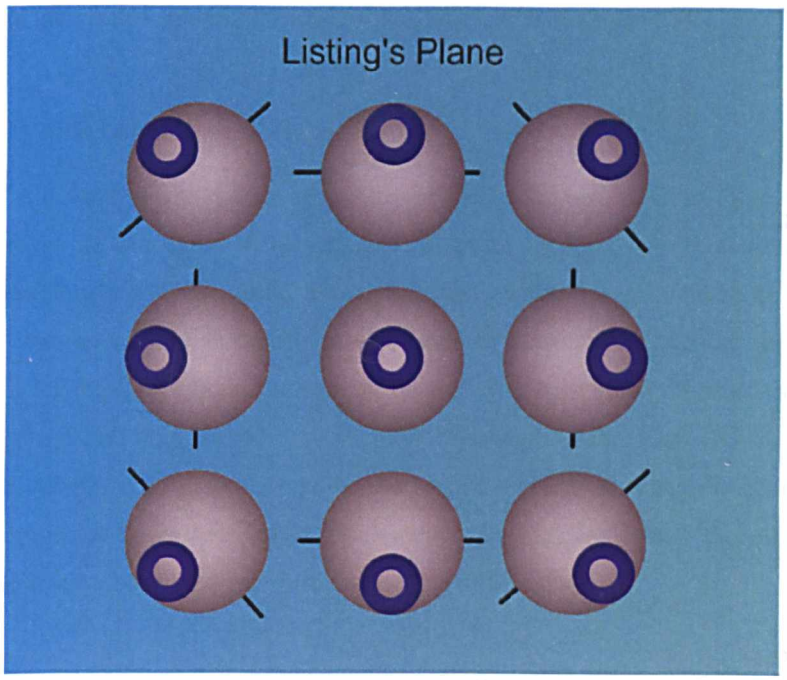

Figure 2.13 Illustrates Listing's Law, note the axes (the bars protruding from the eyes) used to rotate from centre to various eccentric positions are confined to a common plane (green). [adapted from (87)]

2.4.4 Internal Representation

The Internal Representation (IR) of the world coordinate greatly influences the outcome of the motor program generated, there is still ongoing debate as to where the origin and in which coordinate the IR is based on. Many authors propose that specific sensorimotor maps are used for internal representations, and these maps are probably unique to the tasks performed (89,90). Some studies suggested that this is based on the context of the movement required and others provided evidence that some movements no IR is needed, where there is a direct transformation from the eye-centred coordinates to hand centred coordinates (91). As evidence emerges that the IR develops from birth, and continues to adapt when the required motor functions and the environment changes, even in adults and in extreme conditions (89).

It is hypothesised that that during skilled manual movement in MIS, the point of reference (hence, the IR) adapts to produce the most efficient motor control. Measurement of the

trajectories of the effectors, *e.g.*, the hands or instruments, can provide an insight into the final coordinate transformation utilised for this specific visuomotor tasks.

2.4.4.1 Which coordinates: eye or hand?

There are still some disputes between which reference frames are used for sensorimotor transformation, as illustrated in Figure 2.14. In order to point to a remembered visual target, transformation of binocular visual information into appropriate motor output is necessary. However, there is conflicting evidence to demonstrate whether the reference frame used for the motor output is centred on the eye or the hand. Furthermore, studies have suggested that this motor output could be coded in the Cartesian or the Polar coordinate system. As the final stage of movement generation is muscular contraction, and these muscles have their stable attachment points either around the elbow or the shoulder, perhaps a shoulder-coordinate system is utilised instead.

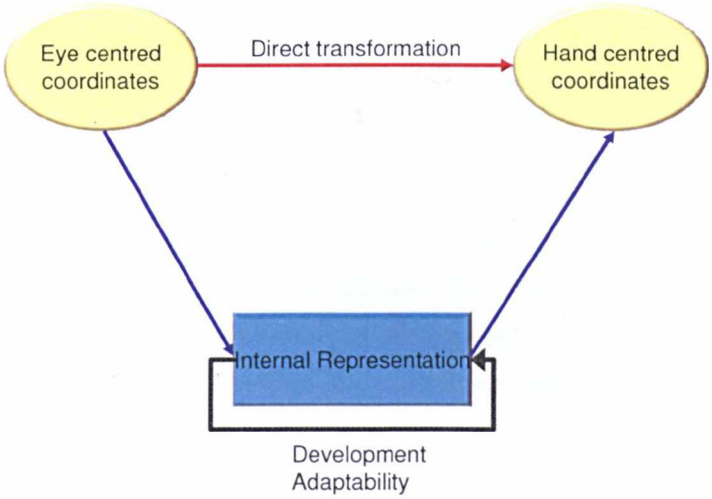

Figure 2.14 Illustrates the options of coordinate transformation.

Eye centred coordinates

An eye-centred target must be linked to initial hand position before a motor program can be formulated that brings the hand towards the target. In a comprehensive set of experiments, where subjects were required to perform pointing movements to remembered 3D targets after a memory delay using different distances and directions of the target, starting positions, head rotation and the effector hand. McIntyre *et al*

demonstrated that the orientation of pointing errors indicated an eye-centred Cartesian reference was used by the subjects (83).

In a single cell recording experiment, measuring Posterior Parietal Cortex (PPC) neuronal activities in monkeys performing a reaching task, where the targets locations are manipulated in eye and limb reference frames. They found that in a specific area in the PPC, the parietal reach region, neuronal activation was more consistently correlated with an eye-centred rather than a limb-centred reference frame (81).

Hand centred coordinates

Analysis of movement errors in reaching tasks has also been examined by others where the initial hand position is unknown, in 2D (92,93) and 3D (94,95) reaching tasks. They found that the estimation of hand position without vision is consistently biased, and the resulting pointing errors reflect significantly on the extent this erroneous estimation of the initial hand position. Thus concluding that aimed hand movements are planned as direction and distance in reference to the initial hand position, hence the hand-centred Polar reference was used by the subjects.

Context based coordinates

Although there seems to be conflicting reports on which frame of reference is used for motor planning, evidence seems to support a gradual transformation from an eye-centred to body-centred and hand-centred coordinates. However, the frame of reference ultimately used depends on the availability of the information of the position of the hand and/or the target, whether the movements are constrained (see Figure 2.16), and whether cognitive influence is involved.

Carrozzo *et al*, in a virtual reality 3D pointing task, found that the availability of visual feedback of the hand affects the frame of reference used (96). By analysing the end-point of each trajectory, the spatial distribution of the variance is calculated and modelled into ellipsoid distributions, where their orientation and shape were determined by eigenvector and eigenvalues of the matrix respectively. As illustrated in Figure 2.15, the Seen Hand task, the axes of maximum variability and of maximum contraction of the ellipsoid distribution converge toward the mid-point between the eyes (the virtual cyclopean eye). This shows that with visual information of the hand available, an **eye-centred reference** was used. In the Unseen Hand task, the axes of maximum variability rotate anti-clockwise

around the body and the effector arm, hence a **body-centred or arm-centred** egocentric representation was used.

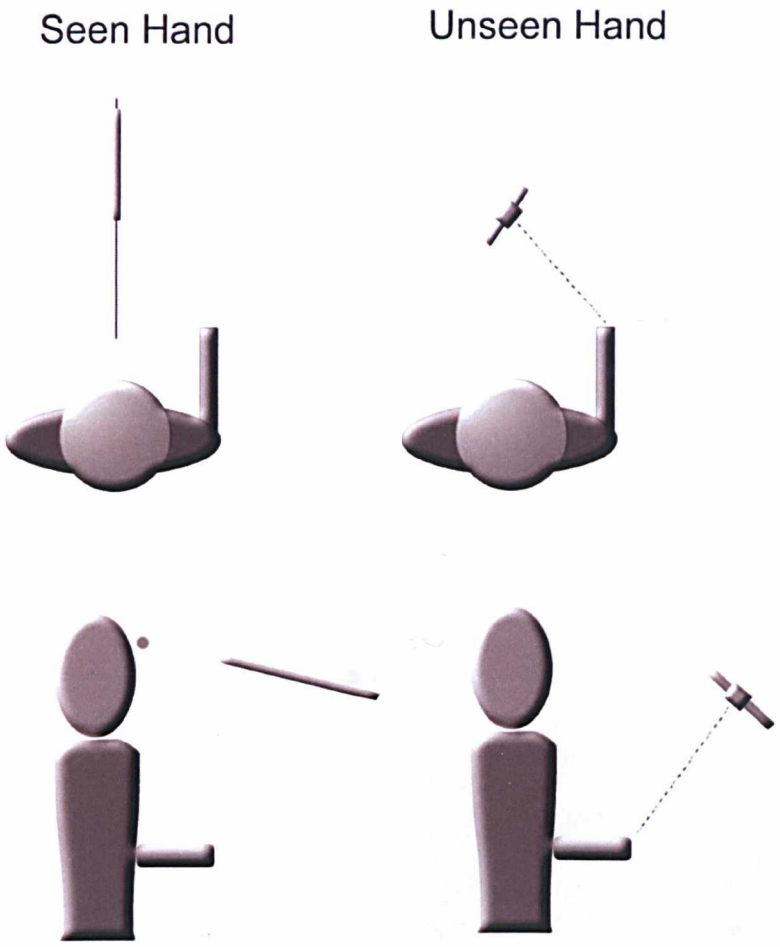

Figure 2.15 Average variable errors in the Seen (left) and Unseen Hand (right) conditions, viewed from above (top) and from the left side (bottom). Ellipsoids represent the 95% tolerance region of all responses. They are centred on the average final finger position. The dark bars emanating from each ellipsoid indicate the axes of maximum variability. Note that these axes converge toward the eyes (in Seen Hand), but rotate anti-clockwise around the body and the effector arm in the Unseen hand. [Adapted from (96)]

Desmurget *et al* demonstrated other factors that influence the control strategies used for goal-directed movement. They first measured pointing movements in two conditions: in the "compliant" condition, the subjects held a hand-held cursor that they displaced towards the target; in the "unrestrained" condition, they used their index fingertip freely towards the target. It was found that the spatiotemporal characteristics of the compliant and unconstrained movements were fundamentally different, where the former produced

straight spatial path, and the latter trajectories were curved. This demonstrated the control strategies differ according to the task.

Interestingly, when the subjects were instructed to move along a straight-line path, this modified the characteristics of the "unrestrained movement" as shown in Figure 2.16 (97). Even in the "unrestrained" movement, near straight trajectories were produced. It seems that both "external constraint factor" and "cognitive factor" can modify the outcome of the coordinate transformation.

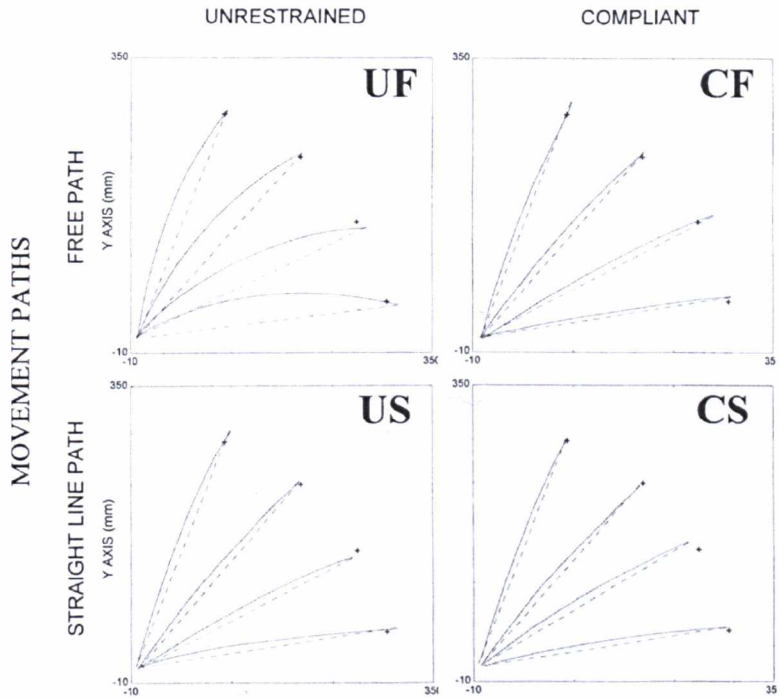

Figure 2.16 Movement trajectories with unrestrained and compliant movements, and when asked to follow a straight line path (2nd row). Notice the change in trajectories when the movement was unrestrained, but following a straight path. UF – unrestrained free path, CF – compliant free path, US – unrestrained straight line, CS – compliant straight line. [reproduced from (97)]

2.4.4.2 Direct transformation from eye coordinates to hand coordinates

Some authors have suggested that coordination transformation via the IR is redundant, and that perhaps a more direct approach would be more efficient. Buneo *et al*, in a unit recording task in monkeys, suggested that this comparison is done at an earlier stage in eye-centred coordinates (91). They found that in area 5 in the PPC, remembered target locations are coded in both eye and hand coordinates. This suggests that the PPC transforms target locations directly between these two reference frames by-passing the

body centred reference. Data obtained in an adjacent area indicated that this transformation may be achieved by computation of a hand "motor error" vector in gaze coordinates, as illustrated in Figure 2.17.

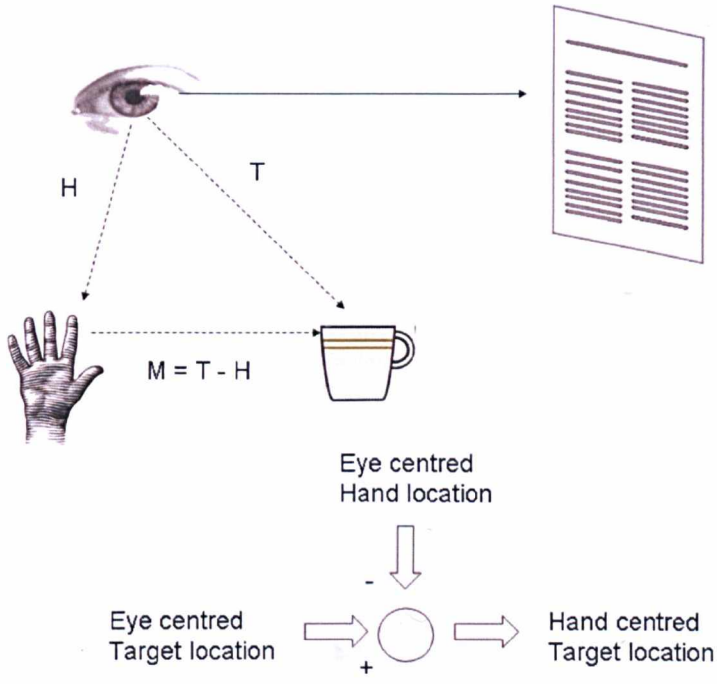

Figure 2.17 Visuomotor transformation schemes. An example of reaching for a cup while fixating on a newspaper. The position of the cup is represented in the brain in terms of its location on the peripheral retina (T). To reach for the cup, its position with respect to the hand must be known (M). This information could be acquired by directly subtracting hand position (H) from target position (T) in eye coordinates. [Adapted from (91)]

This has important implications, as eye-centred representation of hand error could be directly used as motor commands for hand motion without requiring further comparison with eye and head positions. This gives rise to a direct transformation for reaching.

2.4.4.3 Internal Representation: allocentric or egocentric?

In an ***egocentric*** frame of reference, locations are represented with respect to the particular perspectives of a perceiver, whereas an ***allocentric*** reference frame locates points within a framework external to the holder of the representation and independent to their positions (98).

Egocentric frame of reference is adequate for movement control, provided the eyes (or the head) stay still. However, as soon as the eyes shift in gaze, a disruption is introduced in the spatial relationship between the sensory apparatus and the external world. To overcome this, representations of the external world must be stored either independent to the eye movement or are continuously updated to compensate for eye movement. It is thought that the hand-eye coordination system constructs both egocentric and allocentric representations of visual space, depending on various factors including the available sensory information, task constraints, visual background, memory interval, and the cognitive context (86).

In the 3D virtual reality pointing task described earlier by Carrozzo *et al*, the target was presented in the complete absence of allocentric visual cues as to its position in space. The subjects were informed prior to the experiment that all targets would fall on a "virtual" straight line, although this was never explicitly shown to them. The pattern of their movement errors revealed both egocentric and allocentric components of this "virtual" object, consistent with the hypothesis that target information can be defined concurrently in both egocentric and allocentric frames of reference, resulting in two independent, coexisting representations (96,99).

Interestingly, through studies on generating allocentric mental maps by blind and sighted subjects, Eardley *et al* found that subjects who were born totally blinded had no difference between their performances with sighted subjects (100), implying that vision is not necessary in generating allocentric spatial representation.

2.4.4.4 Development of sensorimotor maps in children

Sensorimotor maps are not in born, but are slowly learnt and refined from birth. Prior developmental studies have shown an important developmental milestone between 6 and 8 years of age (101,102), by measuring accuracy of movement, reaction and movement time.

In a study where four, six and eight-year-old children performed line drawing tasks from a common centred position ("centre out task"), under normal (pre and post-exposure), and rotated (exposure) visual feedback conditions. During the tasks, direct vision of the hand was occluded, and indirect feedback of the pen position was provided by a computer screen with a ***45° clockwise rotation***. The movement parameters (time, accuracy, smoothness) improved with increasing age. But on introduction of screen cursor rotation,

a counter clockwise "spiral-like" pattern emerged in all groups (columns 2 and 3 in Figure 2.18).

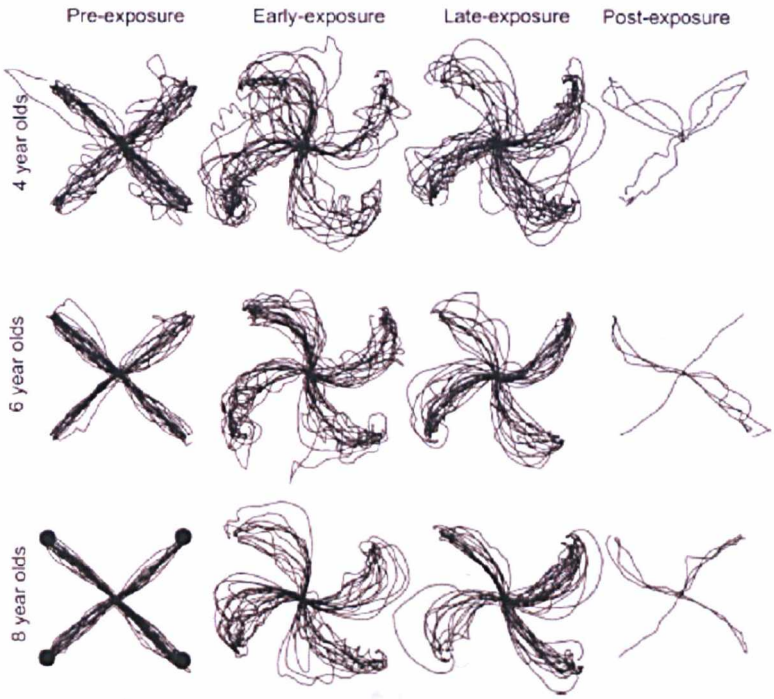

Figure 2.18 Movement trajectories during pre-, early-, late-, and post-exposure to the 45° screen rotation, as a function of age. [reproduced from (101)]

In the post-exposure period, only the older children appeared to show consistent "spiral-like" patterns that **mirrored** those seen during exposure (column 4 in Figure 2.18). The absence of after-effects observed in the younger age group (4 year olds) suggest that these children might have less developed (*i.e.* more broad) internal visuomotor representations for hand movements, and that their IR's are sharpened (*i.e.* tuned) with visuomotor experience (101). The older groups seemed to be able to learn new transformations in response to the rotated visuomotor mappings.

2.4.4.5 Adaptability of sensorimotor maps in adults

The tuning and acquisition of new visuomotor mappings are also evident in adults. A study examined the directional biases in reaching movements on the initial position (left, middle or right) of the unseen hand, again using a "centre out task" (92). It was found that when the initial hand positions were to the right of midline, movements were systematically biased clockwise. Biases were counter clockwise for starting points to the

left. These biases were temporarily eliminated with vision of the hand before the movement or the cursor during the movement. This indicates that visual information about hand location relative to the body is necessary for accurate specification of movement direction.

In the second stage of the experiment, the subjects were trained to reach accurately from either the right or the left starting positions. After training, movements showed a reduction of bias reflecting an adaptive change in feed-forward commands generating the movements. Interestingly, biases were still found, but now the centres of the new biases were shifted to the trained position (92).

2.5 Measuring Hand-eye coordination

Normal hand-eye coordination involves the synergistic function of several sensorimotor systems including the visual system, vestibular system, proprioception, and the eye, head and arm control systems, plus aspects of cognition such as attention and memory. Even if the measure of hand-eye coordination is the understanding of the sum of all these system, the task would appear rather daunting. Hand-eye coordination is in fact even more than these, it evokes combinatorial problems that do not arise when individual systems are studied in isolation. Ultimately, the aim of the study of hand-eye coordination is straightforward: the use of the visual system to guide movements of the hand. This fundamental fact is the best tool in understanding the function of the whole system (86).

2.5.1 Eye movement measurement only

Land studied eye movements in everyday tasks, and found that in general the eyes search ahead, gathering information for the effectors to carry out tasks (103). However, in skilled motor tasks, eye movements are sometimes different from the actions that are generated, these are usually to fulfil specific cognitive needs. For example, when driving a car round a corner, the gaze lands on the inside lane edge (tangent point) of the next bend, as its location relative to the driver provides direct information about road curvature, and hence how much to turn the wheel (104). Comparing professional and amateur cricket batsman, expert eye movement pattern seemed to have faster and more accurate predictions of where the ball lands and hence determining the timing and arc of the swing of the cricket bat (77).

2.5.2 Spatiotemporal relationship between hand and eye

It is still a fundamental question, whether visual information is used predictively or reactively. If the eye movements predict the hand movement, vision must be used as a feed forward system, and *vice versa*.

2.5.2.1 Feed forward or feedback

In robotic controlled systems, feedback systems significantly reduce the complexity of visuospatial transformation, the effector essentially driving to the point where the visual error is reduced to zero (86,105). However, this is only possible when the movements are relatively slow, and robotic sensory feedback is limited by the speed of electrical flow and computer processing time.

Miall and Reckess examined the role of the eye movement as a feed forward system (106). A forward model is a representation of the object being controlled, and is used to predict the outcome of actions, a prediction of the future state of the effector (*e.g.* hand) after a motor command that has been issued. As temporal delays in the sensory and motor pathways suggest that for many human movements, coordination must depend on a predictive state estimate, rather than on feedback signals. They further implied that signals from the manual control system appear to be sent to the ocular control system, to provide predictive information about the required hand movement, and this may be coordinated in the cerebellum when observed in patients with cerebellar disease.

In a series of cursor tracking experiments, Miall and Reckess showed that the performance is best when the hand followed the same trajectory as the eyes, but with a latency of about 75-100ms. If the eyes lead by a greater time offset, manual tracking performance is degraded, presumably because the hand then moves too early, compared to its own target. If the eyes lead by a lesser amount, or lag behind the hand, performance is again degraded; presumably because the predictive information is not available early enough to be used by the manual tracking system. These experiments, however, assumed that the eyes follow the visual cue provided, and eye tracking data was not confirmed in all experimental conditions (106).

This is not to say that sensory feedback has no role in hand-eye coordination. Ghaharamani *et al* examined local remapping of coordination transformation (80), when visual feedback of finger position was limited to one or two locations, at which

discrepancies were introduced between the actual and visually perceived finger position. They found that these local remapping changed the pointing behaviour, and these changes are maximal near areas where the discrepancies were introduced.

2.5.2.2 Measuring spatiotemporal relationship

In order to measure this combinatorial behavioural pattern of hand-eye coordination directly, the relationship between the ***position*** of the hand and the eye in the observation or actual performance of a task is studied.

Numerous studies have implied that in visual guided actions, task specific proactive eye movement (feed forward system) are crucial for planning and control (107,108). However, the eye movements become reactive, when the movements of the objects are not predictable (109). In a study with a block stacking task, when the subjects are **performing** the task, the coordination between their gaze and the hand is **predictive**, rather than reactive. It was found that on average, gaze arrived 150ms earlier than the index finger. An interesting observation arose when the subjects **observed** the block stacking task being performed by an actor, where no part of whom could be seen by the subjects. The actions of the actor became unpredictable, and the gaze exhibited a very similar spatial pattern, but eye movements were markedly **delayed** relative to the index finger as shown in Figure 2.19 (110).

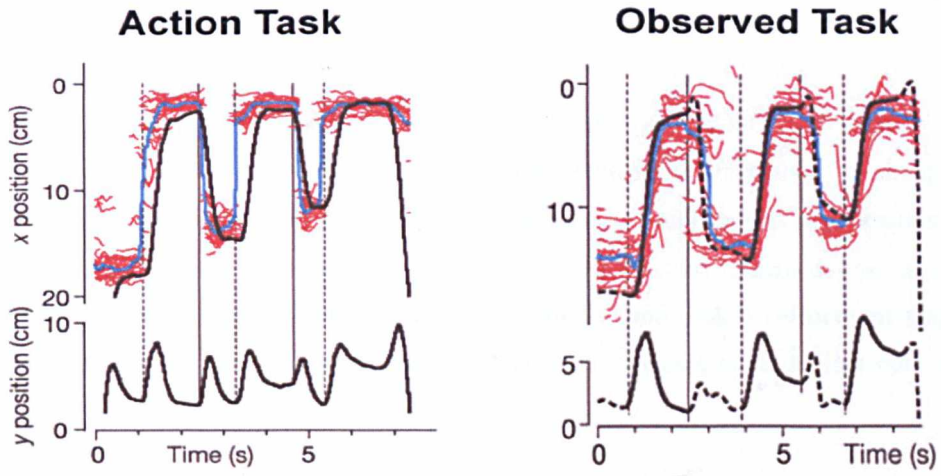

Figure 2.19 **Right**: gaze behaviour in block **observation** task. **Left**: gaze behaviour in **action** task. Median horizontal (x) positions of gaze (blue) and the index finger (black) as a function of time. Red traces represent the raw saccadic data on the x axis. Median vertical (y) positions of the index finger are shown in the bottom subplots in black. [reproduced from (110)]

2.5.2.3 Learning a new visuomotor task

Sailer *et al* further extended this framework in learning of novel visuomotor task (72). They used a novel apparatus to control the position of a cursor on the screen, in order to hit a series of consecutively appearing target boxes. This apparatus is a freely held box with a handle at each end, and the subjects were required to apply opposite rotational torque to the two handles to move the cursor up and down, whilst compressing and stretching forces moved the cursor horizontally. Making oblique cursor movements with the tool was very difficult, and required a combination of longitudinal forces and torques. This is illustrated in Figure 2.20.

Figure 2.20 Illustrating the apparatus used by the subject to control a cursor on the screen. Compressing and stretching forces (green and yellow arrows) control horizontal cursor movements, and torques applied around the longitudinal axis of the apparatus (red and blue arrows) control vertical movements. [adapted from (72)]

The results showed that the subjects learnt this new task in about 20 minutes. This learning process seemed to occur in stages that could be distinguished by changes in performance (target-hit rate) as well as by gaze behaviour and hand-eye coordination. The three distinct stages: an **exploratory stage** in which hit rate remained low; a skilled **acquisition stage** in which the rate increased rapidly; and a **skill refinement stage** in which the hit rate increased more slowly. These three phases have distinct patterns of both cursor control and gaze movement.

Exploratory stage

During the exploratory phase, most cursor and gaze movements were either horizontal or vertical, reflecting the difficulties in making oblique movements. Gaze generally

followed the cursor, with occasional glances to the target. The saccades sizes were generally small 3-4°.

Skill acquisition stage

In the skill acquisition phase, the subjects learnt to move the cursor obliquely. The gaze continued to lag behind the cursor at the beginning, although this changed to a predictive behaviour, leading the cursor by up to 300ms. Saccades sizes were larger (4° to 12°), and more were directed towards the target. This indicates that subjects started to program spatially congruent eye and hand motor commands.

Skill refinement stage

The final skill refinement stage, gaze went directly to the next target, with either a single saccade, or a large and one small saccade.

It is hypothesised that during the exploratory stage, the learner attempts to establish basic rules for visuospatial transformation. As these rules are established in the acquisition stage, a gradual transition from a monitoring to an anticipatory role for the eyes is observed, when vision acquired a feed forward rather than a feedback role.

2.5.2.4 Using spatiotemporal relationship to examine specific motor tasks

Similar hand-eye behaviour has been observed in a study comparing tracing and drawing tasks. During tracing, the pen tip and eyes were tightly coupled, with participants making a series of small saccades just in front of the moving pen interspersed with periods of smooth pursuit. However, during drawing, saccades were fewer and larger and pursuit was less frequent (111).

Tracing depends on external cues such as visual feedback from the eye which is used to monitor the pen tip position in relation to the traced line. Drawing employs internal cues such as memory to a greater extent, guiding the hand direction rather than closely monitoring and comparing its progress.

2.6 Implications for Minimally Invasive Surgery

The complexities of hand-eye coordination were highlighted above, and have been the subject of decades of research. MIS introduces changes in both the sensory input and

motor output aspects of hand-eye coordination. The operating field is transformed through the endoscopic camera into a 2D screen, with intrinsic and extrinsic characteristics of the camera lens introduced into the equation. The hands of the surgeons act indirectly through the endoscopic instruments, the instrument tip movements are subjected to changes in angular rotation through the incisions. In MIS, the coordination is between the surgeons' eyes and the instrument tips projected onto the endoscopic monitor system.

2.6.1 Endoscopic camera system

3D live images of the internal organs are transmitted to a 2D monitor remote from both the operating site and the surgeon. This 3D to 2D transformation introduces 3 broad categories of coordinate changes:

Intrinsic factor of the camera

The internal properties of any camera lens are defined by its magnification, calibrated focal length, principal point location, radial lens distortion, and de-centring lens distortion (see next chapter for further details). Amongst these, magnification of the endoscopic lens is most likely to affect the perceived movements of the instrument tips, in relation to the surgeons' hands. Although magnification was introduced to improve the visibility of the finer details of the operating scene, this also increases the scale of instrument tip movement when projected onto the endoscopic monitor. Radial distortion describes the non-uniform magnification of the lenses, as a compromise to increase the field of view. In wider angled lenses, especially in fisheye lenses, the image magnification decreases with distance from the optical axis, as in Figure 2.21. Theoretically, as the surgical instrument tip moves further away from the centre of the lens, it has to travel a further distance to produce the same displacement on its image projection on the endoscopic monitor. There has been some attempts to correct this distortion, but the relationship between this and hand-eye coordination was not explored (112).

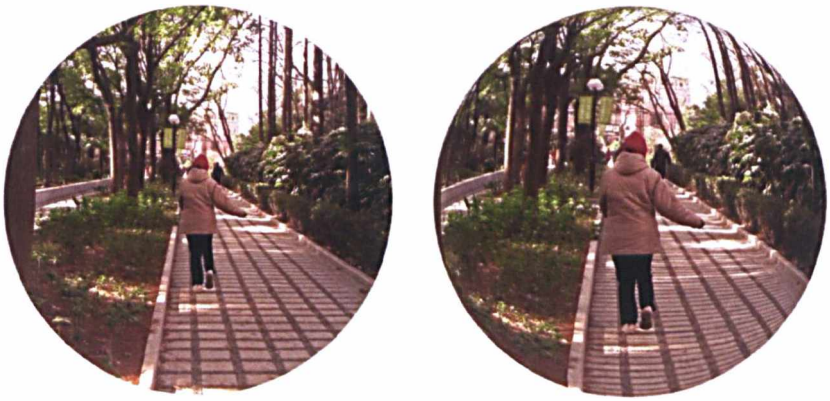

Figure 2.21 The effect of lens distortion; the left image shows a picture without distortion, and the right image illustrates the effect of radial distortion.

Extrinsic factor of the camera

This describes the spatial position of the camera in relation to the operating scene, and this relationship is highly variable in MIS. During surgery, the endoscope has 4 degrees of freedom for movement inside the body cavity. As the endoscope moves, the operating scene can be subject to magnification, translational, rotational changes. Of these, the effect of rotation is most likely to affect hand-eye coordination in MIS, and will be further discussed in Chapter 4.

Visuomotor axis of the surgeon and the monitor

As illustrated in Figure 2.1, the surgeons are not located directly in front of the laparoscopic monitor, often the motor axis (the direction of the instrument) is in 90° angle from the visual axis (the direction of the surgeons' gaze) producing further coordinate changes (113). As described earlier, when the working space is in the left or right of the midline, movements are systematically biased counter-clockwise and clockwise respectively. In time, this movement is adapted and recalibrated to the intended motor output program.

2.6.2 Endoscopic instruments

The effect of the indirect translation of the surgeons' hand movements through the endoscopic instruments have been described before. Briefly, the insertion point of the laparoscopic instrument acts as a pivot, and hence producing the "fulcrum" effect (27). If the handle and the tip of the instrument are fixed to the incision, the kinematic link between the handle and the tip is spherical. The scale of the instrument tip movement is also dependent on the relative length of the instrument inserted through the incision, as

the instruments are inserted further, the surgeons' movements are more magnified. Lastly, the incision of the abdomen does not provide a perfect pivot for the instruments, as the length of this incision increases, the pivot can be subject to translational movements. These are summarised in Figure 2.22.

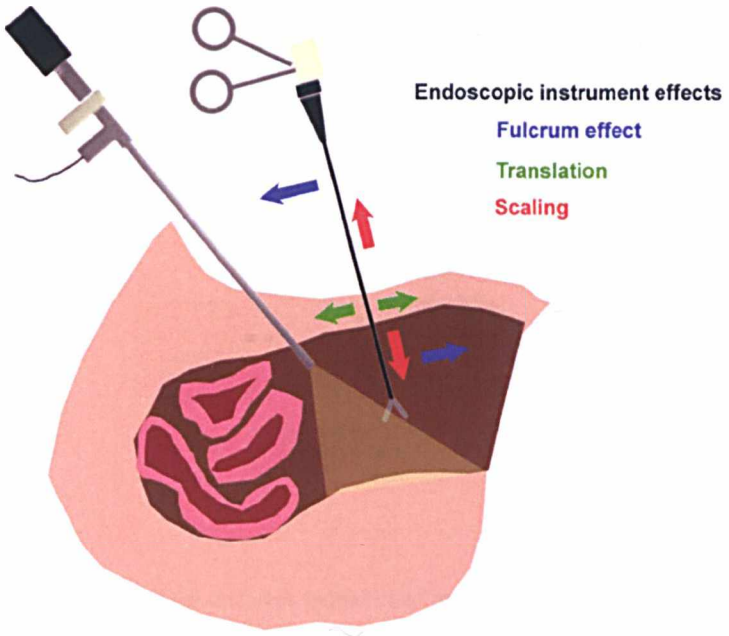

Figure 2.22 Illustrating the effect of the endoscopic instrument on coordinate transformation: blue arrows show the "fulcrum" effect, red arrows show the effect of instrument length on scaling of its movement, and green arrows show the effect of translation at the incision site.

2.6.3 Internal Representation of MIS

It is not the scope of this thesis to model all the coordination system transformation necessary to perform laparoscopic surgery, as shown in Figure 2.23; however, this section has illustrated the complexity of the problem. It is logical that multiple sensorimotor maps are generated during MIS training, and these maps are fine tuned to improve performance. Previously in this chapter, it has been shown that the IR can be both egocentric and allocentric, it is not inconceivable that one of the sensorimotor maps is referenced on the image projection of the instrument tips, and perhaps this will explain the steep motor learning curve in early laparoscopic training.

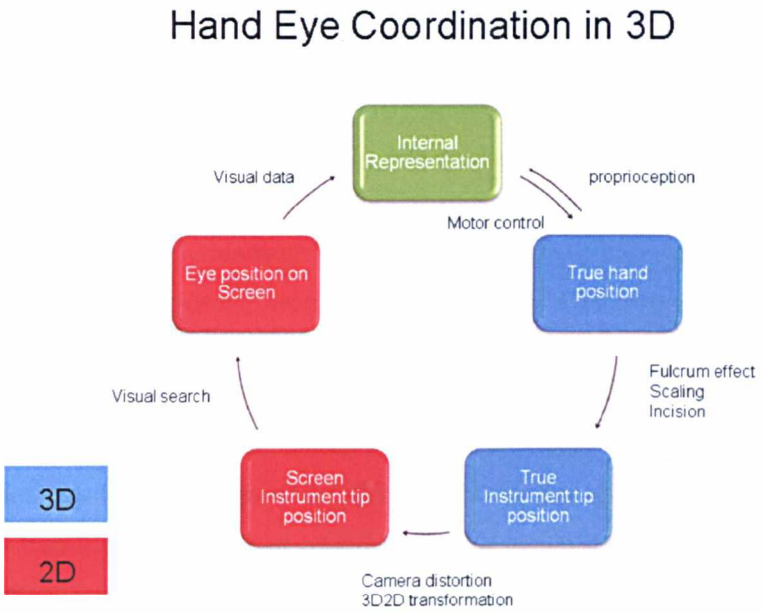

Figure 2.23 Summary of the different spatial transformations required in MIS

2.7 Conclusions

The historical development of MIS is presented in this chapter, its popularisation in the last decade meant that experienced surgeons had to relearn different skills set in order to be competent in surgery. Despite its advantages, in the wrong hands MIS can become potentially dangerous. The steep learning curve is due to the intrinsic constraints of the visualisation method and the ergonomics of the instrument design as highlighted.

Assessment in surgery is still largely based on traditional methods established centuries ago, despite recent development of objective skills measurement methodologies, their widespread adoption remains elusive. The selection process, so far, lacks data to demonstrate predictive validity in technical skills development. Indeed, development of large longitudinal trials following trainee surgeons' progress would provide the most convincing evidence, however it is likely that technology would have moved on by the completion of these trials. Objective assessments of basic technical skills have been reasonably well established, but the lack of cognitive component in these tests may have

deterred its general use in the surgical community. Perhaps a more in depth research into the cognitive aspect of technical skills can provide an alternative solution.

Eye movement experiments done previously in the Department have uncovered specific gaze parameters which reflect on the cognitive need to assist in MIS. Visual cues, such as shadow perception, have been found to be useful in depth perception, overcoming the constraints of the 2D environment (31). Other studies have concentrated on the spatial coupling between the eyes and instrument tips (114), understanding the reliance on the visual system to generate movements in the laparoscopic instruments.

This chapter has introduced two concepts in relation to hand-eye coordination: reference transformation and spatiotemporal relationship. In order to perform a visuomotor task, hand and target locations have to transform from a visual-centred reference frame to a motor-centred reference frame. Evidence shows that these so-called sensorimotor maps develop with age, and continue to adapt when novel visuomotor tasks are learnt. Further study suggests the involvement of the PPC in the direct transformation from eye-centred to hand-centred reference. In MIS, it is likely that a novel sensorimotor map is developed through training, where camera distortion and laparoscopic instrument ergonomics are accounted for. This development can be defined by measuring the quality and error of movement trajectories (96).

The spatiotemporal relationship between the eye and the hand can be used to measure hand-eye coordination. In learning a new visuomotor task, initial reactive eye movements are used for feedback mechanisms for motor error. This is later replaced by predictive eye movements implying improved initial visuomotor prediction accuracy, through better developed sensorimotor maps. A detailed framework of this transition has been described in the chapter, dividing learning into three distinct phases. These methods may provide an objective measure of the cognitive development of hand-eye coordination.

Currently, the measure of hand-eye coordination remains confined to the laboratory settings. Studies demonstrating reference transformation rely on analysing movement trajectories, whilst visual input is either distorted or occluded, the generated movement error and variability are then measured. Although this provides a perfect platform for validating the concept of eye- or hand- centred reference frame, the analysis is limited to simple reaching tasks where the dimensionality of the data is low. Complex bimanual procedures, encompassing multiple reaching and manipulation tasks, remain too complex

for this approach. In MIS, additional variables such as camera distortion, physical constraints of the instruments, and inter-task variability all combine to influence the final movement trajectories.

Only observational data has been presented in measuring the spatiotemporal relationship between the hands and the eyes. Eye movement patterns are highly variable, not only subject to physical but also psychological influences. These movements are characterised by large high velocity saccadic jumps, followed by smooth trajectories of pursuit eye movements. Currently, there are no algorithms for automatic calculations of the relationship between the eye and the intricately related hand movements. Hence, the feasibility for application of this measurement is low outside the laboratory.

Future work needs to focus on the development of algorithms for these measurements. Highly variable data is difficult to evaluate using simple statistical tests, other approaches have been used in biological data, as in speech or gesture recognition. These approaches include the use of stochastic processes, where events are classified as states, and the transitions between them are governed by probabilities rather than deterministically. Furthermore, time series data comparisons are not only confined to medical sciences, other disciplines such as computer sciences and econometrics frequently employ signal matching and causality computations. These can be applied to the calculation of the spatiotemporal relationship between the eye and the hand movement data streams. Finally, cognition and attention are important influences on hand-eye coordination, although difficult to assess directly, functional brain imaging can add further dimension and contextual meaning to the interpretation of the results.

The next chapter will focus on the basic principles of eye and hand position tracking, and the development and adaptation of the hardware best suited for the experimental setup in this thesis. One of the greatest challenges is to transform the tracked 3D instrument position into its location projected onto the 2D laparoscopic monitor screen, applying the distortion produced by the camera and taking into account its position and orientation. Interestingly, this 3D to 2D transformation closely mirrors the processing in the surgeons' brain, where the 2D screen trajectories of the instrument tips are interpreted into 3D hand movements. In order to understand the basis of eye movement studies, a brief description of the human visual system will first be provided.

Chapter 3

Eye and Instrument movement tracking

In the previous chapter, an overview of the development and difficulties in MIS was presented. Despite its obvious advantages, the ergonomic factors intrinsic to MIS have led to a much steeper learning curve than that of traditional open surgery. This, coupled with the changes in service and training structures, has led to calls for tighter regulations in the selection and assessment of trainees aspired to perform MIS. Further arguments were presented to integrate the assessment modalities with cognitive measurements, as this could theoretically become a more comprehensive assessment, and perhaps a more direct measurement of the intrinsic ability of the individuals.

In order to quantify hand-eye coordination, direct measurements of the eye and hand movements need to be obtained individually. This chapter will provide an account of the development and selection of the hardware and software components, and their constant improvement in an attempt to measure the subtle relationship between the hand and eye movements accurately and in a completely unobtrusive way during MIS.

There are three main challenges for the measuring device: first, eye gaze and instrument tip position data need to be measured individually; second, the two data streams then need to be synchronised temporally; third, the instrument tip positions in the 3D world need to be translated into 2D data as projected on the laparoscopic screen, this requires simultaneous tracking of the laparoscopic camera position. Each of these challenges is a topic of research by itself, the final measuring device is a combination of various commercially available components coupled with bespoke software implementation of visual processing algorithms.

As described in the previous chapter, during MIS there are different eye movement behaviours, some related to cognitive search for target selection and others to smooth pursuit for feedback of instrument handling. To understand these behavioural patterns, it is necessary to understand the human vision system, and the purpose of different types of eye movements.

3.1 Purpose of eye movement

The human fovea is an area located in the retina with the highest concentration of photoreceptors, hence providing the best visual acuity. The foveal vision normally represent around 1-2° of visual angle, and the visual acuity drops off dramatically from the centre of focus, this is illustrated by the distribution of photoreceptors in Figure 3.1.

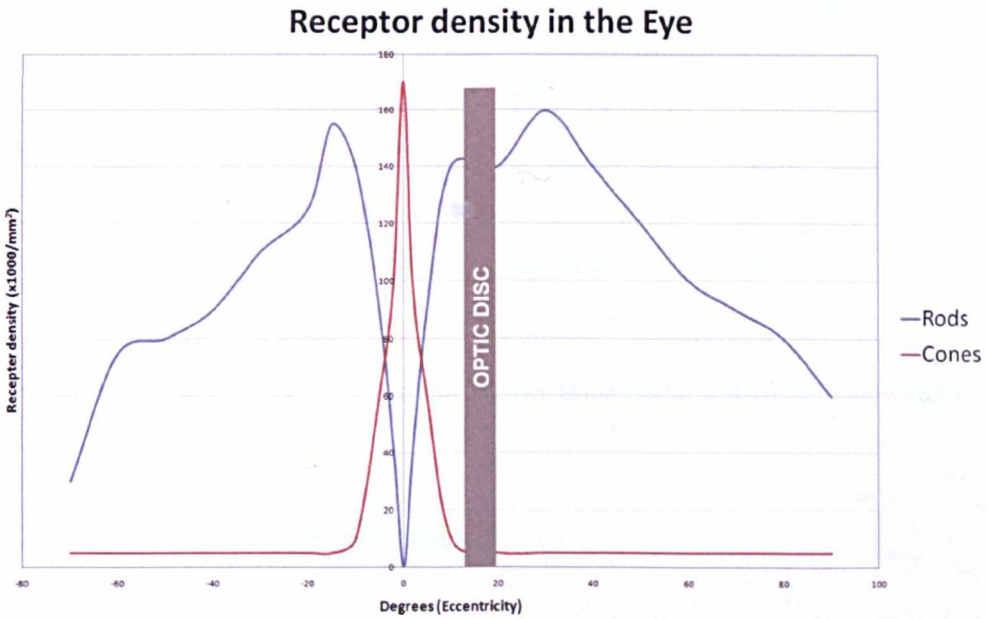

Figure 3.1 Diagram illustrating the distribution of photoreceptors as a function of distance from the fovea (0 degree). [redrawn from (115)]

Human vision is reliant on the small angle of the fovea, this highlights the necessity of the ability to direct the gaze to targets of interests quickly and accurately. Thus the resultant saccadic eye fixations are a very instructive behaviour, revealing much about the underlying cognitive mechanisms that guide them (116).

The eyes move within three degrees of freedom, *i.e.* three rotations within the eye sockets. There are six extrinsic muscles responsible for eye movement: the medial and lateral recti (for lateral movements), the superior and inferior recti (for vertical movements), and the superior and inferior oblique (for rotating on the visual axis). These are illustrated in Figure 3.2.

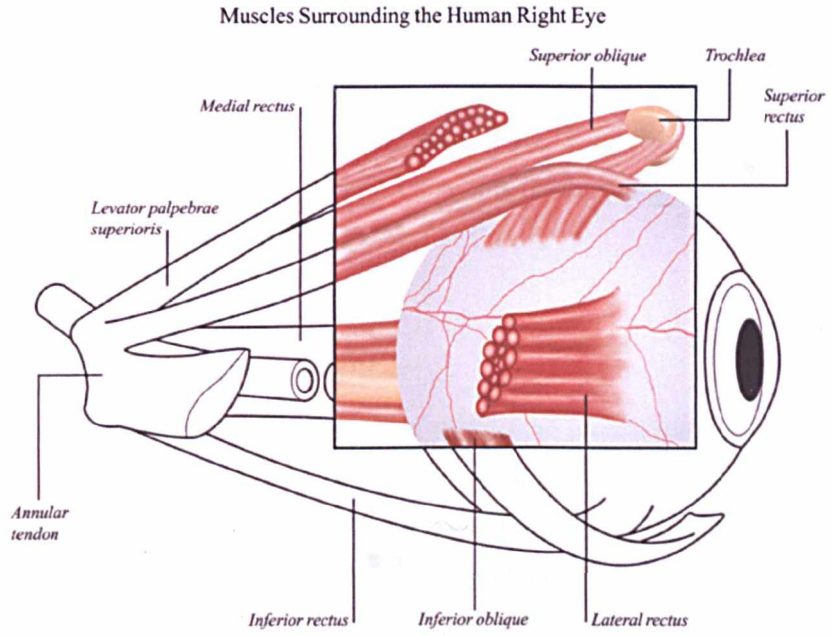

Figure 3.2 Illustration of the six extrinsic muscles of the eye for controlling its movements.

3.2 Physiological eye movements

3.2.1 Saccadic eye movements

The common eye movements include saccadic, smooth pursuit, miniature and optokinetic (65). The term *saccade* comes from an old French word meaning "flick of a sail". These are the voluntary rapid eye movement to direct the fovea to a specific point of interest. It has fast acceleration at approximately 40,000°/s^2 and a peak velocity of 600°/s (65). These eye movements are thought to be ballistic, as there is insufficient time for visual feedback to guide the eye to its final position. The speed is achieved by the control system in the brainstem, the excitatory code consists of a burst of very high-frequency activity, that throws the eye to its new position, followed by a tonic signal that holds the eye in position once it's there. However, others have suggested an internal copy of head, eye and

target position is used as feedback to guide the eyes during a saccade. There is evidence that the intermediate and deep layers of the superior colliculus may have critical components of the neural circuitry initiating and controlling saccadic movements (117).

The speed of saccadic eye movement can be explained by Listing's law (see Chapter 2) which specify an essentially null torsion component in eye movements, requiring virtually only two degrees of freedom for saccadic eye movement.

Saccadic eye movement provides important information that implies the underlying cognitive process. In static imagery search, the patterns of saccadic eye movement and fixations define the visual search process. Using this information, visual search theories can be quantified, as studied in Appendix A.

3.2.2 Smooth Pursuit Movement

These are smooth involuntary eye movement that acts to track a moving object. Pursuit movements provide an example of a control system with built-in negative feedback. Visual search for a moving target usually involves a fast saccadic jump for location of the target, followed by smooth pursuit movement for tracking.

3.2.3 Miniature eye movement

These involuntary eye movements have long been the characteristics of fixation. There are three types: tremor, drift and microsaccades. Tremor is a periodic wave-like motion of the eyes occurring at a frequency of 90Hz and is thought to be independent in the two eyes. Drifts are slow motions of the eye which occur simultaneously with tremors, and occur between microsaccades. Microsaccades are movement signals that are more or less spatially random varying over 1 to 2 minute of arc in amplitude. It has been thought that these miniature eye movements may present "noise" in the control system, or used for compensation or corrective eye movement. However, recently these small positional variations of eye movement are thought to contribute to the hyper acuity in human visual system, when compared to the physical constraints of the eye (*i.e.* the relatively low resolution of the fovea) (118).

3.2.4 Optokinetic nystagmus

These are sawtooth-like eye movements with a combination of pursuit movement interspersed with saccades invoked to compensate for the retinal movement of the target. Vestibule-ocular reflex is a similar type of eye movement compensation for the movement of the head, when an image is moved back and forth on the retina, the eyes compensate by equal and opposite movement to hold the gaze on the moving object.

3.3 Eye tracking hardware

One of the first documented studies of eye tracking was published in a psychology journal in 1901, based on corneal reflection techniques (119). However, the most famous early studies using eye tracking was probably by Alfred Yarbus in 1967. In his experiment, subjects viewed Repin's painting, the "unexpected visitor" returning home after a long absence, to demonstrate the task-dependence of eye movement patterns. Yarbus reported that subjects' eye movement patterns varied dramatically with instruction, depending on whether subjects were simply examining the painting or were preparing to answer a specific question (120). The resulting eye movement data is illustrated in Figure 3.3.

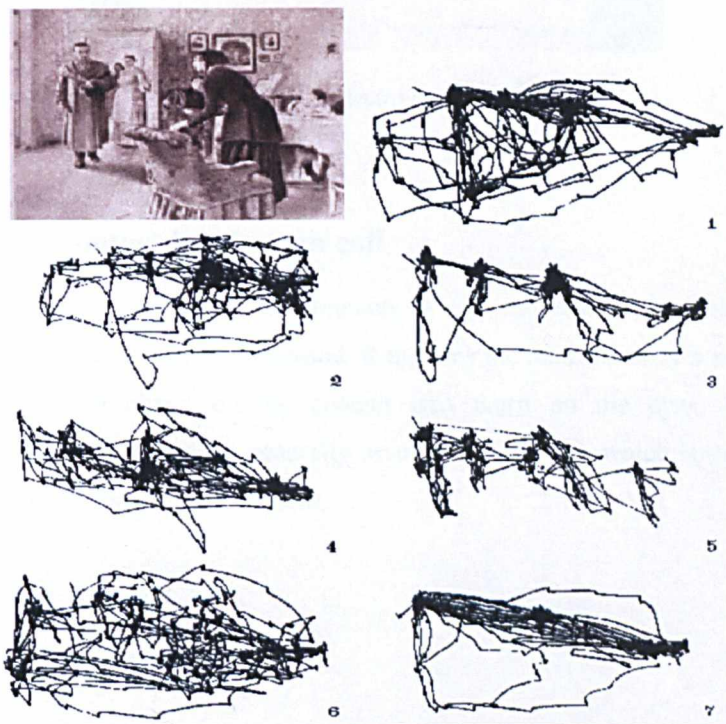

Figure 3.3 Illustrating Yarbus' experiment in 1967 when subjects were asked to look at the painting and were given specific tasks. [reproduced from (120)]

There are three main categories of techniques for measuring eye movements: Electro-oculography, scleral contact lens/search coils, and techniques based on light reflected from the cornea such as Video-OculoGraphy (VOG) (65).

3.3.1 Electro-oculography

Electro-oculography is based on the measurement of electric potential difference of the skin around the eyes using electrodes attached, as illustrated in Figure 3.4. Its accuracy is quoted to be around 1.5°, and was once a popular method for eye tracking. Head movement has to be measured concurrently if the point of gaze is to be determined. There are, however, less invasive and more accurate systems available now.

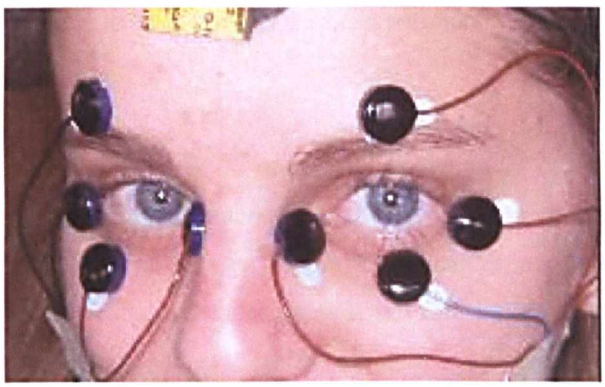

Figure 3.4 Electro-oculography.

3.3.2 Scleral contact lens/search coil

It is one of the most accurate measurements of eye movements, and also relatively straight forward if 3D eye data is required. It requires the attachment of a mechanical or an optical reference object on the contact lens worn on the eyes. The modern implementation of this technique generally involves a wire coil, which is then measured moving through an electromagnetic field.

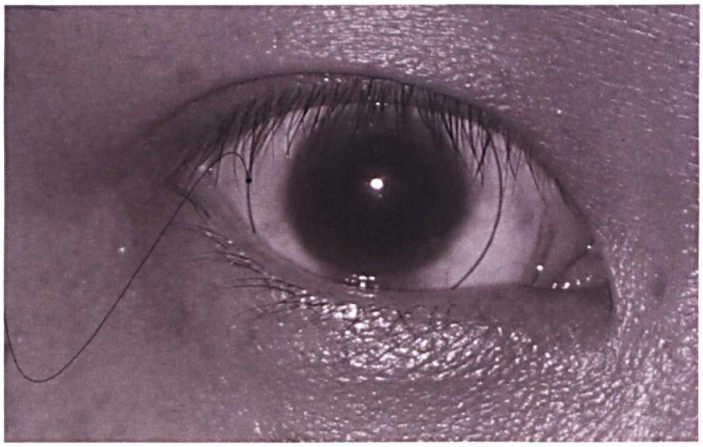

Figure 3.5 Illustrating scleral contact lens with search coil.

It is clearly shown in the illustrated photo in Figure 3.5 that this technique is the most invasive, and is less likely to be acceptable to the experimental subjects.

3.3.3 Video-oculography

This is perhaps the most popular modern method of eye tracking due to its accuracy and non-invasiveness. The general category of this technique involves measurement of distinguishable features of the eye under rotation and translation. These features include the limbus (the iris-sclera boundary), corneal reflections of often a closely positioned Infra-RED (IRED) light source, or the pupil.

Automatic limbus tracking can be achieved using photodiodes, and is only effective in horizontal movement, as the eyelids obscure vertical movement of the limbus. Iris tracking poses less of a problem, and is generally better defined optically.

A more robust method of tracking is the combination of pupil/iris interface tracking and corneal reflection. The corneal reflection of a light source (generally IRED) is known as the Purkinje reflection, and is relatively stable with rotation of the eyes. Since the relative position between the pupil and the Purkinje reflection remains constant with minor head movements, it is therefore possible to determine the point-of-regards without physically fixing the head (65). This setup can tolerate a minor amount of head movement, whilst other techniques for tracking the head (optically or electromagnetic) are sometimes used to supplement the accuracy of the eye tracking data.

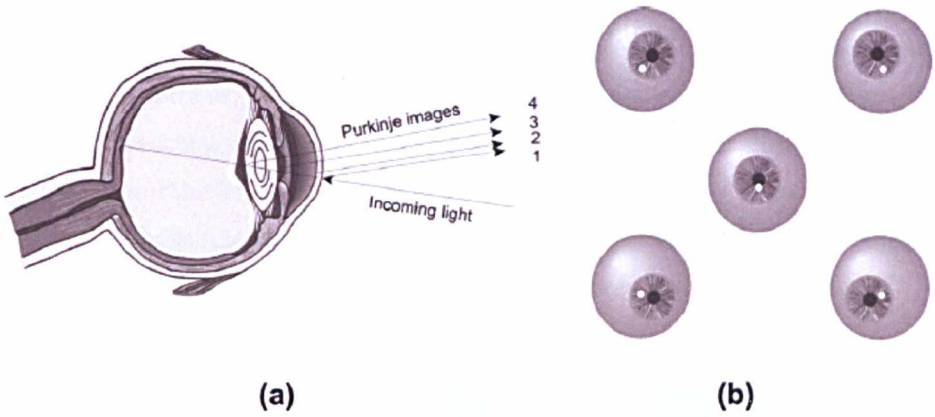

Figure 3.6 (a) Four Purkinje images formed by different layers of the eye. (b) Calibration of the eye. [reproduced from (65)]

Figure 3.6 illustrates the formation of four Purkinje reflections, by measuring the position of the first and the fourth reflections, it is possible to separate out the rotation and translation of the eye. It also illustrates the calibration process, where the relative positions of the pupil and the reflection changes with movement of the eyes.

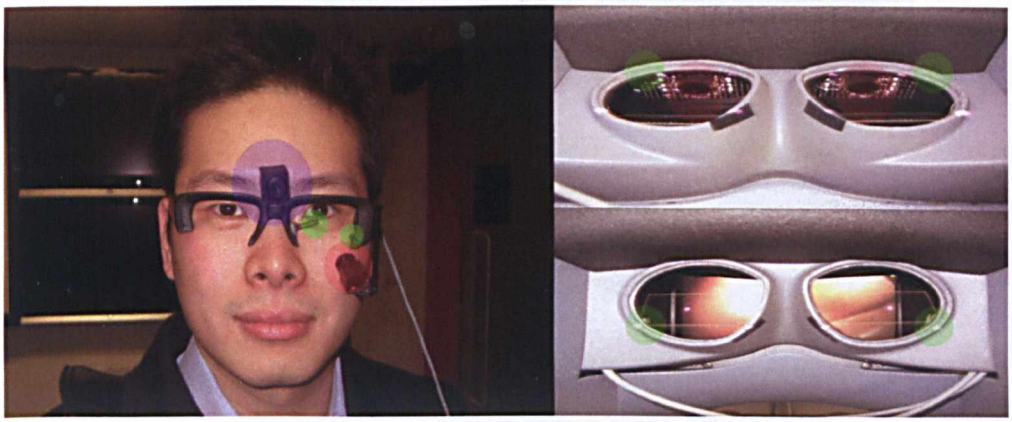

Figure 3.7 Left: head mounted eye tracker. Right: remote eye tracker built into a surgical robotic system. Blue shade – scene camera, green shade – infrared LED, red shade – infrared camera. [remote eye tracker picture courtesy of Mr George Mylonas]

Eye tracking using VOG can be done using a remote eye tracker or a head mounted eye tracker, both setups are shown in Figure 3.7. The remote VOG method has been used for experiments in this thesis, as it allows moderate head movement using a completely non-invasive technique, whilst acquiring reasonably accurate eye movement data. The remote system can also avoid distracting the surgeon's line of sight and introducing extra weight on their heads (121).

Specifically, the Tobii 1750 eye tracker (Tobii Technology, Stockholm, Sweden) was used throughout this thesis, as shown in Figure 3.8. It uses binocular video-oculography technique with an accuracy of 0.5 degrees and a sampling rate of 30Hz-50Hz, integrated with a 17 inch Thin-Film Transistor (TFT) display with a resolution of 1280×1024 pixels. It can tolerate moderate head movement within a 30×15×20 cm volume at 60 cm in front of the device, thus providing a relatively natural environment for experiments.

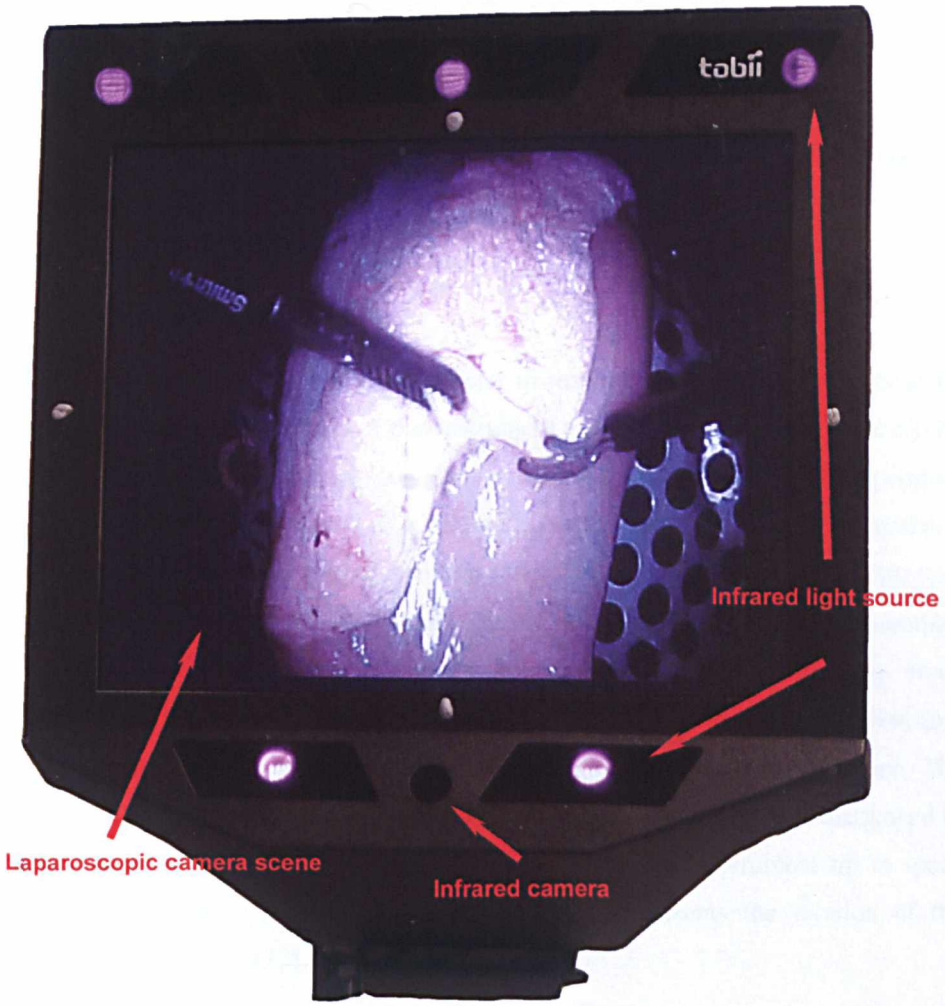

Figure 3.8 Tobii 1750 eye tracker (Tobii Technology, Stockholm, Sweden) with a laparoscopic experiment scene shown.

3.4 Instrument movement

In laparoscopic surgery, the effectors of the actions produced are the laparoscopic instruments which act as extensions of the hands, hence the tips of these instruments are of particular interest in this thesis. However, the actual instrument tips in the 3D world are not intimately linked with eye movements, as direct vision of the instruments is not possible. Therefore, the projection of the instrument tips on the laparoscopic monitor screen (2D) is needed for analysis with eye movements.

3.4.1 Instrument tracking

In general, there are three options to track the instruments: mechanical, electromagnetic and optical tracking.

3.4.1.1 Mechanical tracking

The attachment of a physical link to a surgical instrument allows its movements to be followed and the position of the tip of that instrument relative to a calibration point can be calculated. This principle is used for robotic assisted and virtual reality operating environments when the movements of a surgeon's hands need to be tracked to control a remote manipulator or interact with a virtual task. The Advanced Dundee Endoscopic Psychomotor Tester (ADEPT) is a dual 'gimbal' mechanism that accepts standard laparoscopic instruments each consists of four potentiometers to measure three translations and the rotation about the axis of the instruments. The laparoscopic instrument passes through an opening in the gimbal that acts as a pivot point. The restricted freedom of movement of the gimbal mechanism is similar to that experienced in real laparoscopic surgery. The exact position of a laparoscopic instrument tip in space could be recorded in x, y, and z co-ordinates, and θ represents the rotation of the instrument about its axis (122).

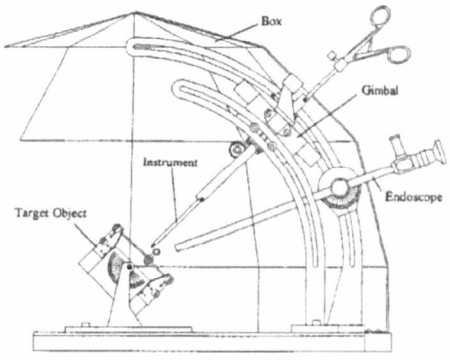

Figure 3.9 ADEPT [reproduced from (122)]

Mechanical trackers are immune from external disturbance such as electromagnetic fields or occlusion of line of sight. Mechanical trackers are very accurate, have high repeatability, and operate well within a small workspace. However, most existing systems are bulky and obtrusive, the mechanical parts produce noise and friction and complete freedom of movement is restricted.

Mechanical trackers were not used in this thesis due to the friction and size of the system. Similar to laparoscopic surgery, the point of incision becomes a pivot point for the instrument, however, there are still possibilities for translation movement of the pivot point in real laparoscopic surgery (see Chapter 2), unlike a gimbal system which create a less realistic simulated environment.

3.4.1.2 Electromagnetic tracker

Electromagnetic tracking devices function by measuring the strength of the magnetic fields generated by three small wire coils, oriented perpendicular to one another. These coils are embedded in small units, and are transformed into electromagnet as electric current flows through them. By sequentially activating each of the wires, and measuring the magnetic fields generated on each of the perpendicular wire coils, it is possible to determine the position and orientation of the sending unit.

These tracking units may experience interference operating in the vicinity of monitor screens or other devices that produce magnetic fields, as well as metal objects such as the operating table that disrupt magnetic fields (123). Another disadvantage to these tracking devices is that the working volume tends to be rather small.

The Fastrak system from Polhemus (Vermont, USA) is the electromagnetic sensor technology used in the Imperial College Surgical Assessment Device (ICSAD), where the coils are attached to the surgeons' wrists. This system provides extremely low latency (in the order of 5 milliseconds) and has the ability to track multiple objects concurrently.

Electromagnetic trackers would be ideal for this thesis, as this eliminates the line of sight problem in optical trackers. The main drawback is the fact that these sensors have to be mounted on metal laparoscopic instruments, which will distort the accuracy of the system.

3.4.1.3 Optical Tracker

Optical trackers are most commonly used in virtual environments and computer guided surgery at present. They are generally divided into four types:

- Active Tracking use InfraRed Light Emitting Diodes (IR-LEDs) as targets observed by infrared cameras. IR-LEDs are pulsed sequentially and detected by three linear sensors located in the tracker.

- Active Cameras emit infrared light from a ring surrounding each infrared camera lens and use balls or discs as targets. The balls or discs are coated with a retro-reflective material containing small spheres that mirror the light back to the lenses, causing the targets to appear as bright spots in the infrared images.

- Fully Passive Cameras use available light in the visible spectrum to detect and pinpoint high-contrast targets printed or painted on any suitable surface. Standard video lenses and sensors are used to capture the images, which are then fed into the host computer or an extra computing unit for processing. Computer vision software detects the presence of target patterns.

- Markerless Passive Cameras do not require wearing any kind of target or markers. Computer vision is used to extract the object's motions and positions. Delivered measurements are in this case only raw approximations.

Optical trackers in general have high update rates, and sufficiently short lags. However, they suffer from the line of sight problem, in that any obstacle between sensor and source seriously degrades the tracker's performance. Ambient light and infrared radiation also

adversely affect optical tracker performance. As a result, the environment must be carefully designed to eliminate these causes of uncertainty.

3.4.1.4 Equipment used in this thesis

NDI Polaris

The NDI Polaris System (Northern Digital Inc, Ontario, Canada) was used for earlier experiments. Active cameras with passive markers were used, these markers are retro-reflective balls mounted on planar rigid body, as illustrated in Figure 3.10. As translation and rotation information are needed to calculate the offset to the instrument tip, at least 3 passive markers are used for each instrument tracked. The maximum update of the system was 60 Hz, which is comparable to the eye tracking system, and the accuracy was quoted as 0.33mm Root Mean Square (RMS) (124).

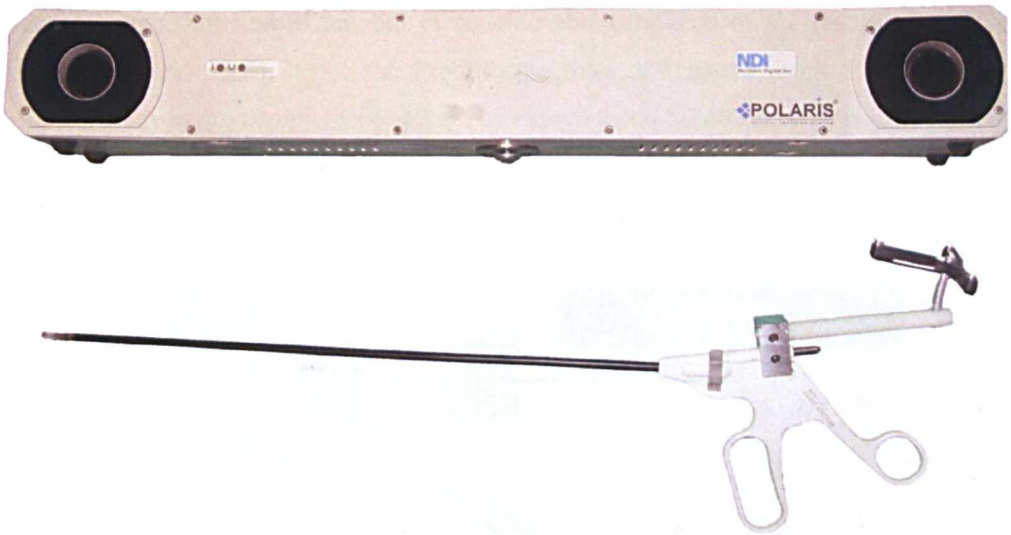

Figure 3.10 The NDI Polaris system with passive markers rigidly attached at the handles of the instruments.

In order to track the instrument tips, which are always hidden inside the cavity where the operation (or simulated operation) is carried out. The tracking markers have to be mounted near the handles of the instruments, without being obtrusive to the surgeons' hands. The offset from the markers to the instrument tips are then calculated using the pivot function in the NDI software.

The accuracy of the instrument tip tracking arrangement was further assessed by mounting the laparoscopic instrument on a Stäubli RX60L robotic arm with repeatability accuracy of +/- 0.02mm and six degrees of freedom. Ten points for calibration were set up using the robotic arm to manipulate the instrument whilst the IRED markers were being tracked.

Earlier experiments (Chapters 4 and 5) were designed around the technical specifications of the NDI Polaris, the measuring volume is in a shape of a pyramid, and at the base of the pyramid the widest dimension is 1.5×1.2 meters. This is just enough volume for most laparoscopic instrument handle movements, but at extreme movements, the markers lost the line of sight from the infrared cameras. The laparoscopic camera was held rigidly throughout each experimental episode, as movement of the camera will change its extrinsic properties and affect 3D-2D transformation (see 3.4.1.5 – Camera Calibration), which meant that the camera had to be re-calibrated after each repositioning. The camera sensor was positioned in the ceiling, where most of the movements of the two instruments were captured, but limited amount of rotation about the axis of the instruments were allowed (around 120°), hence the experiments were designed to limit the amount of rotation necessary to complete the tasks. The experimental hardware setup is illustrated in Figure 3.11.

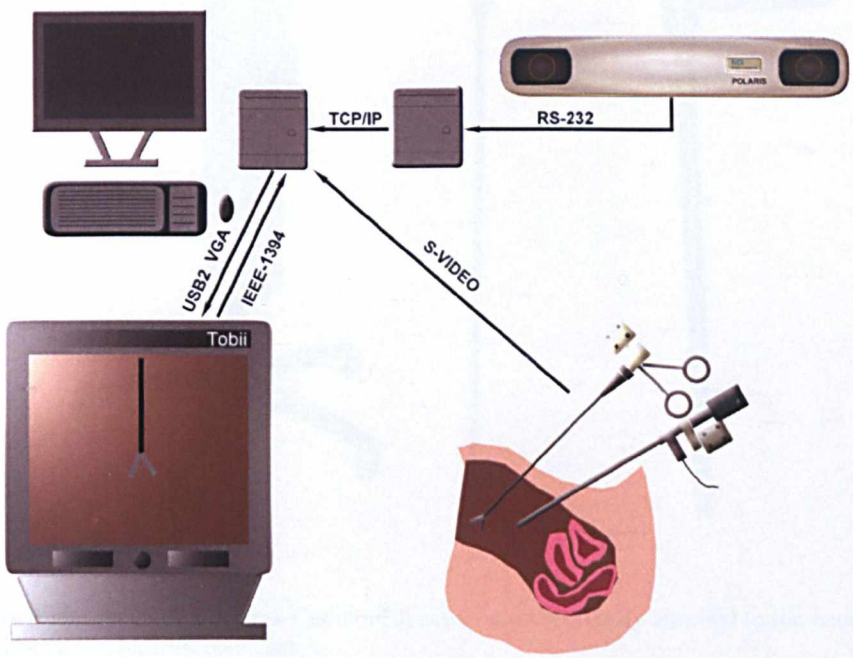

Figure 3.11 Polaris tracking system setup for the experiments conducted in this thesis.

NDI Optotrak Certus

The NDI Optotrak Certus (Northern Digital Inc, Ontario, Canada) is an active marker system demonstrated in Figure 3.12. Passive markers were not used when both instruments and the laparoscopic camera were tracked (see 3.4.1.6 – Robotic Hand/eye calibration), this is due to the higher chance of passive markers overlapping each other. As the passive markers are not unique in their individual properties, the camera system recognises each rigid body by its unique geometric relationships, the more number of objects tracked are likely to confuse the system. Active markers eliminate this problem, as the IR-LEDs can be identified individually. The accuracy of the system is 0.15mm RMS and the volume of tracking extends to 3.6×2.6 meters at the base of the pyramid. The maximum update rate is 4600 Hz, depending on the number of markers used (125). Bespoke rigid bodies were designed to house 4 markers each, shown in Figure 3.12, to increase the amount of rotation allowed for the instruments.

Figure 3.12 The NDI Optotrak Certus with active markers rigidly attached to the handles of the instruments and laparoscopic camera.

The other advantage of the Optotrak Certus system is the ability to use two camera sensors in conjunction, this both increased the volume of tracking and the rotation allowed for each instrument (around 300°). This system was used for Chapters 6 – 8, where more complex laparoscopic tasks were designed in the experiments. Both instruments and the camera are tracked at all times, which allow for free laparoscopic camera movement throughout the experiment after robotic hand-eye coordination.

3.4.1.5 Camera calibration

Physical camera parameters are commonly divided into extrinsic and intrinsic parameters. **Extrinsic parameters** are needed to transform object coordinates to a camera-centred coordinate frame. The pinhole camera model is based on the principle of collinearity, where each point in the object space is projected by a straight line through the projection centre into the image plane. The origin of the camera coordinate system is in the projection centre at the location (X_0, Y_0, Z_0) with respect to the object coordinate system, and the z-axis is perpendicular to the image plane. The rotation is represented using Euler angles ω, φ and κ that define a sequence of three elementary rotations around x, y, z axis respectively. In order to express an arbitrary object point P at location (X_i, Y_i, Z_i) in image coordinates, it needs to be transformed to the camera coordinates (x_i, y_i, z_i). This transformation consists of a translation and a rotation, and it can be performed by using the following matrix equation (126,127):

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$$

where

$$\begin{aligned} m_{11} &= \cos\varphi \cos\kappa \\ m_{12} &= \sin\omega \sin\varphi \cos\kappa - \cos\omega \sin\kappa \\ m_{13} &= \cos\omega \sin\varphi \cos\kappa + \sin\omega \sin\kappa \\ m_{21} &= \cos\varphi \sin\kappa \\ m_{22} &= \sin\omega \sin\varphi \sin\kappa + \cos\omega \cos\kappa \\ m_{23} &= \cos\omega \sin\varphi \sin\kappa - \sin\omega \cos\kappa \\ m_{31} &= -\sin\varphi \\ m_{32} &= \sin\omega \cos\varphi \\ m_{33} &= \cos\omega \cos\varphi \end{aligned}$$

The **intrinsic camera parameters** usually include the effective focal length **fc**, the image centre also called the principal point **cc**, skew coefficient **alpha_c**, and distortions (radial and tangential distortion) **kc**.

Once the camera is calibrated, the instrument tip positions can be transformed into the camera reference frame, this then eliminates the z-axis (the axis projecting directly in front of the camera lens) making a **3D to 2D transformation**. The intrinsic parameters can then be applied, where the distortions intrinsic to the camera are taken into account.

The camera calibration toolbox for Matlab (128) was used, as the algorithm is accurate and simple to perform. A calibration grid is printed and mounted on a "reasonable" planar surface, and the camera is required to observe this pattern from a few different orientations (10 in this thesis), as shown in Figure 3.13. This setup is also used for the Robotic Hand/Eye calibration.

3.4.1.6 Robotic Hand/Eye calibration

The setup described before could only be used in experiments where the laparoscopic camera is rigidly held throughout the whole experiment, and also between experiments, or a laborious recalibration of the extrinsic parameters needed to be performed. An improvement in the design would be to calculate the extrinsic parameters of the camera by tracking its 3D position and rotations. This was done by attaching infrared trackers on the laparoscopic camera, as shown in Figure 3.13.

In order to calculate the offset between the infrared markers and the camera reference frame, the robotic hand/eye calibration was performed. This is the task of computing the relative 3D position and orientation between the camera and the robot gripper (in this case, the infrared markers) in an eye on hand configuration, meaning that the camera is rigidly connected to the robot gripper.

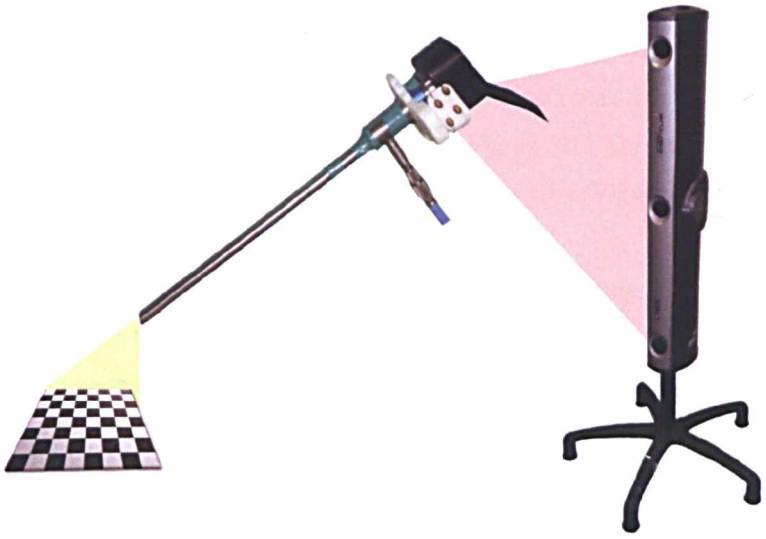

Figure 3.13 Setup for performing the Robotic Hand/eye calibration

The algorithm outlined by Tsai and Lenz was used (129,130), where the infrared marker coordinate system is provided by the tracking device, the endoscopic camera coordinate system is calculated from the camera extrinsic calibration as outlined above, and the tracker (NDI Optotrak Certus) world coordinate system is also known. With this setup shown in Figure 3.13, the calibration grid remains stationary, and the endoscopic camera/infrared marker complex was moved to at least three different positions (10 positions were used in the experiment), in order to estimate the unknown transformation from the tracker coordinate system to the calibration pattern coordinate system, as well as the transformation from the camera to the marker coordinate system. The output of the algorithm is the rotation and translation (homogenous transformation) between the marker and the endoscopic camera.

3.5 Data Synchronisation

There are three data streams simultaneously recorded for each experiment, the video data from the laparoscopic camera, the instrument data from the NDI tracking hardware, and the eye movement data from Tobii.

The ClearView software was provided by Tobii for recording and simple analysis of the data from the Tobii 1750 eyetracker. Analogue video data from the laparoscopic camera was sent via S-Video data cable to the video capture card (ATI 9600 AIW card or the 3D Connect Radeon X800 GTO graphics card), where the data was first digitalised and then displayed on the Tobii Eyetracker, as illustrated in Figure 3.11 This digitalised video data was recorded in synchrony with the eye tracking data by ClearView.

In order to synchronise between the eye and instrument tracking data streams, bespoke software was written to achieve this. In earlier experiments, a two-computer configuration was used, to ensure low processing demand. One computer was used to run the Tobii Eyetracker and the video capture, and another computer was used to run the NDI Polaris hardware for instrument tracking. Every 300ms, the NDI computer sent a Polaris data frame to the Eye tracking computer, which would be stamped on the eye tracking data file. This interface was achieved using the Transmission Control Protocol and the Internet Protocol (TCP/IP) standard, with a near constant delay of 83ms.

For later experiments, only one computer was used, due to the increased speed of the processor. The NDI Optotrak data was captured by the NDI First Principles software, and the eye tracking data by the Tobii ClearView software. As the Optotrak started or stopped recording data, a signal was sent to an 18F452 microcontroller based circuit board, which in turn provided a timestamp on the eye tracking data file, via RS-232.

3.6 Sample data and pre-processing

Figure 3.14 illustrates a sample data collected in Chapter 5. The raw output data from the instrument tip tracking system is shown in the top subplots. Here, 3D data demonstrates the experimental task, as the subject locates two fixed targets arranged horizontally in a simulated MIS environment. The 3D y-axis data is most representative here, as the left instrument starts the task, followed by the right. The x- and z- axes data is a combination of the instrument vertical and depth movements respectively, which are important features in assessing the quality of the movement trajectories and the skills of the operator. Furthermore, eye movement data is collected in 2D, as located on the laparoscopic monitor, hence its incorporation is not possible.

After 3D to 2D transformation, the data is shown in middle subplots. As the targets are arranged in a horizontal fashion, the x-axis data is most representative of the surgical

workflow. Eye movement data is plotted in pink, and is tightly coupled with the active instrument (first left, then right). Vertical movements are illustrated in the y-axis subplot, again these represent deviations from the ideal pathway between the two targets, and are important features for qualitative assessment.

Further dimension reduction is done by converting the (x, y) coordinate into Cartesian distance from the target, and is shown in the bottom subplot. This is otherwise known as the Target Distance Function (TDF), converts the data into single dimension time series signals, whilst retaining most of the features of the movement trajectories. The other advantage of the TDF is its invariance in direction, hence data points approaching 0 always represents target reaching behaviour. It is important to note that the instrument and eye movements decelerate towards 0 (target), and high density eye movement data represents heavy visual burden at the end of each target reaching trajectory.

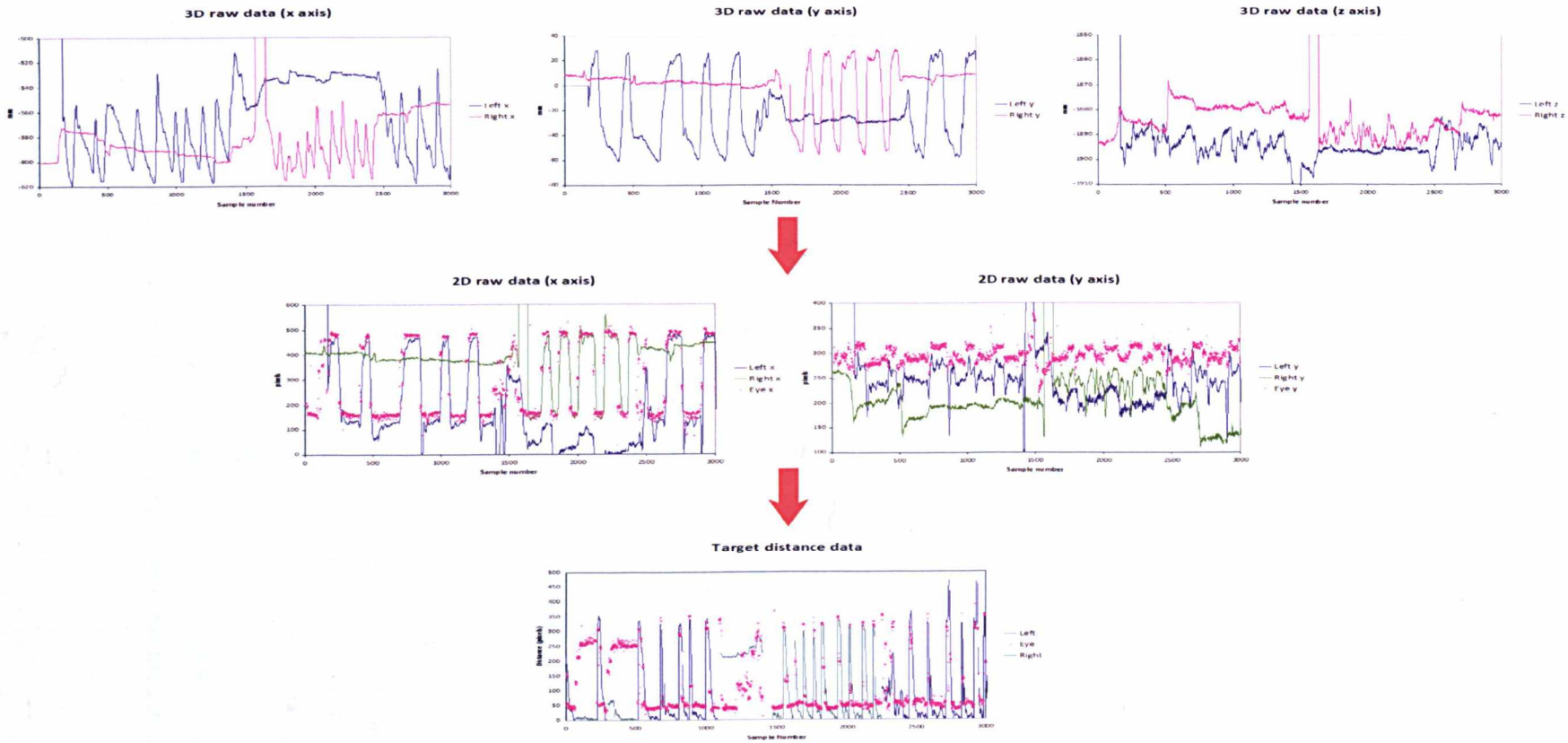

Figure 3.14 An example data stream demonstrating dimension reduction from 3D raw data (top) to 2D screen projection (middle), and to distance data (bottom). Eye movement data is not available in 3D. Red arrows show the direction of the dimension reduction.

3.7 Conclusions

Unlike open surgery, the environment in MIS produces many constraints to the operator, the restricted information provided by the laparoscopic camera results in much higher demands of the visual system to compensate. Hence, eye movement studies provide the implication of underlying cognitive load that underpins the generation of skilled surgical hand movements. Studies of hand or instrument movement and eye movement in surgery have been done previously, but as individual components rather than concurrently (23,48,49,51,114).

This chapter has highlighted the technical challenges in synchronous tracking of instrument and eye movements, the final setup chosen was based primarily on accuracy, availability and suitability for the individual experiments. The use of optical based equipment allowed relatively unobtrusive tracking of the eye and the instrument tips, and produced less disruption on the natural visuomotor behaviour of the subjects. The coordination of the different commercially available hardware setups remained a significant challenge, bespoke additional hardware and software implementations were necessary to provide accurate synchronisation between them.

The application of the robotic hand/eye coordination algorithm has allowed accurate positioning of the 2D instrument tip projection on the laparoscopic screen, by using the information of the camera calibrations and position, and the instrument tip position. This mirrors the transformation that the human brain calculates during laparoscopic surgery. The algorithm, as its name implies, was originally developed for robotic vision and guiding robotic object manipulation.

Overall, the experimental setup has been satisfactory, particularly the high accuracies and update frequency of both systems. It has been suggested that the temporal difference between the eye and the effector is up to 300ms (72), and the equipment in this thesis used has a sampling rate of at least 50Hz (each data frame being 20ms apart), so this should be enough to detect this relationship. However, higher update frequencies might improve the performance of the signal matching algorithm used in later chapters, as will be discussed further.

For the current system setup, the use of wired tracking IREDs has its limitations. Wired IREDs precludes their use in the sterile operating environment, although there are now commercially available wireless and sterile reusable active optical trackers; however, the NDI system allowed more flexible applications due to its available Application Programming Interfaces (API), and has been ideal in the laboratory settings. Furthermore, the combined cost of the Tobii eyetracker and the NDI Optotrak system was high, this is excluding the complete laparoscopic system and disposable materials. The use of wearable eye tracker should be explored in future work, as this has the advantage of measuring important eye movement data outside the laparoscopic screen, which may include occasional hand position referencing, communication with the surgical assistants, and disturbances of the operating theatre. The main disadvantage remains the relative bulk of the equipment worn on the surgeon's head, which may interfere with the natural head and body movement during data recording.

Current work is being done on minimising the calibration steps prior to data collection. Using the current setup, a brief nine point calibration of the eye position is performed at the beginning of the experiment, and instrument tracking calibrations can be performed offline. This is acceptable in the laboratory setting, however, these extra steps may disrupt the flow of live operations. Computer vision based techniques for instrument tracking are also being refined, where automatic segmentation of the instrument tips can be obtained directly from the video data stream. This will certainly be a welcomed addition to the current setup, as hardware based instrument tracking can be eliminated.

In summary, this chapter has provided the basic software and hardware setup for this thesis. The next chapter will present an early experiment where only instrument tracking data is used, and the quality of these movement trajectories are calculated using computer learning algorithms. The experimental procedure is designed around the limitations of the earlier tracking system, as the tracking volume of the NDI Polaris system is small, hence only relatively restricted surgical manoeuvres are allowed to ensure the quality of the data.

Chapter 4

Assessment of the Quality of Movement Trajectory in Laparoscopic Surgery in different screen rotations

4.1 Introduction

In Chapter 2, it has been shown that movement trajectories are affected by the reference coordinates used for motion. Furthermore, the quality of visual information can also change movement patterns. In the experiments where visual information was restricted or distorted, errors were introduced in the trajectories made. However, it seems that this sensorimotor map is developed from birth and is continuously subject to adaptation.

Minimally Invasive Surgery (MIS) was introduced over 20 years ago (17,131), however, many surgeons are still restricted to performing relatively simple procedures, for example laparoscopic cholecystectomy and diagnostic arthroscopy. The uptake of advanced MIS procedures, such as laparoscopic colectomy for cancer (25) and arthroscopic soft tissue repair, is still very slow in many countries. This may have been resulted from the difference in individual's adaptability to the spatial transformation required in MIS.

The constraints imposed by the MIS environment have been discussed in Chapter 2, which include a lack of 3D vision (132), limited haptic feedback (133) and the "fulcrum" effect (134) restricting the variety of surgery performed. It is, however, quite clear that some surgeons are superior in performing these tasks than others (135). This has motivated extensive research into the objective assessment of surgical skills (42). For this purpose, the methodology has now evolved from subjective qualitative assessment by the trainers and knowledge assessment using post-graduate examinations (44), to objective quantitative approaches using time or movement parameters (49). Quantitative methods

for assessing surgical dexterity have been widely validated for a number of open and laparoscopic procedures (136). The validation for these methods, however, relies upon prior definition of expertise, and this classification is mostly based on the assumption that experience equals technical excellence.

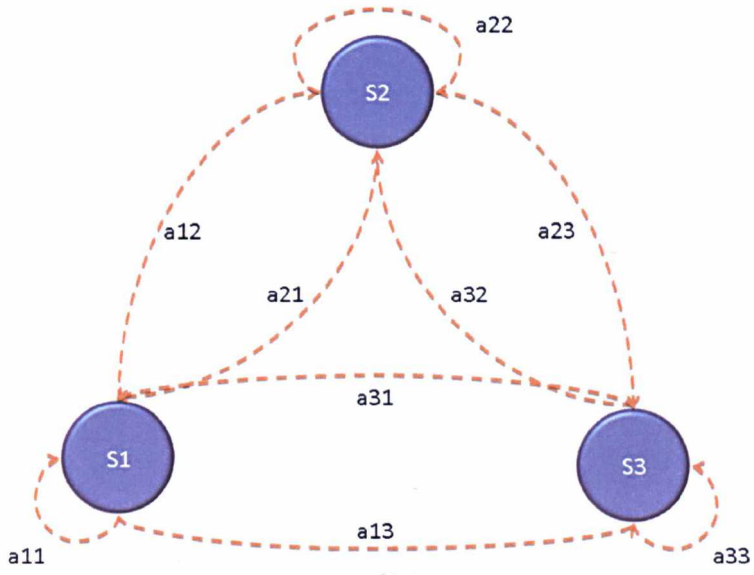

Figure 4.1 Diagram illustrating the Markovian Process. S1, S2 and S3 are the states; a11, a12, a13 ... a33 are the transitional probabilities.

The mathematical tool to be used in this chapter will be based on Hidden Markov Model (HMM), which is a statistical model that has been widely used for pattern recognition, *e.g.*, speech, gesture and hand-writing (137-139). It models a system that is assumed to be a Markovian Process, as illustrated in Figure 4.1, which is a stochastic process where the likelihood of a given future state (S1, S2, and S3) depends only on its present state, and not on any past states. The chances of transition between each state are governed by the transitional probabilities (a11, a12, a13 ... a33). A typical example of a Markovian process is the weather pattern and each state representing the weather conditions (sunny, rainy and cloudy) with the transitional probabilities between them.

In HMM, as illustrated in Figure 4.2, the state is hidden from the observer, but each state generates a number of possible observations (O1, O2, and O3), with their respective output probabilities (b11, b12, b13 ... b33). Using the previous example, the observations can be seen as the dampness of the grass, for example, dry, dryish, damp, and soggy. There are three main usages of HMMs:

1. Given the output sequence of observations, find the most likely transitional and output probabilities; this is generally used for training the HMM.
2. Given the parameters of the model, find the most likely sequence of hidden states that generated the output sequence. In speech recognition, once the HMM is trained, the spoken sounds (observation) is then used to deduce the words they represent (hidden state).
3. Given the parameters of the model, calculate the probability of a particular output sequence is then generated by the HMM. This is used in this chapter as described further.

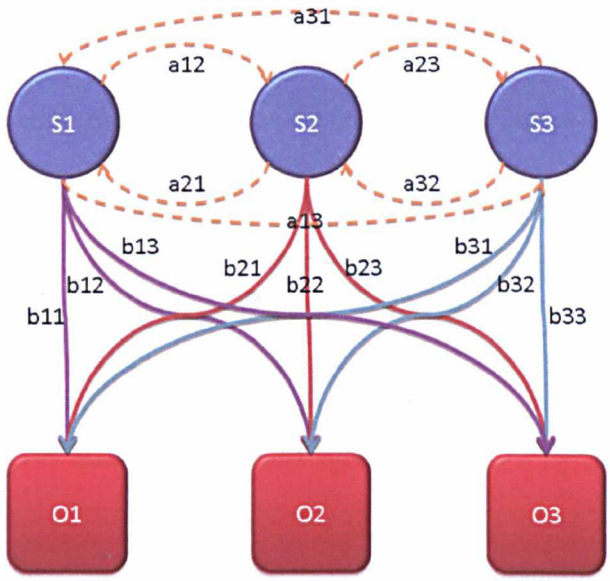

Figure 4.2 Diagram illustrating the Hidden Markov Model. S1, S2, S3 are the hidden states; a11, a12, a13 ... a33 are the transition probabilities; b11, b12, b13 ... b33 are the output probabilities; and O1, O2, O3 are the possible observations.

Rosen *et al* applied HMMs using a series of 14 finite states defined by the surgeons' instrument tissue interactions (140). These states were based on the Forces and Torques (F/T) signatures collected by using two sensors measuring the forces and torques applied at the interface between the surgeon's hand and the instrument. The HMMs classified the actions into specific manoeuvres and the transitions between them. The skill level of each surgeon was then calculated by the statistical distance of their HMMs from those of the expert surgeons. The method, however, relied on a previous definition of expertise and

the scoring system was based on the individual difference from this definition. To obtain the F/T signatures, each video frame was also visually analysed by two experts.

The purpose of this chapter is to examine the use of HMMs based on view invariant trajectory representations for assessing complex MIS tasks, as the effects of the many coordinate transformations in MIS are difficult to quantify by simple means. Identical surgical motions performed by different subjects are used as input into the HMM, which is treated as a "black box", this is regarded as the initial **training** phase. The trained HMM will then "recognise" the most represented surgical motions (or the "ideal" trajectory), as defined by the initial training datasets. Finally, the output of the HMM is the similarity of the testing surgical trajectory from the "ideal" trajectory. Obviously, the "ideal" trajectory is influenced by the initial training data, and this is discussed further below.

In order to increase further complex spatial transformations affecting hand-eye coordination, view rotation tasks have been introduced and the motion trajectories are measured. This effect has been studied previously (27,28,134), which indicates a detrimental effect on the performance of surgeons and novices. Furthermore, the ability to handle mental rotation tasks has been suggested to be indicative of the innate ability in mastering laparoscopy (141). In this chapter, the proposed method is demonstrated where a probabilistic framework can be formulated to allow the observation of trajectories without prior, arbitrary classification of the subjects' abilities.

4.2 Methods

4.2.1 Modelling Instrument motion trajectory

4.2.1.1 Instrument tip tracking and calibration

In order to obtain the trajectory of the instrument tips in Euclidean space and not interfere with the experimental task, a tracking device was attached rigidly to the handles of the laparoscopic instruments. For accurate positioning of the tracking device, a Polaris (Northern Digital Inc, Ontario) 6 degree-of-freedom InfraRED (IRED) tracker was used. The Polaris tracker is able to track a number of passive, active, wired and wireless IRED tools in real time simultaneously. Data interfacing was achieved through RS-232 and the provided tracking accuracy is 0.33mm RMS at a sampling rate of 60Hz. The offset of the

instrument tips from the IRED markers was calculated using the Pivot function in the NDI ToolViewer Software version 3.02.01.

The accuracy of the instrument tip tracking arrangement was further assessed by mounting the laparoscopic instrument on a Stäubli RX60L robotic arm with repeatability accuracy of +/- 0.02mm and six degrees of freedom. Twelve points for calibration were set up using the robotic arm to manipulate the instrument whilst the IRED markers were being tracked, an average of 550 measurements were taken for each point divided in 3 separate sessions to demonstrate the accuracy and repeatability of the tracking system. The summary of the results are shown in Table 4.1, where the Root Mean Squared Error (RMSE) was around 0.5mm.

Point	Number of Measurements	RMSE (mm)
1	653	0.45
2	572	0.48
3	582	0.50
4	513	0.46
5	533	0.45
6	575	0.45
7	577	0.47
8	580	0.55
9	573	0.53
10	570	0.52
11	580	0.49
12	312	0.44

Table 4.1 Summary of tracking error using 12 robotic calibrated reference points.

4.2.1.2 Subjects and experimental setup

Eleven subjects were recruited for the study (9 medical student and 2 practicing surgeons, 2 subjects were left-handed). None of the medical students had previous simulated or real laparoscopic experience. They were then randomised into two groups. All subjects were consented prior to the study, and were given a short introduction of laparoscopic surgery and the instruments used. Both groups were required to perform a laparoscopic task in a box trainer, after familiarisation with the instruments and environment. A Karl Storz laparoscopic stack with an S1 camera head, Xenon Nova light source, and Hopkins II 0° endoscope was used for the experiment, along with two laparoscopic graspers with IRED tracking devices rigidly attached.

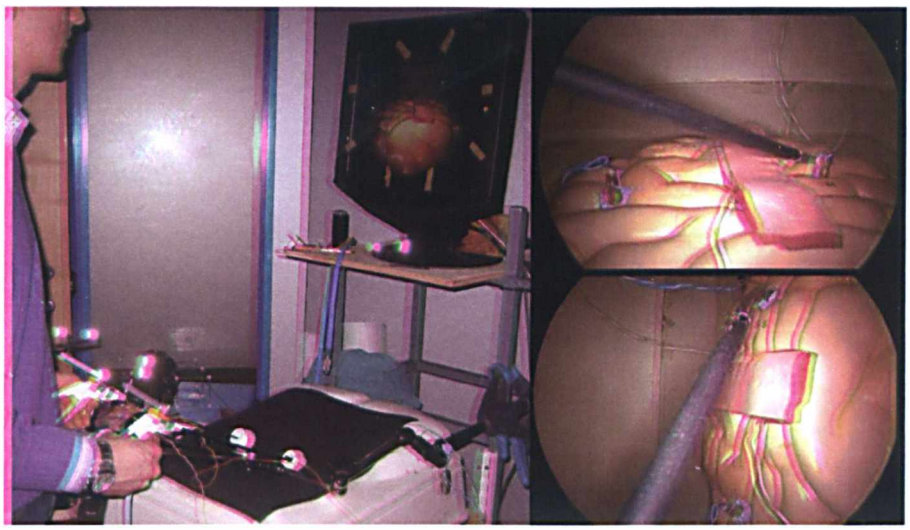

Figure 4.3 Experimental setup showing the arrangement of the IRED markers in relationship to the laparoscopic tools, and the plastic small bowel model with a simulated omental flap with and without the camera view rotated.

For both groups, the first task was to locate two standardised points (A and B) on a simulated plastic small bowel model, as illustrated in Figure 4.3, using laparoscopic graspers. Each point was attached to an in-house designed touch sensitive circuit switch to mark the beginning and end of each trajectory. Each time the circuit was completed, an alarm will indicate a successful contact. The subjects were then asked to locate alternatively the points A and B with the left instrument 10 times, and then the right instrument, and this step was repeated. A total of 36 trajectories were obtained between the two points. The subjects were then required to perform a complex task: first, lift a simulated omental flap using the left instrument to grasp the left corner, and touch a third point (C) hidden underneath the flap with the right. This was repeated 3 times, however the data for this task was not analysed due to its complexity. For the second task, subjects of Group 1 were required to repeat the first task with the laparoscopic camera rotated 90 degrees counter clockwise and Group 2 with the camera rotated 90 degrees clockwise.

Five subjects (3 from Group 1 and 2 from Group 2) underwent further training where a total of three sessions with the laparoscopic camera rotated 90 degrees counter clockwise and three clockwise were completed, with a final post-training assessment using normal camera orientation.

The left column in Figure 4.4 shows the typical instrument tip trajectories of a novice (pre-training, first rotation and post-training tasks) and a surgeon (normal and rotation

tasks). The actual captured data is a Euclidean representation of the 3D instrument tip position expressed as mm from the camera origin (0, 0, 0). To simplify the illustration, distance from the origin was calculated and whole data sets are plotted in Figures 4.5 and 4.6 using:

$$k = \sqrt{x^2 + y^2 + z^2}$$

where k is the distance, and x, y, and z are the coordinates collected.

For the analysis of the trajectories, the experiments were divided into: (1) left instrument motion from point A to B; (2) left instrument motion from point B to A; (3) right instrument motion from point A to B; (4) right instrument motion from point B to A.

4.2.1.3 View Invariant Representation of 3D trajectories

Prior to HMM analysis, the 3D trajectories were mapped to a view invariant representation based on the Centroid Distance Function (CDF) (142). The instrument tip positions from IRED tracking after offset correction were modelled as a parametric curve:

$$r[t] = \{x[t], y[t], z[t]\}$$
$$t = 0...N-1$$

CDF is a feature that is affine invariant, and is also widely used in image retrieval applications (142). The centroid is defined as the weighted average of all the points of each particular trajectory, and in essence, CDF describes a time series of the Cartesian distance of each point in the trajectory from the centroid. Scale normalisation transforms the CDF values into standard normal distribution which are rotational and translational invariant. Other affine invariant representations based on local angle and velocity measurements (143) were considered, however, these methods were sensitive to local variations of the trajectories and were not used for this study. Previous study compared CDF with curvature scale space, and showed that CDF with HMMs yielded better recognition of trajectories (142). Both methods, however, offered rotation invariant representations of trajectories. CDF significantly simplifies the subsequent HMM classification by foregoing cumbersome pre-processing steps. The right column of Figure 4.4 shows the CDF projection of the instrument trajectories for a surgeon and a novice.

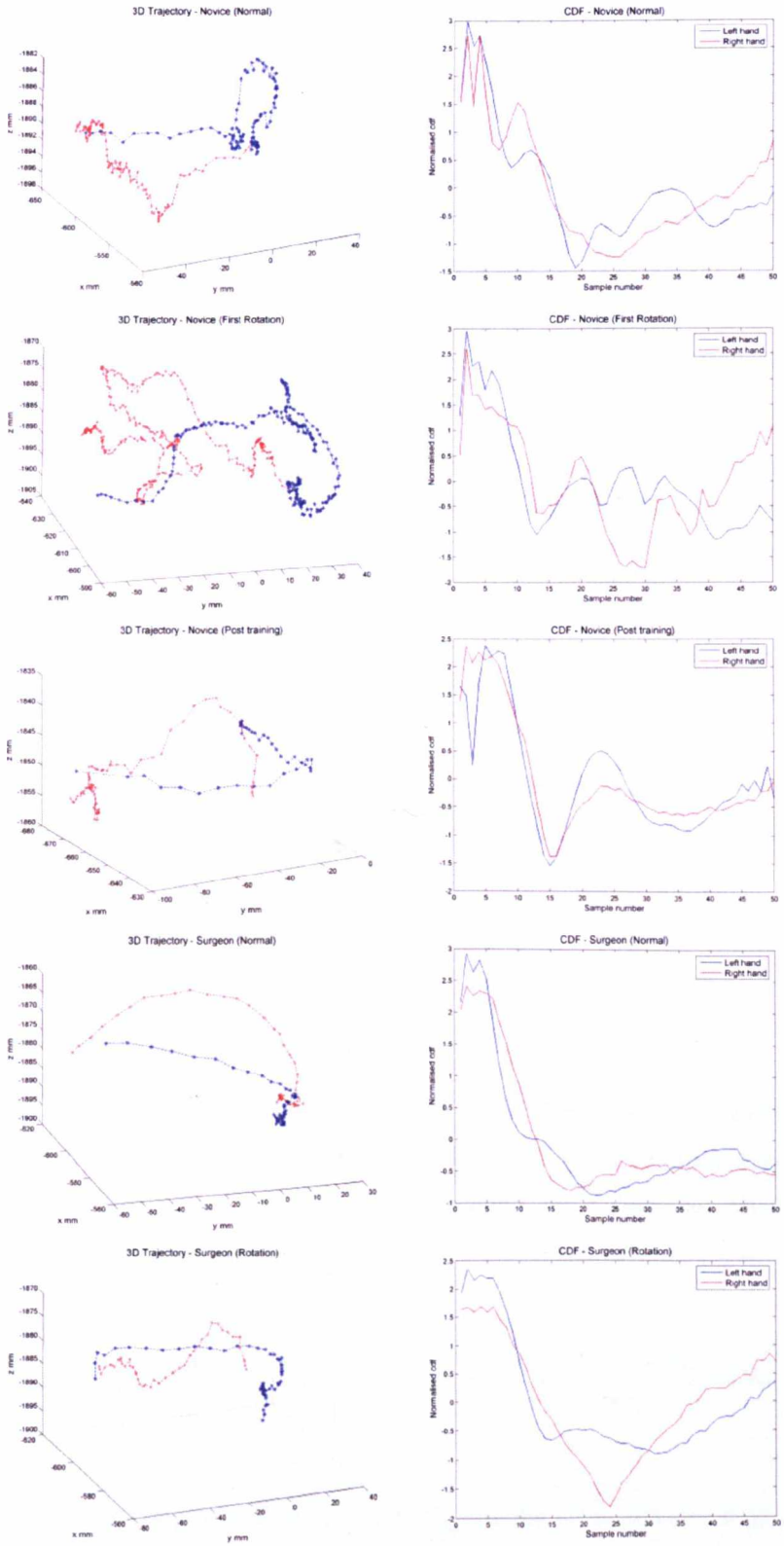

Figure 4.4 Left Column: 3D trajectories of the surgeon and novice, where blue shows the left hand and red shows the right. Right column: CDF representation of the same trajectories. In order from top to bottom: novice pre-training, novice first rotation, novice post-training, surgeon normal camera orientation, and surgeon rotated camera orientation tasks.

4.2.1.4 Hidden Markov Modelling

As mentioned earlier, HMMs are finite state stochastic machines that allow dynamic time warping for modelling time series data. HMMs have been used to classify movement trajectories, however, segmentation was necessary to avoid violating the Markovian property which assumes independence of a current state from past states given the previous one (144). Each trajectory was regarded as one independent action, and the CDF for the trajectory was used as input signals to the HMM, hence the notion was made that each trajectory adhere to the Markovian assumption. In this study, HMM was used to learn each trajectory of a given experiment and view rotation. The leave-one-out method was used to train the HMM from all subjects excluding the test subject's data. The trained HMM was then used to calculate the log likelihood of the test subject and indicate similarities or differences to the learned model. For example, for testing Subject 1, the HMM was trained using the trajectories from Subjects 2 – 11. This trained HMM was then used to calculate the log likelihood of Subject 1's dataset. Subsequently, Subject 2's dataset would be tested using the HMM trained by trajectories from Subjects 1, 3 – 11.

An HMM can be described by three model parameters representing the relationship between the hidden states (h) and the observed data (x). These parameters are:

$$\left[\pi_i, a_{ij}, p(x \mid h) \right]$$

where π_i is the initial state probability, a_{ij} the transition matrix between the hidden states and $p(x \mid h)$ the probability of generating an observation given the hidden state. In this study, a fixed number of states were used. However, in order to have a more flexible model, the observation probabilities were modelled by a Gaussian Mixture Model (GMM). Thus $p(x \mid h)$ can be defined as:

$$p(x \mid h) = \sum_{m=1}^{M} c_m \frac{1}{(2\pi)^{P/2} \left|\Sigma_m\right|^{1/2}} \exp\left(-\frac{(x-\mu_m)^T \Sigma_m^{-1} (x-\mu_m)}{2} \right)$$

where c_m is the mixing parameter, μ_m and Σ_m are the mean and covariance matrix of the component m of the GMM.

A K-means algorithm was used to initialise the parameters of the observation GMM (mainly the means μ_{mi}). Two versions of the K-means algorithm were implemented to compare performance, using Euclidean distance and Derivative Dynamic Time Warping (145) to calculate the similarity between two trajectories. The second option can take signal 'warping' into consideration when finding cluster centres, and the length of the 'warping' was then used to normalise the distance between the two trajectories. As the resulting cluster centres were identical with both methods, Euclidean distance was used as it was more computational efficient. The Expectation Maximisation (EM) algorithm (146) was used to calculate the maximum likelihood of the parameters of the HMM, namely the means and covariances of the components of the GMM and the parameters π_i and a_{ij}.

The parameters of the HMM including the number of Gaussians in the observation GMM, as well as the number of the hidden states of the HMM were selected experimentally. Parameters that lead to the least variation in the values of the test data likelihoods were selected as the parameters that can provide good data representation for this dataset. The number of hidden nodes was selected as 4 with a mixture of 2 Gaussians for the GMM.

4.2.2 Categorised observational score

The video of all the tasks were scored by two independent observers who were blinded to the identity of the subjects, the scoring system used was a modified version of the OSATS global rating scale. This is a widely validated score developed by Martin *et al*, using 8 categories each with a Likert scale of 1-5 anchored by descriptors (45). The modification was necessary as 3 of the categories did not apply to this particular experiment (suture handling, knowledge of procedure and quality of final product), see Table 4.2. Inter-observer reliability was calculated using Cronbach's alpha test (147).

Minor modifications of the original OSATS global rating scale have been employed for evaluating specific procedures, for example in knee arthroscopy (148), laparoscopic salpingectomy (149) and corneal suturing (150). Grantcharov et al shortened the OSATS score to only four of the original categories, and validated this video-based scoring system in laparoscopic cholecystectomy (52). This is similar to Table 4.2, except an added fifth category for evaluating overall performance in this Chapter. The performance of this abridged OSATS score and the original generic OSATS global rating scale have

been compared previously (151), which showed comparable inter-rater and intertest reliability.

GLOBAL RATING EVALUATION OF PERFORMANCE				
Respect for tissue				
1 Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments.	2	3 Careful handling of tissue but occasionally caused inadvertent damage.	4	5 Consistently handled tissues appropriately with minimal damage.
Time and Motion				
1 Make unnecessary moves.	2	3 Efficient time/motion but some unnecessary moves.	4	5 Clear economy of movement and maximum efficiency.
Instrument handling				
1 Frequently asked for the wrong instrument or used an inappropriate instrument.	2	3 Competent use of instruments although occasionally appeared stiff or awkward.	4	5 Fluid moves with instruments and no awkwardness.
Flow of operation				
1 Frequently stopped operating or needed to discuss the next move.	2	3 Demonstrated some forward planning and reasonable progression of procedure.	4	5 Obviously planned course of operation with efficiency from one move to another.
Overall				
1 Very poor	2	3 Competent	4	5 Clearly superior

Table 4.2 Modified version of the Objective Structured Assessment of Technical Skill (OSATS) global rating scale.

4.3 Results

Figures 4.5 and 4.6 illustrate the complexity of the raw data, although in reduced dimensions for simpler demonstration. The general shape of the data remains similar, where the instrument tips travelled between two points producing a pseudo-sinusoidal pattern. The most notable difference is the time scale: in the novice data, the time-to-complete the first rotation task more than doubled that of the pre-training task; for the surgeon, the task completion times were similar. Another difference is as the instrument tip approaches the targets, the peaks and troughs of the trajectories become less distinct in the rotated task, this effect is seen in both the surgeon and the novice. The initial part of each trajectory, from point A to point B (or *vice versa*), is mainly controlled by feed forward mechanisms, where previously stored proprioceptive information and target location were used as a rough guide to movement generation. As the instrument tip approaches the target, larger errors were detected due to the effect of the screen rotation, visual input was then used as feedback mechanism to reduce the error to zero, when the instrument tip arrived at the target.

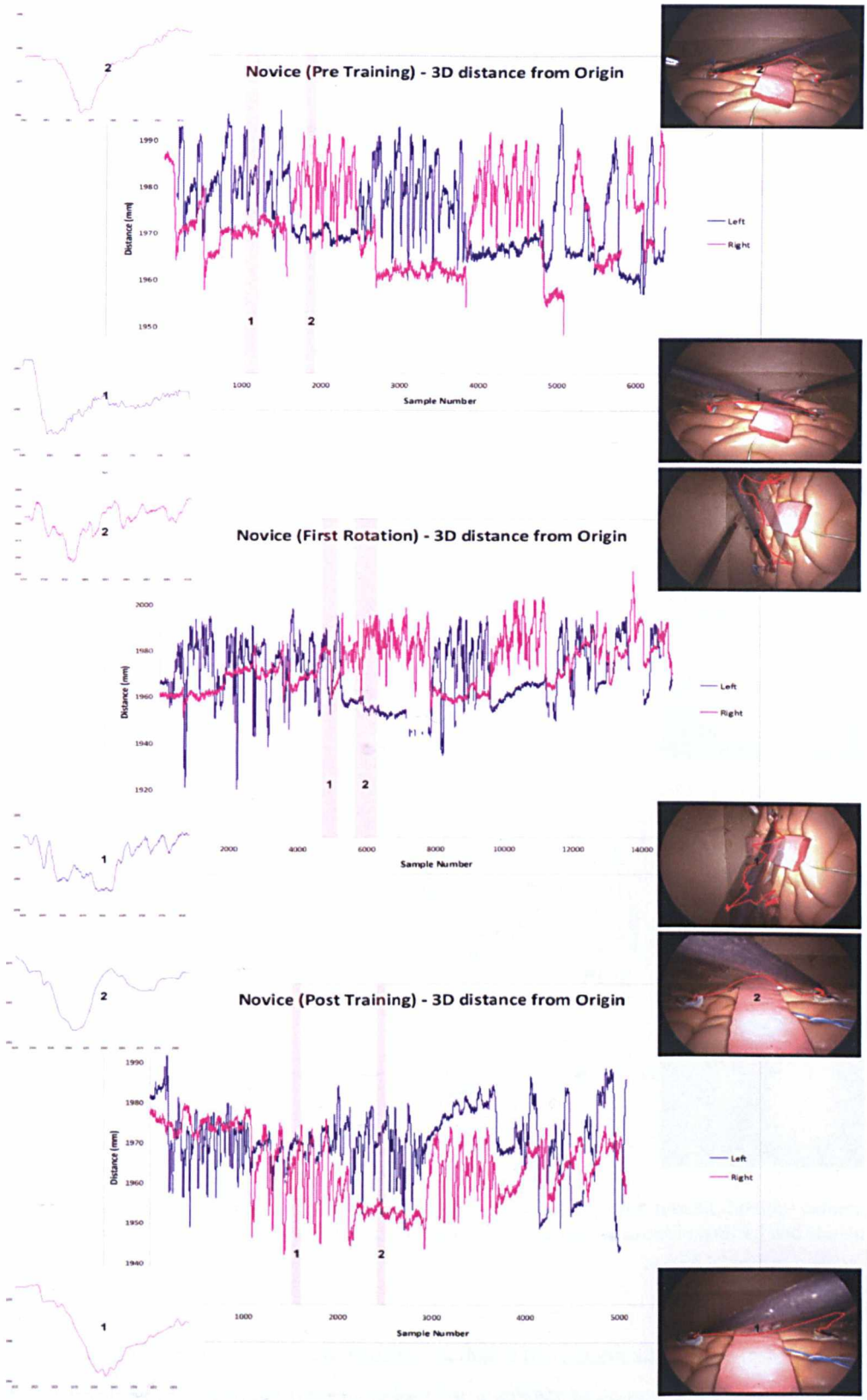

Figure 4.5 Motion trajectory of a novice performing pre-training (top), first rotation (middle) and post-training tasks (bottom). Illustrated in raw 3D distance data, with pink shaded areas magnified and shown in laparoscopic scene as numbered.

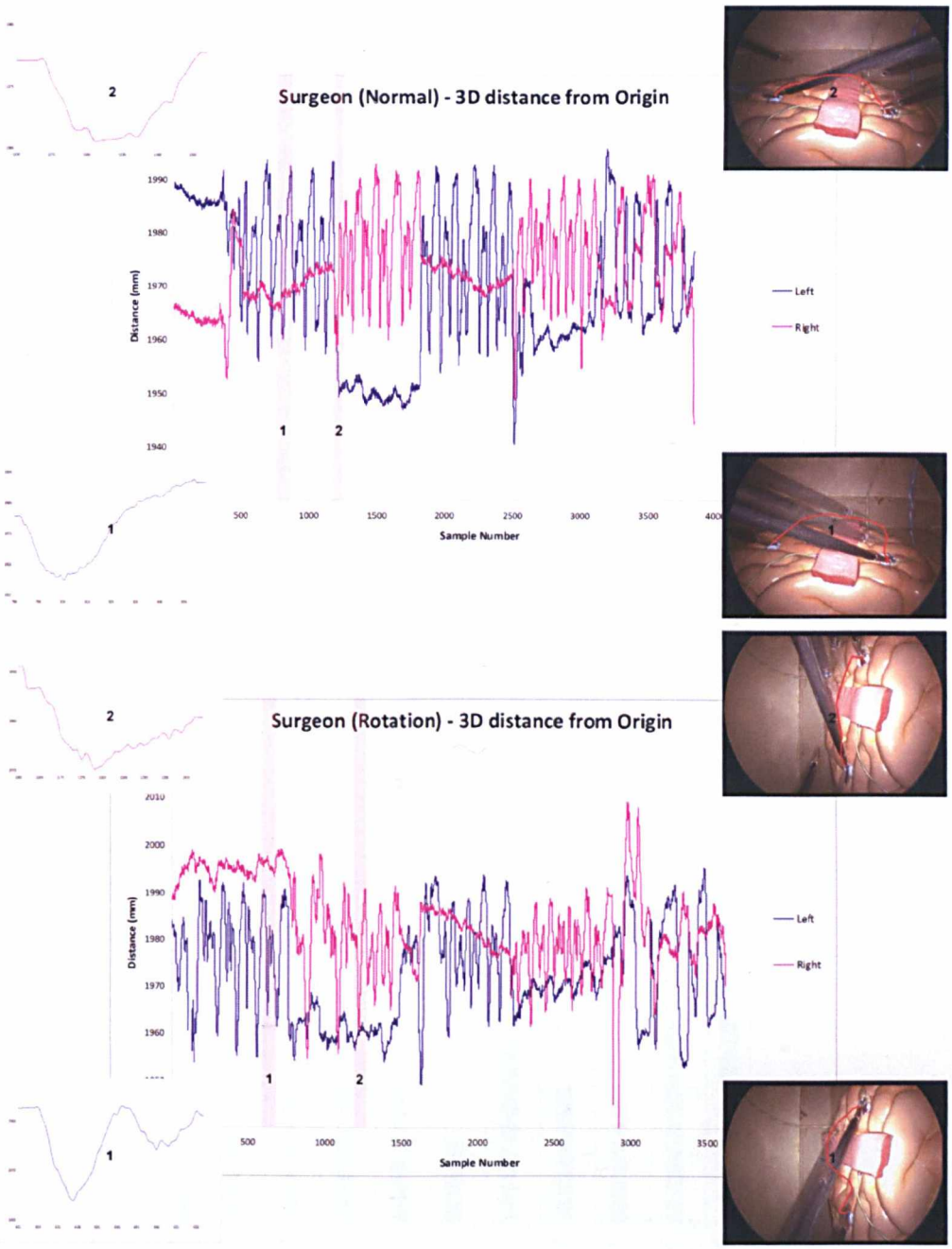

Figure 4.6 Motion trajectory of a surgeon performing normal (top) and rotated (bottom) camera orientation tasks. Illustrated in raw 3D distance data, with pink shaded areas magnified and shown in laparoscopic scene as numbered.

To demonstrate the extent of view rotation increases the complexity of the tasks, Figure 4.7(a) illustrates the average time increased for a subject to complete a trajectory when the view is rotated. Figure 4.7(b) demonstrates the OSATS global rating scale (with strong inter-observer agreement, α = 0.914) decreases with view rotation, however this

does not completely correlate with the time measurement (r = -0.818, p = 0.002). For example, Subject 2 in the rotated task had the third lowest score in OSATS, but had the fifth longest average time of trajectory. There is a significant correlation between mean (unfiltered) path length and time taken (r = 0.916, $p < 0.001$), and in the rotated task, between path length and OSATS score (r = -0.873, $p < 0.001$). Interestingly, one subject improved slightly in the rotated task (subject 7).

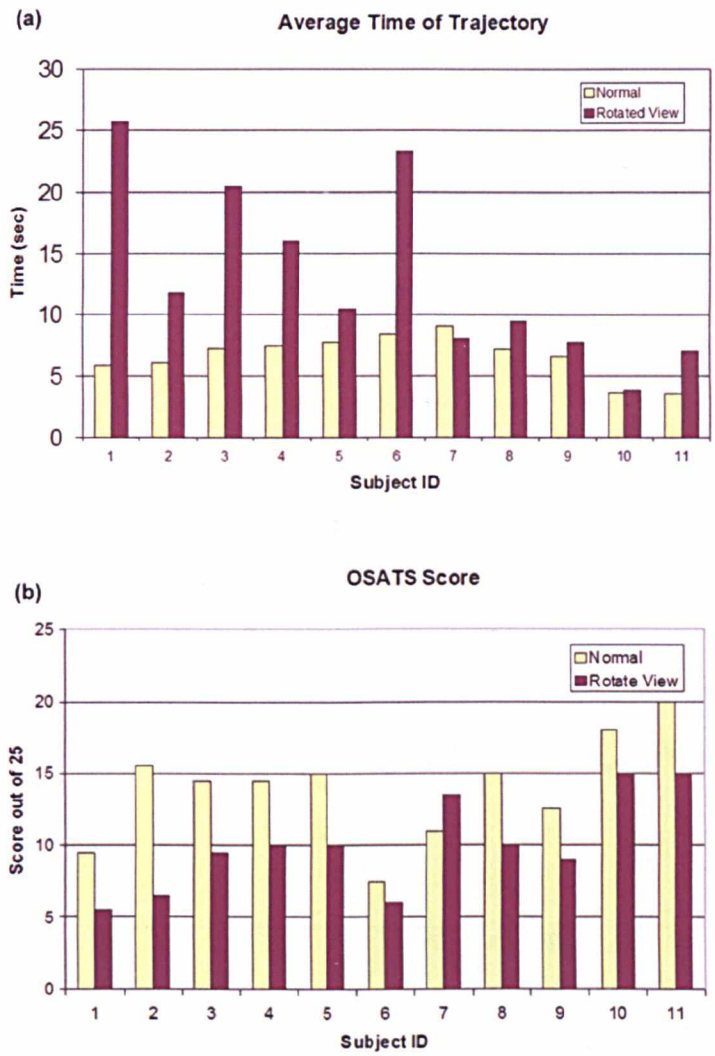

Figure 4.7 (a) The average time for the trajectories of each subject. (b) The modified OSATS score for all the subjects involved in the experiments.

CDF representations of motion trajectories are illustrated in Figure 4.8(a). This figure shows that the surgeon's trajectory generally lies closer to the mean of all the CDF than the novice's trajectory in this particular experiment. However, the average CDF trajectory has a distinctively different shape to either of the sample trajectories.

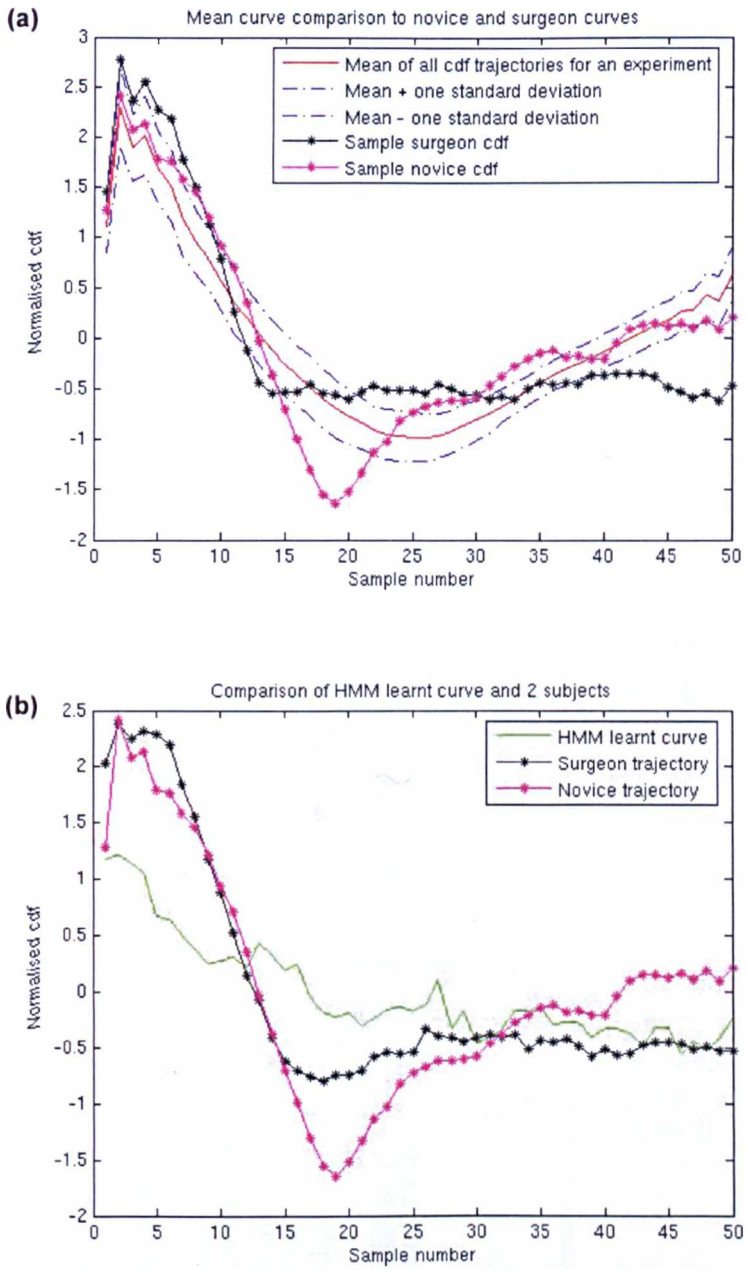

Figure 4.8 (a) The mean of CDF trajectory for one of the experiments with standard deviation. An example of a practicing surgeon's trajectory is shown in black and a novice in pink. (b) HMM learned curve shown in green.

In Figure 4.8(b), the HMM learnt curve is plotted. This is the average of ten observations after training the HMM. Examples of the surgeon's and novices trajectories are plotted for comparison, it shows that the shape of the surgeon's trajectory is more similar to the HMM learnt curve. However, the set of trajectories showed a large variability with effects

of time warping. By calculating the likelihood to the HMM, a measure of similarity of a certain trajectory to the whole group can be obtained.

The log likelihoods of the subjects' trajectories belonging to the training data sets by using the leave-one-out method are illustrated in Figures 4.9(a) and (b). Lower values indicate the subjects' trajectories are more likely to match the models learnt by the HMMs. The log likelihood values are negative. The effect of view rotation accentuates the difference between the subjects as shown in Figure 4.9(b).

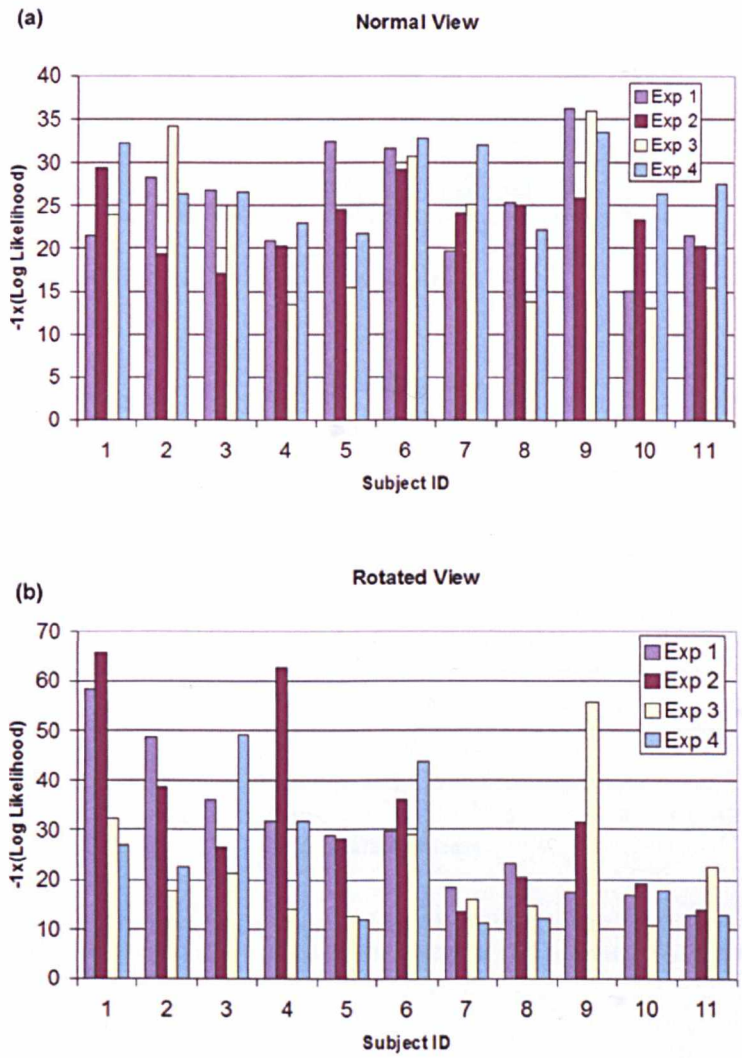

Figure 4.9 (a) The negative log likelihood of the subject in each experiment to belong to the group in the normal view orientation and (b) rotated.

In this chapter, the log likelihood values are ranked in each experiment independently, rank number 1 represents the most similar and *vice versa*. The mean rank of experiment

1-4 for each subject is calculated in both normal and rotated view in Figure 4.10(a). In order to compare this with the most validated surgical rating scale, this rank plotted against the ranking in the OSATS score is shown in Figure 4.10(b). There is a very significant correlation between them ($r = 0.93$, $p < 0.001$).

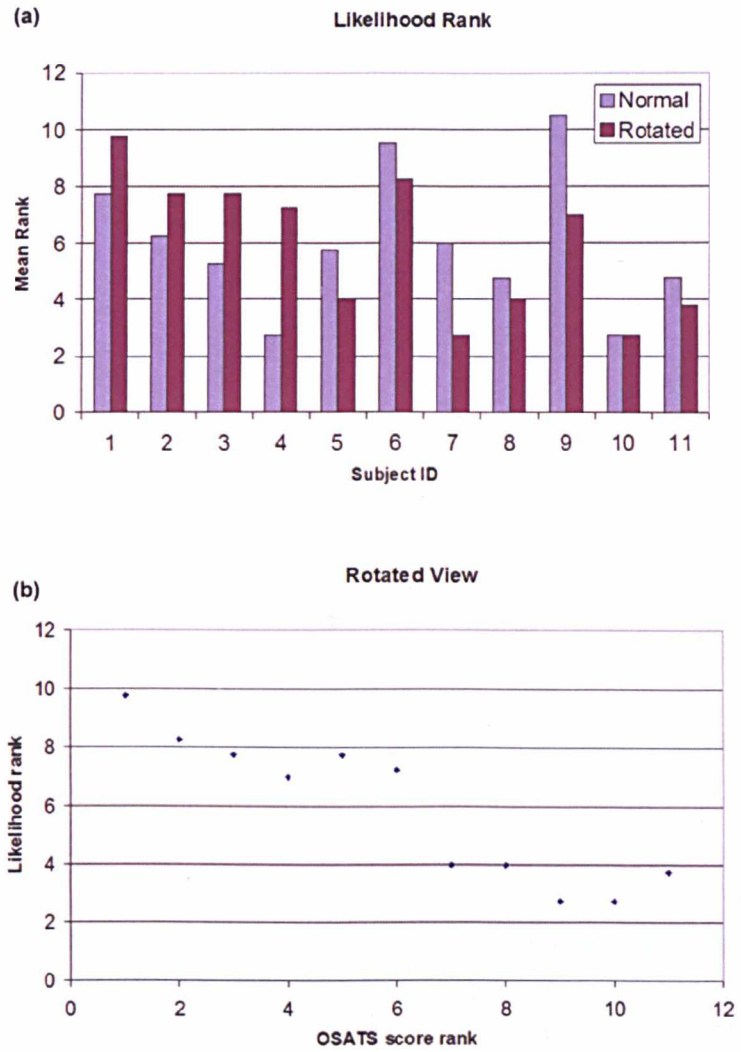

Figure 4.10 (a) The mean rank of subjects' likelihood of belonging to the test group, (b) scatter plot of the rank of likelihood generated by the trained HMM against OSATS score ranking in the rotated tasks.

Pre- vs. Post-training

Figure 4.11 shows the CDF representations of the post-training motion trajectories. This series of figures show that the means of the post-training trajectories generally lie closer to the means of all the CDF, this appears to be more pronounced in Trajectories 3 and 4, which are the movements of the right hand.

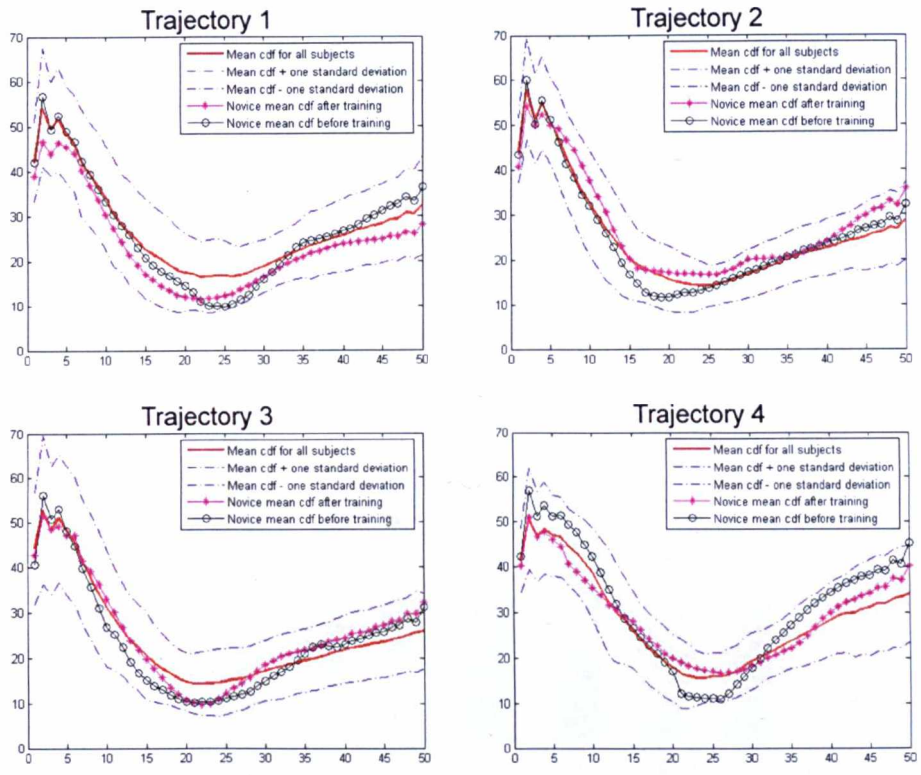

Figure 4.11 The means of CDF for all four trajectories with standard deviation. The means of the pre- and the post-training data are plotted in black and pink respectively.

The log likelihoods in Figure 4.12 were calculated using the leave-one-out method as described above, however only the five novices' pre- and post-training and the two surgeons' trajectories were used as the training data sets. Figure 4.12(a) shows that in all 4 experiments, the log likelihood is lower (more similar to the training data sets) in the post-training sessions, however this isn't statistically significant. Figure 4.12(b) illustrates the likelihood rank of individual subjects' trajectories before and after training.

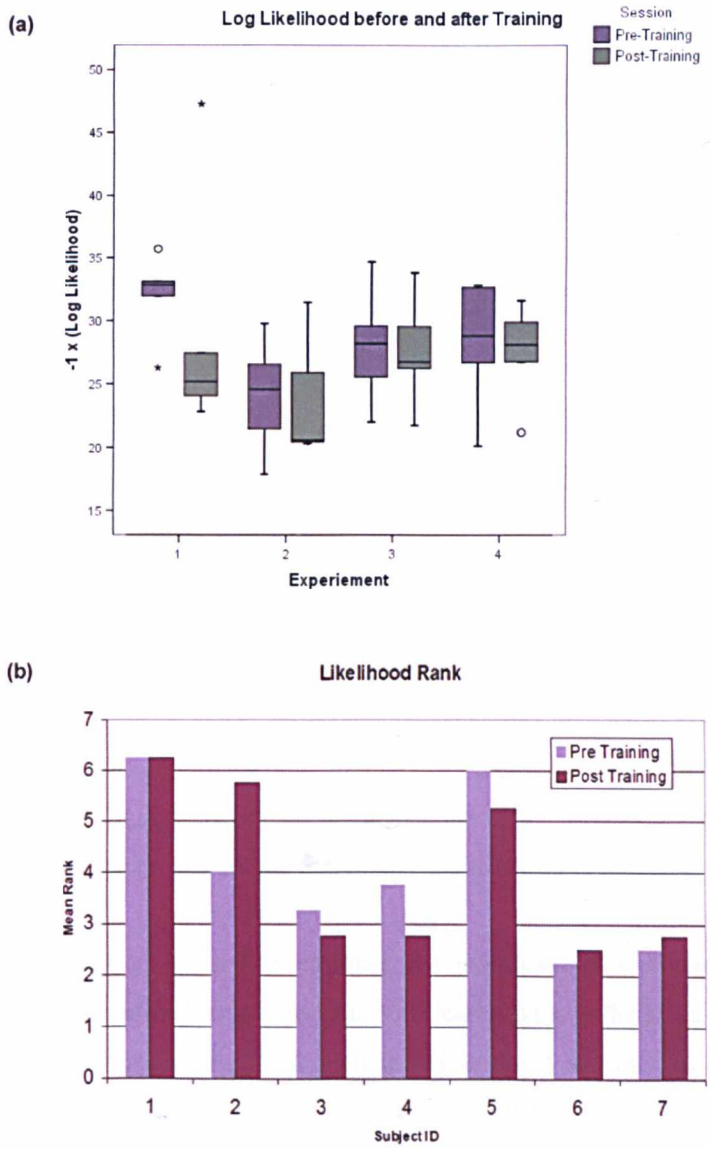

Figure 4.12 (a) Box plot of the negative log likelihood of the subjects pre- and post-training data belonging to the test group where the median (line), inter-quartile range (shaded box), range of the data (whiskers), outliers (circles) and extreme (stars) are plotted; (b) the likelihood ranks of individual subjects' trajectories before and after training.

4.4 Discussion

This chapter has shown that HMM can be used to learn models of surgical motion trajectories even in a group of subjects with mixed abilities without prior classification of technical skills levels. The proposed models learn all the trajectories between two targets from the subjects, and then rank the subjects in terms of consistencies to the trajectories.

As the subjects' skill levels improve through adaptation to the new sensorimotor transformation, their movements between the two targets become more consistent, and hence lead to a lower likelihood output. This seems to be particularly effective when the tasks are performed in difficult laparoscopic environments, *e.g.*, camera rotation, where the difference in skills performance is more pronounced. However, even in such extreme conditions, the subjects seemed to learn the visuomotor map needed to improve in the task. Furthermore, this technique is shown to be sensitive enough to distinguish the effect of the short duration of skills training on the subjects.

Although objective quality scoring systems such as OSATS have been validated and shown to be reliable, these methods are labour intensive and require invaluable time from surgical experts. The speed of a surgeon has long been used as a benchmark of skills, however time is considered a crude metric for performance, whereas OSATS concentrates on the quality of technique. Even in this study, discrepancies are shown between the ranking of OSATS and time performance. Dexterity analysis has been developed to provide a more efficient way for surgical skills assessment, again, parameters such as path length and number of movements may have neglected the importance of the quality aspect of skills evaluation. It remains an effective tool for assessing the performance of simple and highly standardised procedures.

In this chapter, HMM is used to calculate the likelihood of similarity between each individual subject and the learnt trajectories of the group using the leave-one-out method. This likelihood correlated very well with the observation scores, and perhaps this can provide an automated quality scoring system. The second advantage of HMM is the ability to learn models of surgical motion trajectories in a group of subjects with mixed abilities. The method successfully calculated the likelihood of the practicing surgeons as amongst the most representative trajectories (subjects 10 and 11 in Figure 4.10(a)). Although Subject 7 had a lower likelihood rank than Subject 11 (surgeon), and this was reflected in Figure 4.10(b) where the OSATS score was ranked at 9. This was due to the following two factors. First, there may be a 'ceiling' effect where the subject's performance was limited by the methodology used, and the scoring system was not adequate to discriminate the performance further due to the simplicity of the task (152). Second, Subject 7 had unusual spatial awareness, most likely due to her previous training as an aeroplane pilot.

A potential criticism would be the small study population, and the biased numbers of novices when compared to surgeons. It should be noted that even with 11 subjects, a significant correlation could be found between the HMM calculated likelihood with OSATS, which is considered as the 'gold standard' in skills assessment. This further illustrates the sensitivity of the current method. Furthermore, the inclusion of the data of the two surgeons was to demonstrate its ability to differentiate subjects that were both within the same and in different skills groups.

One of the important considerations of applying the proposed HMM scheme is the feature representation of the trajectories. In general, it should be invariant to affine transformation, as this can cope with trajectories with different starting points, rotations and approaching directions. In this chapter, CDF was used as a means of invariant feature representation. Other approaches based on extrema in acceleration measured by high frequency wavelet coefficients are also applicable (153). This technique should be explored in complex laparoscopic procedures and validated in a larger scale study.

This chapter has shown that even by measuring the quality and the shape of the trajectories, a difference was shown between individual's ability in performing in the MIS environment, and the effect of further coordinate transformation introduced when the camera view was rotated. In the next chapter, eye movement data will be included in the analysis, and the focus will shift to the relationship between the instruments and the eyes during the performance of this experiment. This relationship has been shown to change in learning a novel visuomotor task, and has been implied in the measurement of hand-eye coordination.

Chapter 5

Spatiotemporal relationship between eye and instrument tip position on the laparoscopic screen: a pilot study

5.1 Introduction

In the previous chapters, evaluation of the spatiotemporal relationship between the hands and the eyes as a measure of hand-eye coordination has been described (72,108,111). Briefly, in order for the visual system to be used as a feed forward mechanism for movement control, precise visuospatial mapping would have to be established, and the eyes tend to search ahead of the hand.

In a task comparing drawing and tracing, where drawing was seen as a feed forward process and movements are generated internally, whereas tracing depends mainly on external cues and constant visual feedback. Gowen and Miall found that in the drawing task, the eyes move further ahead from the pen tip; however, in the tracing task, eye movements are tightly coupled with the pen tip (111).

In learning a novel visuomotor task, Sailer *et al* found three stages of learning as defined by performance of the task and also the gaze-cursor relationship (72). During the early exploratory phase, gaze generally followed the cursor, with occasional glances to the target. The saccades sizes were generally small (3 to 4°). During the skill acquisition phase, this changed to a predictive behaviour, leading the cursor by up to 300ms. Saccades sizes were larger (4° to 12°), and more were directed towards the target. Finally, at the skill refinement stage, gaze went directly to the next target, with either a single saccade, or a large and one small saccade.

In order to extend this same spatiotemporal relationship in MIS, certain coordination transformations need to be performed. In MIS, the visual information is provided indirectly by the endoscopic camera and displayed onto a 2D monitor screen. Hand movements are used to move the instrument tips, however, these movements are affected by the "fulcrum effect", movement scaling, and the effect of incision as described in Chapter 2.

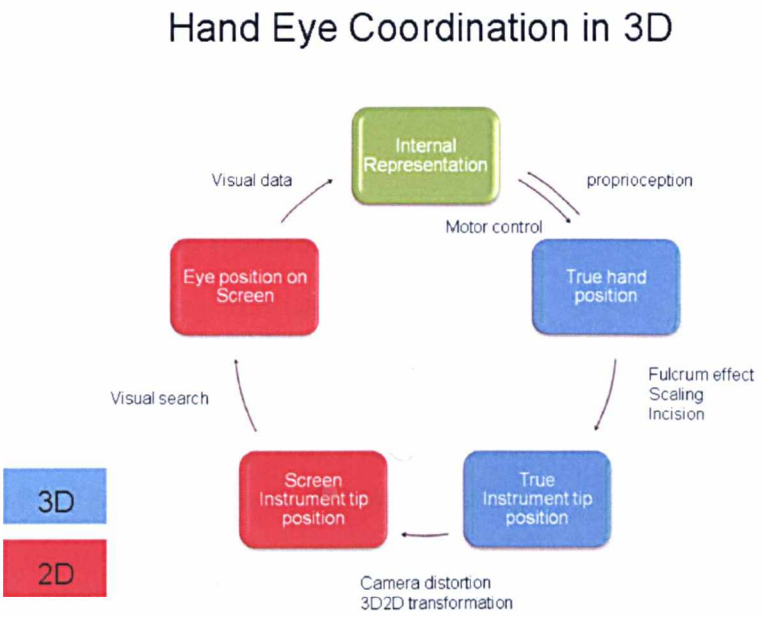

Figure 5.1 Schematic illustration of physical object positions and factors affecting their representations in a 2D monitor screen and the Internal Representation.

In order to measure the spatiotemporal relationship between the eyes and the effector (instrument tips), eye movement parameters together with the screen projection of the instrument tips must be obtained on the laparoscopic screen, as illustrated in Figure 5.1.

5.2 Methods

5.2.1 Eye tracking

A Tobii 1750 eye tracker (Tobii Technology, Stockholm, Sweden) was used to display the laparoscopic video. It is a remote eye tracking device using the standard binocular

VOG technique with an accuracy of 0.5 degrees and a sampling rate of 50Hz, integrated with a 17 inch TFT display with a resolution of 1280x1024 pixels. It can tolerate moderate head movement within a 30x15x20 cm volume at 60 cm in front of the device.

The experimental hardware setup was similar to that of Section 4.2. Additionally, S-Video data obtained from the Karl Storz laparoscopic stack was streamed to a PC with an ATI 9600 AIW graphics capture card at a resolution of 640x480 at 30 Hz refresh rate. This video data was then displayed live on the Tobii eye tracker using the external video function provided with ClearView 2.5.1. Pixel coordinates of the eye tracking data were acquired by using the software provided with ClearView 2.5.1. Raw eye movement data was used for this study.

5.2.2 Instrument tip tracking and screen projection

The most unobtrusive way to obtain the true instrument tip position (3D) is by rigidly attaching a tracking device to the handles of the laparoscopic instruments, as described in 4.2.1.1. Electromagnetic tracking is not feasible in the presence of metal interference in this experimental setup. The tip position can be calculated by applying the rotation matrix on a known translation between the tracking device and the instrument tip.

In order to calculate the two dimensional screen projection of the instrument tip, the S1 camera head attached with the Hopkins II 0° endoscope was calibrated using the principles described by Bouget (128). The extrinsic and intrinsic parameters were then obtained after the camera setup was rigidly fixed, using two laparoscopic camera holders. The 3D instrument tip data was then projected onto 2D data corresponding to its position on the laparoscopic monitor screen.

5.2.3 Data Synchronisation

Two computers were used for obtaining instrument position and eye tracking data to improve performance of the setup. Synchronisation of the two data streams was achieved by using a bespoke software written in Visual Basic, simultaneously collecting the Polaris instrument position data and sending regular timestamp (every 300ms) to the eye tracking computer using TCP/IP standard (average delay 83ms). This is illustrated in Figure 3.11.

5.2.4 Subject and experimental protocol

Data from all 5 subjects who completed all eight tasks from Chapter 4 was used in this pilot study, all subjects were medical students with no previous simulated or real laparoscopic experience.

The pre-training task was to locate two standardised points (A and B) on a simulated plastic small bowel model as illustrated in Figure 4.3. Each point was attached to an in-house designed touch sensitive circuit switch to mark the beginning and end of each trajectory. Each time the circuit was completed, an alarm will indicate a successful contact and a unique marker will be added to the data file. The subjects were then asked to touch alternatively the points A and B with the left instrument 10 times, and then the right instrument, and this step was repeated. A total of 36 trajectories were obtained between the two points.

For the training period, the subjects were required to repeat the first task with the laparoscopic camera rotated 90 degrees counter clockwise, this was repeated twice. The task was then repeated three more times with the camera rotated 90 degrees clockwise. The post-training task was identical to the pre-training task. The subject was allowed to perform a maximum of two tasks per day and each task at least one hour apart.

5.3 Data Analysis

5.3.1 Qualitative analysis

5.3.1.1 Completion time

Using time as a crude measure of performance, the median time for task completion for the pre-training task was 239.4 seconds and for the post-training task was 158.4 seconds (Wilcoxon Signed Ranks Test, $z = -2.02$, $p < 0.05$). There appears to be a gradual improvement throughout the three training tasks with one camera rotation (90 degrees counter clockwise first); however, with the introduction of a new rotation (90 degrees clockwise), the performance worsened for the first session and then gradually improved throughout the rest of the training period.

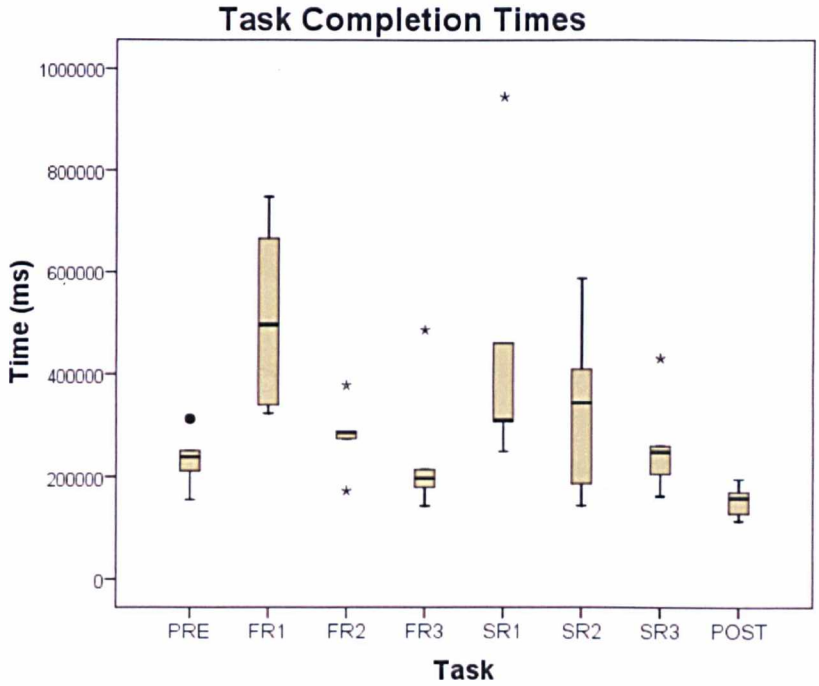

Figure 5.2 Box plot representing the time to complete each task in milliseconds, where the median (line), inter-quartile range (shaded box), range of data (whiskers), outliers (circles) and extreme (starts) are plotted. PRE – Pre Training task, FR – First Rotation, SR – Second Rotation, and POST – Post training task.

5.3.1.2 Target distance function

The screen positions of each target (points A and B) were easily located, as each time the targets were touched, a unique electronic mark was stamped on the data file. Figures 5.3 – 5.5 illustrate the Target Distance Function (TDF), where the distance of the instrument tips (and eye fixation point) from the target is plotted against time (or sample number). As the distance approached to 0, it indicated that the instrument tips or the eye fixations landed on the target. Qualitative analysis of the example data from the pre-training, first rotation, and post-training tasks was made. With the introduction of the first rotation task, the completion time was more than doubled, whereas the post-training completion time was almost halved when compared to the pre-training time.

Furthermore, just examining the instrument tip data, similar to Figures 4.5 and 4.6 using 3D instrument tip distance data. The trajectories near the targets (distance 0) become more disorganised and the troughs are less distinct in **the first rotation task**, reflecting the increased motor errors generated when screen rotation was first introduced. The

instrument trajectory was clearly not coordinated. In the post training task, the downward slopes of the trajectories are steeper, as the instrument tips approached the target, which represents faster and less error-prone feed forward movements.

The pink crosses in Figures 5.3 – 5.5 represent the eye position data. The most pronounced difference is seen in **the first rotation task**, here the overall density of data points is much higher and tightly coupled with the instrument; however, more importantly this density increases as the eye position approaches the targets (distance 0). As described in the instrument tip data, large movement errors were generated in the final parts of the trajectories towards the target, when the screen was rotated. This is when visual feedback was most required, to guide the fine tuning of the instrument movements. After training, the sensorimotor map was better developed and less visual input was required in the final stage of each trajectory, this is reflected by the reduced number of data points near the targets (distance 0).

Finally, the post-training eye movement, when compared with the pre-training task, larger saccadic jumps were seen with fewer fixations between the targets. It appears that after training, the requirement of visual input in movement generation was much less.

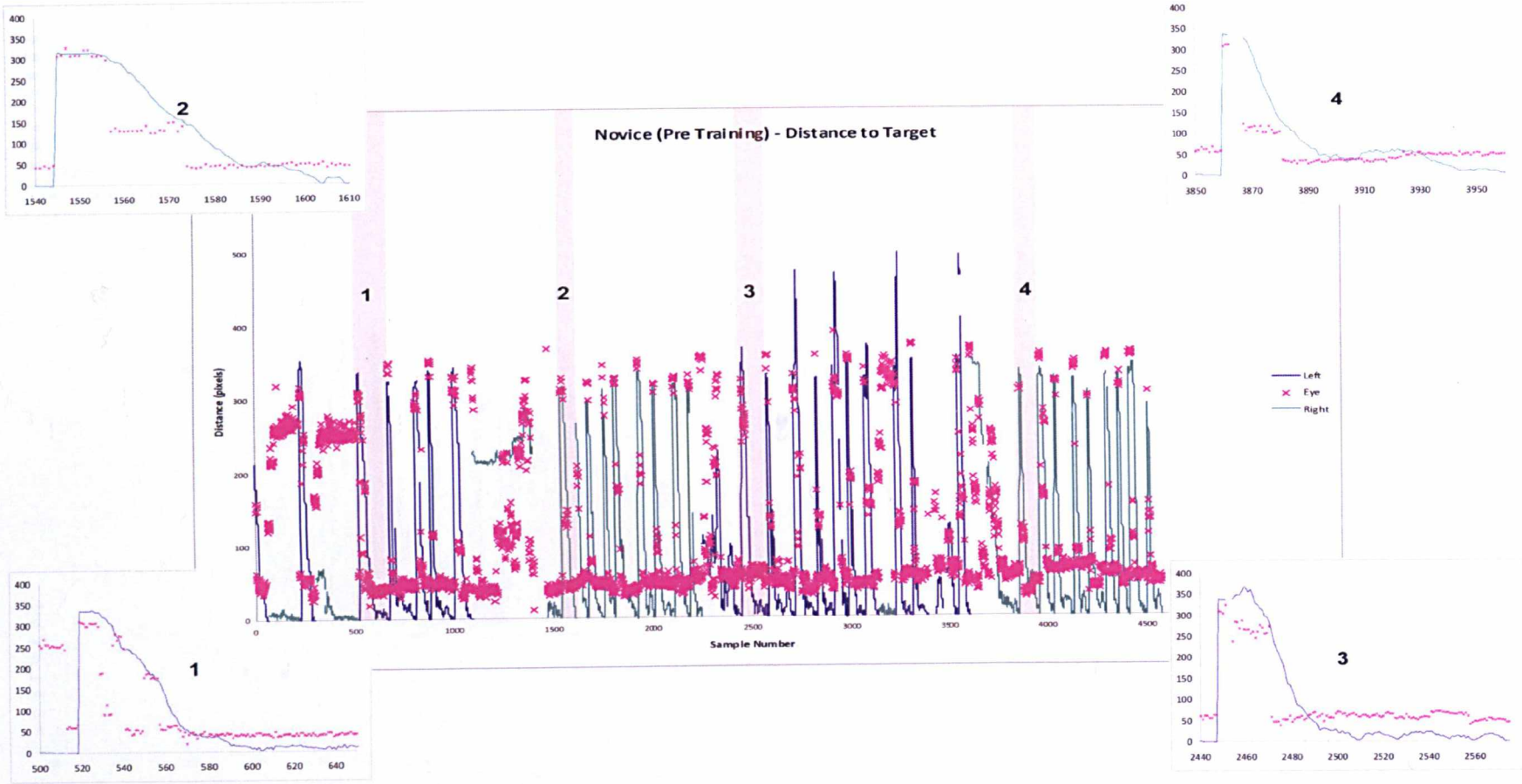

Figure 5.3 An example of novice **pre-training data** showing the spatiotemporal relationship between the pre-training, first rotation, and post-training tasks Blue line – left instrument tip, Green line – right instrument tip, and Pink crosses – eye position on screen. Pink highlighted areas are magnified as numbered.

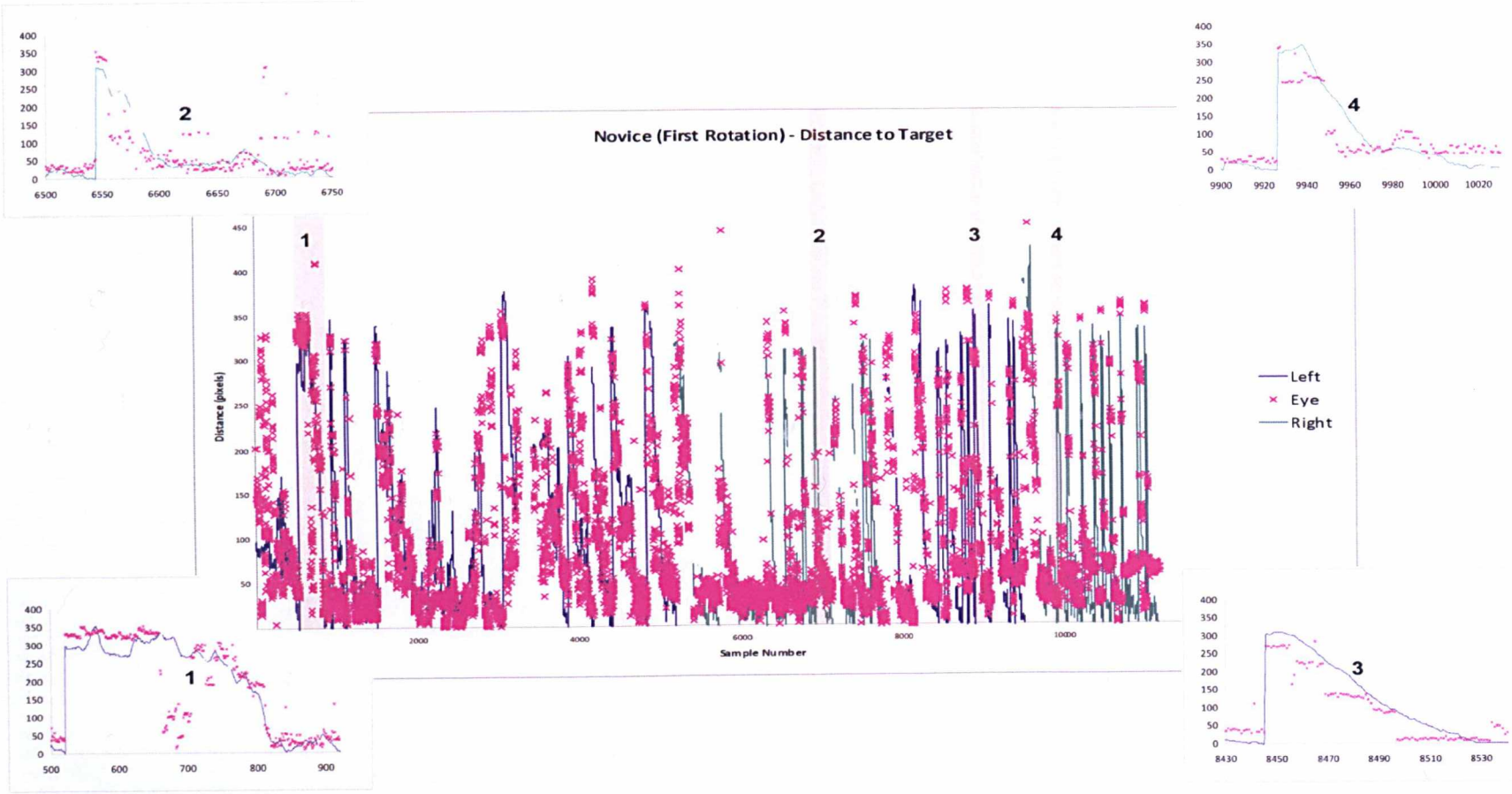

Figure 5.4 An example of novice **first rotation data** showing the spatiotemporal relationship between the pre-training, first rotation, and post-training tasks Blue line – left instrument tip, Green line – right instrument tip, and Pink crosses – eye position on screen. Pink highlighted areas are magnified as numbered.

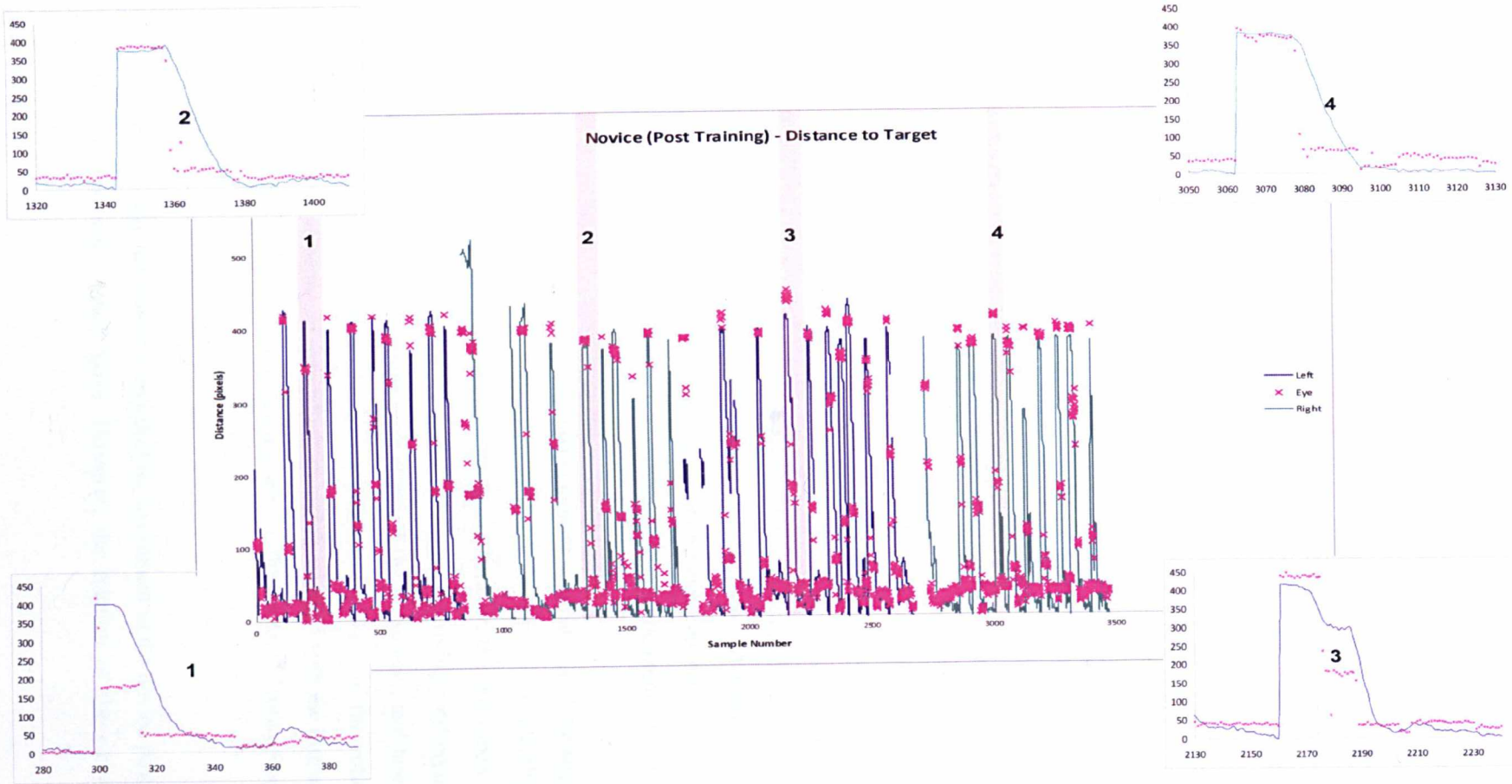

Figure 5.5 An example novice **post-training data** showing the spatiotemporal relationship between the pre-training, first rotation, and post-training tasks Blue line – left instrument tip, Green line – right instrument tip, and Pink crosses – eye position on screen. Pink highlighted areas are magnified as numbered.

5.3.2 Spatiotemporal relationship

To illustrate the temporal difference between the eye and the instrument position, Figure 5.6 represents typical trajectories between Points A and B. The red dots on the right represent the eye fixations, as they approach the target (black line) with time. The blue dots represent the instrument tip positions. These trajectories can be divided into two parts: the initial fast ballistic movement towards the target, followed by slower fine adjustments near the target for final location.

Novice data is presented in the top three subplots in this figure, showing the pre-training, first camera rotation, and post-training tasks. Comparing the **pre- and post training tasks**, smaller eye saccades that are more tightly coupled with the instrument are seen in the initial phase of the trajectory in the pre-training task. This is followed by a much longer second phase for reaching the target, where the small movements of the instrument and the eyes represent fine corrections of motor errors by heavy involvement of the visual feedback mechanisms. In the **first rotation task**, the trajectory is the longest. No distinct separation is seen between the first and second phase of movement, and it clearly shows mirroring eye pursuit movements with the instrument throughout the whole trajectory. Instrument movement errors are characterised by increasing distance from the target, and these are presented frequently in this trajectory. In the **post-training task,** two large eye saccades are used to reach the target, and the distance between the eye fixation and the target is consistently shorter than the instrument, displaying predictive eye behaviour.

The trajectories in the normal and rotated camera orientations of the **surgeon** are shown in the bottom two graphs in Figure 5.6. Although the time to complete the trajectories is similar, with no obvious qualitative differences between the instrument movements are seen in the two tasks. The eye movement behaviour is strikingly different: in the **normal rotation task**, the eye movements are ahead of the instrument, and large predictive eye saccades are seen similar to the novice post-training data; in the **rotated task**, again pursuit eye movements are seen which are tightly coupled with the instrument throughout the trajectory. Movement errors are not seen in the surgeon's rotated task, unlike in the novice data.

In summary, eye movement is ahead of the instrument in the novice post-training and the surgeon's normal rotation tasks. However, the rotation of the camera changes this

relationship, where the distance between the eye fixation and the target remains the same as the reciprocal distance between the instrument and the target.

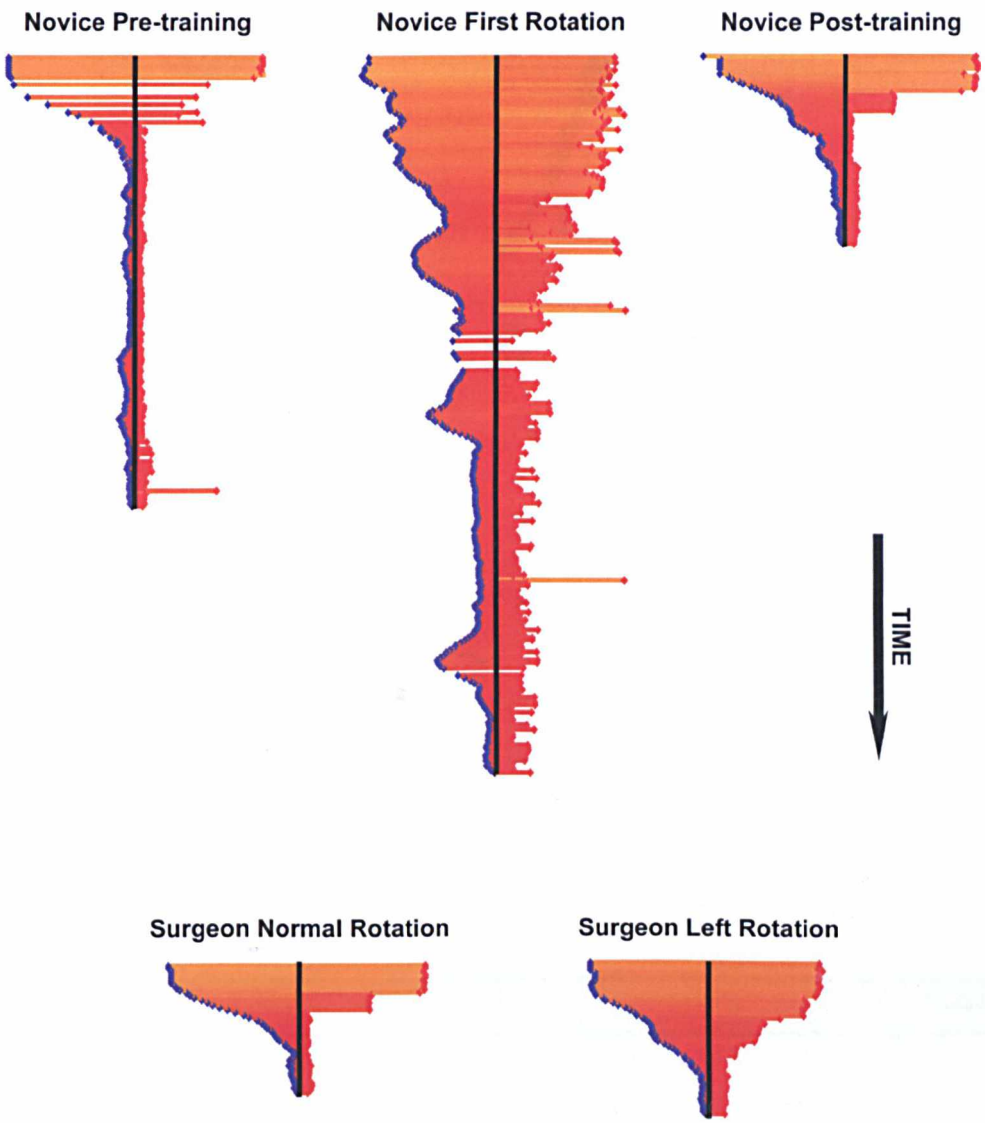

Figure 5.6 Typical trajectories illustrating the spatiotemporal relationship between the instrument tip and the eye. Blue dots – instrument tip position, Red dots – eye position, Black line – target.

To investigate whether the eyes lead or lag the instrument tip, the Cartesian distance between the instrument and the target was compared to the distance between the eye fixations to the target. As the instrument tip approaches the target, if the eyes are lagging behind, then the distance between the target and the eye fixation would be longer, as illustrated in Figure 5.7.

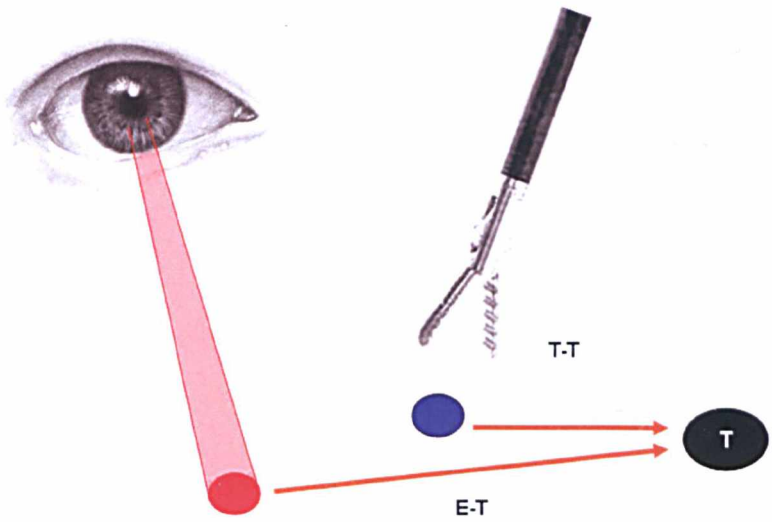

Figure 5.7 Blue dot illustrating instrument tip moving towards target (T), and Red dot representing the eye fixation. If the eyes are lagging behind, E-T would be longer than T-T.

In both the pre-training and the first rotated task, the eye-target (E-T) distance was significantly longer than the tip-target (T-T) distance. This represents instrument leading behaviour, in other words, the eyes are lagging behind. The T-T distance becomes significantly longer after FR1, with FR2 and SR2 having no significant difference between T-T and E-T. This is summarised in Table 5.1.

	PRE	FR1	FR2	FR3	SR1	SR2	SR3	POST
Median E-T	55.17	67.98	42.42	46.71	53.11	61.02	43.86	48.11
Median T-T	43.86	61.22	42.15	54.92	57.52	57.42	51.91	48.78
WSR test	z=-8.08 p<0.001	z=-9.55 p<0.001	z=-0.5 p>0.05	z=-8.91 p<0.001	z=-35.3 p<0.001	z=0 p>0.05	z=-29.4 p<0.001	z-6.93 p<0.001

Table 5.1 Summarising results of eye-target and instrument tip-target distance comparisons. Yellow highlight signifies statistical significance, and green highlight signifies the larger distance. E-T and T-T distances are measured in pixels. PRE – Pre training, FR – First Rotation, SR – Second Rotation, POST – Post training tasks. WSR – Wilcoxon Signed Rank

To illustrate this further, Figure 5.8 represents the error bars with 95% confidence interval of the difference of T-T from E-T, *i.e.*, E-T minus T-T. Positive values denotes E-T longer than T-T, and symbolises instrument leading behaviour, and *vice versa*. This clearly shows that during the Pre-training and First Rotation tasks, the ET-TT difference is larger than in the Post Training task. The data was averaged for clearer illustration, hence the absolute numerical values on the *y*-axis are less meaningful, as the data is non-parametric.

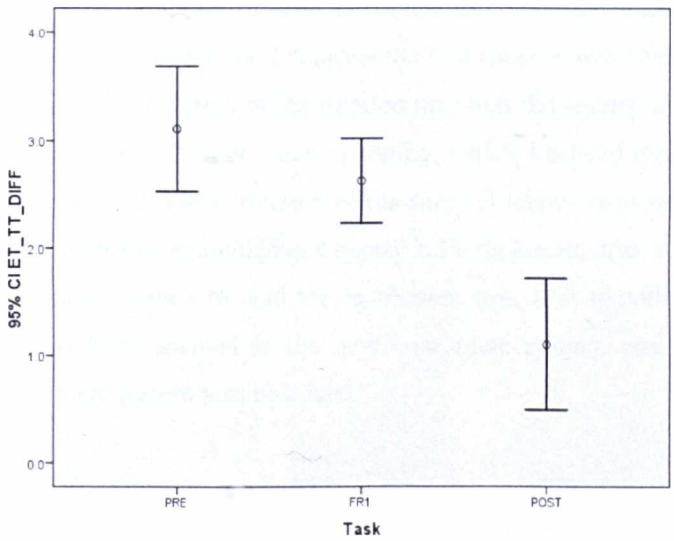

Figure 5.8 Error bars plotting the mean ET minus TT distance, with 95% Confidence Intervals. PRE – Pre training. FR1 – First Rotation (1), POST – Post training tasks.

5.4 Discussions

The effects of novel coordinate transformations due to the natural setup of laparoscopic surgery and the introduction of screen rotations were examined in this chapter. The increased complexity of the task was confirmed by simple time analysis of task completion, which increased as screen rotation was introduced, but decreased as training continued. Qualitatively, the eye movement and instrument trajectories varied independently with these effects. However, the intrinsic spatiotemporal relationship between the eyes and the effector was further examined.

This chapter demonstrated the feasibility of calculating the spatiotemporal relationship between the eyes and the instrument tip using simple Cartesian distance. It has shown that

the E-T distance was greater than the T-T distance in the pre-training and the first rotation task. In the pre-training task, the subject encountered the laparoscopic coordinate system for the first time, hence the poor performance in the task. Interestingly, during the first rotation task, where further changes in the visuospatial transformation were introduced, similar effects were observed. This seemed compatible with the hypothesis that the spatiotemporal relationship between the eyes and the instrument tip projection is affected by novel visuospatial transformation. However, when the second camera rotation was introduced, no significant effect on the spatiotemporal relationship between the eyes and the instrument was seen. It should be noted that the direction of camera rotation was randomised for the subjects, hence for 3 subjects the first rotation was counter clockwise, and 2 were clockwise. The direction of the rotation may bias the results, as most subjects tended to tilt their heads to the right when operating, which rendered clockwise camera rotation (causing counter clockwise rotation of the surgical scene) tasks more difficult to perform. This effect is further examined in Chapter 7. Furthermore, after a brief period of training, eye movement seemed to lead the instrument tips. It is hypothesised that the visuospatial map has been adapted to the new coordinate system, and hence a more predictive eye movement pattern was observed.

This analysis technique can only be used when the location of the target is known, and that each movement is aimed at locating the two fixed targets. The experiment was setup specifically to test the hypothesis, but it did not represent real laparoscopic surgery. During an operation, there are multiple targets which vary between and even within operations. The screen locations of these targets are difficult to deduce automatically: first these targets are chosen intuitively by the surgeon, and prediction of these targets would be impossible, without even considering the order that they would be attended to; second, imaging technique for anatomical soft tissue tracking has been proven to be difficult, due to the complexity and the highly deformable nature of the structures; third, it is also highly possible that the targets can be moved during the procedure, either by the surgeon intentionally or un-intentionally; finally, each movement trajectory during a complex laparoscopic surgery does not necessarily move towards the targets, for example changing instruments, or lifting a flap of tissue to reveal the target.

Despite the weaknesses of the analysis technique, this study remains the first measurement of the hand-eye spatiotemporal relationship as a surrogate calculation of hand-eye coordination in surgery. In order to apply these principles in general, advanced analysis algorithms need to be developed, where prior trajectory segmentation is not

needed for analysis. Instrument and eye movement can be seen as time series data, and algorithms have been developed for forecasting future events, especially in econometrics where the opening stock market share price is dependent on past performances. This predictability, or causality, may be useful in examining the relationship between the eyes and the hand.

The next chapter presents the development of a new analysis framework where the target locations are assumed to be unknown. The general pattern of the instrument and eye movements will be used for deduction of their spatiotemporal relationships. In order to ensure general applicability of the method, a more detailed experiment is to be designed to mimic a live laparoscopic procedure, where realistic subtasks are included.

Chapter 6

Investigation of Partial Directed Coherence for Hand-Eye Coordination in Laparoscopic training

6.1 Introduction

Hand-eye coordination is a complex combinatory problem, involving many sensorimotor coupling including the visual, proprioceptive, tactile, attention and motor systems. However, the study of hand-eye coordination can be simplified to a model describing the use of the visual system to guide/validate the movements of the hands. This fundamental fact implies a black box approach in the understanding of the function of the whole system (86).

Predictive, rather than reactive, eye movements have been implicated in the feed forward paradigm of visual assistance in well rehearsed movement generation. In learning a new visuospatial task, the spatiotemporal relationship of the eyes and the effector has been used to provide a framework to divide learning into distinct phases; where vision is first used as a feedback mechanism until the new coordinate transformation becomes automated (72). A number of studies have confirmed that the prediction of the sensory consequences of actions defines early stages of motor learning (73).

The hand and eye movement data can be depicted as complex interrelated time series, and thus far, a generalised quantification of their mutual predictability has not been described. Several techniques have been investigated to describe the interdependency between multivariate series. For example, Graphical interaction models describe the interrelationships among the components of a time series as undirected graphs, where vertices depict the components and edges indicate possible dependencies between the

components (154). Parametric models have been used to illustrate the dependence structure of a process as a model selection problem (155). In the frequency domain, cross spectrum analysis and coherence have also been used to describe the interdependency between two time series.

However, when causality between time series is of relevance, or when more than two multivariate time series are being observed, the distinction of indirect and direct relationships between them becomes important. While ordinary coherence measures the linear dependence between two signals, **directed** coherence (156-158) focuses on the feed forward and feedback aspects of relationships between signals.

Causality in general has three essential criteria: first, the cause must precede the effect in time; second, the two variables must be associated; third, the correlation cannot be explained by a third variable. It was also defined that if the prior knowledge of a series predicts another, the former Granger-causes the latter. The last chapter described the use of simple Cartesian distance to measure the relationship between the eyes and the instrument tips with known target locations, and the effect of screen rotation and learning changed this relationship. The purpose of this chapter is to evaluate the hypothesis of the interdependency of hand-eye movement along with the variability of their temporal relationships. Partial Directed Coherence (PDC) is used as a frequency domain description of Granger-causality for measuring the predictability of hand and eye motions.

6.2 Measuring partial directed coherence

6.2.1 Autoregressive Models

The AutoRegressive model (AR) is a linear prediction formula that predicts an output x_t based on the weighted sum of its previous values and an error. The order of the AR denotes the effect of the number of previous values that have an effect on the current value, hence a first order AR, or AR(1) could be described as:

$$x(t) = kx(t-1) + e$$

where k is constant with an absolute value of less than 1, and e represents the error with a mean zero and a finite variance (often a normal distribution). Here, only the immediate past value of x has an effect on the current value.

Hence, AR[p] is:

$$x(t) = k_1 x(t-1) + k_2 x(t-2) + ... + k_p x(t-p) + e$$

In this univariate model, $x(t)$ can be influenced by the all past values of the system to the order p. The multi-equation extension of AR where more than one variable is described, the Vector AutoRegressive model (VAR) can be use, a two-variable VAR[1] is:

$$x_1(t) = A_{1,1} x_1(t-1) + A_{1,2} x_2(t-1) + e_1$$
$$x_2(t) = A_{2,1} x_1(t-1) + A_{2,2} x_2(t-1) + e_2$$

or

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

6.2.2 Partial directed coherence

PDC was introduced for the inference of Granger-causality in the frequency domain (156-158). Granger-causality by definition states that an observed time series $x_c(n)$ Granger-causes another series $x_i(n)$ if the knowledge of $x_c(n)$'s past significantly improves prediction of $x_i(n)$. This type of predictability is not reciprocal. Assessing Granger-causality provides a measure of the strength of interaction between time series under the rational that predictable variations in a series take place if their mechanisms of generation are intrinsically linked. In a linear framework, Granger-causality is related to VAR.

If $(x_1(t),....x_n(t))'$ is a stationary n-dimensional time series with mean zero, a vector autoregressive model of order p, VAR[p] can be abbreviated by:

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} = \sum_{r=1}^{p} \mathbf{a}_r \begin{pmatrix} x_1(t-r) \\ \vdots \\ x_n(t-r) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \vdots \\ \varepsilon_n(t) \end{pmatrix}$$

The vector $(\varepsilon_1,\ldots,\varepsilon_n)'$ denotes independent Gaussian white noise, and $\mathbf{a}_r$ is the coefficient matrix of the VAR. To guarantee stationarity of the model, where the averages, variances and covariances remain constant with time, the assumption was made that $\det(I - \mathbf{a}(1)z - \cdots - \mathbf{a}(p)z^p) \neq 0$, for all $z \in C$ such that $|z| \leq 1$ (159).

In the above equation, the coefficients $a_{ic}(r)$ describe how the present values of x_i depend linearly on the past values of the components x_c. Thus, if all entries $a_{ic}(r)$ are zero for $r = 1....p$, then x_c does not Granger-cause x_i. In other words, linear prediction of $x_i(t+1)$ based on the past and present values of all variables but x_c cannot be improved by adding past and present values of x_c.

In this chapter, p_{opt}, the optimal order of an AR model was chosen as the optimizer of Schwarz's Bayesian Criterion. This was investigated by Lütkepohl (159) who found that this selection criterion leads, on average, to the smallest mean squared prediction error of the fitted model. Baccala and Sameshima (156) introduced the concept of the frequency-based PDC, where the difference between the n-dimensional identity matrix and the Fourier transform of the coefficient series is:

$$\mathbf{A}(w) = \mathbf{I} - \sum_{r=1}^{p} \mathbf{a}(r)e^{-iwr}$$

and the PDC is defined as (157):

$$\pi_{i\leftarrow c}(w) = \frac{|\mathbf{A}_{ic}(w)|}{\sqrt{\sum_k |\mathbf{A}_{kc}(w)|^2}}$$

PDC takes a value between 0 and 1 due to the normalisation in the equation above. This provides a measure of the influence of previous samples of x_c on the present samples of x_i with the effect of x_c on other variables. Thus, it can measure the strengths of interactions between two signals in a directed manner. ARfit was used in this study for the estimation of the parameters of the AR model (160).

Figure 6.1 provides an illustration of the causal influences between the eyes and the two hands with their corresponding PDC calculations. Focusing on the off-diagonal diagrams, when causal relationships are present, significant PDC values are highlighted in red. Notice that this causality is directional.

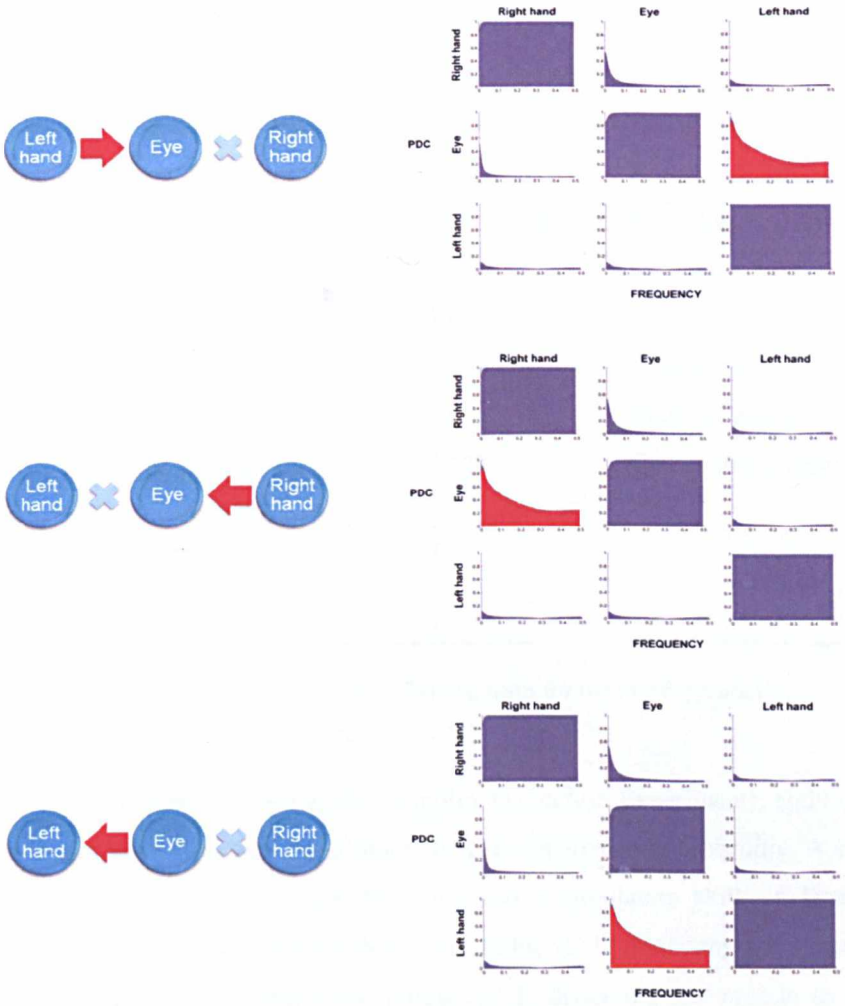

Figure 6.1 A schematic illustration of PDC calculations. Left column – red arrow demonstrates the direction of the influence between the eyes and the two hands. Right column – PDC values plotted against frequency, significant PDC values are in red.

6.2.3 Experiment setup

For assessing basic hand-eye coordination involved in MIS, two experimental datasets were used. The first dataset (referred as the Target Location Experiment in this chapter) derives from Chapter 4 and 5, involved a task of locating two standardised points on a simulated plastic small bowel model using laparoscopic instruments, each point was attached to a circuit switch to mark the beginning and end of the trajectories. The subjects were asked to locate alternatively the two points with the left instrument 10 times, and then the right instrument (Task T.1). View rotation tasks were introduced in order to increase the complexity of hand-eye coordination, as the ability to handle mental rotation tasks has been suggested to be indicative of the innate capacity in mastering laparoscopy (27,28,134). The subjects were required to repeat the task with the laparoscopic camera rotated 90 degrees counter clockwise for three training sessions (Tasks T.2 – T.4), and clockwise for three further sessions (Tasks T.5 – T.7). Finally, a post training assessment (Task T.8) using normal camera orientation was completed. Data from nine complete novices was used for this experiment, although only 5 subjects completed all eight tasks.

Target Location Experiment		**Complex Dissection Experiment**	
T.1	Pre-training	C.1	Skin dissection
T.2	First rotation 1	C.2	Muscle dissection
T.3	First rotation 2	C.3	Tissue removal
T.4	First rotation 3	C.4	Skin reposition
T.5	Second rotation 1	C.5	Return instruments
T.6	Second rotation 2		
T.7	Second rotation 3		
T.8	Post-training		

Table 6.1 Summarises the individual tasks for the two experiments.

For the second dataset (referred as the Complex Dissection Experiment), eight complete novices were recruited to perform a more complex laparoscopic procedure. A cadaveric avian model was used to simulate dissection and manipulation skills in laparoscopic surgery. The procedure was broken down into tasks: (C.1) dissecting the subcutaneous connective tissue over the pectoralis muscle, (C.2) dissecting the muscle to reveal a simulated tumour tissue, (C.3) removal of the simulated tissue, (C.4) repositioning the dissected skin layers, and (C.5) returning the instruments to the start positions. The tasks for the two experiments are summarised in Table 6.1. Each subject completed the

procedure **ten** times in 3 separate sessions. Figure 6.2 illustrates the experimental setup and screenshots of the key tasks.

For both experiments, instrument tip positions were obtained by bespoke designed IRED tracking devices that were attached rigidly to the handles of the laparoscopic instruments. They were tracked by the Polaris (Target Location Experiment) and Optotrak Certus (Complex Dissection Experiment) systems (Northern Digital Inc, Ontario, Canada). The offsets of the instrument tips from the IRED markers were calculated using the Pivot function of the NDI software. Data interfacing was achieved through RS-232 and the provided tracking accuracy was 0.33 mm RMS at a sampling rate of 60Hz for Polaris, and 0.15 mm RMS at a sampling rate of 50Hz for Optotrak Certus. As in Chapter 5, A Tobii 1750 eye tracker (Tobii Technology, Stockholm, Sweden) was used to display the laparoscopic scene. It is a remote eye tracking device using the standard binocular VOG technique integrated with a 17 inch TFT display. It can tolerate moderate head movement thus providing a relatively natural environment for laparoscopic tasks (65,141).

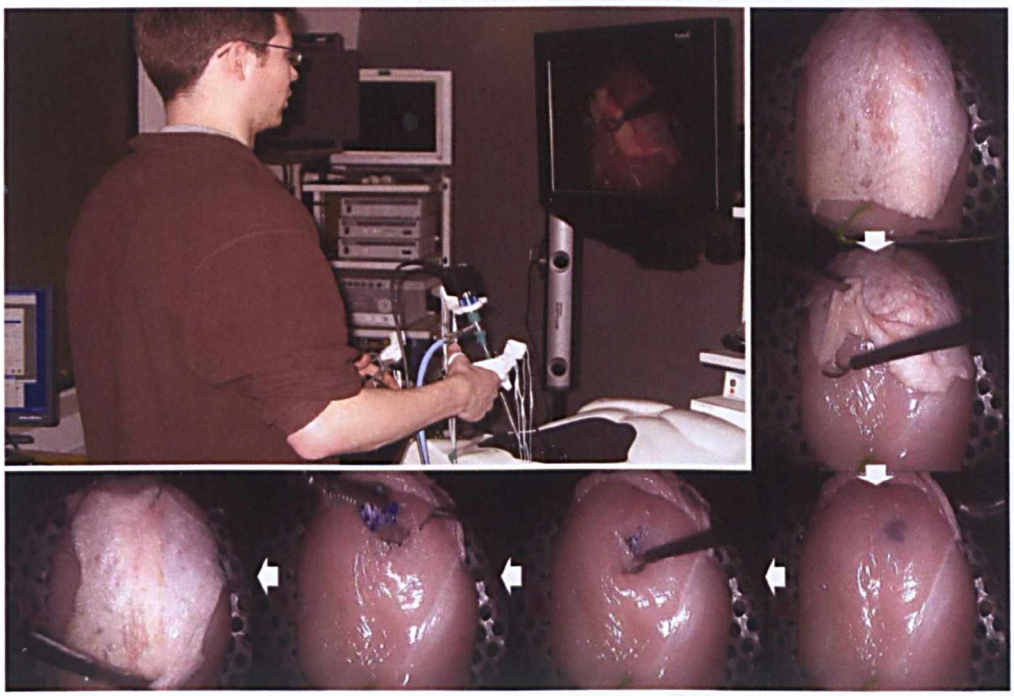

Figure 6.2 Experiment setup with screen shots showing the key manipulation steps involved in the laparoscopic workflow of Experiment 2.

6.2.4 Pre-processing

In order to calculate the 2D screen projection of the instrument tips for the Target Location Experiment, the laparoscopic camera was calibrated using the camera calibration toolbox. The extrinsic and intrinsic parameters were then obtained after the camera setup was rigidly fixed, allowing 3D instrument tip data to be projected onto 2D data corresponding to its position on the laparoscopic monitor screen. For the analysis described above, the Target Distance Function (TDF) was used to reduce the dimensions of the data, where the instrument and eye positions were expressed as the normalised Euclidean distance from the target. This converts each data stream into single time series for PDC analysis.

For the Complex Dissection Experiment, key steps of the workflow were marked manually. The time to complete each trial was compared across the trials by using repeated Friedman Test for non-parametric related samples. When the trials are no longer significantly different from each other, a performance improvement plateau is reached. Non-parametric comparisons between 2 related samples were calculated using Wilcoxon signed rank test, and Spearman's rank correlation was used to detect trends. The significance level of $p < 0.05$ was used throughout the study. TDF was also used for dimension reduction of the data, however, as there were multiple targets with unknown locations, the origin (0, 0) was used as an arbitrary target.

Figure 6.3 presents a sample instrument movement data with simulated eye movement data to illustrate 3 different conditions: (1) the simulated eye data is identical to the instrument data, but lagging temporally by 8 samples (top); (2) the simulated eye data is completely unrelated to the instrument data (middle); (3) the simulated eye data is 8 samples ahead of the identical movement data (bottom). The corresponding actual PDC calculations are presented in the right column.

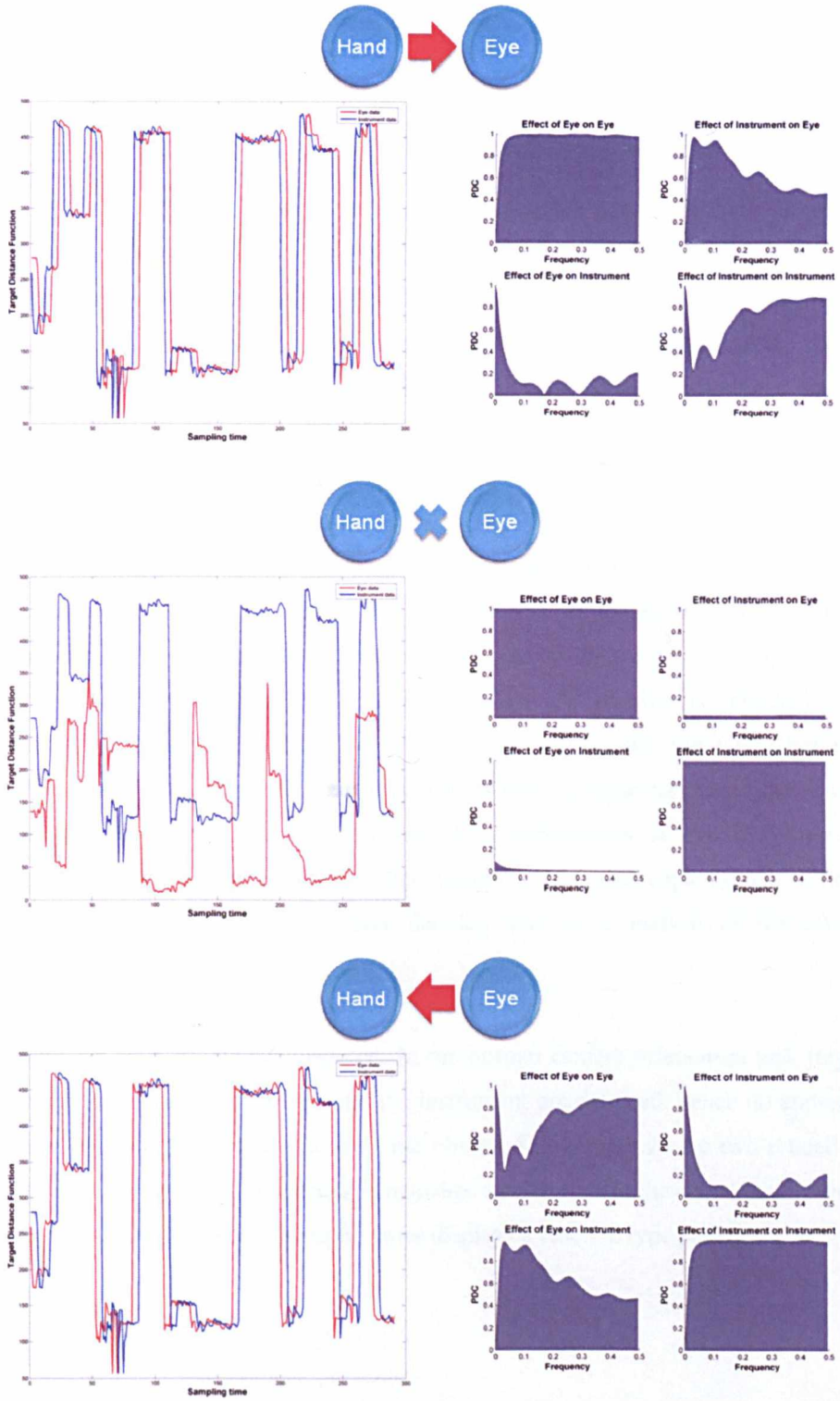

Figure 6.3 Graphical illustration of PDC results related to instrument and simulated eye movements: top – instrument leading behaviour, middle – no casual relationship between instrument and eye, bottom – eye leading behaviour. Right column illustrates the PDC results, and left column illustrates the data illustrations (red – eye, blue – instrument).

6.3 Results

6.3.1 TDF and PDC for the Target Location Experiment

The TDF and the corresponding PDC calculations of the novice data are presented in Figure 6.4. In the pre-training task (**T.1** as shown in the top figures), the TDF data shows tight coupling between the instrument and eye movements, where small eye saccades are displayed throughout each movement trajectory. The corresponding PDC shows a significant effect of the **instrument on the eye** data, describing reactive type eye movements. The middle figures correspond to the first camera rotation task (**T.2**). This is the most difficult task in terms of coordination, as the dissociation of the eye-hand axis was introduced the first time to the subject. The instrument movement is shown to precede the eye movements when approaching the target (TDF = 0). Here, the corresponding PDC analysis detects causal influences of the **instrument on the eye movements** again. However, in the bottom figure, illustrating the post training task (**T.8**), the subject's improved hand-eye coordination after the training is reflected by the predictive eye behaviour, where the eye movement approaches the target before the instrument tip. Higher corresponding PDC values when comparing the influence of **eye movements on instrument** are observed. This corresponds to the hypothesis that predictive eye behaviour develops after training in laparoscopic skills, reflecting improved hand-eye coordination. More detailed qualitative analysis of the data was performed in Chapter 5.

In Figure 6.5, expert data is illustrated. In the normal camera orientation task (top), no causal influences between the eye and the instrument are detected, hence no appreciable temporal between the two data streams are observed. However, in the two rotated tasks (middle and bottom), PDC analysis distinguishes a difference in the coordination between the instrument and the eyes, where the latter displays a reactive type pattern.

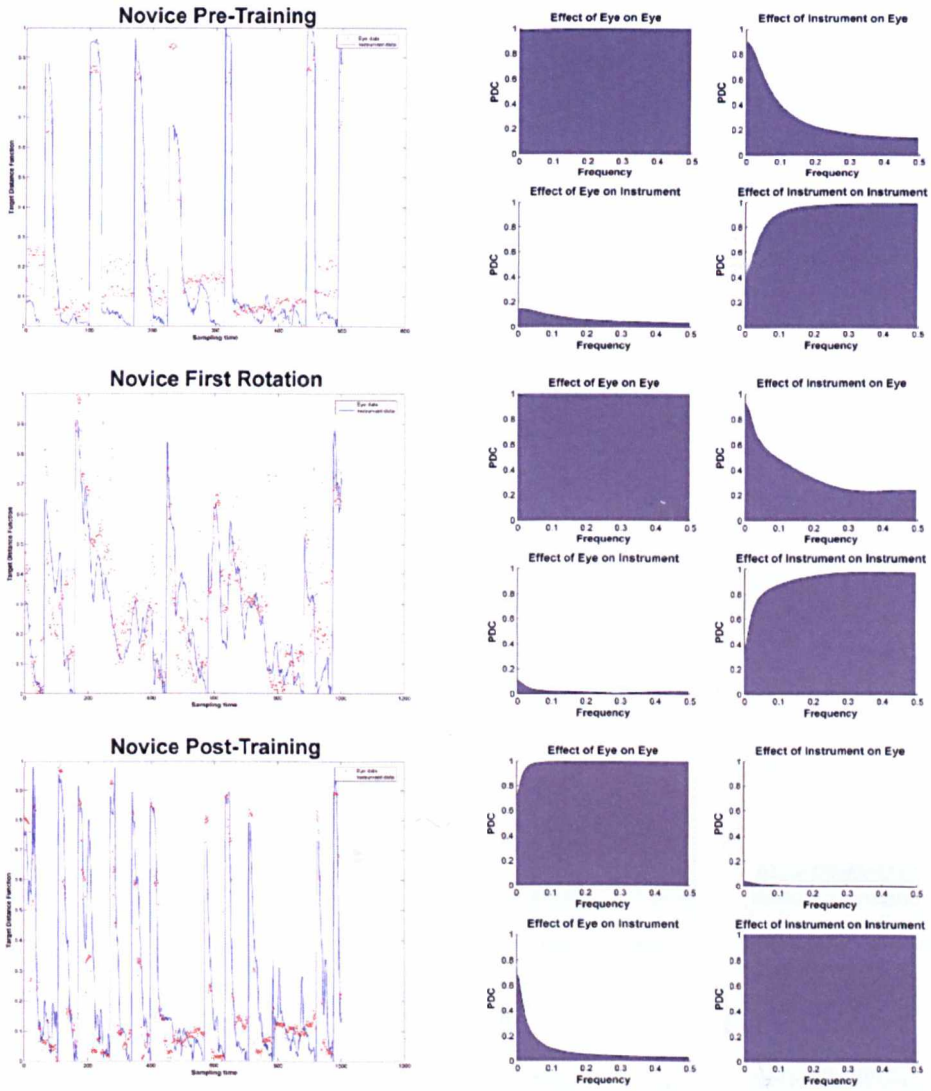

Figure 6.4 TDF of the instrument and eye data. Left column: example traces from **novice** pre-training (top), first rotation (middle), and post-training (bottom) data. Right column: the corresponding PDC plots showing the interdependence of the eye and instrument data.

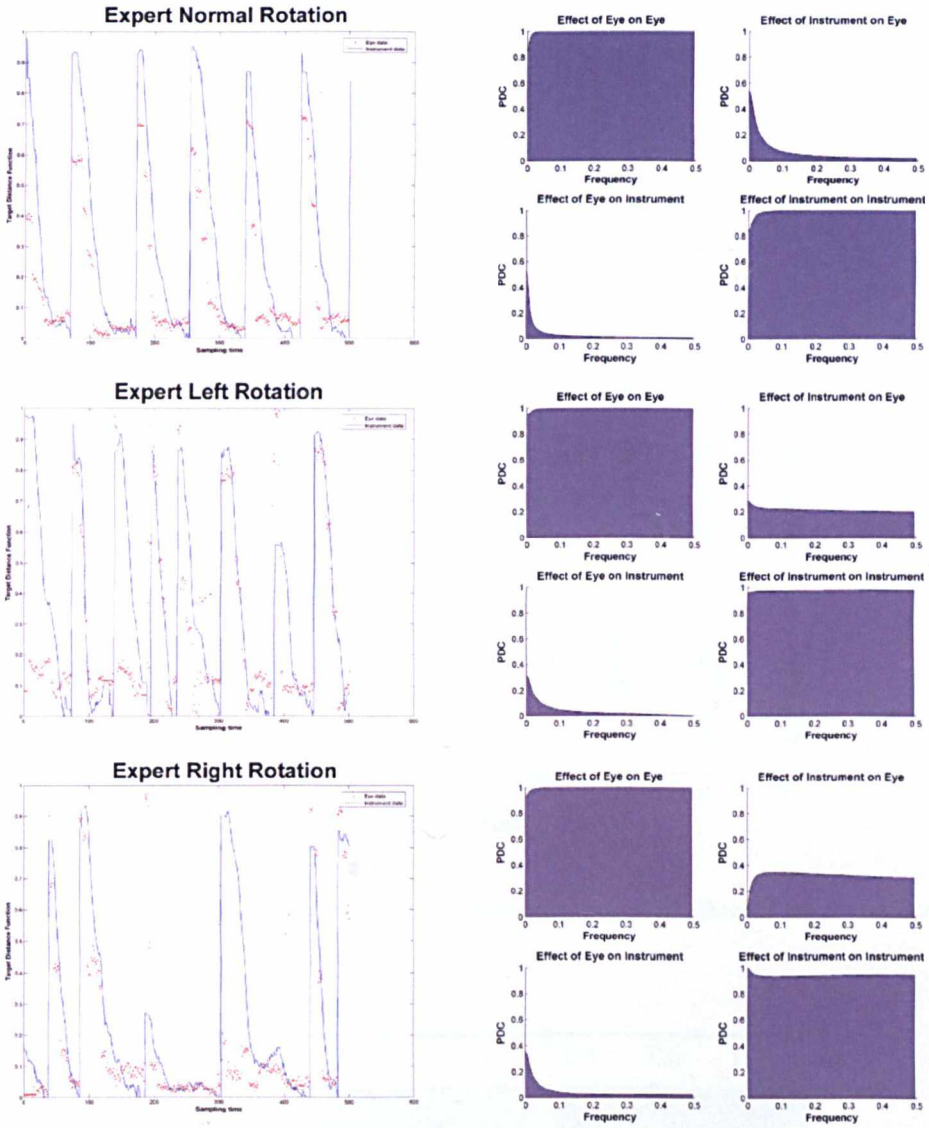

Figure 6.5 TDF of the instrument and eye data. Left column: examples from **expert** normal camera orientation (top), left rotation (middle), and right rotation (bottom) data. Right column: the corresponding PDC plots showing the interdependence of the eye and instrument data.

Figure 6.6 summarises the PDC analysis for all the subjects of the Target Location Experiment. In an attempt to quantify the change in PDC throughout the experiment, the PDC values examining the causal **influence of the instrument on the eyes** for all frequencies are summarised for all the tasks. Using non-parametric repeated measure tests, there is a significant **decrease** in PDC between Task T.1 and T.8 ($p < 0.001$) when compared within the subjects at the same frequencies. A corresponding **increase** in PDC from Task 1.1 to 1.8 was also observed, when the **effect of eyes on instrument motion** was examined ($p < 0.001$) as shown in Table 6.2. This shows that after training, the eye

movements are more likely to predict the instrument trajectories, whereas the causal influence of the instrument on the eye was stronger before training.

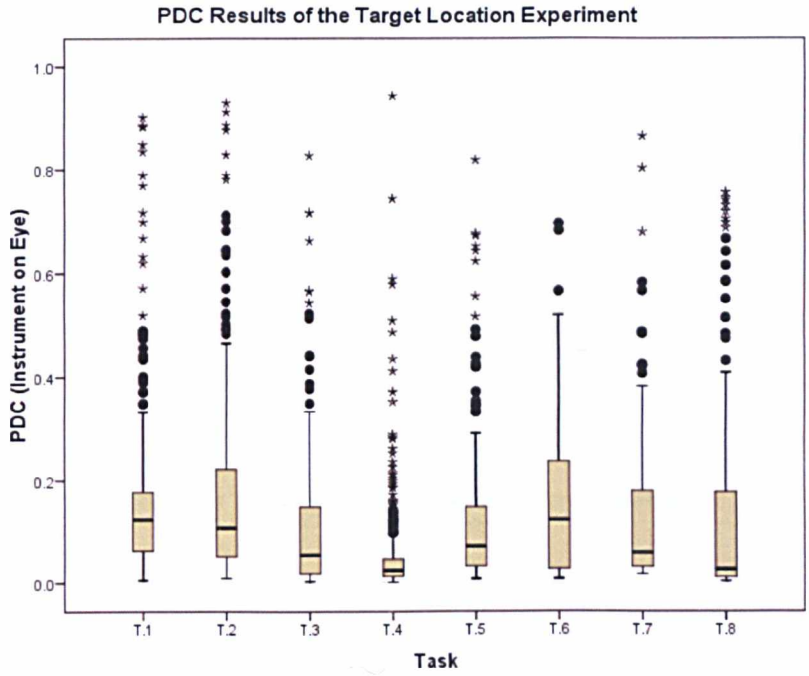

Figure 6.6 PDC Results of Target Location Experiment, where the effect of the instrument on eye is examined.

Task	T.1	T.2	T.3	T.4	T.5	T.6	T.7	T.8
Instrument on Eye *	123.6	107.7	54.7	24.6	70.5	122.4	58.6	26.5
Eye on Instrument *	37.6	29.8	71	35.4	33.8	23.7	23.7	46.5

* *Median PDC* $\times 10^{3}$

Table 6.2 Summary of PDC analysis for all the subjects in the Target Location Experiment. Significant difference between T.1 and T.8 using Wilcoxon signed rank test are highlighted in yellow. Spearman's rank correlation shows significant negative correlation from T.2 – T.4 and T.5 – T.7 are in different shades of green.

Moreover, there is a significant negative correlation between PDC values in Tasks T.2 to T.4 (r = -0.48, $p < 0.001$) and Tasks T.5 to T.7 (r = -0.15, $p < 0.001$) using Spearman's rank correlation, which represent the three successive training sessions with the first and second camera rotations respectively. This describes the decreasing causal influence of

the instrument on eye movement, as the subjects improved in one camera rotation condition and the other. However, the opposite correlation is not observed with the PDC measures of the eyes' influence on instrument movements, which is explained by the fact that the subjects were complete novices who required further training to reveal an increased eye predicting (feed forward) behavioural model.

6.3.2 PDC analysis for the Complex Dissection Experiment

Figure 6.7 summarizes the time performance improvement of the whole procedure (C.1 to C.5), showing a plateau effect at the 5th attempt, where there is no further statistical significant improvement in time. For individual task analysis, only Task C.3 showed no statistical improvement throughout the ten attempts.

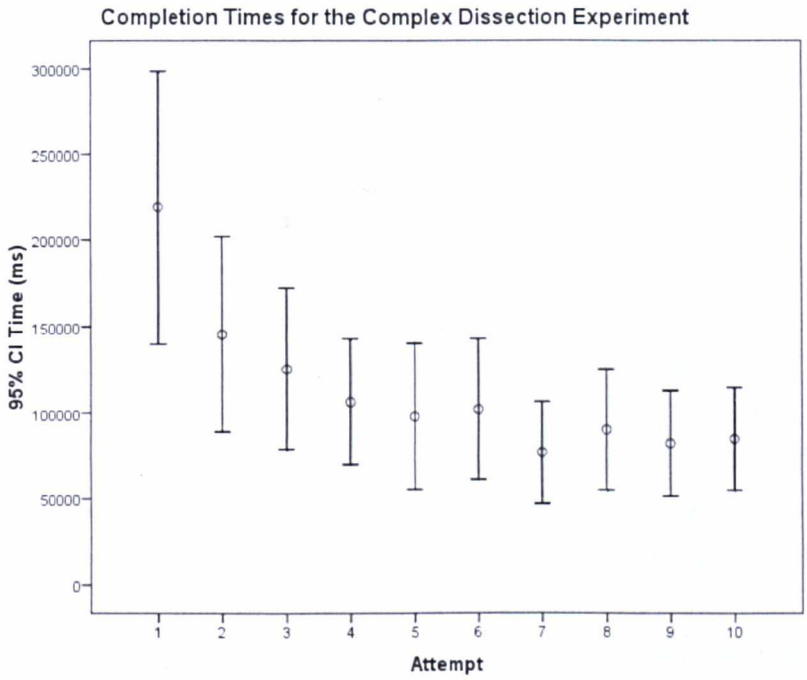

Figure 6.7 Completion times of the Complex Dissection Experiment across all 10 attempts. Time expressed in ms, with dots representing mean, and error bars representing 95% confidence interval.

Figure 6.8 illustrates the complexity of the data streams, displaying the whole procedure using distance to origin (0, 0) as a simple method to reduce the dimensionality of the data set. In general, the right and left instruments worked in synchrony around the areas of interest, with frequent separations of the instruments, for example at sample 3000, 4000

and 5000. There were large amplitude movements, especially in the left hand, whereas the right hand seemed to concentrate mainly in the middle part of the surgical scene. The eye fixations (red dots) alternated between the right and left instrument throughout the procedure, exhibiting a tightly coupled behaviour with them.

In Task C.1, the subjects were instructed to dissect the skin flap from the right side of the screen to the left, this was performed with the intention to produce similar pseudo-sinusoidal data as in the Target Location Experiment, for simpler data analysis. This was achieved with partial success; however, with the introduction of this bimanual task, the eye movement patterns seemed to alternate between the right and left instrument frequently. The left instrument was used to lift the skin flap, whilst the right instrument started dissecting. At sample 3000, 4000, and 5000, it clearly shows that visual input was important for the left hand to grasp the skin flap.

In Task C.2, the subjects used mainly the right instrument to dissect out the simulated tissue from the muscle layer. Here, the eye fixations and the right instrument worked closely together, whilst little attention was paid to the left instrument.

Task C.3 was a simple task requiring the left instrument to grasp the simulated tissue, and remove it from the surgical scene and drop it in an untargeted area at the bottom of the laparoscopic training box. It showed that little visual cue was needed, as minimal coordination was required, and hence a primarily feed forward ballistic movement.

Task C.4 showed large movements with the instruments and less organised eye fixations. This task was to replace the skin flap, and observation of the video data confirmed that no systematic way was used to complete this task, and the data was difficult to interpret.

Task C.5 was to place both instruments to the start positions, and interestingly, eye movements followed the early part of the right instrument and the latter part of the left, implying that more confidence in the control of the right, as the subject was right-handed.

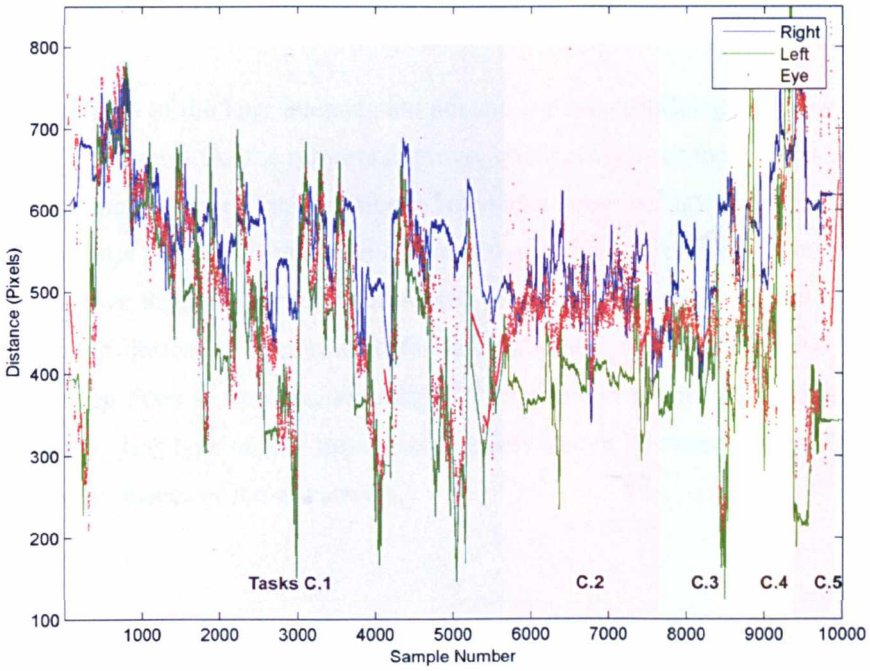

Figure 6.8 The complete procedure for the Complex Dissection Experiment, blue and green lines representing the right and left instrument tip, and red dots shows the eye fixation. All data expressed as distance from the origin (0,0). Each task was shaded with a different colour as labelled.

Figure 6.9(a) illustrates part of the data streams, with TDF of the eye and **left** instrument data plotted against time for Task C.1 for the first attempt. The highlighted windows in red show areas where the instrument movements precede eye movements. The corresponding PDC analysis is shown where high PDC values are expressed in light colours, this is consistent with higher influence of the instrument on eye movements. Screenshot examples of two of the highlighted areas are shown in the bottom two rows, where the left instrument is shown to lead the saccadic eye movements.

Figure 6.9(b) shows the same TDF and PDC analysis of the effect of the **right** instrument on the eye movements. Here, the blue highlighted windows correspond to low PDC values, and are represented by darker colours on the PDC graph in almost all the frequencies. Screenshot examples of the two highlighted areas are shown in the bottom two rows, representing one high (top) and one low (bottom) PDC window. The screenshots of the low PDC window shows that the visual system was used to guide the left instrument, whilst the right instrument remained stationary. This substantiates the

reason for the low causal influence of the right instrument on the eye movements during this period, and the corresponding low PDC values.

The PDC analysis of the later attempts did not show a causal influence of the eye on the instrument movements. As the subjects improve, visual analysis of the video data showed a significant improvement of predictive eye behaviour in the 10^{th} attempt when compared to the 1^{st} attempt. However, this type of predictive behaviour is non-sequential: as the subjects improve in this complex laparoscopic procedure, the visual system is used for target selection, instead of simple feed forward guidance of the instruments. Fixations often landed on three to four potential targets, before the instrument was utilised on the selected target. This type of non-sequential causality cannot be measured by PDC due to the cognitive influence of the manoeuvre.

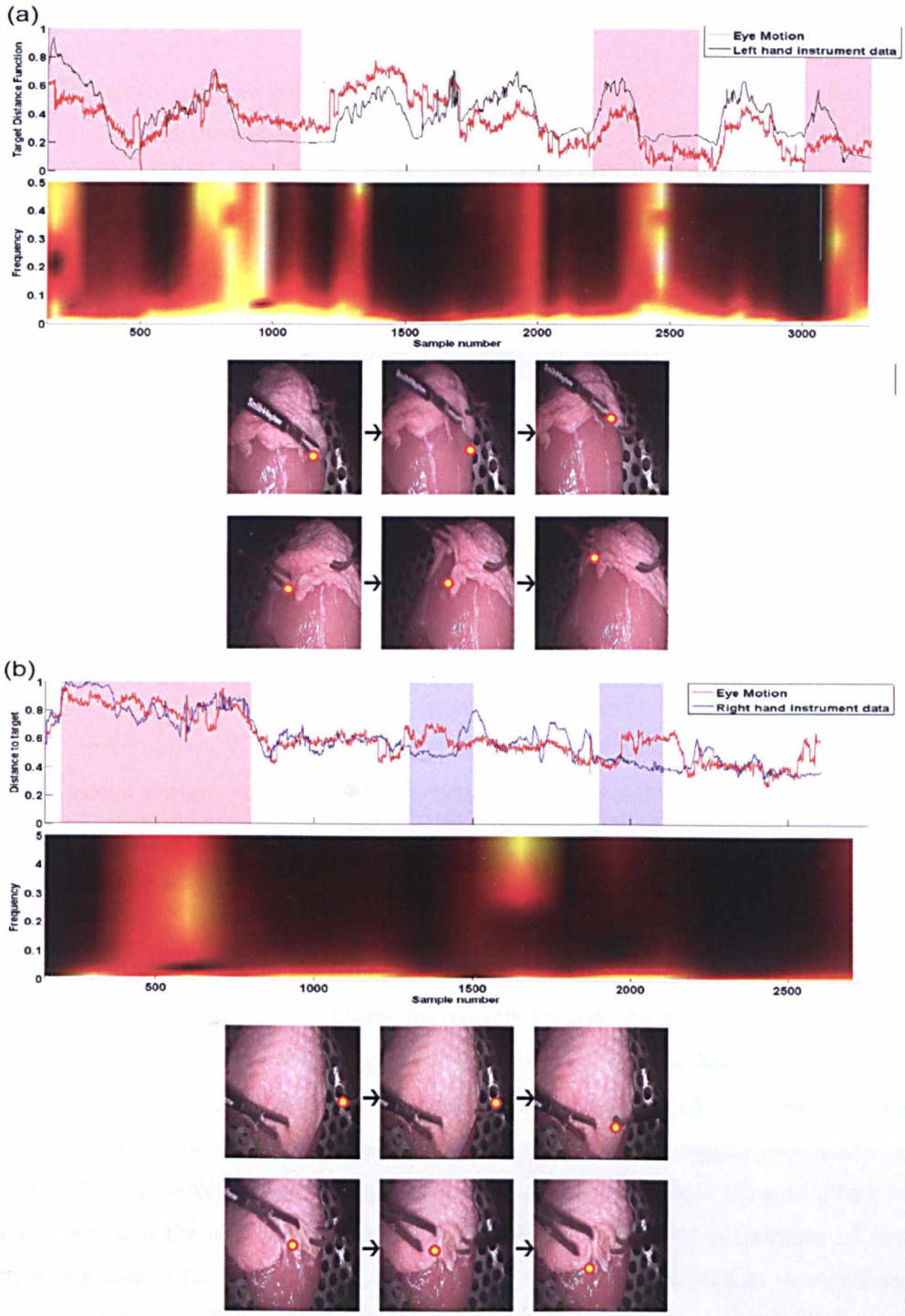

Figure 6.9 (a) TDF of eye and **left** instrument data against time (expressed as sample number) for a subject performing Task C.1, and moving window PDC analysis (300 time samples with 50% overlap) of the causal influence of the left instrument on eye movement, the high PDC values of the specific frequencies are expressed in light colours, these areas are highlighted in red. Screen shots of two high PDC windows are illustrated in the bottom two rows (red/yellow dot = gaze position). (b) TDF and moving window PDC of the causal influence of the **right** instrument on eye movement. Blue highlighted areas represent low PDC levels. Screen shots of one high and one low (bottom) PDC windows are illustrated.

6.4 Discussion and Conclusions

This chapter investigates the interdependency of the hand and eye movements along with the variability of their temporal relationships in laparoscopic surgery. PDC reveals the subtle effects of improvement in hand-eye coordination, where the causal relationship between instrument and eye movements gradually reverses during simple laparoscopic tasks in the Target Location Experiment. This method of analysis has been used to describe time series data in economics and its forecast based on the concept of Granger-causality. PDC is the frequency domain description of Granger-causality, and the data presented with the Target Location Experiment is ideal for this type of analysis. Low frequency data represents the fluctuation of the time series data between points A and B, whilst high frequency jitters in the data correspond to random local variations of instrument and eye movements.

In contrast to previous analysis, PDC provided an automatic algorithm for signal matching (coherence) and temporal delay detection, this required little data pre-processing. Furthermore, the data set was analysed as a whole, in contrast to Chapters 4 and 5 where each trajectory was analysed individually, allowing for less constraints in experimental designs and artificial data markings. Importantly, the information of the target becomes less relevant for PDC to be implemented, as the algorithm matches the shape and trajectory of the data series.

In the laparoscopic task in the Complex Dissection Experiment, PDC also successfully identified areas where the instrument movements precede the eye movements in early skills acquisition of a complex procedure. However, in this bimanual task, both instruments often moved in synchrony, and visual input was used to guide each hand alternatively as necessary. By definition, at least half of the instrument data would not match the eye movement data, which generated significant random noise to affect the robustness and the accuracy of the algorithm. Due to these large proportions of non-matching data, it became apparent that very high frequency data, such as miniature eye movement, tremor of the instrument tips and error generated by the tracking equipment, were classified as matching signals by the algorithm.

As the novices improved in this complex task, the visual behaviour also changed to an intricate non-sequential predictive relationship with the instruments, which is difficult to be measured using simple PDC analysis. Expert surgeons often visualise multiple potential targets prior to definitive instrument movement generation, these eye

movements may be related to target selection and attention to potential areas of hazard. Further investigation is warranted for measuring this complex combinatory behavioural pattern by incorporating high-level cognitive inferencing data.

A more robust signal matching algorithm needs to be developed to improve the signal-to-noise ratio for this analysis. As in Appendix A, eye movements can be modelled into probability distributions, where search patterns are compared mathematically. The theory of relative entropy measures the statistical distance between two probability distributions, and the lower distance represents higher similarity. This type of analysis may be useful for the hand and eye datasets.

The next chapter will concentrate on improving the signal matching technique, using principles based on relative entropy to analyse this complex laparoscopic procedure. Application results of this technique will be presented, as it is necessary to improve its performance for the hand-eye data. Furthermore, improvements on the experimental design and hardware setup will also be discussed, along with data from expert surgeons to provide validation.

Chapter 7

Multiscale Jensen-Shannon distance to measure the hand-eye relationship in Laparoscopic surgery

7.1 Introduction

The spatiotemporal relationship between the hands and the eyes has been implicated in measuring the mental ability to perform visuospatial transformation. In performing a novel visuomotor task, reactive eye movements provide feedback mechanisms to correct motor errors. However, during the learning of this task, a transition to predictive eye movements has been observed. In this scenario, eye saccades tend to move ahead of the hand and land on the target without further referencing of the hand position. Previous measurements of this relationship have relied on qualitative observations, as the complexity of the data deluded its precise measurement (72,73). In the last chapter, Partial Directed Coherence (PDC) was used to measure the spatiotemporal interdependency between the instrument tips and eye motions during two laparoscopic experiments. This technique was initially successful in identifying areas of casual relationship between them. However, as the movement trajectories became more complex with large areas of non-matching signals, PDC analysis appeared more ambiguous where high frequency noise interfered with the results. Furthermore, the numerical results of PDC provided an arbitrary implication of causality, as meaningful interpretation remained limited whilst no significance levels had been predefined (158).

Camera rotation has been shown to deteriorate surgical performance previously, although training substantiates adaptation to the new rotated environment, and performance gradually returns to previous levels. The amount of training needed for this adaptation is unknown, and is likely to vary between individuals. It is felt that the experimental

protocol and hardware setup was suboptimal in the last chapter, and multiple variables needed to be controlled to ensure higher quality analysis. The main modifications in experimental designs are highlighted below:

In the Two Target Experiment (previously Target Location Experiment):

1. A more robust experimental protocol is designed, with the effect of learning compared to the effect of screen rotation further examined.
2. More natural surgical environment is used, where no signal is given to the subjects on completion of each trajectory.
3. Using a time-based block designed experiment, instead of task-based, where subjects are required to perform and repeat a task within the same time constraints, rather than performing the same number of repetitions without the constraints of time (which can vary dramatically between individuals, and may affect the amount of learning achieved between them).
4. Longer training periods are used in an attempt to achieve learning plateau, however, this is limited by factors such as boredom and fatigue.
5. The instrument tracking setup is standardised, where the upgraded hardware is used. This allows for more natural movements of the subjects' hands, as the tracking volume is increased and almost complete rotation of the instruments is allowed without any data loss.
6. Identical hardware setup is used for both experiments.

In the Complex Dissection Experiment:

1. Only Task C.1 was analysed in the last chapter, this was due to the complexity of the data. In this particular task, the data resembled the Target Location Experiment naturally when the instruments moved systematically across the screen.
2. The whole procedure is analysed in its totality in this chapter. The proposed method of analysis seems to be more versatile and extra steps are taken into account to select suitable signals automatically
3. The introduction of laparoscopic expert data to provide validation of the measurement methods.

The purpose of this chapter is to use a novel analysis framework robust enough to analyse both experiments automatically. The ultimate goal is to find the spatiotemporal relationship between the eyes and the instrument tips in any laparoscopic procedure without artificial experimental constraints. In the last chapter, the results were biased due to the performance of PDC, where causality relationships were found even in sections of poor eye-instrument signal matching

After detailed observations of the data in both experiments, it appears that in the Two Target Experiment, a less demanding matching algorithm is required, as the patterns of the two signals resemble sinusoidal waves of similar frequencies. However in the Complex Dissection Experiment, the eye and instrument signals tend to match in different frequencies. There is a general shape matching in the low frequency domain, where the eye and instrument data largely follow each other during the whole experiment. However, the more relevant parts of the signal lie in the higher frequency domain, where targets are selected by a brief visual search and followed by actions of the instruments, these areas are usually around 400 – 2000ms. In the signal windows between 10ms and 400ms, microsaccades of the eyes and tremor of the instruments become predominant in the data streams, where they are regarded as non-purposeful movements.

A different approach is used in this chapter, where the analysis is divided into two passes: first, the signals are modelled as probability distributions, and signal matching is achieved by measuring their statistical distances using principles based on the Kullback-Leibler (KL) distance (161) (see Appendix A); second, the temporal relationship is calculated between the instrument tip and the eye position in these matching areas.

7.2 Methods

7.2.1 Hardware setup

This was identical in both experiments. The Optotrak Certus system (Northern Digital Inc, Ontario, Canada) described in the earlier chapters was used to track both laparoscopic instruments and the endoscopic camera, as shown in Figure 7.1. The use of active IRED tracking device rigidly attached to the instruments allowed for simultaneous tracking of multiple objects. The offsets of the instrument tips from the IRED markers were calculated using the Pivot function of the NDI software. Data interfacing was achieved through RS-232 and the provided tracking accuracy was 0.15mm RMS at a

sampling rate of 50 Hz. As before, a Tobii 1750 eye tracker (Tobii Technology, Stockholm, Sweden) was used to display the laparoscopic scene whilst tracking the binocular eye movements unobtrusively. The updated version of the eye tracker provided an accuracy of 0.5 degrees with a sampling rate of 50 Hz.

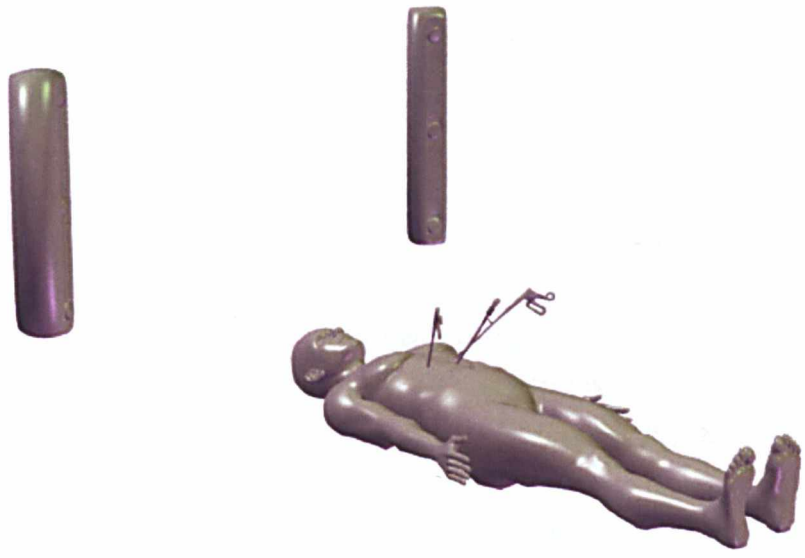

Figure 7.1 Illustrating the experimental setup involving two Optotrak Certus tracking systems for providing better line of sight and rotation coverage.

7.2.2 3D to 2D transformation

As the IRED markers were rigidly attached to the laparoscopic camera, once the spatial relationship between them was obtained, the extrinsic parameter of the camera can be continuously updated whilst the camera moved. A bespoke modification of the Matlab camera toolbox (128) was used to calibrate the camera, where the intrinsic and extrinsic parameters were obtained using an algorithm based on the Tsai and Lenz method (129,130,162). With this setup, a stationary calibration grid was used, and the laparoscopic camera/IRED marker complex was moved to 10 positions capturing images of the calibration grid, this is illustrated in Figure 3.13. This provided the estimation of the unknown transformation from the Optotrak Certus coordinate system to the calibration grid coordinate system, as well as the transformation from the camera to the IRED marker coordinate system. The output of the algorithm was the rotation and translation (homogenous transformation) between the IRED marker and the camera.

As the laparoscopic camera position was updated every 20ms in synchrony with the instrument positions, 3D to 2D transformation was calculated for each data point using this information with the known intrinsic parameter of the camera.

7.2.3 Experimental Setup

7.2.3.1 Two Target Experiment

14 right-handed subjects (mean age ± SD = 20.77 ± 1.15 years) participated. All subjects were complete novices in laparoscopic surgery, with only an initial introductory session to familiarise the subjects with the instruments used. They were instructed to locate two standardised points, repeatedly and accurately, on a simulated plastic small bowel model using only their right hand holding a laparoscopic grasper. A block design experiment was conducted where motor tasks were separated with blocks of motor rest. In Tasks NR1, NR3 and NR5 (recording tasks), after an initial period of motor rest (30s), 3 motor task blocks (20s) were interspersed with 2 rest blocks (20s). In Tasks NR2 and NR4 (training tasks), after the same initial period of rest, the 6 motor task blocks (20s) were separated by 5 rest blocks (3s). The length of the tasks were chosen after pilot data suggesting learning plateau occurred around the 4^{th} and 5^{th} session, effects of fatigue and boredom became apparent afterwards. The subjects were then randomised to either left rotation (90 degrees counter clockwise camera rotation) or right rotation (90 degrees clockwise camera rotation), using the sealed envelope method. The left rotation group repeated the 5 tasks with left rotation first, then right rotation, and *vice versa*. A final experiment was run with normal camera rotation as a recording task. Table 7.1 summarises the experimental protocol for the two groups.

	Left rotation group	**Right rotation group**
NR1 – NR5	Pre-training tasks 1 – 5	Pre-training tasks 1 – 5
LR1 – LR5	Left rotation tasks 1 – 5	Left rotation tasks 1 – 5
	(First rotation, FR 1 – 5)	*(Second rotation, SR 1 – 5)*
RR1 – RR5	Right rotation tasks 1 – 5	Right rotation tasks 1 – 5
	(Second rotation, SR 1 – 5)	*(First rotation, FR 1 – 5)*
NR6	Post-training task	Post-training task

Table 7.1 Summarises the Two Target Experiment protocol. For the First Rotation (FR) tasks in the left rotation group represents the Left rotation, and Right rotation in the right rotation group. The opposite applies for the Second Rotation (SR) tasks.

The design of this experiment enables the analysis of three effects:

1. the effect of learning on the hand-eye coordination measurement.
2. by analysing the difference between left and right rotation tasks (which were randomised as 1st or 2nd rotation blocks), the effect of 90 degrees of camera rotation is illustrated, whilst the difficulty levels and amount of rotation are assumed to be the same.
3. by analysing the difference between 1st rotation and 2nd rotation tasks (which were randomised as left or right rotation blocks), the effects of 90 degrees (in the 1st block) and 180 degrees of rotation (in the 2nd block) can be compared, assuming that there is no difference between left and right rotations above. This is summarised in Table 7.2.

	Normal	**Left rotation**	**Right rotation**
Difficulty	N	+	+
Rotation	N	+	+

	Normal	**1st Rotation**	**2nd Rotation**
Difficulty	N	+	+
Rotation	N	+	++

Table 7.2 Summarising the effects on task difficulty and coordinate transformation by analysis based on the direction of the rotation and the order of the rotation. N = normal.

7.2.3.2 Complex Dissection Experiment

In addition to the eight complete novices, 5 laparoscopic experts (with performance of over 100 laparoscopic operations) were recruited to perform a more complex bimanual laparoscopic procedure. A cadaveric avian model was used to simulate dissection and manipulation skills in laparoscopic surgery. The procedure was broken down into tasks: (C.1) dissecting the subcutaneous connective tissue over the pectoralis muscle, (C.2) dissecting the muscle to reveal a simulated tumour tissue, (C.3) removal of the simulated tissue, (C.4) repositioning the dissected skin layers, and (C.5) returning the instruments to the start positions. Each novice completed the procedure ten times in 3 separate sessions, and each expert completed 5 repetitions. Table 6.1 summarises the task breakdown of the experiment.

7.2.4 Analysis Method

7.2.4.1 Pre-processing

The distances from the eye and instrument screen coordinate data to a fixed point (0, 0) on the screen were calculated, resulting in an eye signal $E(t)$ and instrument signals $L(t)$ and $R(t)$, for the left and right instruments respectively. This Target Distance Function (TDF) was calculated similar to the last chapter, and thus reducing the dimensionality of the data.

7.2.4.2 Problem statement

The aim of the proposed algorithm is to locate matching hand-eye data with a temporal delay, different methods have been tested. Linear correlation would provide the simplest methodological solution, however it did not account for temporal shifts. Here, an extension of the Jensen-Shannon Divergence (JSD) method (163-165) is proposed. JSD is a standard metric for matching probability distributions, in order to apply this to the experimental data in different frequency domains, a multiscale modification has been developed. The main purposes of the proposed method are:

1. Locate matching areas between the time series.
2. Segment the matching areas by deducing the optimal time windows (signal lengths) that provide areas of maximum match, given the temporal delay between these signals.
3. Calculate the actual temporal delay.

7.2.4.3 The Jensen-Shannon Divergence

For probability distributions P and Q of a discrete random variable, the KL distance (161) of Q from P is defined as:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

For distributions of a continuous nature, the summations give way to integrals:

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx$$

where p and q are the densities of P and Q respectively.

JSD can be considered a natural extension of the KL distance to a set of distributions, where the JSD of a set of distributions is the average KL distance to the mean of the set (166). Unlike KL distance, JSD is a bounded metric. Consider the set $M_+^1(A)$ of probability distributions where A is a set provided with some $\sigma -$ algebra. The Jensen-Shannon divergence, $JSD : M_+^1(A) \times M_+^1(A) \rightarrow [0,1]$ is defined by:

$$JSD(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel Q) + \frac{1}{2} D_{KL}(Q \parallel M)$$

where $M = \frac{1}{2}(P + Q)$, the JSD can be thought of a square of a metric, as the square root of JSD fulfils the triangle inequality (167), thus it can be used as a metric for signal matching. The $JSD(X,Y)$ between 2 signals X and Y uses the distributions of signal values, so P and Q in the above formula would be the probability distributions of X and Y respectively. Figure 7.2 illustrates signal examples of high JSD (poor matching) and low JSD values (good matching).

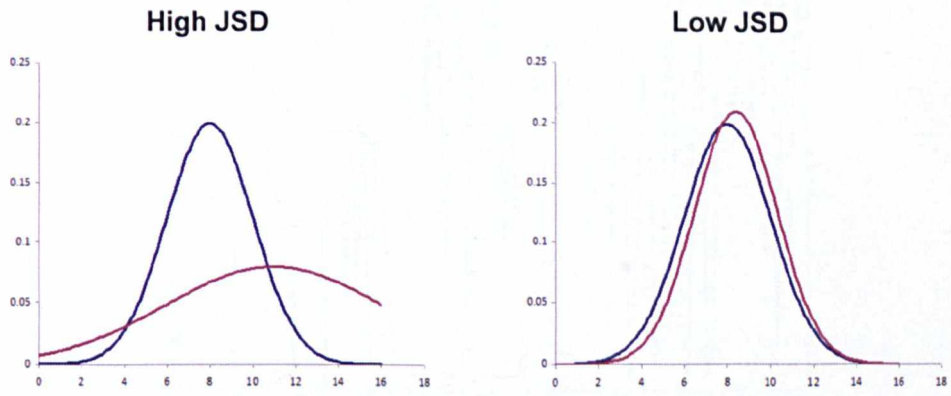

Figure 7.2 Signal examples illustrating poor signal matching (left) with a high JSD value, and good signal matching (right) with a low JSD value.

7.2.4.4 Jensen-Shannon Divergence Modifications

However, as JSD compares distributions but not time series data; signals as demonstrated in Figure 7.3, with similar value distributions but a completely different temporal order, would result in a low JSD value (good matching). To overcome this, modifications were made to incorporate linear correlation into the algorithm, namely the correlation coefficient *r*, which for 2 signals X and Y can be defined as:

$$r(X,Y) = \frac{\sum_i (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum_i (X - \bar{X})^2 (Y - \bar{Y})^2}}$$

The correlation coefficient *r*, ranges between -1 and +1, with high values indicating correlating signals. Thus, this new measure of signal similarity $JSDr$, for signals X and Y can be defined as:

$$JSDr(X,Y) = \frac{1}{2}(JSD(X,Y) - |r(X,Y)|)$$

This metric weighs against non-correlating signals and aims at finding signals that both correlate and consist of matching distributions.

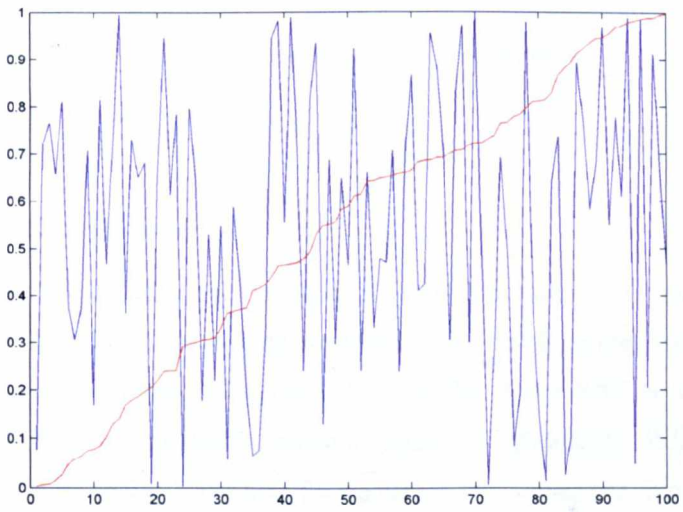

Figure 7.3 Example illustrating signals (blue and red) with exactly the same values, but in completely different temporal order. JSD analysis would indicate high probability distribution matching (low JSD value), but with a low correlation coefficient, *r*.

7.2.4.5 Multiscale *JSDr* to locate matching windows

The multiscale approach was used as a solution to overcome the frequency problem as described in Section 7.1, as the length of the matching trajectories was variable throughout the experiments, depending on the nature of the targets selected for each movement episode. In this method, $JSDr$ was calculated in segments with incremental windows sizes from 400ms to 2000ms between the two signals, hence called the multiscale approach. As highlighted above, typical movement trajectories were observed to last between these time scales.

In order to locate the optimally matching window size (signal length), with the lowest $JSDr$, a moving window algorithm was used across the two signals whilst computing the multiscale $JSDr$ (middle subplot of Figure 7.5). Thus for 2 signals X and Y, the $JSDr$ at time t for a window size w would be:

$$JSDr(t,w) = JSDr(X[t-w:t+w], Y[t-w:t+w])$$

The multiscale values of $JSDr(t,w)$ offer a method of observing signal similarities at different scales, as well as at different times. Low values of $JSDr(t,w)$ indicate areas of high matching between the 2 signals. In order to find areas with the closest match, the local minima of $\sum_{w} JSDr(t,w)$ were located by using a hill climbing algorithm (bottom subplot of Figure 7.5). For each of these minima t^*, the window size corresponding to the lowest value was found (w^*).

7.2.4.6 Simulated analysis

Figures 7.4 and 7.5 illustrate the analysis of simulated time series data using this algorithm. Two signals with 4 matching areas and 1 non-matching area were created for this analysis, Gaussian white noise was added in the background to simulate high frequency errors of the system, this is shown in Figure 7.4. Multiscale $JSDr$ is shown in the *middle* subplot of Figure 7.5, where four areas of low values are illustrated in dark blue at varying window sizes. The averaged $JSDr$ is plotted in blue in the *bottom* subplot, and local minima are highlighted with red stars. These local minima

corresponded exactly with the 4 areas of matching signals, as shaded in blue in the *top* subplot, illustrating the performance of the signal-matching algorithm.

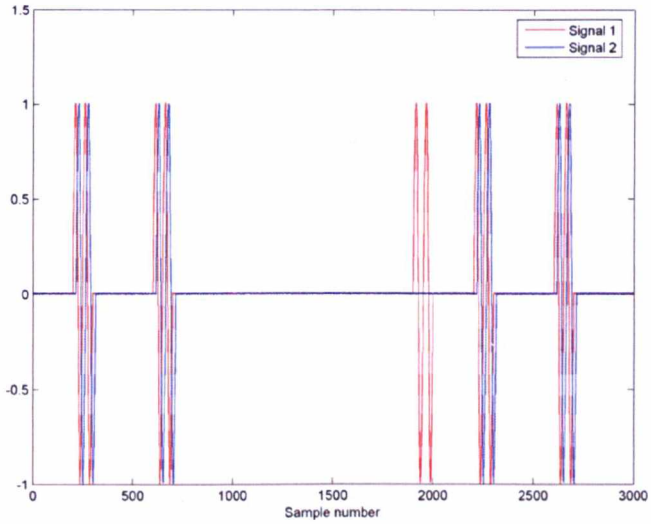

Figure 7.4 Shows two simulated signals containing four matching areas with a time-shift. Background Gaussian white noise is added simulating high frequency error signals.

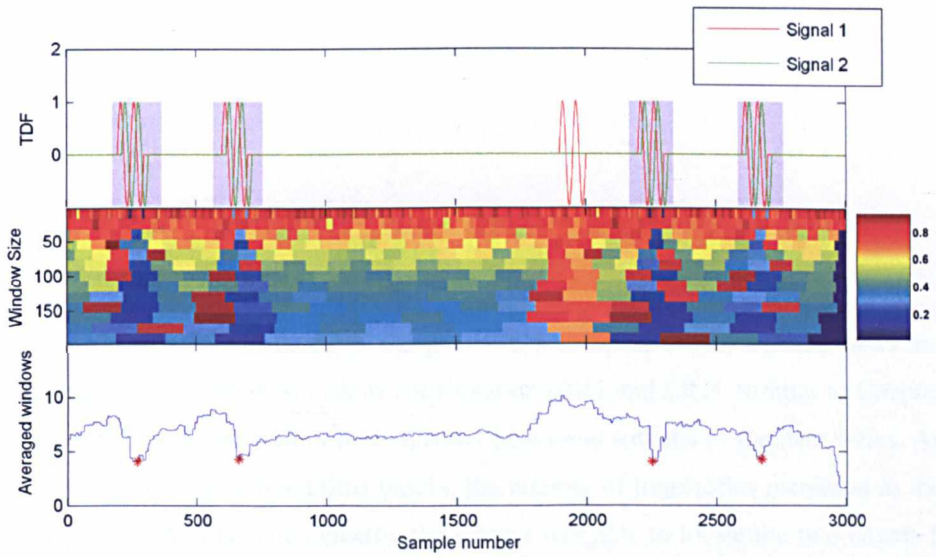

Figure 7.5 Example traces illustrating the *JSDr* algorithm. The 2 original signals (top), the corresponding multiscale *JSDr* (middle) and the average of the *JSDr* and locations of the minima (bottom). For each of these minima (red stars), the window size with the lowest value was selected, and the matching areas with the windows are highlighted in blue (top).

7.2.4.7 Temporal delay

The time delays between two signals were calculated within the output signal windows from the above algorithm. As each window represented two matching trajectories, the common points between these signals were found by locating the lowest maximum and the highest minimum between the two signals. A line was drawn at the midpoint and its intersection with each signal was calculated. The length of this segment would indicate the time delay between the signals as shown in Figure 7.6.

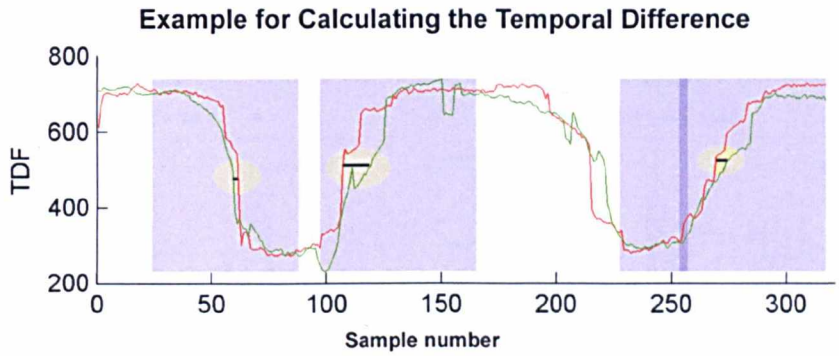

Figure 7.6 Temporal differences are calculated by the difference in value in the *x*-axis between the two signals. These are black lines highlighted in yellow.

7.3 Results

7.3.1 Two Target Experiment

7.3.1.1 Qualitative analysis

The raw data is shown in Figure 7.7 the pre (NR1) and post (NR6) training tasks are illustrated together with the first task of each rotation (RR1 and LR1). Similar to Chapter 5, the effect of training resulted in more distinct peaks and troughs of the data series. As the study was performed in fixed time blocks, the number of trajectories increased as the subjects improved. As illustrated clearly, the subject was able to locate the two targets 5 times each in the post training task, compared to 3 in the pre training task.

Camera rotation produced a similar effect as in Chapter 5, when the movement trajectories and eye movements became severely disrupted. As a consistent observation throughout the data collection period, right rotation tasks created more coordinate

transformation difficulties for these right-handed subjects, and this is also demonstrated in Figure 7.7.

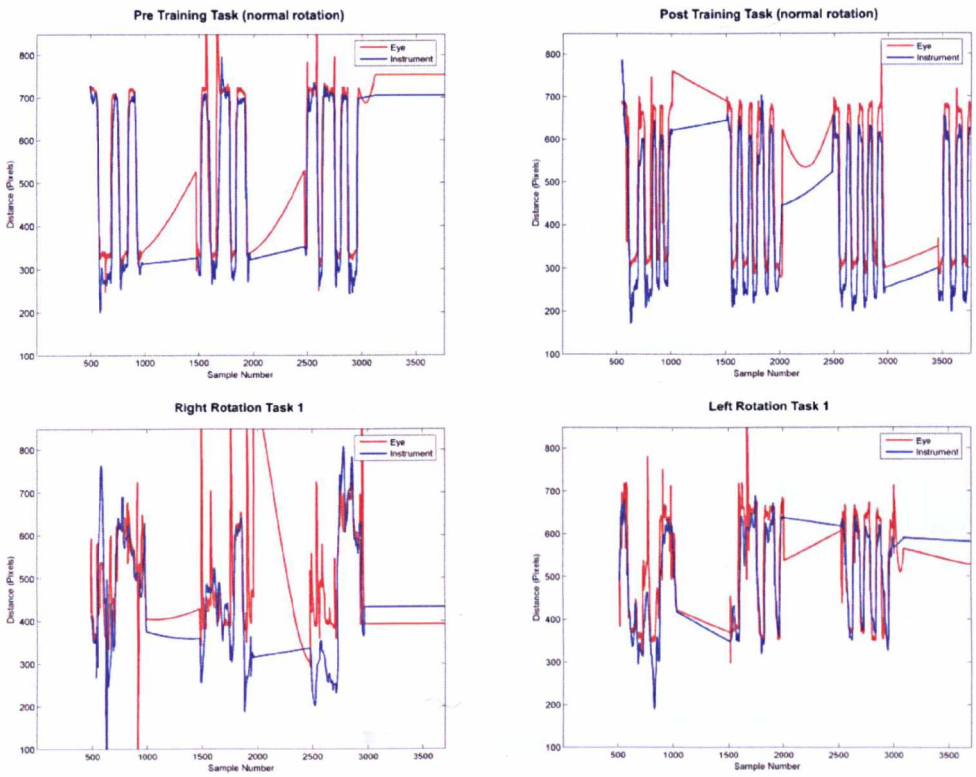

Figure 7.7 Illustrating the raw data plotting the distance from origin against sample number (time). Top left – Pre training task (NR1), top right – Post training task (NR6), bottom left – first right rotation task (RR1), and bottom right – first left rotation task (LR1). Red line – eye position, blue line – instrument position.

7.3.1.2 Multiscale JSDr

Figures 7.8 and 7.9 illustrate the performance of multiscale $JSDr$ in signal matching. Similar to Figure 7.5 but using actual experimental data, areas of matching signals are shaded in blue in the top subplots. This is achieved by finding the local minima of the averaged multiscale $JSDr$ as illustrated in the bottom two subplots. In Figure 7.8, NR1 and NR6 representing the pre- and post training tasks with normal camera orientation, the instrument and eye movements correspond visually throughout the example. However, areas of matching were located with high specificity, whilst the algorithm missed out other areas where the signals also appeared to be similar. This lowered sensitivity of the algorithm was accepted to improve the signal-to-noise ratio, which was problematic with PDC in the last chapter. In Figure 7.9, where left (LR1) and right (RR1) rotation tasks are examined, right rotation task again appears to be more detrimental to the task

performance by the subjects, generating more erroneous movement trajectories. The algorithm successfully located 3 matching areas even with these difficult signals, primarily due to the lower frequency matching signals. The middle subplots show the JSDr values in different window sizes, and the lower values (green or blue) indicating better signal matching, are concentrated in the longer signals (window size 40 – 60 samples), or lower frequency data.

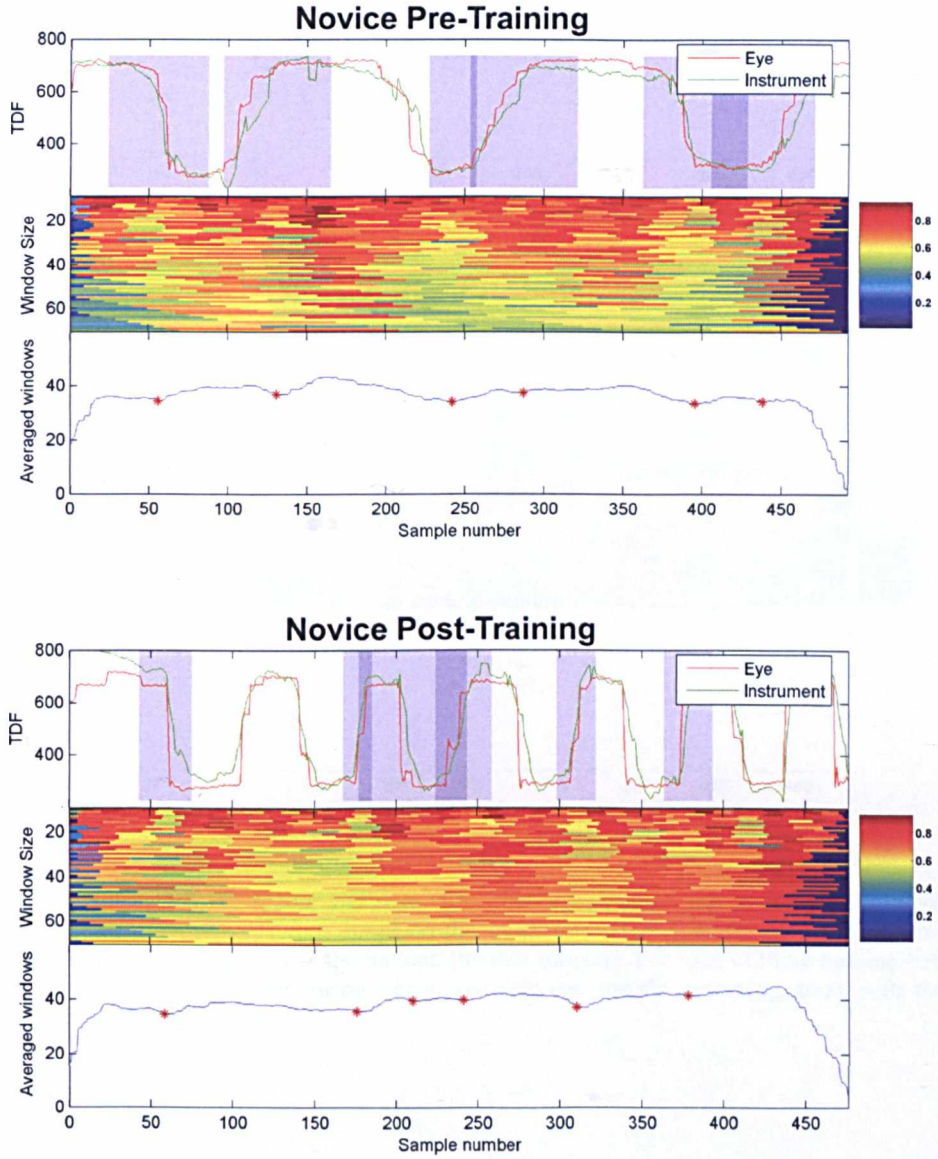

Figure 7.8 Novice pre- (NR1, top) and post- (NR6, bottom) training data showing TDF of the eye and instrument movements (top subplot), the multiscale *JSDr* (middle subplot), and the average of the *JSDr* and locations of the minima (bottom subplot). For each of these minima (red stars), the window size with the lowest value was selected, and the matching areas with the windows are highlighted in blue.

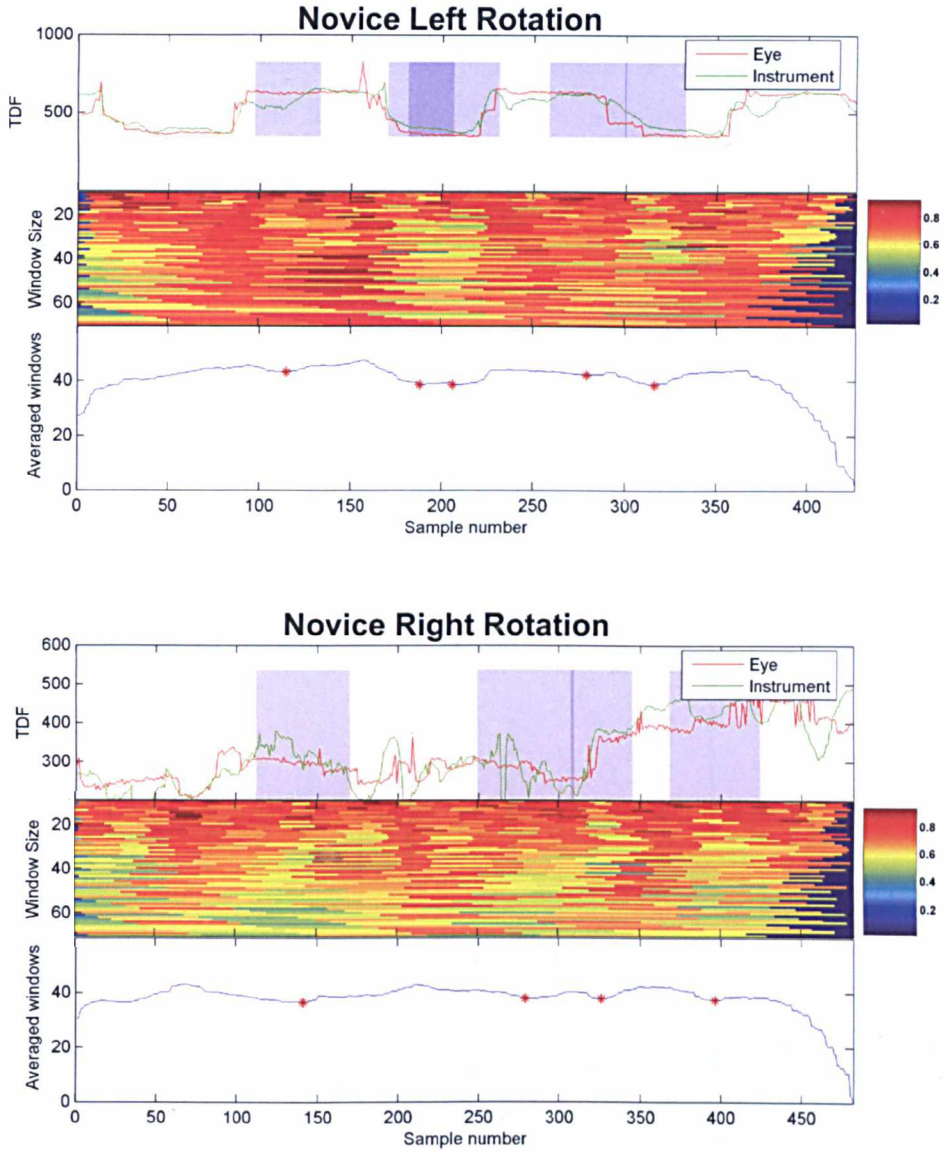

Figure 7.9 Novice left (LR1, top) and right (RR1, bottom) rotation data showing TDF of the eye and instrument movements (top subplot), the multiscale *JSDr* (middle subplot), and the average of the *JSDr* and locations of the minima (bottom subplot). For each of these minima (red stars), the window size with the lowest value was selected and the matching areas with the windows are highlighted in blue.

7.3.1.3 Examination by rotation

Figure 7.10 illustrates the temporal relationships between the instrument and the eyes, after matching windows were found using the $JSDr$ algorithm. Only positive values were included, which examined the amount of delay of the eye movement from the instrument tip. The analysis is based on the **direction** of camera rotation.

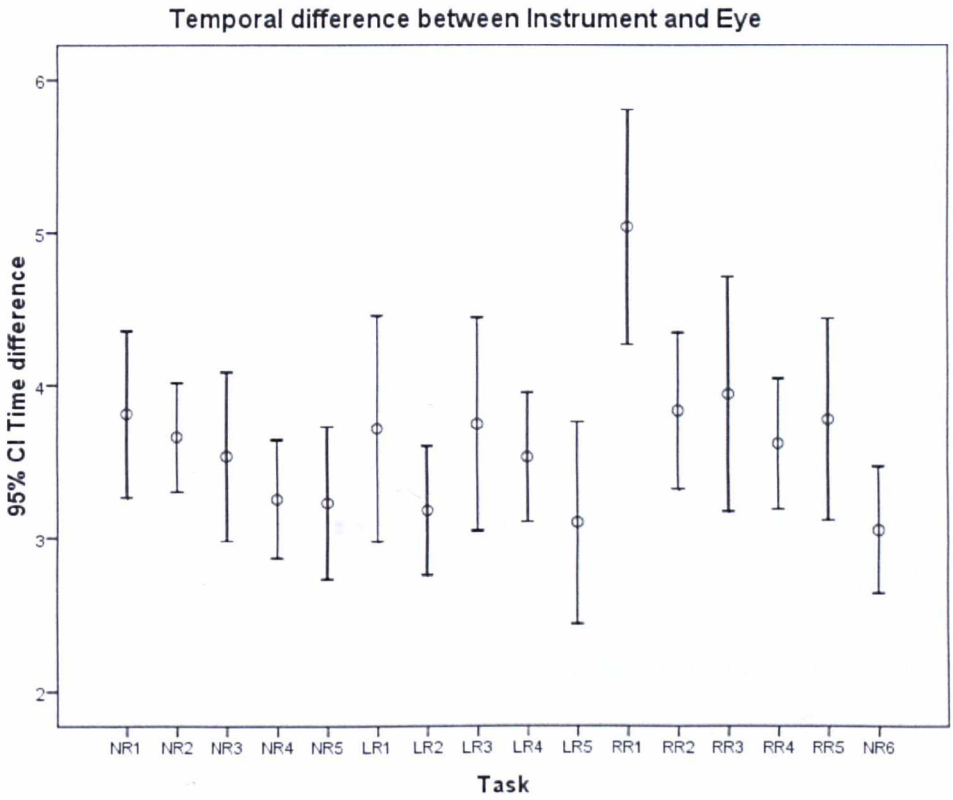

Figure 7.10 Showing mean and 95% confidence interval of the temporal differences between the instrument and eye positions throughout the experiment. NR – normal rotation, LR – left rotation, RR – right rotation.

NR 1 – 5 represent the first 5 consecutive tasks with normal camera orientation, this illustrates the significantly decreasing temporal delay between the instrument tip and the eye position (Spearman's correlation, $r = -0.118$, $p < 0.01$). Higher values represent larger delays of the eye positions from the instrument tip positions, where 1 unit equals 20ms. LR 1 – 5 represent the 5 consecutive left rotation tasks, which was randomised to either the subjects' first or second rotation. There was an increase in eye movement delay when counter clockwise rotation of the camera was introduced for the first time (LR1), and as the subjects learned throughout the 5 tasks, the delay decreased (not statistically

significant). RR 1 – 5 represent the right rotation task, similarly an initial increase of the eye movement delay was observed (Mann-Whitney Test, $z = -3.77$, $p < 0.001$), followed by a gradual improvement (Spearman's correlation, $r = -0.113$, $p < 0.05$).

Interestingly, the analysis shows that right rotation affects the hand-eye coordination more, and this was also observed during data collection (LR1 *versus* RR1, Mann-Whitney Test, $z = -2.62$, $p < 0.01$). This may be unique to the handedness of the subjects, and will be discussed further below.

NR6 represents the last experiment block with normal camera orientation. This shows the lowest time difference between the eye and the instrument tip movements, when compared to NR1 (Mann-Whitney Test, $z = -1.97$, $p < 0.05$), despite subjects had training in camera rotated tasks in between, perhaps this represents the adaptability of motor learning throughout the experiment, as the motor task was constant throughout the whole experiment.

7.3.1.4 Examination by time order

Figure 7.11 illustrates the temporal relationships between the instrument and the eyes, after matching windows were found using the *JSDr* algorithm. Again, only positive values were included, which examined the amount of delay of the eye movement from the instrument tip. The analysis here is based on the **order** of rotation, rather than the direction of rotation as above.

NR 1 – 5, again represent the first normal camera rotation tasks. FR 1 – 5 in this graph illustrate the first rotation task (randomised to left or right camera rotation), and SR 1 – 5 represent the second rotation task. NR6 is the last normal camera orientation task.

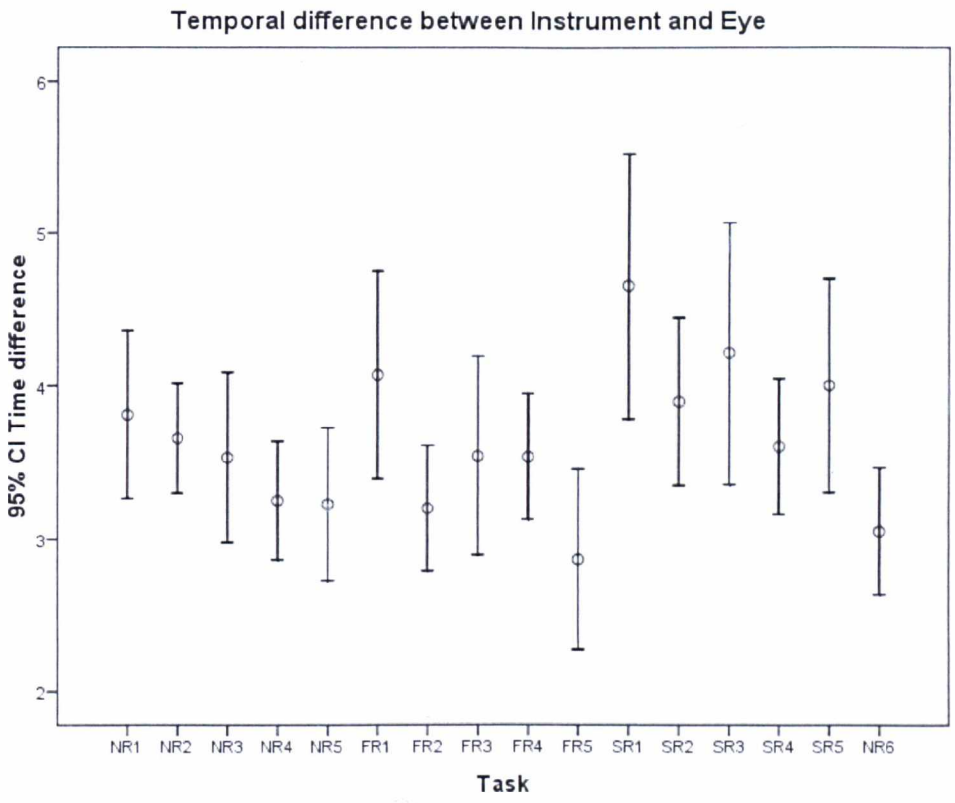

Figure 7.11 Showing mean and 95% confidence interval of the temporal differences between the instrument and eye position throughout the experiment. NR – normal rotation, FR – first rotation, SR – second rotation.

Similar trends are observed with the eye movements becoming less delayed from the instrument tip movements as the subjects trained, except for the disruptions when camera rotations are first introduced. The contrast with the above analysis is the lack of difference between the 1st and the 2nd rotation tasks (FR1 *versus* SR1, Mann-Whitney Test, $p > 0.4$). This is interesting, as it seems that the orientation of the rotation affects the subjects' hand-eye coordination more than the order of the rotation, contrary to the hypothesis stated in Table 7.2.

7.3.2 Complex Dissection Experiment

7.3.2.1 Qualitative analysis

Figure 7.12 shows the raw data of a novice's first and last attempt of the procedure, compared to an expert surgeon's eye and instrument data, again expressed as a distance function. The second attempt of the expert data was used for illustration, this is due to the

inevitable small learning curve in the first attempt to acclimatise to the new simulated surgical environment. The general shape of the time series are similar in all of them, especially in the first half of the data sets, which represents the first subtask of dissecting the subcutaneous tissue of the model (C.1).

The most notable difference is the amount of time taken to complete each procedure, with the novice taking around 880 and 360 seconds to complete the first and tenth tasks respectively, and the expert finished at around 120 seconds. Second, the high frequency data, which represents small movement corrections and tremor, decreases according to experience; although the time scale difference between the three graphs should be taken into account. Third, the left instrument movement trajectories appear smoother in the novice's last attempt compared to the right, and less tightly coupled with the eye movement data. This is reflected in the subsequent $JSDr$ analysis where a more notable improvement in the left hand is shown. This effect is also seen in the expert data, where the left instrument is used as an adjunct to the right, for example to manipulate objects surrounding the targets for better exposure, and becomes less reliant on visual guidance.

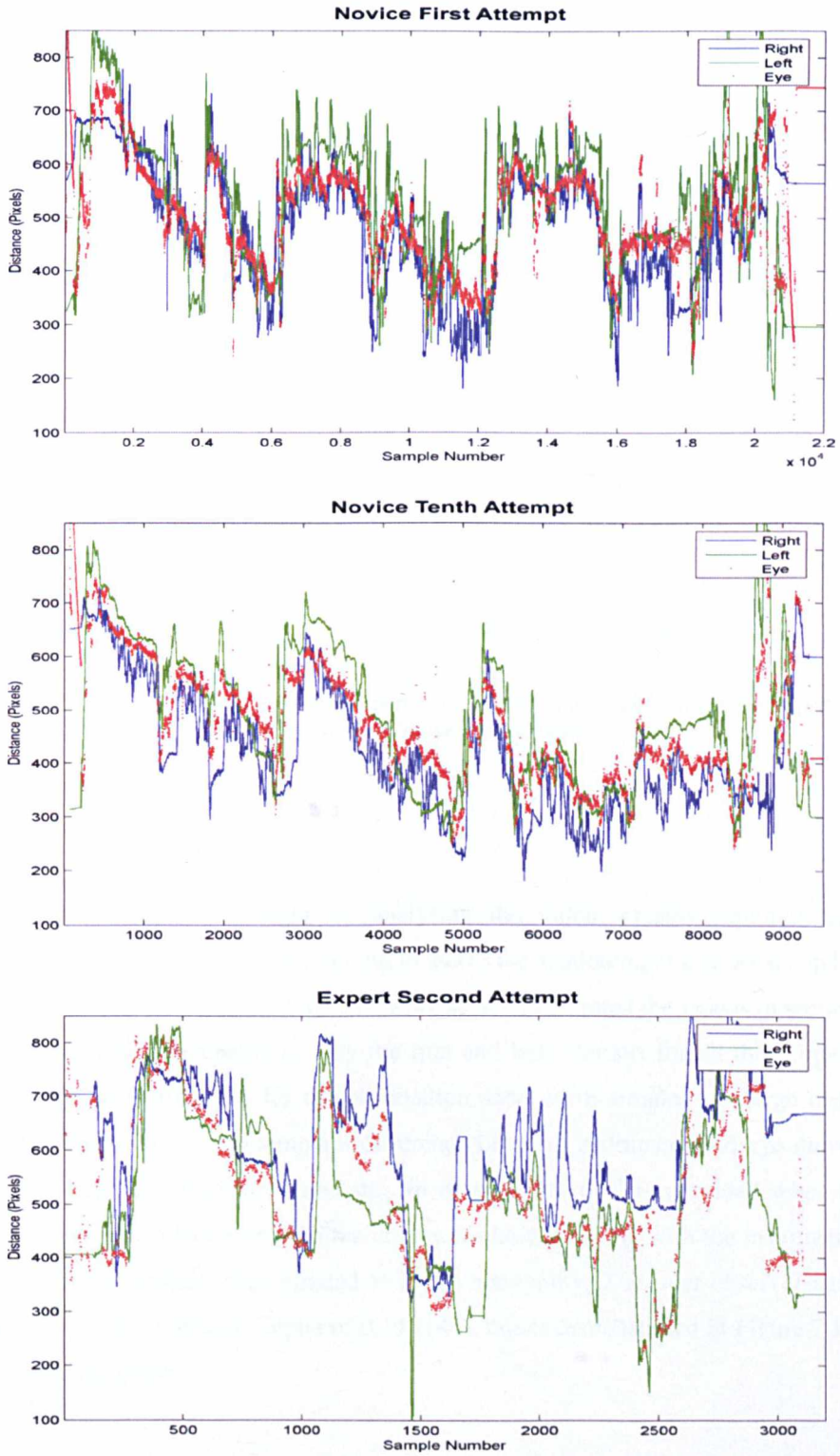

Figure 7.12 Illustrating the raw data for the whole procedures in the Complex Dissection Experiment, where the data is expressed as a distance function. Blue line – right instrument, green line – left instrument, red dot – eye fixations. Three figures are shown, top – novice's first attempt, middle – novice's last attempt, bottom – expert's second attempt.

The temporal relationship of the data streams is difficult to demonstrate due to the scale of the graphs in Figure 7.12. Selected typical examples are shown in Figure 7.13, which illustrates clearly novice's reactive eye movement, compared to the predictive behaviour of the expert's eye movement.

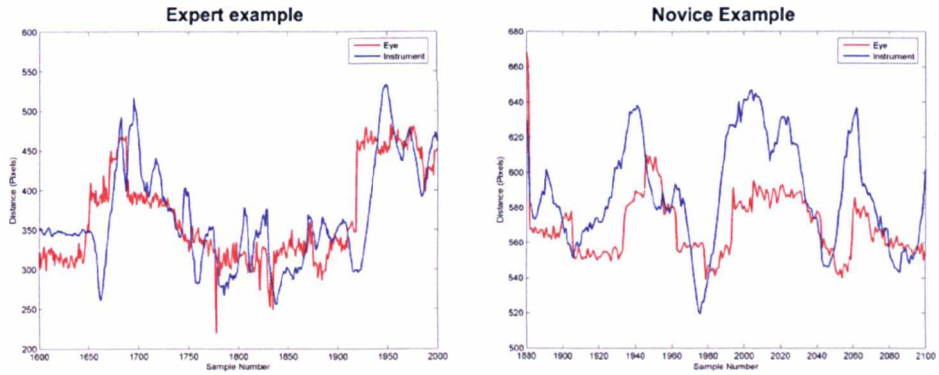

Figure 7.13 Typical examples of the expert's and novice's data sets, where the expert displays predictive and the novice shows reactive eye movement patterns.

7.3.2.2 Video analysis

Further validation was sought by analysing the video streams captured from the experiments, with eye fixations overlaid to assess the spatiotemporal relationship between the instruments and the eyes. Two independent observers rated the videos in sections, due to the abundance of repetition, only the first and last attempts for all the subjects were analysed. The instructions for the observation score were simple: 1 – large instrument leading behaviour, 3 – no temporal difference between instrument and eye movements, and 5 – large eye leading behaviour. An extra score, 6, was provided when the eye fixations appear to fixate only on the targets, without coupling with the instruments. The identity of the subjects was blinded to retain anonymity. The inter-observer agreement was good, with Cronbach's alpha of 0.79 (147), this is demonstrated in Figure 7.14 using a stacked line graph.

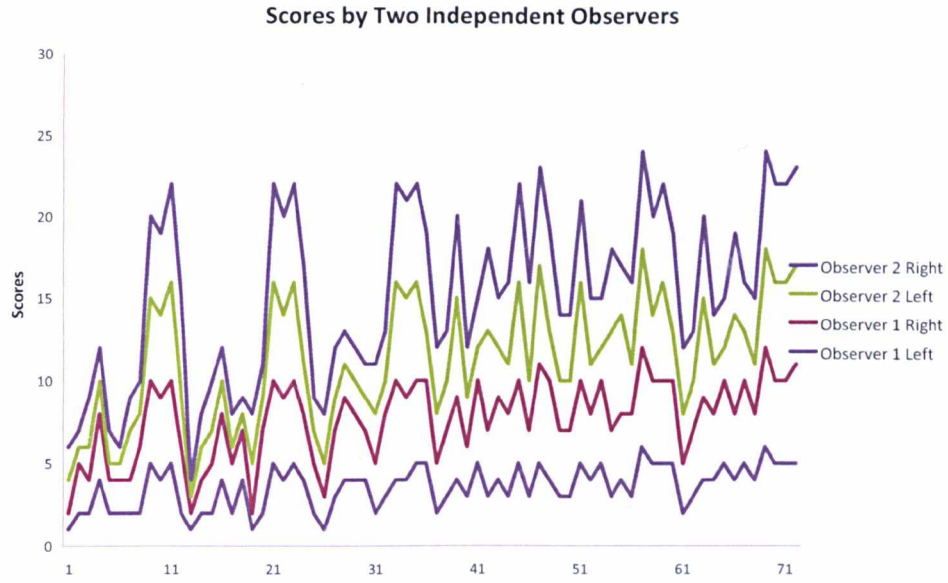

Figure 7.14 Stacked line graph showing the independent observation scores, showing good inter-observer agreement.

The observation score significantly increased for the novices' right instrument from a mean score of 2.5 to 4.1 (z = -5.4, $p < 0.001$) and left from 2.1 to 3.7 (z = -5.39, $p < 0.001$) analysed using Wilcoxon's signed rank test. This is clearly demonstrated as a box plot in Figure 7.15. For the experts, the mean left instrument scores were 4.7 and 5.0 for the first and last attempts respectively, with no significant difference. However, the mean right instrument scores increased from 5.4 and 5.7 (z = -2.11, p = 0.04). This may be due to the lack of familiarity of this specific procedure in the first attempts for the experts.

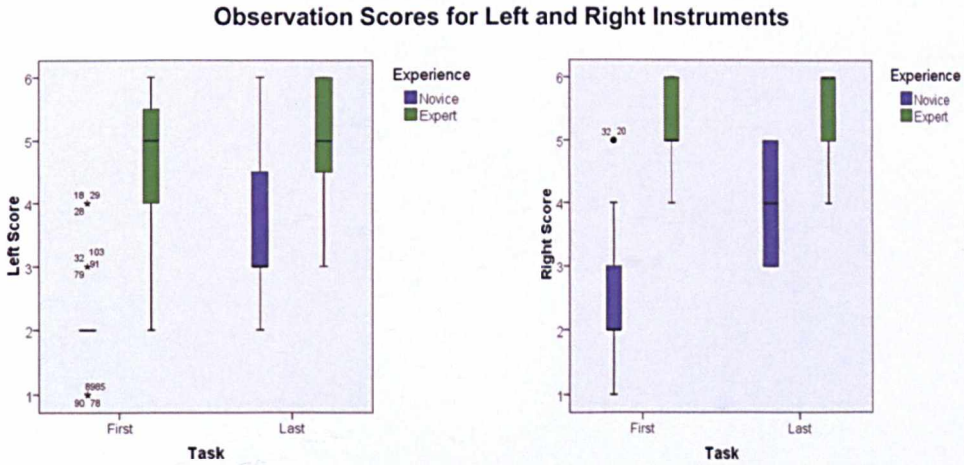

Figure 7.15 Box plots of the Observation Scores for the left and right instruments. Higher score represents eye leading behaviour. Dark line – median, shaded box – interquartile range.

7.3.2.3 Multiscale JSDr

Figure 7.16 illustrates the multiscale $JSDr$ analysis on the Complex Dissection Experiment. The novice's first and last attempts are shown in each hand separately, together with the expert's data. The length of the signal matching areas are restricted between 400ms and 2000ms as defined in the analysis algorithm, and are presented throughout the tasks as blue shaded areas in the top subplots. Even allowing for the time scale differences in the three sets of data, areas of matching appears to decrease in density with experience. This represents eye and instrument movement uncoupling with experience, and was described in Chapter 6, where non-sequential eye movements develop after training in MIS, as multiple target referencing becomes more prominent.

Examining the middle subplots, where the individual $JSDr$ values are plotted in colour for each signal window. The lower values (green and yellow), indicating better signal matching, are seen in the windows between 30 – 50 samples, which correspond to 1200ms to 2000ms during the experimental tasks.

The performance of the matching algorithm appears to be reasonable, where obvious non-matching areas not selected. Although some visually matching areas are not highlighted, again this lowered sensitivity is accepted to compensate for the higher specificity in the signal matching algorithm.

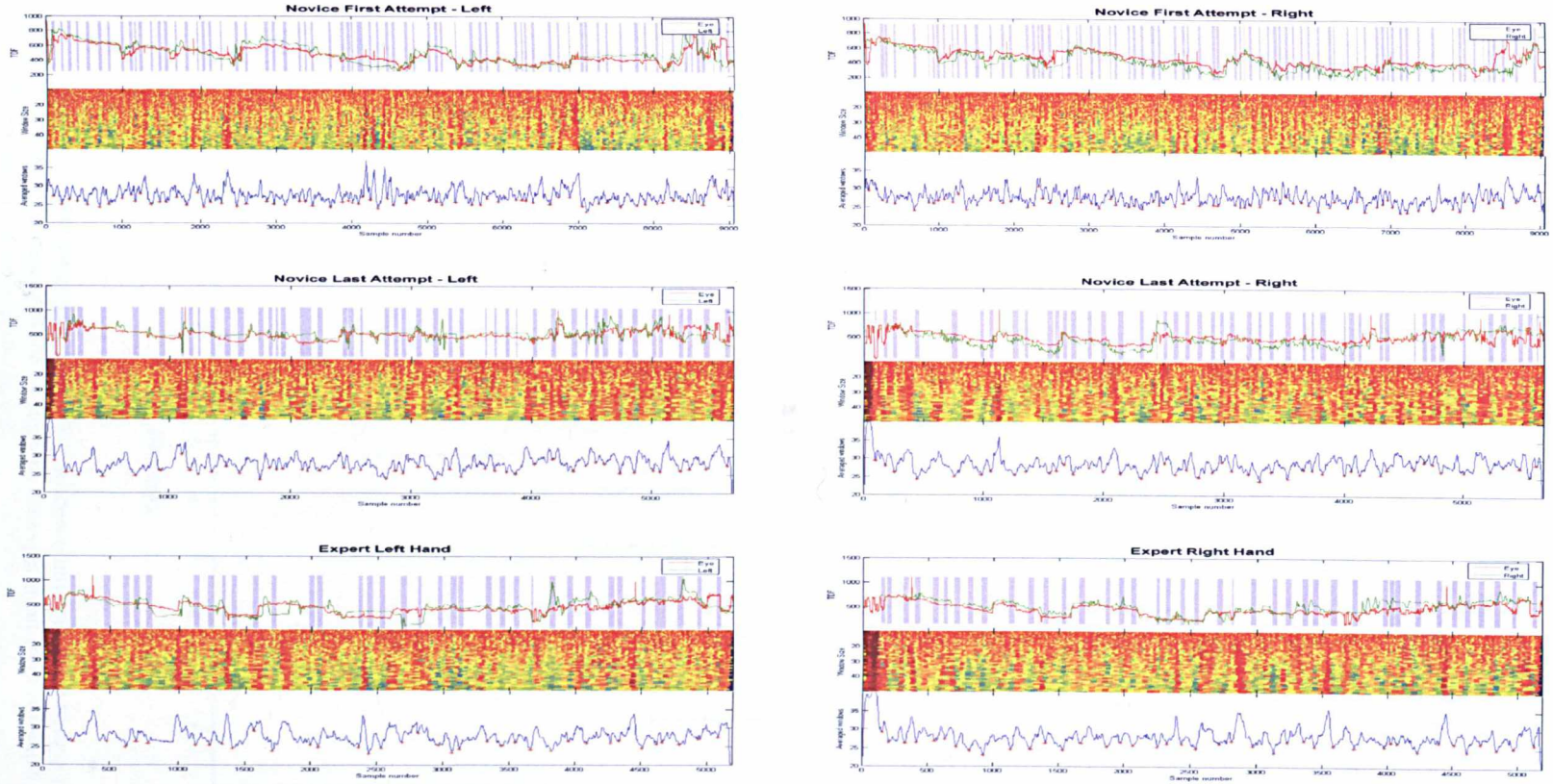

Figure 7.16 Novice (top 2) and expert (bottom) data from the Complex Dissection Experiment, showing TDF of the eye and instrument movements (top subplot), the Multiscale *JSDr* (middle subplot), and the average of the *JSDr* and locations of the minima (bottom subplot). For each of these minima (red stars), the window size with the lowest value was selected, and the matching areas with the windows are highlighted in blue. Left column – left hand, right column – right hand.

7.3.2.4 Temporal relationship

Figure 7.17 illustrates the temporal relationship between the **right** instrument tip and the eye movements. Positive values represent instrument leading eye movements, and *vice versa*. Blue error bars represent the 10 repetitions of the novices, and green bars represent the 5 from the experts. This clearly demonstrates that experts consistently display an eye leading behaviour (Expert versus Novices, Mann-Whitney Test, $z = -3.52$, $p < 0.001$), with little change throughout the 5 runs (Spearman's correlation, $p > 0.4$). On the other hand, the novices significantly improved throughout the first five repetitions (Spearman's correlation, $z = -0.06$, $p < 0.01$), from instrument leading behaviour to around 0 (no lead or delay between the instrument and the eye movements). Data from tasks 6 – 10 are more variable, which may be the effect of boredom and fatigue (Spearman's correlation, $p > 0.4$).

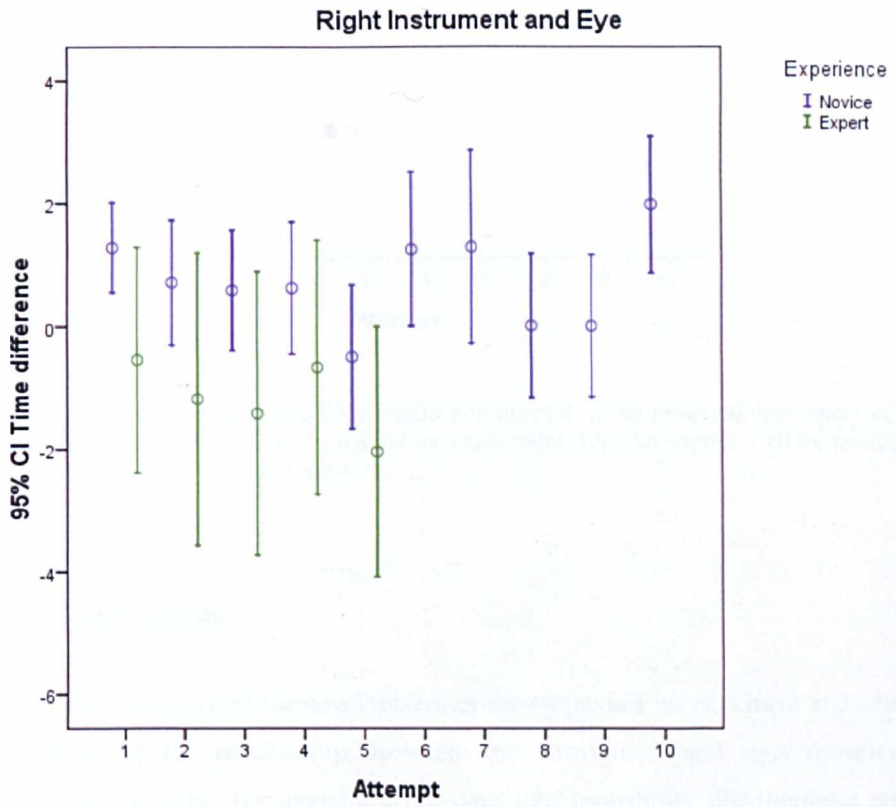

Figure 7.17 Showing mean and 95% confidence interval of the temporal differences of the **right** instrument and the eye position throughout the experiment from Attempts 1 – 10 for novices (blue) and Attempts 1 – 5 for experts (green)

Figure 7.18 illustrates the relationship between the **left** instrument and the eye movements. Interestingly, the expert left instrument movements centre around 0 with quite wide variance. A consistent slope towards 0 is observed with the novices left instrument movements.

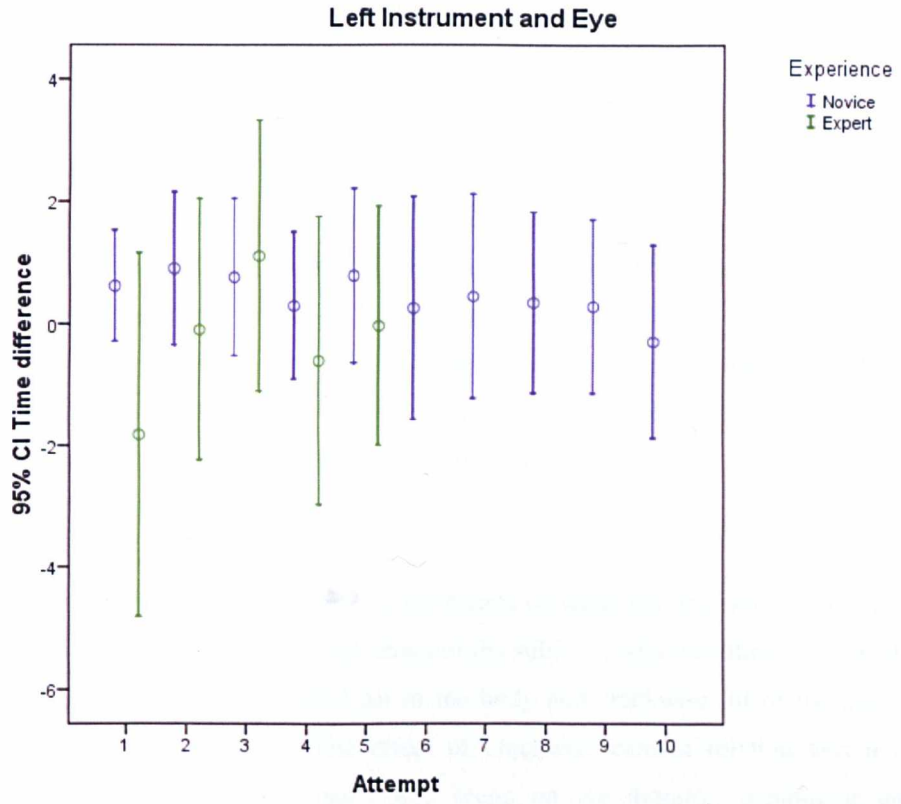

Figure 7.18 Showing mean and 95% confidence interval of the temporal differences of the **left** instrument and the eye position throughout the experiment from Attempts 1 – 10 for novices (blue) and Attempts 1 – 5 for experts (green)

7.4 Conclusions

This chapter summarised the development of the method of measurement and analysis of the spatiotemporal relationship between the instrument and eye movements in laparoscopic surgery. By modelling the data into probability distributions, principles based on the JSD proved to be a more robust signal matching algorithm. Novel modifications were made to the technique in order to suit the purpose of this study. First, in order to segment each trajectory automatically, a multi-scale approach was proposed. This allowed for the accommodation of variable movement trajectory lengths and time

order, and also created a natural filter against high frequency physiologically irrelevant movements. Second, the intrinsic problem of modelling time series data into probability distributions was addressed. As JSD on its own only considers the distribution of the data but not its time order, hence two signals with the same y-axis values but in completely different temporal order are considered a perfect match using JSD. In this modified technique, the correlation coefficient between the two data sets was included to address this problem.

Thus far, this was the first account of an automatic quantitative method for measuring hand-eye coordination in complex motor tasks, both within and outside surgery. In the Two Target Experiment, the effect of learning was accompanied by a decreasing time difference between the eye and the instrument. Echoing the findings of Chapters 5 and 6, the introduction of camera rotation significantly disrupted this coordination, and the eye movements became more reactive in nature, reflected by the larger time differences between the instrument and eye movements. The improved experimental design and hardware setup resulted in much higher quality data, and the effects of training and camera rotation were accentuated.

It was rather unexpected to observe a difference between left and right camera rotations. This is likely due to the right-handedness of the subjects, which is often associated with a natural tendency of a right sided tilt of the body and clockwise tilt of the head during intense work (*e.g.* writing). The effect of clockwise camera rotation was a counter clockwise rotation of the laparoscopic scene on the monitor, combining with the clockwise tilt of the subjects' heads, would generate more coordinate rotation than the opposite camera rotation. Nevertheless, this would require further examination with left handed subjects, and the measurement of head tilt during the experiment.

In the Complex Dissection Experiment, a high fidelity bimanual procedure was designed. Although specific instructions were given to the subjects on performing each subtask, large individual variations were observed in terms of surgical techniques, both within and between experience levels. Furthermore, variations in animal tissues were unavoidable, these challenges were similar to conducting experiments on live operations. Despite this, the proposed measurement algorithm successfully identified areas of matching signals throughout the procedure, a relatively high threshold of matching was chosen to increase the signal-to-noise ratio, at the expense of sacrificing some useful data.

In novices, a general decreasing trend was observed in the time difference between the instrument and the eye in both the left and right hands, this was consistent with improvements in hand-eye coordination as eye movements became less reactive. In contrast, experts displayed predictive eye movements consistently in the right hand, and less so in the left. Correlating with the observation results, some improvement was seen in the right hand, but not the left. It was an unexpected observation, although this may represent some adaptations in surgical techniques in performing this novel artificial animal model experiment. Another explanation could attribute to the relationship between the experts' left instrument and eye movements, which was likely to be non-sequential. As the experts required less visual information of the instrument tips, due to their well developed motor and proprioceptive systems, eye and instrument movements became more independent. As the left instrument was generally used to support, stretch and stabilise surrounding structures to facilitate target manipulation with the right instrument; the relationship of the eye and the left instrument may be more subtle to measure with its more peripheral role.

Although this method of analysis successfully highlighted the effect of learning, expertise and change in coordination in both simple and complex laparoscopic tasks, further validation work is needed to include larger subject populations with different surgical experience levels. Complete distinction between the experts and novices was not achieved using the current measure, as there was significant overlap between their results in the Complex Dissection Experiment. This may be due to the "ceiling effect" (152) when no camera rotation was introduced, hence a more complex task would have to be chosen for surgical skills assessment. However, it can be argued that the important difference was the switch from reactive (positive time difference) to predictive (negative time difference) eye movements, as this represented the complete adaptation of the MIS environment. The actual magnitude of difference may not be critical. Furthermore, one parameter of the analysis algorithm had to be chosen contextually, which is the range of the window sizes measured. In the laparoscopic tasks, it is unlikely that individual movement trajectories extend beyond 2000ms, after observation of the experimental recordings. This may change according to the nature of the tasks, although the basic framework seemed to be robust enough for two very different laparoscopic tasks.

This improvement in hardware setup has allowed non-intrusive tracking of the eyes, instruments and endoscopic camera, whilst allowing complete freedom of movement of the laparoscopic tools. The calibration process, although relatively quick and could be

performed partially offline, may still interrupt the flow of live operations. The experiments were mostly conducted with two operators, but further simplification is possible. The main practical limitations of the current setup are: the use of wired and non-sterile active IRED markers, which can be corrected with a slight modification of the setup; the relatively bulky and expensive Optotrak Certus system; and the line of sight problem with IRED trackers, although there are ceiling mounted tracking systems with multiple cameras to address this.

Future work should focus on bringing this measurement into live operating theatres, although further improvements in the setup are needed to be in place. First, instrument tracking hardware can be completely eliminated, as computer vision techniques in direct instrument segmentation from the video data are becoming available. Second, wearable eye trackers can provide vital data in eye movements outside the surgical scene, where human interactions, hand position referencing and outside disturbances can be factored in. Third, higher update frequency data collection could improve the performance of JSD, as the increased data points can provide better distribution matching in relative entropy. Finally, the inclusion of cognitive data could add further dimensionality in measuring hand-eye coordination. Cortical areas, such as the Posterior Parietal Cortex (PPC), are known to be involved in this coordination, and further results validation can be provided if cortical activation patterns are similar to the measures provided here.

This method appears robust enough to measure and analyse high fidelity simulations of laparoscopic tasks, with no artificial constraints of the operating environment. By slight modifications of the equipment, its use in live laparoscopic operations is theoretically possible. Even in its current form, this method provides a valid and reliable assessment of the subjects' familiarity of the MIS environment, by measuring their hand-eye coordination. It also demonstrates the improvement of hand-eye coordination with practice, and its deterioration with camera disorientation. This relationship is intrinsic to the brain's ability to perform the coordinate transformations necessary in MIS, and its use can be extended into other forms of surgery, such as NOTES and robotic assisted surgery.

Chapter 8

Conclusions and future work

8.1 Achievements of the Thesis

The practice of surgery continues to change in a rapid pace; with the introduction of new technologies, the increasing public awareness of clinical outcomes, and economical pressures for higher efficiency of the delivery of health care. The most notable change in surgery is the introduction of MIS over 20 years ago, where smaller incisions and quicker recovery are welcomed by both patients and surgeons. However, this has come at a price of technical challenges intrinsic to the arrangement of MIS; the misalignment of the visuomotor axis, the lack of 3D vision, and the "fulcrum" effect all contribute to the disruption of hand-eye coordination in MIS. Although new technologies have been developed to overcome some of these restrictions, for example, the da Vinci Surgical System (Intuitive Surgical, California, USA), a master-slave surgical robotic system with stereoscopic vision and direct movement translation of the surgeons' hands.

Surgical apprenticeship is also undergoing drastic transformations; the introduction of the European Working Time Directive (EWTD) and the shortened training years by the Modernising Medical Careers (MMC), surgical training has to be competence-based with performance measured objectively. Trainee selection should be more all-rounded, include technical skills evaluation which could predict learning curve and future performance. The challenge is to make the process more transparent and evidence-based, to ensure acceptance by the surgical community.

A large body of work has been published on technical skills assessment in the last decade, although the need for assessment is obvious, its actual implementation remains sparse. To this end, surgical skills assessment still relies heavily on labour intensive processes from experts in the field, whilst automatic scoring remains possible only on simple measures of hand movement kinetics. Visual assessment using category-based scoring system remains the most widely used method, although this type of manual scoring is difficult to be

completely devoid of subjectivity, and can be error prone when large numbers are involved. Most importantly, whilst these scores are useful for monitoring personal progression, they remain less meaningful when used strictly to define competency. Most validation studies can group experts, intermediates and novices in different score brackets, however there is often overlap between the three score distributions. Hence, it is difficult to impose a score threshold to define competency.

To facilitate the measurement of hand-eye coordination, several movement based methods have been proposed. These methods depend on modelling the centre of reference and the error generated by the movement trajectories, although provided convincing proof of the concept of sensorimotor transformation, are difficult to apply in complex skilled manoeuvres. Furthermore, these reference points are context-based, and can change according to the nature of the tasks, the availability and the quality of the visual input. Others have used a more qualitative approach to measure the spatiotemporal relationship between the hands and the eyes in relative complex tasks, although automatic quantification of this relationship remains unsolved.

The use of mathematical models has enabled the analysis of hand-eye coordination in this thesis. Although these models have been used in many different biological data, they have eluded researchers in surgical skills assessment. Stochastic (probabilistic) models are ideal for modelling human behaviour, as they introduce uncertainty in the human output based on probability distributions. In complex human tasks, like MIS, the number of variables influencing the ultimate goal directed behaviour is overwhelming, furthermore, cognitive influence due to prior experiences would be impossible to model precisely. It is perhaps the success of these stochastic models over previous attempts, where no strict definitions of behaviour patterns are defined.

In order to model the behavioural changes in MIS during the initial exposure to this unique environment, Chapter 4 used HMM to model each movement trajectory of the novices and surgeons performing a simple laparoscopic task. To further accentuate the difference between the novices and the surgeons, screen rotation of 90 degrees was introduced. Although this has successfully produced sufficient information to distinguish individuals of different abilities, the screen rotation tasks represent rather extreme conditions during MIS. Further training of the novices showed an improvement measured by this method, which further validates the model. However, HMM remains an indirect method of modelling the complex behaviour to represent the effects of the disruption of

mental sensorimotor transformation, in contrast to previous study where an exact point of reference was calculated, here the problem has become too complex due to the setup of the laparoscopic environment. The application of this algorithm in formal skills assessment is difficult, as it is an unrealistic task with low face validity.

In Chapter 5, eye movement data was introduced in the analysis for the first time. Assumptions were made that eye movements were only used to direct the active instrument, and each instrument trajectory was directed towards the target. These assumptions were accurate in this particular experimental setup, where the subjects were required to locate alternatively between two fixed targets. With this in consideration, using a simple target distance function, the eye movements were shown to change from lagging behind the instrument (reactive) to preceding it (predictive) with training. Without the known target locations and automatic trajectory segmentation this analysis would not be possible, although this chapter provided a proof of concept for the chapters to follow.

Based on the findings in Chapter 5, PDC was used to analyse the instrument and eye data in Chapter 6. PDC is a variation of Granger-causality, which is an econometric technique to calculate the relationship between two time series, and whether one can be used to forecast another. This technique eliminated the requirements of the target coordinates, and analysed the whole data streams in the frequency domain. The nature of the data collected presented with high frequency noise, which was a combination of tracking error, tremor and miniature eye movements, and can be identified in frequency based analysis. However, its performance was suboptimal on complex bimanual laparoscopic tasks, due to significantly large areas of non-matching data streams.

Chapter 7 modelled the instrument and eye movements into probability distributions, and JSD calculated the similarity between the two distributions. JSD is based on the Kullback – Leibler (KL) distance which is used in Appendix A to calculate consistencies in visual search patterns. By applying the modified JSD analysis to data of variable lengths enabled signal matching at different frequencies, hence allowing filtering of high frequency noise to improve the signal-to-noise ratio. This method seemed more robust, and validation with expert data was demonstrated.

In order to reach a target, initial visual search is needed to obtain its physical location before movement generations. Appendix A presented a study on the quantitative analysis

of the quality of visual search, based on a previously established cognitive model. KL distance and Gaussian curve fitting algorithms successfully established the changes of the quality and consistency of visual search, as a function of the effects of training and clinical specialisation.

8.2 Critical Comments and Discussions

In conclusion, this thesis has introduced the concept of measuring hand-eye coordination in surgery in the context of visuospatial transformation. This measurement implies the operator's familiarity of the specific movement parameters imposed by MIS, although this implication does not equate superior surgical performance, it is certainly a component of it. Unlike other scoring systems, it is shown that as soon as this mental transformation is achieved, an opposite spatiotemporal relationship occurs between the hands and the eyes. This creates a natural threshold of competency. Although only laparoscopic surgery is considered throughout the thesis, this measurement can extend to other aspects of surgery, including NOTES and robotic assisted surgery.

In terms of instrument trajectory quality measurements, the leave-one-out method was used in this thesis to train the HMM from all subjects excluding the test subject's data. The trained HMM was then used to calculate the log likelihood of the test subject and indicate similarities or differences to the learned model. This method can be influenced by the subjects' data as a group, hence the trained HMM would be substantially different if a group of laparoscopic experts were used. In order to use this method reliably, test subjects should have similar abilities and movement parameters, or a large reference movement trajectory database will need to be collected prior to testing. Furthermore, the measurement lacked context, hence its use in complex laparoscopic procedures would be limited. Contextual data could be obtained by using HMM as surgical gesture recognition, where basic manoeuvres such as grasping, cutting and lifting can be identified.

In this thesis, simple Cartesian distance measurements provided a simple proof of concept for the eye-instrument spatiotemporal relationship, its use was confined to simple direct reaching tasks with known target locations. The measurement also only provided an estimate of the temporal difference between them, as actual eye movement data was more complex. In a reaching task, eye saccades first landed on the target for initial location coding, and then smooth pursuit eye movement followed the instrument tip for motor

error feedback. In well rehearsed tasks, the smooth pursuit part of the eye movement was no longer necessary. In order to analyse this qualitative observation, segmentation of the eye movement into smooth pursuit and saccadic jump can provide further insight.

In order to reveal the intrinsic causality of the time series data collected for this thesis, PDC has been used. The major drawback of the technique is the lack of a well defined significance level to infer causal relationships, and the absolute value of PDC is of limited value. Nevertheless, this technique successfully identified areas of eye-instrument spatiotemporal relationship in the experiment with simple reaching tasks, and partially in complex bimanual tasks. It was found that large areas of non-matching data significantly increased the error of the analysis. Hence, signal pre-processing can improve the performance of PDC, where signal matching and segmentation are performed prior to PDC analysis.

JSD, thus far, has provided the most robust analysis technique for the instrument-eye spatiotemporal relationship. However, this method required a prior arbitrary definition of window size range, *i.e.*, only signal lengths of 400 to 2000ms were analysed, although this was chosen after careful observations of the data. The accuracy of the method is likely to improve, if the sampling rate of the tracking system is more frequent, with higher density probability distribution functions for JSD calculations. Furthermore, in order to provide better distinction between the novice and expert results, functional brain imaging data could provide specific areas of interest for analysis. As discussed further below, activation of certain cortical areas may represent intense hand-eye coordination processing, and highlight the instrument and eye data for closer scrutiny.

During the development of this thesis, further advances in computer vision have facilitated segmentation of the instrument tips directly from the laparoscopic video. This allows for direct quantification of the instrument tip position, without the need of optical tracking devices and fucidal markers. This will certainly simplify the experimental setup, providing a completely non-intrusive environment, whilst eliminating the problem of sterilisation of equipment in live operations. However, vision based techniques still rely on the visibility of the instrument tips, which is not always possible during surgery.

Further work should also include the use of anatomical target locations, as certain index targets, such as the cystic duct/artery and the common bile duct in a laparoscopic cholecystectomy, are known areas of focused dissection and avoidance respectively.

Analysis of gaze patterns around these areas can further reveal the underlying cognitive influences on hand-eye coordination, and provide contextually meaningful data. Advances in anatomical features and soft tissue deformation tracking can provide reasonably accurate information of these targets, these are based on invariant feature extraction techniques, such as Scale-Invariant Feature Transform (SIFT), and are being refined for such use.

8.3 Future perspectives

8.3.1 Robotic assisted MIS

Due to the constraints of laparoscopic surgery, other forms of MIS have emerged to improve the "usability" of this technique. Whilst ergonomic factors such as the "fulcrum" effect, monoscopic vision and visuomotor misalignment in laparoscopic surgery greatly affect normal hand-eye coordination of the operator; robotic assisted MIS, such as the da Vinci Surgical System (Intuitive Surgical, California, USA), was developed to counter some of these effects. The master-slave robotic design allows direct translation of instrumental movements to mimic the surgeon's hand movement eliminating the "fulcrum" effect. The surgeon is seated at the master console, whilst his fingers grasp the master control directly below the surgical scene display, thus creating a more natural visuomotor axis. Gross hand movement and tremor can be reduced by downscaling the movement generated through microprocessor control, further advantages include the additional degree of freedom through the "wristed" instruments, and the addition of stereoscopic endoscope. Current work at the Department focuses on further enhancements on the robotic system to improve patient safety through improved ergonomics. In cardiothoracic off-pump surgery, continuous movement of the heart greatly impedes the performance of the surgeon. The concept of gaze-contingent perceptual docking describes the use of the surgeon's eye perceptual behaviour to assimilate with the robot, whereby the eye movement vergence information is fed back into the system for 3D depth recovery (168) and motion stabilisation (169). Haptic and tactile feedback enhancement in the robotic system can also reduce the risk of tissue damage, whilst the current system would only allow for direct movement translation disregarding the force generated (170). Finally, the use of augmented reality can overlay computer-tomographic images onto the surgical scene, through 3D anatomical registration, these additional information can be used to define areas of safety and risk to alert the operator during surgery (171). The proposed framework of hand-eye coordination measurements can provide an objective method to illustrate the ergonomic advantages of robotic assisted MIS.

8.3.2 NOTES

Other technological advances have driven the move towards "invisible" surgery, like the Natural Orifices Translumenal Surgery (NOTES), where incisions are made through the walls of the stomach, vagina or rectum, where no external wounds are created (172). Access to internal organs is achieved by specially adapted flexible endoscopes, where instruments are tunnelled through additional lumens of the endoscope. Further challenges in terms of visual information restrictions, off-axis visualisation, spatial disorientation and difficult triangulation of the instruments need to be overcome. Undoubtedly, surgery without scars will be welcomed by patient groups, although tight regulations need to be in place to control the safety of the procedures, and the necessary technical skills of the surgeon to perform such tasks need to be defined (173).

The Natural Orifice Ssimulated sSurgical Environment (NOSsE™) is a high fidelity inanimate model developed in the Department, specifically designed to simulate the various ergonomic challenges in NOTES disposing the need of animal models. Solid and luminal organs within the abdominal cavity are modelled using foam and liquid latex respectively, with the addition of simulated diaphragmatic movements to be more realistic. The endoscope is passed through the model rectum, whilst the NOTES instruments are inserted whereby the usual "fulcrum" effect in laparoscopic surgery is lost. Figure 8.1 illustrates the NOSsE™ setup in the laboratory, which also depicts the glove sensors worn on the subject's hands for hand gesture tracking during the simulated tasks.

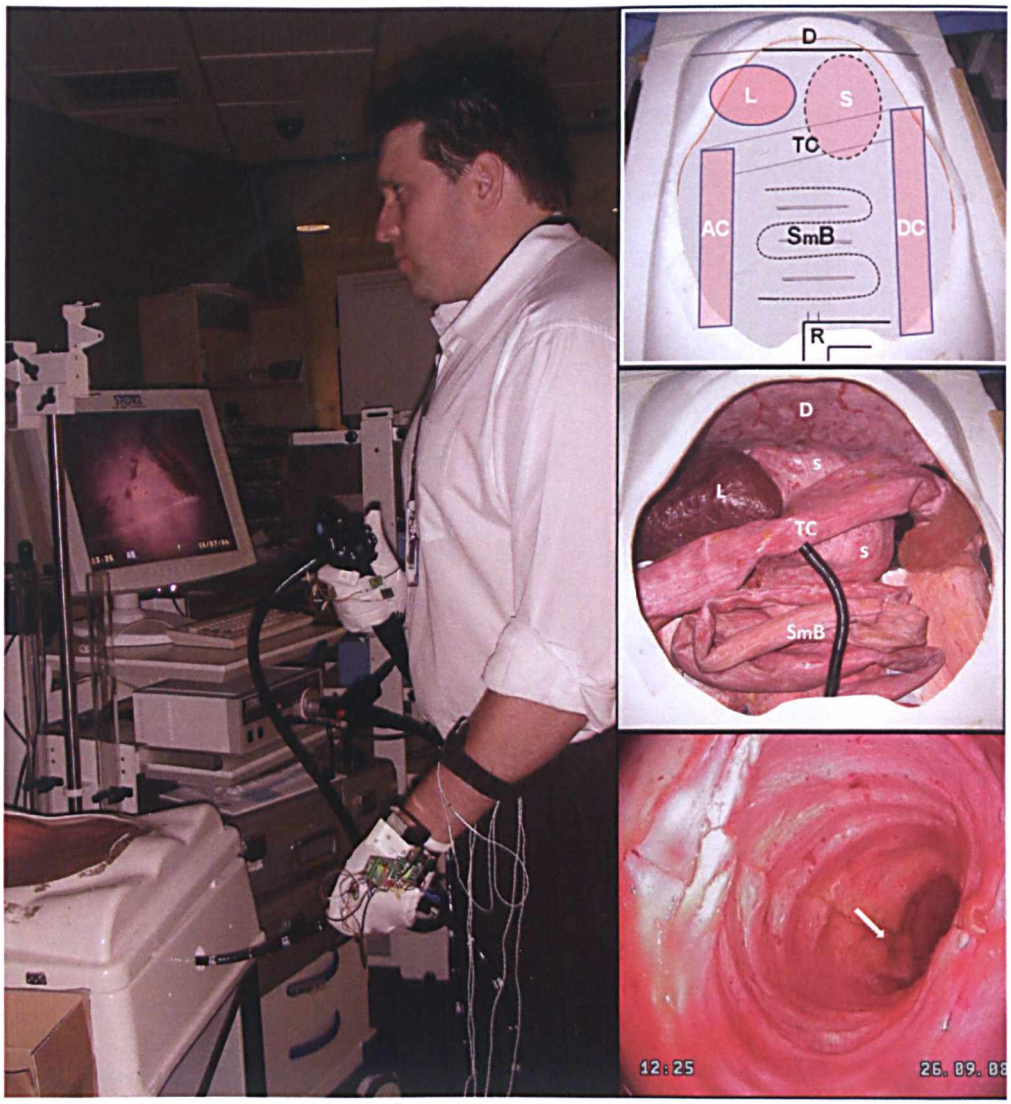

Figure 8.1 Illustration of the setup for NOSsE. Left – subject performing simulated surgical task wearing glove sensors. Top and middle – inanimate models of abdominal organs inside a laparoscopic training box (D – diaphragm, TC – transverse colon, SmB – small bowels, AC – ascending colon, DC – descending colon, L – liver, S – spleen). Bottom – illustrating the endoscopic view inside the simulated bowel. [picture courtesy of Mr James Clark]

Using this model, detailed studies of the ergonomic design are possible in the laboratory settings. In addition to the usual constraints in laparoscopic surgery, spatial disorientation within luminal organs remains most difficult to overcome, as illustrated in bottom right picture of Figure 8.1. This is akin to the camera rotation experiments presented in this thesis, where hand-eye coordination is impeded. The proposed framework of hand-eye coordination measurement can be used to assist in improving the designs of the setup of NOTES.

8.3.3 Functional brain imaging

Figure 8.2 Picture of the Hitachi ETG-400 Optical topography system, 24-channel optodes are placed on the table on the left. [picture courtesy of Mr Daniel Leff]

The Posterior Parietal Cortex (PPC) is known to be an area involved in sensorimotor transformation (91), whilst the prefrontal cortex is important for complex motor skills acquisition. As an extension of Chapter 7, brain activation data was collected during the experiment using functional Near InfraRed Spectroscopy (fNIRS), a non-invasive optical neuroimaging technique. Oxyhaemoglobin (HbO_2) levels are known to increase during functional brain activation, which is usually accompanied by a corresponding decrease in deoxyhaemoglobin (HHb) levels. These levels were recorded by the ETG-4000 Optical Topography System (Hitachi Medical Co., Japan) using a 24-channel array of optodes, illustrated in Figure 8.2. These optodes emit near infrared light at 695nm and 830nm which penetrate through to the cortical surface, the attenuated light levels detected are then used to calculate the HbO_2 and HHb levels using the modified Beer-Lambert law

(174,175). The optodes were placed over the left pre-frontal cortex and the right PPC during data collection, as illustrated in Figure 8.3.

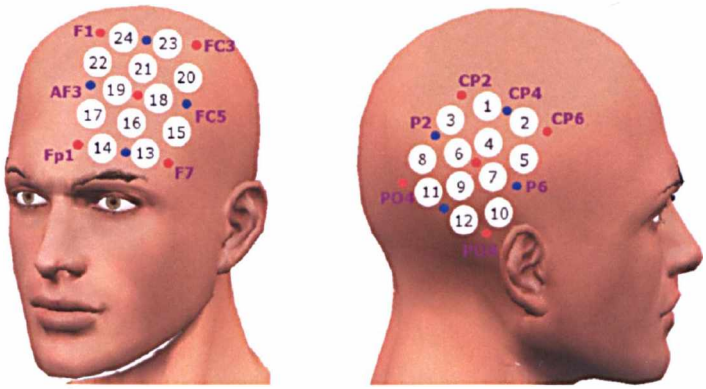

Figure 8.3 Illustrating channel placement of the optodes. [illustration courtesy of Mr Daniel Leff]

Data from Channel 4 of the pre- and post-training tasks of a subject is shown in Figure 8.4. The averaged HbO_2 concentration is plotted, where the shaded central area represents the task period, with the rest periods on either side. An increase in HbO_2 concentration is seen in the pre-training task, but not in the post-training task.

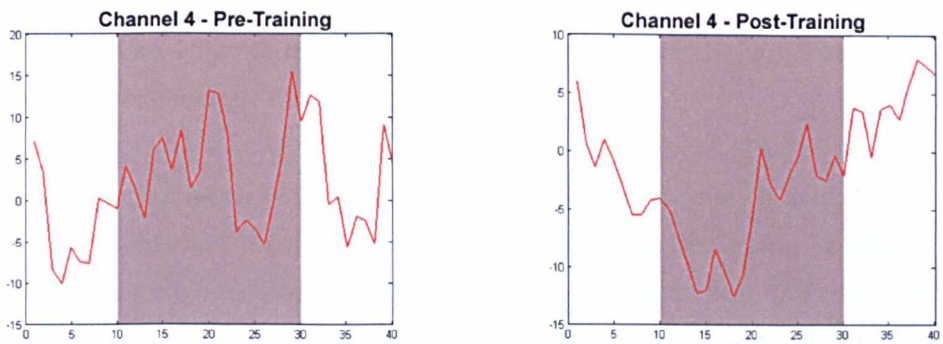

Figure 8.4 Averaged HbO_2 concentration measured by Channel 4 plotted in red. Left – pre-training task, and Right – post-training task. Shaded areas indicate motor task performance.

Figure 8.5 illustrates a subject's PPC activations during performance of the Two Target Experiment in Chapter 7. Here, a false-colour map is overlaid on the brain model, where the in-task data recordings show clearly an increased activation in the PPC during the pre-training task compared to the post-training task. Although it should be noted that this only provides a snapshot of the dynamic changes in the cortical areas throughout the tasks.

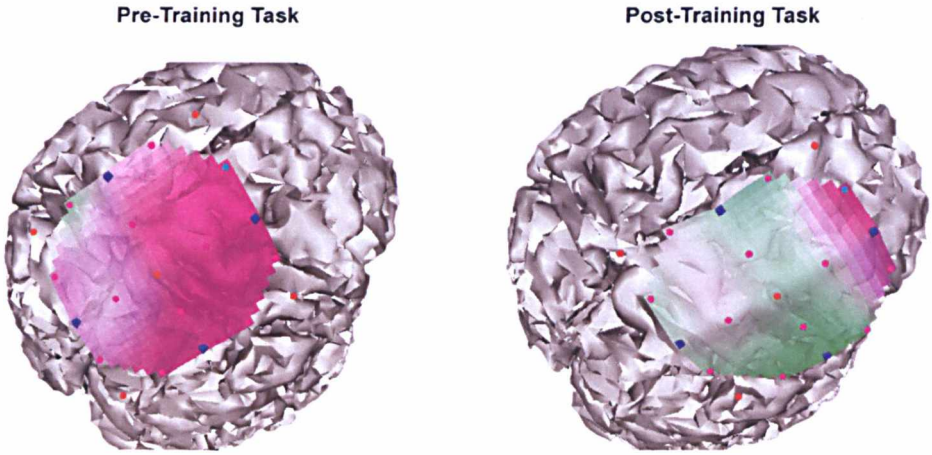

Figure 8.5 False-colour map overlaid on the PPC of the brain model. Left – pre-training task, right – post-training task. Pink illustrates higher haemoglobin concentrations compared to green. [illustration courtesy of Mr Daniel Leff]

Figure 8.6 shows the channels displaying decreasing trends in HbO_2 concentration, when compared with baseline rest periods, as the subjects progressed through the experiments (Channels 2, 4, 5 and 7, Spearman's rank correlation r = -0.22, -0.23, -0.19 and -0.23 respectively, and $p < 0.05$). These channels are placed on the PPC as illustrated in Figure 8.3, and this cortical area has been shown to be involved in visuospatial transformation. In this preliminary analysis, cortical activations in certain areas of the PPC decrease through training, although the introduction of camera rotation (L1 to L3 and R1 to R3) has not shown to increase this activation.

Due to the high dimensionality of the data, simple linear correlation is unlikely to reveal the intricate activities of cortical areas. Further work in the Department has shown that dimension reduction techniques, such as manifold embedding, have been successful in identifying individual channel and task activations (176). Extensive validation work has shown that pre-frontal cortex activation decreases with experience in simple surgical knot-tying tasks (177). Furthermore, after training, novices display more similar brain activation behaviour to expert surgeons (176). Inclusion of functional imaging data can provide a powerful tool in surgical skills assessment, cortical activations represent increases in the "work-load" of the brain during surgery, and this information can be used to highlight areas of interest in the eye-instrument data for further scrutiny.

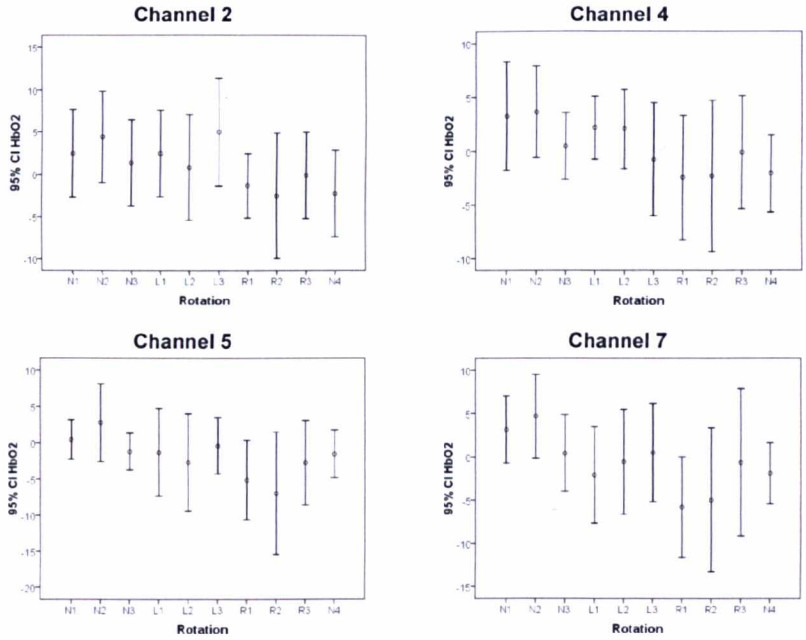

Figure 8.6 Mean and error bars (95% confidence interval) displaying the HbO_2 – baseline concentration throughout the training tasks. Channels 2 (top left), 4 (top right), 5 (bottom left), and 7 (bottom right) are displayed.

8.4 Conclusions

In conclusion, this thesis has provided a framework for quantitative analysis of hand-eye coordination in surgery. The resulted measurement implies the subject's familiarity of the specific MIS environment, and may prove to be invaluable for objective testing of the ergonomics of new surgical technologies. Validation of the technique has been presented through data from expert and novice surgeons performing simple and complex laparoscopic procedures. The key contributions of the thesis include the proposal of a novel concept for measuring hand-eye coordination in MIS, establishing the spatiotemporal relationship between the hand and eye movement in simulated laparoscopic tasks, and quantification of subtle differences in visual search pattern as a function of experience and training. To our knowledge, this is the first comprehensive attempt into automatic calculation of hand-eye coordination within and outside surgery; further consolidation of the method is needed through larger scale, long-term longitudinal trial. It should be noted however, in order to assess surgical skills comprehensively, hand-eye coordination only represents a small but important aspect of the essential abilities towards surgical competency.

Appendix A

Visual Search Behaviour in Skeletal Radiographs: experience *vs.* training

A.1 Introduction

Errors in the interpretation of radiographs in Accident and Emergency Departments are estimated to be 1.5% (178). In a busy unit, most radiographs are read by the treating physicians as well as radiologists to ensure good consensus, although often with a significant delay between the two interpretations. In a report by Williams *et al*, 671 cases were found with discrepant radiographic reports between the emergency and radiological staff in 1 year, of these 286 cases required further actions (179). Factors influencing diagnostic accuracies include the duration of training (180,181), and the difference in training methods between clinical specialities (182).

It has been estimated that up to 40% of radiographs taken in the hospitals are musculoskeletal images (183). Again, most of these radiographs have duplicate readings by radiologists and orthopaedic surgeons, and the discrepancies between their interpretations are also shown to be significant (184). The use of eye tracking methodologies may provide a possible means of understanding the factors involved in these inconsistencies.

One of the first documented studies of eye tracking was published in a psychology journal in 1901 (119). The technology has since evolved from being invasive, *e.g.* the use of a scleral contact lens with embedded search coil, to accurate and non invasive video based eye tracking devices using bilateral VOG methods (65). Existing research has shown that eye tracking data implies visual attention, and can provide further insight into

the cognitive process of image understanding and aberrant or idiosyncratic visual search behaviour (185-189).

Kundel and Nodine postulated the global-focal model for describing the behaviour of radiograph interpretation, they suggested four stages of search that include: (i) **global impression** which is defined as the initial search using mainly peripheral vision guidance and lasts for less than 200ms; (ii) **discovery search** which utilises the information from step one, and involves a detailed inspection of the target; (iii) **reflective search** which involves gathering evidence from cross referencing other potential targets; and (iv) **post search recall** which describes the period when the image is no longer available, and is recalled from memory (67). The first and last phases of the model are difficult to capture by eye tracking, whereas the discovery and reflective stages can be influenced by the ambiguity of the targets. This was demonstrated in a previous study using a low-contrast lung nodules detection experiment (67).

The purpose of this study is to provide a detailed quantitative analysis of the **discovery** and **reflective** stages of the visual search involved in identifying focal fracture sites in skeletal radiographs. It is aimed at establishing a numerical framework for the practical application of the global-focal model, and evaluating the effects of specialisation and training duration on visual search behaviour.

A.2 MATERIALS AND METHODS

A.2.1 Selection of Radiographs

A total set of 33 digital radiographic images were obtained from a London hospital, which consisted of 12 images of the hand (including 2 practice ones), 9 images of the knee, and 12 images of the shoulder. All images were converted from the Digital Imaging and Communications in Medicine (DICOM) standard to Tagged Image File Format (TIFF) format using lossless conversion, and only anteroposterior view was used. The images were standardised in size to fit to a screen resolution of 1280×1024 pixels, and patient information was removed. All images were reported by a consultant radiologist prior to the study. One shoulder, two knee, and two hand radiographs had no fractures, and four images had more than one fracture. Data from three images were discarded due to ambiguity of the diagnosis.

A.2.2 Eye Tracking Experiment Setup

A Tobii 1750 eye tracker (Tobii Technology, Stockholm, Sweden) was used to display the images. It is a remote eye tracking device using the standard binocular VOG technique with an accuracy of 0.5 degrees and a sampling rate of 30Hz, integrated with a 17 inch TFT display with a resolution of 1280×1024 pixels. It can tolerate moderate head movement within a 30×15×20cm volume at 60cm in front of the device, thus providing a relatively natural environment for radiograph interpretation.

A total of 25 subjects: five consultant radiologists (Rad), six consultant orthopaedic surgeons (Con), five orthopaedic specialist registrars (SpR), four orthopaedic senior house officers (SHO) and five accident and emergency department senior house officers (A&E) were recruited for the study. Ethical approval was obtained from St Mary's Local Research Ethics Committee, and all subjects signed written consents prior to the study.

The instructions were explained in writing and displayed on screen, all experiments were carried out in a darkened room with minimal noise disturbance, and the subjects were positioned 60 +/- 10cm in front of the screen, as illustrated in Figure A.1. After written consent, and standardised 5 point calibration on the Tobii eye tracker, repeat instructions were displayed on screen and two slides were used for familiarisation at the beginning of each session. 33 images were displayed sequentially. The subjects were asked to search for the fracture(s), and fix their gaze on the fracture(s) and press a button. The subjects were then required to report aloud a number from 1 to 5 (with 5 being most confident), indicating the confidence level for the diagnosis after each button 'click'. The image was changed when the subjects were satisfied that there was no further fracture. No other interactions were available for the subjects.

Pixel coordinates of the eye tracking data were acquired by using the software provided with ClearView 2.2.0. Fixations were calculated when gaze points fell within a 1.5° visual angle with a minimum duration of 100ms. The location of the fracture identified by the observer was indicated by the fixation point coincided with a button 'click' within 200ms.

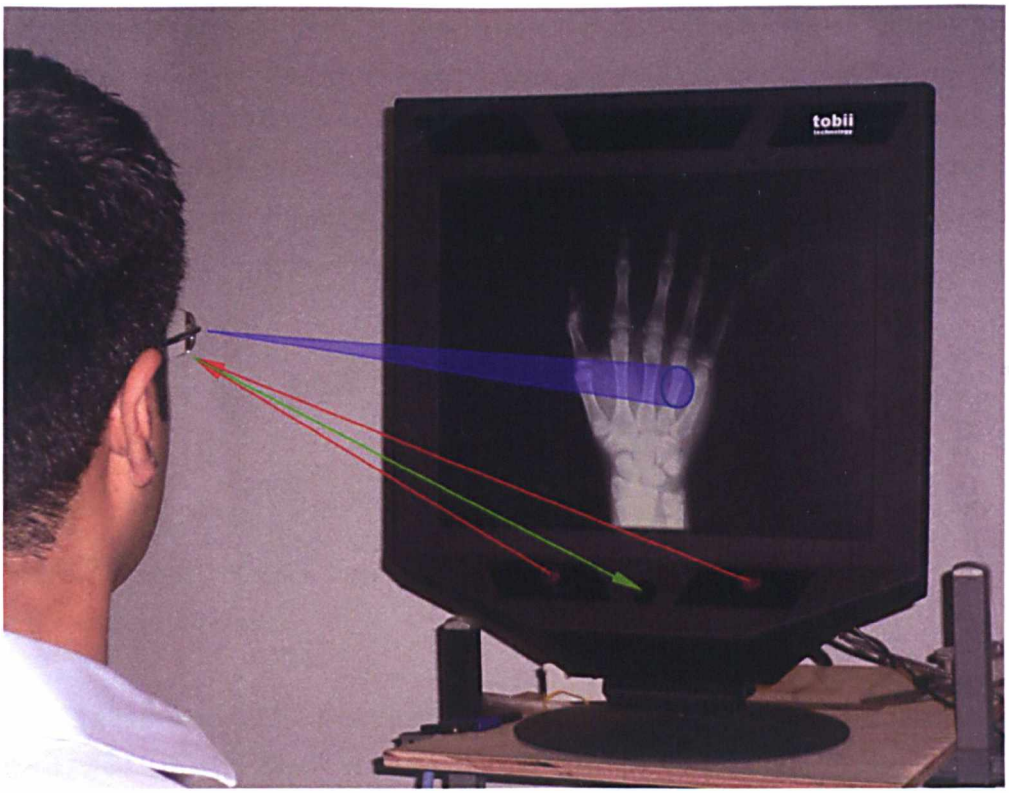

Figure A.1 Illustrating experimental setup. Subject is examining a hand radiograph displayed on the eye tracking screen. The red lines illustrate the infrared light from the unit, and the green line represents the reflection captured by the camera in the eye tracker. The blue cone is the subjects gaze point.

Time taken to interpret the radiograph, diagnostic performance and eye tracking data were analysed for each subject. The dwell time was the amount of time when the subject's fixations fell on the fracture site, and the medians of the group were used for comparison using non-parametric tests. Further analysis was performed by plotting the Cartesian distance between the gaze positions from the centre of the fracture as a function of time, or the Target Distance Function (TDF). The shape of the curves generated was used to assess the consistency of visual search strategy of the observer and the effect of experience on the search patterns as described further.

A.2.3 Statistical analysis

There were 19 radiographs with single fractures, and they were used in the subsequent analyses. Kullback-Leibler (KL) distance has been used to measure the statistical similarity between probability distributions (161). In this study, KL distance was used to assess the intra-observer search consistencies by comparing the TDF projections of

individual subject's eye tracking data, where low KL distances equate high search consistencies. The number of data points were standardised by interpolation, and the time-scale was normalised prior to analysis. The TDF from each radiograph was used to compare with all other 18 datasets from the same subject, this pair-wise comparison created KL distances in a 19×19 matrix. Images with multiple fractures, normal and ambiguous images were discarded for this analysis. KL distance was calculated by using the formula below:

$$d = \sum_k p_k \log_2 \left(\frac{p_k}{q_k} \right)$$

where d is the KL distance from p, the "true" probability distribution, to q, the "target" probability distribution.

Careful observations of the fixation data revealed distinct bimodal distribution in some of the datasets, implying two separate search phases. Hence, the TDF data was fitted with a two-mode Gaussian mixture model (shown in Figure A.3):

$$pg(x \mid c, \theta) = \sum_{i=1}^{2} c_i g_i(x \mid \theta_I)$$

where $\sum_{i=1}^{2} c_i = 1$, and

$$\begin{cases} g(x \mid \theta_i) = \dfrac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\dfrac{(x-\mu_i)^2}{2\sigma_i^2} \right) \\ \theta_i = (\mu_i, \sigma_i) \end{cases}$$

In the above equation, μ_i is the mean of the Gaussian component and σ_i the covariance. These parameters were derived from the Expectation Maximisation (EM) algorithm (146,190), which was solved iteratively through the following set of equations:

$$P_i(x) = \frac{c_i^{old} g_i(x \mid \mu_i^{old}, \sigma_i^{old})}{\sum_{k=1}^{G} c_k^{old} g_k(x \mid \mu_k^{old}, \sigma_k^{old})}$$

$$\begin{cases} c_i^{new} = \dfrac{\sum_x y(x)p_i(x)}{\sum_x y(x)} \\ \mu_i^{new} = \dfrac{\sum_x y(x)p_i(x)x}{\sum_x y(x)p_i(x)} \\ \sigma_i^{new} = \dfrac{\sum_x y(x)p_i(x)\left[(x-\mu_i^{old})^T(x-\mu_i^{old})\right]}{\sum_x y(x)p_i(x)} \end{cases}$$

where $y(x)$ is the original y value at time x.

The "goodness of fit" was then calculated by the Mean Squared Error (MSE) of the curve fitting, with low MSE values indicating better fit. The above analyses were done using bespoke software written in C++.

As the output data (KL distances and MSE values) was not normally distributed, Kruskal-Wallis test was used to demonstrate the difference between more than two groups, and Mann-Whitney test was used to compare between two groups. Non-parametric correlations were calculated using Spearman's rank test. SPSS 11.5 (Chicago, Illinois) was used for statistical calculations.

A.3 RESULTS

A.3.1 Qualitative analysis

Figure A.2 illustrates examples of the fixation distributions of a consultant orthopaedic surgeon and an orthopaedic senior house officer, whilst examining the same hand radiograph with multiple fractures. It illustrates qualitatively the difference in fixation patterns: the experienced consultant systematically surveyed all the individual bony structures of the hand, and successfully identified both fractures; whereas the junior clinician misdiagnosed the fracture and prematurely terminated the search for further targets.

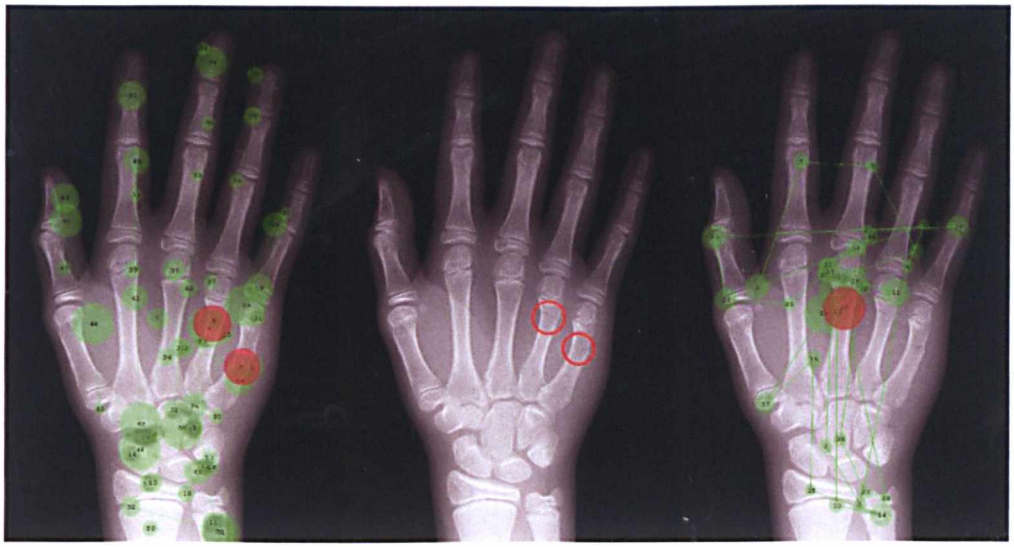

Figure A.2 Fixation plots of a consultant orthopaedic surgeon (left) and an orthopaedic senior house officer (right) viewing a hand radiograph with two fractures (circled red, centre). In these plots, the green circles are fixations (the larger the longer the fixation) and green lines are saccades, red circles indicate fixations when the button was pressed. This shows that the senior house officer missed the two fractures, and misdiagnosed the growth plate (third metacarpal) as a fracture. Furthermore, this subject did an incomplete survey of all the bones in the hand radiograph.

To illustrate the quality of the TDF data used for the Gaussian mixture model fitting, Figure A.3 provides three example plots of a consultant radiologist (A), an orthopaedic surgeon (B), and an A&E senior house officer (C) examining a hand radiograph. The data points (black dots) are gaze measurements, and the lines illustrate the two Gaussian components fitted over the data points. In Figures A.3A and A.3B, the data points fit well with the two Gaussian components, as compared to the poorly fitted Gaussian curves in Figure A.3C due to the disorganised eye fixations. The two Gaussian components are different in sizes in Figures A.3A and A.3B, where the former is composed of a relatively small first component when compared to the second, and *vice versa*. This may be implicated in the relative importance of the two stages of search, as indicated by the amount of time spent in each stage.

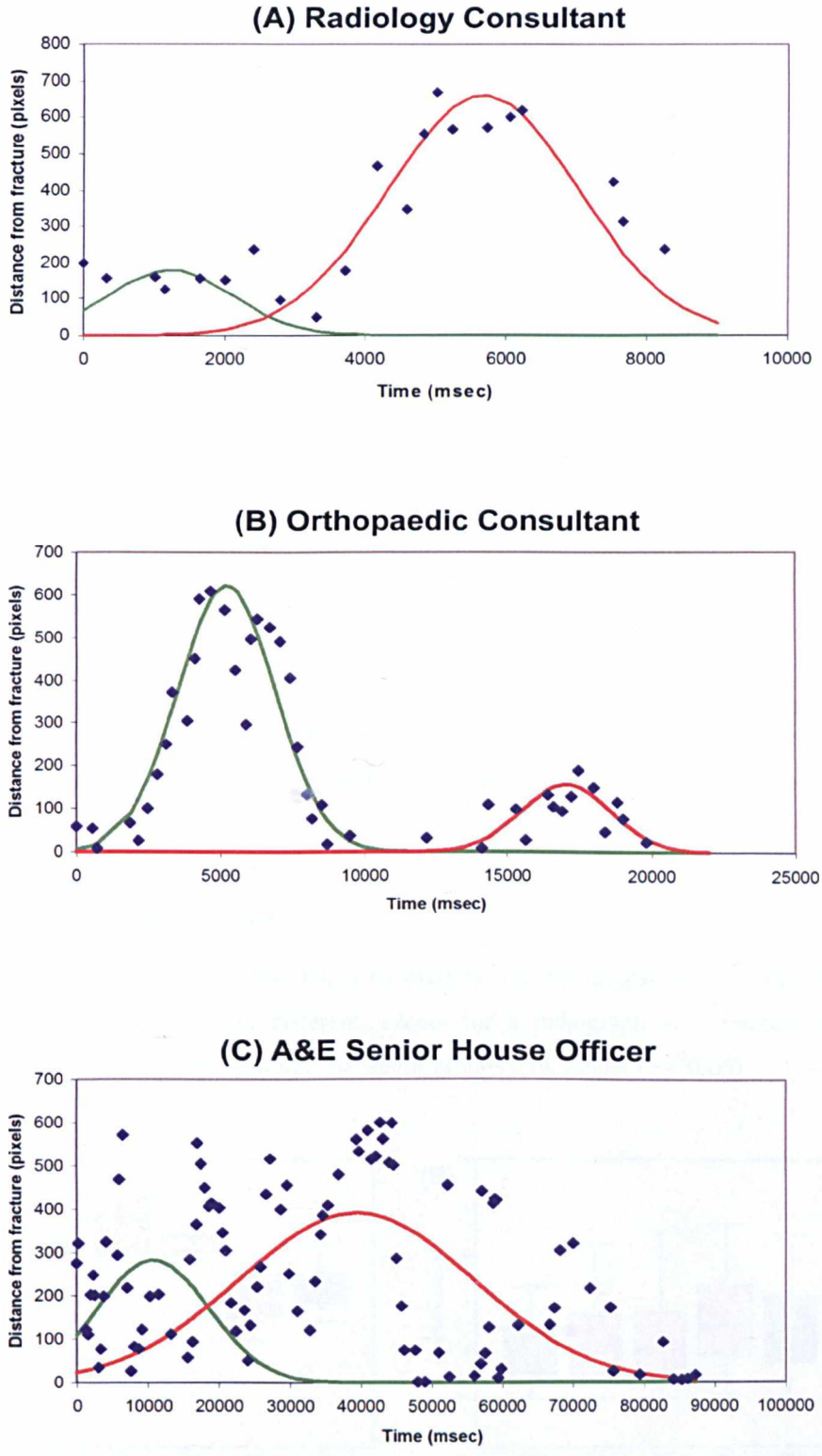

Figure A.3 Plot of TDF versus time of (A) a consultant radiologist, (B) a consultant orthopaedic surgeon and (C) a senior house officer in accident and emergency examining a hand radiograph. The data points (black dots) are gaze points whereas the lines illustrate the two Gaussian components fitted over the data points.

Figure A.4A illustrates the number of True Positives (TP) or identified fractures for all subjects, it is evident that the senior clinicians had much higher number of TP ($p < 0.001$). Lower number of False Negatives (FN) or missed fractures was also found in the senior groups ($p < 0.001$).

	Groups				
	Rad	**Con**	**SpR**	**SHO**	**A&E**
Total Time to finish whole study (s)	(529) **609** (1187)	(550) **741** (1220)	(342) **630** (1144)	(467) **585** (703)	(580) **766** (1114)
TP (total = 29)	(19.5) **20** (21.5)	(20) **24** (26)	(18.5) **22** (22.5)	(12.25) **16.5** (18.5)	(14) **16** (19.5)
Dwell time per fracture (TP) (ms)	(2364) **4407** (7021)	(2347) **4698** (6974)	(1593) **3999** (7410)	(2505) **4811** (7768)	(3445) **6014** (10369)
Dwell time per fracture (FN) (ms)	(0) **584** (3282)	(0) **277** (951)	(0) **228** (3234)	(0) **390** (1023)	(0) **455** (2080)

Table A.1 Summary of the results of the study. Bold = median, brackets = interquartile range

A.3.2 Dwell time analysis

It was found that the total time taken to examine all the images by different subject groups was not significantly different, except for a radiograph of a fracture of the shoulder with immature bone where the senior groups took longer ($p < 0.05$).

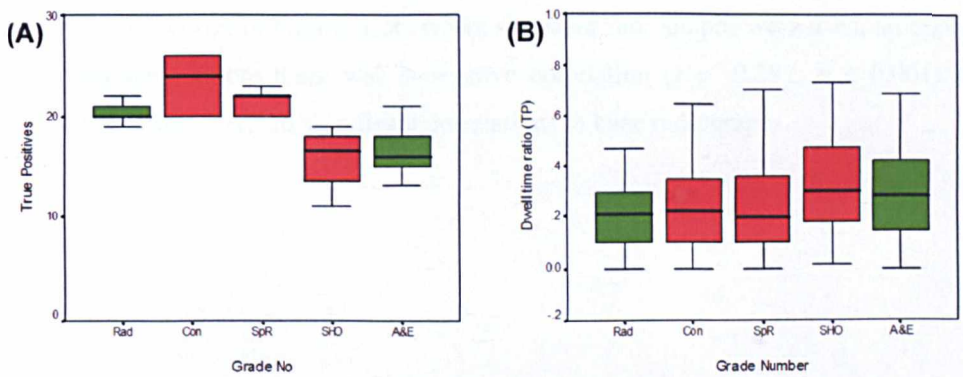

Figure A.4 True positives (A) and dwell time ratio (B) in TP radiographs. Rad and A&E are highlighted in green for easy comparison. (Boxplots show the median (line), interquartile range (shaded box), range of the data (whiskers), outliers and extreme cases not plotted)

Figure A.4B shows the dwell time ratio (which is defined as dwell time on fracture site divided by total time spent on the radiograph) among the five groups. For the TP radiographs, there was a significant difference in dwell time ratios between the groups ($p < 0.001$), with the senior groups spending less time at the fracture sites. In FN radiographs, no significant difference was found between the groups.

A.3.3 Search consistency

Within subject scan path comparison by KL distance was used to assess search pattern consistency, there was a significant difference between the groups ($p < 0.001$). The Rad group had significantly lower KL distances than the A&E ($p < 0.001$) and the Con groups ($p < 0.001$), indicating higher search pattern consistencies. The SHO group also had significantly lower KL distances than the A&E group ($p < 0.001$). The mean ranks of the KL distance for the different groups are illustrated in Figure A.5A. The variance of the KL distance was also compared within each subject, there was no significant difference between the groups; however, when radiologists were compared with non radiologists, there was a significant difference ($p < 0.05$).

A.3.4 Gaussian model fitting

Figures A.5B and C show the MSE of the Gaussian mixture model fitting of the five groups in examining the hand and shoulder radiographs respectively. The MSE of **hand** radiographs correlated with experience levels, in increasing order from Rad, Con, SpR, A&E and SHO ($r = 0.162$, $p = 0.07$). This indicated that the scan paths of the Rad group best fitted with the two-stage model. The Rad group had lower MSE than the A&E group ($p = 0.09$). However, in Figure A.5C where **shoulder** radiographs were used, an opposite trend was seen, where there was a negative correlation ($r = -0.287$, $p < 0.001$) with experience. There were no significant correlations in knee radiographs.

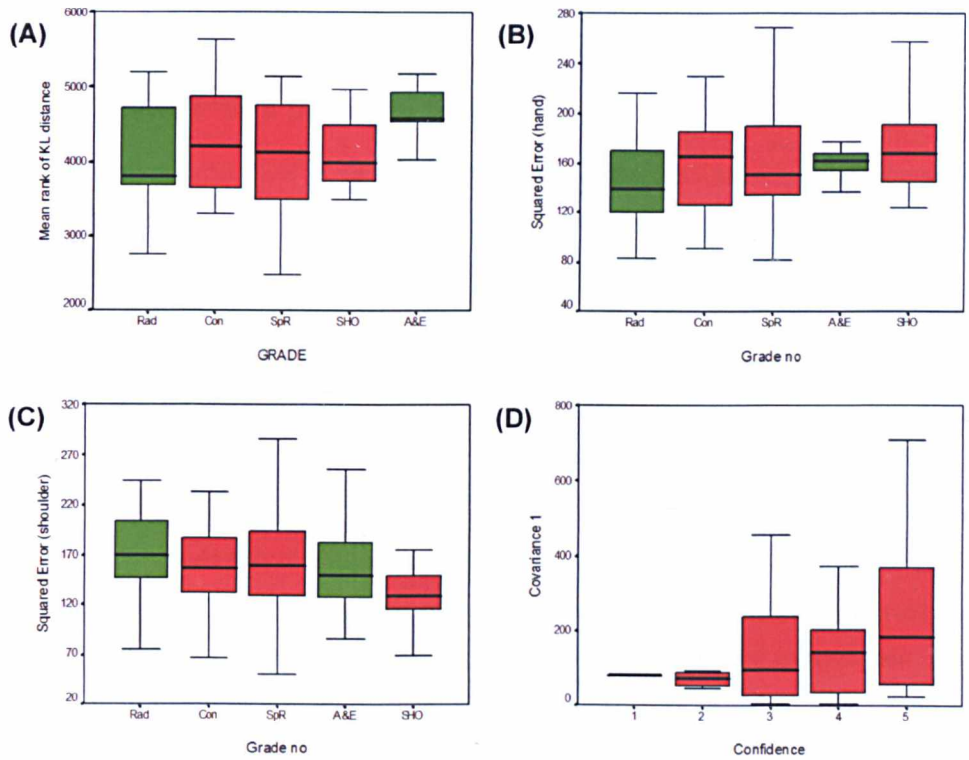

Figure A.5 (A) Mean rank of the KL distance between the five groups. (B) MSE of the Gaussian mixture model fitting in **hand** radiographs. (C) MSE of the Gaussian mixture model fitting in **shoulder** radiographs. (D) Covariance of the first Gaussian curve in hand radiographs plotted against the confidence level. Rad and A&E are highlighted in green for easy comparison. (Boxplots show the median (line), interquartile range (shaded box), range of the data (whiskers), outliers and extreme cases not plotted)

In Figure A.5D where only hand radiographs were examined, the covariance of the first Gaussian curve correlated with the confidence level ($r = 0.266$, $p = 0.07$). In other words, less time was spent on the second stage of visual search as the diagnosis was more confident. This can be referred back to Figure A.3, in the example shown in Figure A.3B the second Gaussian curve has a much lower covariance (narrower distribution) than the first, in contrast with Figure A.3A. Experience also influenced the diagnostic confidence of the subjects, where the confidence level was negatively correlated with the experience levels in the order as above ($r = -0.303$, $p < 0.05$).

A.4 DISCUSSION

A.4.1 Statement of principal findings

Based on the global-focal model, this study aims to provide a quantitative framework for assessing subtle differences in visual search behaviour in locating focal lesions in musculoskeletal images. It confirmed quantitatively that the more experienced observers have higher accuracies in fracture identification than less experienced ones, along with an explanation of the plausible causes.

A.4.2 Dwell time

This study showed that it was the distribution of time to interpret each image which was significantly different, not the total time taken. The dwell time analysis showed that in identified fractures (TP), less time was spent on the fracture site by experts than novices, as shown in Figure A.4B. This implied that with experience less time was needed at the fracture site for identification and decision processing, but more time was spent for cross-referencing or identification of further abnormalities.

A.4.3 Two-stage search

The Gaussian mixture model fitting was used to dissect the search pattern into two stages, this was decided experimentally after observations of all the raw data. It appeared that expert search strategy (especially in hand radiographs) was more consistent with the two-stage search pattern (see Figures A.3). Another interesting observation through the properties of the Gaussian model revealed that the covariance of the first curves increase with confidence, whereas the second covariance exhibit the opposite behaviour, as illustrated in Figure A.5D. The covariance of the Gaussian mixture model is proportional to the width of the curve, and hence more time was spent in the second stage of search when the diagnosis was less obvious. This further confirmed that the second stage was used for cross referencing other potential targets, as described in the global-focal model.

A strikingly different approach was observed in more conspicuous targets, namely shoulder radiographs. Here the number of potential fracture sites is limited, and are generally more obvious when compared with hand radiographs (191). Experts in fracture search should be able to detect the targets in the first stage of search (using only peripheral vision), and the next two stages of detailed search would become redundant.

The contrast with hand radiographs as displayed in Figures A.5B and C, in fact, further confirmed that the model best describes search models in subtle targets only.

A.4.4 Search consistency

Consistency in search strategy was quantified using KL distance in this study, this distance describes the amount of difference between gaze distributions. For each subject, the KL distance was calculated between all the images, so the shorter distance meant that similar strategies (or scan paths) were used throughout the study. The variance of KL distance also showed a similar trend, which reiterated the consistency in the expert groups.

A.4.5 The effect of training and specialisation

Previous reports have shown that A&E doctors have inferior diagnostic performances when compared with radiologists. The first aim of the study was to compare A&E senior house officers to radiology consultants, which should display the most difference in search behaviour. Indeed, it was found that radiologists were significantly more consistent with their search pattern, and seemed to adhere to the two-stage search strategy.

The comparison of clinicians across specialities provides an interesting contrast that may be explained by the difference in training and the primary aim of radiographic interpretation. Although radiology and orthopaedic consultants had similar diagnostic performances, radiologists were more consistent in their approach, and also adhered more closely to the two-stage search model. A&E and orthopaedic senior house officers also had similar accuracies in their interpretations. Interestingly, A&E group were less consistent in their search behaviour.

Furthermore, the effect of training was evaluated by examining the three orthopaedic groups at various levels of training. It seemed that training has neither changed the consistency nor the search strategies into the two-stage model proposed.

A.4.6 Weaknesses of the study

It should be noted that this study only used a relatively small number of single view radiographs without any clinical information given to the subjects. This was: first, to simplify the study design, a two-view study would add the complexity of subjects glancing between the two radiographs; second, if clinical information was given to the subjects, this would influence their search behaviour and bias the results; third, increasing the length of the study would introduce factors such as fatigue and boredom. These factors would further complicate the analysis framework.

The method used to identify fractures in this study required the subjects to prolong their gaze at the fracture site, whilst pressing a button. This would obviously artificially increase the dwell time at the fracture, although this increase should be similar in all groups. This method, however, allowed a more accurate assessment of their diagnostic performances.

A.4.7 Contrasts to previous studies

Although extensive previous research has been conducted in visual search scan path analysis in radiological images, dwell time analysis and time-to-target still remain the most commonly used metrics for analysis (187-189). Other studies have concentrated on extracting image features that influence eye movement pattern, to uncover the underlying cognitive strategies of visual search (185,192,193).

Skeletal radiographs are less studied using eye tracking, as the image features are more heterogeneous in nature. A study on hand and wrist radiographs revealed that radiologists used four different visual search patterns (circular, radial, zigzag, complex), however, this was analysed subjectively and qualitatively (194). Search pattern in radiographs with multiple fractures was also studied by Berbaum *et al*, the aim was to exhibit the effect of premature termination due to satisfaction of search (195). In terms of comparison of performances between specialities, two studies examined skeletal radiographs and eye movements of radiologists and orthopaedic surgeons. However, the studies focused on comparing the presenting media of the radiographs only (196,197), rather than comparing across clinical specialisations.

A.4.8 Meaning and implications of this study

The comparison of the radiologists with A&E doctors provided possible explanation of the difference in accuracies in diagnosis. The contrasts between orthopaedic surgeons and radiologists may be explained by their usual clinical practice. Orthopaedic surgeons tend to examine the patients prior to reading the radiographs, hence search strategy is heavily influenced by prior knowledge and clinical judgement, this is known as the "top-down" approach. In contrast, radiologists usually receive the radiographs with an insufficiently brief summary of the clinical picture; they also search for all abnormalities in the radiograph (not just fractures), this is called the "bottom-up" approach (198). The difference between A&E and orthopaedic senior house officers in their search consistency is interesting. A&E junior doctors usually have a mixed interest in their future careers; however, the majority of orthopaedic senior house officers will have an interest in developing a surgical career, and usually have stronger background knowledge in anatomy and surgical pathology.

The postgraduate training of radiologists is also very different from orthopaedic surgeons: radiologists tend to be taught formally how to interpret radiographs and usually have their results audited periodically, this is not the case in orthopaedics. This may explain the relatively unchanged search strategies between the three experience groups in orthopaedics. Formal education in orthopaedic surgeons in radiographic interpretation may be beneficial to their search consistencies.

A.4.9 Future research

This study first included radiographs with multiple fractures, but the behavioural analysis proved to be too complex. Further studies of radiographic images with multiple pathologies are warranted, where the effect of satisfaction of search may be quantified mathematically (195).

Eye tracking may prove to be useful for training in radiographic interpretation. However, its routine use will require further improvements of the eye tracking technology in being truly pervasive and not affecting the usual behaviour of the observers. Further development in the analysis framework is also necessary to account for the idiosyncrasy of cognitive visual search strategies used, as this study primarily focused on the analysis of spatiotemporal scan path patterns.

Appendix B

Ethics Committee Approval

Ethics Committee	Approval date	Reference number
St Marys LREC	30/09/03	EC No. 03.77 R&D No. 03/CD/002E
St Marys LREC	31/03/06	Extension of project MS/VP approved

Reference List

[1] Hawasli A, Lloyd LR. Laparoscopic cholecystectomy. The learning curve: report of 50 patients. *Am Surg*, 1991. **57**(8): pp. 542-4.

[2] Peters JH, Ellison EC, Innes JT, Liss JL, Nichols KE, Lomano JM, Roby SR, Front ME, Carey LC. Safety and efficacy of laparoscopic cholecystectomy. A prospective analysis of 100 initial patients. *Ann Surg*, 1991. **213**(1): pp. 3-12.

[3] Sidhu RS, Grober ED, Musselman LJ, Reznick RK. Assessing competency in surgery: where to begin? *Surgery*, 2004. **135**(1): pp. 6-20.

[4] Ren L, Khan AZ, Blohm G, Henriques DY, Sergio LE, Crawford JD. Proprioceptive guidance of saccades in eye-hand coordination. *J Neurophysiol*, 2006. **96**(3): pp. 1464-77.

[5] Madan AK, Frantzides CT, Park WC, Tebbit CL, Kumari NV, O'Leary PJ. Predicting baseline laparoscopic surgery skills. *Surg Endosc*, 2005. **19**(1): pp. 101-4.

[6] Lerch HA. Developmental motor activities for all children. Dubuque, IA: 1992.

[7] Modlin IM, Kidd M, Lye KD. From the lumen to the laparoscope. *Arch Surg*, 2004. **139**(10): pp. 1110-26.

[8] Zajaczkowski T, Zamann AP. Julius Bruck (1840-1902) and his influence on the endoscopy of today. *World J Urol*, 2004. **22**(4): pp. 293-303.

[9] Reuter MA, Reuter HJ. The development of the cystoscope. *J Urol*, 1998. **159**(3): pp. 638-40.

[10] Litynski GS. Laparoscopy--the early attempts: spotlighting Georg Kelling and Hans Christian Jacobaeus. *JSLS*, 1997. **1**(1): pp. 83-5.

[11] Litynski GS, Paolucci V. Origin of laparoscopy: coincidence or surgical interdisciplinary thought? *World J Surg*, 1998. **22**(8): pp. 899-902.

[12] Hatzinger M, Kwon ST, Langbein S, Kamp S, Hacker A, Alken P. Hans Christian Jacobaeus: Inventor of human laparoscopy and thoracoscopy. *J Endourol*, 2006. **20**(11): pp. 848-50.

[13] Calder I, Ovassapian A, Calder N. John Logie Baird--fibreoptic pioneer. *J R Soc Med*, 2000. **93**(8): pp. 438-9.

[14] Hopkins HH, Kapany NS. A Flexible Fibrescope, using Static Scanning. *Nature*, 1954. **173**(39): pp. 39-41.

[15] Van Heel ACS. A New Method of transporting Optical Images without Aberrations. *Nature*, 1954. **173**(39): pp. 39.

[16] Fuchs GJ. Milestones in endoscope design for minimally invasive urologic surgery: the sentinel role of a pioneer. *Surg Endosc*, 2006. **20 Suppl 2**: pp. S493-S499.

[17] Semm K. Endoscopic appendectomy. *Endoscopy*, 1983. **15**(2): pp. 59-64.

[18] Litynski GS. Erich Muhe and the rejection of laparoscopic cholecystectomy (1985): a surgeon ahead of his time. *JSLS*, 1998. **2**(4): pp. 341-6.

[19] Reynolds W, Jr. The first laparoscopic cholecystectomy. *JSLS*, 2001. **5**(1): pp. 89-94.

[20] McMahon AJ, Fischbacher CM, Frame SH, MacLeod MC. Impact of laparoscopic cholecystectomy: a population-based study. *Lancet*, 2000. **356**(9242): pp. 1632-7.

[21] Bingener-Casey J, Richards ML, Strodel WE, Schwesinger WH, Sirinek KR. Reasons for conversion from laparoscopic to open cholecystectomy: a 10-year review. *J Gastrointest Surg*, 2002. **6**(6): pp. 800-5.

[22] Keus F, de Jong JA, Gooszen HG, van Laarhoven CJ. Laparoscopic versus open cholecystectomy for patients with symptomatic cholecystolithiasis. *Cochrane Database Syst Rev*, 2006. (4): pp. CD006231.

[23] Nicolaou M. The assessment of visual behaviour and depth perception in surgery. London. Imperial College London, 2006. PhD Thesis.

[24] Bingener J, Buck L, Richards M, Michalek J, Schwesinger W, Sirinek K. Long-term outcomes in laparoscopic vs open ventral hernia repair. *Arch Surg*, 2007. **142**(6): pp. 562-7.

[25] Harinath G, Shah PR, Haray PN, Foster ME. Laparoscopic colorectal surgery in Great Britain and Ireland--where are we now? *Colorectal Dis*, 2005. **7**(1): pp. 86-9.

[26] Haveran LA, Novitsky YW, Czerniach DR, Kaban GK, Taylor M, Gallagher-Dorval K, Schmidt R, Kelly JJ, Litwin DE. Optimizing laparoscopic task efficiency: the role of camera and monitor positions. *Surg Endosc*, 2007. **21**(6): pp. 980-4.

[27] Crothers IR, Gallagher AG, McClure N, James DT, McGuigan J. Experienced laparoscopic surgeons are automated to the "fulcrum effect": an ergonomic demonstration. *Endoscopy*, 1999. **31**(5): pp. 365-9.

[28] Jordan JA, Gallagher AG, McGuigan J, McClure N. Randomly alternating image presentation during laparoscopic training leads to faster automation to the "fulcrum effect". *Endoscopy*, 2000. **32**(4): pp. 317-21.

[29] Hanna GB, Cresswell AB, Cuschieri A. Shadow depth cues and endoscopic task performance. *Arch Surg*, 2002. **137**(10): pp. 1166-9.

[30] Mishra RK, Hanna GB, Brown SI, Cuschieri A. Optimum shadow-casting illumination for endoscopic task performance. *Arch Surg*, 2004. **139**(8): pp. 889-92.

[31] Nicolaou M, James A, Lo BP, Darzi A, Yang GZ. Invisible shadow for navigation and planning in minimal invasive surgery. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 2005. **8**(Pt 2): pp. 25-32.

[32] Hanna GB, Shimi SM, Cuschieri A. Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy. *Lancet*, 1998. **351**(9098): pp. 248-51.

[33] Munz Y, Moorthy K, Dosis A, Hernandez JD, Bann S, Bello F, Martin S, Darzi A, Rockall T. The benefits of stereoscopic vision in robotic-assisted performance on bench models. *Surg Endosc*, 2004. **18**(4): pp. 611-6.

[34] Taffinder N, Smith SG, Huber J, Russell RC, Darzi A. The effect of a second-generation 3D endoscope on the laparoscopic precision of novices and experienced surgeons. *Surg Endosc*, 1999. **13**(11): pp. 1087-92.

[35] Savescu A-V, Cheze L, Wang X, Beurier G, Verriest J-P. A 25 degrees of freedom hand geometrical model for better hand attitude simulation. SAE Digital Human Modelling for Design and Engineering Symposium, 2004. p. 2196.

[36] Pellegrini CA. Surgical education in the United States: navigating the white waters. *Ann Surg*, 2006. **244**(3): pp. 335-42.

[37] The effects of the Calman reforms on training, service provision, and recruitment. *Arch Dis Child*, 1999. **80**(1): pp. 88-90.

[38] Catto G. Specialist registrar training. Some good news at last. *BMJ*, 2000. **320**(7238): pp. 817-8.

[39] Morris-Stiff GJ, Sarasin S, Edwards P, Lewis WG, Lewis MH. The European Working Time Directive: One for all and all for one? *Surgery*, 2005. **137**(3): pp. 293-7.

[40] Neville E. Modernising medical careers. *Clin Med*, 2003. **3**(6): pp. 529-31.

[41] Devey L. Will modernised medical careers produce a better surgeon? *BMJ*, 2005. **331**(7528): pp. 1346.

[42] Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. *BMJ*, 1999. **318**(7188): pp. 887-8.

[43] History of the College. *The Royal College of Surgeons of Edinburgh*, 2008. Available from URL: http://www.rcsed.ac.uk/site/345/default.aspx

[44] Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Jones DB. Evaluating surgical competency with the American Board of Surgery In-Training Examination, skill testing, and intraoperative assessment. *Surgery*, 2000. **128**(4): pp. 613-22.

[45] Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*, 1997. **84**(2): pp. 273-8.

[46] Kopta JA. An approach to the evaluation of operative skills. *Surgery*, 1971. **70**(2): pp. 297-303.

[47] Likert R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932. **22**(140): pp. 1-55.

[48] Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg*, 2001. **193**(5): pp. 479-85.

[49] Smith SG, Torkington J, Brown TJ, Taffinder NJ, Darzi A. Motion analysis. *Surg Endosc*, 2002. **16**(4): pp. 640-5.

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

[51] Francis NK, Hanna GB, Cuschieri A. The performance of master surgeons on the Advanced Dundee Endoscopic Psychomotor Tester: contrast validity study. *Arch Surg*, 2002. **137**(7): pp. 841-4.

[52] Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg*, 2004. **91**(2): pp. 146-50.

[53] Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*, 2002. **236**(4): pp. 458-63.

[54] Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform*, 1998. **50**: pp. 124-30.

[55] Pedowitz RA, Esch J, Snyder S. Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy*, 2002. **18**(6): pp. E29.

[56] Smith S, Wan A, Taffinder N, Read S, Emery R, Darzi A. Early experience and validation work with Procedicus VA--the Prosolvia virtual reality shoulder arthroscopy trainer. *Stud Health Technol Inform*, 1999. **62**: pp. 337-43.

[57] Nielsen DM, Ricketts DM. Short-listing for orthopaedic specialist registrar posts--what is important? *Ann R Coll Surg Engl*, 2005. **87**(3): pp. 185-7.

[58] Bernstein AD, Jazrawi LM, Elbeshbeshy B, la Valle CJ, Zuckerman JD. Orthopaedic resident-selection criteria. *J Bone Joint Surg Am*, 2002. **84-A**(11): pp. 2090-6.

[59] Black KP, Abzug JM, Chinchilli VM. Orthopaedic in-training examination scores: a correlation with USMLE results. *J Bone Joint Surg Am*, 2006. **88**(3): pp. 671-6.

[60] McManus IC, Smithers E, Partridge P, Keeling A, Fleming PR. A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ*, 2003. **327**(7407): pp. 139-42.

[61] Tooke J. Aspiring to excellence: Final report of the independent inquiry into Modernising Medical Careers. London: MMC Inquiry; 2007.

[62] GMC. Tomorrow's doctors. London: General Medical Council; 2003.

[63] Tansley P, Kakar S, Withey S, Butler P. Visuospatial and technical ability in the selection and assessment of higher surgical trainees in the London deanery. *Ann R Coll Surg Engl*, 2007. **89**(6): pp. 591-5.

[64] Treisman AM, Gelade G. A feature-integration theory of attention. *Cognit Psychol*, 1980. **12**(1): pp. 97-136.

[65] Yang GZ, Dempere-Marco L, Hu XP, Rowe A. Visual search: psychophysical models and practical applications. *Image and Vision Computing*, 2002. **20**: pp. 291-305.

[66] Wolfe JM, Cave KR, Franzel SL. Guided search: an alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform*, 1989. **15**(3): pp. 419-33.

[67] Kundel HL, Nodine CF. The cognitive side of visual search. In: O'Regan JK, Levy-Schoen A., editors. Eye movements: from physiology to cognition. Elsevier; 1987. p. 573-82.

[68] Chase WG, Simon HA. Perception in Chess. *Cognit Psychol*, 1973. **4**(1): pp. 55-81.

[69] Reingold EM, Charness N, Pomplun M, Stampe DM. Visual span in expert chess players: evidence from eye movements. *Psychol Sci*, 2001. **12**(1): pp. 48-55.

[70] Reingold EM, Jolicoeur P. Perceptual versus postperceptual mediation of visual context effects: evidence from the letter-superiority effect. *Percept Psychophys*, 1993. **53**(2): pp. 166-78.

[71] Eckstein MP, Abbey CK, Pham BT, Shimozaki SS. Perceptual learning through optimization of attentional weighting: human versus optimal Bayesian learner. *J Vis*, 2004. **4**(12): pp. 1006-19.

[72] Sailer U, Flanagan JR, Johansson RS. Eye-hand coordination during learning of a novel visuomotor task. *J Neurosci*, 2005. **25**(39): pp. 8833-42.

[73] Flanagan JR, Bowman MC, Johansson RS. Control strategies in object manipulation tasks. *Curr Opin Neurobiol*, 2006. **16**(6): pp. 650-9.

[74] Magill RA. Motor learning and control : concepts and applications. 8th ed. New York: McGraw-Hill Higher Education; 2007.

[75] Fitts PM, Posner MI. Human performance. Belmont, USA: Brooks/Cole Pub. Co; 1967.

[76] Gentile AM. Skill acquisition: Action, movement and neuromotor processes. In: Carr JH, Shepherd RB, editors. Movement Sciences: Foundation for Physical Therapy in Rehabilitation. 2nd ed. Maryland: Aspen Press; 2000. p. 111-87.

[77] Land MF, McLeod P. From eye movements to actions: how batsmen hit the ball. *Nat Neurosci*, 2000. **3**(12): pp. 1340-5.

[78] Nagano T, Kato T, Fukuda T. Visual behaviors of soccer players while kicking with the inside of the foot. *Percept Mot Skills*, 2006. **102**(1): pp. 147-56.

[79] Rodrigues ST, Vickers JN, Williams AM. Head, eye and arm coordination in table tennis. *J Sports Sci*, 2002. **20**(3): pp. 187-200.

[80] Ghahramani Z, Wolpert DM, Jordan MI. Generalization to local remappings of the visuomotor coordinate transformation. *J Neurosci*, 1996. **16**(21): pp. 7085-96.

[81] Batista AP, Buneo CA, Snyder LH, Andersen RA. Reach plans in eye-centered coordinates. *Science*, 1999. **285**(5425): pp. 257-60.

[82] Krakauer JW, Pine ZM, Ghilardi MF, Ghez C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J Neurosci*, 2000. **20**(23): pp. 8916-24.

[83] McIntyre J, Stratta F, Lacquaniti F. Viewer-centered frame of reference for pointing to memorized targets in three-dimensional space. *J Neurophysiol*, 1997. **78**(3): pp. 1601-18.

[84] Flanagan JR, Nakano E, Imamizu H, Osu R, Yoshioka T, Kawato M. Composition and decomposition of internal models in motor learning under altered kinematic and dynamic environments. *J Neurosci*, 1999. **19**(20): pp. RC34.

[85] Weiss P, Jeannerod M. Getting a Grasp on Coordination. *News Physiol Sci*, 1998. **13**: pp. 70-5.

[86] Crawford JD, Medendorp WP, Marotta JJ. Spatial transformations for eye-hand coordination. *J Neurophysiol*, 2004. **92**(1): pp. 10-9.

[87] Tweed D, Cadera W, Vilis T. Computing three-dimensional eye position quaternions and eye velocity from search coil signals. *Vision Res*, 1990. **30**(1): pp. 97-110.

[88] Wong AM. Listing's Law: Clinical Significance and Implications for Neural Control. *Survey of Opthth*, 2004. **49**(6): pp. 563-75.

[89] Marotta JJ, Keith GP, Crawford JD. Task-specific sensorimotor adaptation to reversing prisms. *J Neurophysiol*, 2005. **93**(2): pp. 1104-10.

[90] Toussaint M. A sensorimotor map: modulating lateral interactions for anticipation and planning. *Neural Comput*, 2006. **18**(5): pp. 1132-55.

[91] Buneo CA, Jarvis MR, Batista AP, Andersen RA. Direct visuomotor transformations for reaching. *Nature*, 2002. **416**(6881): pp. 632-6.

[92] Ghilardi MF, Gordon J, Ghez C. Learning a visuomotor transformation in a local area of work space produces directional biases in other areas. *J Neurophysiol*, 1995. **73**(6): pp. 2535-9.

[93] Gordon J, Ghilardi MF, Ghez C. Accuracy of planar reaching movements. I. Independence of direction and extent variability. *Exp Brain Res*, 1994. **99**(1): pp. 97-111.

[94] Vindras P, Viviani P. Frames of reference and control parameters in visuomanual pointing. *J Exp Psychol Hum Percept Perform*, 1998. **24**(2): pp. 569-91.

[95] Vindras P, Desmurget M, Prablanc C, Viviani P. Pointing errors reflect biases in the perception of the initial hand position. *J Neurophysiol*, 1998. **79**(6): pp. 3290-4.

[96] Carrozzo M, McIntyre J, Zago M, Lacquaniti F. Viewer-centered and body-centered frames of reference in direct visuomotor transformations. *Exp Brain Res*, 1999. **129**(2): pp. 201-10.

[97] Desmurget M, Jordan M, Prablanc C, Jeannerod M. Constrained and unconstrained movements involve different control strategies. *J Neurophysiol*, 1997. **77**(3): pp. 1644-50.

[98] Klatzky RL. Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In: Freksa C, Habel C, Wender KF, editors. Spatial cognition - An interdisciplinary approach to representation and processing of spatial knowledge (Lecture Notes in Artificial Intelligence 1404).Berlin: Springer-Verlag; 1998. p. 1-17.

[99] Battaglia-Mayer A, Caminiti R, Lacquaniti F, Zago M. Multiple levels of representation of reaching in the parieto-frontal network. *Cereb Cortex*, 2003. **13**(10): pp. 1009-22.

[100] Eardley AF, Edwards G, Malouin F, Michon PE. Challenging the importance of vision for the development of an extrinsic spatial framework: evidence from the blind and sighted. *Cogn Process*, 2006. **7 Suppl 5**: pp. 30-1.

[101] Contreras-Vidal JL, Bo J, Boudreau JP, Clark JE. Development of visuomotor representations for hand movement in young children. *Exp Brain Res*, 2005. **162**(2): pp. 155-64.

[102] Olivier I, Hay L, Bard C, Fleury M. Age-related differences in the reaching and grasping coordination in children: unimanual and bimanual tasks. *Exp Brain Res*, 2007. **179**(1): pp. 17-27.

[103] Land MF. Eye movements and the control of actions in everyday life. *Prog Retin Eye Res*, 2006. **25**(3): pp. 296-324.

[104] Land MF, Lee DN. Where we look when we steer. *Nature*, 1994. **369**(6483): pp. 742-4.

[105] Hager GD, Wen-Chung C, Morse AS. Robot feedback control based on stereo vision: towards calibration-free hand-eye coordination. Proceedings of the IEEE International Conference on Robotics and Automation, 1994. p. 2850-6.

[106] Miall RC, Reckess GZ. The cerebellum and the timing of coordinated eye and hand tracking. *Brain Cogn*, 2002. **48**(1): pp. 212-26.

[107] Land MF, Furneaux S. The knowledge base of the oculomotor system. *Philos Trans R Soc Lond B Biol Sci*, 1997. **352**(1358): pp. 1231-9.

[108] Johansson RS, Westling G, Backstrom A, Flanagan JR. Eye-hand coordination in object manipulation. *J Neurosci*, 2001. **21**(17): pp. 6917-32.

[109] Rotman G, Troje NF, Johansson RS, Flanagan JR. Eye movements when observing predictable and unpredictable actions. *J Neurophysiol*, 2006. **96**(3): pp. 1358-69.

[110] Flanagan JR, Johansson RS. Action plans used in action observation. *Nature*, 2003. **424**(6950): pp. 769-71.

[111] Gowen E, Miall RC. Eye-hand interactions in tracing and drawing tasks. *Hum Mov Sci*, 2006. **25**(4-5): pp. 568-85.

[112] Doolin EJ, Strande L. Calibration of endoscopic images. *Ann Otol Rhinol Laryngol*, 1995. **104**(1): pp. 19-23.

[113] Hanna GB, Shimi SM, Cuschieri A. Task performance in endoscopic surgery is influenced by location of the image display. *Ann Surg*, 1998. **227**(4): pp. 481-4.

[114] Nicolaou M. Determining laparoscopic strategies of novices with eye tracking. Medical Image Perception Society, Sep 2005; Bowness on Windermere, England 2005.

[115] Osterberg G. Topography of the layer of rods and cones in the human retina. *Acta Ophthal*, 1935. **S6**: pp. 1-103.

[116] Ballard DH, Hayhoe MM, Li F, Whitehead SD. Hand-eye coordination during sequential tasks. *Philos Trans R Soc Lond B Biol Sci*, 1992. **337**(1281): pp. 331-8.

[117] Liversedge SP, Findlay JM. Saccadic eye movements and cognition. *Trends Cogn Sci*, 2000. **4**(1): pp. 6-14.

[118] Zhaoping L, Herzog MH, Dayan P. Nonlinear ideal observation and recurrent preprocessing in perceptual learning. *Network*, 2003. **14**(2): pp. 233-47.

[119] Dodge R, Cline TS. The angle velocity of eye-movements. *Psychological Review*, 1901. **8**: pp. 145-57.

[120] Lipps M, Pelz JB. Yarbus revisited: task-dependent oculomotor behvior. *Journal of Vision*, 2004. **4**(8): pp. 115a.

[121] Duchowski AT. Eye tracking methodology: theory and practice. New York: 2003.

[122] Hanna GB, Drew T, Clinch P, Hunter B, Cuschieri A. Computer-controlled endoscopic performance assessment system. *Surg Endosc*, 1998. **12**(7): pp. 997-1000.

[123] Meskers CG, Fraterman H, van der Helm FC, Vermeulen HM, Rozing PM. Calibration of the "Flock of Birds" electromagnetic tracking device and its application in shoulder motion studies. *J Biomech*, 1999. **32**(6): pp. 629-33.

[124] Polaris Spectra & Polaris Vicra Technical Specifications. *Northern Digital Inc*, Jan 2009. Available from URL: http://www.ndigital.com/medical/polarisfamily-techspecs.php

[125] Optotrak Certus Motion Capture System: Technical Specifications. *Northern Digital Inc*, Jan 2009. Available from URL: http://www.ndigital.com/lifesciences/certus-techspecs.php

[126] Heikkila J, Silven O. A four-step camera calibration procedure with implicit image correction. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. p. 1106-12.

[127] Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations. The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999. p. 666-73.

[128] Bouguet JY. Camera Calibration Toolbox for Matlab [computer program]. 2007. http://www.vision.caltech.edu/bouguetj/calib_doc/

[129] Tsai RY, Lenz RK. Real time versatile robotics hand/eye calibration using 3D machine vision. Proceedings of the IEEE International Conference on Robotics and Automation, 1988. p. 554-61.

[130] Tsai RY, Lenz RK. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 1989. **5**(3): pp. 345-58.

[131] Litynski GS. Profiles in laparoscopy: Mouret, Dubois, and Perissat: the laparoscopic breakthrough in Europe (1987-1988). *JSLS*, 1999. **3**(2): pp. 163-7.

[132] Moorthy K, Munz Y, Dosis A, Hernandez J, Martin S, Bello F, Rockall T, Darzi A. Dexterity enhancement with robotic surgery. *Surg Endosc*, 2004. **18**(5): pp. 790-5.

[133] den Boer KT, Herder JL, Sjoerdsma W, Meijer DW, Gouma DJ, Stassen HG. Sensitivity of laparoscopic dissectors. What can you feel? *Surg Endosc*, 1999. **13**(9): pp. 869-73.

[134] Jordan JA, Gallagher AG, McGuigan J, McGlade K, McClure N. A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. *Am J Surg*, 2000. **180**(3): pp. 208-11.

[135] Neumayer LA, Gawande AA, Wang J, Giobbie-Hurder A, Itani KM, Fitzgibbons RJ, Jr., Reda D, Jonasson O. Proficiency of surgeons in inguinal hernia repair: effect of experience and age. *Ann Surg*, 2005. **242**(3): pp. 344-8.

[136] Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *BMJ*, 2003. **327**(7422): pp. 1032-7.

[137] Nel EM, du Preez JA, Herbst BM. Estimating the pen trajectories of static signatures using hidden Markov models. *IEEE Trans Pattern Anal Mach Intell*, 2005. **27**(11): pp. 1733-46.

[138] Rose RC, Juang BH. Hidden Markov models for speech and signal recognition. *Electroencephalogr Clin Neurophysiol Suppl*, 1996. **45**: pp. 137-52.

[139] Shanableh T, Assaleh K, Al-Rousan M. Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language. *IEEE Trans Syst Man Cybern B Cybern*, 2007. **37**(3): pp. 641-50.

[140] Rosen J, Solazzo M, Hannaford B, Sinanan M. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. *Comput Aided Surg*, 2002. **7**(1): pp. 49-61.

[141] Nicolaou M, James A, Darzi A, Yang GZ. A Study of Saccade Transition for Attention Segregation and Task Strategy in Laparoscopic Surgery. In: Bariollot C, Haynor DR, Hellier P, editors. Lecture Notes in Computer Science. 3217 ed. St Malo, France: 2004. p. 97-104.

[142] Bashir FI, Khokhar AA, Schonfeld D. View-invariant motion trajectory-based activity classification and recognition. *Multimedia Systems*, 2006. **12**(1): pp. 45-54.

[143] Vlachos M, Gunopulos D, Das G. Rotation invariant distance measures for trajectories. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining; Seattle, USA: ACM Press; 2004. p. 707-12.

[144] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989. **77**(2): pp. 257-86.

[145] Keogh EJ, Pazzani MJ. Derivative Dynamic Time Warping. 1st SIAM International Conference on Data Mining (SDM'2001), 2001.

[146] Akaho S. The EM algorithm for multiple object recognition. *Proceedings of IEEE International Conference on Neural Networks (ICNN'95)*, 1995. **5**: pp. 2426-31.

[147] Bland JM, Altman DG. Cronbach's alpha. *BMJ*, 1997. **314**(7080): pp. 572.

[148] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL. Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. *J Bone Joint Surg Br*, 2008. **90**(4): pp. 494-9.

[149] Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *BJOG*, 2008. **115**(7): pp. 908-16.

[150] Ezra DG, Aggarwal R, Michaelides M, Okhravi N, Verma S, Benjamin L, Bloom P, Darzi A, Sullivan P. Skills acquisition and assessment after a microsurgical skills course for ophthalmology residents. *Ophthalmology*, 2009. **116**(2): pp. 257-62.

[151] Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Ann Surg*, 2008. **247**(2): pp. 372-9.

[152] Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA. Ceiling effect in technical skills of surgical residents. *Am J Surg*, 2004. **188**(3): pp. 294-300.

[153] Chen W, Chang SF. Motion Trajectory matching of Video Objects. Proceedings of SPIE Storage and Retrieval for Media Databases, 2000. p. 544-53.

[154] Dahlhaus R, Eichler M. Causality and graphical models in time series analysis. In: Green N, Hjort N, Richardson S, editors. Highly structured stochastic systems. University Press, Oxford; 2003.

[155] Eichler M. Fitting graphical interaction models to multivariate time series. Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, AUAI Press, Virginia; 2006.

[156] Baccala LA, Sameshima K. Partial directed coherence: a new concept in neural structure determination. *Biol Cybern*, 2001. **84**(6): pp. 463-74.

[157] Baccala LA, Sameshima K. Partial directed coherence and neuronal connectivity inference. Proceedings of the 25th Annual International Conference of the IEEE, 2003. p. 2151.

[158] Schelter B, Winterhalder M, Eichler M, Peifer M, Hellwig B, Guschlbauer B, Lucking CH, Dahlhaus R, Timmer J. Testing for directed influences among neural signals using partial directed coherence. *J Neurosci Methods*, 2006. **152**(1-2): pp. 210-9.

[159] Lutkepohl H. Comparison of criteria for estimating the order of a vector autoregressive process. *J Time Ser Anal*, 1985. **6**: pp. 35-52.

[160] Schneider T, Neumaier A. ARfit: A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models [computer program]. 2001. http://www.gps.caltech.edu/~tapio/arfit/

[161] Kullback S, Leibler RA. On information and sufficiency. *The Annals of Mathematical Statistics*, 1951. **1**(22): pp. 79-86.

[162] Tsai R. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 1987. **3**(4): pp. 323-44.

[163] Fuglede B, Topsoe F. Jensen-Shannon divergence and Hilbert space embedding. Proceedings of the International Symposium on Information Theory, 2004. p. 31.

[164] Lin J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 1991. **37**(1): pp. 145-51.

[165] Topsoe F. Some inequalities for information divergence and related measures of discrimination. *IEEE Transactions on Information Theory*, 2000. **46**(4): pp. 1602-9.

[166] Melville P, Yang S, Saar-Tsechansky M, Mooney R. Active Learning for Probability Estimation Using Jensen-Shannon Divergence. Proceedings of the 16th European Conference on Machine Learning, 2005. p. 268-79.

[167] Endres DM, Schindelin JE. A new metric for probability distributions. *IEEE Transactions on Information Theory*, 2003. **49**(7): pp. 1858-60.

[168] Stoyanov D, Mylonas GP, Deligianni F, Darzi A, Yang GZ. Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 2005. **8**(Pt 2): pp. 139-46.

[169] Mylonas GP, Darzi A, Yang GZ. Gaze-contingent control for minimally invasive robotic surgery. *Comput Aided Surg*, 2006. **11**(5): pp. 256-66.

[170] Mylonas GP, Kwok KW, Darzi A, Yang GZ. Gaze-contingent motor channelling and haptic constraints for minimally invasive robotic surgery. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 2008. **11**(Pt 2): pp. 676-83.

[171] Lerotic M, Chung AJ, Mylonas G, Yang GZ. Pq-space based non-photorealistic rendering for augmented reality. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 2007. **10**(Pt 2): pp. 102-9.

[172] Swanstrom L, Zheng B. Spatial orientation and off-axis challenges for NOTES. *Gastrointest Endosc Clin N Am*, 2008. **18**(2): pp. 315-24.

[173] Zacharakis E, Purkayastha S, Teare J, Yang GZ, Darzi A. Natural orifices translumenal endoscopic surgery (NOTES)--who should perform it? *Surgery*, 2008. **144**(1): pp. 1-2.

[174] Jobsis FF. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*, 1977. **198**(4323): pp. 1264-7.

[175] Koh PH, Glaser DE, Flandin G, Kiebel S, Butterworth B, Maki A, Delpy DT, Elwell CE. Functional optical signal analysis: a software tool for near-infrared spectroscopy data processing incorporating statistical parametric mapping. *J Biomed Opt*, 2007. **12**(6): pp. 064010.

[176] Leff DR, Orihuela-Espina F, Atallah L, Athanasiou T, Leong JJ, Darzi AW, Yang GZ. Functional prefrontal reorganization accompanies learning-associated refinements in surgery: a manifold embedding approach. *Comput Aided Surg*, 2008. **13**(6): pp. 325-39.

[177] Leff DR, Elwell CE, Orihuela-Espina F, Atallah L, Delpy DT, Darzi AW, Yang GZ. Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual co-ordination task: an fNIRS study. *Neuroimage*, 2008. **39**(2): pp. 805-13.

[178] Benger JR, Lyburn ID. What is the effect of reporting all emergency department radiographs? *Emerg Med J*, 2003. **20**(1): pp. 40-3.

[179] Williams SM, Connelly DJ, Wadsworth S, Wilson DJ. Radiological review of accident and emergency radiographs: a 1-year audit. *Clin Radiol*, 2000. **55**(11): pp. 861-5.

[180] Rhea JT, Potsaid MS, DeLuca SA. Errors of interpretation as elicited by a quality audit of an emergency radiology facility. *Radiology*, 1979. **132**(2): pp. 277-80.

[181] Tachakra S. Level of diagnostic confidence, accuracy, and reasons for mistakes in teleradiology for minor injuries. *Telemed J E Health*, 2002. **8**(1): pp. 111-21.

[182] Eng J, Mysko WK, Weller GE, Renard R, Gitlin JN, Bluemke DA, Magid D, Kelen GD, Scott WW, Jr. Interpretation of Emergency Department radiographs: a comparison of emergency medicine physicians with radiologists, residents with faculty, and film with digital display. *AJR Am J Roentgenol*, 2000. **175**(5): pp. 1233-8.

[183] Cimmino CV. The radiologist and the orthopedist. *Radiology*, 1970. **97**(3): pp. 690-1.

[184] Anglen J, Marberry K, Gehrke J. The clinical utility of duplicate readings for musculoskeletal radiographs. *Orthopedics*, 1997. **20**(11): pp. 1015-9.

[185] Krupinski EA, Berger WG, Dallas WJ, Roehrig H. Searching for nodules: what features attract attention and influence detection? *Acad Radiol*, 2003. **10**(8): pp. 861-8.

[186] Krupinski EA. Visual search of mammographic images: influence of lesion subtlety. *Acad Radiol*, 2005. **12**(8): pp. 965-9.

[187] Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Invest Radiol*, 1978. **13**(3): pp. 175-81.

[188] Kundel HL, Nodine CF, Krupinski EA. Searching for lung nodules. Visual dwell indicates locations of false-positive and false-negative decisions. *Invest Radiol*, 1989. **24**(6): pp. 472-8.

[189] Nodine CF, Mello-Thoms C, Weinstein SP, Kundel HL, Conant EF, Heller-Savoy RE, Rowlings SE, Birnbaum JA. Blinded review of retrospectively visible unreported breast cancers: an eye-position analysis. *Radiology*, 2001. **221**(1): pp. 122-9.

[190] Dempster A, Laird N, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society B*, 1977. **39**(1): pp. 1-38.

[191] van Onselen EB, Karim RB, Hage JJ, Ritt MJ. Prevalence and distribution of hand fractures. *J Hand Surg [Br]*, 2003. **28**(5): pp. 491-5.

[192] Dempere-Marco L, Hu XP, MacDonald SL, Ellis SM, Hansell DM, Yang GZ. The use of visual search for knowledge gathering in image decision support. *IEEE Trans Med Imaging*, 2002. **21**(7): pp. 741-54.

[193] Hu XP, Dempere-Marco L, Yang GZ. Hot spot detection based on feature space representation of visual search. *IEEE Trans Med Imaging*, 2003. **22**(9): pp. 1152-62.

[194] Hu CH, Kundel HL, Nodine CF, Krupinski EA, Toto LC. Searching for bone fractures: a comparison with pulmonary nodule search. *Acad Radiol*, 1994. **1**(1): pp. 25-32.

[195] Berbaum KS, Brandser EA, Franken EA, Dorfman DD, Caldwell RT, Krupinski EA. Gaze dwell times on acute trauma injuries missed because of satisfaction of search. *Acad Radiol*, 2001. **8**(4): pp. 304-14.

[196] Krupinski EA, Lund PJ. Differences in time to interpretation for evaluation of bone radiographs with monitor and film viewing. *Acad Radiol*, 1997. **4**(3): pp. 177-82.

[197] Lund PJ, Krupinski EA, Pereles S, Mockbee B. Comparison of conventional and computed radiography: assessment of image quality and reader performance in skeletal extremity trauma. *Acad Radiol*, 1997. **4**(8): pp. 570-6.

[198] van Zoest W, Donk M. Bottom-up and top-down control in visual search. *Perception*, 2004. **33**(8): pp. 927-37.