Mechatronics in Medicine Research Group

Computer Vision Techniques for a Robot-Assisted Emergency Neurosurgery System

by

Mark Ovinis

Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of

Doctor of Philosophy

Loughborough University

June 2011

To my wife, Saineh, my daughter, Alexis, and my son, Jonathan

Acknowledgements

To my supervisors, Dr David Kerr and Dr Kaddour Bouazza-Marouf, for giving me the opportunity to pursue a PhD, and for your constant guidance throughout, I owe you both my deepest gratitude. I am hopeful that MISEN will one day make the transition from the research laboratory to the operating theatre, and change the medical landscape for the provision of emergency neurosurgical services.

I would also like to acknowledge gratefully the financial support from Universiti Teknologi Petronas, Malaysia, and Loughborough University, UK, without which my PhD studies would not have been possible.

To my collaborators in this project, Dr Michael O'Toole and Dr Mahendra Gooroochurn, who did much of the preliminary groundwork that enabled me to hit the ground running, thank you for this and for your friendship.

Special thanks go to the radiographic staff at Johor Specialist Centre, Malaysia, in particular Dr Muhamad Razif Ibrahim, who kindly provided me with a large dataset of CT images for use in my research, and to Prof. Michael Vloeberghs, for his advice on the neurosurgical aspects of the project.

Last but by no means least; I would like to acknowledge my parents, for instilling in me the value of education. I hope that I have made the both of you proud.

Copyright Agreements

"Portions of the research in this thesis use CT and MRI image data sets collected under the Visible Human Project sponsored by the National Library of Medicine, Department of Health and Human Services, USA”

CERTIFICATE OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgments or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.

…… (Signed)

…… (Date)

Abstract

This thesis presents the development of computer vision techniques for a robot-assisted emergency neurosurgery system that is being developed by the Mechatronics in Medicine group at Loughborough University, UK, and situates them within the context of the overall project. There are two main contributions in this thesis. The first is the development of a registration framework, to establish spatial correspondence between a preoperative plan of a patient (based on computed tomography images) and the patient. The registration is based on the rigid transformation of homologous anatomical soft tissue point landmarks of the head, the medial canthus and tragus, in CT and patient space. As a step towards automating the registration, a computational framework to localise these landmarks in CT space, with performance comparable to manual localisation, has been developed. The second contribution in this thesis is the development of computer vision techniques for a passive intraoperative supervisory system, based on visual cues from the operative site. Specifically, the feasibility of using computer vision to assess the outcome of a surgical intervention was investigated. The ability to mimic and embody part of a surgeon's visual sensory and decision-making capability is aimed at improving the robustness of the robotic system. Low-level image features to distinguish the two possible outcomes, complete and incomplete, were identified. Encouraging results were obtained for the surgical actions under consideration, which have been demonstrated by experiments on cadaveric pig heads. The results obtained are suggestive of the potential use of computer vision to assist surgical robotics in an operating theatre. The computational approaches developed, to provide greater autonomy to the robotic system, have the potential to improve current practice in robotic surgery. It is not inconceivable that the state of the art in surgical robotics can advance to a stage where it is able to emulate the ability and interpretation process of a surgeon.

Abbreviations

CCD: Charge-Coupled Device
CIE: Commission Internationale de l'Eclairage (International Commission on Illumination)
CIELAB: CIE 1976 (L*, a*, b*) colour space
CPP: Cerebral Perfusion Pressure
CMM: Coordinate Measurement Machine
CNC: Computer Numerical Control
CSF: Cerebrospinal Fluid
CT: Computed Tomography
DGH: District general hospital
FDA: Food and Drug Administration
FLE: Fiducial Localisation Error
FRE: Fiducial Registration Error
FTE: Full Time Equivalent
GCS: Glasgow Coma Scale
GDG: Guideline Development Group
GLCM: Grey-Level Co-occurrence Matrix
HES: Hospital Episode Statistics
HSI: Hue, Saturation and Intensity
HU: Hounsfield Unit
ICP: Intracranial Pressure
LDA: Linear Discriminant Analysis
LED: Light Emitting Diode
LOOCV: Leave-One-Out Cross-Validation
MISEN: Mechatronic Intervention System for Emergency Neurosurgery
MRI: Magnetic Resonance Imaging
NCEPOD: National Confidential Enquiry into Patient Outcome and Death
NHS: National Health Service
NICE: National Institute for Health and Clinical Excellence
OPCS: Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures
PCA: Principal Component Analysis

QDA: Quadratic Discriminant Analysis

RGB: Red, green and blue

RMS: Root Mean Square

RNU: Regional Neurosciences Units

SBNS: Society of British Neurological Surgeons

TBI: Traumatic Brain Injury

TRE: Target Registration Error

Symbols

I: Transmitted intensity

I_o: Incident intensity

μ: Attenuation coefficient

Δx: Voxel thickness

FRE: Fiducial registration error

$\mathcal{T}$: Transformation matrix

x: x coordinate

y: y coordinate

FLE: Fiducial localisation error

N: Number of fiducials

TRE: Target registration error

d_k: Distance of the target from principal axis, k

f_k: Distance of the fiducials from principal axis, k

ω: Non-negative weighting factor

R: Rotation matrix

t: Translation matrix

$\tilde{x}_i$: Mean centred x coordinate

$\tilde{y}_i$: Mean centred y coordinate

$\bar{x}$: x coordinate of weighted centroid of the fiducial configuration

$\bar{y}$: y coordinate weighted centroid of the fiducial configuration

H: Weighted fiducial covariance matrix

U: Unitary matrix

Λ: Diagonal matrix with nonnegative real numbers on the diagonal

V^t: Conjugate transpose of unitary matrix V

I: Identity matrix

λ: Eigenvalue

θ: Orientation

μ_{pq}: $(p+q)$ order central moment

e: Edge of a triangular mesh

$\Im(v)$: Curvature tensor at a vertex on a triangular mesh

v: Vertex on a triangular mesh

B: Region on a triangular mesh

$|B|$: Surface area around a vertex on a triangular mesh over a curvature tensor is estimated

$\beta(e)$: Signed angle between normals to two oriented triangles incident to edge of a triangular mesh

$|e \cap B|$: Length of an edge of a triangular mesh within a region

$\bar{e}$: Unit vector of an edge of a triangular mesh in the same direction as the edge

$C_{gaussian}$: Gaussian curvature

κ_{min}: Minimum principal curvature

κ_{max}: Maximum principal curvature

C_{mean}: Mean curvature

μ: Mean

σ: Standard deviation

$p(x|w)$: Probability that x comes from class w

C: Covariance matrix

m: Mean vector

d: Decision function

$P(w)$: Likelihood of class w appearing

ΔE_{ab^*}: CIE delta-E colour difference

L^*: Coordinate in CIELAB colour space representing lightness of colour

a^*: Coordinate in CIELAB colour space representing position between red/magenta and green

b^*: Coordinate in CIELAB colour space representing position between yellow and blue

I: Greyscale intensity image

Δ: Laplacian

∇: Gradient

div: Divergence operator

c: Conduction coefficient

List of publications

Conference papers (see Appendix A)

1. M. Ovinis, D. Kerr, K. Bouazza-Marouf, M. Vloeberghs, "Machine Vision for the Inspection of Surgical Tasks: Applications to Robotic Surgery Systems", In Proceedings of the International Conference on Machine Vision, Amsterdam, The Netherlands, 2010, vol. 70, pp. 338-344, ISSN 1307-6892

This publication describes the use of machine vision to inspect the outcome of surgical tasks for a common neurosurgical procedure and presents the experimental results. Subsequent work is presented in this thesis, based on the ideas presented in this paper. The corresponding work is presented in Chapter 6.

2. M. Ovinis, D. Kerr, K. Bouazza-Marouf, M. Vloeberghs, "Localisation of Anatomical Soft Tissue Landmarks of the Head in CT Images", In Proceedings of the International Conference on Machine Vision, Amsterdam, The Netherlands, 2010, vol. 70, pp. 752-758, ISSN 1307-6892

This publication describes a computational approach to the automatic localisation of anatomical soft tissue landmarks of the head in CT images, for an automated image to patient registration system. The corresponding work is presented in Chapter 4.

Journal Paper

1. M. Ovinis, D. Kerr, K. Bouazza-Marouf, “The automatic localisation of the medial canthus and tragus in CT images for image to patient registration”, submitted to Computerized Medical Imaging and Graphics.

This Publication presents further results obtained in the second publication. Specifically, a curve skeleton extraction algorithm was incorporated to improve localisation accuracy. The corresponding work is also presented in Chapter 4.

Table of Contents

List of figures

List of tables

Glossary of Terms

Anterior	Nearer the forward end, front of a human
Axial Plane	A plane roughly perpendicular to spine that divides the body into upper and lower parts
Burr hole	Small window in the skull made to access the brain
Canthus	Either corner of the eye, where the eyelids meet
Catheter	A small tube inserted into a body cavity
Cauterise	To burn, sear, or freeze tissue
Cerebrospinal Fluid	A clear bodily fluid that occupies the subarachnoid space in the brain
Cerebrum	The upper part of the brain, which is divided into the two cerebral hemispheres
Coronal Plane	A vertical plane roughly parallel to spine that divides the body into anterior and posterior sections
Coronal Suture	The line of junction between the frontal bone and the two parietal bones of the skull
Dura Mater	The tough and inflexible outermost of the three layers of the meninges
Epidural	Situated on or outside the dura mater
Glasgow Coma Scale	A neurological scale to assess level of consciousness of acute medical and trauma patients
Haematoma	A swelling of blood, usually clotted, which forms as a result of a broken blood vessel
Intracerebral	Occurring or situated within the cerebrum
Intraparenchymal	Within the brain
Kocher's point	A common entry point for an intraventricular catheter to drain cerebral spinal fluid from the cerebral ventricles
Lateral	Pertaining to the left or right of the body; further from the midline
Medial	Pertaining to the inside; closer to the midline
Meninges	The three membranes that envelop the brain and spinal cord
Oedema	An excessive accumulation of serum in tissue spaces or a body cavity
Pericranium	A membrane that covers the outer surface of the skull
Periosteum	A membrane surrounding a bone
Pia mater	Inner most of the meninges membranes that surround the brain and spinal cord
Posterior	Nearer the back end; nearer the dorsal end in humans
Radiography	The process of making an image, often a photographic negative, produced by radiation other than normal light
Sagittal Plane	A plane roughly parallel to the spine that divides the body in left and right sections

Sagittal Suture	A dense, fibrous connective tissue joint between the two parietal bones of the skull
Subarachnoid	Located or occurring below the arachnoid membrane, often specifically between the arachnoid membrane and the pia mater
Subcutaneous fat	Fat found beneath the dermis layer of the skin
Subdural	Located beneath the dura mater and above the meninges.
Suture	To sew up or join edges of skin usually by means of a thread
Tragus	The small piece of thick cartilage of the external ear that is immediately in front of the ear canal
Trauma	Any serious injury to the body
Ventricle	One of four cavities in the brain

Chapter 1

Introduction

“No head injury is so serious that it should be despaired of nor so trivial that it can be ignored.”

Hippocrates

1.1 Background

Head injury is significant cause of fatality among trauma patients. In a prospective study [1] of 22,216 major trauma patients with head injury i.e. patients with abnormal CT brain scan and those with skull fractures, between 1989 and 2003, the overall mortality rate was 44%. This mortality rate is ten times the mortality rate for major trauma patients without head injury. Although patient outcome after a head injury depends largely on the severity of the initial injury, the management of the injury and any subsequent complications is important for a good outcome [2].

In patients with head injuries, neurosurgery is aimed at preventing or minimising irreversible damage to the brain. There are two types of demand for neurosurgery. The first is emergency, urgent and trauma, the second is demand for neurosurgery that is not immediately life threatening. Two thirds of the neurosurgical workload in the United Kingdom (UK) and Ireland is emergency, urgent and trauma [3], a demand which is not always met. A report by the National Confidential Enquiry into Patient Outcome and Death (NCEPOD) 2007 [2] highlights some of the inconsistencies in the provision of emergency neurosurgical services in the UK. The following case studies [4] are examples of the poor management of some patients with severe head injuries and symptomatic of the scarcity of neurosurgical resources in the UK:

Case study 1

“A young patient was admitted to hospital after sustaining a head injury. Admission GCS [Glasgow Coma Scale] was 5. CT [Computed Tomography] head scanning revealed a large intracerebral haemorrhage, cerebral oedema and midline shift. Referral was made to the regional neurosurgical centre at 01:15. The neurosurgical SpR [Specialist Registrar] stated that there were no beds in the ICU [Intensive Care Unit] and advised that the patient be referred to

another neurosurgical centre. After some difficulty in contacting the on call neurosurgeon, the patient was discussed with the neurosurgical SpR at a second centre. This occurred at 02:15. The neurosurgical SpR stated that they had beds available and would be prepared to take the patient but that the opinion of whether transfer should occur should be made by the local neurosurgical centre that had no capacity to take the patient. A further telephone call was made to the local neurosurgical centre at 02:30 who felt that as they had no capacity they could not comment on the patient care. At 03:00 a second telephone call was made to the second neurosurgical centre and at that point the neurosurgical SpR requested that hard copies of the CT be sent to him. Six more telephone calls occurred over the next few hours between the initial hospital and three different neurosurgical centres. By 06:30 the patient, who was still in the emergency department, had fixed and dilated pupils and it was believed that death was the inevitable outcome..." [2]

Case study 2

"A teenager was involved in a road traffic accident. On admission, they had a Glasgow Coma Score (GCS) of 14/15. A CT scan demonstrated a subdural haematoma. An emergency department specialist registrar discussed the patient with a neurosurgical SpR and a further CT was ordered. Transfer was not accepted despite deterioration in the patient's GCS to 12/15 over the next two hours. Following a further deterioration over another hour to GCS 8/15 the patient was intubated and following further discussion with a neurosurgical specialist registrar a third CT scan was ordered. During the scan, the patient's endotracheal tube became blocked and the patient became hypoxic which lead to raised intracranial pressure. Thirty six hours later the patient was declared brain dead and ventilation withdrawn" [5]

As exemplified by these case studies, when demand for neurosurgery, in particular emergency neurosurgery, is not met, the consequences can be fatal. An expert review of the case study 2 questioned whether scarce neurosurgical resources were a contributing factor. A snapshot audit [6] of 223 critical care units in 187 hospitals and 24 regional neurosciences units (RNUs) in England and Wales showed that demand for neurocritical care has outstripped supply by

9% - 22%, and stated that '*…. in some cases at least, access to neurocritical care is determined by capacity rather than by clinical need*'. As a step towards improving the supply side of the equation, the Mechatronics in Medicine group at Loughborough University, UK, is developing a robot-assisted system for emergency neurosurgery where immediate surgery is required, but for various reasons, the patient cannot be operated on conventionally by a neurosurgeon. This robot-assisted system, the Mechatronic Intervention System for Emergency Neurosurgery (MISEN), is conceived as a supervisory-controlled system, where the robot performs the surgery autonomously based on a preoperative plan by an off-site neurosurgeon, with an on-site surgeon, usually without specialist neurosurgical experience, providing some degree of monitoring and control.

1.2 Image to patient registration

An often-overlooked aspect of robotic surgery is that the surgical intervention by the robot is only one aspect of the surgery. There are typically three phases for robotic surgery: (a) preoperative planning, (b) intraoperative surgical intervention, (c) postoperative assessment [7]. In the preoperative planning phase, a surgeon uses a computer to plan the surgical intervention to provide guidance to the robot. This preoperative plan, based on a 3D model of the patient reconstructed from medical images such as CT or magnetic resonance imaging (MRI), usually specifies an entry point, target point(s), and a trajectory. To establish spatial correspondence between the preoperative plan and the patient, the preoperative plan has to be 'aligned' to the patient using a technique known as image to patient registration. The output of the image to patient registration is a geometric transformation that will relate 'medical image space' and 'patient space'. This geometric transformation will provide a one-to-one mapping of points in the CT/MRI image coordinate system to points in the patient coordinate system, enabling the robot to follow a trajectory on the patient based on the preoperative plan.

1.3 Intraoperative assessment

As MISEN would be performing actions that would otherwise be carried out by a neurosurgeon, the ability to assess the progress of the operation is an important aspect of the surgical intervention. One possible solution is to mimic and embody a surgeon's visual sensory and decision-making capability using computer vision. As a step towards this goal, the use of visual cues to inspect the outcome of surgical actions is investigated. The incorporation of visual feedback is intended to improve the reliability and robustness of robotic surgery systems by improving its perception. The use of computer vision should therefore not be viewed in isolation.

1.4 Objectives

This thesis has two objectives. The first is to develop an automatic image to patient registration framework for MISEN that is sufficiently accurate for the targeted neurosurgical interventions. For the registration, the error in localising a target anywhere in a head should be within required clinical accuracy of the procedures. Because the targeted neurosurgical procedures are normally performed using a freehand technique i.e. without image-guidance, the accuracy requirements for the procedures are within the centimetre range, similar to that achieved by neurosurgeons. The goal is to develop a registration framework that is capable of sub-centimetre accuracy.

The second objective is to investigate a computer vision based approach to facilitate the surgical intervention by the robot. The aim is to use computer vision to augment/complement other sensory information e.g. force, for a more comprehensive assessment of a surgical action. When multiple sensory inputs are available, various control schemes such as traded, hybrid and shared control are commonly used. However, the type of control scheme used is beyond the scope of this thesis.

Both the registration framework and computer vision system should be cost effective and practical. In terms of cost-effectiveness, the use of specialised hardware, such as the use of projectors to provide structured

lighting or the use of 3D laser scanners, which is prohibitive in terms of cost, should be avoided. In terms of practicality, there should not be unrealistic requirements on the robotic system in terms of its overall architecture and design. Existing workflow should not be modified. For example, in CT imaging, a secondary scan should not be required.

1.5 Contribution

There are two main contributions in this thesis. The first is the development of a retrospective image to patient registration framework based on a rigid body transformation of two homologous anatomical soft tissue landmarks of the head, the medial canthus and tragus, in CT and patient space. The proposed framework avoids resorting to the use of fiducial markers, which is the method of choice for image to patient registration but is a prospective technique that requires a secondary scan. In contrast to surface-based registration methods, which do not require a secondary scan, the use of anatomical landmarks (if judiciously chosen and correctly localised) is a potentially more accurate method. It also avoids the use of expensive 3D scanners needed to obtain surface scans of the patients. As a step towards automating the registration, a computational framework to localise these landmarks in CT images, with performance comparable to manual localisation, has been developed.

The second contribution in this thesis is the development of computer vision techniques for the visual inspection of the outcome of surgical actions with applications to robotic surgery systems. To the author's knowledge, the use of vision in this context has never been considered before and this work represents the first step in this direction. Admittedly, the surgical environment can be highly unstructured, in contrast to the more controlled conditions typically found where computer vision is more often applied. However, because the surgical actions being inspected are intended to be performed robotically, their context is known and visual inspection is feasible. The use of computer vision in this context, to provide greater situational awareness to the robotic system and confidence to the end user, has been the goal of this research.

1.6 Thesis structure

There are seven chapters in this thesis. Chapter 2 describes the motivation behind developing a neurosurgical robotic system and introduces MISEN. Chapter 3 gives a background on image to patient registration, and presents the proposed registration framework for MISEN. Chapter 4 presents a computational approach towards automating the registration. Chapter 5 proposes the use of computer vision for the inspection of the outcome of surgical actions, as a step towards greater autonomy in robotic surgery systems. Chapter 6 describes the experiments and results on the use of computer vision for this purpose. Chapter 7 concludes with a summary and areas for future work.

Chapter 2

A Mechatronic Intervention System for Emergency Neurosurgery (MISEN)

"Our future in surgery lies not in blood and guts, but in bits and bytes!"

Colonel Richard M. Satava, one of the pioneers of robotic telesurgery

2.1 Background

In the UK, patients with head injuries are usually admitted to a hospital nearest to the scene where the injuries were sustained, typically a district general hospital (DGH) and frequently with no neurosurgery department [8]. The majority of DGHs do not have neurosurgical services, with neurosurgery usually performed at one of the 35 RNUs around the country. A survey [9] of 161 hospitals by the Trauma Committee of The Royal College of Surgeons of England found that only 12% (19) of the hospitals had on-site neurosurgery. Of the 170 hospitals involved in the NCEPOD study, only 38 hospitals (1 out of every 5 hospitals) perform neurosurgical procedures. In the same study, more than half of the patients (58/114) that required neurosurgery were taken to hospitals without a neurosurgery department. In hospitals without a neurosurgeon, traditionally and for historical reasons, emergency department consultants, general surgeons or orthopaedic surgeons are usually responsible for the early management of these patients [10], through collaboration with a RNU. Typically, a CT examination is performed and the patient's case and their CT images will be discussed with a RNU. If neurosurgery is required, the patient would have to undergo a secondary transfer to a RNU.

2.1.1 Secondary transfer

The secondary transfer of patients is necessary under the present system of admitting patients to DGH in the first instance, and transferring patients who require specialist care, where it is unavailable, to a specialist centre. One out of every four patients (194/795) in the NCEPOD study underwent a secondary transfer, with neurosurgery by far the speciality in which most patients were subsequently transferred (see Figure 1).

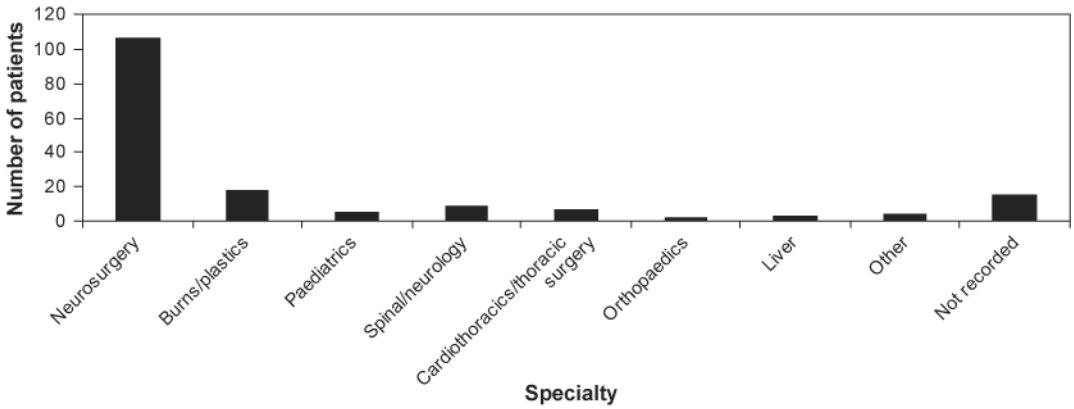


Figure 1: Specialty to which patients in the NCEPOD study were transferred [2]

There are many disadvantages and risks associated with the secondary transfer of patients. One of the greatest drawbacks of secondary transfers is that they inevitably delay time to surgery.

2.1.2 Time to surgery

In patients with severe head injury, patient outcome can be significantly improved by the speed at which proper care is administered. For example, the single most important factor in improving survival rates of patients with an expanding intracranial haematoma, often a consequence of a severe head injury, is its immediate surgical evacuation [11, 12]. An intracranial haematoma is a blood clot within the skull resulting from internal bleeding or haemorrhage and is often fatal. Delay in the evacuation of a haematoma is associated with an increased risk of morbidity and mortality.

In patients who underwent surgery for traumatic acute subdural haematoma within the first four hours (from the time of injury to the evacuation of the haematoma), the mortality rate was 30% compared to a mortality rate of 90% in patients who had surgery after four hours [11]. A study [12] on the effect of delayed treatment for a haematoma suggested an even smaller two-hour window from diagnosis to evacuation of a haematoma. However, in a study [9] involving 161 hospitals, only 8% of hospitals where neurosurgery is not available on-site were able to consistently evacuate an intracranial haematoma within four hours, compared to 41% of hospitals when neurosurgery is available on-site.

Because the severity of head injuries can vary greatly, it is difficult to determine a maximum permissible delay between time of injury and surgery that is critical for a good outcome. The Association of Anaesthetists of Great Britain and Ireland advocates a maximum of four hours between time of injury and surgery in their recommendations for the safe transfer of patients with brain injury [8]. However, this is rarely achieved in practice. Of the 43 patients in the NCEPOD study that required a secondary transfer to a RNU, only one out of seven patients (6/43) had neurosurgery within four hours of injury. In contrast, where an on-site neurosurgeon was available, two out of three patients (22/33) had an operation within four hours of injury.

A report [6] from the Neurocritical Care Stakeholder Group, established to offer expert advice to the Department of Health and those that commission neurocritical care locally, concluded that the existing system of care does not achieve prompt surgical evacuation of an intracranial haematoma for all patients. A 12-month audit on access to emergency neurosurgery for severely head injured children in the UK concluded that the system of care does not achieve surgical evacuation of a significant haematoma within the critical four hours [13], primarily due to delays resulting from secondary transfer.

2.1.3 Delays

A study [14] on the secondary transfer times of 81 patients with haematomas from a DGH to a RNU in Greater Manchester found a mean transfer time (from arrival at a DGH to arrival at the RNU) of 5.25 and 6 hours respectively. Another study [15] involving the secondary transfer of 50 consecutive head-injured patients found a mean transfer time from a DGH to a RNU of 7.4 hours. Long distance emergency transfers by ambulance from Shrewsbury to Manchester and from Colchester to London have been reported [16]. In addition to travel times, delays were often caused by not administering appropriate treatment for injuries prior to the transfer, not recognising the need for transfer, waiting for an ambulance, etc [15, 17].

2.1.4 Compliance

Arrangements for transfer were also often found to be haphazard, lacking compliance to existing recommendations and guidelines on the transfer of critically ill and severely injured patients [2]. Approximately one out of every three (57/183) hospitals in the NCEPOD study did not have protocols for the secondary transfer of patients with serious head injuries. The Intensive Care Society, in their 2002 guidelines for the transport of critically ill adults [18] observed that many secondary transfers were performed in an ad hoc manner and warned of the possibility of potentially serious complications. Many studies have shown serious flaws (e.g. lack of coordination, inadequate provision of equipment and training, variability in transfer quality, injuries that are not properly diagnosed or managed) in the secondary transfer of patients with head injuries [18-21].

2.1.5 Equity of access

Closely related to the issue of secondary transfer is equity to access. This is because a long distance to a RNU may affect the likelihood of patients being subsequently transferred. The Society of British Neurological Surgeons (SBNS) has advocated a maximum surface journey time of two hours to a RNU if equity of access is to be achieved [3]. In Scotland, some examples of typical travel times by road that exceeds this recommendation include three hours from Inverness, and five hours from Wick and Fort William [22]. Figure 2 illustrates the travel time to the nearest RNU for the Scottish population.

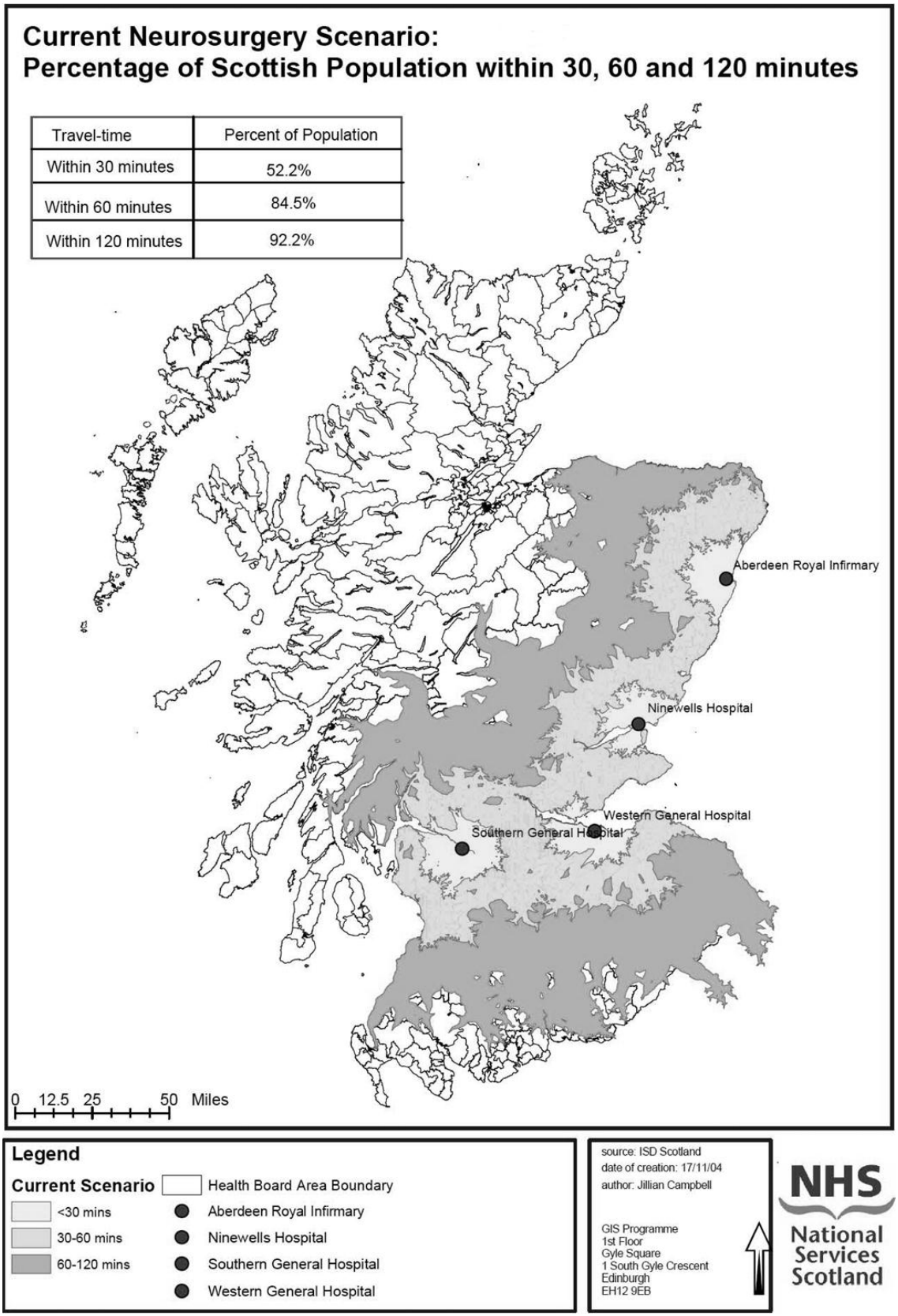


Figure 2: Surface journey time to the nearest RNU for the Scottish population[1] [23]

[1] Image reproduced after permission from NHS Scotland

The geographical distribution of the neurosurgical centres in Scotland (see Figure 2) and indeed, around the country (see Figure 3) means that patients living in remote areas in the country do not have equity of access to neurosurgical services. Of the 35 RNUs in the UK, nine centres alone are in the Greater London area [24] with only four RNUs in the whole of Scotland, and just one in Northern Ireland. The Royal College of Surgeons (RCS) Report of the Working Party on the Management of Patients with Head Injuries [25] found that although the national average distance of emergency departments without on-site neurosurgery facilities to a RNU was 23 miles, there were significant regional variations. Finally, secondary transfers are not always possible, as patient transfer may also be refused by a RNU, as neurosurgical capacity is limited.

Figure 3: Geographical location of neurosurgical centres in the United Kingdom and Ireland [26][2]

[2] Image reproduced after permission from BMJ Publishing Group Ltd.

2.1.6 Neurosurgical capacity

In the survey [9] of 161 hospitals by the Trauma Committee of The Royal College of Surgeons of England, the vast majority of hospitals without neurosurgical facilities (91%) did not transfer all patients with a severe head injury to a RNU. This was primarily due to RNU policy of only admitting patients requiring an operation (72%) and/or because of a lack of capacity by the RNU to accept patients (25%). To determine the capacity at RNUs to accept patients requiring emergency neurosurgery, Crimmins and Palmer [26], conducted a survey to provide a snapshot view of 36 RNUs in the UK and Ireland. This survey concluded that there is a severe shortage of available emergency neurosurgical capacity as the RNUs could only admit an additional 43 patients (at the time of the survey) for an overall estimated population of 63.6 million. Although the number of head injuries was projected at 28.6 injuries per day, these form only part of the neurosurgical workload. Even the London area, with nine RNUs, faced a shortage of available emergency neurosurgical capacity.

2.1.7 Demand for emergency neurosurgical procedures

The exact number of patients who undergo emergency neurosurgical procedures each year is unknown, partly because the data available, the Hospital Episode Statistics (HES) data, reflects the number of procedures and interventions performed, rather than the number of patients involved. An operation on a patient can actually be made up of several separate procedures and interventions, with different aspects of each operation (the actual procedures and interventions that are used) individually recorded. For example, a patient who undergoes neurosurgery to evacuate a haematoma could have more than one intervention recorded in the operation.

The HES inpatient data for England for the 2008/2009 annual dataset recorded 21816 interventions, of which 7290 were emergency, to evacuate an extradural haematoma (Classification of Interventions and Procedures - Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS) code A401). 2864 interventions, of which 1187 were emergency, were performed to evacuate a subdural haematoma (OPCS code

A411). The insertion of intracranial pressure monitors accounted for 1088 interventions, of which 533 were emergency interventions. These figures are conservative, as they have not been adjusted to reflect shortfalls in the number of records received from National Health Service (NHS) trusts and/or missing/invalid clinical data. They would also vary because emergencies are unpredictable. In any case, there is very little excess capacity in the system to cater for emergencies when they do happen [27]. It should also be noted that emergency neurosurgery is only part of neurosurgical workload.

2.1.8 Neurosurgical workload

The SBNS has stated that the neurosurgical workload in the UK is achieved by working at levels of resource that are potentially unsafe [8]. Statistics from the NHS Information Centre for Workforce for 2009 show that there are currently only 213 full time equivalent (FTE) consultant neurosurgeons in England [28] for an estimated population of 51.8 million people [29], a ratio of approximately one FTE consultant neurosurgeon to 243,000 people. This is far below the European average of one FTE consultant neurosurgeon to 125,000 people[3] [30]

A more representative indicator of neurosurgical workload is the average number of operations performed by consultant neurosurgeons. Because there is no available data on the number of operations performed by a consultant annually, statistics on the number of procedures and interventions performed by consultant neurosurgeons are used to estimate the total number of operations. In 2008/2009, consultant neurosurgeons in England performed an average of 1376 main procedures and interventions each [31]. Assuming an average of five procedures and interventions in an operation, a consultant would have performed an average of 275 operations a year, a figure that exceeds the recommended 180 – 250 number of operations [3]. Additionally, because of various factors such as geographical spread and population density, there would inevitably be an imbalance in neurosurgical workload. As such, there

[3] No recent statistics available, figures are for 2000

would be consultant neurosurgeons that greatly exceed the recommended number of operations, leading to potentially unsafe conditions.

2.1.9 Discussion

The NCEPOD report has highlighted the suboptimal care for some patients with severe head injuries in the UK. The report, on the current acute care of 795 patients who sustain a traumatic brain injury (TBI), gives a snapshot view of the sometimes poor state of affairs of emergency neurosurgery in the UK. A study on the trends in trauma care in England and Wales from 1989 to 2000 [32] found that survival rates for patients with severe head injuries have not shown much improvement since 1994, suggesting that any progress under the existing system of care may have plateaued. This lack of improvement can be attributed to a number of deficiencies under the present management of patients with severe head injuries. These obstacles to care will at best delay surgical intervention or at worst prevent a patient from receiving definitive care.

The first problem with the existing system of care for patients with head injuries is the requirement for secondary transfers to a RNU for neurosurgical care. Secondary transfers have been shown to delay time to surgery, and in some instances, result in avoidable deaths. A possible solution to the problems associated with secondary transfers is to transfer all severely head injured patients directly to a RNU as a matter of routine i.e. bypassing DGHs. While potentially eliminating the need for secondary transfers, the London Severe Injury Working Group [34] has concluded that direct transport might actually worsen outcomes where journey time to a RNU is more than 20 minutes.

Nevertheless, some ambulance trusts have a policy of bypassing DGHs for RNUs to avoid the possibility of a secondary transfer. These direct transfers however are the exception rather than the rule [2]. Currently the decision to transport patients directly to a RNU is at the discretion of clinicians at the scene [33] based on the perceived need for neurosurgery. However, the condition of some of the ‘non-surgical’ patients may deteriorate to an extent that they would also require neurosurgery, necessitating a secondary transfer.

In the UK, the management of patients with head injuries, including transfers, is informed by guidelines. A report [35] commissioned by National

Institute for Health and Clinical Excellence (NICE), addressed two key issues regarding the transfer of patients. The first is whether ambulances should “bypass” DGHs i.e. direct transport to RNUs. The second is what the indications for secondary transfer are, if patients were continued to be sent to DGHs (with is currently the norm).

On the issue of ‘bypassing’ local hospitals, the NICE Guideline Development Group (GDG), representing all relevant professional and patient parties in the UK, at present does not recommend that DGHs be ‘by passed’ [2]. This recommendation was based on the lack of evidence to recommend otherwise, as well as the significant impact on resources required for its implementation. As such, most head injured patients will continue to be sent to DGHs.

On the indications for secondary transfer, GDG now recommends that all patients with severe head injury be transferred to a RNU after receiving initial treatment from a DGH, irrespective of the need for surgical intervention [36]. A systematic review [37] of the implications of these guidelines concluded that there is a lack of capacity in the existing system and expressed reservations about the recommendation to transfer all severely head injured patients to a RNU, citing the inability of RNUs to even cope with current demand.

The requirement to transfer every patient with a severe head injury will not only stretch the already strained capacity of RNUs, it threatens equity to access to neurosurgical services for other groups such as patients scheduled for elective surgery as emergency surgery is often performed at the expense of elective surgery [3]. Therefore, it is unlikely that this recommendation can be fully implemented, a fact even acknowledged by GDG [2], with the lack of resources being a major constraint.

As such, although present guidelines [36, 38] recommend that all patients with a severe head injury be transferred from a DGH to a RNU, many patients are currently not transferred, and those that are transferred experience significant delays, in addition to risks associated with secondary transfers. Furthermore, secondary transfers may not always be possible. The condition of some patients may have deteriorated since the initial transfer and may be too unstable to be transferred.

The fact is that the most severely head injured patients do not require surgical intervention [2]. Although the majority of patients (62% of the patient population or 493 patients) in the NCEPOD study suffered severe head injuries, only one in every five patients (114 patients) required neurosurgery. A more liberal policy in transferring patients would overwhelm a system that is operating at or close to full capacity.

A possible solution to the lack of immediately available capacity at RNUs is to develop neurosurgical expertise within DGHs. However, training consultants involved in the care of severely head injured patients in DGHs without a neurosurgeon i.e. emergency department consultants, general surgeons or orthopaedic surgeons, to perform neurosurgical procedures would not be feasible, as most will deal with only few severe head injuries. Alternatively, the number of RNUs could be increased to cater for a demand for neurosurgical services, particularly emergency neurosurgery. However, the associated cost would be prohibitive. The NHS currently spends over an estimated £1 billion on acute hospital care alone, excluding rehabilitation and community care, for every 10,000 patients with significant TBI [35].

Given the known problems associated with secondary transfers and the fact that the majority of patients with head injury will continue to be sent to a hospital without on-site neurosurgery, performing robot-assisted surgery at DGHs may be a feasible alternative to moving a patient to a distant facility. Performing robot-assisted surgery at DGHs may be the only alternative if transfer to a RNU is not possible or if the transfer would put the patient at a greater risk. The use of a robotic-assisted surgery system will allow patients to be treated locally, eliminating the need to transfer a patient. The ability to perform the required surgical intervention at DGHs is especially important if transfer to a regional neurosurgical centre is not possible. It is also important if immediate surgical intervention is required but a transfer would result in significant delays to surgery. At present, no such robotic surgery system exists.

Robotic surgery is particularly suited for neurosurgery, since immobilising the head keeps the brain relatively stationary with respect to the robot. Admittedly, neurosurgery carries with it the risk of severe blood loss, brain damage, paralysis, or even death. Therefore, while it may not be feasible to perform all types of neurosurgery robotically, certain ‘low’ risk but lifesaving

neurosurgical procedures that does not involved deep brain incursions or complex manipulation may be particularly suitable to be performed robotically.

To this end, a robot-assisted system for emergency neurosurgery that is capable of performing routine but lifesaving neurosurgical procedures such as the evacuation of a haematoma, the drainage of cerebrospinal fluid (CSF) from the brain ventricles and the placement of an intracranial pressure monitor, is proposed. A more detailed discussion of these targeted surgical procedures is given in Section 2.3. The next section provides an overview of robotic surgery in general and robotic neurosurgery in particular.

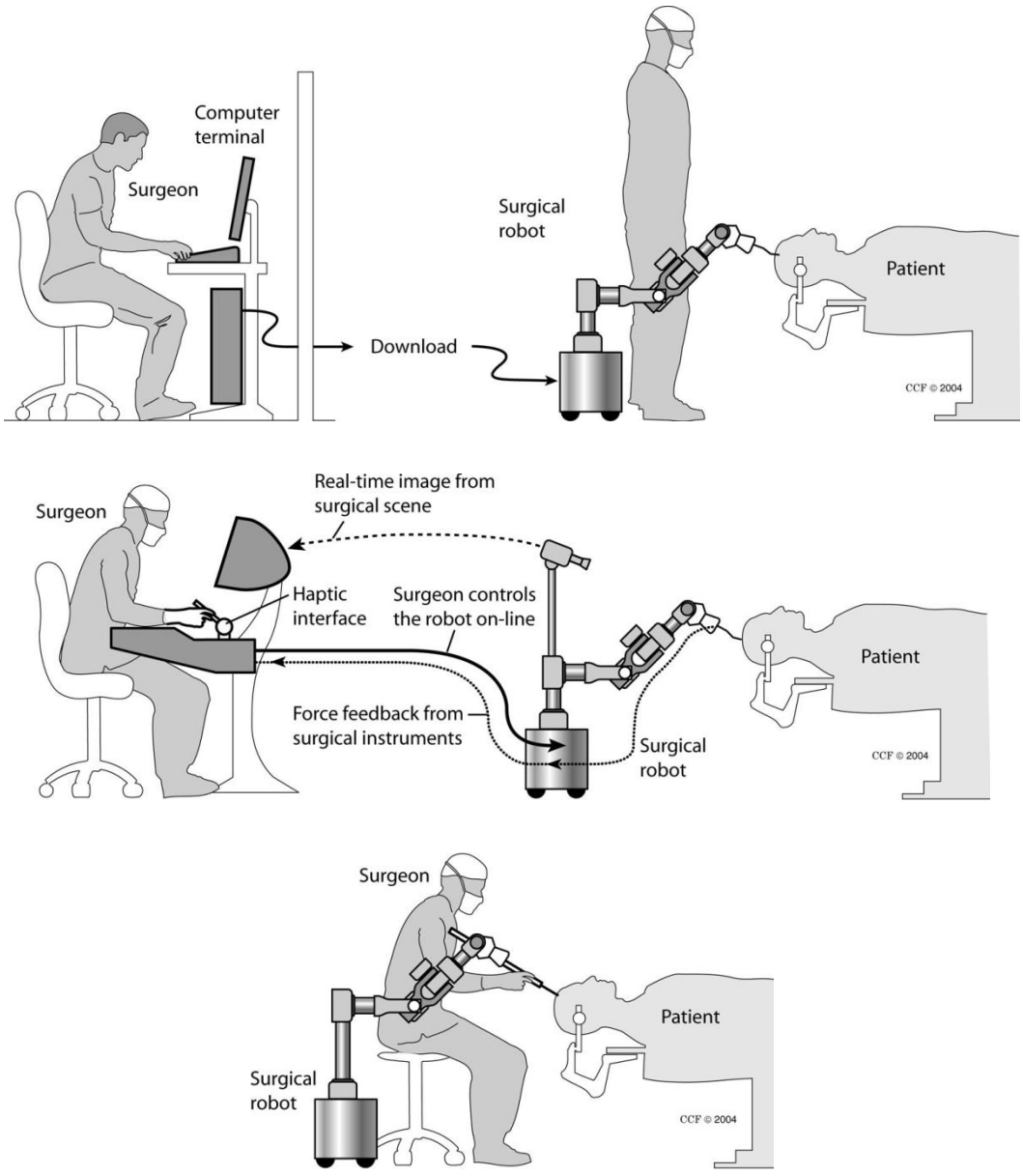
2.2 Robotic surgery

‘‘Robotic surgery’’ originated as an imprecise term, but it has been widely used and is now generally accepted by the medical community [39]. It refers to the use of a robot(s) to perform surgery. A different but somewhat related concept is computer-assisted surgery. In robotic surgery, a surgeon operates on a patient using a powered electromechanical device. In computer-assisted surgery, devices are generally manually powered by the surgeon [7]. Computer assisted surgery is therefore inherently safer than robotic surgery, although without some of the benefits of robotic surgery, such as the ability to perform procedures autonomously.

There are many robotic surgery systems at various levels of research and development. Some of these have already made the successful transition from research laboratories to the operating theatre. These commercial robotic surgery systems, with clinical applications in general surgery (the da Vinci Surgical System), neurosurgery (NeuroMate, Pathfinder, SurgiScope, EVOLUTION 1), and orthopaedic surgery (ROBODOC Surgical Assistant, SpineAssist) have different levels of autonomy, ranging from telesurgical (da Vinci), to supervisory controlled systems (ROBODOC).

2.2.1 Classification of robotic surgery systems

Nathoo et al [40] classified robotic surgery systems according to their levels of autonomy, using a broad classification based on their control strategies in terms of robot–surgeon interaction. Their classification divides robotic surgery systems into three categories (see Figure 4).



a
b
c

Figure 4: Technical classification of robotic surgery systems: (a) supervisory controlled system, (b) telesurgical system and (c) shared control system[4].

[4] Image reproduced after permission from Wolters Kluwer Health

The first is supervisory controlled robotic systems, where the robot performs the surgery autonomously under the direction and supervision of the surgeon. The second is telesurgical systems, where the surgeon performs the surgery remotely with a robot. The third is shared control systems, in which the robot and surgeon share control of the surgical instrument.

2.2.2 Robotic neurosurgery systems

In 1985, Kwoh et al. [41] pioneered the use of robots in neurosurgery, using a Unimation PUMA industrial robot as an aid to perform brain tumour biopsies. This was the first clinical application of a robot, and was motivated by the fact that conventional stereotactic surgery was, in their words, tedious and time-consuming. Stereotactic surgery is the use of positional information to locate targets within the body and is a minimally invasive form of surgery. Kwoh's system essentially automated the stereotactic frame adjustments required in traditional stereotactic surgery for localisation of a target. The robot was used to position a guide for drilling a burr hole[5] through the skull close to the site of the tumour. The system had the advantage of being faster and more accurate compared to previous methods that used stereotactic frames. The system was eventually abandoned, as the robot manufacturer did not want its robot, which was design for industrial applications, to be used in this manner [7]. The use of an industrial robot, with its large work envelope and corresponding huge range of motions, was intrinsically less safe than a special-purpose mechanism whose motions and forces were designed specifically for the action [7].

2.2.2.1 Neuromate

In principle similar to Kwoh's system, NeuroMate is the first robotic neurosurgical system to be approved by the Food and Drug Administration (FDA) [42], as well as the first to be commercially available. It is perhaps the most widely used neurosurgical robot [43]. The NeuroMate system consists of a

[5] A small opening in the skull to provide access to the brain

floor mounted modified 5-DOF industrial robot interfaced with a computer. A surgeon plans the procedure by identifying target(s) on preoperative CT or MRI scans, and a passive robotic arm (an active robot with power removed) aligns itself with respect to the target and serves as a guide to perform surgical procedures. Clinical and phantom studies of the NeuroMate system [44-46] have reported a target localisation error of within 3 mm. For all its complexity and associated hardware, the system cannot rival accuracies of within 1 mm for standard stereotactic frame-based techniques [45]. It is also limited to performing single dimension incursions and cannot be used to perform procedures involving complex manipulation. The current version has been modified from the original design to incorporate specific stereotactic requirements and to improve on safety issues.

2.2.2.2 CranioAssist

CranioAssist is the neurosurgical version of SpineAssist, an FDA approved robotic system for orthopaedic surgery [47]. It is a minimally invasive, image-guided, miniature robot system that is placed directly on a patient's head. The robot is capable of repositioning itself to within millimetres of predefined surgical targets. Its attached mechanical guide serves as a guide for the insertion of a needle, probe, or catheter by a surgeon. By mounting the robot directly on the head/skull, the requirement to immobilise the patient is avoided. Image to patient registration is performed by aligning a surface reconstruction of a patient's preoperative CT/MRI images with a surface reconstructed from an intraoperative 3D scan image. The registration system has an RMS registration error of 1.0 mm (standard deviation of 0.95 mm) [47] and an overall target registration error of 1.7 mm (standard deviation of 0.7 mm) [48].

2.2.2.3 ROBOCAST

ROBOCAST, an acronym for ROBOt and sensors integration for Computer Assisted Surgery and Therapy, is a recently launched three year

€4.5 million collaborative project (see Figure 5) by a consortium of European institutions scheduled for completion in 2011 [49]. The project is aimed at developing a proof of concept robot assisted minimally invasive neurosurgery system. The system is designed to perform procedures such as endoscopy, biopsy and deep brain simulation, autonomously under the supervision of a surgeon.

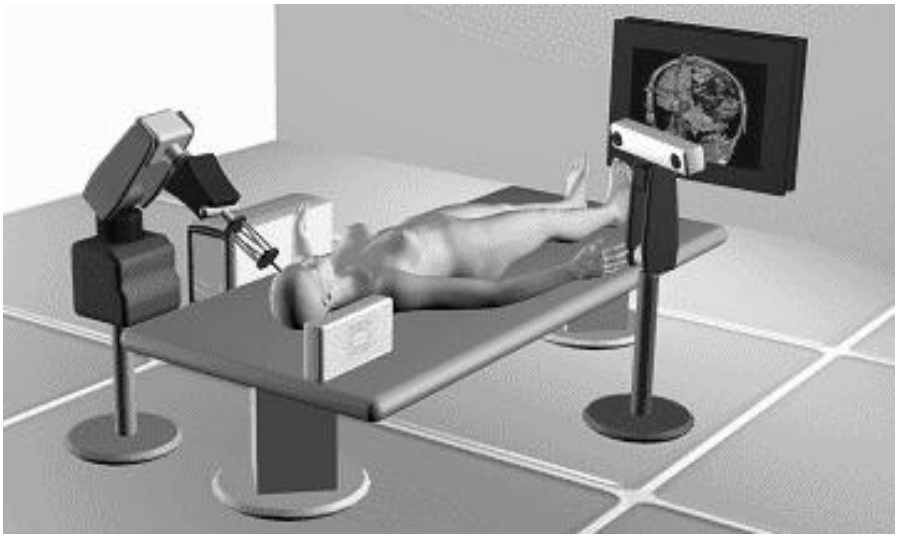


Figure 5: The ROBOCAST project[6] [53]

It will feature a modular architecture consisting of two robots, an active biomimetic probe, an optical tracker, an endoscope and a probe equipped with position and force sensors. A gross positioning six-axis revolute robot will support a miniature parallel robot holding the probe to be introduced through a "keyhole" opening into the skull of the patient, adopting a concept first used by Davies [50, 51]. Electromagnetic position and force sensors, mounted on surgical instruments, provide position and force feedback to the system while an imaging neuroendoscopic camera provides visual information of the surgical field and optical trackers track patient movement.

The integration of a multitude of sensors aims at increasing the level of autonomy of the robot by improving its perception. The ROBOCAST controller plans the entire robot trajectory within the brain based on an atlas that incorporates critical structures to be avoided. Execution of these procedures,

[6] Image reproduced after permission from the Project Coordinator of the ROBOCAST project, Prof. Giancarlo Ferrigno

which include trajectory planning right up to the intervention is with minimal or no input from a surgeon. There are, however, provisions for change intraoperatively. Semi-autonomous plan updating, to respond to changes that occur during surgery as detected by intraoperative sensors, will be proposed by the system to the surgeon for final decision-making [52]. This system, if successful, promises to be the closest thing to autonomous neurosurgery by a robotic surgery system.

2.2.2.4 neuroArm

Today, the state of the art in robotic neurosurgery systems is the MRI-compatible, image-guided, telesurgical, neuroArm system developed at the University of Calgary, Canada [54]. This system is capable of performing complex procedures, such as microsurgery and biopsy, inside the bore of an open magnet with near real-time guidance. It was used for the first time to remove a tumour in the brain of a 21-year old patient on the 12th of May, 2008 [55]. However, the neuroArm system, while capable of performing complex procedures, is a telesurgical system, relying on a surgeon for every move.

2.2.2.5 Minerva

The Minerva system [57] (see Figure 6), developed by the Swiss Federal Institute of Technology, is a six-axis stereotactic neurosurgical robot and is the only supervisory controlled robotic system for neurosurgery developed thus far. With the exception of target identification and trajectory selection, the system is capable of performing stereotactic neurosurgery without human intervention. The first four axes of the robot are for gross positioning and are not back driveable. To ensure the safety of the system, these axes are locked in place and switched off after reaching the desired position. The fifth axis brings the end effector close to the operative site and it is subsequently disabled. The sixth axis is a linear axis and is the only active axis during the operation.

Figure 6: ‘Minerva’ neurosurgery robot[7]

The use of traditional tools adapted and fitted with electric motors enables autonomous incision of the skin, drilling of a burr hole through the cranium, perforation of the dura mater and manipulation of surgical tools [58]. These tools are located on a rotary carousel and locked into position on the end effector. The end effector, which moves in a single axis only, then advances the tool linearly into the region of the patient’s head. Force and torque characteristics were used to determine the completion of a surgical action [58].

A tube with a sharpened end cuts the skin at the surgical site, by simultaneous rotation and advancement of the end effector. The start and end of skin incision is determined by monitoring the current consumption of the motor driving the linear axis and the motor rotating the knife/tube. Removing the incised skin requires the use of a corkscrew through the tube. The tube is left in place and serves as a mechanical guide for a drill to create a burr hole in the skull.

Similar to skin incision, the start and end of drilling is determined by monitoring the current consumption of the motor driving the drill bit. As a precaution, drilling will not continue for more than 15 mm, regardless of the current profile to prevent plunging i.e. to stop the drill bit from advancing further beyond the inner table of the skull. Forward movement of a sharpened tool through the guide-tube incises the dura.

[7] Image reproduced after permission from Professional Engineering Publishing Ltd

The entire operation was performed inside a CT scanner, allowing the surgeon to monitor the progress of the operation e.g. to verify the position of tools, intraoperatively. Clinical trials of the system involving 13 patients have been reported [59, 60]. The system was used to operate on two patients with tumours (malignant intracerebral cystic lesions) using an aspiration biopsy probe [59]. A further 11 operations, including nine biopsies, and two cyst evacuations were also reported [61]. Development of the system officially ended in 1998 [61], with the authors stating that the Minerva project had achieved its goal of demonstrating the feasibility of robotic neurosurgery.

Very few technical details of the Minerva system and of the clinical trials are available. It is unclear to the author if the system was used as described or was used in a reduced capacity, as some of the methods described were not entirely feasible. For example, although it was claimed that the current consumption and the torque profile of the motors was used to determine the start and end of an incision, the method was not described in any detail. The reported accuracy of 0.3 mm is because the surgery was performed inside a CT scanner, which greatly improves 3D localisation and accuracy. However, the use of a dedicated CT scanner is prohibitive in terms of cost. The use of CT intraoperatively is also highly unconventional, in part because the patient would be exposed to a significant amount of radiation.

2.2.3 Discussion

At present, patients with severe head injuries who require neurosurgery are transferred to a RNU for surgical intervention if neurosurgery is not available on-site. However, secondary transfer delays time to surgery, is not always possible, and is associated with increased risk to the patient. A possible solution is the use of supervisory controlled robotic surgery systems such as Minerva that are able to perform surgery autonomously under the direction and guidance of a surgeon (who would intervene in the event of a problem). The MISEN system, which aims to perform some of the same procedures as MINERVA, is introduced in the next section, along with the targeted neurosurgical procedures and the proposed surgical protocol.

2.3 Mechatronic Intervention System for Emergency Neurosurgery (MISEN)

The Mechatronics in Medicine group at Loughborough University, UK, has proposed the use of a robot-assisted surgery system to perform emergency neurosurgical procedures at DGHs [62] (see Figure 7). The system, referred to as MISEN, is to be used in emergencies only when immediate surgery is required, but for various reasons, the patient cannot be operated on conventionally by a neurosurgeon. The term “emergency” is used to differentiate time critical surgery where immediate intervention significantly improves a patients’ prognosis, from elective surgery. It is intended as an aid to surgeons who would not usually perform these procedures.

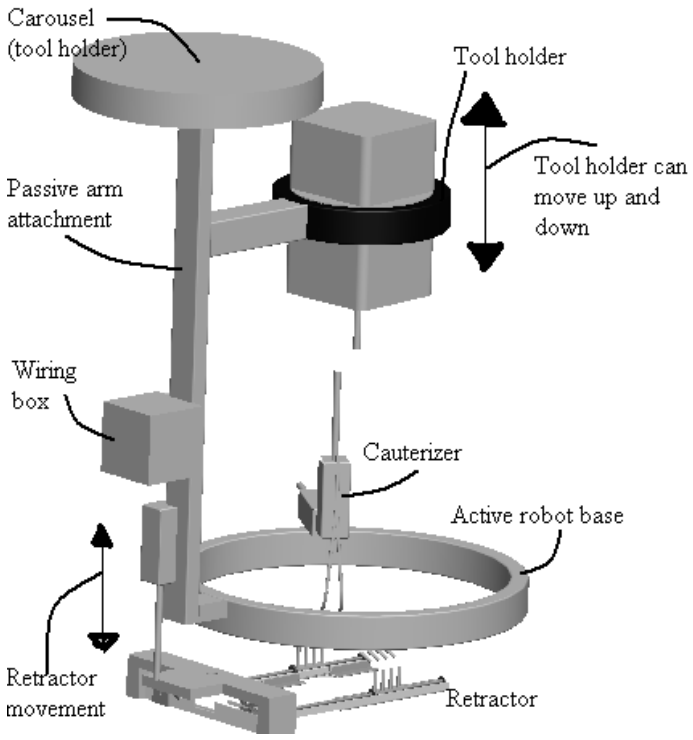


Figure 7: Schematic of MISEN robot manipulator[8]

2.3.1 Targeted procedures

Brain injury can be classified as either primary or secondary. A primary injury is an injury that is a direct consequence of trauma, stroke, etc. while a

[8] Drawn by former Mechatronics in Medicine group member, Salman Hussain

secondary injury is an injury that results from events following the initial injury. In patients with brain injuries, neurosurgery is aimed at preventing or minimising irreversible damage to the brain from secondary injuries. While the range of neurosurgical procedures is extensive, three procedures were targeted through discussions with a collaborating neurosurgeon[9]. The three procedures are the evacuation of an intracranial haematoma, the drainage of cerebrospinal fluid (CSF) from the brain ventricles, and the placement of an intracranial pressure (ICP) monitor. An ICP monitor is a diagnostic tool useful in monitoring ICP. The evacuation of a haematoma and the drainage of CSF from the brain ventricles are classified as emergency procedures [24] while an elevated ICP is a common secondary injury in traumatic brain injury (TBI) patients. Of the 114 patients in the NCEPOD report that had a surgical procedure because of head trauma, the majority of patients (42%) had surgery to insert an ICP monitoring device. Statistics from the NCEPOD study indicate that the majority of the neurosurgical procedures performed on patients with TBI are the evacuation of a haematoma and placement of an ICP monitor. These procedures alone account for four out of every five procedures performed on severely injured head patients requiring neurosurgery [2].

2.3.1.1 Intracranial pressure monitoring

Increased ICP may compromise cerebral perfusion (blood flow to the brain) and lead to cerebral ischemia (lack of blood flow to the brain). Extreme ICP may cause brain hernia, where the brain shifts across structures within the skull and is often fatal. Although the brain has the ability to regulate ICP, it loses this ability after a traumatic injury. Therefore, an important aspect in the treatment of TBI patients is to monitor and manage their ICP. ICP is used to calculate cerebral perfusion pressure (CPP), an indirect measure of cerebral perfusion. CPP is defined as the difference between mean arterial pressure (average blood pressure) and ICP. A CPP above 70 mm Hg is necessary to maintain sufficient cerebral blood flow [24].

[9] Professor Michael Vloeberghs, Queen's Medical Centre, Nottingham, UK

Although ICP monitoring may be performed in hospitals without on-site neurosurgery, this procedure is not widely available in the UK [1]. While there are several guidelines or recommendations related to traumatic brain injuries [2, 8, 24, 35, 63], there are currently no published national guidelines for the use of ICP monitoring in the UK. However, ICP monitoring is generally used in severe TBI patients[10], moderate head injuries[11] [64] or abnormal CT scans [36]. Compared to the evacuation of a haematoma, where immediate surgery significantly improves patient outcome, intracranial pressure monitoring is less urgent, but important nevertheless as part of the overall management of brain injury. ICP related therapy such as the drainage of CSF could improve outcome after severe head injury [1].

An ICP monitor may be placed at several locations (see Figure 8). The most accurate, low-cost, and reliable method of monitoring ICP is intraventricular monitoring i.e. a ventricular catheter connected to an external pressure transducer e.g. strain gauge [24]. It is considered the 'gold' standard in monitoring ICP and is the preferred method as it allows for CSF drainage [65]. It is placed in over 50% of severely injured patients [68]. A study of 136 patients with severe TBI concluded that intraventricular ICP monitoring is a low-risk procedure and may facilitate an early and accurate intervention in these patients [66].

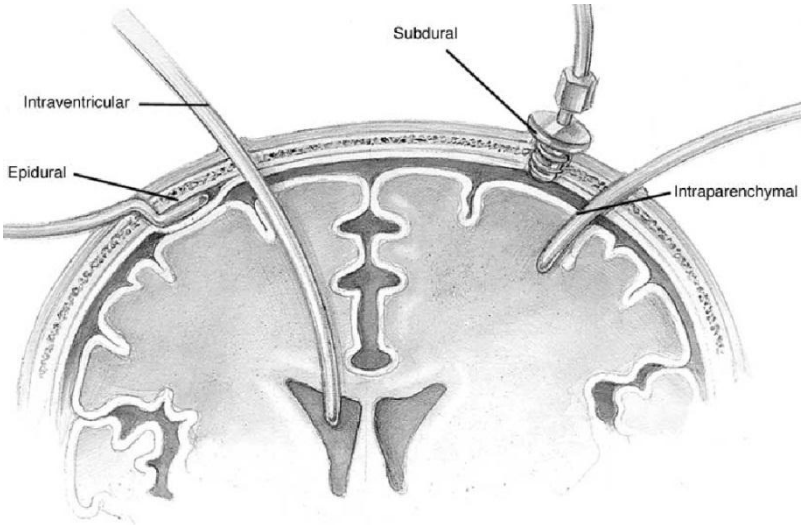


Figure 8: Various locations for ICP transducer placements[12]

[10] A patient with a GCS of 8 or less

[11] A patient with a GCS of 9 - 12

[12] Image reproduced after permission from Elsevier

If intraventricular ICP monitoring is not feasible, for example, if the ventricles are very small, an intraparenchymal catheter may be used instead. Subarachnoid, subdural, and epidural monitors are known to be less accurate [24] and the exact positioning of these monitors needs to be able to accommodate a nearby ventriculostomy incision [67]. The following steps outline the general procedure for the placement of an ICP monitor:

1. An anatomical landmark for the placement of an ICP monitor called the 'Kocher's point'[13] is identified (see Figure 9). The entry site must at least 1 cm anterior to the coronal suture to avoid the motor strip.
2. The site is shaved, followed by incision and retraction of the skin to expose the underlying skull.
3. The pericranium is removed prior to drilling.
4. A burr hole is drilled through the skull to allow access into the brain.
5. The dura is perforated while avoiding damage to the underlying structures.
6. Depending on the type of ICP monitor used (see Figure 8), it may then be inserted into the relevant part of the brain.

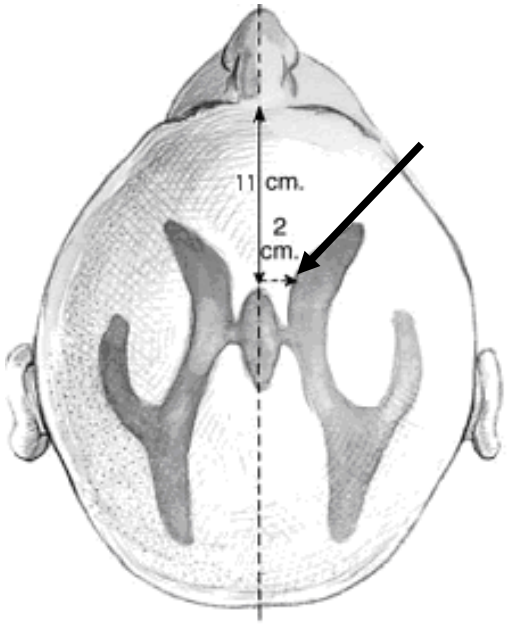


Figure 9: Location of ICP monitor placement[14]

[13] 2–4 cm lateral to the midline and 11 cm posterior to the nasion

[14] Image reproduced after permission from Elsevier

2.3.1.2 Drainage of cerebrospinal fluid

The drainage of CSF allows for the reduction of elevated ICP. An ICP above 20 - 25 mmHg is generally accepted as the upper threshold above which treatment to lower ICP is necessary [24]. One method of reducing ICP is through the drainage of CSF from the brain ventricles, where a resulting reduction in volume in the ventricles will result in a decrease in intracranial pressure. This is the basis of the Monro-Kellie hypothesis. The Monro-Kellie hypothesis states that total volume of the constituents of the cranial i.e. brain, CSF and blood, is constant and in a state of volume equilibrium. An increase in volume of one constituent must be compensated by a decrease in volume of another [68].

CSF is usually drained into a closed collection system outside the body through a catheter inserted into the brain ventricles. The placement of a catheter in the brain ventricles for drainage of CSF is known as a ventriculostomy. A ventriculostomy is time critical, potentially life-saving [69] and one of the most common emergency neurosurgical procedures [70]. It is usually performed using a freehand technique based on anatomical landmarks [71], which is the standard of care and is the method of choice due to its simplicity and efficiency [69].

2.3.1.3 Evacuation of an intracranial haematoma

An intracranial haematoma is characterised by an immediate and rapid internal bleeding that may be fatal. It is a medical emergency where prompt intervention makes a significant difference to patient outcome. The sooner an expanding haematoma is evacuated, the better the outcome. For patients with severe brain injury e.g. intracranial haematoma, its evacuation is time critical and requires immediate surgery. One method for the evacuation of a haematoma is through a burr hole craniotomy [63]. In many respects, the evacuation of a haematoma is similar to a ventriculostomy. Therefore, for brevity, the procedure will not be explained.

2.3.2 On surgical complications

Aside from complications arising from infections, the only other reported risk of a ventriculostomy, the most invasive of the targeted procedures, is haemorrhaging. However, this complication is rare. When it does occur, it is unlikely to be life threatening. A review of 1,790 ventriculostomies found an overall risk of haemorrhage of 5.7% with less than 1% clinically significant haemorrhage [73]. In terms of complications due to infections, a retrospective study of 368 ventriculostomies performed over a 4-year-period [72] found an infection rate of less than 3%. However, these statistics are based on a neurosurgeon performing the procedure. To infer from these statistics the likelihood that a robotic system will achieve similar complication rates is difficult.

2.3.3 On clinical accuracy

In a retrospective study of 24 patients who underwent a ventriculostomy using a free-hand technique [76], the accuracy of catheter placement, defined as the distance between the catheter tip and the foramen of Monro (channels that connect the lateral ventricles and the third ventricle at the midline of the brain), was 9.7 ± 6.3 mm. In another study [69] of 97 patients on the accuracy of a ventriculostomy using a free-hand technique, catheters were misplaced by an average of 1.6 cm ± 0.96 cm from the target ventricle, with an average of two passes. A high percentage (22.4%) of catheter tips was actually placed in extraventricular/nonventricular CSF spaces. These placements were considered ‘successful’ as they were still able to drain CSF. Drainage would still be possible in these instances as ventricular catheters typically have multiple holes located away from the tip.

2.3.4 Discussion

The average volume of ventricles is 30.9±5.7 ml [75] with the average thickness of a haematoma requiring surgical evacuation is more than 10 mm [63]. These ‘targets’ are in general are quite large, allowing for lower localisation accuracy. In fact, neurosurgeons frequently use fingerbreadths to

localise the site haematoma [54]. Therefore, in terms of clinical accuracy, a localisation accuracy of within a centimetre, based on the levels of accuracy achieved by neurosurgeons performing the neurosurgical procedures under consideration, should be sufficiently accurate.

In terms of surgical complications, the targeted procedures are in general safe with very little risk of complications and may be made safer by incorporating image-guided techniques. A study [74] involving 16 patients in an image-guided robotic placement of a ventricular catheter using a preoperatively defined trajectory reported no complications. The catheter successfully reached the ventricles in a single pass each time, which reduces the risk of haemorrhaging.

A natural progression of image-guided surgery is to perform these procedures robotically. The accuracy and number of passes for a robotic system performing a ventriculostomy is expected to be comparable to that of an image-guided technique. Therefore, it is likely that a ventriculostomy by a robotic system would be more accurate than neurosurgeons performing the procedure using a freehand technique, potentially reaching the ventricles in a single pass. The number of passes is especially important in a ventriculostomy, as an increase in the number of passes increases the likelihood of haemorrhaging.

2.3.5 Proposed surgical protocol

Under the proposed surgical protocol, MISEN will perform the required surgery as prescribed by the neurosurgeon, with a surgeon providing some degree of monitoring and control. This includes all surgical actions such as skin incision, retraction of the incision, cauterisation, drilling of a burr hole and perforation of the dura. A neurosurgeon would determine the need for surgery and where surgical intervention is indicated would determine if the patient is to be operated on robotically. If the decision to operate on the patient robotically is made and patient consent is received, the neurosurgeon should additionally provide a preoperative plan i.e. specify targets and entry points, based on CT images of the patient, to guide the robot. The neurosurgeon should also ensure that the straight line trajectory from the entry point (burr hole) to the target point

(haematoma or ventricles) avoids critical brain structures e.g. vascular regions and motor strip, which may otherwise cause bleeding and functional deficit respectively.

A passive positioning arm is used to position the robot manipulator close to the entry point. The use of a passive arm prevents gross motions during surgery. Once near the entry point, the robot moves to the specified chosen entry point, and is orientated based on the trajectory towards the target point. At the entry point, the necessary surgical procedures to create a burr hole, such as incision and retraction of the scalp, drilling a burr hole and cauterising the dura mater is performed. A motorised drive, with a single linear axis, will advance an instrument e.g. a catheter, towards the target. The surgical protocol for the system is divided into two stages, preoperative and intraoperative [62, 77].

Preoperative

- CT image acquisition
- Identification of an entry point and target by a neurosurgeon, ensuring that the straight line trajectory avoids critical brain structures

Intraoperative

- Gross localisation by manually positioning the passive arm of the robotic system
- Localisation of the entry point by a smaller active robot
- Perform the surgical action autonomously i.e. skin incision, retraction of the incision, cauterising of any bleeding, drilling of a burr hole through a skull, perforation of the dura, and insertion of an ICP monitor or catheter.

2.3.6 Discussion

The majority of existing neurosurgical robotic systems, with Minerva being a notable exception, requires that intracranial access to the brain via a burr hole be made by a neurosurgeon prior to surgical intervention by the robot. This requirement reduces the overall utility and autonomy of these systems.

Existing robot-assisted surgery systems are therefore fundamentally different from the proposed MISEN system.

The major steps involved in the targeted procedures are incising the skin, retracting the incision, drilling a burr hole through the skull to provide access to the brain and the insertion of an ICP monitor or catheter. These surgical procedures involve straight-line incursions into the brain but no complex manipulation. A closer inspection of these procedures, by decomposing them into the individual sub actions involved, reveals that performing these actions robotically may be feasible. The drilling of a burr hole through the skull may be thought of as analogous to the computer numerical control (CNC) machining of a work piece, with the patient's head, the work piece, and the robot, the CNC machine. Although a very simplistic view of the procedure, it is conceivable that these procedures can be performed using a supervisory controlled robotic surgery system such as the proposed MISEN system.

An important step prior to the intraoperative surgical intervention in robotic surgery is the localisation of the intervention site based on a preoperative plan by a surgeon, who typically specifies an entry point, target point(s), and a trajectory to provide guidance to the robot. The preoperative plan, based on a virtual model of the patient reconstructed from medical images such as CT or MRI, has to be 'registered' to the patient to establish spatial correspondence in a technique known as image to patient registration, which is the subject of the next chapter.

Chapter 3

A Registration Framework for MISEN

3.1 Introduction

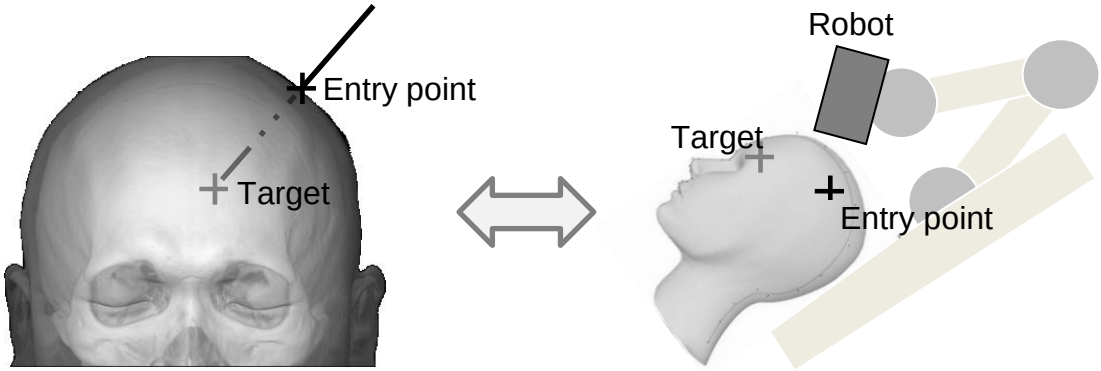
Before the advent of CT by Hounsfield in 1971, diagnosis of neurological problems was primarily based on the symptoms presented by a patient. Today, radiological images such as CT or MRI are routinely used for diagnosis in all but the most trivial head injuries. Because the need for surgery can normally only be established after a CT or MRI scan, these images are usually available prior to surgical intervention. In addition to their diagnostic utility, CT and MRI images are useful as a means of localising an area of interest in a patient. Neurosurgeons, for example, would locate a particular brain structure by referring to its position with respect to bony landmarks in radiological images, based on their knowledge of the anatomy of a skull. This ability is important especially where there is no direct visualisation of the operative site, such as keyhole neurosurgery and burr hole procedures.

An enabling technology for robotic surgery is the ability to localise 'targets' identified in medical images, within a patient, independent of a surgeon. To localise an area of interest on a patient based on medical images, the geometrical relationship between the images and patient must first be determined. Unfortunately, unless special measures are taken during imaging, spatial correspondence between the image and patient is lost. In medical imaging, this spatial correspondence may be established through a technique known as image to patient registration.

3.2 Image to patient registration

Image to patient registration is a method where the position of an anatomical in an image is correlated with the position of the anatomy itself. This would facilitate the determination of a position identified in 'image space' in 'patient space'. Mathematically, image to patient registration is the determination of a geometric transformation matrix that aligns corresponding points from these two spaces. The determination of this relationship will enabled target(s) and trajectory specified in a preoperative plan of the surgical procedure to be mapped on a patient, and is an enabling technology for robotic interventions that are guided by these preoperative plans. The robotic insertion

of a catheter into the brain ventricles for example, may be guided by a trajectory specified on a virtual model of the patient reconstructed from radiological images and registered to the patient. Figure 10 illustrates the registration chain between preoperative plan and the patient/robot. Medical imaging, which is the starting point in the registration chain, is detailed in the next section.

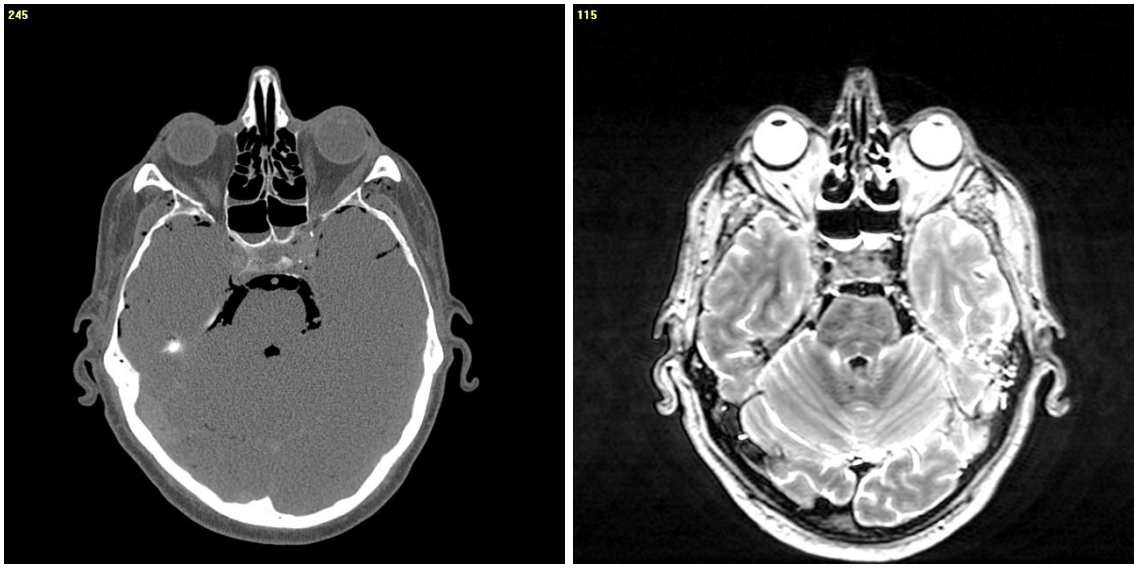


a | b

Figure 10: Registration chain between (a) preoperative plan and (b) patient/robot.

3.3 Medical imaging

Common medical imaging modalities include X-Rays, CT and MRI (see Figure 11). In contrast to conventional X-Rays, where three-dimensional structures are superimposed to form a two-dimensional image that is often geometrically distorted, CT and MRI provides a geometrically correct three-dimensional volumetric and two-dimensional cross-sectional visualisation of these structures. In their clinical guidelines on head injury [35], NICE has indicated CT imaging of the head as the primary investigation for the detection of acute brain injury, as well as advocated a more liberal scanning policy. Where CT imaging is indicated but unavailable, arrangements should be in place for urgent transfer to a centre with CT scanning. X-Rays are no longer recommended in the assessment of head trauma [78], with the exception of minor/mild head injuries [35]. For safety, logistic and resource reasons, MRI is also not currently indicated as the primary investigation for clinically important brain injury in patients who have sustained a head injury.



a | b

Figure 11: (a) A typical axial CT and (b) MRI image

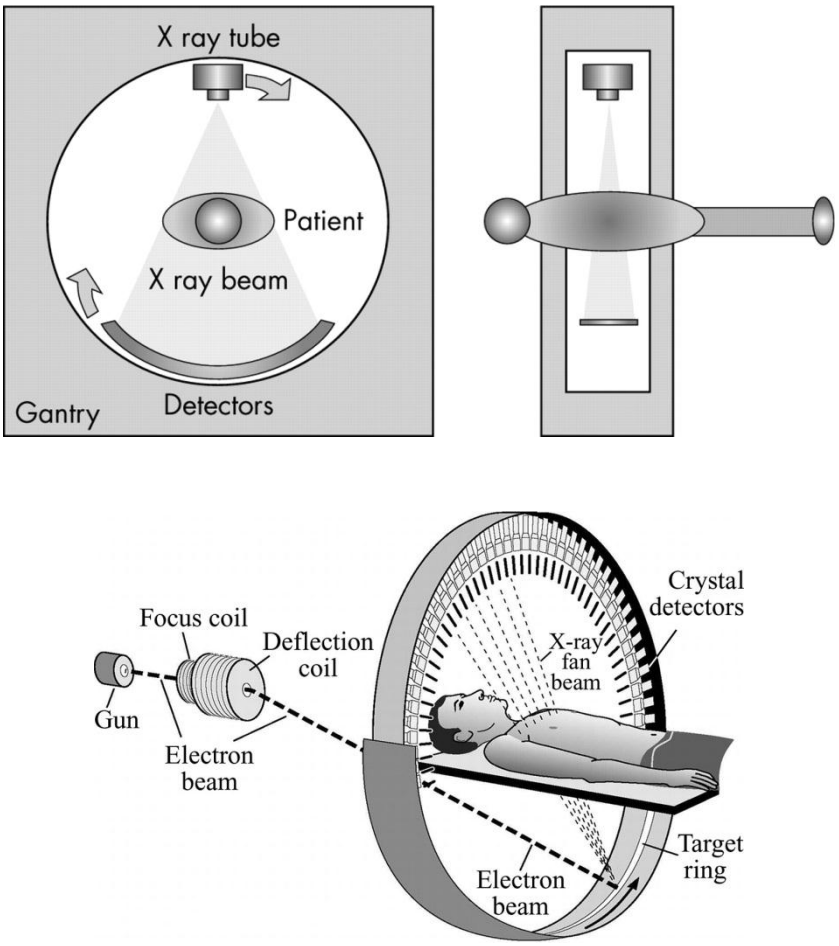
3.3.1 CT vs. MRI

CT uses X-rays and computer reconstruction to generate cross-sectional images of individual structures while MRI uses strong magnetic fields and radiation in the radio frequency range to image parts of the body. Whereas 95% of hospitals have 24-hour access to CT imaging [2], not all hospitals have MRI scanners, due to the associated high cost of procuring as well as maintaining these scanners, as well as the personnel to operate them. Additionally, the use of MRI scanners is contraindicated on patients with cardiac pacemakers, medical or bio stimulation implants e.g. insulin pumps, ferromagnetic foreign bodies e.g. shell fragments, or metallic implants e.g. surgical prostheses, aneurysm clips [33, 79] because of the strong magnetic fields generated during its operation.

Because of the nature of emergency surgery, as opposed to elective type surgery, it is very likely that the medical history of a patient is not known. The use of MRI in emergencies is therefore contraindicated, as the patient's medical history cannot be ascertained with absolute certainty. On the other hand, the only contraindication for use of CT is if the patient is unstable [80], although the patient may be sedated prior to scanning.

3.3.2 CT basics

In older fourth generation CT scanners, a continuous volume of contiguous CT slices is acquired as an X-ray tube/source rotates around a patient, who is moved within a scanner equipped with a stationary circular array of detectors (see Figure 12a). Modern day fifth generation CT scanners use a circular array of X-ray sources instead, and there is no mechanical motion, with the exception of the bed that the patients lays on (see Figure 12b).

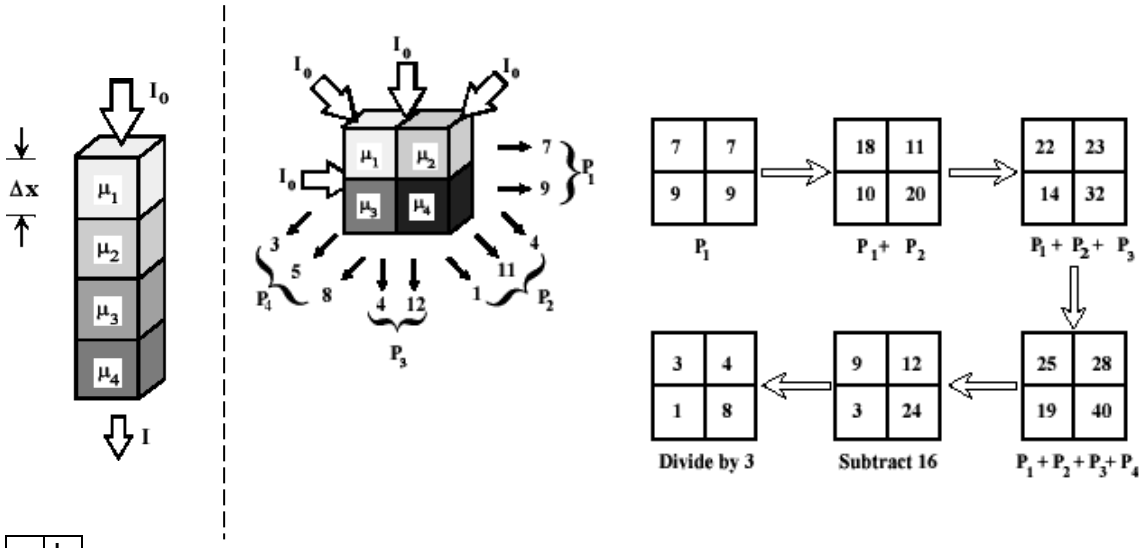


a
b

Figure 12: (a) Fourth generation and (b) fifth generation or electron beam CT scanner architecture[15] [81]

[15] Image reproduced after permission from the Radiological Society of North America, and authors Wolbarst A B and Hendee W R.

Each CT slice consists of voxels that corresponds to the density or attenuation coefficient of the anatomy being imaged, which is traversed during the scan by numerous X-ray beams and acquired by the X-ray detectors. The attenuation coefficient associated with each voxel is usually determined using back projection. Figure 13 illustrates how back projection works.



a | b

Figure 13: (a) X-radiation of incident intensity, I_o attenuated by four voxels each of height, Δx and of different linear attenuation coefficients, μ_1 to μ_4 and the transmitted intensity, I and (b) back projection reconstruction[16].

The transmitted intensity, I for each voxel is given by Equation 3.1:

$$I = I_o{}^{-(\mu_1+\mu_2+\mu_3+\mu_4)\Delta x} \qquad 3.1$$

where I is the transmitted intensity, I_o is the incident intensity, μ is the attenuation coefficient and Δx the voxel thickness. The attenuation of the X-ray beam is assumed to have occurred uniformly along the entire ray path. Back projection reconstruction, to obtain the attenuation coefficient of each voxel, consists of summing the respective profiles, P_1, P_2, P_3 and P_4, subtracting an offset and renormalisation (see Figure 13b).

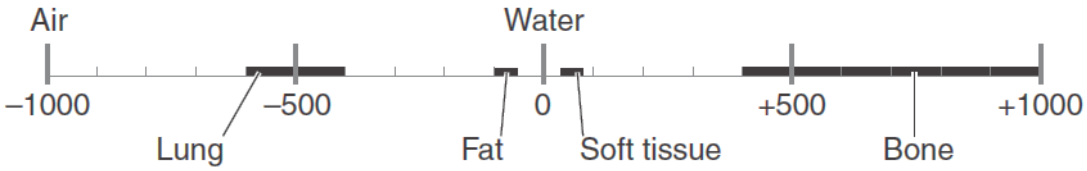
[16] Adapted from the Wikimedia Commons file "Image: The origin of image degeneracy in radiography.gif" and Image: Representation of back projection.gif".

3.3.2.1 Hounsfield units

Hounsfield units (HU) are arbitrary numbers assigned to elements based on their relative density with values ranging from -1000 to + 1000. DICOM greyscale pixel value may be converted to HU using Equation 3.2:

$$HU = \text{Greyscale pixel value} \times \text{RescaleSlope} + \text{RescaleIntercept}. \quad 3.2$$

where the RescaleSlope and RescaleIntercept values may be found in the file header of a DICOM file. The HU of various elements are shown in Figure 14.



Bone	+400 → +1000
Soft tissue	+40 → +80
Water	0
Fat	–60 → –100
Lung	–400 → –600
Air	–1000

Figure 14: Hounsfield units for various elements [82]

3.3.2.2 Noise

Scanning protocols such as applied voltage, tube current, scan time, filtration, slice thickness, slice spacing, gantry tilt, etc., affect the signal-to-noise ratio of CT images. Manufacturer recommended values produce an image with an acceptable level of noise while limiting the radiation dose to the patient. However, the selection of the tube current - exposure time product (mAs) varies significantly even for the same make and model of CT system. Selection of mAs often relies on subjective decisions and predictions, and for most systems, the minimum and maximum values selected for mAs differ by up to a factor of 3-4 [89].

Excessive radiation dose does not provide additional diagnostic information. However, unlike conventional radiography where excessive or inadequate radiation doses may result in overexposed or underexposed respectively, the effect is minimal in a CT image. Because CT is a digital technique, its image may be manipulated to a certain degree to correct for excessive or inadequate radiation dose [89]. However, images will be unacceptably noisy if the chosen mAs are less than optimal. Noise (measured standard deviation) is inversely proportional to the squared reduction in mAs i.e. $\frac{1}{\sqrt{mAs}}$ [90]. Therefore, the level of noise doubles if the mAs are reduced by three quarters.

3.3.2.3 Filtering

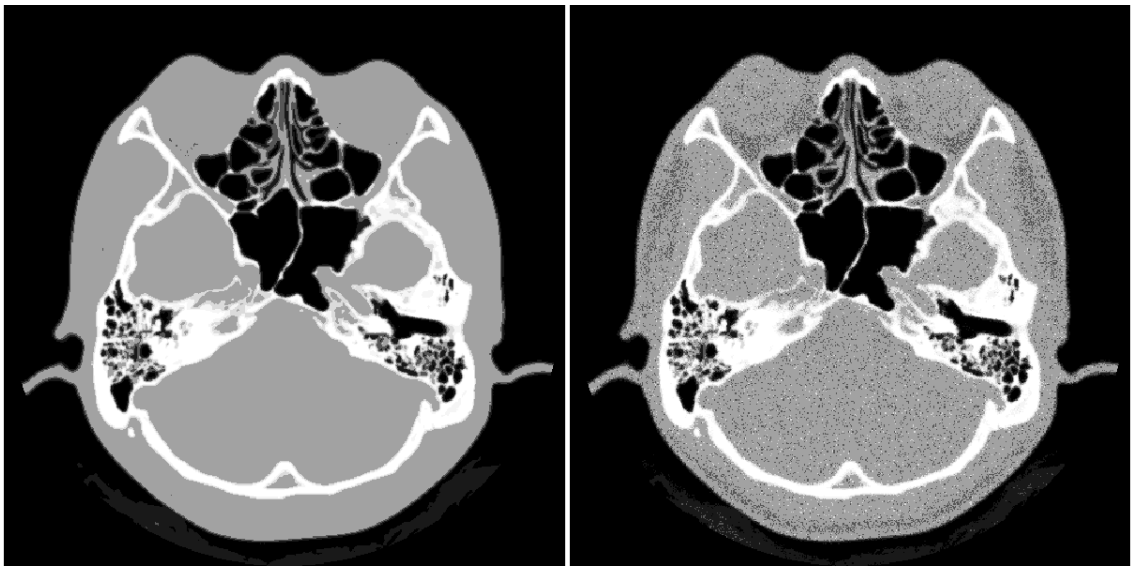
Noise in CT images has many components but is primarily due to the quantum noise in photon detection, electronic noise in the projection data, and noise from the reconstruction of images from projection data [83]. A common technique of removing noise is by averaging a pixel value over a local neighbourhood. Unfortunately, averaging can blur the boundaries of image features by smoothing across edges. In their seminal work on anisotropic diffusion, Perona and Malik [84] developed a technique that removes image noise while preserving boundaries. In anisotropic diffusion, smoothing occurs within regions where the gradient is small, but not across boundaries where the gradient magnitude is high as the diffusion coefficient is reduced near edges estimated by the first derivative, as defined by Equation 3:

$$\frac{\partial I}{\partial t} = div(c(x,y,t)\nabla I) = \nabla c \cdot \nabla I + c(x,y,t)\Delta I \qquad 3.3$$

where I is a greyscale intensity image, Δ is the Laplacian, ∇ is the gradient, div is the divergence operator. c is the conduction coefficient, which controls the rate of diffusion and is given in Equation 3.4:

$$c(x,y,t) = e^{-(\|\nabla I\|/\kappa)^2} \qquad 3.4$$

Gerig et al. [85] was the first to apply an anisotropic non-linear diffusion based technique to filtering MRI data. Anisotropic diffusion is now widely used in medical imaging [86]. A variation of this technique that preserves ramp edges was developed by Gilboa [87] with smoother edges and almost no staircasing effects. Figure 15 shows a CT image and the same image corrupted with Poisson noise. Poisson noise rather than Gaussian noise is more commonly present in imaging applications where images are generated by photon-counting devices such as CT [88]. The noisy image was filtered with a Gaussian low pass filter (3x3 with a standard deviation of 0.5) and a non-linear anisotropic diffusion (c = 20, κ = 0.15), respectively. The image filtered with the Gaussian low pass filter and anisotropic diffusion had 86% and 89% similar boundary pixels to the original image, with a mean absolute image difference of 2.7 and 3.2 greylevels between the original image and filtered image respectively. Anisotropic diffusion therefore maintains more edge pixels than the Gaussian lowpass filter, while reducing noise better.



a | b

Figure 15: CT image (a) Original and (b) corrupted with Poisson noise (shown in eight greylevels for visualisation purposes)

3.3.2.4 CT measurement accuracy

Measurements can be reliably taken from CT if they accurately reflect the actual physical dimensions of the imaged object. The mean error between linear soft tissue measurements taken from 3D volume rendered CT images and physical measurements of ten conventional craniometric anatomical landmarks on 13 cadaver heads was 1.78% [91]. A study [92] involving 20 linear measurements of 16 conventional craniometric anatomical landmarks of 12 cadaver heads found no statistically significant difference between measurements taken from 3D volume rendered CT images and physical measurements. Another study [93] on the accuracy of facial soft tissue thickness measurements under different scanning protocols e.g. various slice thickness (0.5 – 7 mm), pitch (1:1 – 2:1), and types of scanner (conventional, spiral, multidetector) found the mean deviation to be within 0.43 mm in all instances when compared to actual physical measurements.

3.4 Overview of image to patient registration methods

Image registration methods can be classified based on the type of features that are used to effect registration in the two spaces to be co-registered. These can be broadly divided into four categories, point-based methods, surface-based methods, intensity-based methods and non-image based methods [94]. The output of an image registration method may be a rigid or non-rigid geometrical transformation. A rigid transformation preserves all distances and angles, and consists of a translation and rotation component. It is usually used in neurosurgery and orthopaedics. Non-rigid transformations, which are applicable to non-rigid anatomy, are used for registration of anatomy other than the head and bone.

3.4.1 Point-based methods

The simplest registration methods are point-based methods. In point-based methods, corresponding point pairs, whose coordinates in both spaces are known is used to facilitate registration. A minimum of three non-collinear

points is required for registering a volume in space or four [95] if homogeneous coordinates (as is common in robotics) are used. Early approaches employed head frames that were attached to a patient's head. The intersection of Z-shape rods on these heads frames with the image plane appear as points in the medical images. However, these frames were cumbersome to use and restricted access to certain parts of the brain.

Roberts [96] was the first to introduce a frameless registration system, using skin-affixed markers. While less restrictive than frames, they are susceptible to errors due to the markers shifting or even falling. The use of skull-implanted markers reduces the possibility of the markers shifting, and is considered the gold standard in image to patient registration, but is a very invasive procedure. Furthermore, the use of skull-implanted markers for registration is not practical and highly unconventional for the targeted neurosurgical procedures, as they are typically only used for deep brain procedures requiring accuracy in the sub millimetre range such as a biopsy or functional neurosurgery. By contrast, errors in the order of centimetres have been reported without any major complications [69] for a ventriculostomy, the most invasive of the proposed targeted procedure. Alternatively, anatomical landmarks [97] may be used. The disadvantage of using anatomical landmarks is that, unlike skin-affixed markers, the type of landmarks available for use in registration is very much dependent on the anatomy being registered.

3.4.2 Surface-based methods

In surface-based methods, the surface of a patient's anatomy, typically obtained using a laser-contouring device, is matched with a surface of the anatomy reconstructed from medical images. Two of the earliest implementations of surface-based methods for image to patient registration was by Colchester et al. [98], and Grimson et al. [99]. In their methods, a surface model of a patient's head, reconstructed intraoperatively using stereovision and laser scanners respectively, was matched to a surface derived from their medical images.

Unlike point-based methods, there is no closed form solution for surface matching. Surface matching algorithms, such as the 'head and hat' algorithm by

Pelizzari [100] and the iterative closest point method by Besl and McKay [101], are usually used to find an optimal transformation to align the two surfaces. A known problem of surface matching algorithms is that the optimisation may converge to a local minimum, resulting in an inaccurate transformation. Mis-registration or poor registration is likely if surfaces to be matched have no prominent features. Surface matching algorithms are also computationally intensive, as the optimisation problem requires an iterative approach to solve.

3.4.3 Intensity-based methods

Intensity-based methods exploit the full content of an image, and include techniques based on mutual information and photo consistency. As these methods are based on image intensity rather than image features, feature extraction or segmentation is not necessary. These techniques are therefore well suited in applications where features cannot be reliably extracted.

In their seminal work on image registration by maximisation of mutual information, Viola and Wells [102] described an approach whereby a three-dimensional model of an object is registered to an optical image of the object. This was achieved by maximising a measure of mutual dependence between the normal component of the surface model of an object and the image intensity of the object. An optimisation scheme, using mutual information as a basis, was used to find a transformation that best aligned the object and its image. Mutual information is greater when images are aligned than when not aligned. However, the assumption that there is a relationship between the image intensities and surface normals is only valid if the light sources and cameras are located far away from the patient. The requirement that the cameras and light sources be placed far from the patient would place impractical constraints on the design of the registration system. Alternatively, a collimated light source may be used.

A different technique, based on photo consistency, was developed by Clarkson et al. [103] to register CT images to two or more optical images of a patient using a similarity measure based on the alignment of a surface to its optical images. In photo consistency, an unknown surface can be reconstructed from a set of optical images by exploiting the consistency of image intensities of

points in each image. Conversely, given an accurately defined surface, photo-consistency might be used as a measure of alignment of a surface to these optical images. Their technique had a mean error of 1.05 mm when tested on an artificial skull, with a surface error of between 1.45 and 1.59 mm when the initial mis-registration was up to 16 mm/degrees. However, their method assumes a relationship between surface normal and observed intensity. Reflectance, though, is in general dependent on lighting and viewing direction as well. To date, there has been no reported use of intensity-based methods to perform image to patient registration in a clinical environment.

3.4.4 Non-image based registration methods

Non-image based registration methods require that the two devices be calibrated to each other. For example, registration of robot arm to a CT scanner may be achieved by calibrating the robot arm in relation to the CT scanner. The correspondence between the position of surgical tools mounted on the robot arm and a position on CT images would then be known. However, this would mean that the operation would have to be performed inside or adjacent a CT scanner, requiring the exclusive use of the CT scanner for the duration of the operation, which is not a feasible option. This method is also seldom used because of the exacting calibration requirements.

Table 1 summarises the registration method, advantages and constraints as well as the target registration errors (TREs) of some existing commercial robot-assisted systems and navigation systems, with particular emphasis on the registration methods employed. These systems were chosen as illustrative examples, as there are many more similar in design and function. TRE is the registration error of points not used in the registration, and is explained in Section 4.0.

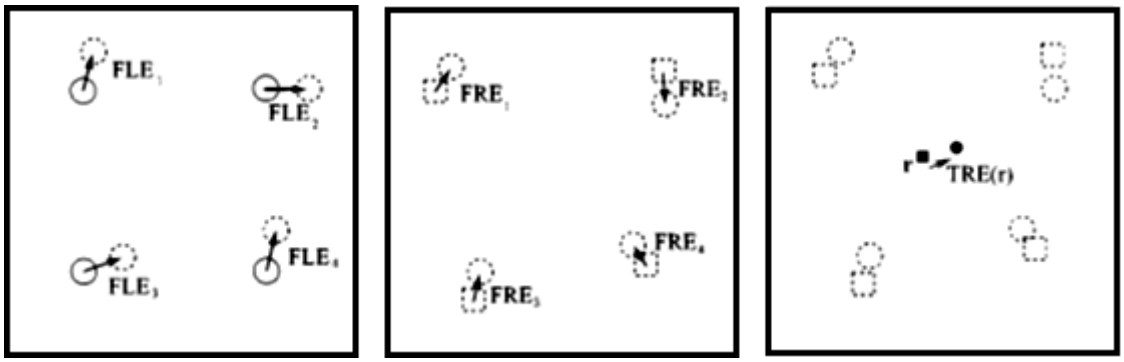
Table 1: Existing commercial robot assisted systems and navigation systems

System		Registration method	Advantages	Constraints	Accuracy
Robot-assisted surgery	Pathfinder	Point-based. Camera mounted on robot end effector determines position of skin affixed fiducial markers (black titanium sphere mounted on a yellow disc, with an adhesive backing for skin mounting) in patient space. Dedicated software automatically localises these markers In CT/MRI image space.	Non-contact measurement method, eliminates errors due markers shifting due to physical contact with markers	Prospective registration method	TRE = 2.7mm (range 1.8 – 3.23 mm) [104]
	NeuroMate	1. Non-imaged based. NeuroMate is precalibrated with respect to a head frame (ZD stereotactic ring). No explicit registration required. 2. Point-based. Uses helicopter-shaped ultrasonic localising device (with corresponding CT and MRI localisers) affixed to patient's skull. Ultrasound probe determines position of fiducial markers in patient space. In image space, dedicated software automatically determines position of the CT and MRI localisers.	1. Accurate 2. Registration is automated	1. Difficult to use, may restrict access to certain parts of the brain 2. Prospective registration method	1. TRE = 0.86 ± 0.32 mm [45] 2. TRE = 1.95 ± 0.44 mm [45] 2.9 mm (0.7 – 3.2 mm) [105]
Navigation	ISG viewing wand (Mechanical-based system)	Point-based. Interchangeable contrast cap fiducials fits into skull implanted base and localised interactively by a neurosurgeon in image space. The contrast caps are replaced with registration caps and are localised in patient space by the neurosurgeon placing the tip of a tracked probe in the straightest possible orientation to a	Precise, reliable	Prospective registration method. Mechanical arm restricts mobility and range.	TRE = 2.76 ± 1.19 mm (skin affixed markers) 2.25 ± 0.95 mm (bone implanted markers) [106]

		small divot on the registration cap.			
	BrainLAB (Optical-based system)	1. Point-based. Array of three reflective markers is rigidly fixed to a patient's head holder with five skin-affixed fiducials attached to the patient's head. These fiducials are either donut-shaped, self-adhesive markers or BrainLab markers (inserted into sockets attached to the patient's scalp and detected automatically in medical images by software). A probe, tracked relative to the marker array, is used to touch the fiducials, to determine the position of patients head. 2. Surface-based. Surface of patients face reconstructed using a laser contouring device and matched with preoperative medical image.	1. Navigation of various surgical instruments possible 2. Retrospective registration possible	1. Registration cannot be performed retrospectively. Skin-affixed markers, when used, may shift or fall off. 2. Localisation accuracy away from frontal facial region is poor	1. TRE = 4 ± 1.4 mm [107] 1.31 ± 0.87 mm [108] 2. TRE = 2.77 ± 1.64 mm [108]
	InstaTrak (Electromagnetic-based system)	Point-based. A plastic headrest with seven metal fiducial balls (with an additional two balls in the earpiece to provide additional accuracy for procedures more posteriorly) and a radiofrequency transmitter is placed on a patients head prior to CT imaging/MRI. Software detects fiducials in patient and image space automatically. A receiver is attached to one of several possible instruments, and this instrument is calibrated by touching the central point on the transmitter	Registration is automatic. Unlike optical systems, no line of sight requirement. Can be used with fiducial or anatomical markers.	Electromagnetic system susceptible to interference by external electromagnetic field or metallic objects Contraindicated for use of patients with pacemakers and cochlear implants	TRE = 2.28 ± 0.91 mm [109]

3.5 Registration error metrics

A registration error metric is a quantitative measure of how well registered images are aligned. In point-based methods, only a least-squares approximation of the correct registration is possible, as there is bound to be some error with respect to their homologues. Maurer et al [110] defined three types of registration errors for point-based methods: fiducial localisation error, (FLE), fiducial registration error, (FRE) and target registration error, (TRE). A fiducial is a point used for the purposes of registration and is known to be reliable. FLE is the error in localising the exact position of a fiducial, which may occur in both spaces and causes errors in registration. FRE is a measure of fiducial misalignment after registration. TRE is the distance of points of interest not used in the registration from their true position. Figure 16 illustrates the different types of errors.



a | b | c

Figure 16: (a) FLE (solid circles represent actual positions while dashed circles represent localised position), (b) FRE (dashed circles represent position in one space, dashed squares represent positions mapped into the other space) and (c) TRE is the distance after registration between an anatomical location (filled square) and the corresponding location in the other space (filled circle)[17] [160].

[17] Image reproduced after permission from IEEE

3.5.1 On fiducial registration errors

FRE is defined as the error between the fiducial and its corresponding position after registration and is given by Equation 3.5:

$$FRE_i = \mathcal{T}(x_i) - y_i \tag{3.5}$$

where x_i and y_i are the x and y coordinates of fiducials in both spaces belonging to fiducial i, and $\mathcal{T}$ the transformation matrix. Since FLE is not usually known, FRE may be used to estimate FLE. The relationship between FLE and FRE [111] is given by Equation 3.6:

$$FRE^2 = (1 - \frac{2}{N})FLE^2 \tag{3.6}$$

where N is the number of fiducials. In general, if there is no systematic bias error in identifying the fiducials, the greater the number of fiducials used, the better the registration accuracy [112].

3.5.2 On target registration errors

For rigid-body point-based registration, with an identical, independent, and isotropic FLE distribution, TRE is given by Equation 3.7:

$$TRE^2(r) = \frac{FLE^2}{N}\left(1 + \frac{1}{3}\sum_{k=1}^{3}\frac{d_k^2}{f_k^2}\right) \tag{3.7}$$

where d_k is the distance of the target(s) from principal axis, k and f_k is the distance of the fiducials from principal axis, k. While FRE is independent of the fiducial configuration, TRE is a function of their configuration. Minimum TRE is achieved by having as many landmarks as possible, a landmark configuration where the target is located at the centroid of the configuration, and localising these landmarks accurately.

Errors get progressively worse as the target deviates away from this centroid i.e. d_k becomes large. Large TREs will also result from a near collinear fiducial configuration i.e. f_k is very small. Although the target location may not be known beforehand, as a general rule of thumb, landmarks should be spaced out evenly and located as far away from each other. Unlike FRE, TRE cannot be determined directly. Therefore, FRE is usually used as a measure of registration accuracy, as feedback to assess whether or not the registration is successful. Shamir et al [120] argues that FLE is not always good estimate of the TRE because of the assumption of identical, independent, and isotropic FLE distribution is not always met. However, TRE estimated based on this analytical expression has been shown to correspond well to that obtained in practice in a clinical study by West et al [121].

3.5.3 On the accuracy of image registration methods

FRE of 1.0 to 3.0 mm are typical of registration systems that use anatomical landmarks, with the number of landmarks used ranging between eight and 16. Systems that use skin-affixed markers may have FRE values less than 2.0 mm, with the number of markers used ranging from six to ten. Systems using bone-implanted markers have FREs less than 1.0 mm, and use three to five markers [114]. A prospective study involving 26 patients [115] found a mean root mean square (RMS) FRE of 3.2 ± 1.0 mm and 2.9 ± 1.0 mm when anatomical landmarks and skin-affixed markers were used.

FRE depends only on the degree to which the chosen points correspond in the two views. However, in medical applications, TRE is a more significant measure of accuracy. A more recent study [116] involving 50 patients found that, independent of target location or imaging modality, the TRE for anatomical landmarks was 4.97 ± 2.29 mm for 6.7 ± 0.5 landmarks, the TRE for skin-affixed markers was 2.49 ± 1.07 mm for 6.9 ± 0.4 markers and the TRE for surface matching was 5.03 ± 2.30 mm. The ‘target’ in the study was a fiducial marker placed on the surface of the head.

3.5.4 On the required registration accuracy

The required registration accuracy is dependent on the surgical procedure. As the ‘targets’ in a ventriculostomy and evacuation of a haematoma are in general quite large (allowing for a large margin of error in positioning) the required accuracy is not very high. For example, the average volume of ventricles is 30.9 ± 5.7 cm^3 [75] while the average thickness of a haematoma requiring surgical evacuation is more than 10 mm [63]. A TRE of not more than 10 mm is therefore considered sufficient for the three targeted procedures.

3.6 Registration framework

The proposed registration framework should meet the accuracy requirements of the targeted neurosurgical procedures, be cost effective, and simple to implement. The registration problem can be stated as follows: Given a preoperative plan based on CT images of a patient, find a transformation that will registere this plan to the patient.

3.6.1 Retrospective versus prospective methods

The use of skin-affixed markers, which is the method of choice for image to patient registration systems, is a prospective technique. The main disadvantage of prospective methods is that they require two separate CT scans, one for diagnostic purposes and the other for registration purposes. On the other hand, registration using anatomical landmarks and surface matching, although less accurate than skin-affixed markers, can be performed retrospectively. The choice of a registration method is therefore a compromise between accuracy and the ability to perform the registration retrospectively.

Prospective registration methods are not appropriate in emergencies, because secondary scans inevitably delay time to surgery. This is in contrast to elective neurosurgery, where both retrospective, but more frequently prospective registration methods, are used. This is because the neurosurgical procedure is scheduled in advance i.e. not time critical, and a secondary CT

scan can usually be performed. There is no requirement for secondary scans when retrospective methods are used, an important consideration in an emergency when immediate surgery is indicated or when the patient may no longer have access to a CT scanner for another scan.

3.6.2 Point based versus surface based methods

The use of a retrospective image to patient registration technique is appealing because no secondary scan is required i.e. a diagnostic scan may be used to effect registration. The choice is therefore between the use of anatomical landmarks and surface based methods. Because surface scans are typically of the facial region, accuracy in localising targets away from the facial region is poor. A study involving 12 patients [117] who underwent a surface based registration with facial surface scans found that TRE increases the further the target is located away from the face.

Additionally, surface based methods require the use of specialised hardware such as 3D laser scanners or structured lighting to obtain a surface model of the patient's anatomy. Surface-based methods are also very much dependent on the similarity of the reconstructed surface, and the ability to reproduce faithfully similar surfaces. However, because the surfaces are acquired at different times, varying facial expressions between the time the 3D scan and medical image was acquired may reduce the accuracy of registration. The requirement for similar facial expression is a lesser constraint when anatomical landmarks that do not change substantially due to changes in facial expression are used, such as the ear tragus. Because of the complexities and hardware cost associated with surface matching, the use of anatomical landmarks as a registration basis is considered.

3.6.3 Anatomical landmarks

Anatomical landmarks of the head commonly used as a registration basis include the medial and lateral canthus (outer and inner corners of the eye), tip of the nose, philtrum (vertical groove in the upper lip), soft tissue

nasion (depressed area between the eyes, superior to the bridge of the nose), and tragus (a small pointed eminence of the ear). However, because routine diagnostic head CT scans start from the base and end at the vertex of the skull, the tip of the nose and philtrum may not always be available in CT images. In addition, due to the superficial nature of the lateral canthus, it is sometimes not very well defined. For example, the position of the lateral canthus depends on whether the patient's eyes are opened or closed.

To ensure consistency in the position of the lateral canthus, a CT scanning protocol requiring that patients be scanned with their eyes opened is needed, which by definition would make this a prospective technique. Deep-set wrinkles can also mask the true location of the lateral canthus. In contrast, landmarks such as the medial canthus and tragus are more distinct. Meanwhile, surface analogues of bony landmarks such as the soft tissue nasion can usually only be found in patient space by palpation, and therefore cannot be used as an anatomical landmark in an automated registration system.

3.6.4 Proposed registration framework

A point-based rigid body registration based on homologous anatomical landmarks in CT and patient space is proposed, with automatic localisation of the landmarks. Currently anatomical landmarks are determined manually based on visual inspection or palpation, requiring user interaction. These landmarks, if robustly detected, could be used as a basis for an automated image to patient registration system. An automated registration system is proposed because it is a user-friendly method, requiring no user intervention. However, the automatic detection of anatomical landmarks is difficult. Unlike skin-affixed markers, which can be designed for automatic detection, anatomical landmarks do not lend themselves to similar techniques. Nevertheless, the automatic localisation of these landmarks in both image and patient space would be an enabling step towards an automated image to patient registration system.

The use of anatomical landmarks as a registration basis in an automated image to patient registration system has not been reported. Clarkson et al. [118] argues that a point-based registration using anatomical landmarks is likely to be inaccurate as landmarks are difficult to localise accurately due to limited image

resolution and contrast, which gets progressively worse in CT as the slice thickness/spacing is increased. However, improvements in the spatial resolution of the CT images have enabled more accurate landmark localisation. Although the typical slice thickness for routine diagnostic CT images is 5 mm [119], CT images can be reconstructed with different slice thickness retrospectively from existing raw data, without re-scanning the patient. The limiting factor in slice thickness is the detector configuration used during the scan. Advances in imaging technology have therefore made localisation of anatomical landmarks feasible.

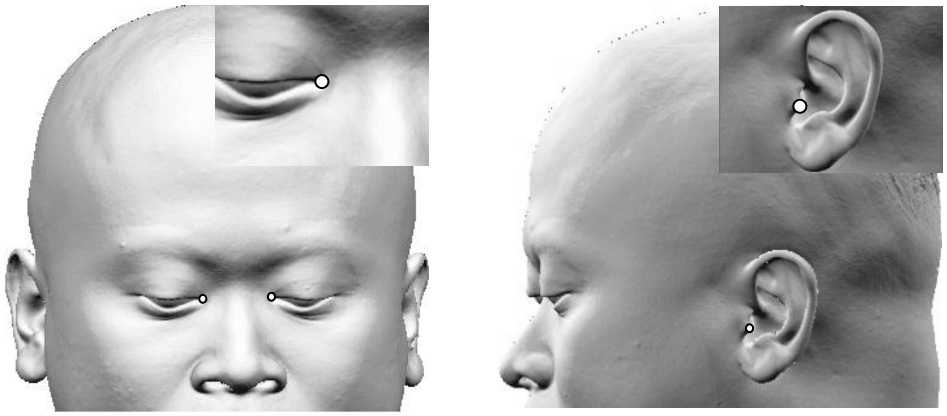
The anatomical landmarks used should be spaced out as evenly, located as far away from each other, and encompass as large a volume of the head, as possible. The mean TRE when anatomical landmarks that were located primarily on the face were used was lower when the target was at a frontal location (4.03 ± 2.14 mm with a range of 0.31 – 6.59 mm) compared to a non-frontal location (6.03 ± 2.70 mm with a range of 2.80 – 11.11 mm) [116]. Therefore, the landmarks chosen should not be confined to the facial region alone. Furthermore, while the use of four fiducials/landmarks instead of three improves registration accuracy and robustness significantly, registration accuracy does not improve considerably after five or six fiducials/landmarks are used [113, 114].

In view of this, four near co-planar anatomical landmarks, the left and right medial canthus i.e. inner eye corners, and the left and right tragus i.e. a small pointed eminence of the ear, were chosen as inputs for a rigid body registration. The anatomical point landmark pairs were also chosen because they can be reliably and robustly identified in patient and CT space. Intraoperatively, in patient space, close range stereo photogrammetry is used to localise these landmarks, because it is relatively inexpensive and allows for non-invasive, non-contact measurements. The same landmarks may be found in CT through a computational approach. The proposed framework does not modify the workflow in terms of CT image acquisition.

As the registration is only valid if the patient does not move during surgery, the patient must be immobilised. Patient movement can be detected by continuously monitoring the position of the anatomical landmarks during surgery, and re-registration performed if any movement is detected.

3.6.5 Definition of the anatomical landmarks used

The anatomical landmarks used are conventional craniometric anatomic landmarks. The medial canthus is the inner corner of the eye where the upper and lower eyelids meet (see Figure 17a). The tragus is a small, pointed, cartilaginous flap in front of the external opening of the ear (see Figure 17b). Because the tragus is a relatively large cone shaped structure, its apex is chosen as the point that uniquely defines its position.

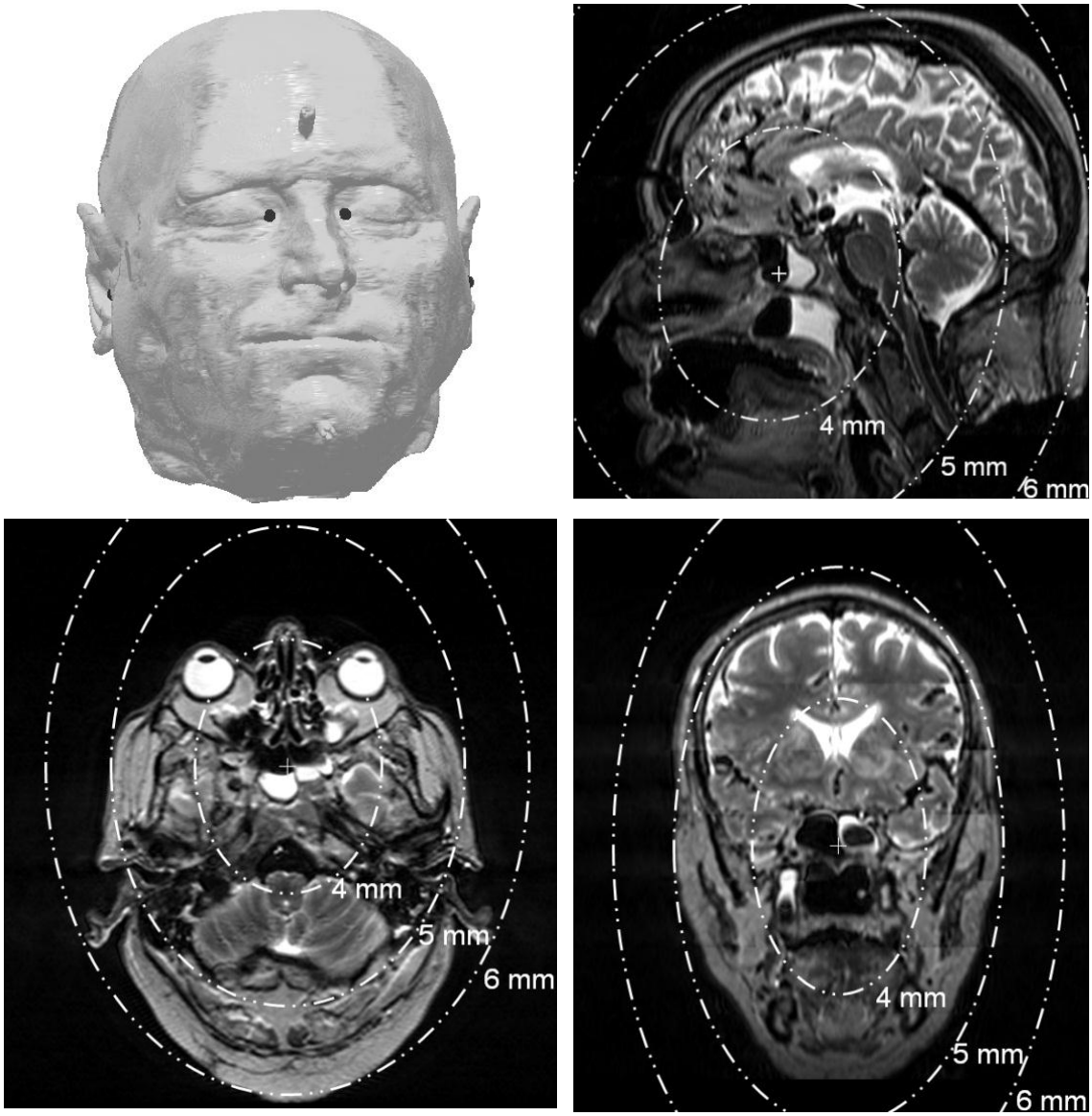


a | b

Figure 17: (a) Medial canthus and (b) tragus

3.6.6 TRE estimated based on analytical expression

The proposed registration framework may be used if the TRE is within the required clinical accuracy. Registration accuracy depends on the degree of correspondence between the point pairs in the two spaces. The accuracy of the proposed registration framework was estimated based on the analytical expression of TRE by Fitzpatrick et al. [111] for a given FLE and fiducial configuration, as given in Equation 3.7. The TRE using the four anatomical landmarks i.e. the medial canthi and both ear tragus, has been simulated for landmark localisation error of 5 mm. Figure 18 illustrates the estimated TRE. The TRE isocountours are ellipsoid and are overlaid in the orthogonal landmark principal axis planes on MRI images, and are similar to that obtained by West et al. [114]. The resulting TRE, within millimetre range for any location within the head, is acceptable as the required clinical accuracy of the targeted procedures is in the centimetre range and is similar to that achieved by a neurosurgeon.



a	b
c	d

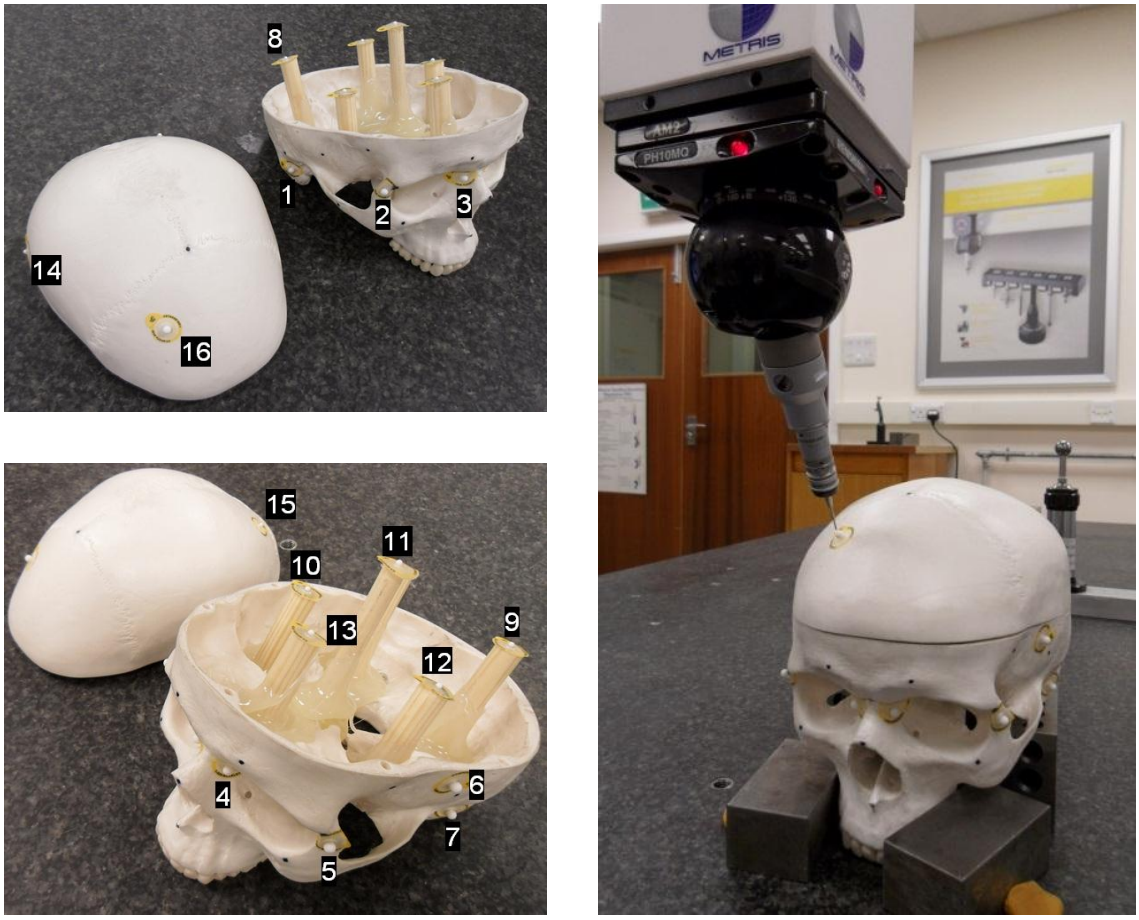
Figure 18: (a) Anatomical landmarks used, the medial canthus and tragus; and the expected TREs isocontours for FLE of 5 mm in the (b) sagittal, (c) axial and (d) coronal planes respectively[18].

3.6.7 TRE estimated on anthropomorphic skull

Because TRE cannot be measured directly, the TRE for the proposed registration framework has been determined experimentally using an anthropomorphic skull with 16 radiopaque markers attached at random

[18] MRI data from US National Library of Medicine's Visible Human Project®

locations inside and on the surface of the skull, to simulate the range of possible target locations (see Figure 19a). The positions of the radiopaque markers are known based on measurements using a coordinate measuring machine (see Figure 19b).



a | b

Figure 19: (a) Anthropomorphic skull and radiopaque markers used as fiducials for assessing TRE and (b) position of radiopaque markers in ‘skull space‘ determined using a coordinate measuring machine

The anthropomorphic skull was subsequently CT scanned (see Figure 20) and a point-based rigid body registration, using bony analogues of the medial canthus and tragus, was performed to register CT images of the skull to the skull itself.

Figure 20: CT scanning of the anthropomorphic skull with radiopaque markers attached

Because four landmarks were used for the determination of a 3x3 dimensional transformation matrix, the calculation is overestimated and has in general no solution. The rigid body transformation matrix, consisting of a rotation matrix, R, and translation matrix, t, are matrices that minimise the square FRE, as given by Equation 3.8:

$$FRE^2 = \frac{1}{N}\sum_{i}^{N} \omega_i^2 |Rx_i + t - y_i|^2 \qquad 3.8$$

where ω_i is the non-negative weighting factor for landmarks i = 1, 2,… N, and x, y are the landmarks locations in the two spaces, used to decrease the influence of less reliable fiducials. A closed form solution, developed by Arm, Huang & Blostein [122] was used to determine the transformation matrix, consisting of a rotation and translation component. The weighted centroid of the fiducial configuration is computed in each space, as given by Equation 3.9:

$$\bar{x} = \sum_{i}^{N} \omega_i^2 x_i / \sum_{i}^{N} \omega_i^2$$
$$\bar{y} = \sum_{i}^{N} \omega_i^2 y_i / \sum_{i}^{N} \omega_i^2 \qquad 3.9$$

The mean centred coordinates of each fiducial, as given in Equation 3.10, is obtained by subtracting the coordinates of each fiducial with the centroid:

$$\tilde{x}_i = x_i - \bar{x}$$
$$\tilde{y}_i = y_i - \bar{y} \qquad 3.10$$

The weighted fiducial covariance matrix, H, as given in Equation 3.9, is computed next:

$$H = \sum_{i}^{N} \omega_i^2 \tilde{x}_i \, \tilde{y}_i^t \qquad 3.11$$

Equation 3.12 gives the singular value decomposition of H:

$$H = U\Lambda V^t \qquad 3.12$$

where U is the unitary matrix, Λ is a diagonal matrix with nonnegative real numbers on the diagonal and V^t is the conjugate transpose of unitary matrix V. $U^tU = V^tV = I, \Lambda = diag(\lambda_1, \lambda_2, \lambda_3)$, and $\lambda_1 > \lambda_2 > \lambda_3$. I is an identity matrix and λ is the eigenvalue. The rotation matrix, R, is given by Equation 3.13:

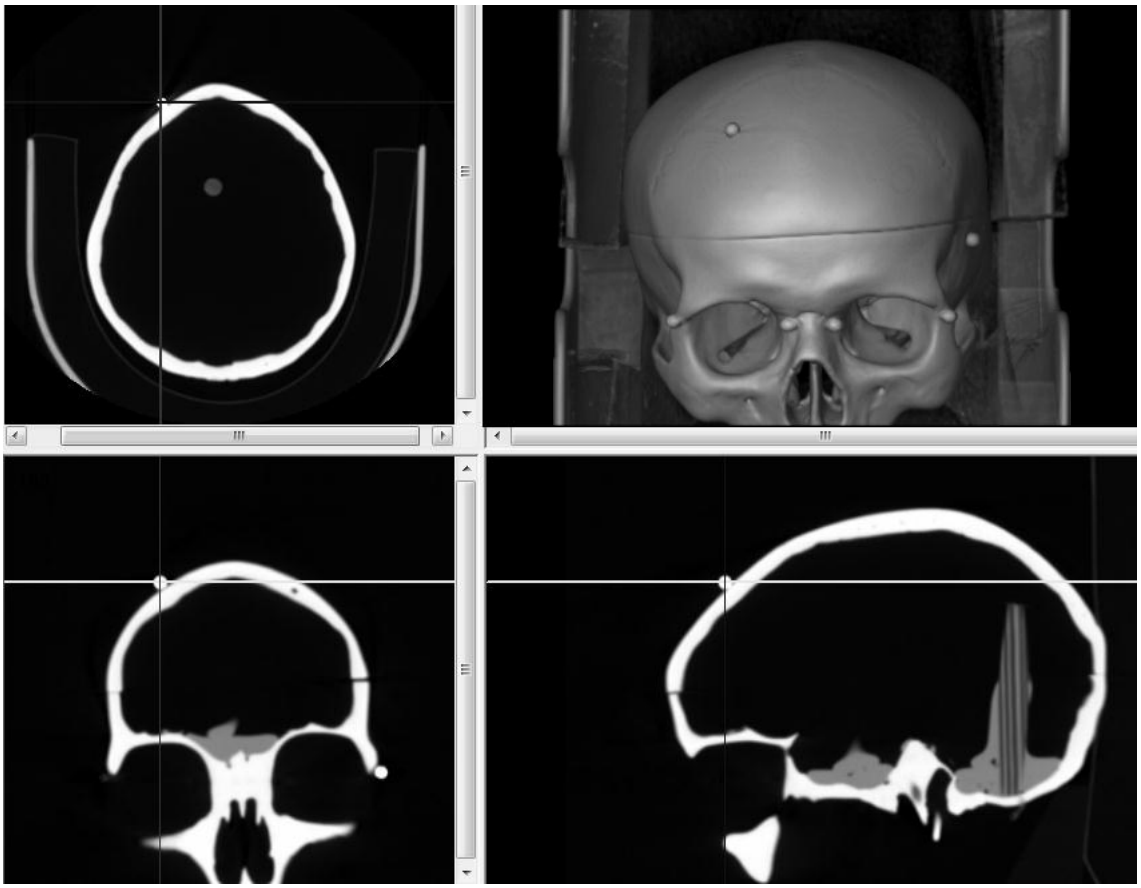
$$R = Vdiag(1,1,\det(VU))U^t \qquad 3.13$$

and the translation matrix, t, is given by Equation 3.14:

$$t = \bar{y} - R\bar{x} \qquad 3.14$$

Four radiopaque markers corresponding to bony analogues of the anatomical landmarks i.e. medial canthus and tragus (see radiopaque markers 1,3,4,7 in Figure 19a) were used as the registration basis for the point-based rigid body registration. The remaining 12 radiopaque markers were used as targets to assess the TRE. The location of the radiopaque markers on the skull was determined using a Metris LK Ultra coordinate measuring machine. The radiopaque markers in CT space were localised based on visual inspection

(see Figure 21). A FLE of 5 mm was incorporated into the localisation of the radiopaque markers in CT space corresponding to the anatomical landmarks by offsetting the localised position by 5 mm in random directions. The CT images of the skull were registered to the skull itself based on minimising FRE, as given by Equation 3.8. Both the FRE and the TRE were determined.



a	b
c	d

Figure 21: Localisation of radiopaque markers in CT space. Crosshairs are placed as close as possible to the centre of the radiopaque markers in the (a) axial view, (c) coronal view and (d) sagittal view. (b) Volume rendered model of the skull with radiopaque markers attached.

3.6.8 Results

Table 2 shows the CT and CMM coordinates of the 16 radiopaque fiducial markers.

Table 2: Radiopaque fiducial markers CT and CMM coordinate

Modality	CT			MRI		
Fiducials	Coordinates (in millimetres)					
	x	y	z	x	y	z
1	71.186	179.116	33.125	-22.528	7.370	-43.199
2	74.853	93.613	40.625	62.457	11.085	-32.158
3	118.887	78.135	39.375	78.187	55.296	-34.863
4	136.865	79.976	39.063	76.614	73.202	-35.871
5	181.265	91.035	42.500	65.489	117.591	-35.063
6	190.434	131.572	72.188	23.915	128.120	-7.663
7	190.801	174.326	38.438	-17.408	126.826	-43.297
8	90.996	201.230	106.250	-47.884	30.583	27.831
9	155.210	197.915	106.250	-44.111	94.797	25.387
10	88.428	139.311	90.625	14.669	27.334	15.144
11	121.836	131.514	124.688	21.286	62.044	48.000
12	160.347	138.208	95.000	15.872	99.054	16.218
13	122.920	101.719	93.750	52.163	61.826	18.656
14	56.875	176.167	102.813	-22.770	-3.792	27.007
15	158.315	216.269	120.000	-63.000	98.746	38.099
16	97.602	90.298	114.688	62.876	37.186	40.955

Figure 22 shows the FRE for radiopaque markers 1,3,4,7, which correspond to the right tragus, right medial canthus, left medial cantus and left tragus respectively. The mean FRE is 0.18 mm with a range of 0.10 mm – 0.26 mm.

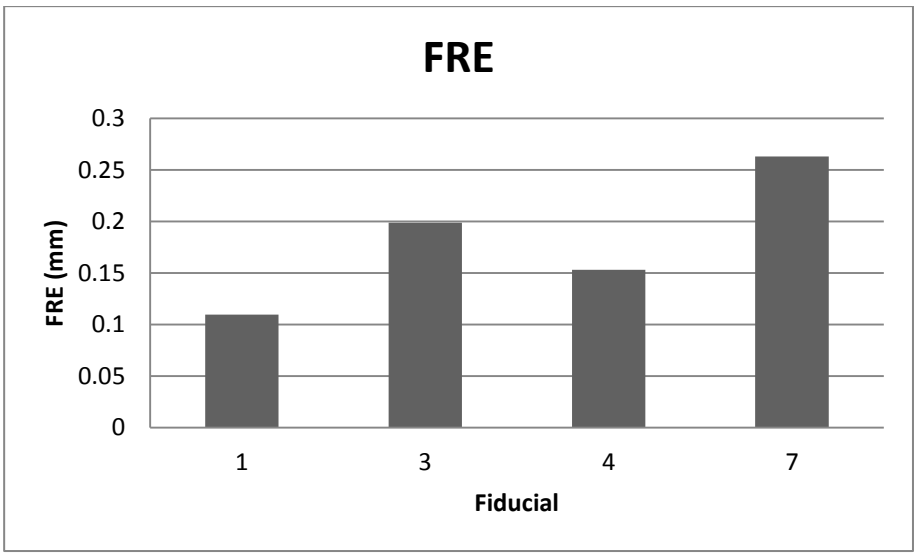


Figure 22: FRE for radiopaque markers 1,3,4,7, corresponding to the anatomical landmarks

Figure 23 shows the TRE for radiopaque markers 2, 5, 6, 8 - 16, corresponding to randomly positioned ‘targets’. The mean TRE is 0.28 mm (standard deviation = 0.11) with a range of 0.09 – 0.46 mm. The TRE compares very favourably to a study by Bickel [123] who obtained a mean TRE of 0.277 ± 0.696 mm and 0.266 ± 0.640 mm when using 5 landmarks to register an artificial and cadaver skull to their CT images.

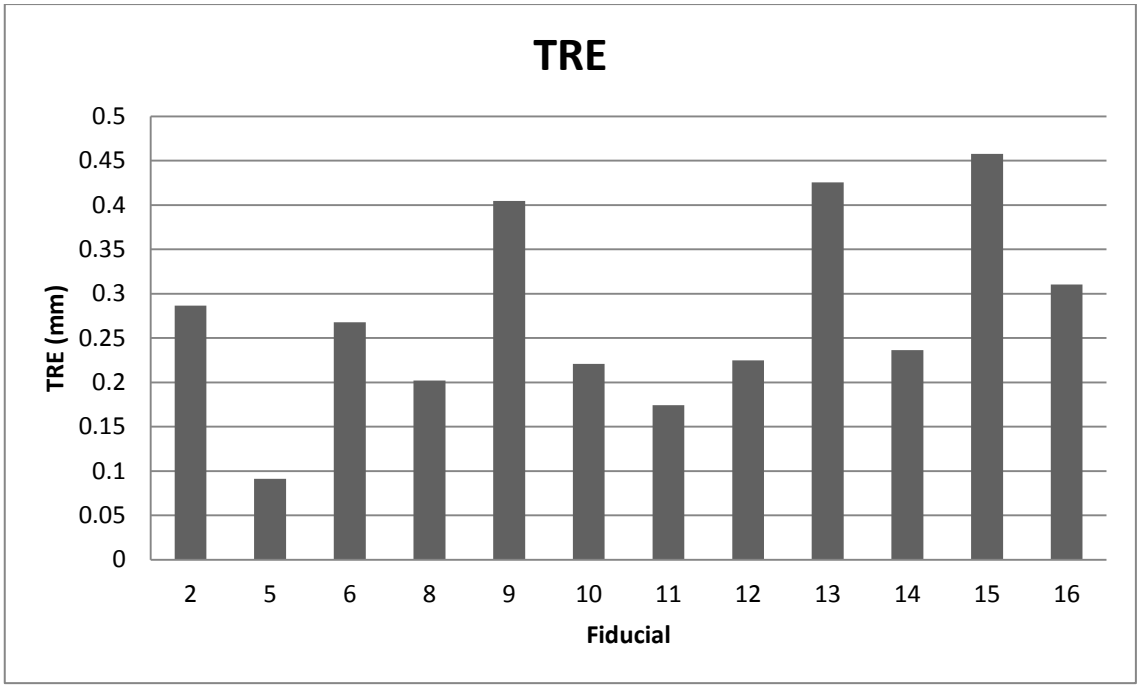


Figure 23: TRE for radiopaque markers 2, 5, 6, 8 - 16, corresponding to the 'targets'

Figure 24 shows the FRE for radiopaque markers 1,3,4,7, which correspond to the right tragus, right medial canthus, left medial cantus and left tragus respectively for a FLE of 5 mm. The mean FRE is 3.4 mm with a range of 3 – 4 mm.

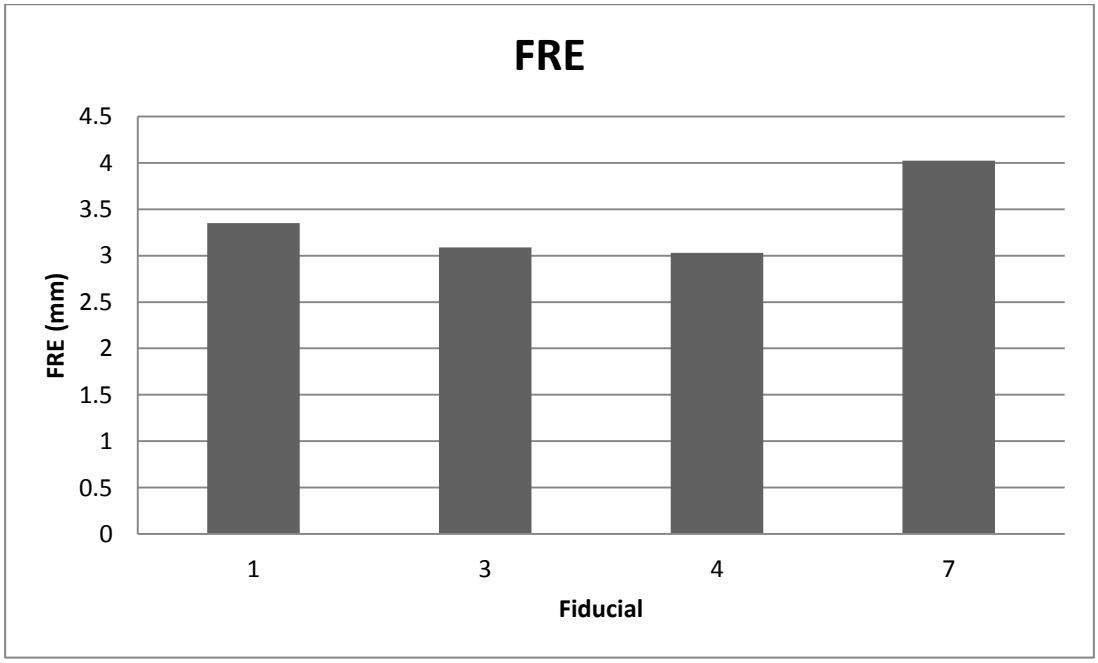


Figure 24: FRE for radiopaque markers 1,3,4,7, corresponding to the anatomical landmarks with FLE of 5 mm

Figure 25 shows the TRE for radiopaque markers 2,5,6,8-16, corresponding to randomly positioned ‘targets’ with a FLE of 5 mm. The mean TRE is 3.4 mm with a range of 2.8 – 3.8 mm. TRE estimated analytically based on the number of landmarks, accuracy in localising the landmarks and landmark configuration underestimates the actual TRE.

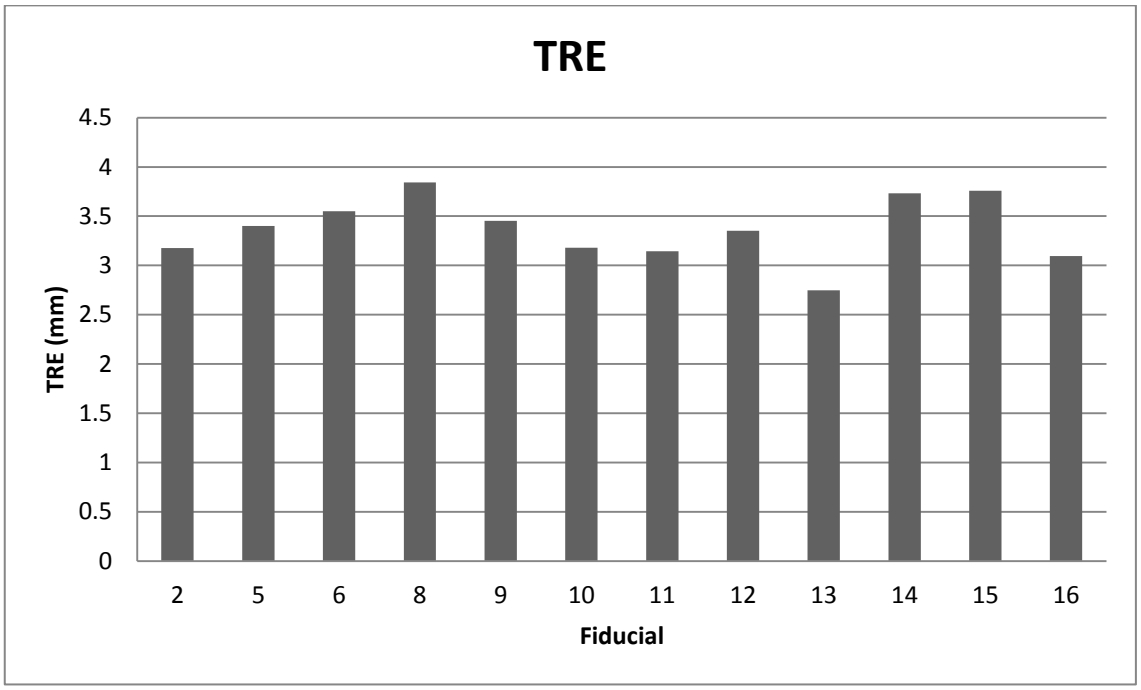


Figure 25: TRE for radiopaque markers 2, 5, 6, 8 - 16, corresponding to the ‘targets’ with FLE of 5mm

3.7 Discussion and conclusions

The estimated TREs for a FLE of 5 mm (obtained experimentally and based on the analytical expression) is in the millimetre range, and compares favourably to the accuracy achieved by a neurosurgeon performing the targeted neurosurgical procedures using a freehand technique, which is sometimes in the centimetre range. The discrepancies between the TREs obtained experimentally and those based on the analytical expression are because of the different assumptions regarding FLEs. TREs determined using the analytical expression were based on the assumption that the FLEs are independent, normally distributed random variables with zero means and equal variances while the FLEs used to determine TREs experimentally (based on

measurements using the anthropomorphic skull) were based on random 5mm errors with a non-zero mean and unequal variance. The TREs obtained in the phantom study are more indicative of the expected errors for a given landmark configuration and localisation error.

It should be noted that the TREs obtained would underestimate the actual errors achieved in practice. This is because several factors are not accounted for, such as patient movement, the accuracy of a robot to find a position in space, etc. However, errors due to these may be reduced by proper calibration of the robot as well as ensuring that the patient's movement is minimised. The patient's head for example, may be held relatively stationary by immobilising it with a clamp, as is common in neurosurgery, thereby reducing potential errors due to patient movement. A landmark-based approach can therefore yield a sufficiently accurate registration for the required clinical accuracy of the targeted procedures if the landmarks can be localised within 5 mm of their true position and others sources of errors can be reduced.

The limitation of the proposed registration framework is that the required anatomical landmarks i.e. medial canthus and tragus may be unreliable or absent, as would be the case in the occurrence of facial/head trauma that distorts a patient's normal anatomy. In this circumstance, the registration would have to be performed using skin-affixed markers, although this would mean that the patient would be required to undergo a secondary scan.

Another drawback of the proposed registration framework is that because the surgical targets are predefined at the preoperative stage, the system cannot compensate for any errors due to imaging, surgical tool deflection and tissue deformation. It may also be prone to errors if the brain shifts after acquisition of preoperative images due to an expanding haematoma, patient positioning or because of the surgery itself. However, this problem is not unique to the proposed registration framework but is common to all registration methods that use preoperative images for registration. Intraoperative brain shift can only be accounted for by intraoperative imaging e.g. intraoperative CT/MRI. Potential errors due to brain shift may be reduced by careful patient positioning e.g. placing a patient in a position similar to their position during imaging. Reducing the time between imaging and surgery may reduce errors due to

brain shift. Fortunately, brain shift is less of an issue with burr hole procedure than with an open craniotomy.

A framework for registering CT images to patient for MISEN has been described. Specifically, the proposed registration has been developed to support three emergency neurosurgical procedures. Because the three targeted neurosurgical procedures are intended to be performed robotically, image guidance is required. The registration system is a part of MISEN where guidance of the robot manipulator is based on registration between an image and the physical anatomy.

Simulation and experimental results of the registration framework based on a FLE of 5 mm showed an estimated TRE of within the millimetre range, which is the required accuracy for these procedures. In contrast, the accuracy of neurosurgeons performing the targeted procedures is sometimes in the centimetre range. Therefore, the proposed registration approach is sufficiently accurate and meets the required clinical accuracy of the targeted procedures. To reduce the subjectivity inherent when localising the landmarks, the automatic localisation of these landmarks is proposed. The next chapter describes the automatic localisation of these landmarks in CT images.

Chapter 4

Automating the Registration

4.1 Introduction

The use of anatomical landmarks as a basis for image to patient registration is appealing because the registration may be performed retrospectively. Two anatomical soft tissue landmarks of the head, the medial canthus and the tragus have been proposed as a registration basis for an automated CT image to patient registration system, as inputs for a rigid body registration algorithm.

In this chapter, algorithms for the automatic localisation of these landmarks in CT images are described. At present, anatomical soft tissue landmarks of the head are localised manually on CT images [124]. However, as CT images are digital in nature, a computational approach to the localisation of these landmarks was developed. The automatic localisation of these landmarks in CT images is an enabling step towards automating image to patient registration, with the aim of reducing the subjectivity inherent in the landmark selection process. A brief review of previous work on anatomical landmark localisation in CT images is presented, with the algorithm described in the subsequent section.

4.2 Previous work

Existing computational approaches to landmark localisation in CT images are predominantly semi-automatic and primarily for localising internal landmarks. The various approaches mainly differ in the types of landmarks localised. Frantz et al [124] used 3D generalisations of 2D differential operators/corner detectors to localise internal anatomical landmarks in head CT images such as the tips of the frontal, temporal, and occipital horns of the ventricular system; saddle point of the zygomatic bone; fourth ventricle, and tip of the external occipital protuberance. Their method however requires the user to specify a coarse position of the landmark of interest, as well as to select the most promising candidate among the detected candidate landmarks.

A more recent approach fits geometric and intensity models of these landmarks to their images [125, 126]. In this semi-automatic approach, the user is required to provide a coarse localisation of the landmarks, which the

algorithm then refines, by fitting models of anatomical structures to their images. Although this technique could potentially be applied to localising the tragus, and to a lesser extent the medial canthus, user intervention, to provide coarse localisation of these landmarks, would still be required.

Recently, Subburaj et al [127] described a method to localise anatomical landmarks of the knee-joint in CT images automatically, using surface curvature properties and the spatial adjacency of landmarks i.e. the relative location of a landmark with respect to other landmarks. However, spatial adjacency cannot be used in localising the medial canthus or tragus because there are no other landmarks close to these landmarks that can be used to facilitate their localisation.

Closely related to this work is the localisation of anatomical landmarks of the face in 3D range images. Deo and Sen [128] localised anatomical landmarks of the face such as the medial canthus on a surface mesh extracted from range images, using curvature analysis and morphological operations. The extremities of regions isolated by thresholding a surface mesh based on a predetermined mean curvature threshold range correspond to the landmarks. Their intended application was in anthropometry, the science and practice of human body measurements. Although they reported good results for the localisation of the medial canthus with their database of about 20 subjects, they were primarily interested in measuring distances between landmarks, rather than in localising the landmarks accurately. When measuring distances between landmarks, incorrect landmark positions may still provide correct inter landmark distance. Furthermore, the results were achieved with subjects whose eyes were open, which makes localisation of the medial canthus easier. In contrast, the dataset used in this study consists of high resolution images of 118 subjects. High resolution images, while affording greater localisation accuracy, are more susceptible to false detection.

Chen et al [129] presented a scale space based approach using integral volume descriptors, for localising anatomical landmarks of the face. Integral descriptors are more robust to noise than curvature based strategies. Similar to the work by Deo and Sen, they used 3D range images. Because of its relatively low resolution, although seemingly similar in nature, the majority of the work in the localisation of anatomical landmarks of the face in 3D range images is in

the detection of a region of interest, rather than a specific point. The context of the localisation is therefore different. Furthermore, unlike surface models rendered from range images, surface models reconstructed from CT are more detailed, and as a result, false detections are more of an issue when localising landmarks in CT images. This is especially true when localising the medial canthus, where the presence of deep-set wrinkles can make localisation of the medial canthus difficult. In this work, the medial canthus and tragus are localised in CT images based on curvature analysis and a rule-based system. To the author's knowledge, no prior work to localise these landmarks in CT images has been reported in the literature.

4.2.1 Localisation in patient space

The localisation of the medial canthus and tragus in patient space have previously been described by Gooroochurn et al [130], who used close range photogrammetry and a neural network approach based on Gabor filters to facilitate the localisation of these landmarks. Five cameras mounted on an arc, placed such that it is approximately facing the patient were used to detect the landmark positions on the patient (see Figure 26). This camera configuration ensures that each landmark is viewed from two different directions, as a minimum of two views are required in photogrammetry. Alternatively, a single camera that moves to predetermined positions on the arc may be used.

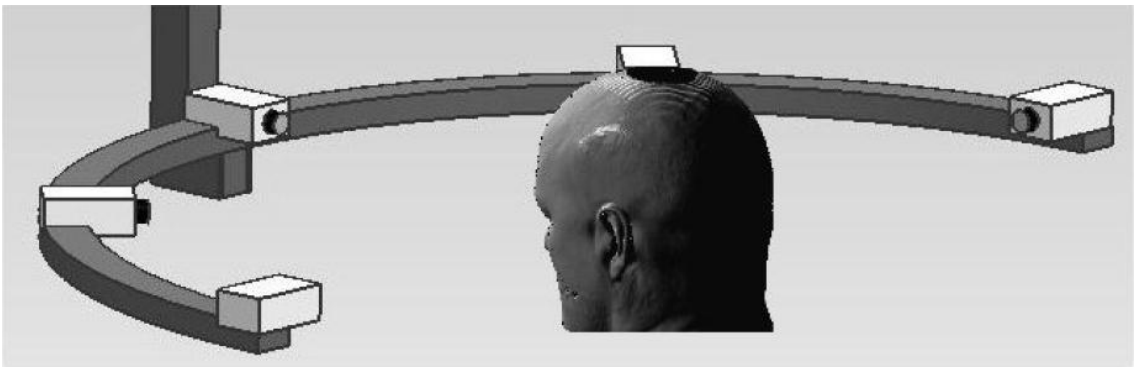


Figure 26: Close range photogrammetry camera setup for localising the medial canthus and tragus in patient space[19]

[19] Used with permission from author

The automatic localisation of these landmarks in CT space would complement the automatic localisation of these landmarks in patient space in an automated image to patient registration system. In the following section, the methodology for the automatic localisation of these landmarks is described.

4.3 Methodology

The medial canthus and tragus are localised on surface models extracted from CT images based on their curvature saliency and using a rule based system that incorporates prior knowledge of their geometric and spatial characteristics. Candidate landmark locations are first isolated on the surface models based on their curvature properties. A rule-based system was subsequently used to facilitate the localisation of these landmarks as the use of curvature properties alone will result in numerous false detections. The landmarks are localised within candidate locations that have a geometric structure and spatial location consistent with the prior knowledge of these landmarks. The various stages of the algorithm are detailed below.

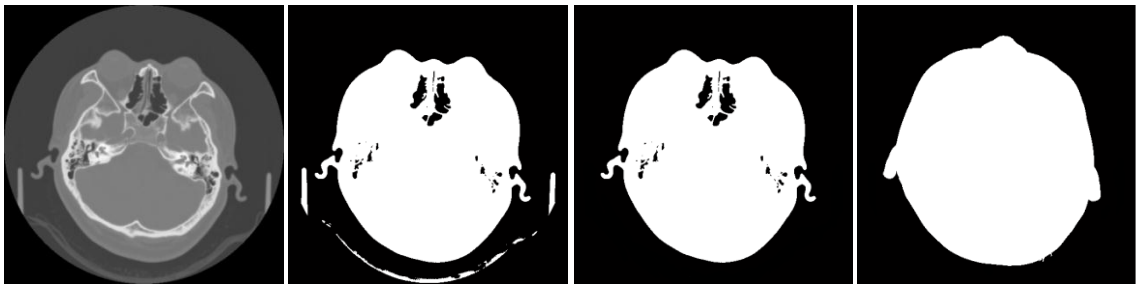
4.3.1 Head orientation

The first step in the algorithm is to detect and correct for any deviation of the head orientation from a forward facing position. This is because the localisation of the eye region is dependent on the head being in an approximately forward facing position. The left and right eye regions are localised based on a nearest neighbourhood clustering, with the centroids for the left and right eye region at the centre of the left and right sides of the bounding box containing the surface model. Clusters belonging to the eye regions may be erroneously localised if the head deviates from a frontal facing position. Figure 27 illustrates the left and right eye regions grouped based on nearest neighbourhood clustering for a deviation of 0º, 15º and 30º from the frontal facing position. While the eye clusters were correctly localised for a deviation of 15º (Figure 27b), they were wrongly localised for a deviation of 30º (Figure 27c). Therefore, any deviation greater than 15º has to be corrected.

a	b	c

Figure 27: Surface model illustrating deviation from the frontal facing position of (a) 0º, (b) 15º and (c) 30º and the corresponding left and right eye regions as grouped based on nearest neighbourhood clustering

Fortunately, it is relatively easy to detect and correct for any deviations from the required orientation. For a given set of axial CT images, a series of operations is performed to isolate the head region (see Figure 28). The head region is segmented in the axial CT images (Figure 28a) by thresholding (Figure 28b). To isolate the head from other structures, such as the headrest, a blob analysis was performed to eliminate spurious regions in the segmented images, whereby the largest connected binary component, which corresponds to the head, is chosen (see Figure 28c). The resulting binary axial CT images are stacked to create a composite image, which is approximately elliptical (see Figure 28d).



a	b	c	d

Figure 28: (a) Original CT axial image, (b) segmented image, (c) largest connected binary component and (d) composite image.

The grey level intensity distribution of a typical CT image of the head exhibits two well-defined peaks, representing the head region and surrounding

air respectively, with a valley separating the two peaks (see Figure 29). Intensity-based thresholding therefore works well in delineating the skin to air boundary, with the threshold level determined automatically using Otsu's method [131].

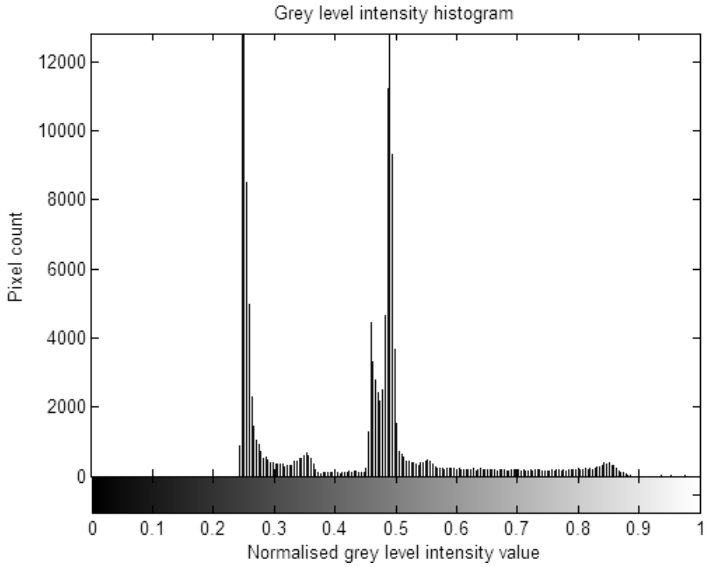


Figure 29: Grey level intensity histogram of CT axial image in Figure 28

The orientation, θ of the head is then estimated as the 2D orientation of an ellipse with the same normalised second central moments as the composite image, and is given by Equation 4.15:

$$\theta = \tan^{-1}\left(\frac{\mu_{02} - \mu_{20} + \sqrt{(\mu_{02} - \mu_{20})^2 + 4\mu_{11}^2}}{2\mu_{11}}\right) \qquad 4.15$$

where $\mu_{20} = \frac{\sum(x-\bar{x})^2}{N}, \mu_{02} = \frac{\sum(y-\bar{y})^2}{N}, \mu_{11} = \frac{\sum(x-\bar{x})(y-\bar{y})}{N}$ are the second order central moments, x and y the x and y pixel coordinates, $\bar{x}$ and $\bar{y}$ the mean x and y pixel coordinates and N the number of pixels

The orientation estimated using this method becomes less accurate the closer the shape of the composite image is to a circular shape. This is because the terms $\mu_{02} - \mu_{20}$ and μ_{11} become smaller, leading to numerical instability. As the head is only required to be approximately forward facing, only deviations of

more than 15° (see Figure 27) from the vertical axis are corrected. If required, a rigid body rotation of the CT images, about its centroid, is used to re-orient the CT images, using a rotation matrix, R given by Equation 4.16:

$$R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \quad 4.16$$

To determine the validity of the proposed approach, a set of CT images were rotated by a known angle and the angle was subsequently calculated. For the given set of CT images that was rotated by 45.000°, the orientation of the composite image was estimated as 45.007°, an error of 0.007°. The orientation of the composite image, estimated as the 2D orientation of an ellipse with the same normalised second central moments, is therefore very accurate. However, this is the accuracy of the relative orientation with respect to a prior orientation, rather than the absolute orientation. For the purposes of reorienting CT images of the head however, a deviation of ±15º from the forward facing position is acceptable, as the head has to be approximately in this position only. Once properly oriented, the surface model of the head is extracted from the CT images.

4.3.2 Surface model

An important technique in visualising volumetric data such as CT in surface form is using an isosurface. An isosurface is a 3D three-dimensional analogue of isocountours, and for CT, is a set of all the points that have the same Hounsfield number i.e. constant intensity. Figure 30 illustrates the extraction of an isosurface where the intensity of each voxel is represented by different grey levels.

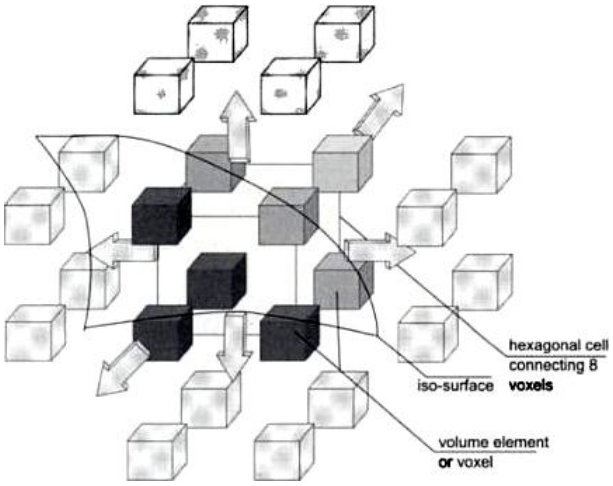
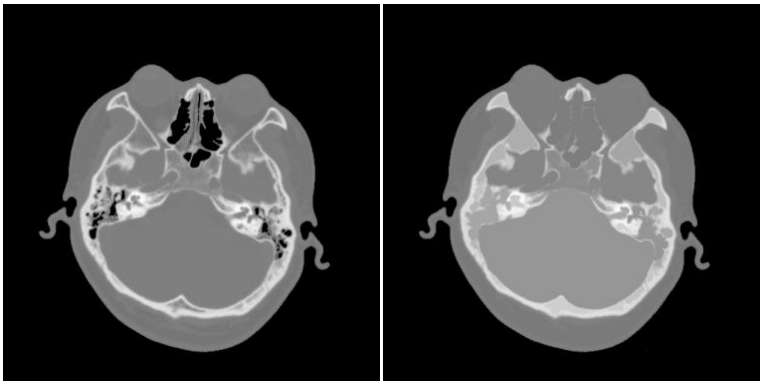


Figure 30: An isosurface extracted from volume data[20] [132]

To obtain a surface model of the head from CT images, masks consisting of binary axial CT images of the segmented head region obtained earlier are multiplied with the original CT axial images (see Figure 28). A grey level morphological fill operation was then performed on the segmented image to remove any holes (see Figure 31), as they presence of these holes would complicate the automatic localisation of the landmarks.



a | b

Figure 31: (a) Head region with 'holes' and (b) without 'holes' after grey level morphological fill operation

A surface model of the head, represented as a triangulated mesh, was then extracted using a custom isosurface algorithm [133], with an isovalue equal to the threshold level used earlier for the segmentation of the head region

[20] Image reproduced after permission from SPIE

in CT images (see Section 4.3.1). For accurate surface reconstruction from CT images, an exact determination of the skin to air boundary is required. However, this boundary is ill defined in CT images, represented as a continuum of grey levels. As such, only an approximation of the actual surface is possible. A sample of surface models of the head for eight subjects, extracted from CT images, is shown as in Figure 32. The curvature of these surface models is computed next.

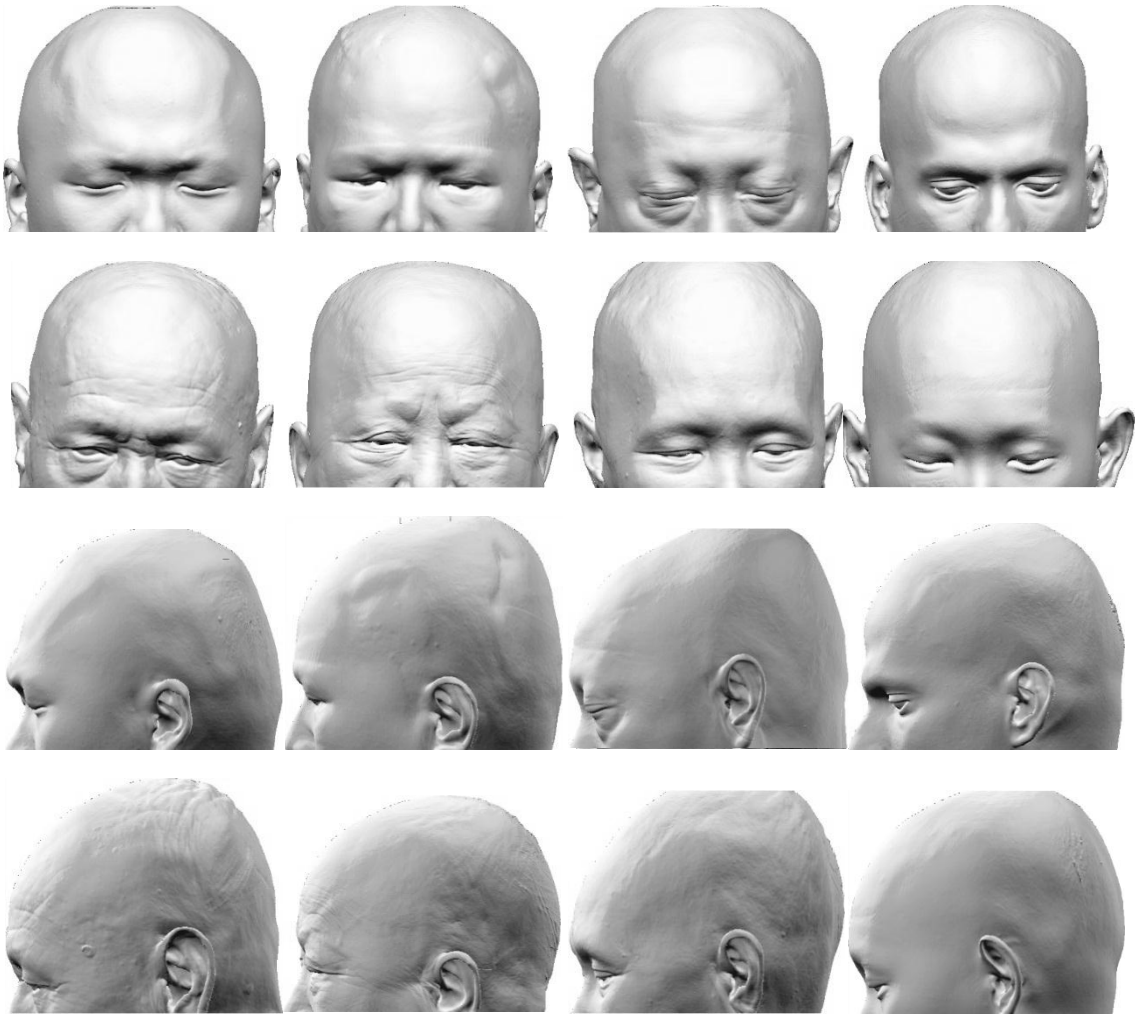


Figure 32: Sample of head surface models extracted from CT images (Frontal view and corresponding profile view)

4.3.3 Surface curvature

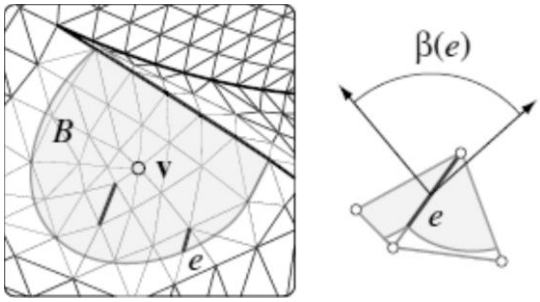
While computing the curvature of a parametric surface is trivial, the computation of curvatures of an isosurface represented as triangular meshes is

difficult. Unlike a parametric surface, the curvature of a triangulated mesh is not well defined, as it is piecewise linear and cannot be differentiated. There are three approaches to calculate the curvature of a surface represented by a mesh: fitting methods, discrete methods, and curvature tensor estimation methods. In fitting methods, which are computationally expensive, an analytic function, whose curvature can be computed, is fitted to a mesh locally. Discrete methods involve a direct estimation of the curvature at each vertex, by summing the curvature of each face or edge associated with the vertex. These methods are appealing because of their speed. However, they are sensitive to noise and mesh resolution. Curvature tensor estimation methods are similar to discrete methods, except that instead of estimating the curvature directly, a curvature tensor is estimated. Curvatures and principal directions are derived from the eigenvalues and eigenvectors of the curvature tensor. These methods are computationally less complex than fitting methods.

A curvature estimation method by Alliez et al [134] was used to estimate the curvature of an isosurface. In their method, a curvature tensor is defined at each point along the edge, by observing that for every edge e of the mesh, the minimum curvature is along the edge, and the maximum curvature is across the edge. A curvature tensor can therefore be defined at each point along an edge. The curvature tensor, $\mathfrak{I}$ of an arbitrary region, B is the average of these individual tensors over the region. Formally, the curvature tensor, $\mathfrak{I}(v)$ at an arbitrary vertex, v on the mesh, over B is given by Equation 4.17:

$$\mathfrak{I}(v) = \frac{1}{|B|} \sum_{edges\ e} \beta(e)|e \cap B|\bar{e}\bar{e}^t \qquad 4.17$$

where $|B|$ is the surface area around v over which the tensor is estimated, $\beta(e)$ is the signed angle between the normals to the two oriented triangles incident to edge e (positive if convex, negative if concave), $|e \cap B|$ is the length of $e \cap B$ (always between 0 and $|e|$), and $\bar{e}$ is a unit vector in the same direction as e (see Figure 33).



a | b

Figure 33: (a) Illustration of a region, B, vertex, v, edge, e, and (b) signed angle between the normals to the two oriented triangles incident to edge, $\beta(e)$ [134]

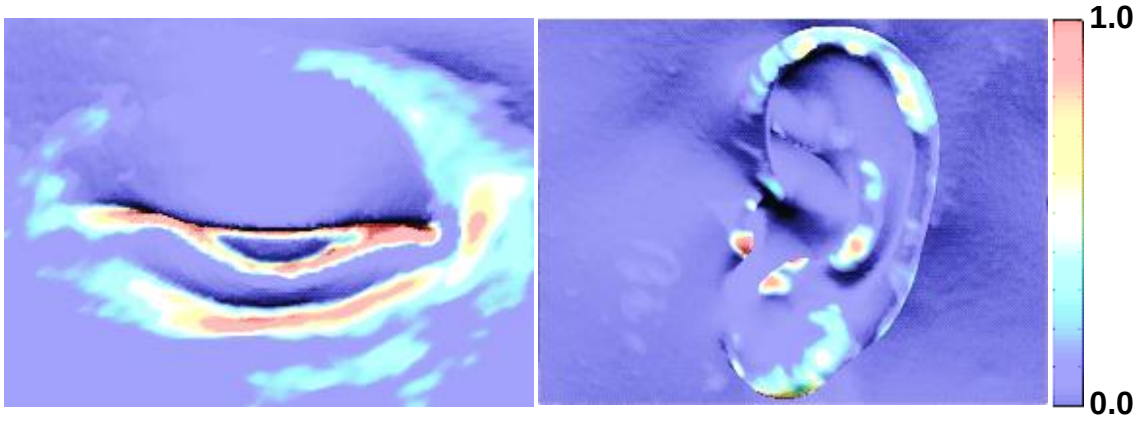
To obtain a continuous tensor field over the whole surface, the piecewise curvature tensor is linearly interpolated at each vertex. The normal at each vertex is the eigenvector of the curvature tensor $\mathfrak{I}$ associated with the minimum eigenvalue. The principal curvatures at v, $\varkappa_{min}$ and $\varkappa_{max}$, are the two other eigenvalues of curvature tensor $\mathfrak{I}$. The Gaussian and mean curvature can be computed from the principal curvatures. The Gaussian curvature, $\mathrm{C}_{gaussian}$ is the product of the principal curvatures, κ_{min} and κ_{max} and is given by Equation 4.18:

$$\mathrm{C}_{gaussian} = \kappa_{min}\kappa_{max} \qquad 4.18$$

The mean curvature, C_{mean} is the average of the principal curvatures, κ_{min} and κ_{max} and is given by Equation 4.19:

$$\mathrm{C}_{mean} = \frac{1}{2}(\kappa_{min} + \kappa_{max}) \qquad 4.19$$

Figure 34 illustrates the mean curvature map of the eye and the Gaussian curvature map of the ear, determined using the curvature tensor estimation algorithm. A neighbourhood B that approximates a disk around v that is within a sphere centred at v , with a radius equal to 1/100th of a bounding box diagonal of the surface (as used by Alliez et al [134]), was used. The curvature values have been normalised and colour coded for display purposes. Red and blue corresponds to areas of high and low curvature respectively.



a | b

Figure 34: (a) Mean curvature map of the eye region and (b) Gaussian curvature map of the ear region

To reduce the effect of noise when estimating surface curvature, a 10-ring (neighbourhood) averaging of the curvature tensor at each vertex was applied. Figure 35 shows the concept of a ring neighbourhood for a triangular mesh and Figure 36 illustrates the effect of the size of this ring when averaging curvature values. Curvature values are smoother as the curvature tensor is averaged over larger neighbourhoods. However, this has the effect of masking surface detail and blurring curvature estimates. Nevertheless, the positions of the landmarks are robust to smoothing using a 10-ring averaging.

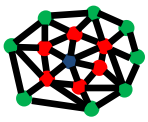


Figure 35: A triangular mesh. The vertices in red and green are the 1-ring and 2-ring neighbourhood of the vertex in blue

a | b | c | d

Figure 36: Thresholded mean curvature map at (a) 0-ring, (b) 1-ring, (c) 5-ring and (d) 10-ring neighbourhood averaging filter sizes

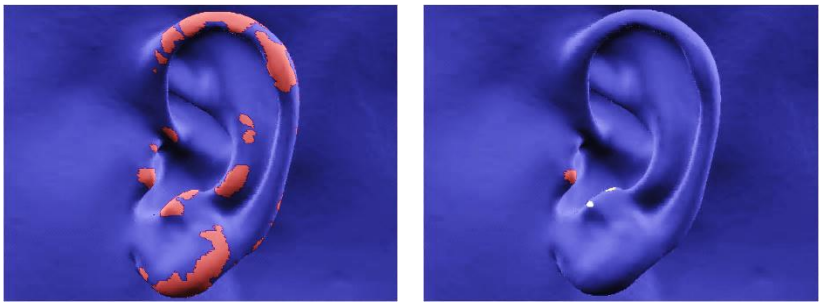
4.3.4 Candidate landmark regions

Because the medial canthus and tragus are regions of high curvature, an obvious way to localise them would be to use curvature properties. By applying an appropriate surface curvature threshold level, the medial canthus and tragus can be localised, due to their location on curvature extremes of contours. Geometrically, the medial canthus is the most medial point of a valley surface formed by intersection of the upper and lower eyelids. This valley surface extends from the outer corner to the inner corner of the eye and is characterised by high mean curvature and zero Gaussian curvature. Candidate locations for the medial canthus are therefore isolated by thresholding the surface curvature based on mean curvature, $C_{mean} \geq threshold$, and Gaussian curvature threshold $C_{gaussian} = 0$ (see Figure 37).

a	b	c	d

Figure 37: Curvature map with $C_{gaussian} = 0$ and $C_{mean} \geq$ (a) 0.005, (b) 0.010, (c) 0.015 and (d) 0.020

The tragus is approximately semi-ellipsoid, and its apex is the tip of this peak structure, a locally maximum point. It is characterised by negative mean curvature and high Gaussian curvature (see Figure 34). Candidate locations containing the tragus are identified by thresholding the surface curvature based on Gaussian curvature, $C_{gaussian} \geq threshold$ and $C_{mean} < 0$ (see Figure 38).



a | b

Figure 38: Curvature map with $C_{mean} < 0$ and (a) $C_{gaussian} \geq 0.0001$ (b) $C_{gaussian} \geq 0.0005$

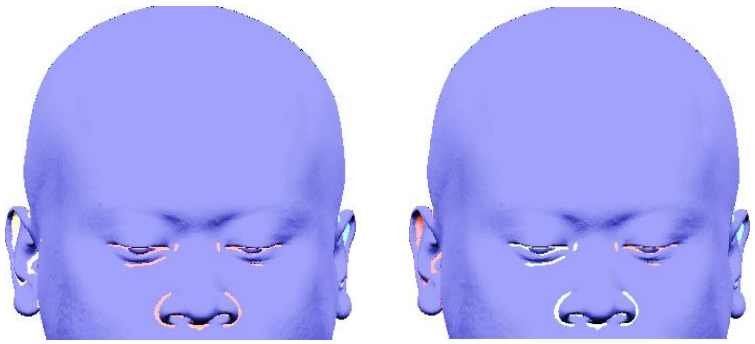
While the amount and the type of curvature of a point on a surface are indicative of its local shape, Deo and Sen [128] used mean curvature only in their work. Using two principal curvatures i.e. mean and Gaussian curvature would yield more information about local shape. The mean and Gaussian curvature properties used are based on the surface geometry of the landmarks and the threshold level used was determined empirically. A threshold based on an arbitrary upper limit of a curvature histogram was considered but the presence of outliers due to imperfect segmentation, e.g. parts of the headrest, greatly affected the resulting threshold value.

For the medial canthus, a mean curvature $C_{mean} > 0.015$ with a Gaussian curvature $C_{gaussian} = 0$ (see Figure 37) worked well in the majority of cases, although in some instances (Section 3.7) the mean curvature threshold was varied to two predefined levels i.e. 0.010 (see Figure 37b) and 0.020 (see Figure 37d). For the tragus, a Gaussian curvature $C_{gaussian} \geq 0.0005$ (see Figure 38) and negative mean curvature $C_{mean} < 0$ worked well. Once isolated, candidate landmark regions were clustered to facilitate landmark localisation.

4.3.5 Clustering

Curvature properties alone are inadequate to localise the landmarks because there will be many false positives, as other structures with similar surface curvature values will be detected as well. To facilitate the localisation, an analysis of the clusters and knowledge of the characteristics of the landmarks were used. Candidate regions that may contain the landmarks are

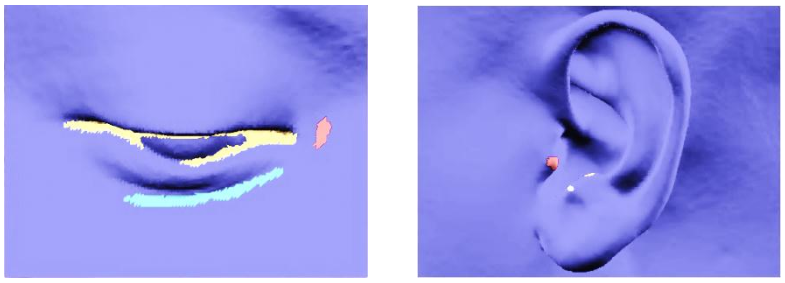
first identified, with the landmarks subsequently localised within these regions. This is because localising the region containing the landmarks is less error prone than localising the landmarks by themselves. The eyes (including the nose) and ears (see Figure 39a) are clustered into larger regions. As these regions tend to have arbitrary shapes, a density-based clustering algorithm [135], which works well with these types of shapes, was used.



a | b

Figure 39: Nearest neighbour clustering, of (a) right ear, eyes and nose, left ear and of (b) right ear, right eye, nose, left eye and left ear.

The regions are subsequently grouped using nearest neighbour clustering, with the eyes and nose area further divided into the right eye, the left eye and the nose (see Figure 39b). The seed points used for the nearest neighbour clustering are based on the relative positions of these anatomies e.g. the seed points for the right and left ear were located at the extreme right and left of the surface model. As each anatomical area will usually have more than one candidate landmark region, these regions were subsequently divided into individual clusters (see Figure 40).



a | b

Figure 40: Density-based clustering of a thresholded surface curvature map of the (a) eye and (b) ear. Each colour represents a separate cluster.

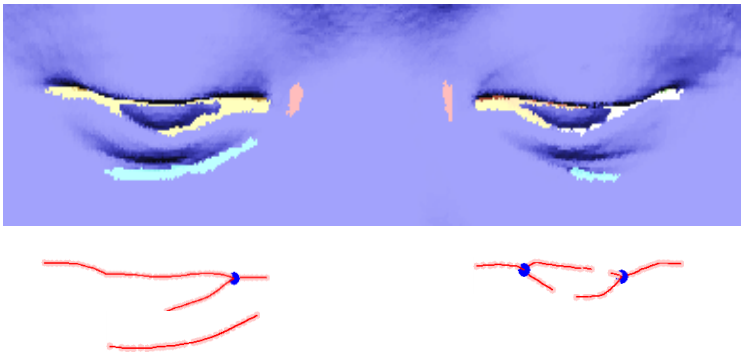
4.3.6 Curve skeleton extraction

When multiple clusters are present in the eye region, the cluster containing the landmark has to be identified. For the medial canthus, part of the difficulty in localising this cluster is that the shape of the cluster is different depending on whether the eyes are closed, opened or partially opened (see Figure 41).

a	b	c

Figure 41: Clusters containing the medial canthus with the eyes (a) closed, (b) opened and (c) partially opened (only main clusters are shown).

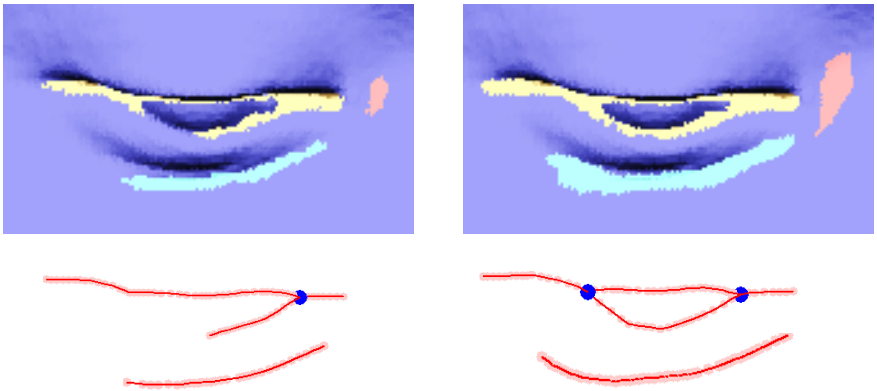
To identify the cluster containing the medial canthus, the curve skeleton of the clusters is extracted, and the number of branch points/nodes computed. A curve skeleton is a one dimensional representation of a three dimensional mesh or point cloud. The number of branch points of a curve skeleton is an indication of whether a cluster contains the medial canthus. Curve skeleton extraction is based either on volumetric or geometric techniques. Volumetric methods require voxelisation/discretisation of the cluster, resulting in lower resolution and loss of detail. On the other hand, geometric methods work directly on triangle meshes. Figure 42 illustrates the curve skeleton of the clusters in the eye region, based on a geometric method by Cao et al [136]. The curve skeletons are extracted directly on point cloud data, without the need to reconstruct the surface of an object.



a
b

Figure 42: (a) Clusters of a thresholded surface curvature map of the eye region and (b) the corresponding curve skeleton. Branch points are indicated as nodes.

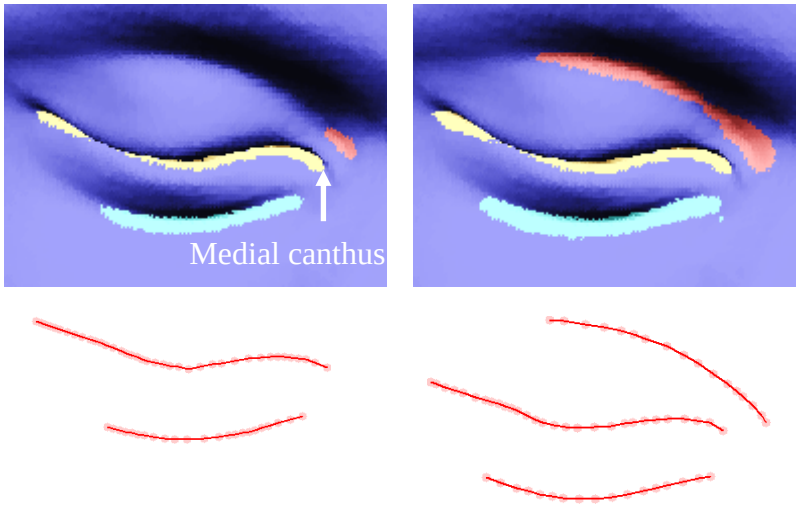
The goal is to determine if any of the clusters have curve skeletons with two branch points, as this corresponds to the geometry of a cluster with the eyes open or partially opened, and is therefore the cluster containing the medial canthus. Figure 43a illustrates the corresponding branch points of the curve skeleton of the clusters around the eye, at $C_{mean} \geq 0.015$ and $C_{mean} \geq 0.010$ for $C_{gaussian} = 0$. Based on the initial mean curvature threshold, $C_{mean} = 0.015$ neither cluster contains two branch points (one cluster contains a branch point and the other cluster contains no branch point). A curve skeleton with a single branch point does not correspond to the cluster containing the medial canthus. Decreasing the mean curvature threshold to 0.010 yields a cluster containing two branch points (see Figure 43b).



a
b

Figure 43: Clusters and the corresponding curve skeletons for $C_{gaussian} = 0$ and mean curvature threshold (a) $C_{mean} \geq 0.015$ and (b) $C_{mean} \geq 0.010$

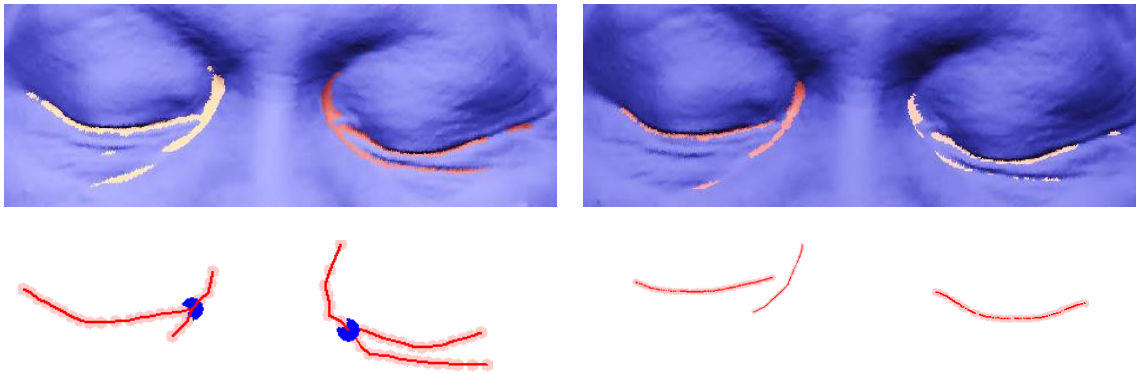
The medial canthus corresponds to the most medial point of the cluster corresponding to the curve skeleton with two branch points. If no cluster with a curve skeleton containing either no branch points or two branch points is found, the process is repeated with $C_{mean} \geq 0.020$. If a cluster with a curve skeleton containing two branch junctions is still not found, the eyes are either closed or partially opened. In this case, the cluster whose curve skeleton is the longest, based on the initial threshold, is the cluster containing the medial canthus (see Figure 44). This simplistic approach to localising the cluster containing the medial canthus works well (for images in our dataset). Curve skeleton length is a more appropriate measure than the Euclidean distance between the two extreme points of a cluster because it takes into account the surface shape.



a | b

Figure 44: Clusters and corresponding curve skeletons for $C_{gaussian} = 0$ and mean curvature threshold (a) $C_{mean} \geq 0.015$ and (b) $C_{mean} \geq 0.010$

Finally, if no cluster containing the medial canthus was found i.e. curve skeleton with no branch points or two branch points based on $C_{mean} \geq 0.010$ and $C_{mean} \geq 0.015$ (see Figure 45a), curve skeletons are extracted from clusters based on $C_{mean} \geq 0.020$ (see Figure 45b). The algorithm for localising the medial canthus is summarised in Section 3.7.



a	b

Figure 45: Curve skeletons for $C_{gaussian} = 0$ and mean curvature (a) $C_{mean} \geq 0.015$ and (b) $C_{mean} \geq 0.020$. Only main curve skeletons are shown.

4.3.7 Algorithm for the localisation of the medial canthus

Input: CT images

Output: Vertices corresponding to the left and right medial canthus $\{v_{lmc}, v_{rmc}\}$.

1. Reorientate the head if the head orientation in the axial plane deviates more than 15^{o} from the forward facing position (see Figure 27).
2. Extract surface model of head from CT images (see Figure 32).
3. Compute surface curvature (see Figure 34)
4. Find candidate vertices, $v_j \in V\{j = 1,2,\dots,n\}$ by thresholding based on $C_{mean} \geq 0.015$ and $C_{gaussian} = 0$ (see Figure 26).
5. Group each vertex, $v_j \in C_i\{i = 1,2,3\}$ using density based clustering.
6. Isolate resulting clusters, corresponding to the left ear, the eye and nose, and the right ear, using nearest neighbour clustering (see Figure 39a).
7. Separate the eye and nose, $d_j \in c_2\{i = 1,2,3\}$ into three clusters, corresponding to the left eye, the nose and the right eye, using nearest neighbour clustering (see Figure 39b).
8. If there is no candidate cluster for the eye i.e. d_1 and/or d_3, go back to step 1 but change the mean curvature threshold to 0.010.
9. If there is only one candidate cluster for the eye, select the cluster, $e_1 \in d_1$ and $e_2 \in d_3$ if it has either no branch points or two branch points.

Otherwise, go back to step 2 but threshold using $C_{mean} \geq 0.010$ in the first instance and $C_{mean} \geq 0.020$ in the second instance.

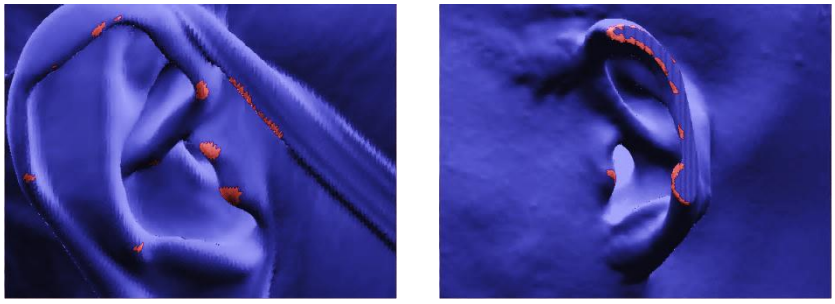
10. If there is more than one candidate cluster (see Figure 42 - Figure 45), select the cluster, $e_1 \in d_1$ and $e_2 \in d_3$ that has two branch points (see Figure 43b). Otherwise, go back to step 2 but change mean curvature threshold to 0.020 in the first instance and 0.010 in the second instance.
11. If no cluster containing two branch points is found, the cluster with the longest curve skeleton length and no branch points, based on $C_{mean} \geq 0.015$, is selected as the cluster containing the medial canthus (see Figure 44a). Otherwise, threshold based on $C_{mean} \geq 0.020$.
12. Select most medial vertex, $v_{lmc} \in e_1$ for left medial canthus and most medial vertex, $v_{rmc} \in e_2$ for right medial canthus.

4.3.8 Localisation of the tragus

To facilitate the localisation of the tragus and to avoid large localisation errors, a coarse location of the tragus is first estimated (see Figure 46).

Figure 46: Estimated tragus positions

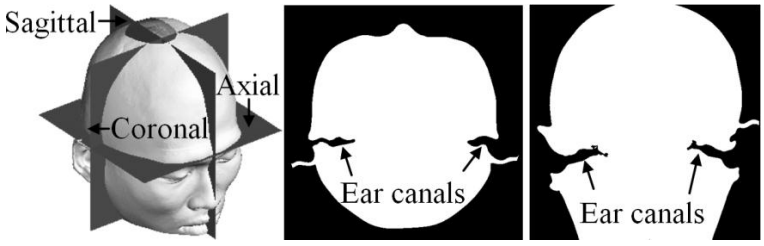
Due to the anatomy of the ear, noise, and the fact that some parts of the ear are sometimes truncated in CT images, the ear region may contain false clusters i.e. not the tragus (see Figure 47).



a | b

Figure 47: Thresholded Gaussian curvature map of the ear with negative mean curvature illustrating (a) noise and (b) part of the ear truncated in the surface model

An estimated location of the tragus is first found by exploiting their proximity to the ear canals, a more easily localised structure. A more accurate localisation is then found close to the estimated location on a surface rendered model. The position of the tragus is estimated by looking for the presence of ear canals in the CT axial and coronal planes (see Figure 48); and the outer ear structure in the CT sagittal plane (see Figure 49).



a | b | c

Figure 48: (a) Anatomical planes of reference, (b) ear canals in CT axial plane and (c) ear canals in CT coronal plane

The sagittal plane where the tragus is approximately located may be found by scanning from the most lateral to the most medial sagittal plane (see Figure 49) until the Euler number of the largest connected binary region is equal to or is less than zero. In practice, a line is inserted at the bottom of the image to compensate for possible partial ear structures. This is because the whole ear does not always appear in CT images, as the scan end location is at the base of the skull.

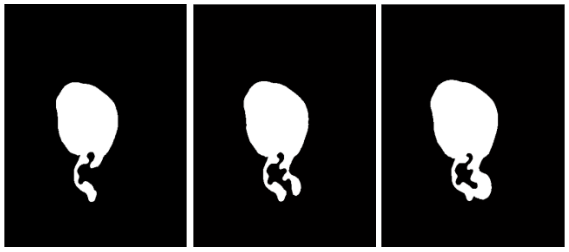
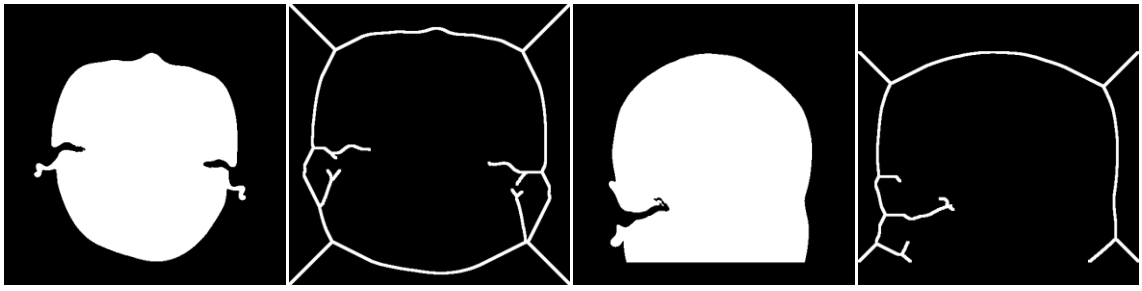


Figure 49: (from left to right) Sequence of the outer ear structure in CT sagittal view i.e. lateral to medial.

The ear canals in the axial and coronal CT planes are prominent features that can be robustly localised. A gross localisation of the tragus may be found by looking for the axial slice where the ear canal is longest. A possible approach to detect the CT slice where the ear canal is longest is to perform a skeletisation of the CT images and look for the medial most terminating point of the skeleton. This approach is illustrated in Figure 50. The axial and coronal planes whose skeleton endpoints are closest to the mid-sagittal plane correspond to the axial and coronal planes where the tragus is approximately located.



a	b	c	d

Figure 50: (a) Ear canals in CT axial plane, (b) corresponding image skeleton, (c) ear canals in CT coronal plane and (d) corresponding image skeleton

The largest cluster, within 10 mm (determined empirically) from the estimated tragus location, is selected as the cluster containing the tragus. The vertex with the highest Gaussian curvature value in this cluster corresponds to the apex of the tragus. The algorithm for the localisation of the tragus is summarised in Section 3.9.

4.3.9 Algorithm for the localisation of the tragus

Input: CT images

Output: Vertices, $\{v_{rt}, v_{lt}\}$, corresponding to the left and right tragus.

1. Reorientate the head if the head orientation in the axial plane deviates more than 15^{o} from the forward facing position (see Figure 27).
2. Extract surface model of head from CT images (see Figure 32).
3. Compute surface curvature (see Figure 34)
4. Estimate location of the tragus (see Figure 46). This corresponds to the axial (see Figure 48b) and coronal (see Figure 48c) plane where the ear canals appears longest based on skeletonisation (see Figure 50), and the first sagittal plane (from lateral to medial), with a region whose Euler number is equal to, or less than, zero (see Figure 49).
5. Find candidate vertices, $v_j \in V_i \{j = 1,2,\dots,n\}$ by thresholding based on $C_{gaussian} \geq 0.0005$, and $C_{mean} < 0$ (see Figure 38).
6. Group each vertex, v_j, into clusters, $C_k \{k = 1,2,\dots.,p\}$ using density based clustering (see Figure 40b)
7. Select the largest cluster, $c \in C_k \{k = 1,2,\dots.,p\}$, within 10 mm from the estimated tragus location.
8. The vertices, $v_{lt}, v_{rt} \in c$ with the highest Gaussian curvature value correspond to the apex of the left and right tragus respectively.

4.3.10 Overview of the automatic localisation algorithm

The various stages in the algorithm are shown in flowchart form in Figure 51.

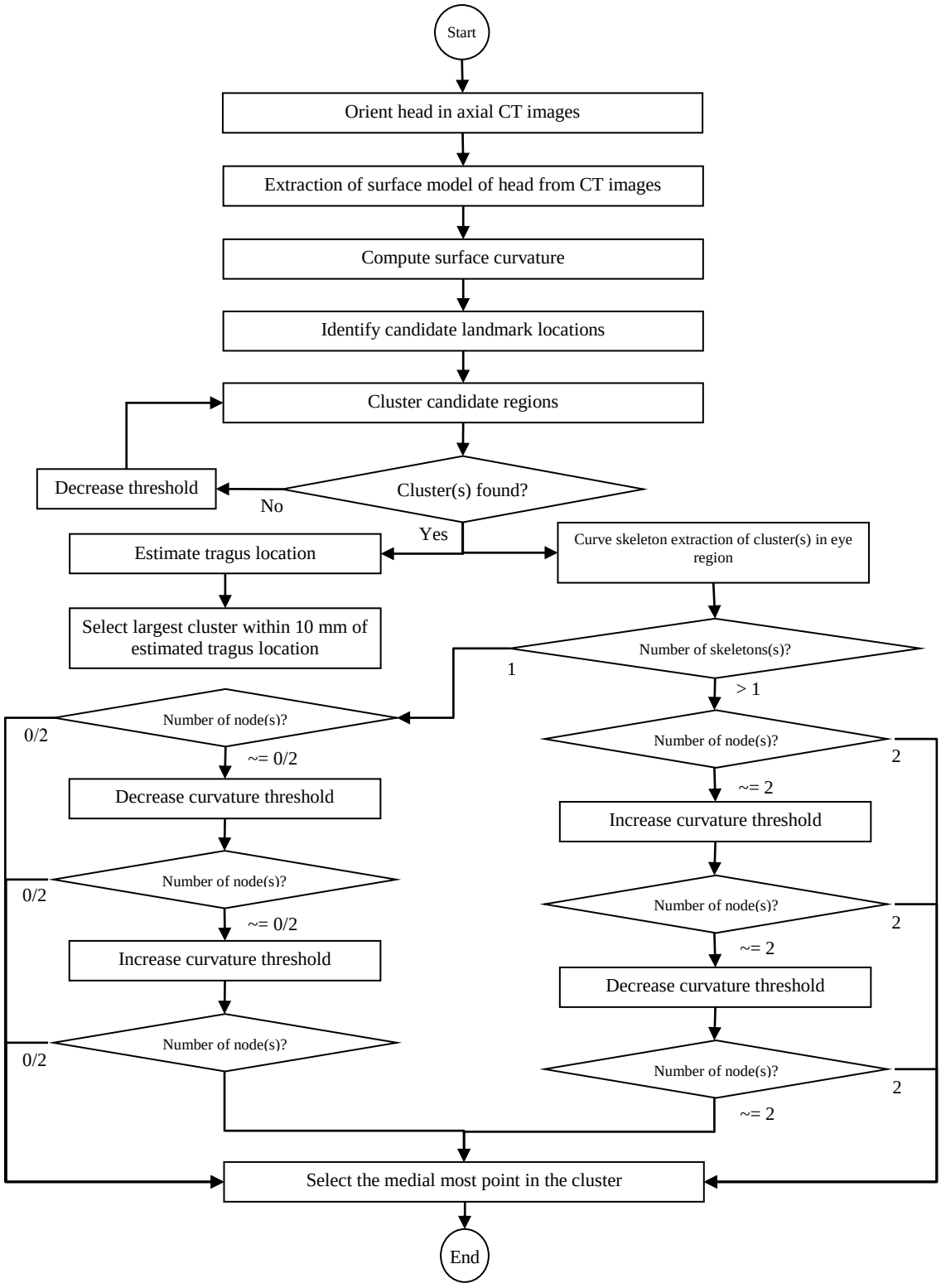


Figure 51: Overview of the algorithm

To validate the approach, the positions of the landmarks localised automatically and manually in near isotropic CT images of 118 patients were compared. These images were acquired using a multislice helical CT scanner (Aquilion 64 Toshiba-America Medical Systems Inc., Tustin, CA) in a private hospital, Johor Specialist Hospital, Malaysia. The images are 512 x 512 pixels with average in-slice pixel spacing and average slice thickness of 0.5 mm each, with no interslice spacing. The start and end scan location was from the skull vertex to at least the skull base, with the gantry oriented parallel to the infraorbitomeatal line. The DICOM compliant CT images were transferred to a personal computer, and the landmarks were localised automatically.

4.4 Results

CT images of 16 patients (out of 118 patients) were discarded from the analysis for the following reasons:

1. The medial canthus or tragus was not present in CT images of nine patients (due to a scan end location above the skull base).
2. One patient's eyes were obscured by a respiration mask
3. Medial canthus was not well defined in CT images of six patients.

Of the 102 patients that were included in the analysis, 52 patients were scanned with their eyes opened or partially opened, with the remaining 50 patients scanned with their eyes closed. There was no significant difference between the localisation errors for the two groups of patients. Figure 52 illustrates the automatically localised medial canthus and the most lateral point of the tragus for eight subjects in the database.

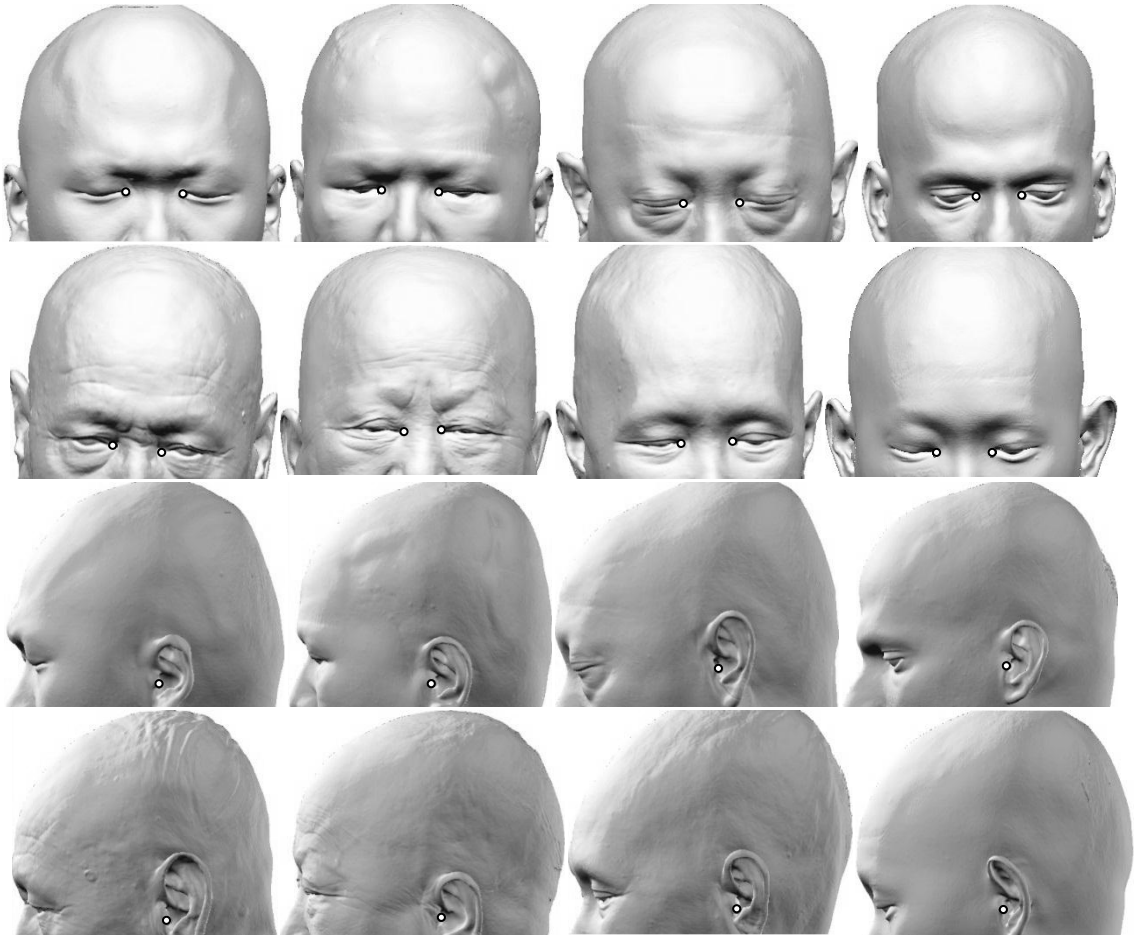


Figure 52: Automatically localised medial canthus and most lateral point of the tragus

Because of the retrospective nature of this validation, there is no anatomical ground truth. As such, the difference in localising the landmarks, defined as the distance between the perceived location of the landmark (as determined by visual inspection on the surface rendered model) and the location of the landmark found automatically based on the algorithm was used for comparison. Because manual localisation is subjective and to minimise bias, the landmarks were annotated twice by the author and their position were averaged. Intra examiner variability was 1.0 mm and 0.7 mm for the medial canthus and tragus, with a maximum variability of 2.9 mm and 2.5 mm respectively.

The mean difference in the localisation of the medial canthus, expressed as the root mean square Euclidean distance, was 1.2 mm (standard deviation = 0.9 mm), with a maximum difference of 4.5 mm. The mean difference for

localising the tragus was 0.8 mm with a maximum difference of 2.6 mm (standard deviation = 0.6 mm). For the medial canthus, automatic and manual localisation agreed to within 1 mm for 51% of the landmarks, within 2 mm for 80% of the landmarks, within 3 mm for 89% of the landmarks, and within 4 mm for 94% of the landmarks. For the tragus, automatic and manual localisation agreed within 1 mm for 75% of the landmarks, and within 2 mm for 97% of the landmarks. Estimates of the tragus location were found to be always within 10 mm of its actual location, with an average difference of 8-mm.

4.5 Discussion and conclusions

The results for the localisation of the medial canthus were compared to the distance between the two medial canthi i.e. intercanthal width with that reported by Deo and Sen [128]. Although not exactly a like for like comparison, the results are nevertheless compared in the absence of a more valid comparison. Deo and Sen obtained an error of 0.85% for the difference between the intercanthal widths determined automatically and manually for one subject. Manual measurements were ground truth measurements, based on measuring the medial canthus on the subject using callipers. In contrast, the average difference was 3.8% for 102 subjects using the proposed algorithm.

Good inter landmark measurement accuracy however, does not necessarily correspond to accurate landmark localisation. This is because landmarks can be erroneously localised but the distance between the landmarks could still be accurate e.g., the localisation of both the medial canthus could be offset by equal amounts and still yield accurate intercanthal width. Because these landmarks will be used as a basis to register images from two different modalities, the localised landmarks have to be spatially consistent between the two modalities. As such, localisation accuracy rather than measurement accuracy is more important.

The maximum difference in positions of landmarks localised manually and automatically of 4.5 mm for the medial canthus and 2.6 mm for the tragus is acceptable, considering the maximum variability in localising these landmarks manually can be as high as 2.9 mm and 2.5 mm respectively. The error in

localising a target, or the TRE, for a 5 mm localisation error, has previously been determined using an analytical expression by Fitzpatrick et al [111] to be within 5 mm for targets located in the brain. As the intended use of the localised landmarks is as a registration basis for a CT-to-image patient registration system for three neurosurgical procedures normally performed using a freehand technique i.e. without image-guidance, this TRE is within the clinical accuracy required for the procedures.

Shamir et al [120] found that FLEs are dependent on the type of anatomical landmarks and attributed the differences in the FLE values to the saliency of anatomical landmarks. The tragus, for example, is a cone-shaped structure, so its apex recognition was relatively accurate, with a low FLE. In contrast, the medial canthus is harder to localise and has a comparatively higher FLE. A possible solution is to weigh the landmarks differently according to their localisation uncertainty. The medial canthus could be assigned lesser weights compared to the tragus to reduce the overall registration error due to the larger medial canthus localisation error.

A methodology to localise the medial canthus and tragus in CT images automatically was presented. Manual localisation is subjective, as it is a function of the examiner's experience and perception. On the other hand, automatic localisation can be performed in a consistent manner, reducing the subjectivity inherent when localising the landmarks manually. The basic approach was to exploit the curvature saliency of the medial canthus and tragus, as they are the most geometrically salient among commonly used anatomical landmarks of the head, and can therefore be more reliably detected. A rule system based on prior knowledge of the landmark geometric structure and spatial location was used to constrain the landmark search, as curvature properties alone result in many false detections. The medial canthus and tragus can be automatically localised in CT images, with a performance comparable to manual localisation, based on the approach presented.

Once the preoperative plan has been registered to the patient using the proposed registration framework, the next phase is the intraoperative surgical intervention. Performing a surgical procedure autonomously using a robot is a safety critical process, as patient safety is paramount. However, there is always a real risk of the intervention going wrong. For example, when creating a burr

hole autonomously, it is entirely possible for a robot to begin drilling even though the incision has not been properly retracted, a highly unlikely event if it were a surgeon performing the operation. In the following chapter, the concept of a supervisory system, to monitor the surgical intervention intraoperatively is discussed, with applications to the MISEN system.

.

Chapter 5

An Intraoperative Supervisory System

5.1 Introduction

The relatively low usage of robotics in surgery can in part be attributed to, whether real or perceived, the lack of safety. Robotic surgery systems, particularly those involving neurosurgery, should therefore be designed to be inherently safe. While safety features (such as redundant sensors, compliance, hardware limited insertion speed by reduced final gear reduction ratio, etc.) are usually incorporated into these systems, they are unable to equal the level of situational awareness of a surgeon overseeing a surgical procedure. Unless this shortcoming is adequately addressed, it is difficult to envisage the widespread use of robotic surgery in place of conventional surgery.

As a step towards emulating the supervisory role performed by a surgeon, an intraoperative supervisory system that is capable of monitoring a surgical procedure performed by a robotic system is proposed. The aim is to reduce the risks associated with robotic surgery and the potential for mistakes by the robotic system. This system would provide reassurance to the end user, typically a surgical team without specialised neurosurgical experience, during the intervention. An important feature of such a system would be the ability to assess the outcome of a surgical action, to validate its successful completion.

A surgical action is a single task that is performed during a surgical procedure, which is usually made up of several surgical actions performed sequentially. The ability to assess the outcome of a surgical action is an important one, as subsequent surgical actions cannot usually be performed prior to the successful completion of the previous surgical action. This capability is therefore essential from both a safety and operational point of view.

Surgeons using the da Vinci telesurgical robot, a robot synonymous with surgical robotics, routinely perform surgical actions primarily based on visual feedback (a force feedback feature, as a substitute for tactile sensation, was a later addition). While these surgeons assess task completion by relying on visual cues from the operative field, there is no reported use of vision to validate the successful completion of a surgical action in robot-assisted surgery systems. The Minerva system for example, relies only on monitoring the current profile of the motor driving the drill bit when drilling a burr hole [58], without any visual observation of the site of the burr hole.

In principle, breakthrough i.e. successful penetration of the cranium can only be assured when it is possible to observe the dura mater (the outermost part of the brain). When the dura mater is detected in image space, breakthrough has occurred, regardless of the current profile. The outcome of the drilling could therefore be assessed using a computer vision-based approach. Apart from slight variations, the outcome of surgical actions, when created in a consistent fashion (as would be the case in robotic surgery systems), would exhibit certain distinguishing visual characteristics. This would make visual assessment of these outcomes, given that the surgical context is known, feasible.

5.2 Computer vision

Computer vision is the science and technology of machines that see, replicating a human's ability to deduce useful information from image. It is often used to imitate the human sense of sight in robotic systems. Not using computer vision in robot assisted surgery systems is somewhat analogous to a surgeon performing a surgical procedure with their eyes closed. An obvious solution to mimic and embody a surgeon's visual sensory capabilities when monitoring a surgical procedure is to use computer vision. Specifically, the use of computer vision to assess the completion of a surgical action, for contextual support when performing a surgical procedure using robotic surgery systems, is proposed. Information from the scene relevant to a task can be analysed and the outcome of the tasks classified, based on their surgical context.

While numerous sensors may be used to assess the outcome of surgical actions, the use of a visual sensor is considered here. Sensors used to perform a specific task should not be used to validate the completion of the same task, to avoid a situation where errors, if any, become compounded. Sensors independent from those used to perform the task being assessed, such as visual sensors, should therefore be used in the validation. For example, the start and end of drilling in the Minerva system is based on monitoring the current consumption of the motor driving the drill bit. It would be necessary to use information from a different sensor to determine breakthrough. A visual

sensor may be used to inspect a burr hole to determine whether breakthrough was achieved, to validate task completion. Different sensory capabilities therefore complement each other, allowing for a more complete assessment of a task. Incorporating visual sensory capability improves the reliability and robustness of robotic surgery systems, which traditionally only employs force, torque and positional measurements, as a visual sensor can often provide information that other sensors do not have.

5.3 Vision-based robotics

There are two different ways of using visual information in a robotic system, for control or for inspection. In control, visual feedback is used, in an open loop or closed-loop manner, to perform a task. In open loop control using visual feedback, an action is performed by a “looking’’ then “acting” approach. To illustrate the use of vision in open-loop control, consider the task of positioning a cauteriser relative to a bleeding source to affect cauterisation. Here, the required robot’s joint–space configurations to position the cauteriser towards the bleeding source are determined based on the position of the bleeding source in an image. It is assumed that the calibration between the task space and the camera, calibration between the camera and the robot, and the robot’s inverse kinematic model i.e. kinematic relationship between the robot and the task space is known. In closed loop visual control, a similar approach to position the cauteriser at the bleeding source is used, the difference being that visual feedback is constantly acquired from the workspace to position the cauteriser.

In inspection, the goal is to examine or evaluate an object or process in terms of certain characteristics or metrics based on a set of rules and to decide on whether the inspection criteria are met. There are three primary stages in visual inspection: Image acquisition, image interpretation and decision-making. In the first phase, image acquisition, a graphical representation of a physical scene is acquired. This image is subsequently analysed to extract and interpret information contained in the scene. Finally, the next course of action is decided, based on rules applied to information extracted from the image interpretation phase. These three stages are exemplified in the automatic inspection of an

incision to determine whether it is within a prescribed length. First, an image of the incision is acquired using a camera. The incision is then segmented from the image and its length determined. Its actual length can be determined If the scale of the image is known. If the required length is known, it can be used to determine if the incision length is within the allowable range.

Visual inspection tasks range in complexity from quantitative to qualitative assessment. Computer vision-based inspection systems for factory automation began in earnest in the mid-1960s with the advent of real-time image processing [138]. The first application of computer vision for automated inspection was the detection of defects in printed circuit boards (PCBs), a remarkable achievement considering that this was achieved using simple logical operations on binary images. The use of greyscale and colour images is more common today, as technological advancements have made great improvements in the performance of both hardware and software possible, while driving cost down. This has led to an exponential growth in the use of computer vision for automated inspection in the manufacturing, food, pharmaceutical and agricultural sectors. Today machine vision is used for automated inspection in a wide variety of applications from the characterisation of microstructural deformations and strains in cortical bone [139] to knowledge intensive application such as the quality inspection of seedlings [140], where expert knowledge is explicitly used.

The use of computer vision for inspection is the subject of much research, and the advances and success stories in this area have motivated this work. In particular, the use of computer vision for the diagnosis of malignant melanoma is especially relevant, where the subject of the inspection is biological specimens. Encouragingly, diagnostic accuracy that rivals those achieved by experts has been reported [156,157], an impressive feat considering that the clinical signs of malignancy are often ambiguous [158], and the use of low-level features only such as asymmetry, colour, texture and size in the inspection. Other uses of computer vision related to biological specimens include the automatic monitoring and analysis of human embryonic stem cells growth, by classifying their morphologic changes based on features extracted from their optical images [159]. Automating the inspection proved a reliable

means of obtaining more data compared to conventional methods, as well as allowing the continuous monitoring of these cells without disturbing them.

5.4 Computer vision in the operation theatre

While computer vision is routinely employed in industrial settings for automated inspection tasks, its use in other less conventional settings such as in the operating theatre has been less widespread. Several authors [141, 142] have investigated the use of vision in the operating theatre. Lo et. al. [141] developed a framework for the classification of surgical episodes using multiple visual cues related to shape, deformation, changes in light reflection and other low level image features. Specifically, their work was on the detection of surgical actions, such as the interaction of surgical instruments with soft tissue, retraction, cauterisation, and suturing, in video sequences of minimally invasive laparoscopic surgeries. An average classification accuracy of 85.3% for five video sequences (1762 video frames) was achieved using a naïve Bayesian classifier, with individual episode accuracy ranging from 60% to 100%. Both retraction and suturing were accurately classified every time. The frame-by-frame classification accuracy was lower, averaging 77%, as temporal information was not used, whilst temporal information was used for episode classification. Misclassification of the interaction of surgical instruments with soft tissue and cauterisation was attributed to the lack of depth perception, owing to the use of a monocular vision system, and the inability to differentiate between these two episodes using a single feature (specular highlight) alone.

Padoy et. al. [142] developed an approach to recognise a subset of surgical actions performed by the surgeon during laparoscopic surgery using visual cues and signals recorded from surgical instruments to monitor the progress of an operation. The objective was to use the information gathered to design a support system for an operating theatre e.g. activation of the operating theatre lights, automatic reporting, etc. A detection accuracy of 93% for eleven surgeries was achieved, although this was achieved by relying on both visual cues and recorded signals.

The above examples on the use of computer vision in the operating theatre are primarily concerned with the detection of surgical actions. A natural progression in the use of computer vision in the operating theatre is for the inspection of these surgical actions. The visually rich operative field presents a good environment for the development of automated visual inspection techniques to assess the outcome of surgical actions. To the author's knowledge, the use of computer vision in this context has not been reported.

5.5 Applications to MISEN

The proposed use of MISEN to perform the targeted neurosurgical procedures in their entirety, all of which require the creation of a burr hole, is a novel application of robotics. While creating a burr hole is trivial for a neurosurgeon, it is a very complex action for a robot to perform. To create a burr hole, a series of surgical actions must be performed. There are three primary surgical actions that are performed in a sequential manner to create a burr hole, namely skin incision, retraction of an incision, and drilling of a hole through a skull. Each of these surgical actions can only be performed upon successful completion of the previous action. For example, the skin has to be properly incised prior to retraction, properly retracted prior to drilling, and properly drilled prior to accessing the brain through a burr hole. As such, the ability to validate the outcome of each of these surgical actions is essential.

The MISEN system is already equipped with close range photogrammetry cameras for registration purposes (see Figure 26). These same cameras can be used for support during surgery, in a supervisory role after the registration is performed. Although MISEN is conceived as a supervisory controlled system i.e. the robot performs surgery autonomously under the supervision of the surgeon, an intraoperative supervisory system to augment a surgeon's supervisory role is needed, as this information is essential to make a decision on whether to proceed with the next surgical action. For example, drilling may only commence if the incision is deemed to have been sufficiently retracted. The three primary actions in creating a burr hole: skin

incision, retraction of the incision, and drilling of a burr hole are outlined next, and areas where vision may play a role are discussed.

5.5.1 Skin incision

When creating a burr hole, the scalp has to be incised at the site of the hole and the incision retracted to provide access to the skull for drilling. Skin incision involves incising the scalp with a scalpel blade along the length of the incision. The variables for an incision are its length, depth, position and orientation. An incision of insufficient length or depth would make it difficult or impossible to place the retractors, or to retract the incision. An incision that is too long is unnecessary and would be more susceptible to bleeding. A wrongly placed incision or incorrectly oriented incision would also complicate the insertion of the retractors.

An incision of a specified length can be made autonomously using a scalpel attached to the end effector of a robotic arm with encoders built into it to provide accurate positional feedback. Visual inspection of the incision can be a useful check to validate incision length and is a good complement for positional feedback, especially if the encoders have failed and gross positioning errors have occurred. To determine the length of an incision from its image, the incision would have to first be segmented from the image. Assuming calibrated cameras are used, the incision length can be determined. An alternative would be to infer the incision length from the scalpel travel, although this would be an indirect means of determining incision length.

Determining the depth of an incision is not as straightforward. Because the depth of the scalp can vary considerably, between subjects of different age, sex and ethnicity, it is virtually impossible to specify a correct depth of cut. However, it may be possible to inspect an incision for sufficient depth of cut if it displays a distinctive visual appearance. Finally, the position and orientation of the incision may be determined by standard measures of region properties such as centroid and second moments (for orientation), assuming that the incision can be reliably and consistently segmented from the image.

5.5.2 Retraction of an incision

One of the goals of the intraoperative supervisory system is to detect whether retraction of the incision is complete, as the incision would have to be sufficiently parted before drilling can commence. An obvious solution would be to infer the amount the incision has retracted based on the actual displacement of the retractors. Although the state of retraction can be determined quite trivially based on positional feedback from encoders that relay information on the extension of the retractors, a potential problem with this approach is that the retractors may have extended without the incision being retracted. This would be case if the retractors were incorrectly inserted into the incision prior to retraction. While this possibility may be ruled out by the use of force sensors to detect whether there is a resistive force that would indicate the incision being retracted, an erroneous force reading may still be possible. For example, the presence of other structures that impede the retraction may result in a resistive force that is wrongly attributed to retraction of the incision, or it might simply be the case of faulty force sensors. Because the use of positional and force sensors is an indirect form of measurement, the use of these sensors alone to assess the retraction of an incision might be inadequate.

An alternative and a direct approach to the problem of determining whether the incision has been retracted would be a visual observation of the incision for evidence of retraction. A visual sensor can therefore be used to complement information from these other sensors to assess the level of retraction of an incision. Calibration between the task space and camera frame will enable the amount of retraction to be measured, assuming the boundaries of the incision are known, although this requires sufficient contrast in the image to segment the boundaries of the retracted incision reliably. The presence of the retractors may also obscure a complete view of the retracted incision.

If the extent to which the incision has retracted proves to be difficult to determine visually, another possibility is to identify gross skin deformation, which is a consequence of the incision being retracted. The vision system may be used to identify the occurrence of such skin deformation, which would be indicative of the incision being retracted. If the amount of skin deformation exceeds a certain threshold, there is a strong likelihood that the incision is

being retracted. A problem with this assumption is that all skin deformations will be interpreted as a consequence of the incision being retracted. Skin deformation may also be the result of retractors pushing into the tissue. As such, it would be necessary to differentiate skin deformation due to retraction and due to other causes. Although less than ideal, one possibility is to relate the extension of the retractors with skin deformation, by relating the extension of the retractors and looking for corresponding skin deformation. The retraction amount can then be subsequently validated with the extension of the retractors.

5.5.3 Burr hole

The visual inspection of a burr hole consists of observation of the dura mater and is the most straightforward of the three surgical tasks under consideration. The inspection of the outcome of the drilling of a burr hole is particularly suitable to be performed visually, as a complete outcome is one where the underlying dura mater i.e. the outermost part of the brain, is visible. There are two stages in the visual inspection of the burr hole: Detecting a burr hole in the image and assessing the burr hole to determine if breakthrough i.e. the skull has been successfully penetrated has occurred. Once isolated, features that can distinguish the two conditions of the burr hole i.e. incomplete and complete can be identified and used as a basis for classification.

5.6 Classification

The automated inspection of the outcome of surgical actions can be formulated as classification problem. In classification, the objective is to group the outcomes of a surgical action into two classes, unsatisfactory/incomplete or satisfactory/complete. As there is no formal description of a complete surgical action, classification is based on features extracted from ground truth images of a surgical action. Features that are well suited as a basis for classification are features that can be reliably detected in image space and are distinct. These features should be invariant features, differentiate the two classes well while being robust to irrelevant variations. To determine features that can differentiate

the two classes, multiple observations of the two possible outcomes are grouped and the features that yielded the best classification rates are selected. In order to obtain a feature space, low level features such as colour, texture and shape descriptors may be used if the two possible outcomes of the surgical tasks under consideration differ in these respects

5.7 Challenges

The use of computer vision in the operating theatre may be fraught with difficulties, as the surgical environment can be highly unstructured, in contrast to the more controlled conditions typically found in an industrial setting when dealing with man-made objects. The inspection of these objects is usually based on few criteria and small deviations thereof. On the other hand, biological subjects have an inherent natural variation and are usually characterised by a combination of not so obvious features. This makes automated inspection more complex than the automated inspection of man-made objects. The correct classification of the outcome of a surgical task may have to rely on these not so obvious features. Nevertheless, certain surgical actions, when performed consistently, usually have similar outcomes. Furthermore, because the surgical actions being inspected are intended to be performed robotically, their context is known. This would make a computer vision based inspection system of a surgical action feasible.

A fundamental limitation of a vision-based approach is that its use is limited in situations where there is no direct visualisation of the surgical action being assessed. For example, in the context of neurosurgery, it is only possible to ‘view’, and thus assess, the outcomes of extracranial procedures such as skin incision, bleeding, the retraction of an incision, and the drilling of a burr hole. Intracranial procedures such as the insertion of the catheter into the brain ventricles cannot ordinarily be viewed and therefore cannot be assessed. Furthermore, even in extracranial procedures, obstruction in the camera’s field of view may complicate inspection. A practical complication of this is the use of vision to detect whether an incision has been sufficiently retracted. Ordinarily, if a complete view of the retracted incision is available, by performing edge based

segmentation and using a priori knowledge on the expected edge separation distance, successful retraction of the incision can be determined. However, the presence of retractors obscures the complete view of the retracted incision and may complicate its inspection.

Another limitation of using vision is its lack of robustness in an unstructured environment. In general, image-processing algorithms are known to be highly susceptible to changes in its environment such as changes in lighting. However, the algorithms used for image processing should be robust enough to be able to work reasonably well within a specified operating range. Image enhancement techniques such as contrast enhancement using adaptive histogram equalisation may be used to extend this operating range.

5.8 Discussions and summary

While on the one hand the ability of robotic surgery systems to perform surgery autonomously has tremendous potential to free up scarce surgical resources and enable surgery that would otherwise not have taken place be performed, there are safety considerations that have to addressed, as performing surgical actions autonomously is not without risk. As a step towards improving the safety aspect of these robotic assisted surgery systems, the automated inspection of the outcome of surgical actions, to provide reassurance during a surgical procedure, was proposed. Automated inspection will afford these robots the added capability to detect anomalies. It will also allow for more robust and autonomous robotic systems. The use of computer vision techniques presents a good framework for the development of this system.

The intraoperative supervisory system will use video images as part of its sensory input for a more comprehensive assessment of the surgical action. A possible way of combining multiple sensory inputs where present is to use the Boolean logical AND operator. The outcome of a surgical action is considered satisfactory if all the sensory inputs that are capable of assessing task completion have indicated that the task is complete. Its output will be used

as a basis for deciding on the next course of action i.e. whether to execute the subsequent surgical action.

A natural progression in the use of computer vision for the inspection of the outcome of surgical actions in the operating theatre is in the area of surgical training. Surgical training is usually based on an apprenticeship model, with the skills of trainees traditionally assessed by surgeons themselves. There are two main problems with this model for assessing surgical skills. The first is that the assessment can be highly subjective [143]. The second is that the amount of resources involved can be considerable, especially the time a surgeon has to spend assessing a trainee's skill, which could otherwise have been spent on performing surgical procedures.

The alternative is the use of medical simulators based on virtual reality-simulated environments [144]. While these simulators have been used by the medical community for surgical training and skills assessment, they are not realistic enough as the sole source for the training and assessment of surgical skills, in particular advanced skills [141]. Performing surgery on patients in the operating theatre or cadaveric dissection is therefore still necessary. As such, the only other alternative is to reduce the amount of time available for training, although the danger is that trainees may not have acquired all the requisite surgical skills, as they would be required to achieve the same level of proficiency in a lesser amount of time.

If the outcome of a surgical action is used as a reflection of surgical skill, this could be used as a basis for surgical skill assessment. An alternative to the assessment of a trainee's surgical skill by a surgeon therefore could be the use of computer vision to assess the outcome of a surgical action, and by extension the surgical skill of the trainee. A computer vision based assessment of surgical skills has the potential to afford trainees another avenue for training, in addition to being a more objective means of assessment. Although there is more to surgical competence than technical skills e.g. decision making, communication skills, and leadership skills; technical proficiency is considered the most important attribute among surgical trainees [145].

In this chapter, the concept of an intraoperative supervisory system for robotic surgery was discussed and a niche where this system may be used for MISEN was identified. A vision-based approach to assess the outcome of three

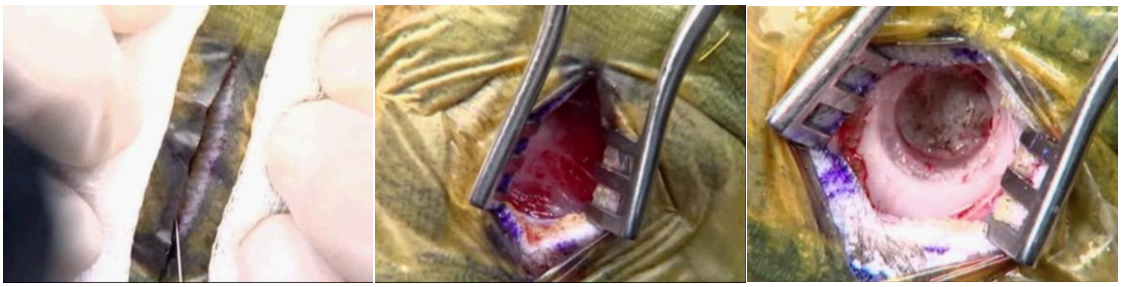
primary surgical actions involved in creating a burr hole (skin incision, retraction of an incision and drilling) was proposed and possible approaches to assess the outcome of surgical actions using computer vision techniques were presented. The visual inspection of these surgical actions was investigated, to prove the feasibility of the approach. Preliminary results are presented in Chapter 6.

Chapter 6

Computer vision for the inspection of surgical tasks

6.1 Introduction

The use of computer vision to inspect the outcome of surgical tasks for a common neurosurgical procedure was investigated, with the aim of incorporating this approach in MISEN. Practical examples of three primary surgical tasks for creating a burr hole, namely skin incision, retraction of the incision, and drilling of a burr hole (see Figure 53) were used to illustrate the concept.



a	b	c

Figure 53: (a) Skin incision, (b) retracted incision, and (c) burr hole

The feasibility of creating a burr hole robotically has previously been demonstrated by the Minerva system [61]. In the Minerva system, the force and torque characteristics of the motors driving purpose-built surgical tools were used as the control variables to perform the surgical actions [58]. For example, the start and end of drilling is determined by monitoring the current consumption of the motor driving the drill bit. An important distinction to be made in the use of these sensory capabilities is in terms of their purpose. While force and torque sensors were used to aid in the performance of a task i.e. to determine when to stop the drilling, the use of a visual sensor is for the inspection of its outcome i.e. to determine whether there was breakthrough.

The ability to inspect the outcome of the three primary surgical tasks in creating a burr hole is an important functionality, from an operational and safety point of view, as each of these tasks can only be performed upon completion of the previous task. For example, the skin has to be properly incised prior to retraction, the incision has to be properly retracted prior to drilling, and the cranium has to be properly drilled prior to accessing the brain through the burr hole. Performing a task before a prior task has been successfully completed

may not be possible and would be dangerous. For example, attempting to create a burr hole in the skull without incising and retracting the scalp will inflict irreparable damage to the scalp, as well as cause heavy bleeding.

6.2 Methodology

Experiments were performed using cadaveric pig heads to simulate the two possible outcomes i.e. complete/satisfactory and incomplete/unsatisfactory, of the three primary surgical tasks involved in making a burr hole. Practical and ethical considerations preclude experimentation on humans or live animals. As such, animal cadavers were used as a substitute. In terms of the selection of an animal cadaver, pig head cadavers were used for the experiments because of their anatomical similarity to humans. With the exception of primates, the anatomy of the temporoparietal region (site of a burr hole) of pigs most closely resembles that of a human [146]. As access to the pig's brain is easier through the frontoparietal region (due to its proximity to the brain and structure of the pigs skull), this region was chosen as the site of the burr hole, instead of the temporoparietal region.

A charge-coupled device (CCD) video camera (see Table 3 for technical specifications), with a white light emitting diode (LED) ring light (see Table 4 for technical specifications) mounted around the optical lens, was used to obtain video images of the surgical tasks (see Figure 54). Ring lighting reduces the effects of shadows by providing an even illumination in the field of view. The ring light used had an adjustable intensity and a four-section quadrant control, enabling each quadrant to be individually controlled. For a greater depth of focus, the camera aperture was kept to a minimum, by using the maximum intensity setting for the ring lighting unit. Both the camera focus and exposure settings were set to automatic.

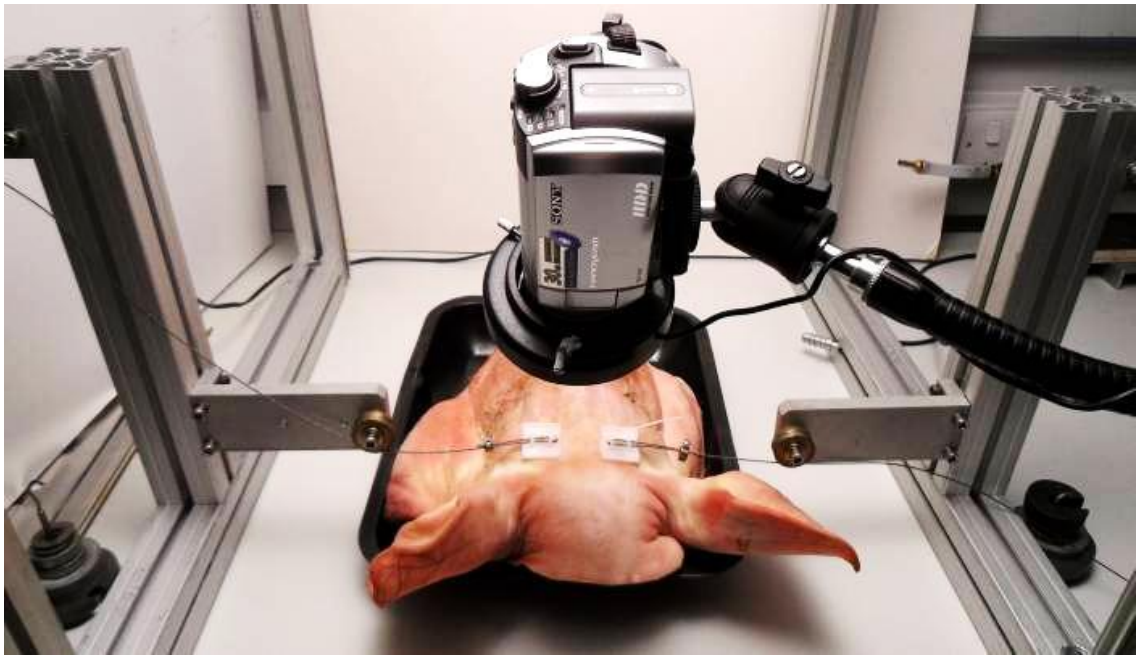


Figure 54: Experimental setup

Table 3: Video camera specifications

Optical Sensor Size	1/6”
Optical Sensor Type	CCD
Focal Length	2.3 mm-46 mm,
Lens Aperture	F/1.8 - 3.1
Optical Zoom	20x
Lens system type	Zoom lens
Focal Length Equivalent to 35mm Camera	44 - 440 mm
Focus Adjustment	Manual, Automatic
White Balance	Presets, Automatic
Digital Video Format	MPEG-2

Table 4: Ring light specifications

Illumination control	Each quadrant controlled separately
Working Distance	40-250mm
Number of LED	144
Brightness	20000 Lux at a height of 100mm
Light colour temperature	6400K
Outer diameter	98 mm
Inner diameter	61 mm

The surgical procedure for creating a burr hole in a pig, as described by Kaiser and Fruhauf [147], is similar in many respects to that for creating a burr hole in a human and was adopted with some modifications. A midline incision, approximately 3 cm long, centred over the site of the burr hole, was made with a #10 scalpel blade, at different depths up to the pericranium to simulate incomplete incisions, and a single incision through the layers of the scalp to simulate complete incisions (see Figure 63). Incisions are said to be incomplete if all the layers of the scalp including the pericranium have not been incised, and are classified as complete if all the layers of the scalp including the pericranium have been incised. The complete incisions were subsequently retracted at varying degrees using a self-retaining retractor to expose the cranium, until the incision was wide enough to accommodate a burr hole. A burr hole 10 mm in diameter was then drilled in the cranium using a cordless drill at high speed.

To facilitate the placement of retractors, the incision is usually pulled apart by applying tension parallel to the plane of the skin. In MISEN, a robotic tool would apply a similar constant force on either side of the incision. To simulate this parting of the skin, two square tabs (29 x 29 mm) were attached to the skin with cyanoacrylate adhesive [148] (see Figure 54) and pulled apart by two 600 gram weights using a pulley and weight system (see Figure 55). The distance between the tabs was 29 mm.

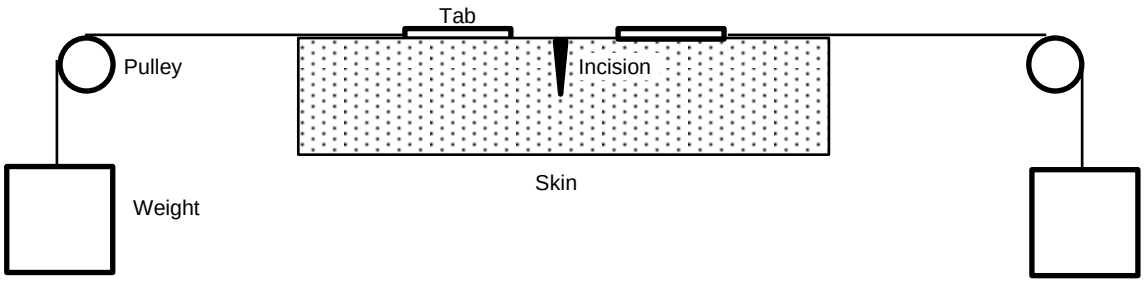


Figure 55: Schematic of the parting of the skin

Over fifty experiments (one pig head for each experiment) were performed to obtain a representative sample of the outcomes of surgical actions, owing to the variability associated with biological subjects. Images of these outcomes were analysed to identify appropriate visual cues and characteristic features to distinguish the two possible outcomes of each surgical action.

6.3 Characterisation of an incision

The first illustrative example is the inspection of the outcome of a skin incision. When surgeons incise the scalp, a scalpel blade is pushed into the scalp, through the layers of the scalp, until it has touched the cranium. A complete incision should be both sufficiently long and deep i.e. an incision up to the cranium along the length of the incision (see Figure 56). An incision that is either too short or not sufficiently deep would make it difficult or impossible to retract the incision or place the retractors into the incision. Unless the incision is completely retracted, the burr hole cannot be drilled, as the cranium will not be sufficiently exposed.

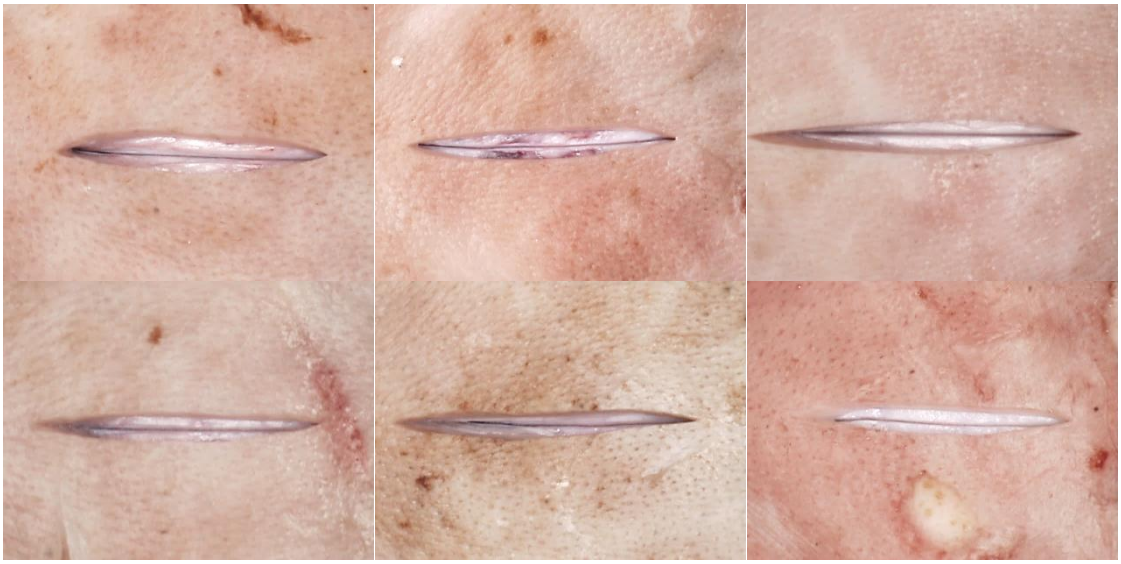


Figure 56: Illustrative examples of complete incisions

In this work, computer vision was used to inspect an incision for sufficient depth i.e. an incision that was incised up to the cranium. The incisions were created such that they were of the required length, the only variable being the depth of incision. It is assumed that the position, length and orientation of the incision are satisfactory, although these properties could be determined as well using computer vision techniques. For example, assuming that the incision can be segmented from the image and that the scale of the image is known, the properties of an image region corresponding to the incision such as position, length and orientation can be measured. The position of the incision is given by the centroid or the centre of mass of the image region. The length of the incision is the Euclidean distance between the two left and right extreme points of the image region and the orientation (relative to the camera axis) is the angle between the horizontal axis and the major axis of the ellipse that has the same second-moments as the image region.

Two types of incisions were created, complete and incomplete incisions. Complete incisions are incisions incised through all the layers of the skin up to the cranium. Incomplete incisions were divided into superficial (see Figure 57) and marginal incisions (see Figure 58). Incisions up to the dermis layer of the skin were considered superficial incisions, while incisions beyond the dermis but up to the pericranium were classified as marginal incisions. These incisions

were subsequently analysed to determine if a distinction could be made between complete and incomplete incisions based on their image properties.

Figure 57: Illustrative examples of superficial incisions

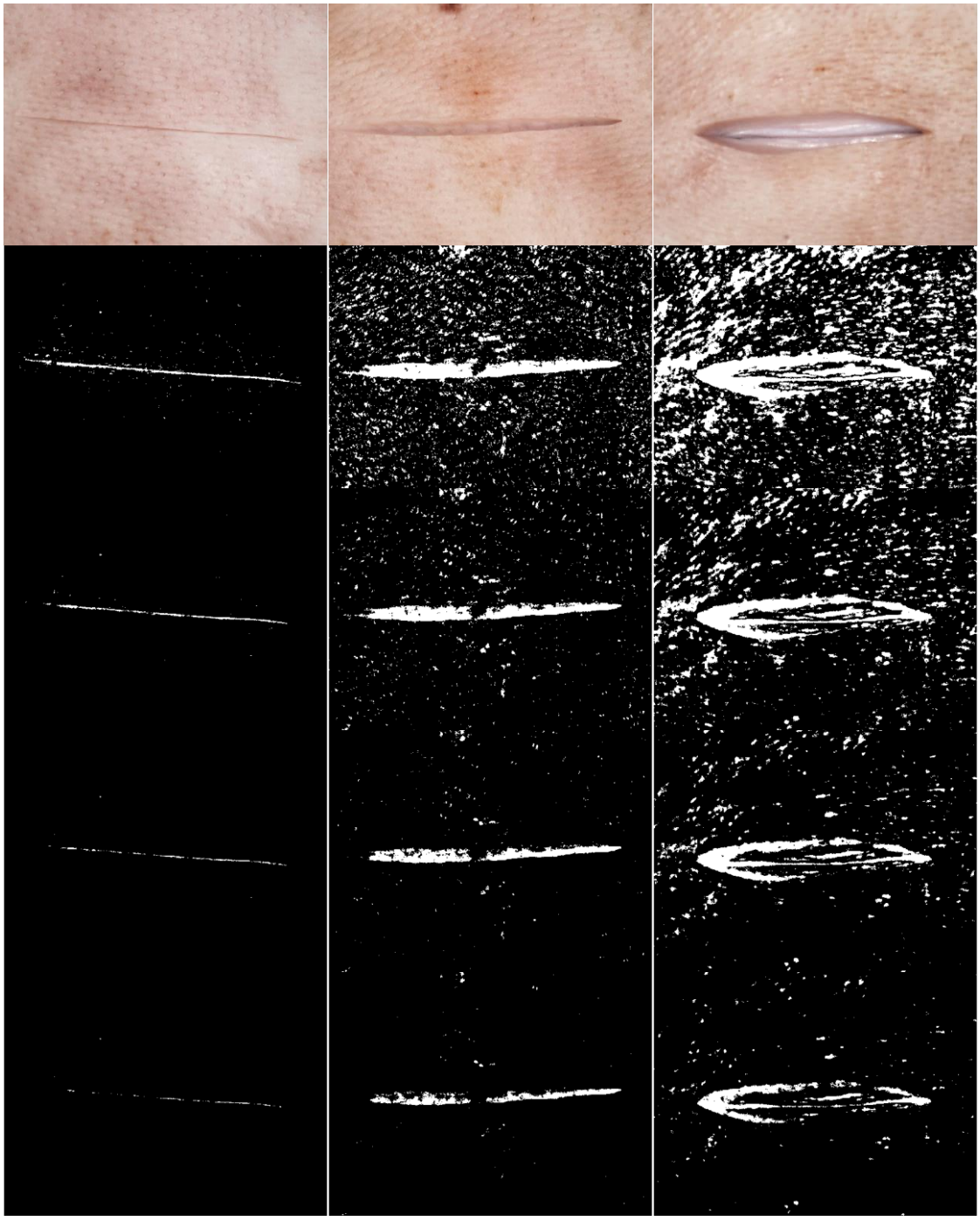
Figure 58: Illustrative examples of marginal incision

6.3.1 Image differencing

When parted (by applying tension parallel to the plane of the skin, see Figure 55), complete incisions tend to cause the skin surrounding them to deform more than incomplete incisions. Although a pig's skin is not as elastic as

the human skin, it is anatomically similar. In terms of its biomechanical properties, it exhibits similar behaviour to human skin under uniaxial tension in terms of ultimate stress–strain, as concluded in a study by Ankersen et al. [149].

To determine if the amount of skin deformation could be used as a basis to classify incision type i.e. incomplete and complete incision, the amount of skin deformation under constant uniaxial load by the robot would first have to be quantified. A simple measure of skin deformation, the absolute difference in intensity of greyscale images before and after an incision was made, was considered. To compensate for differences due to slight changes in intensity levels, each pixel is only considered changed if the difference in intensity value exceeds an empirically determined threshold. The mean number of pixels that changed by more than 25 greyscale intensity levels was used as the threshold (see Figure 59e). This threshold was chosen as it reduces the amount of noise i.e. pixels that do not correspond to the incision, while keeping most of the pixels corresponding to the incision. The mean number of changed pixels was used rather than the mean of the absolute difference in greyscale intensity value because it is a more representative measure of deformation than intensity difference, as large intensity difference does not necessarily correspond to large skin deformation.



a
b
c
d
e

Figure 59: Illustrative examples of (a) incisions and (b-e) their thresholded difference images corresponding to an absolute difference in greyscale intensity levels of 10, 15, 20 and 25.

6.3.1.1 Results

The mean number of changed pixels of the difference image for a total of 17 superficial incisions, 8 marginal incisions, and 18 complete incisions were evaluated. Figure 60 and Figure 61 show the boxplot of the mean number of pixels that changed by more than 25 greyscale intensity levels for superficial and non-superficial incisions i.e. marginal and complete incisions, and marginal incisions and complete incisions respectively. The central mark in each box is the median value, the edges of the box are the lower and upper quartiles, and the ends of the whiskers are the extreme values.

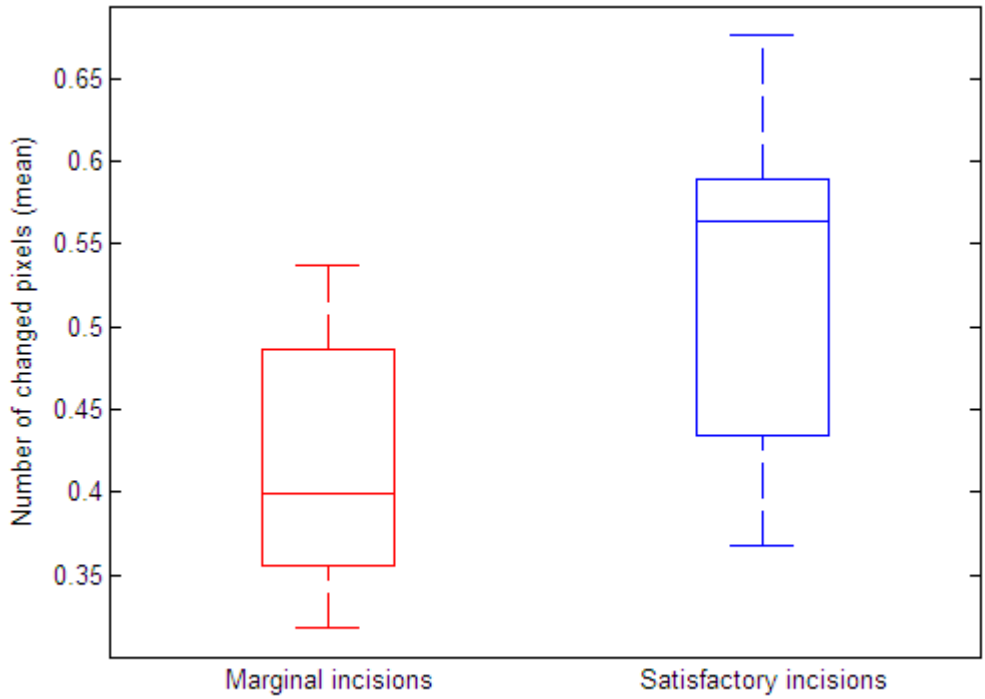


Figure 60: Boxplot illustrating the number of changed pixels (mean) versus incision type

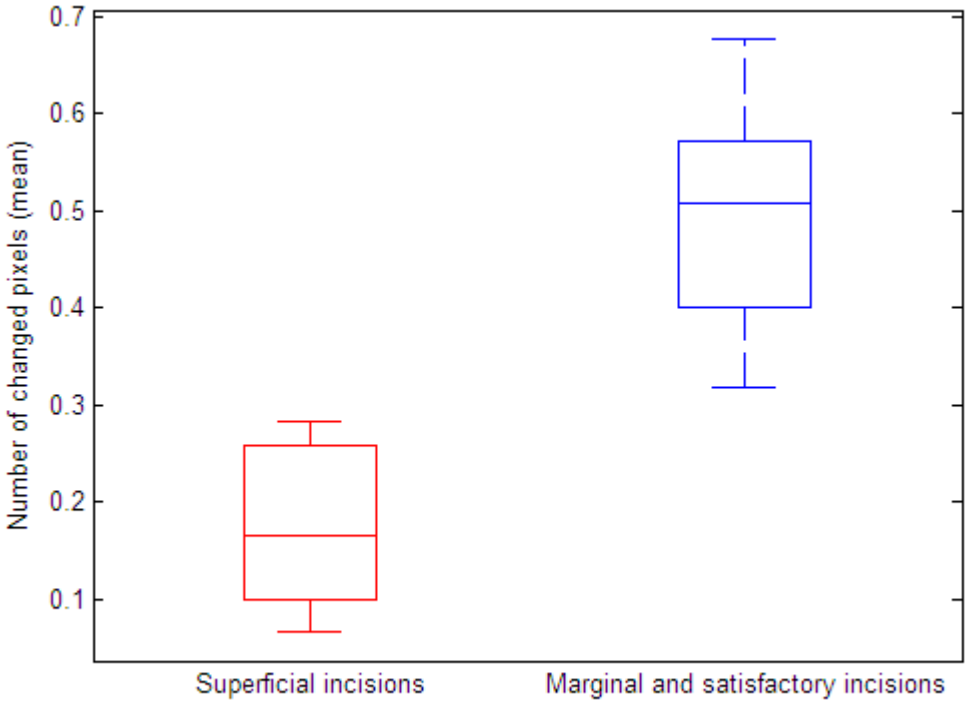


Figure 61: Boxplot illustrating the number of changed pixels (mean) versus incision type

6.3.1.2 Discussion

A naïve approach to segmenting an incision based on image differencing is to classify any pixels in the difference image (before and after an incision is made) that change in greyscale intensity levels as belonging to the incision. However, when we consider the fact that the skin deforms when incised, and the fact that the skin is not homogenous (in appearance), the problem is not as simple as detecting any changes greyscale intensity levels. To account for the non-homogeneity in the appearance of skin, and the deformation of the skin when incised, the difference image is thresholded based on a change of 25 greyscale values as discussed previously in Section 6.3.1 and illustrated in Figure 59.

As the mean number of changed pixels of superficial incisions and non-superficial incisions do not overlap (see Figure 61), image differencing can potentially be used to differentiate these two types of incision, using a nearest neighbour classifier based on the mean number of changed pixels. However,

although superficial and non-superficial incisions can be differentiated based on image differencing, it is not possible to differentiate marginal and complete incisions based on their difference image. There is a significant amount of overlap in mean number of changed pixels for these two types of incision (see Figure 60). This is because unlike superficial incisions, marginal incisions are closer in terms of size to a complete incision and therefore their difference image would be similar to that of a complete incision. Because image differencing is unable to differentiate marginal and complete incisions, the shape properties of these incisions are considered next, to determine if they could be used as a basis to differentiate these two types of incisions.

6.3.2 Shape properties

It was hypothesised that an incision may exhibit sufficiently different shape properties when skin is incised at different depths and subject to a constant lateral force by the robot, i.e. there is a correlation between the shape properties of an incision and incision depth. Specifically, in relation to the human anatomy, the extension of an incision and hence its shape properties may be a function of the layers of the scalp that were incised. The scalp is a heterogeneous tissue composed of five layers (skin, subcutaneous fat, galea, loose connective tissue and pericranium) with distinct biomechanical properties [150]. Figure 62 illustrates the different layers of a scalp.

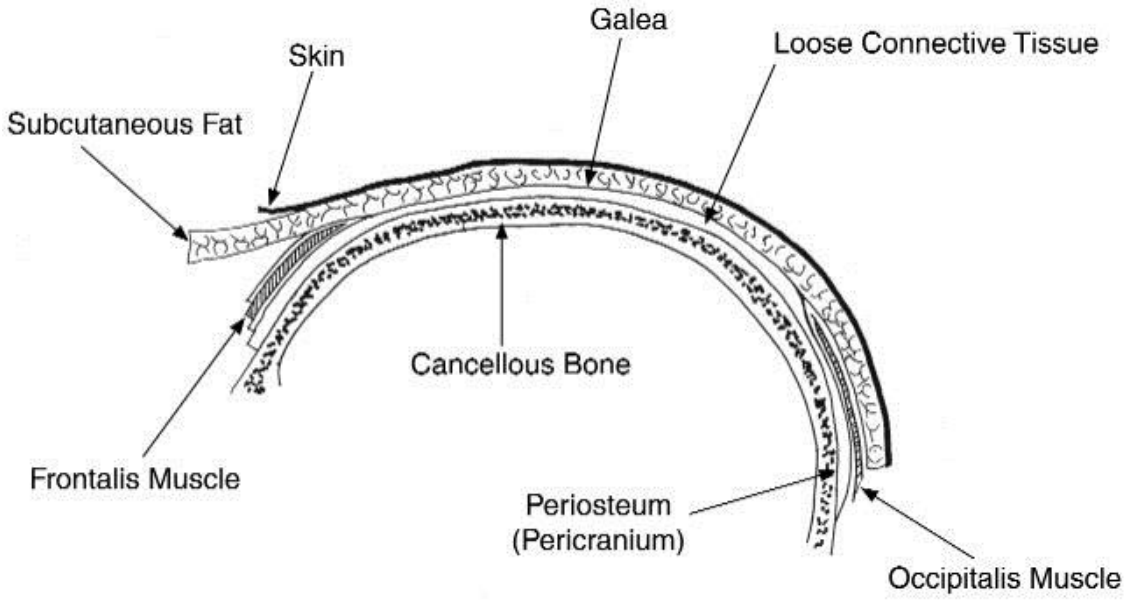


Figure 62: The anatomy of a scalp[21]

In particular, the galea, which is a tough layer of dense fibrous tissue (due to the presence of collagen fibres), significantly affects the tensile properties of the scalp [150]. Therefore, incising the galea should cause the incision to extend considerably (see Figure 63).

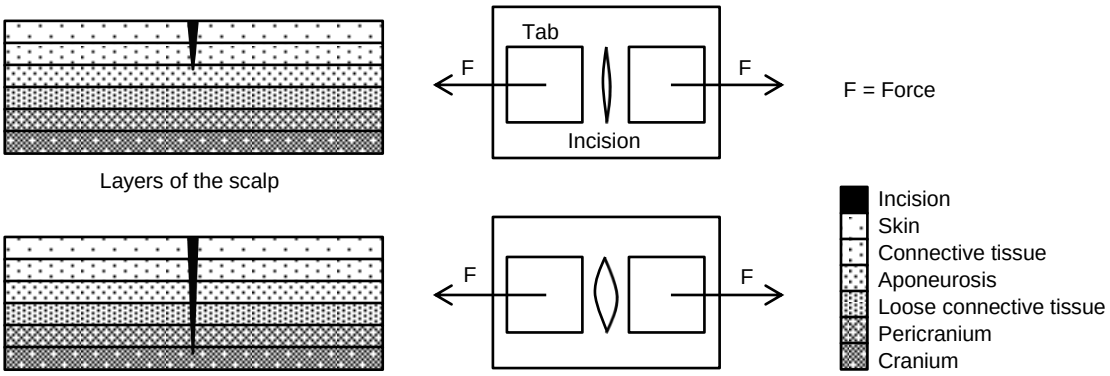


Figure 63: Schematic illustrating the expected behaviour of an incision under tension as the different layers of the scalp is incised

[21] Image reproduced after permission from Medscape

6.3.2.1 Segmentation of an incision

To determine the shape properties of an incision, it must first be segmented. Due to variations in the appearance of skin, and because the edges of the incisions may have low contrast in an image, segmentation of the incision is not trivial. Segmentation using a region-based and edge-based approach was attempted to segment the incision.

6.3.2.1.1 Image differencing

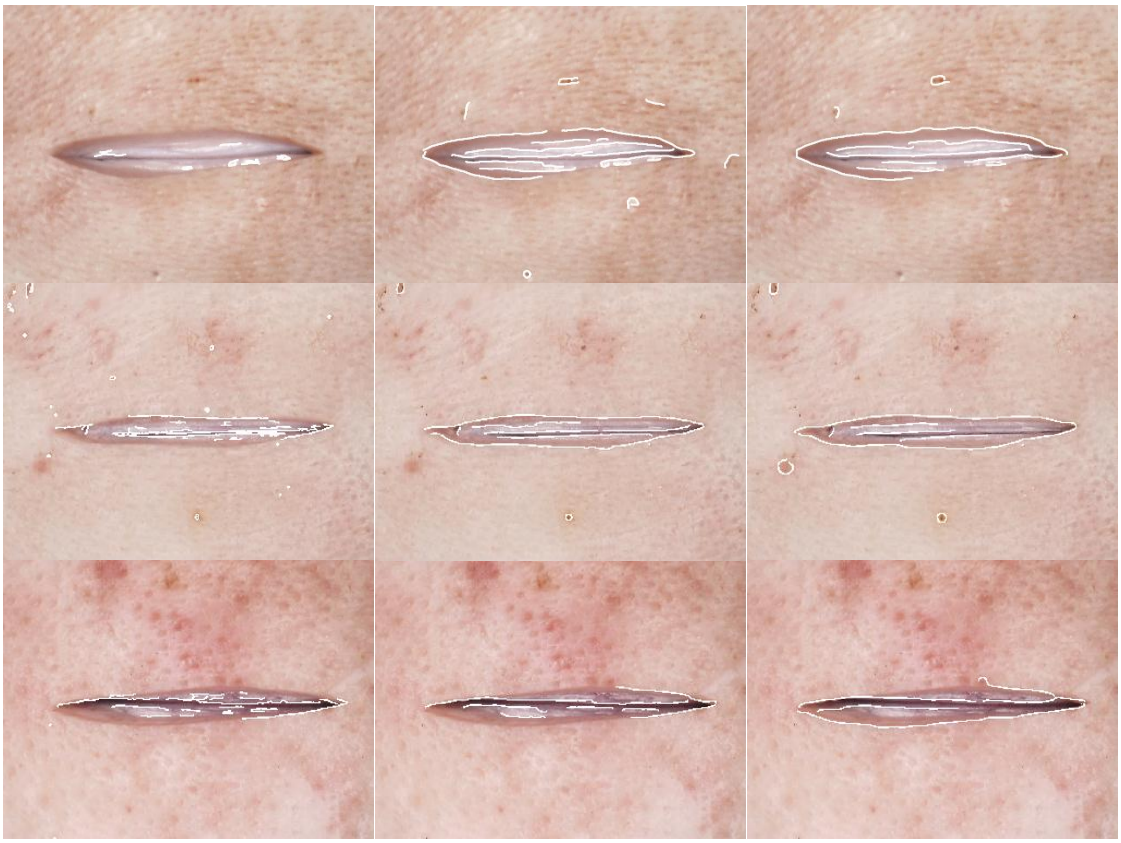
To segment the incision, image differencing, obtained by subtracting a greyscale image of an incision with a greyscale image prior to the incision being made, was used. The thresholded difference image, with the threshold level determined automatically using Otsu's method [131], is shown in Figure 59. Black areas in the thresholded difference image indicate small differences between the two images, and white areas indicate large differences.

The incision is ill defined in the thresholded difference image and image differencing cannot be used to segment an incision from the background skin. This is because of the high levels of noise in the difference image, a consequence of skin deformation (when incised). Additionally, the grey level intensity of the underlying skin layer after an incision is made is sometimes similar to the grey level intensity of the uppermost skin layer. As a result, the thresholded difference image may not register any noticeable differences for some portions of an incision (see Figure 59).

6.3.2.1.2 Edge detection

The well-known 'Canny' edge detector was subsequently used to try to segment the incision by delineating its boundaries. A 31 x 31 Gaussian filter with a standard deviation of five, and a low and high threshold of 0.2 and 0.4 respectively, was used. These values were determined empirically and were chosen as they worked best for the set of images analysed. More of the edges of the incision were detected (see Figure 64c) compared to those found based on a 4 x 4 (see Figure 64a) and 19 x 19 (see Figure 64b) Gaussian filter. A

morphological opening operation was used to remove noise i.e. all edges (connected components) that have fewer than 100 pixels, in the edge map. An 8-connected neighbourhood was used when establishing connectivity.



a	b	c

Figure 64: Incisions segmented with Canny edge detector using a Gaussian filter with a standard deviation, σ of (a) 1, (b) 3 and (c) 5. The filter size is 6σ +1.

As can be seen in Figure 64, the Canny edge detector was unable to segment some of the boundaries of the incision, primarily because of the low contrast at the edges of the incision. To facilitate edge detection, these images were contrast enhanced by linearly mapping the greyscale image intensity values to new values such that 1% of intensity values in the contrast enhanced image are saturated at low and high intensities. Figure 65 shows the segmented incision using Canny edge detection on contrast enhanced images. Although edge detection on contrast enhanced images is able to segment some previously undetected boundaries, it is unable to do so consistently for all boundaries. Additionally, not all edges of the incisions are properly detected.

Figure 65: Segmented incision using Canny edge detection on contrast enhanced images

6.3.2.1.3 Edge detection and directional lighting

To improve the detection of the edges of the incision, directional lighting was used. The top and bottom half of the ring light was alternately switched on and images of the incision were acquired. Figure 66 illustrates this technique.

a	b
c	d

Figure 66: Incision overlaid with edges, detected using Canny edge detection on image illuminated with (a) all quadrants of ring light, (b) top half of the ring light, (c) bottom half of the right light, and (d) incision overlaid with edges from (b) and (c)

Figure 66 a-c shows the incisions overlaid with edges detected using Canny edge detection when illuminated by all four quadrants, top half and bottom half or the right light respectively. Figure 66d shows the combined detected edges from Figure 66 b-c. The edges of the incision are clearly enhanced by controlling the direction of the light source.

6.3.2.1.4 Discussion

The use of directional lighting to facilitate the segmentation of an incision in an image is a promising technique that could potentially be used in the MISEN system. The technique works by casting shadows on the edges of the incision, resulting in a sufficiently high contrast between the edges of the incision and the background skin, enabling the detection of ill-defined edges with poor contrast. However, this technique is only applicable if the incision is sufficiently deep such that the use of directional lighting would be able to cast a shadow along its edges. As such, while potentially useful, it may not work on superficial incisions. Image differencing (discussed in Section 3.1) should therefore be used in combination with directional lighting, whereby directional lighting may be used to facilitate the segmentation of an incision in the image in the case of non-superficial incisions.

6.3.3 Shape measurement

To obtain the shape properties of an incision, it must first be segmented. Because the majority of images of incisions were obtained without using directional lighting, the incisions were manually segmented by fitting a polygon to their edges using Matlab's *impoly* function. The vertices of the polygon were specified at or near the edge of the incision by inspection such that it approximated as close as possible the incision's actual shape. The number of vertices used varied depending on the complexity/curvature of the shape, with a greater number of vertices used for highly complex shapes. Figure 67 illustrates the manual segmentation of an incision in an image. The blue dots are the vertices of the polygon

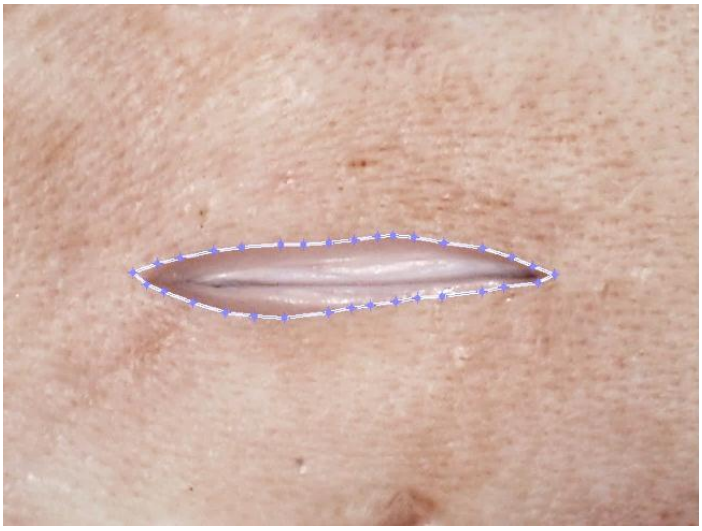


Figure 67: Manual segmentation of an incision.

Once segmented, the incisions are represented as an image region and their shape properties consisting of regional descriptors such as statistical moments, aspect ratio and eccentricity were computed. The aspect ratio of the incision is defined as the ratio of the major and minor axis of an ellipse that has the same normalised second central moments as the image region i.e. approximate regions by ellipses. The eccentricity of the incision is the eccentricity of the ellipse that has the same second-moments as the image region. It is the ratio of the distance between the foci of the ellipse and its major axis length, with a value between zero and one, with zero and one being degenerate cases (a circle has an eccentricity of zero while a line segment has an eccentricity of one). Skewness and kurtosis are the third and fourth shape moments respectively. Skewness is a measure of the asymmetry of the shape around the centre line, and is given by Equation 6.1

$$Horizontal\ skewness = \frac{(x-\mu)^3}{\sigma^3}$$
$$Vertical\ skewness = \frac{(y-\mu)^3}{\sigma^3} \qquad 6.1$$

where μ is the mean and σ is the standard deviation of the x and y pixel coordinates of the image region respectively. If skewness is negative, the shape is skewed more to the left of the centre than to the right. If skewness is

positive, the shape is skewed more to the right. The skewness of a perfectly symmetric shape is zero. Kurtosis is a measure of how irregular a shape is, and is given by Equation 6.2.

$$Horizontal\ kurtosis = \frac{(x-\mu)^4}{\sigma^4}$$
$$Vertical\ kurtosis = \frac{(y-\mu)^4}{\sigma^4} \quad 6.2$$

An image is rotated such that the orientation of the incision is approximately horizontal prior to computing its skewness and kurtosis. The orientation, θ of an incision is estimated as the 2D orientation of an ellipse with the same normalised second central moments, based on Equation 4.15. If the orientation of the incision deviates from a horizontal position, a rigid body rotation of the image region about its centroid is used to re-orient it, based on Equation 4.16.

6.3.3.1 Results

A linear classifier, linear discriminant analysis (LDA), was used to classify images of 54 incisions (27 complete incisions selected at random from 37 complete incisions and 27 marginal incisions) consisting of two classes i.e. complete and incomplete incisions, based on their shape properties. Because of the relatively small sample size, a leave one out cross validation technique, which uses a single observation for validation and the remaining observations as the training data, was used. This cross validation is repeated such that each sample is validated once. Table 5 shows the LDA leave one out misclassification error rate for classifying marginal and complete incisions based on various shape properties.

Table 5: LDA leave one out incision misclassification error rate based on various shape properties

LDA leave one out misclassification error rate (%)											
Shape property	Iteration										Average
Eccentricity	31.9	27.8	27.8	29.2	30.6	27.8	33.3	41.7	30.6	29.2	31.0
Aspect ratio	51.4	36.1	34.7	34.7	61.1	38.9	36.1	34.7	50.0	37.5	41.5
Vertical kurtosis	23.6	25.0	25.0	22.2	25.0	26.4	26.4	25.0	22.2	26.4	24.7
Horizontal kurtosis	18.1	20.8	19.4	18.1	18.1	18.1	19.4	22.2	16.7	23.6	19.5
Vertical skewness	25.0	23.6	23.6	25.0	23.6	23.6	23.6	25.0	22.2	25.0	24.0
Horizontal skewness	29.2	25.0	23.6	25.0	25.0	22.2	25.0	26.4	26.4	29.2	25.7

Figure 68 shows a boxplot of the various normalised shape property values (for an iteration of the classification) for the two types of incision with outliers shown as crosses. An outlier is defined as a value outside the range: $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ where Q_1 and Q_3 are the lower and upper quartiles respectively. For normally distributed data, the constant 1.5 corresponds to approximately 99% of the data. Incomplete and complete incisions are denoted in red and blue respectively.

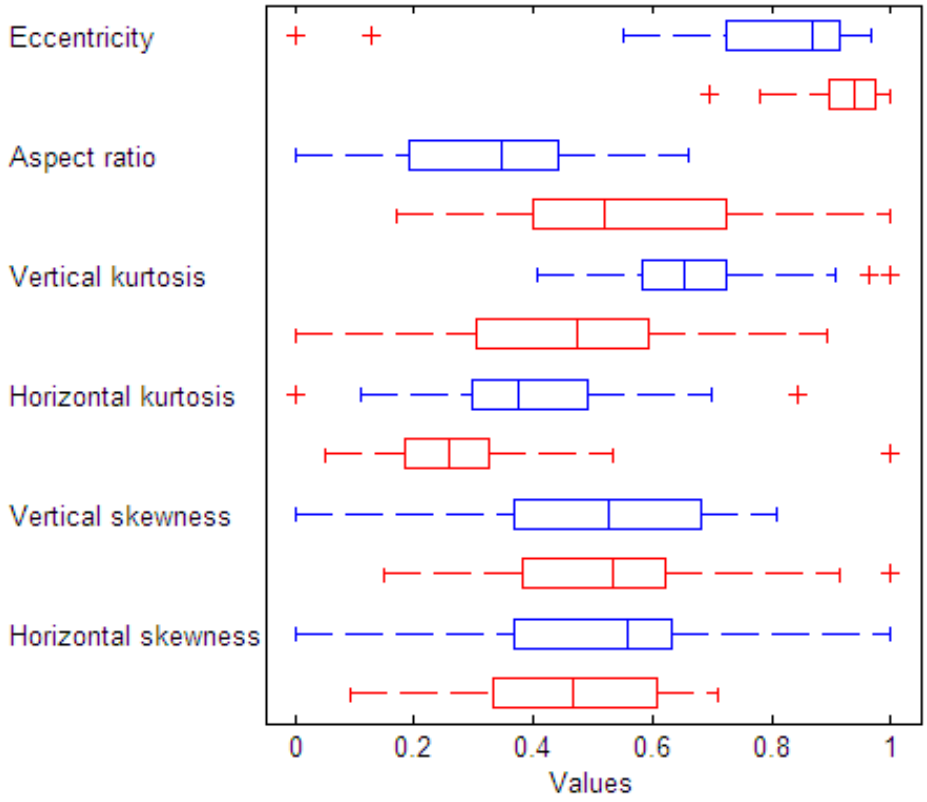


Figure 68: Boxplot of various normalised shape property values versus incision type. Incomplete incisions are in red and complete incisions in blue.

6.3.3.2 Discussion

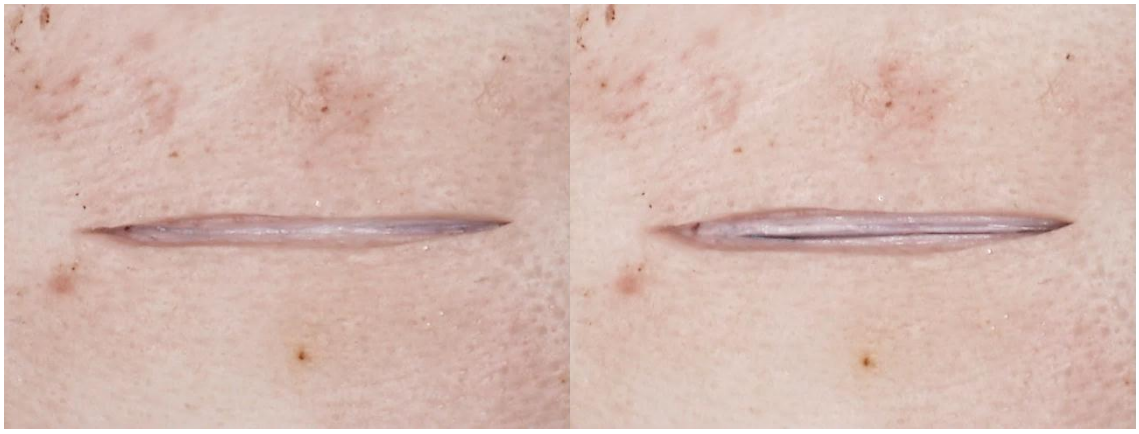
The average misclassification error rate based on the six shape features range from 19.5% - 41.5%. In particular, the worst misclassification error rate (average of 41.5% with a range of 34.7% – 51.4%) was based on the use of aspect ratio i.e. the extension of an incision, as a basis for classification. Therefore, the earlier assumption about a correlation between incision depth and incision type was not valid. While the aspect ratio of individual incisions is related to incision depth, the in-class variation of an incision's aspect ratio is sometimes larger than its between class variation i.e. two complete incisions can have aspect ratios that are less alike than a complete and incomplete incision.

In general, shape properties are not discriminatory enough to classify incisions, as there is significant overlap in the feature space of the six shape properties of the two classes of incisions (see Figure 68). Because shape

properties were not discriminatory enough to distinguish marginal and complete incisions, the visual appearance of these incisions were examined next.

6.3.4 Visual cues

It was observed that complete incisions exhibit a characteristic dark line, made as the scalpel blade incises the skin up to the cranium, which is not present in incomplete incisions (see Figure 69). As such, the presence of a dark line along the length on an incision might be used to identify a complete incision.



a | b

Figure 69: (a) Incomplete incision and (b) complete incision

The domestic pig, Sus scrofa, is the most appropriate model for all types of dermatological and surgical wound investigation, as well as a standard model for skin and plastic surgical procedures [151]. However, the use of cadavers meant that any bleeding resulting from the skin being incised could not be simulated, as cadaveric skin has no blood supply. Nevertheless, as any bleeding will normally be cauterised and irrigated, the visual appearance may not be too dissimilar and the techniques developed may potentially be applicable during actual surgery.

6.3.4.1 Line intensity profile

To detect the presence of the dark line corresponding to a complete incision, the line intensity profiles of six equally spaced 13-pixel wide line segments across the incision were determined (see Figure 70).

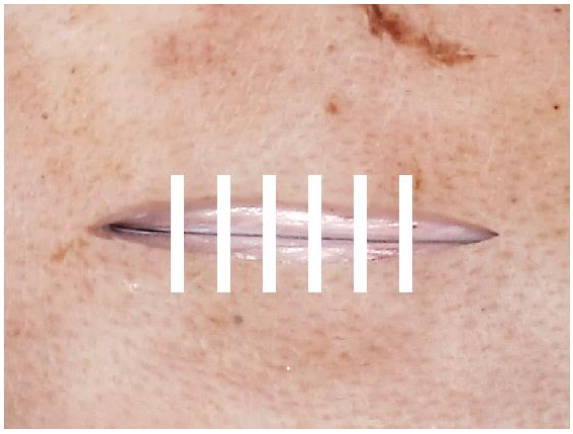


Figure 70: Line segments across an incision (line segments are drawn to scale)

This method does not require that the edges of an incision, which are difficult to segment, to be known. However, the start and end position of the incision is required. As the incision would be made robotically, it is assumed that the start and end positions of the incision would be known. A pixel on the complete incision line (characteristic dark line indicative of a complete incision) corresponds to a pixel along the line segment where the second derivative of the line intensity profile is maximum and exceeds an empirically determined threshold of 0.015 (see Figure 71). The threshold level was determined based on visual inspection of the resulting images, and is a compromise between false detection and under detection.

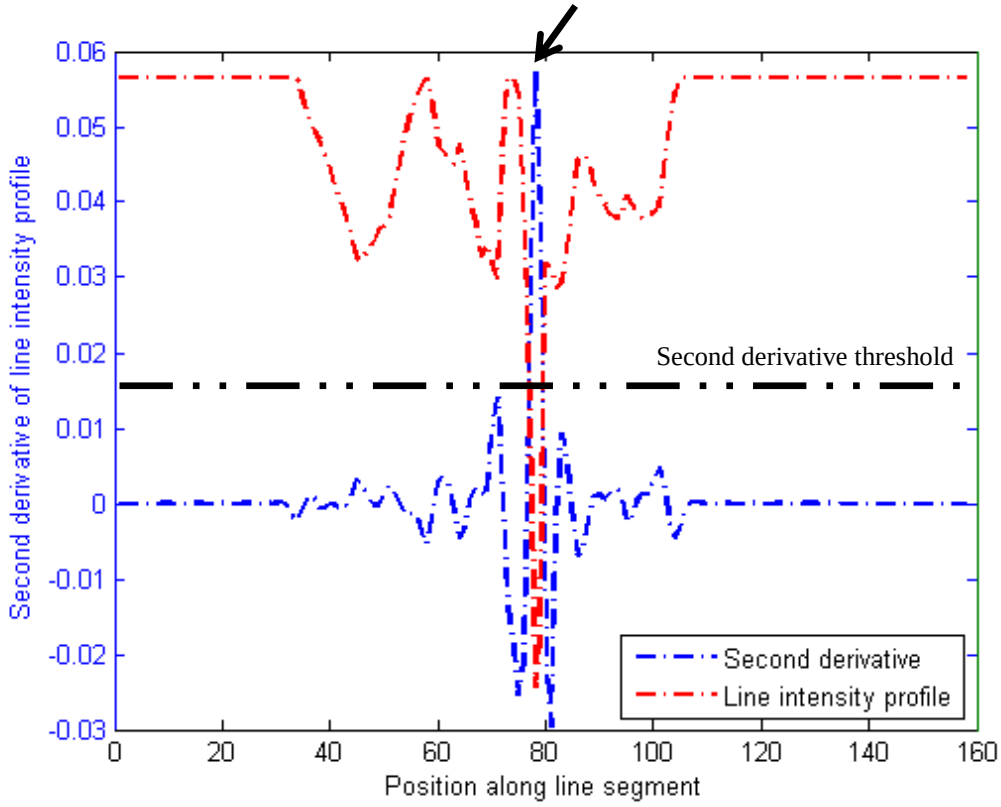


Figure 71: Intensity profile of a line segment across an incision and its second derivative. The position of the pixel on the complete incision line along the line segment is denoted by an arrow.

Because techniques based on intensity profiles and derivatives are susceptible to noise, several noise reduction techniques were used. First, the image was filtered with a low pass 3-by-3 Gaussian filter with a standard deviation of 0.5. The mean value of the 13-pixel wide line intensity profile of each line segment was used, and fitted to a cubic spline (see Figure 71), to reduce the effects of any outliers by way of smoothing.

The use of a relatively wide line segment also reduces the potential of erroneously detecting the boundaries of the incision, as the line intensity profile of a straight incision line would be closer to the profile of a complete incision line than that of the profile at the curve edges of an incision. To reduce the effects of specular reflection, which distorts second derivative values of the line intensity profile, the greyscale intensity values of pixels along a line segment greater than the mean of the line intensity profile was replaced with its mean value (see Equation 6.6).

$$I(x) = \begin{cases} \bar{I} & I(x) > \bar{I} \\ I(x) & I(x) \leq \bar{I} \end{cases} \quad 6.3$$

where $I(x)$ is the greyscale intensity value of a pixel along the line segment at position x and $\bar{I}$ is the mean greyscale intensity value of all pixels along the line segment. Additionally, techniques such as temporal filtering could be used where video images are available. However, in this work, only still images (snapshot) of the incision were used.

6.3.4.2 Results

Figure 72 and Figure 73 illustrate the output of the complete incision detection method, based on the detection of a complete incision line from its line intensity profile, for complete and marginal incisions. The crosses represent the pixels that correspond to the perceived complete incision line.

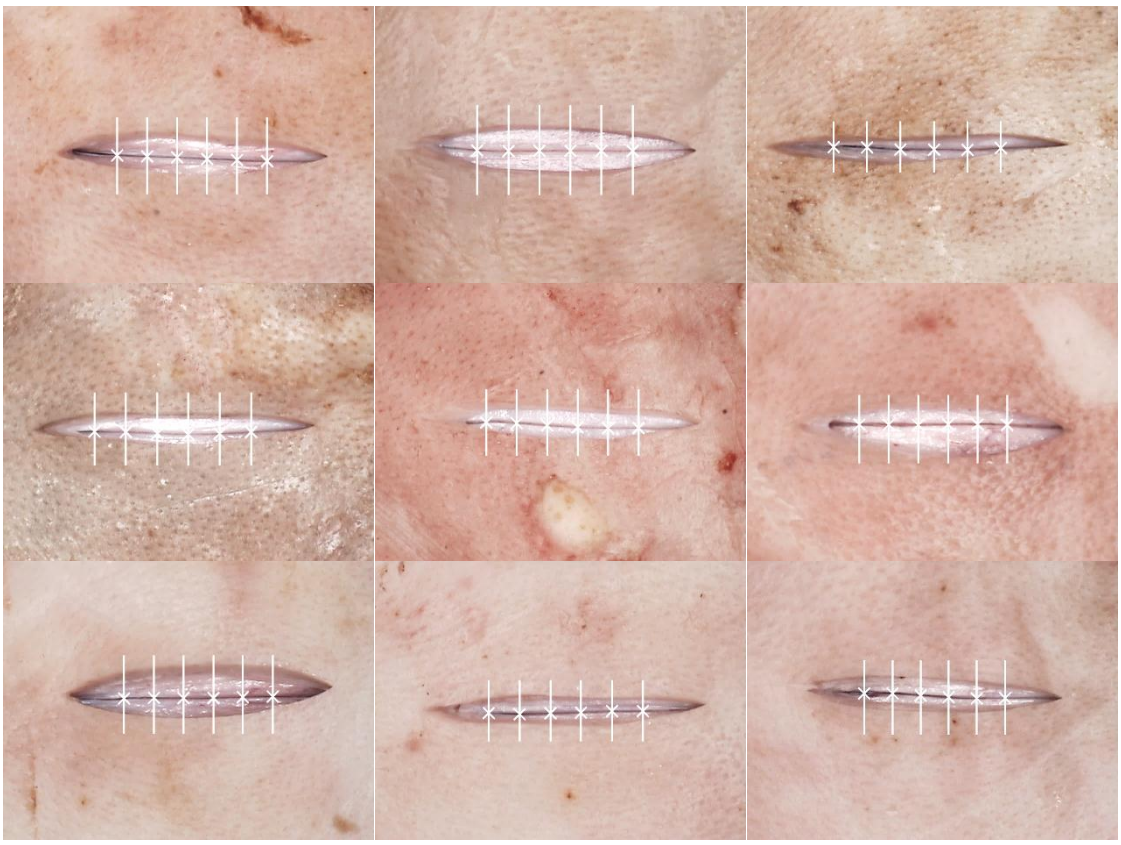


Figure 72: Illustrative examples of the output of the complete incision detection method for complete incisions. Further examples are included in Appendix B.

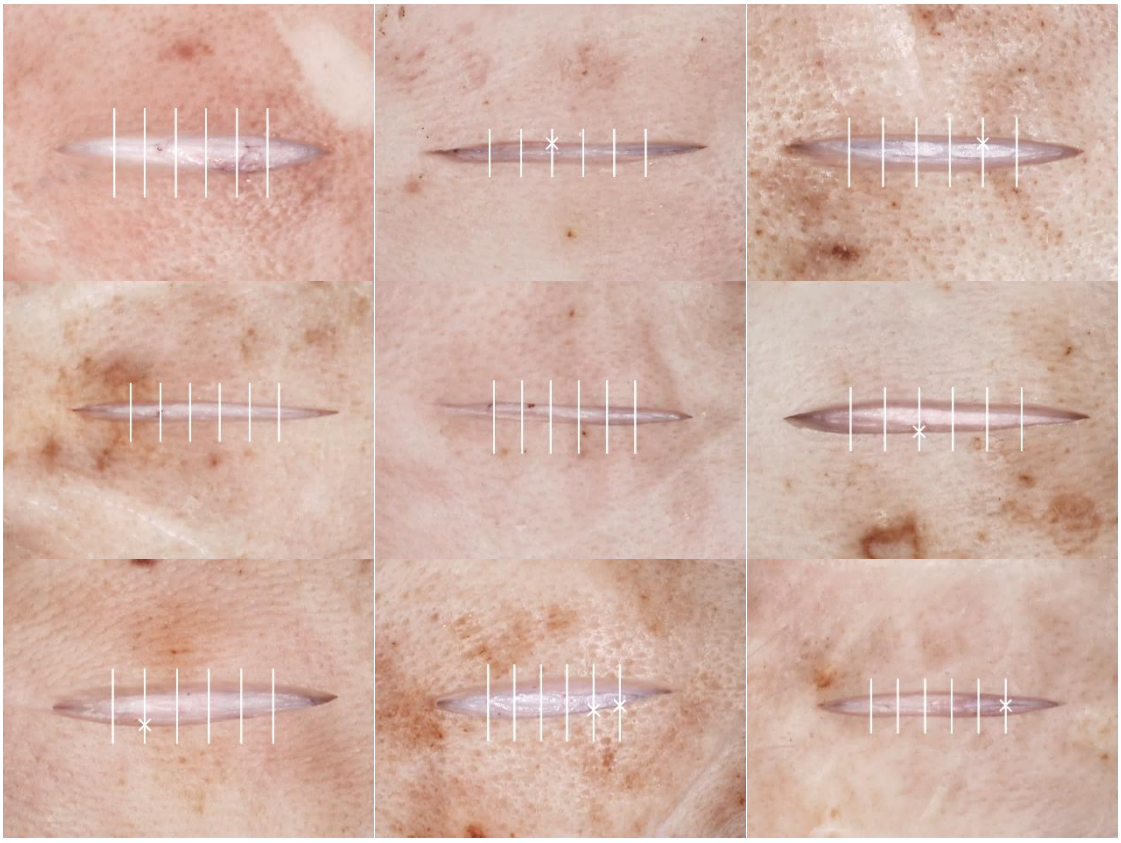


Figure 73: Illustrative examples of the output of the complete incision detection method for marginal incisions. Further examples are shown in Appendix C.

Sixty-four incisions, consisting of 27 marginal incisions and 37 complete incisions, were analysed. Figure 74 shows the number of lines segments where a complete incision line is detected, for complete and incomplete incisions. Marginal and complete incisions can be partially differentiated based on the number of line segments with a complete incision line detected. An incision is classified as incomplete if three or less line segments where a complete incision line was detected were found.

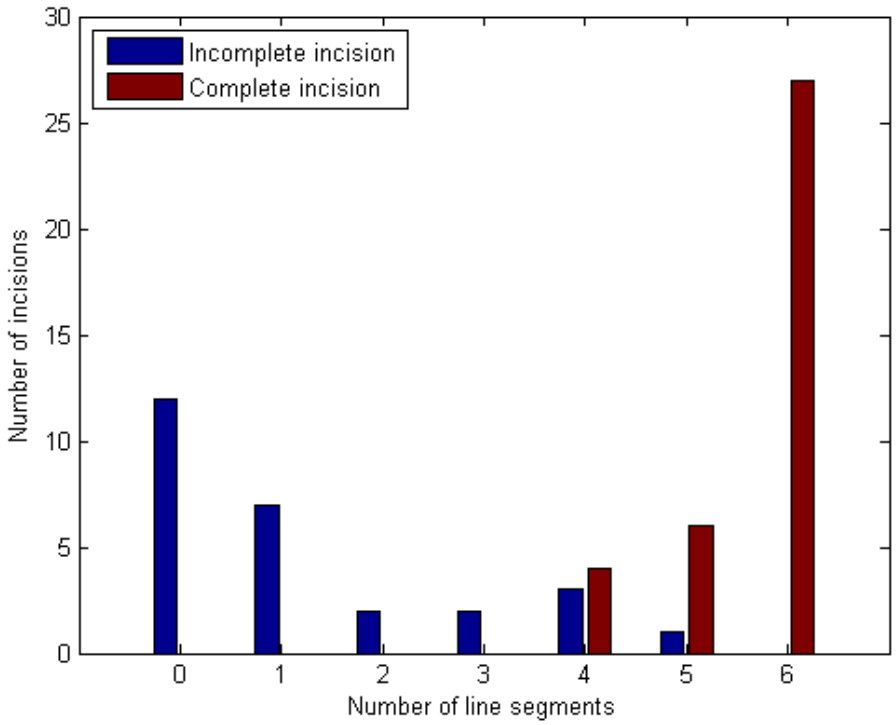


Figure 74: The number of line segments with a complete incision line detected, for incomplete and complete incisions.

The mean residual of a best-fit line in the least squares sense fitted to the pixels corresponding to the perceived complete incision line (see Figure 75) was determined and used to differentiate marginal and complete incisions further. To account for the different scales in the images, the ratio of the residual to the distance between the line segments was used.

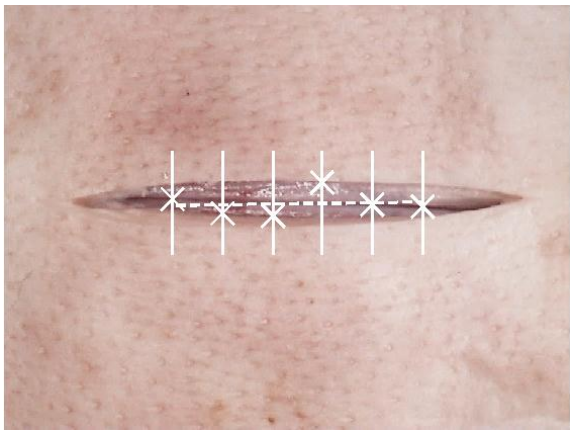


Figure 75: Line fitted to the pixels corresponding to the perceived complete incision line

The mean residual of this fitted line for marginal incisions (with a complete incision line detected in four or more line segments) and complete incisions are shown in Figure 76 and Figure 77.

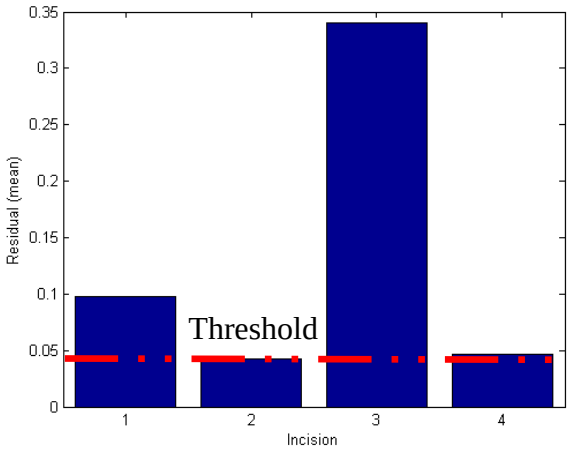


Figure 76: Mean residuals for marginal incisions (with a complete incision line detected in four or more line segments). The red dotted line is the mean residual threshold above which an incision is considered incomplete.

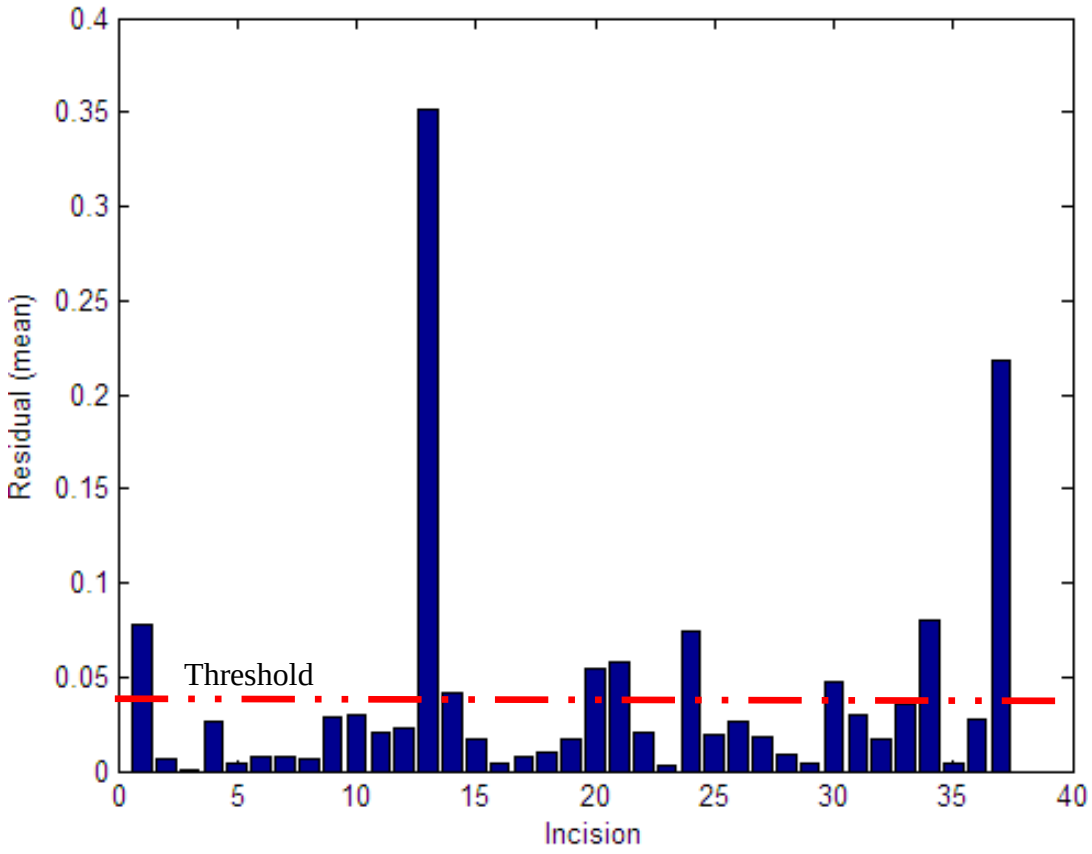


Figure 77: Mean residuals for complete incisions. The red dotted line is the mean residual threshold in Figure 76

6.3.4.3 Discussion

By setting a mean residual threshold of 0.04 for an incision to be considered as a complete incision, all the marginal incisions in this work can be precluded, although the overall misclassification error rate would increase. Nine out of the 64 sample incisions were misclassified, a misclassification error rate of 14.1%. Crucially however is that all the misclassifications are false negatives i.e. complete incisions classified as incomplete incisions (see Figure 78). In the visual inspection of an incision, false negatives i.e. complete incisions classified as marginal incisions are preferable to false positives i.e. marginal incisions classified as complete incisions. Erroneously classifying an incomplete incision as complete will result in an attempt by the robotic system to retract the incomplete incision. On the other hand, a false negative would only require further inspection before the surgical procedure could be continued, a minor inconvenience compared to the possible complications that may arise by attempting to retract an incision that has not been properly incised.

Misclassifications were due to the mean residual of complete incisions exceeding the mean residual threshold of 0.04. This threshold level was selected to avoid misclassification of marginal incisions. The high mean residual of misclassified complete incisions were due to pixels being incorrectly identified as belonging to a complete incision line. This is usually due to the complete incision line having lower contrast compared to the edges/boundaries of the incision.

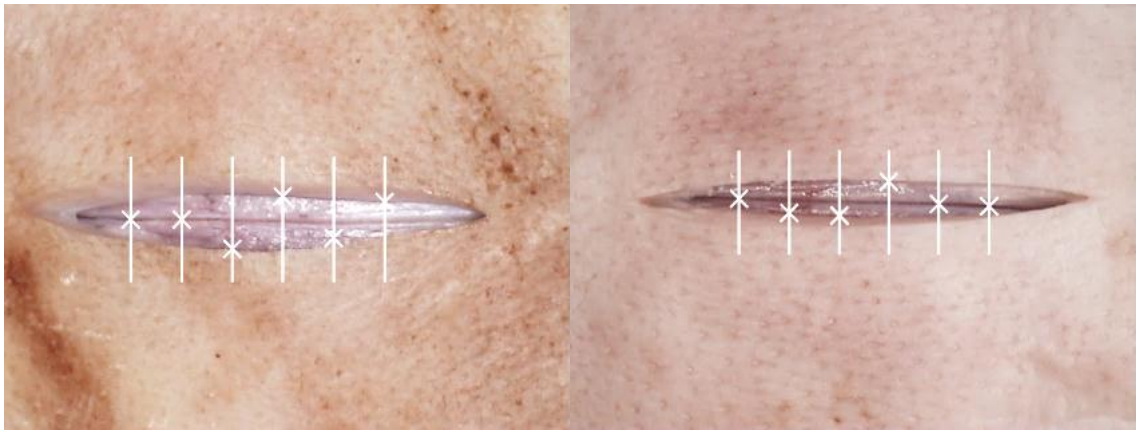


Figure 78: Examples of misclassified complete incisions

There are two main problems with this approach. The first is that the complete incision line may not always be obvious in complete incisions, making its automatic detection difficult. A possible solution is to enhance the contrast of the complete incision line by using techniques such as histogram equalisation. However, contrast enhancement inevitably increases the contrast of other features as well, such as the edges of the incision, which would invariably be detected instead of the complete incision line.

The second is that the maximum second derivative of the line intensity profile does not always correspond to the complete incision line (see Figure 79). A possible solution would be to assign greater weightage to the line intensity profile closer to the middle of the line segment, to improve the detection of an incision line over the edges of an incision, although this technique may not work if the incision line deviates from the middle of the line segment. Nevertheless, the visual inspection of an incision based on the detection of a complete incision line using a line intensity profile is the most promising technique of all the methods considered.

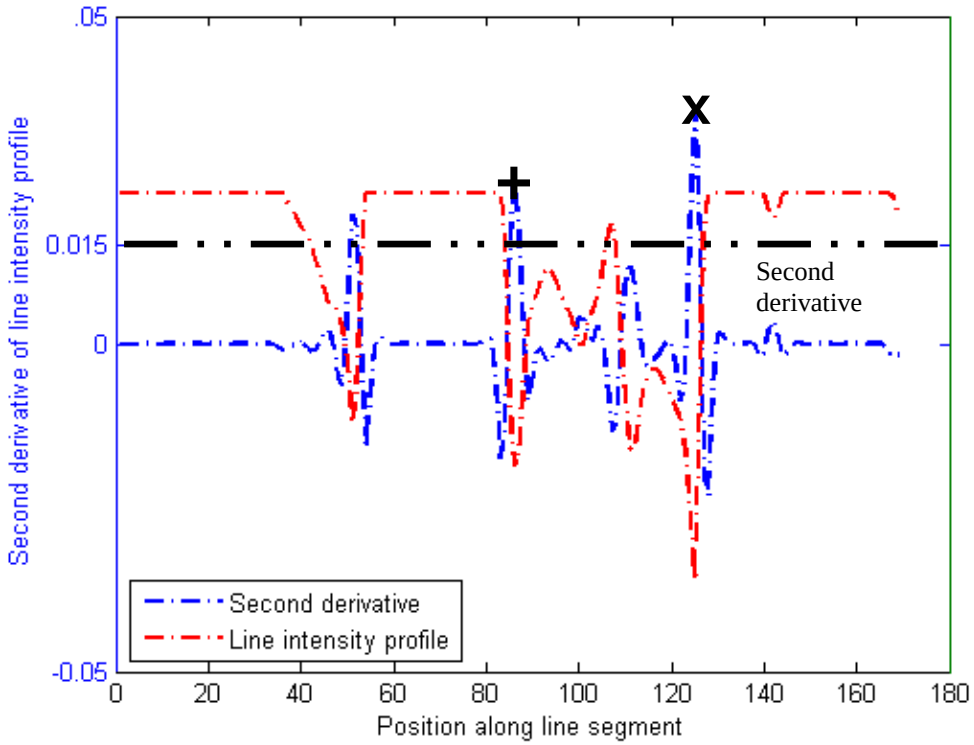
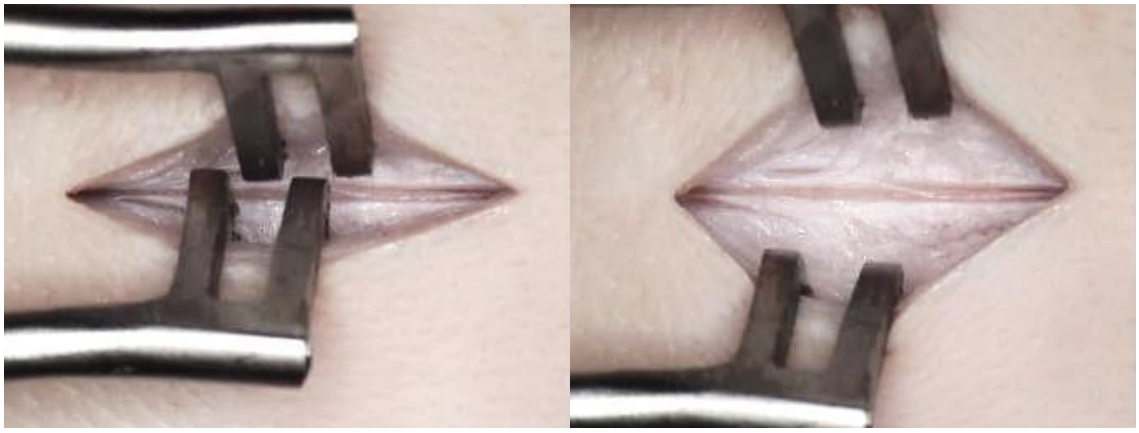


Figure 79: Intensity profile of a misclassified complete incision. 'x' indicates the perceived position of the complete incision line and '+' indicates its actual position.

6.4 Characterisation of a retracted incision

Once the skin has been properly incised, the next step involves retraction of the incision. An incision is considered completely retracted if it is able to accommodate a burr hole of a given size. While the retraction of an incision may be determined indirectly from the separation distance of the prongs of the retractors (see Figure 80), a potential problem with this approach is that the retractors may have extended without the incision being retracted. This would be the case if the prongs of the retractor have not engaged the edges of the incision, leading to the erroneous assumption that the incision is sufficiently retracted.



a | b

Figure 80: (a) Incomplete retraction and (b) complete retraction

The approach adopted instead, to assess if an incision has been sufficiently retracted, was to determine the maximum circular free space within the retracted incision. This was found by fitting the largest possible circle inside a region bounded by the edges of the incision and the prongs of the retractor (see Figure 81).

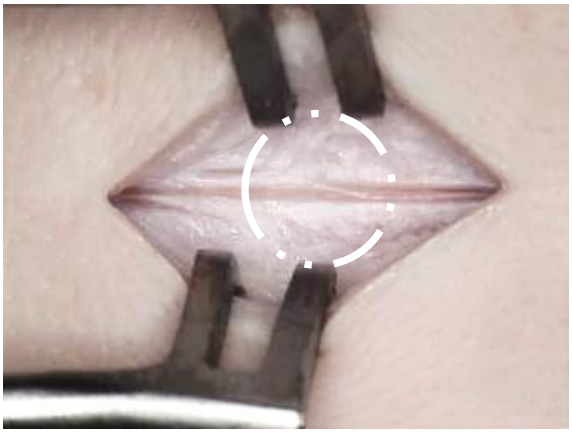


Figure 81: The largest circular free space area within the retracted incision

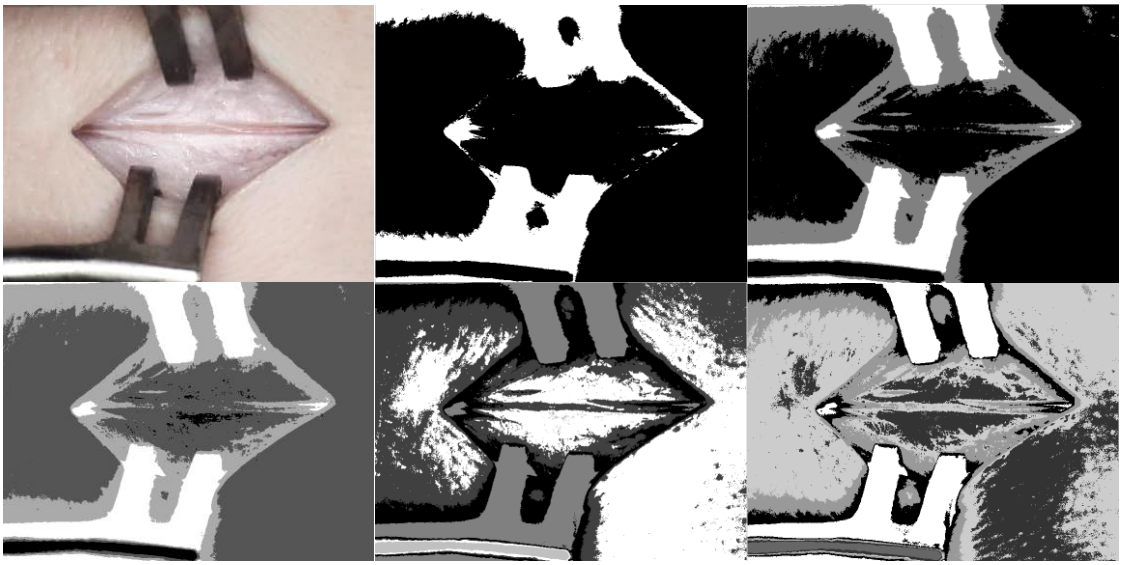
If the diameter of this circle is larger than the diameter of the prospective burr hole (within a specified tolerance), the incision is considered sufficiently retracted. The actual diameter of the fitted circle can be found if the scale of the image is known. A possible solution to determine image scale is by using the pixel separation distance of the prongs on each side of the retractor as a guide. Assuming that the actual distance is known a-priori, the scale of the image and hence the actual diameter of the fitted circle can be found.

6.4.1 Colour-based segmentation of the retractors

The edges of the incision and retractors have to first be detected in the image to define a region bounded by these edges. As the retractors are made from surgical stainless steel and have a dull grey colour that is distinct in the operative site, the difference in colour of the retractor and the background was used as a basis to segment the retractors. A colour-based segmentation using k-means clustering was used. K-means clustering groups objects together based on their spatial location in a feature space. The clustering is performed such that the distances of all objects to their respective cluster centroid are the least among all clusters. Objects belonging to the same cluster have the least distance to the centroid of their cluster compared to the centroid of other clusters.

6.4.1.1 Results

Figure 82 shows the segmented image containing the retractors using a k-means clustering colour-based segmentation with 2, 3, 4, 5, and 6 clusters respectively. The clustering was repeated three times at different randomly selected set of initial centroids to avoid local minima.



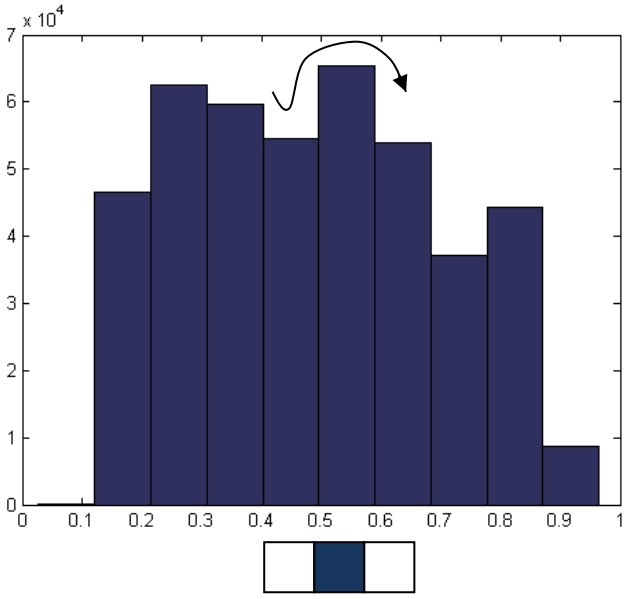
a	b	c
d	e	f

Figure 82: (a) Original colour image, and colour-based segmentation using k-means clustering with (b) 2,(c) 3,(d) 4,(e) 5, and (f) 6 clusters respectively.

6.4.1.2 Discussion

A disadvantage of colour-based segmentation using k-means clustering is that the number of clusters to be partitioned, which is in general not known, has to be specified. A way of automatically determining the number of clusters to be partitioned is based on the number of local maxima of a three-dimensional colour histogram of the image, which corresponds to the number of regions with the most dominant colour in the image [152]. These local maxima may be determined using a hill-climbing optimisation technique. To illustrate the hill-climbing optimisation technique, consider the one-dimensional histogram in Figure 83. The local maxima is be found by searching across the histogram using a three-neighbourhood search window. The k seeds are the number of

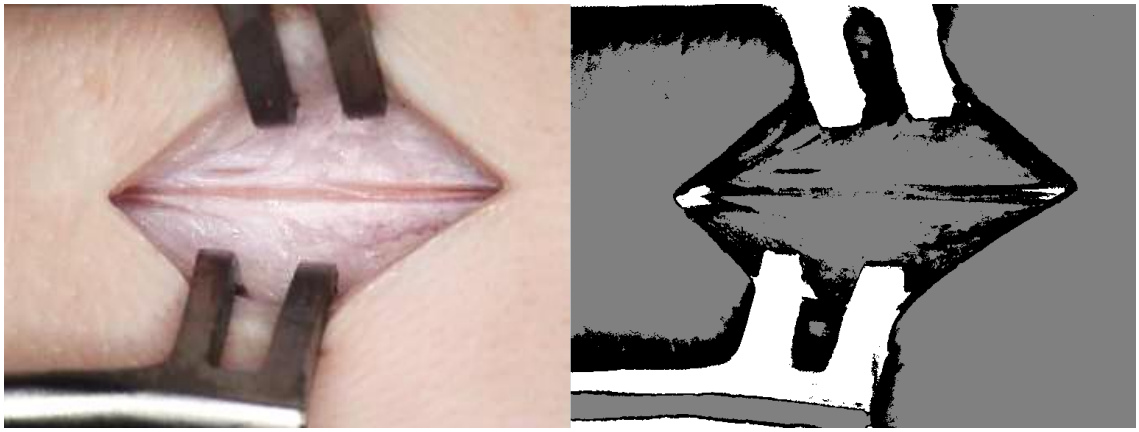
local maxima, with the initial seeds corresponding to the position of the histogram bin centre. The corresponding three-dimensional case has a 26-neighbourhood search window. Specifying the initial cluster centroid positions or ‘seeds’ improves the performance of the k-means clustering.



a
b

Figure 83: (a) A one-dimensional histogram and the local maxima found using (b) a three-neighbour search window

Figure 84b shows the segmentation results, with the different grey levels indicating the different clusters. A colour based segmentation using a k-means clustering algorithm was used, with the number of clusters determined automatically with the initial cluster centroid positions or ‘seeds’ specified. The ‘seeds’ are the local maxima (determined using a hill-climbing optimisation technique) of a three-dimensional CIELAB colour space histogram of the image, with 12 histogram bins used for each dimension. Although segmentation results may differ depending on the number of histograms bin used, the algorithm is relatively robust to small deviations in the number of histogram bins used e.g. 10 - 15.



a | b

Figure 84: (a) Original colour image and (b) segmented image of the retractors

6.4.2 Colour recognition of the retractors

Colour-based segmentation has successfully segmented the retractors by first dividing the colour image into regions of the most dominant colour. The next step after colour-based segmentation is colour recognition of the segmented regions to identify the retractor. Because of the distinctive colour of surgical instruments compared to background tissue, the retractors can be differentiated by colour recognition. A representative colour value for the retractors was obtained by averaging the colour value of the pixels for a sample region of the retractor. This colour value was subsequently used to select the region whose average pixel colour value is closest to the representative colour value for the retractors based on the nearest neighbour rule.

6.4.2.1 Results

Sample colours in CIELAB colour space for the retractor and background were obtained from 51 images. The mean sample colour values for the retractor and background is shown in Table 6. Illustrative examples of the colour recognition of retractors based on these values are shown in Figure 85.

Table 6: Retractor and background mean L*, a* and b* sample values

Pattern class	Dimensions		
	L*	a*	b*
Retractor	76.1729	7.8286	5.7752
Background	58.0143	3.8930	2.7737

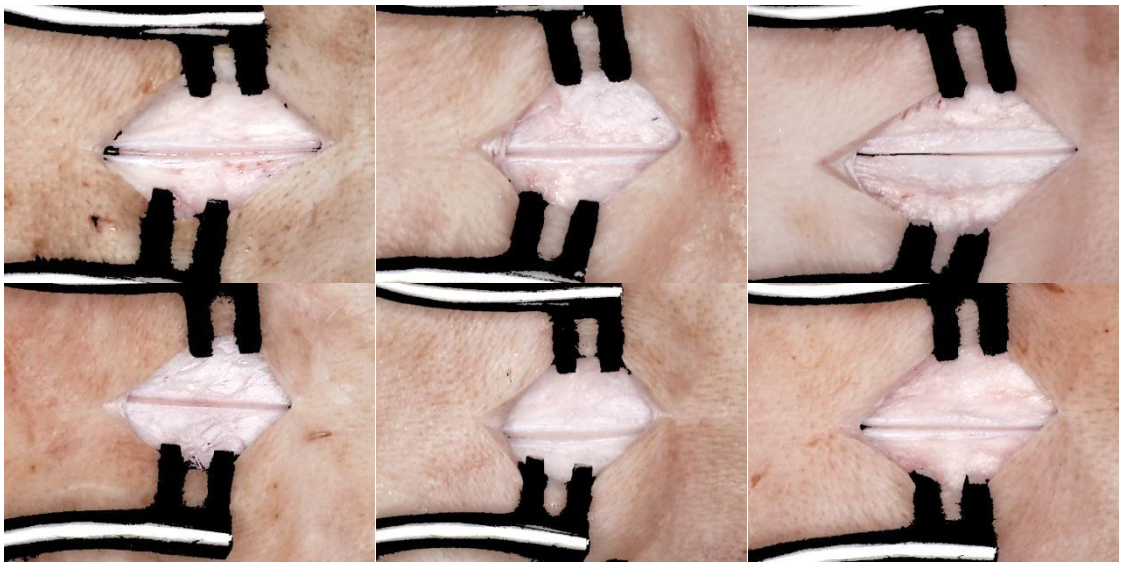
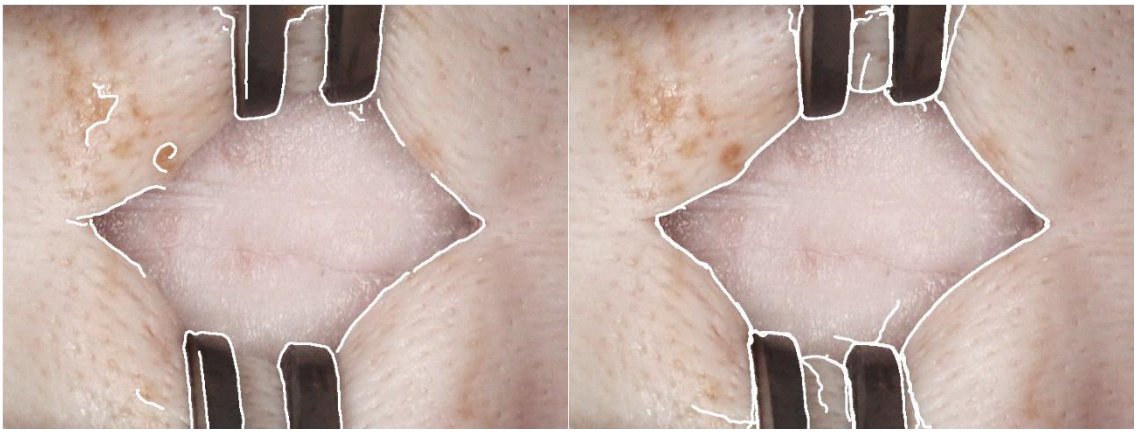


Figure 85: Illustrative examples of retractors identified based on colour recognition

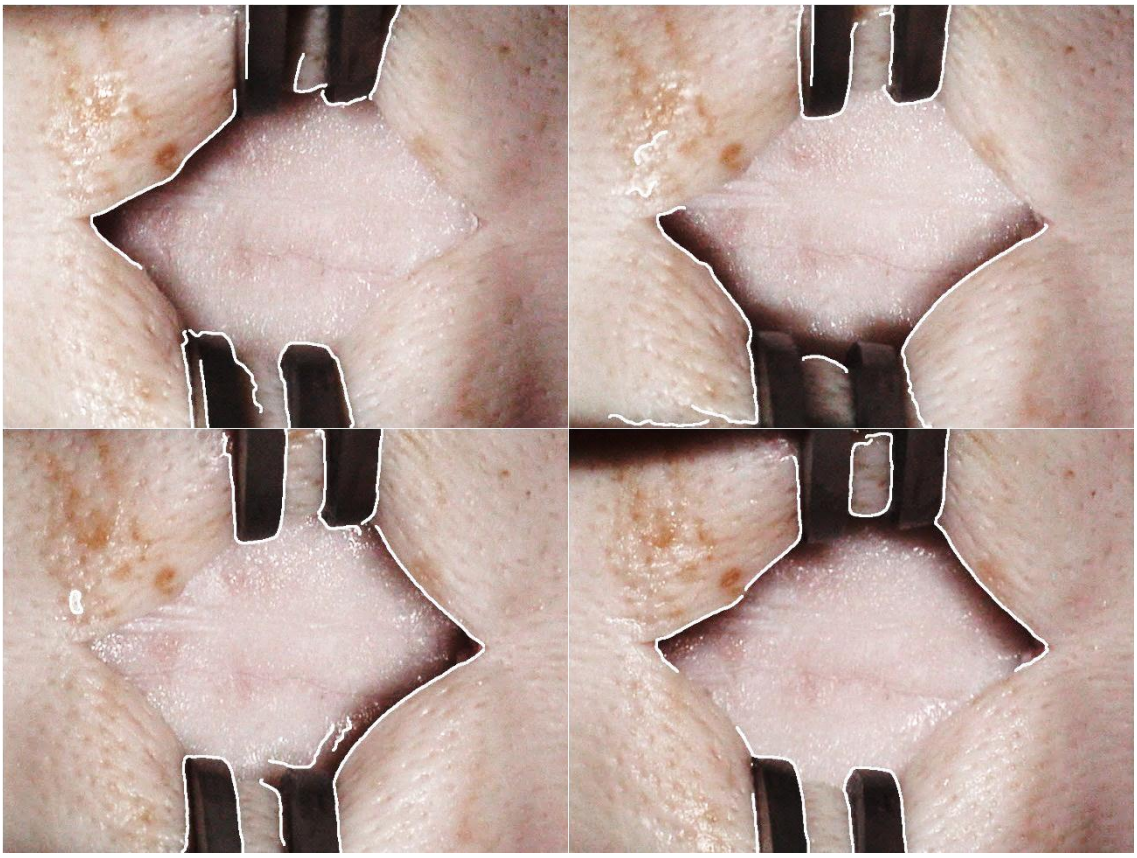
6.4.2 Segmentation of the retracted incision

The retracted incision was segmented next using a Canny edge detector to detect its edges. However, the edges of the retracted incision cannot always be detected because of the sometimes low contrast at the boundaries of the incision (see Figure 86). To improve edge detection, sections of the ring light were turned off to cast shadows at the boundaries of the incision to increase their contrast. Figure 87 illustrates the technique.



a	b

Figure 86: Retracted incision overlaid with edges detected using a Canny edge detector and illuminated (a) with all quadrants of the ring light and (b) by alternating each quadrant of the ring light

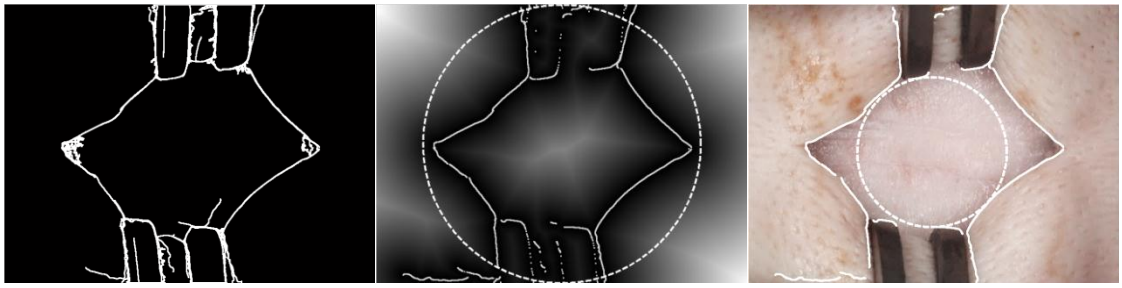


a	b
c	d

Figure 87: Detected edges (using Canny edge detector) of a retracted incision when illuminated with (a) left half, (b) bottom half, (c) right half and (d) top half of the ring light respectively. The composite image is shown in Figure 86b.

6.4.3 Distance map

Once the edges of the incision and retractors were detected, a distance map (see Figure 88b) based on the distance transform of all the pixels in the image to an edge map of the retractors and the retracted incision (see Figure 88a) was determined. The pixel with the maximum distance within a circular region of interest (see Figure 88b) is the centre of a circle that corresponds to the maximum free circular space available to accommodate a burr hole. The radius of the fitted circle is the distance of this pixel to either the nearest edge pixel of the prongs of the retractors or the edge pixel of the retracted incision. It is important that pixels within the edges of the incision be removed, as the position and diameter of the fitted circle will be affected otherwise (see Figure 90). Pixels within the edges of the incision were removed by selecting the first and last pixel of an edge map when traversing vertically i.e. vertical scanning of the edges (see Figure 88b).



a	b	c

Figure 88: (a) An edge map, (b) a distance map overlaid with the edge map and a circular region of interest and (c) the fitted circle

6.4.3.1 Results

Figure 89 and Figure 90 illustrate examples where the technique successfully fitted the largest circle within the retracted incision and examples where the technique failed respectively.

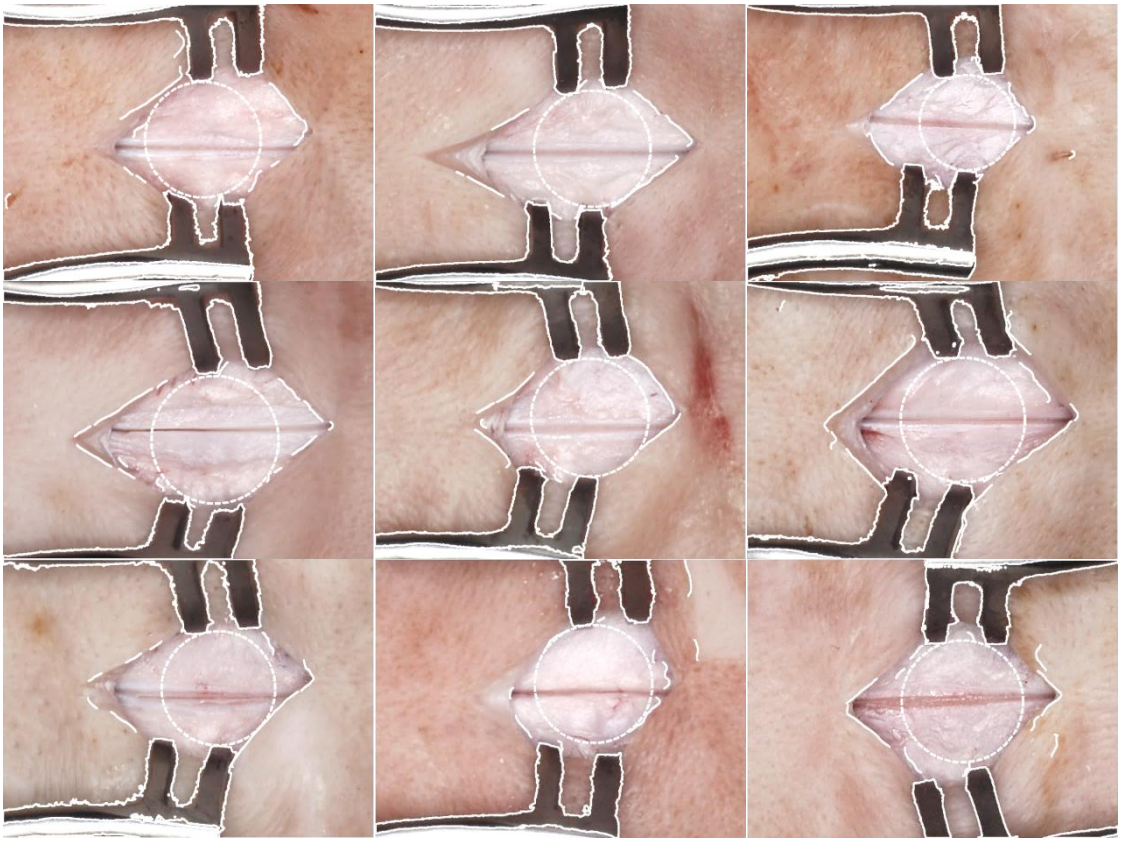


Figure 89: Illustrative examples of a correctly fitted circle within the retracted incision. Further examples are included in Appendix D.

Figure 90: Examples illustrating an incorrectly fitting circle within the retracted incision

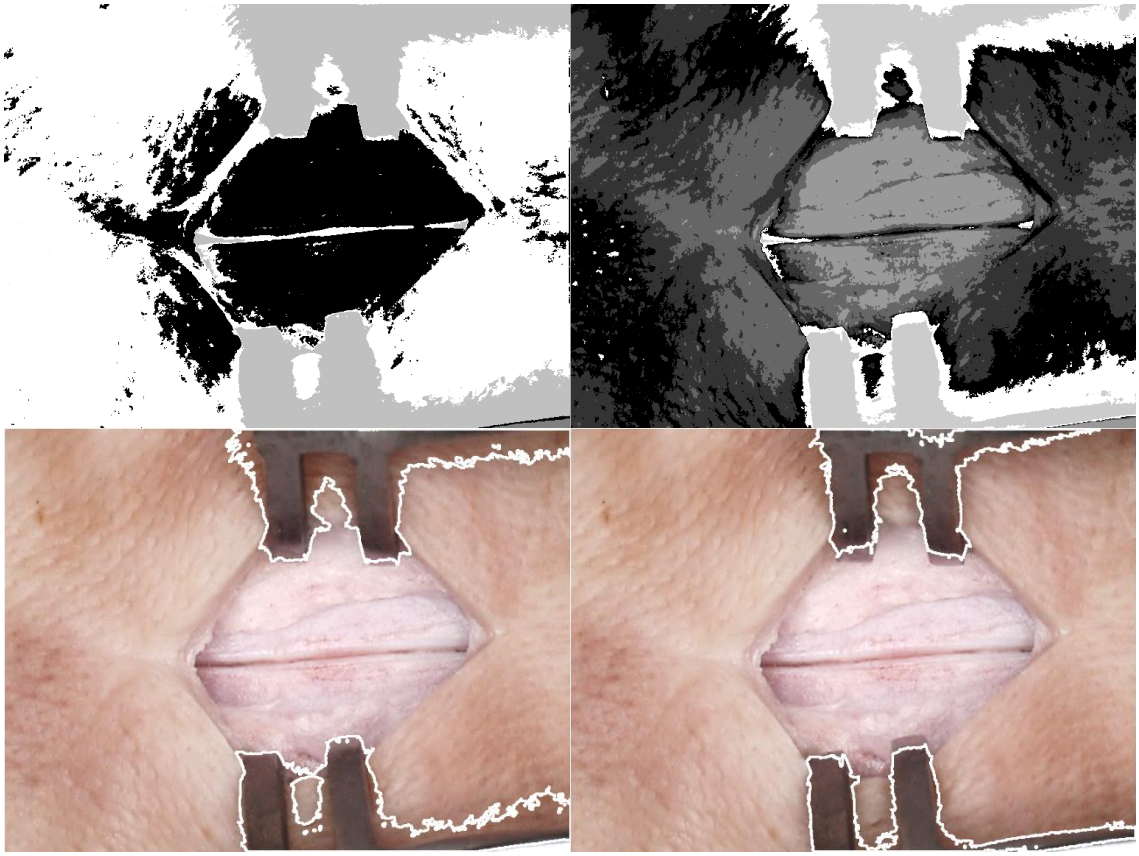
6.4.3.2 Discussion

In all examples where the technique failed (see Figure 90), this was because the edges of the retracted incision or the prongs of the retractor were not segmented well. The segmentation of the retractors failed in some images because the number of clusters (determined automatically) used for colour-based segmentation was inadequate. As a result, due to the proximity of the colour of the retractor with part of the background skin and/or shadows, the retractors are sometimes segmented along with these regions (see Figure 91c).

Segmenting the retractors robustly is difficult because of the different skin/background colours, the presence of shadows and specular reflections results in a complex scene that can complicate segmentation of the retractor based on colour alone. Specular reflections, due to the reflectivity of the retractor surface, can be reduced by using retractors with a matt surface (instead of retractors with a mirror like surface) as well as by using a diffuser to provide softer illumination of the scene. This also reduces shadows in the scene.

A possible solution to improve segmentation of the retractors is to over-segment the scene (see Figure 91). In the example below, it was possible to differentiate between the retractors and shadows in the scene by “over-segmenting” e.g. k-means clustering using six clusters instead of the five clusters that were originally specified based on the local maxima of the three-dimensional colour histogram of the image.

Alternatively, the colour segmentation and recognition of the retractors can be contrived to be simple by using a retractor with a distinct colour that will not be mistaken with its background and is not present in the operative field, such as the colour cyan. However, this would involve either the use of non-standard retractors, or painting of the retractors (which may not be feasible in terms of sterility requirements).



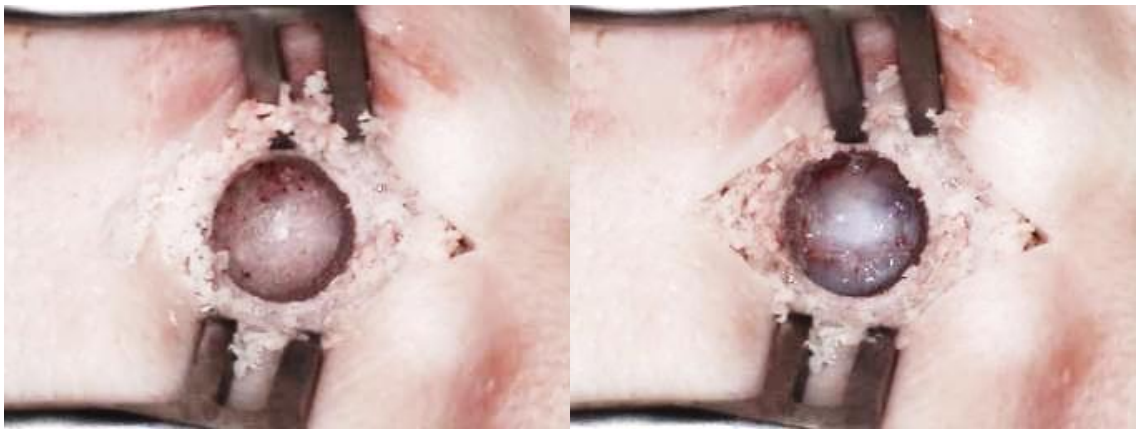
a	b
c	d

Figure 91: Colour based segmentation using k-means clustering algorithm with (a) five clusters and (b) six clusters. (c) – (d) The original colour images overlaid with edges corresponding to the retractors in (a) and (b) respectively.

Although the edges of the incision were not always detected in the majority of the images (ring lighting was not used to facilitate detection of the edges), this technique has successfully fitted relatively large circles within the retracted incision. The technique does not always require that all the edges of the incision be found. Circles have been fitted successfully within the retracted incision although some of the edges were missing (see Figure 89). Nevertheless, as shown in Figure 90, where edges of the incision are not adequately detected, the position and size of the fitted circle can be erroneous. Where the edges of the retracted incision were not properly detected, the use of directional ring lighting would facilitate its detection (see Figure 86).

6.5 Characterisation of a burr hole

The inspection of a burr hole is the most straightforward of the three surgical tasks under consideration. An incomplete or partial burr hole is where the skull has not been penetrated. Depending on whether the partial burr hole has been drilled to the cortical or cancellous layer of the skull, it would have either a smooth or a 'spongy' appearance, although both will have a predominantly white colour (see Figure 92a). A complete burr hole is one where the underlying dura mater i.e. the outermost part of the brain, is visible. The dura mater is primarily greyish in colour and has a membrane like appearance and texture (see Figure 92b). Possible features to distinguish the two types of burr holes may therefore include colour and texture.



a | b

Figure 92: (a) Incomplete/partial burr hole and (b) complete burr hole.

6.5.1 Segmentation of the burr hole

For the visual inspection of a burr hole, it is first necessary to detect its presence in an image. Because of its circular shape, a Hough transform to detect circles was used for the detection of a burr hole. The Hough transform accumulator cell is a set of all points (x, y) in 3D parameter space consisting of the circle centre coordinates (a, b) and radius, r (range of radius values specified) such that (Equation 6.4):

$$(x - a)^2 + (y - b)^2 = r^2 \qquad 6.4$$

High accumulator cell counts represent circles corresponding to a potential burr hole. A circle is disregarded if the ratio of its accumulator cell count i.e. the number of detected edge pixels, to the circle perimeter is less than 10%. This ratio eliminates false positives i.e. a burr hole is sometimes ‘detected’ when there is none. The 10% ratio was found to represent an acceptable compromise between false detection and under detection, as determined by visual inspection. The circle with the highest ratio of detected edge pixels i.e. accumulator cell count, to the number of pixels along the perimeter of the circle (calculated based on the radius of the circle), was selected. This is so that the algorithm was not biased towards larger circles, which may have a higher accumulator count over smaller circles. Figure 93 and Figure 94 show examples of the detected edge of incomplete and complete burr holes respectively, using this technique.

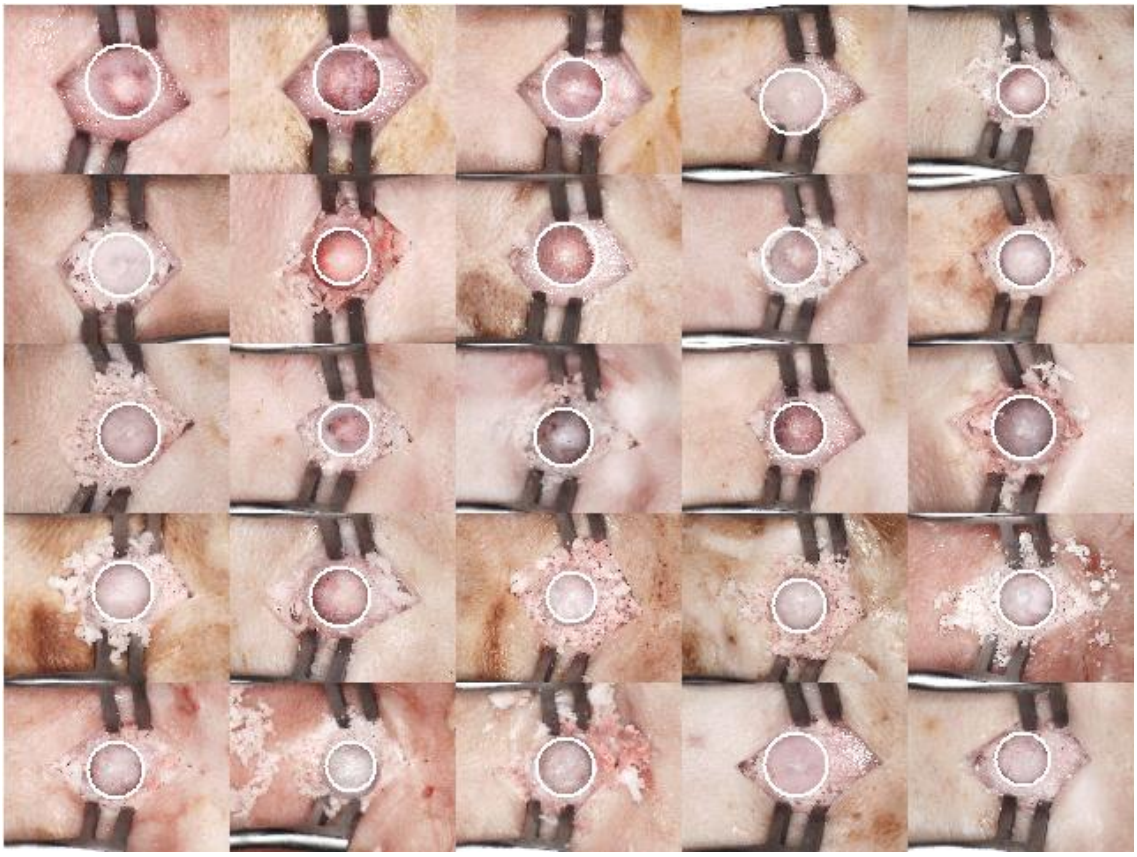


Figure 93: Detected incomplete burr holes

Figure 94: Circles corresponding to complete burr holes

6.5.2 Classification

Once isolated, features that can distinguish the two classes of burr hole i.e. incomplete and complete were identified. Based on the visual appearance of the burr hole, it is likely that colour and texture properties might be used to distinguish the two (see Figure 92). The visual inspection of the burr hole was formulated as a classification problem with two classes i.e. incomplete and complete, using colour and texture features used as a basis for classification. Images of burr holes were labelled for ground-truth and classified using a naïve Bayes classifier, a linear classifier, LDA and a quadratic classifier, QDA.

6.5.2.1 Naïve Bayes classifier

A naïve Bayes classifier was used to perform the classification, assuming two pattern classes i.e. incomplete and complete, in three dimensions L*, a*, and b*. The patterns of each class were assumed to have a Gaussian distribution in CIELAB colour space. In the three-dimensional case, the probability density function, $p(x|w_j)$ of class w_j ($w_0 = incomplete, w_1 = complete$) and feature,$x$ where $x = (L^*, a^*, b^*)$ is given by Equation 6.5:

$$p(x|w_j) = \frac{1}{(2\pi)^{3/2}|C_j|^{1/2}} e^{-\frac{1}{2}(x-m_j)^T C_j^{-1}(x-m_j)}$$

$$Incomplete(x,y) = \begin{cases} 1 & p(x|w_0) > p(x|w_1) \\ 0 & otherwise \end{cases} \quad 6.5$$

$$Complete(x,y) = \begin{cases} 1 & p(x|w_1) > p(x|w_0) \\ 0 & otherwise \end{cases}$$

where each pattern class is governed by Gaussian densities with covariance matrix C_j and a mean vector m_j whose values are estimated using Equation 6.6:

$$m_j = \frac{1}{N_j} \sum_{x \in w_j} x$$

$$C_j = \frac{1}{N_j} \sum_{x \in w_j} xx^T - m_j m_j^T \quad 6.6$$

The form of the decision function used, $d_j(x)$ is given in Equation 6.7:

$$d_j(x) = \left(lnP(w_j) - \frac{1}{2} ln|C_j| - \frac{1}{2}\left[(x-m_j)^T C_j^{-1}(x-m_j)\right] \right) \quad 6.7$$

with $P(w_j)$the likelihood of a class appearing. The decision function minimises the penalty associated with misclassification.

6.5.2.2 Linear discriminant analysis

A linear discriminant classifier is similar to a naive Bayes classifier, in that both assume normally distributed features. However, LDA does not assume that the features are independent (it uses a pooled estimate of covariance instead of a diagonal covariance matrix estimate). LDA is closely related to principal component analysis (PCA) in that both can be used to find a linear combination of features that is most representative of the data and separates two or more classes. The difference between the two however is that LDA takes into account differences between the classes of data while PCA does not.

LDA for two classes assumes normally distributed conditional probability density functions $p(\vec{x}\,|y) = 0$ and $p(\vec{x}\,|y) = 1$ where $\vec{x}$ is a set of features for each sample of an object or event with a known class y. Given only a feature, $\vec{x}$, the classifier attempts to find a good predictor for the class y of any sample of the same distribution. The decision criterion to predict points as being from the second class is a threshold on the dot product given by Equation 6.8 for mean (μ_0, μ_1) and some threshold c:

$$\sum^{-1}(\mu_1 - \mu_0) \cdot \vec{x} < c \tag{6.8}$$

6.5.2.3 Quadratic discriminant analysis

A natural extension of the LDA, the quadratic discriminant analysis (QDA) fits to normally distributed data better, although its performance degrades for non-normal data. The Bayes optimal solution for some threshold T is given by Equation 6.9:

$$\begin{aligned}(\vec{x} - \vec{\mu}_0)^T \sum_{y=0}^{-1} (\vec{x} - \vec{\mu}_0) + ln\left|\sum_{y=0}\right| \\ - (\vec{x} - \vec{\mu}_1)^T \sum_{y=0}^{-1} (\vec{x} - \vec{\mu}_1) + ln\left|\sum_{y=1}\right| < T\end{aligned} \tag{6.9}$$

where $(\vec{\mu}_0, \sum_{y=0})$ and $(\vec{\mu}_1, \sum_{y=1})$ are the mean and covariance for each class respectively. Both LDA and QDA classify data by fitting a multivariate normal density to each group. However, the covariance estimate is pooled in LDA while the covariance estimate is stratified by group in QDA.

6.5.3 Colour space

An important aspect when using colour as a feature is the use of an appropriate colour space. There are many colour spaces available, with the more common ones being RGB, CIELAB and HSI. While the RGB colour space is the most commonly used colour space, its main drawback is that its brightness and chromaticity (hue and saturation) components are not separable, meaning that changes in lighting will affect colour information. The use of normalised RGB components can remedy this to some extent by reducing the effects of variations in intensity.

Generally, hue, saturation and intensity are better at distinguishing one colour from another than RGB components. The HSI (hue, saturation, intensity) colour space for example, decouples the intensity component from the colour-carrying component (hue and saturation) in a colour image. Intensity is analogous to intensity in a greyscale image, while hue and saturation refers to the dominant colour and 'purity' or amount of white light of the colour respectively. Owing to the different ways in these which colour spaces represent colour, the use of RGB for image processing using colour is limited [153]. RGB is well suited for image generation e.g. image capture by colour camera or image display in a monitor screen while the HSI model is more suited for colour image processing.

Both the RGB and HSI colour spaces lack perceptual uniformity and do not adequately separate the colour components. Unlike the RGB and HSI colour models, which model the output of physical devices rather than human visual perception, the CIELAB colour space is a device-independent model that approximates the perceptual uniformity of human vision i.e. a change of the same amount in a colour value produces a change of about the same visual importance. Because it is a perceptually uniform colour space, CIELAB is

mainly used in image segmentation. It also conforms to international standards, allowing measurements to be compared easily.

6.5.3.1 CIELAB colour space

The coordinates of the CIELAB colour space are dimension L^* for lightness and a^* and b^* for the colour components. L^*, a^* and b^* are represent lightness of colour (L^* = 0 is black and L^* = 100 is diffuse white; specular white may be higher), position between red and green (positive a^* values indicate red and negative a^* values indicate green) and position between yellow and blue (positive b^* values indicate yellow and negative b^* values indicate blue) respectively. Chroma and hue correlates to the radial distance $\sqrt{(a^*)^2 + (b^*)^2}$ and angular position $\tan^{-1}\left(\frac{a^*}{b^*}\right)$ in the a^*b^* plane respectively. The CIELAB space is relative to the white point and does not define absolute colours unless the white point of the data they were converted from is also specified. The absolute colorimetric International Color Consortium (ICC) CIELAB values are relative to CIE standard illuminant D50.

6.5.4 Colour difference

The difference in colour between a burr hole and a surrounding circular area, with a radius of 1.5 times the radius of the hole (see Figure 95) was computed to determine if colour difference could be used as a basis to differentiate between incomplete and complete burr holes. A circular area of radius 1.5 times the radius of the burr hole was chosen so that it included the skull surrounding the burr hole, while keeping within the confines of the retracted incision. Colour difference is the difference or distance between two colours in a colour space based on a distance measure such as Euclidean distance. The CIE delta-E colour difference, ΔE_{ab^*}, as given by Equation 6.10, is the colour difference between two colours in CIELAB space. A $\Delta E_{ab^*} \approx 2.3$ corresponds to a just noticeable difference [154], below which the two colours are indistinguishable (unless placed adjacent to one another).

$$\Delta E_{ab^*} = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \qquad 6.10$$

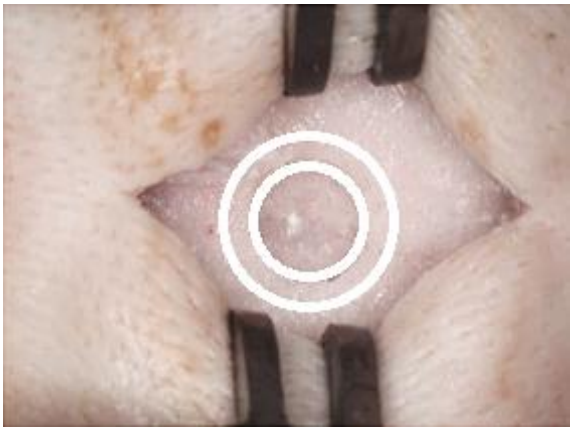


Figure 95: Circular areas used for the determination of CIE delta-E colour difference

6.5.4.1 Results

A random subset of 108 images (54 images of complete burr holes and 54 images of partial burr holes) was classified based on the CIE delta-E colour difference, a measure of the 'distance' between two colours. The sRGB[22] images of burr holes were first converted into the CIELAB colour space (based on ITU-R Recommendation BT.709 using the CIE illuminant D50 white point reference). The error in transforming sRGB to CIELAB to sRGB is approximately 10^{-5}. Three classifiers, a naïve Bayes classifier, LDA and QDA were used.

The leave one out naïve Bayes, LDA and QDA misclassification error rate based on CIE delta-E colour difference of incomplete and complete burr holes are shown in Table 7. In leave-one-out classification, each image from the dataset is used as validation data, and the remaining images used as training data. This process is repeated such that each observation in the sample is used once as the validation data. The mean and maximum classification error was obtained by repeating the training and testing of the images ten times.

[22] Standard RGB colour space, a device independent colour definition specified in IEC 61966-2.1

Table 7: Naïve Bayes, LDA and QDA leave one out (mean/maximum) misclassification error rate (%) based on CIE delta-E colour difference pixel values of incomplete and complete burr holes

Classifier		
Naïve Bayes	LDA	QDA
19.5 / 24.1	19.5 / 24.1	18.7 / 22.2

6.5.4.2 Discussion

When computing colour difference, images from earlier experiments were excluded from the analysis due to incorrect illuminating conditions at the site of the burr holes. This was because the ring lighting unit was place too close to the site of the burr hole, outside its working distance, resulting in poor illumination. Because intensity is a component of the colour difference computation, it is important that the ring lighting unit be used within its 40mm - 250mm working distance. This effect was more pronounced in complete burr holes that appeared darker than incomplete ones because the amount of illumination was reduced considerably at increasing depths.

The misclassification error rates using the three different classifiers are quite similar, with QDA slightly outperforming the classification performance of naïve Bayer classifier and LDA. The best QDA leave one out misclassification error rate is 18.7% (for 10 iterations) with a maximum error rate of 22.2% based on the CIE delta-E colour difference (median). The maximum error rate, which is a more meaningful measure in terms of classification error rates, is relatively high. The colour difference between the burr hole and a circular area surrounding it seemed like an obvious choice to differentiate between incomplete and complete burr hole. Although on average the colour difference for complete burr holes is higher than incomplete burr holes, there is some degree of overlap in the colour difference values of the two types of burr holes, as shown in Figure 96. As such, colour difference is not discriminatory enough to be used as a basis to classify the two types of burr holes. A different feature space, the colour components of the pixels of a burr hole, was considered to determine if better classification performance could be achieved.

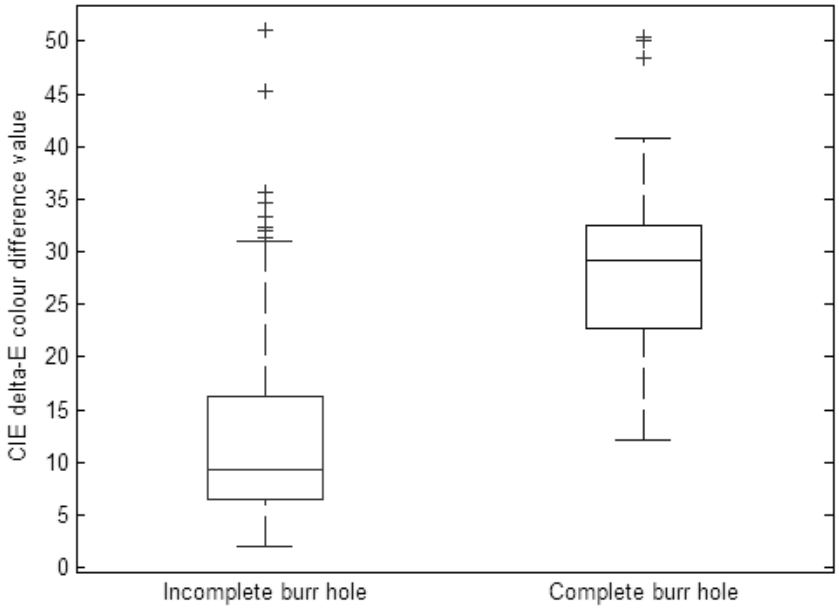


Figure 96: Boxplot of CIE delta-E colour difference (median) for incomplete and complete burr hole

6.5.5 Colour components

To determine if the misclassification error rate of 18.7% and a maximum of 22.2% for burr holes based on colour difference could be improved, the use of three different statistical measures i.e. mean, median, and mode, to represent the colour component values of incomplete and complete burr holes in CIELAB colour space were considered. The use of colour as a feature space was based on the observation that complete burr holes had a different colour to incomplete burr holes.

Because colour can vary greatly depending on illumination, in particular the colour temperature of the illuminant, the white balance setting of the camera was manually calibrated using the CCD video camera's built-in calibration function to account for the 6400K colour temperature of the white LED.

6.5.5.1 Results

Colour components in CIELAB colour space were used to classify the burr holes. Table 8 – Table 11 shows the leave one out misclassification error rates based on colour values of incomplete and complete burr holes in CIELAB colour space for all three classifiers.

Table 8: Naïve Bayes leave one out misclassification error rate (%)

Naïve Bayes leave one out misclassification error rate (%)							
Statistical Measure	Colour space						
	L*	a*	b*	L*a*	L*b*	a*b*	L*a*b*
Mean (average/maximum)	13.8 14.8	35.2 38.9	10.8 13.0	9.7 11.1	3.7 6.5	15.6 17.6	4.5 7.4
Median (average/maximum)	13.3 15.7	36.6 39.8	13.2 15.7	8.1 10.2	4.1 5.6	18.6 20.4	5.6 8.3
Mode (average/maximum)	14.4 18.5	34.8 37.0	19.3 21.3	9.6 12.0	5.9 9.3	24.6 29.6	6.8 9.3

Table 9: LDA leave one out misclassification error rate (%)

LDA leave one out misclassification error rate (%)							
Statistical Measure	Colour space						
	L*	a*	b*	L*a*	L*b*	a*b*	L*a*b*
Mean (average/maximum)	13.8 14.8	35.2 38.9	10.8 13.0	8.9 10.2	3.7 6.5	10.7 13.9	2.3 3.7
Median (average/maximum)	13.3 15.7	36.6 39.8	13.2 15.7	7.0 10.2	4.1 5.6	13.1 15.7	4.1 4.6
Mode (average/maximum)	14.4 18.5	34.8 37.0	19.3 21.3	9.8 13.0	6.1 9.3	19.4 21.3	6.0 9.3

Table 10: QDA leave one out misclassification error rate (%)

QDA leave one out misclassification error rate (%)							
Statistical Measure	Colour space						
	L*	a*	b*	L*a*	L*b*	a*b*	L*a*b*
Mean (average/maximum)	13.9 14.8	35.1 38.0	10.7 13.0	10.0 13.0	4.5 6.5	6.3 8.3	3.0 3.7
Median (average/maximum)	13.2 15.7	35.0 38.0	13.3 15.7	8.5 12.0	5.4 7.4	10.8 13.0	3.8 5.6
Mode (average/maximum)	14.8 18.5	35.8 40.7	20.4 23.1	9.8 13.0	5.5 7.4	19.7 22.2	5.2 7.4

Table 11: LDA leave one out misclassification error rate (%) for each iteration

LDA leave one out misclassification error rate											
Statistical measure	Iteration										Average/ Maximum
Mean	3.7	3.7	1.9	1.9	3.7	2.8	3.7	3.7	1.9	2.8	2.3 3.7

6.5.5.2 Discussion

The best mean misclassification error rate is 2.3% (for 10 iterations) with a maximum error rate of 3.7% based on mean L*a*b* values using LDA. The good classification performance can be attributed to the good separability of the two types of burr hole in CIELAB colour space (see Figure 97 - Figure 99). Figure 97 - Figure 101 shows the separation of the mean pixel values of all incomplete and complete burr holes in CIELAB colour space. The error rate shown in the figures is the percentage of samples that are misclassified, weighted by the prior probabilities for the group. The mean CIE delta-E colour difference, ΔE_{ab^*} between the two classes is 34.4, which is considerably greater than the just noticeable difference $\Delta E_{ab^*} \approx 2.3$, below which the two colours are indistinguishable [154].

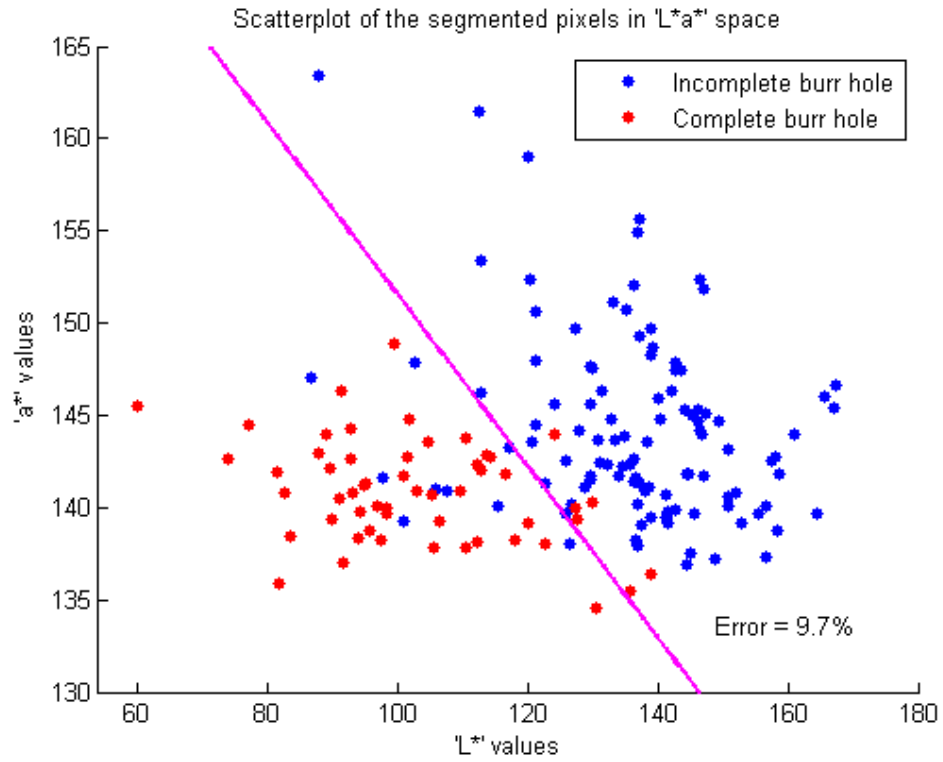


Figure 97: Scatterplot of mean L*a* values of incomplete and complete burr holes in CIELAB L*a* colour space

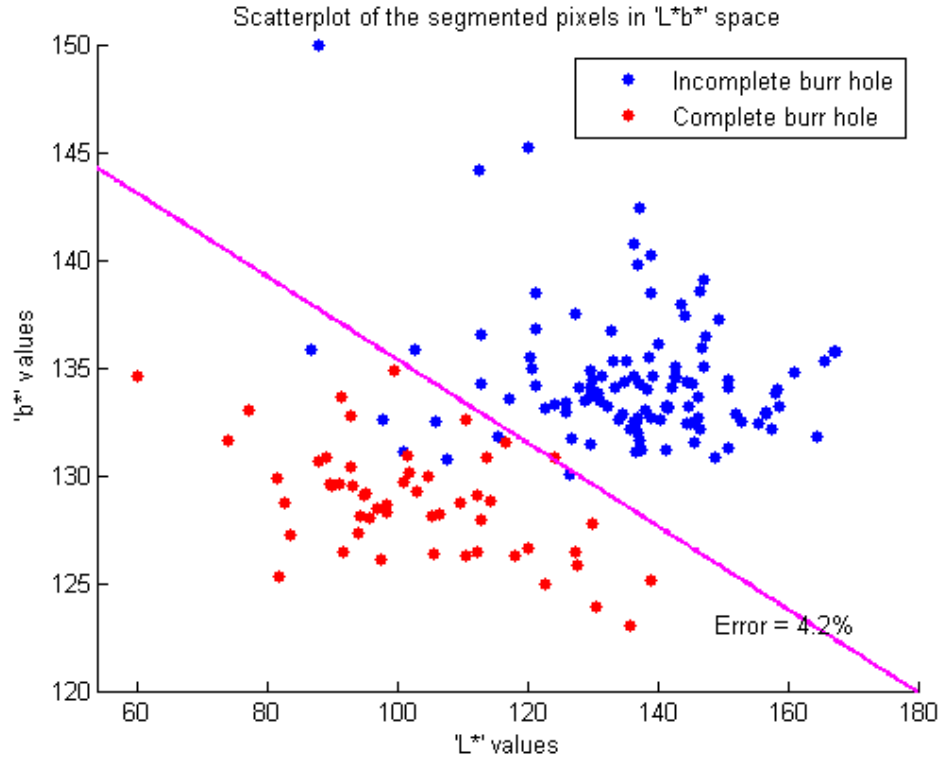


Figure 98: Scatterplot of mean L*b* values of incomplete and complete burr holes in CIELAB L*b* colour space

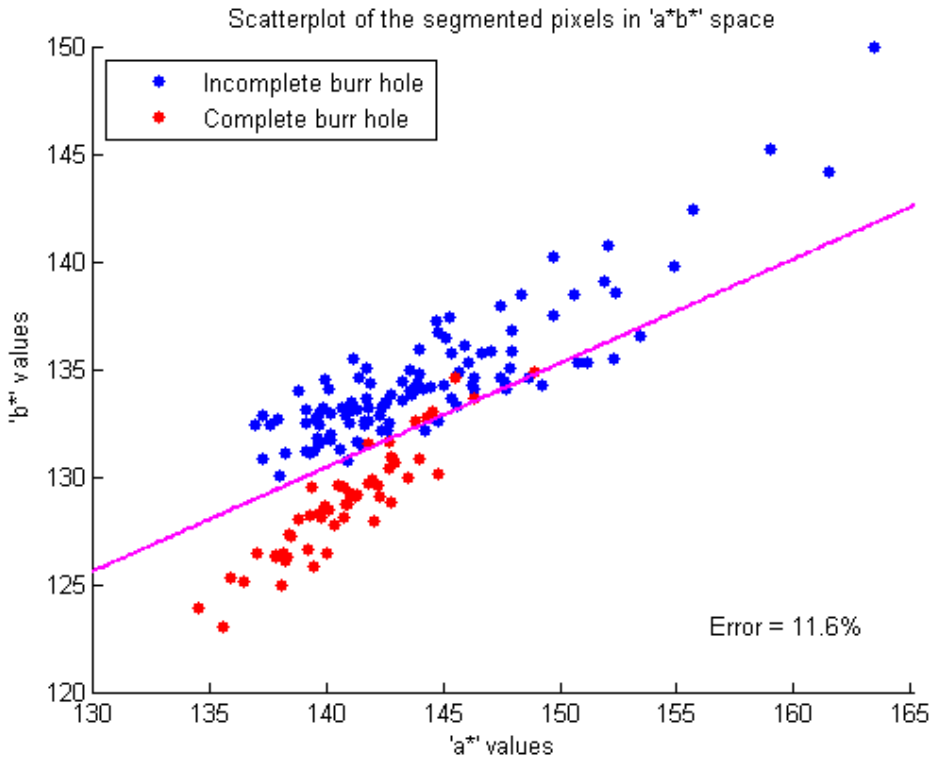


Figure 99: Scatterplot of mean a*b* values of incomplete and complete burr holes in CIELAB a*b* colour space

Figure 100 and Figure 101 show the classification of the burr holes based on the CIELAB L*a*b* feature space with the two classes separated by a plane and quadric surface respectively.

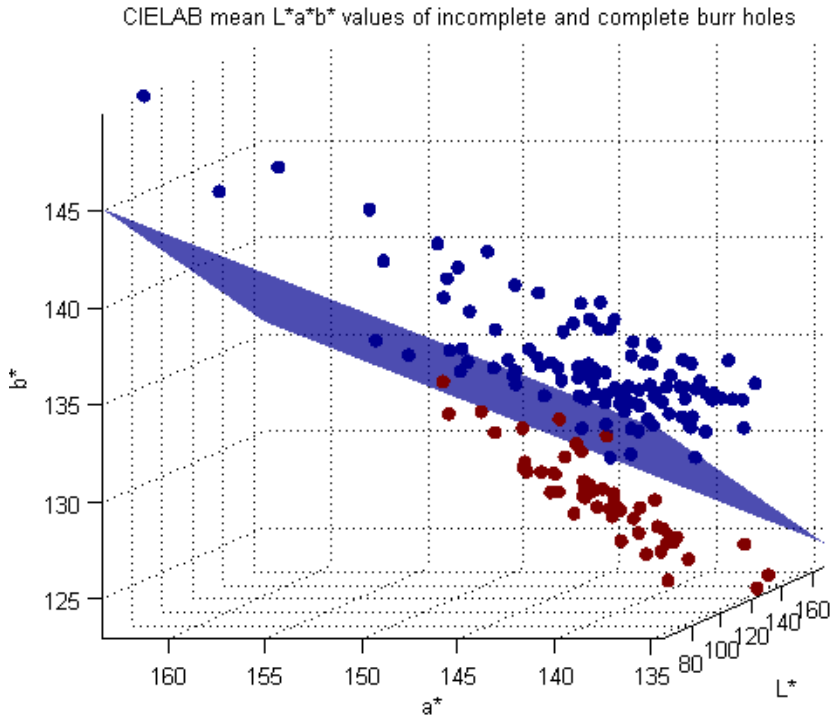


Figure 100: Scatterplot of pixels values (mean) of incomplete and complete burr holes in CIELAB colour space and the LDA decision boundary separating the two classes. The misclassification error rate is 1.9%.

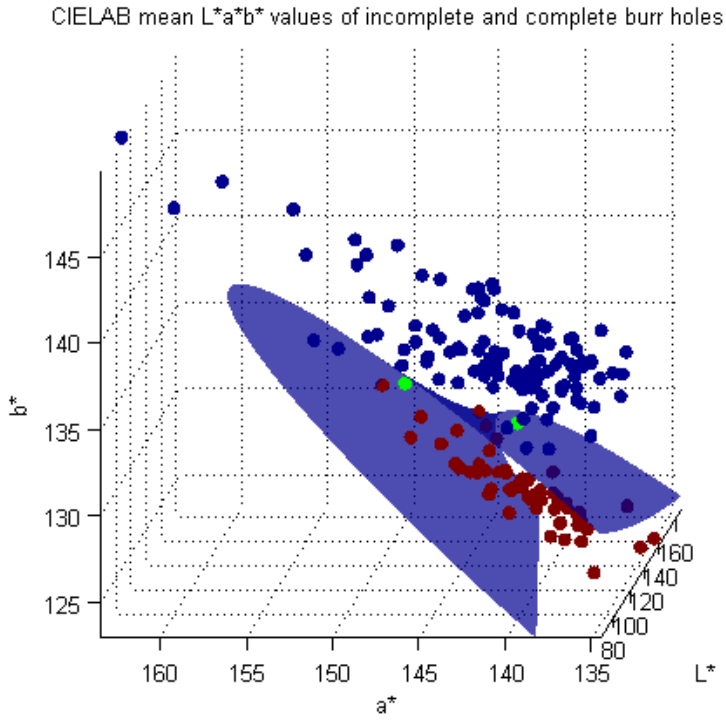


Figure 101: Scatterplot of pixels values (mean) of incomplete and complete burr holes in CIELAB colour space and the QDA decision boundary separating the two classes. Misclassification error rate is 2.8%.

Although the images from the CCD video camera are in sRGB colour space, a device-independent colour model, it is not known if the images acquired are colour-accurate sRGB because the CCD was not calibrated for colour. Nevertheless, although the colour values are not absolute colorimetric values, assuming similar colour calibration for all images, this work has shown that the two types of burr holes can be separated in terms of their colour.

In general, all the classifiers performed equally. However, classification performance e.g. sensitivity and specificity can be changed using different LDA and QDA decision criterion threshold value c and T respectively in Equation 6.8 and Equation 6.9. Geometrically, these threshold values correspond to the location of the decision boundary i.e. the plane and quadratic surface in Figure 100 and Figure 101. Although these decision boundaries give the best misclassification error rate, in this work, a false positive i.e. partial burr holes incorrectly classified as a complete burr hole would result in the subsequent surgical task being performed. A receiver operating characteristic curve could be used to identify threshold values such that there are no false positives. Although this will inevitably reduce the sensitivity i.e. number of true positives correctly identified, of the inspection and increase the amount of false negatives i.e. complete burr holes misclassifies as partial burr hole, a higher overall misclassification rate without any false positives is preferable to a lower misclassification error rate with false positives.

6.6 Texture

The use of textural information to improve further classification performance was considered. Textural information may prove to be useful feature for discriminating the two types of burr holes as incomplete and complete burr hole are predominantly smooth and rough respectively. Four of Haralick's coefficients [155] for a grey-level co-occurrence matrix (GLCM), contrast, correlation, energy and homogeneity, were computed and used as features for classification. Each element (i, j) in the GLCM corresponds to the frequency pairs of adjacent pixels (for a given spatial relationship) have grey level intensity values i and j respectively. A k x k x 8 GLCM was used (k =

8,16,32,64,128 and 256), with a distance of one pixel in each direction for a pixel of interest and its corresponding pair/neighbour.

Contrast (also known as variance) is a measure of intensity contrast between pixel pairs over the entire image and is given by Equation 6.11. It is zero for a constant image.

$$Contrast = \sum_{i,j} |i-j|^2 p(i,j) \tag{6.11}$$

Correlation is a measure of how correlated a pixel pair are over the entire image and is given by Equation 6.12.

$$Correlation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \tag{6.12}$$

Energy is a measure of uniformity and is given by the sum of squared elements in the GLCM over the entire image, as shown in Equation 6.13. Energy ranges from 0 to 1 and is 1 for a constant image.

$$Energy = \sum_{i,j} p(i,j)^2 \tag{6.13}$$

.

Homogeneity is a measure of the closeness of the distribution of elements in the GLCM to the GLCM diagonal over the entire image and is given by Equation 6.14.

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{6.14}$$

6.6.1 Results

Table 12 shows the leave one out misclassification error rates for all three classifiers respectively, based on all the possible permutations of the four measures of texture of incomplete and complete burr holes for a 256 x 256 x 8 GLCM. The best maximum misclassification error rate ranges from 13.0 – 13.9% with the average misclassification error rate ranging from 9.7 – 12.0%.

Table 12: Mean and maximum misclassification error rate (%) based on permutations of the four different texture measures of incomplete and complete burr holes using a Naïve Bayes classifier, LDA and QDA. (1 = Contrast, 2 = Correlation, 3 = Energy, 4 = Homogeneity) for a 256 x 256 x 8 GLCM

GLCM properties	Classifier		
	Naïve Bayes	LDA	QDA
1	39.4 41.7	39.4 41.7	39.7 42.6
2	22.8/ 25.0	22.8 25.0	23.0 25.9
3	43.9 48.1	43.9 48.1	42.4 45.4
4	40.1 46.3	40.1 46.3	37.9 44.4
1,2	20.6 22.2	22.1 25.0	21.6 23.1
1,3	42.1 45.4	38.0 40.7	30.7 36.1
1,4	22.3 28.7	18.3 20.4	19.0 21.3
2,3	24.0 25.9	21.8 24.1	19.4 25.0
2,4	19.8 23.1	19.9 23.1	18.2 20.4
3,4	26.6 32.4	16.3 21.3	17.0 25.9
1,2,3	21.8 24.1	15.0 17.6	15.6 18.5
1,2,4	17.5 21.3	10.3 13.9	11.9 16.7
1,3,4	25.6 32.4	11.9 14.8	13.2 16.7
2,3,4	20.5 22.2	15.7 19.4	15.1 21.3
1,2,3,4	19.1 22.2	9.7 13.0	10.5 15.7

6.6.2 Discussion

Classification was performed using all permutations of the different texture measures. The lowest leave one out misclassification error rate was obtained using LDA based on all four texture measures i.e. contrast, correlation, energy and homogeneity, with a mean misclassification error rate of 12.0% and a maximum error rate of 13.9%. This misclassification error rate is considerably higher than the mean misclassification error rate of 2.3% and a maximum error rate of 3.7% achieved using colour information alone.

Figure 102 is the bivariate scatterplots between the four texture measures that shows the separation of the two classes between pairs of texture measures. There is a considerable amount of overlap, although separation in four dimensions would be better, based on the classification results.

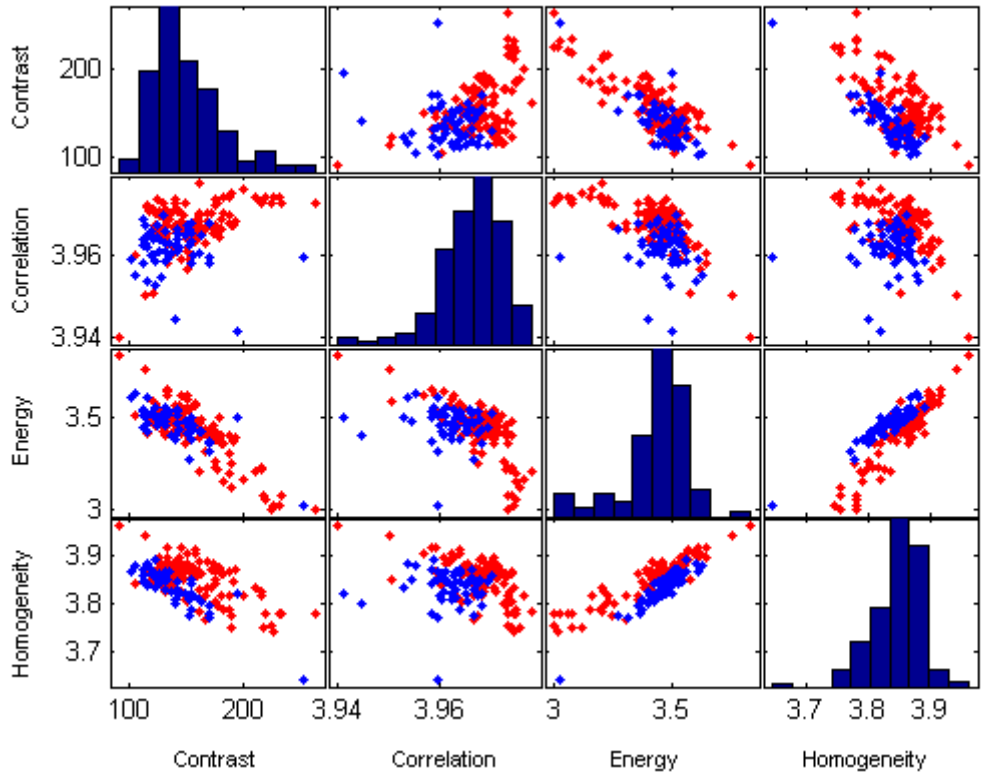


Figure 102: Scatterplot matrices of all the bivariate scatterplots between the four texture measures, along with a univariate histogram for each texture measure. Incomplete and complete burr holes are in red and blue respectively.

Although texture is a weaker feature in discriminating the two classes of burr holes compared to colour, the combined use of texture and colour was considered. Table 13 shows the LDA leave one out misclassification error rates based on colour and texture properties of incomplete and complete burr holes.

Table 13: LDA leave one out misclassification error rate (%) based on mean colour values in CIELAB colour space and texture properties of incomplete and complete burr holes

LDA leave one out misclassification error rate (%)											
Feature Space	Iteration										Average/ Maximum
Colour	3.7	3.7	1.9	1.9	3.7	2.8	3.7	3.7	1.9	2.8	2.3 3.7
Colour + Texture	3.7	2.8	2.8	3.7	3.7	0.9	2.8	2.8	2.8	3.7	3.0 3.7

6.7 Discussion and conclusions

The average misclassification error rate based on combined colour and texture information is 3.0% with a maximum misclassification error rate of 3.7%. On the other hand, the lowest average and maximum misclassification error rate are 2.3% and 3.7% respectively, using colour information only. Combining textural information with colour information did not change classification performance in terms of the maximum misclassification error rate. Further testing with a greater number of sample images is needed to determine if the use of texture measures will aid in classification performance. Figure 103 shows the images that the classifier could not discriminate based on colour and texture alone. All misclassified burr holes were false positives i.e. partial burr holes incorrectly classified as a complete burr hole.

Figure 103: Misclassified burr hole based on mean CIELAB pixel value and texture properties.

In terms of the best feature to represent a burr hole, the mean pixel colour value (in CIELAB colour space) of all its pixels gave the best misclassification error rate with a better classification performance compared to using colour difference. In terms of classifier performance, the misclassification error rates were comparable for naïve Bayes classifier, LDA and QDA, with LDA performing slightly better, suggesting that the feature space is linearly separable. Although different classifiers such as neural networks might give better classification results, because of the relatively small dataset, the emphasis was on finding appropriate features to perform the classification rather than strength of the classifiers. Furthermore, neural networks could overfit the data and may not generalise well to a different set of data.

Although this work has shown promising results on the use of computer vision for inspection of the outcome of the three surgical tasks, there are several challenges regarding the use of computer vision for this purpose. The first is the inherent natural variation in appearance of biological matter. A case in point is the inspection of incisions based on the presence of an incision line. Because of the large variations in appearance of the incision line, the reliance on a single feature was unable to capture all the differences between a complete and incomplete incision. The use of computer vision should therefore not be viewed in isolation. Indeed, the aim is to use computer vision to

augment/complement other sensory information, although it is not inconceivable that the state of the art in computer vision can advance to a stage where it is able to emulate the visual recognition ability and interpretation process of a human expert. The second is misclassification. False positives in the detection are more of an issue than false negatives. Erroneously classifying an incomplete surgical task as complete will result in the subsequent task being performed. When performing classification, the emphasis should therefore be on reducing the number of false positives, rather that achieving the best overall misclassification rate.

This chapter described a vision-based approach to inspect three primary surgical tasks for the creation of a burr-hole, a common neurosurgical procedure, as a step towards performing this procedure robotically. To validate the approach, experiments were performed on cadaveric pig heads, to simulate different fault scenarios, and to assess completion by using visual cues from the scene. While a burr hole is a relatively simple procedure performed by a neurosurgeon, it is a very complex problem to perform robotically. A subset of this problem is to determine the successful completion of a given surgical action. Towards this end, the use visual cues from the operative field were investigated.

The use of computer vision to inspect surgical tasks has never been considered before. Encouraging results were obtained for the three surgical tasks under consideration, which has been demonstrated by experiments on cadavers. For example, the classification rate for incisions was 85.9% while classification rate for burr holes was 97.7%. The low burr hole misclassification rates can be attributed to incomplete and complete burr holes having characteristic colours, and these being consistent through a wide range of subjects. Classification simply becomes a matter of discriminating the two classes based on their characteristic colours.

As the use of images from actual surgeries was beyond present resources, cadaveric dissection of pig heads were used to generate images of the outcome of a surgical action. Similar techniques may not necessarily extend to humans in a clinical environment. As such, clinical testing under more realistic conditions are necessary before this system can be deployed in an

operation theatre. Nevertheless, these findings are encouraging and suggestive for the potential use of computer vision for a more comprehensive approach in robotic surgery.

Chapter 7

Conclusion and Future Work

7.1 Conclusions

In this thesis, the concept of MISEN, a robot-assisted surgery system for the provision of emergency neurosurgical procedures, was presented. MISEN was proposed as an alternative to conventional surgery, as an engineering solution to what is essentially a medical problem. It is designed to perform certain routine but lifesaving neurosurgical procedures such as the evacuation of a haematoma and the drainage of cerebrospinal fluid from the brain ventricles. MISEN will enable these procedures to be performed at hospitals without on-site neurosurgery, where the majority of patients with head injury are sent. This would eliminate the need for a secondary transfer and avoid the problems associated with it. Crucially, it will reduce time to surgery, an important consideration where immediate intervention would significantly improve a patient's outcome. Currently, no similar robotic surgery system exists.

Preoperative planning and intraoperative intervention are two main phases of the surgical protocol for MISEN, as well as for robotic surgery in general. The aim of this thesis was to contribute in these two areas. Important components of both these phases are image to patient registration and intraoperative assessment respectively. Computational approaches were developed to perform these two aspects of the surgery that would ordinarily be performed by a neurosurgeon, to afford greater autonomy to MISEN. Specifically, a computational approach for automating the image to patient registration and computer vision techniques for intraoperative assessment, were developed. In this concluding chapter, these techniques are discussed and areas for future work proposed.

7.1.1 Image to patient registration

An image to patient registration framework for MISEN was proposed, based on a rigid body transformation of homologous anatomical soft tissue landmarks of the head, the medial canthus and tragus, in CT and patient space. The registration framework meets the required clinical accuracy of the targeted procedures, is cost effective, and practical.

In terms of accuracy, the errors in localising a target anywhere in a head, obtained experimentally and estimated based on an analytical expression, was shown to be within the millimetre range (if the landmarks can be localised to within 5 mm of their true position). This accuracy compares favourably to that achieved by a neurosurgeon performing the targeted neurosurgical procedures using a freehand technique, which is sometimes in the centimetre range. Errors due to patient movement and robot calibration notwithstanding, the registration framework can therefore yield a sufficiently accurate registration for the required clinical accuracy of the targeted procedures.

The registration framework is cost effective because it does not require the use of specialised hardware, such as projectors to provide structured lighting or the use of 3D laser scanners, relying instead on comparatively inexpensive optical cameras. It is practical as it is a retrospective technique, avoiding the requirement for a secondary scan. In contrast, the use of fiducial markers, which is the method of choice for image to patient registration, is a prospective technique that requires a secondary scan.

Developing an inherently safe system is the design philosophy behind the MISEN system. As the registration framework is an important aspect of the system, failing which the robot may not be able to locate a point in space, it was deemed essential that the registration be automated, removing the human element and the potential for a mistake, especially if automatic localisation can be shown to be reliable. In fact, one of the motivations behind automating the registration was to improve the safety aspect of the registration. Although the landmarks may be localised manually, a conscious decision was made to relegate the human's role to one of validating the localisation of the landmarks. By reducing human involvement in the registration to one of validation, a check and balance is created, as it is highly unlikely that both a human and computer will commit the same error. Manual localisation is also subjective, as it is a function of the examiner's experience and perception.

As a step towards automating the registration, a computational approach to localise automatically these landmarks in CT space was developed. This approach lends itself well to CT images, which are digital in nature. More importantly, the technique does not require changes to the existing workflow in terms of CT image acquisition. Although the localisation of anatomical

landmarks such as the medial canthus and tragus is well researched in 2D intensity images, its localisation in 3D images is less well researched. To the author's knowledge, no work on the automatic localisation of the medial canthus and tragus in CT images have been reported, and the work on this is a novel contribution, with applications to image to patient registration.

The medial canthus and tragus were localised automatically on surface models extracted from CT images based on their curvature saliency and using a rule based system that incorporates prior knowledge of their characteristics. with performance comparable to manual localisation. The maximum difference in positions of landmarks localised manually and automatically of 4.5 mm for the medial canthus and 2.6 mm for the tragus is within the 5 mm FLE for a TRE of within 5 mm, and compares well to a maximum variability of 2.9 mm and 2.5 mm respectively in localising these landmarks manually.

An important consideration for any registration technique is the speed at which the registration can be performed. The advantage of a landmark based approached compared to retrospective techniques based on mutual information and photo consistency is that it is an inherently faster technique, as the localisation of landmarks in CT space can be performed offline preoperatively, reducing the amount of time required for the registration. In contrast, registration based on mutual information and photo consistency is time consuming, as the solution (optimal) requires a computationally intensive iterative approach. More importantly, for a landmark based approach, the time for reregistration (required if there is patient movement), is only a function of the time taken to localise the landmarks in patient space intraoperatively. This is in contrast to the time taken in finding a new optimal solution for registration based on mutual information and photo consistency. A lengthy reregistration process would negate any potential benefits afforded by automating the registration.

Another advantage in terms of the practicality of using anatomical landmarks as a registration basis is that it is an intuitive technique and can therefore be easily understood. Errors, when they occur, can be corrected, even by someone without specialised neurosurgical experience or engineering expertise. For example, if the algorithm has erroneously localised the medial canthus or tragus, this will be immediately obvious to the person monitoring the

registration, assuming that the localised landmarks are overlaid on the surface rendered CT images and optical images of the patient, and can be easily corrected. On the other hand, errors from image to patient registration techniques based on mutual information and photo consistency (these methods work with optical images as well) are not apparent i.e. they cannot be observed directly, and therefore cannot be corrected when mistakes occur in the registration. Furthermore, these techniques have not been used clinically, in contrast to anatomical landmarks, which is a well-established registration basis in image to patient registration.

The medial canthus and tragus were localised in near isotropic high resolution CT images. As these are point landmarks, the localisation of these landmarks is not feasible in lower resolution CT images, as they would not be well defined, making their localisation impossible. Advances in CT imaging technology such as isotropic or near isotropic sub-millimetre image resolution mean that CT images could now achieve a level of detail that has made possible the detection of point landmarks that would be unfeasible in low-resolution images. These high resolution CT images can be reconstructed retrospectively from existing raw data, without re-scanning the patient, although this is limited by the detector configuration used during the scan. If the CT scanning protocol for head injury/trauma is used, the image resolution of routine CT scans is sufficient to localise these landmarks automatically.

In conclusion, the proposed registration framework is perfectly suited for the MISEN system as it meets all the stated requirements, the most important one being the ability to perform the registration retrospectively. The ability to perform a registration retrospectively is important where immediate surgical intervention would improve a patient's prognosis, as a secondary scan may not always be possible and will inevitably delay time to surgery. This registration framework may also be used in other robotic neurosurgery systems where the emphasis is on being able to perform the registration retrospectively, rather than performing it to a high degree of accuracy. This registration framework is therefore particularly suitable for emergency neurosurgical procedures that are ordinarily performed using a freehand technique, such as the targeted procedures of the MISEN system. However, this would exclude certain neurosurgical procedures such as deep brain stimulation and biopsies.

7.1.2 Intraoperative assessment

In this thesis, the concept of an intraoperative supervisory system was conceived. When developing a robotic surgery system such as MISEN, the design of an inherently safe system is the responsibility of the system developer. In the context of MISEN, an intraoperative supervisory system that is capable of monitoring the surgical intervention by the robot was deemed essential, as MISEN would be performing actions that would otherwise be carried out by a neurosurgeon. This is especially true considering that MISEN is intended for use in hospitals without on-site neurosurgery, by a surgical team with little or no neurosurgical experience and engineering expertise.

An important component of the supervisory system is the ability to assess the outcome of a surgical action, which is an important aspect of the surgical intervention. The use of computer vision was proposed to emulate a surgeon's visual sensory and decision-making capability when assessing task completion. Specifically, a computer vision based inspection of the outcome of a surgical action based on visual cues was investigated. The aim was to incorporate computer vision based inspection into the MISEN system, to improve safety as well as to facilitate the surgical intervention by the robot. The use of vision is particularly well suited for surgery where sterility is an important consideration, as it is a non-contact form of inspection i.e. no part of the vision system will be in direct contact with the patient.

In the context of the MISEN system, the introduction of an intraoperative supervisory system will not require a change in its architecture, as it is already equipped with photogrammetry cameras used for image to patient registration. These same cameras may be used for the visual inspection of the surgical intervention. Unlike conventional surgery, which can be highly unstructured, the surgical actions and their context are known in robotic surgery, making their visual inspection feasible. An approach similar to the use of machine vision for inspection tasks was adopted for the visual inspection of the outcome of a surgical action. Images of the outcomes were acquired and representative features based on visual cues that best differentiate the two possible outcomes were identified. Computer vision techniques were subsequently developed to detect these features.

The use of computer vision for the inspection of the outcome of surgical tasks, while somewhat unconventional, has its advantages over human experts. For example, computer vision is able to quantify differences in colour, a subjective measure that is usually described with adjectives. Computer vision is also able to resolve small differences in colour, while humans can only differentiate colours with a CIE delta-E colour difference, ΔE_{ab} greater than 2.3 (unless these colours are placed adjacent to each other). Therefore, in the visual inspection of a burr hole for instance, computer vision is able to detect differences in colour that might not be apparent to the human observer. This capability is probably best exemplified with the over 97% classification rate achieved when computer vision was used to classify incomplete and complete burr hole based on colour

A difficulty when dealing with biological specimens was inherent natural variation. The variation in biological specimens can be large, for example the variation in terms of its biomechanical properties, which explains why features based on shape alone are not discriminating enough to differentiate incision types. Nevertheless, it was still possible to distinguish between superficial and non-superficial incisions for example, using image differencing. From a visual inspection point of view, one difficulty encountered is that features used for classification may not always be apparent. For example, in determining the shape properties of an incision, the sometimes low contrast of the incision in an image made its segmentation difficult, although the use of directional ring lighting facilitated its segmentation. In the visual inspection of an incision, the presence of a dark line that is characteristic of a complete incision is sometimes not obvious. Nevertheless, the use of this feature was able to give reasonably good misclassification error rates, and crucially without any false positives.

The use of computer vision techniques for the visual inspection of the outcome of surgical actions has shown promising results, suggestive of its potential use in an operating theatre. Although only low-level image features were used, it was possible to distinguish the two possible outcomes, complete and incomplete, for the surgical actions under consideration, which have been demonstrated by experiments on cadaveric pig heads. In particular, the visual inspection of a burr hole, to distinguish between incomplete and complete burr holes, had a detection rate of over 97%. The visual inspection of a burr hole is

therefore sufficiently accurate to be potentially used in an automated inspection system.

For the assessment of the retraction of an incision, the technique proposed, based on determining the maximum circular free space within the retracted incision by fitting the largest possible circle inside a region bounded by the edges of the incision and the prongs of the retractor, worked well in the images used in this study. It could potentially give better results if used in conjunction with directional lighting to facilitate edge detection. Although similar results may not necessarily extend to humans in a clinical environment, it is not inconceivable that the techniques developed can be applied in a clinical setting. The current work has laid the groundwork and foundation for the future development of these techniques.

The concept of an intraoperative supervisory system with applications to robotic surgery was conceived and a niche where this system may be used for MISEN was identified. To the author's knowledge, the use of vision in this context has never been considered before, and is a novel contribution, representing the first application of computer vision in this area. The use of computer vision however, should not be viewed in isolation, as it is intended to augment/complement these other sensory sources for a more comprehensive assessment of a surgical action. This system has the potential to afford robotic surgery systems greater autonomy and a greater level of safety. The use of a visual sensor will also improve the reliability and robustness of these systems, which traditionally only employs force, torque and positional sensors, by introducing redundancy in its sensor space.

The science in automated inspection has advanced to a stage where its application in the operating theatre is no longer a distant reality. The future of the proposed concept of an intraoperative supervisory system is exciting, has tremendous potential and holds great promise. Intraoperative supervisory systems and similar systems, for greater situational awareness and confidence, will probably be increasingly developed and applied to robotic surgery systems in the near future, as it has the potential improve current practice in these systems.

7.2 Recommendation for future work

In the next section, the future directions for the two strands of research in this thesis are suggested, to further their development for eventual use in the MISEN system. While the concepts and techniques suggested have been developed with MISEN in mind, they may extend to other areas such as the earlier example on the use of the intraoperative assessment system for surgical training.

For the automatic localisation of anatomical soft tissue landmarks in CT images, a potential application is in the processing of large databases of CT images such as craniofacial databases. These databases provide a standard craniofacial measurement of the normal population and stores craniofacial information such as standard anthropomorphic and anatomic measurement, typically merging data from a variety of image acquisition modalities such as close range photogrammetry, 3D close range laser scanning, and CT. A typical patient model would incorporate data from all these modalities to provide a comprehensive craniofacial model. To merge CT data with close range photogrammetry data for example, an expert would have to identify surface analogues of bone landmarks on CT and corresponding points on close range photogrammetry images for rigid point based registration. Automation of this annotation step would facilitate the registration of these two modalities.

7.2.1 Image to patient registration

While the medial canthus and tragus are the most geometrically salient among the commonly used anatomical landmarks of the head, the use of other landmarks in addition to these should be considered, as registration accuracy is a function of the number of landmarks used. However, as registration accuracy does not improve considerably after five or six fiducials, an additional two landmarks should be considered. Potential landmarks that may be used include the lateral canthus and the soft tissue nasion. However, the problems associated with the use of these landmarks for registration, such as deep-set wrinkles masking the true location of the lateral canthus and the fact that the

soft tissue nasion can usually only be found in patient space by palpation, have yet to be resolved.

A limitation of the proposed registration framework is that the required anatomical landmarks i.e. medial canthus and tragus may be unreliable or absent, as would be the case in the occurrence of facial/head trauma that distorts a patient's normal anatomy. In this circumstance, the registration may be performed using skin-affixed markers, although this would mean that the patient would be required to undergo a secondary scan. Future work in this area therefore would have to focus on an alternative image to patient registration framework that is not based on anatomical landmarks. Techniques based on mutual information and photo consistency are the most promising, and research should be focused on a means of validating the registration.

On registration accuracy, while the user can validate and correct for an error in the localisation of the medial canthus and tragus prior to the surgical intervention, the target registration error, which is a more significant measure, is unknown. Future work should look into a means of determining the expected target errors.

7.2.2 Intraoperative assessment

In this thesis, the visual inspection of the outcome of surgical action was based on the use of pig head cadavers, specifically images of the outcome of a surgical action from cadaveric dissection of pigs head. Although pigs are anatomically similar to humans in many respects, certain important aspects of the surgery could not be simulated. For example, any bleeding resulting from the skin being incised could not be simulated, as cadaveric skin has no blood supply. Therefore, although cadaveric dissection is the most realistic alternative to live surgery, the limitation in the use of cadavers is that it may not be representative of the conditions in actual surgery. In this case, the techniques developed here may not work in a clinical setting. Clinical trials or testing under conditions that are more realistic is therefore an obvious area of future work and essential before this system can be deployed in an operation theatre.

Although images from live surgery would be the most realistic, unlike cadaver testing where an incomplete outcome may be performed, ethical and

practical considerations would prevent the same from being performed on humans. Alternatively, human cadavers, which would be more representative, could be used. However, this would mean that potentially important physiological processes such as bleeding could not be simulated as well. The use of live animals may be another possible avenue for testing, if ethical approval can be obtained.

Another possible area of future work is to use temporal information from video images. In this work, only a snapshot view i.e. still image of the outcome was used. The use of temporal information has the potential to stop a surgical action being performed if it is deemed unsafe. In the current work, a surgical action would have to be already performed prior to visual inspection, with the goal being to validate task completion only. The subset of surgical tasks considered could also be expanded to include other tasks that are particularly suited for visual inspection.

In this work, image scale was assumed to be known. However, this would not be the case in practice. The use of the prongs of the retractors has been suggested as a possible means to determine image scale. However, this is only applicable during the retraction of an incision, where the retractors are present in the image. When assessing the length of an incision, the retractors are not available in the image and an alternative method to determine image scale must be found. A possible solution is to introduce a calibration object for the purposes of determining scale in the image. The solution is less than ideal, as it would require a change in the process workflow. Another possible solution is to track instrument movement and infer image scale based on cross validating instrument travel with positional information obtained from encoders.

Finally, the potential for using a similar system in the context of surgical training for teaching trainee surgeon was mentioned. The use of such a system has the potential to replace a trained surgeon, for assessing the skills of a trainee. This was based on the assumption that the outcome of a surgical action can be used as an indication of surgical skill. An intraoperative system such as the one proposed in this thesis could be a viable alternative to the subjective assessment of surgical skills by a surgeon, or where a skilled surgeon may not be available. For assessing surgical skill, the inspection criteria has to be defined explicitly and features representative of the different

skills be identified. More demanding and complex surgical tasks would require expert knowledge to identify these features as well as the basis for assessment, as they may not always be known explicitly.

References

[1] H. C. Patel *et al.*, "Trends in Head Injury Outcome from 1989 to 2003 and the Effect of Neurosurgical Care: An Observational Study," *Lancet*, vol. 366, pp. 1538-1544, 2005.

[2] G. Findlay *et al.*, "Trauma: Who Cares? A Report of the National Confidential Enquiry Into Patient Outcome And Death," National Confidential Enquiry into Patient Outcome and Death, London, 2007.

[3] M. J. Nelson, "British Neurosurgical Workforce Plan 2000-2015: A Report on the Staffing Requirements of Neurosurgical Units in the British Isles," The Society of British Neurological Surgeons, London, 2000.

[4] K. Ashkan *et al.*, "Crisis in Resources: A Neurosurgical Prospective," *Br. J. Neurosurg.*, vol. 15, pp. 342-346, 2001.

[5] H. Cooper *et al.*, "Deaths in Acute Hospitals: Caring to the End?," National Confidential Enquiry into Patient Outcome and Death (NCEPOD), London, 2009.

[6] P. Barnes *et al.*, "Neurocritical Care Capacity and Demand," The Neurocritical Care Stakeholder Group, London, 2010

[7] B. Davies, "A Review of Robotics in Surgery," *P. I. Mech. Eng. H.*, vol. 214, pp. 129-140, 2000.

[8] P. A. Farling *et al.*, "*Recommendations for the Safe Transfer of Patients with Brain Injury*," The Association of Anaesthetists of Great Britain and Ireland, London, 2006.

[9] J. Browne *et al*., "High Quality Acute Care for the Severely Injured is not Consistently Available in England, Wales and Northern Ireland: Report of a Survey by the Trauma Committee, The Royal College of Surgeons of England," *Ann. R. Coll. Surg. Engl*., vol.88, no.2, pp. 103-107, Mar. 2006.

[10] I. J. Swann and A. Walker, "Who Cares for the Patient with Head Injury Now?" *Emerg. Med. J.,* vol. 18, pp. 352-357, 2001.

[11] J. M. Seelig *et al.*, "Traumatic Acute Subdural Hematoma: Major Mortality Reduction in Comatose Patients Treated Within Four Hours," *N. Engl. J. Med.*, vol. 304, pp. 1511-1518, 1981.

[12] A. D. Mendelow *et al.*, "Extradural Haematoma: Effect of Delayed Treatment," *Br. Med. J.*, vol. 1, pp. 1240-1242, May 1979.

[13] R. C. Tasker *et al.*, "Severe Head Injury in Children: Emergency Access to Neurosurgery in The United Kingdom," *Br. Med. J.*, vol. 23, pp. 519-522, 2006.
[14] P. Leach *et al.*, "Transfer Times for Patients with Extradural and Subdural Haematomas to Neurosurgery in Greater Manchester," *Br. J. Neurosurg.*, vol. 21, pp. 11-15, 2007.
[15] S. J. Price *et al.*, "Have ATLS and National Transfer Guidelines Improved the Quality of Resuscitation and Transfer of Head-Injured Patients?: A Prospective Survey from a Regional Neurosurgical Unit," *Injury*, vol. 34, pp. 834-838, 2003.
[16] Modernisation Agency/Department of Health, "The Neuroscience Critical Care Report," Department of Health, Leicester, United Kingdom, Aug. 2004.
[17] H. Marsh *et al.*, "Closed Head Injuries: Where Does Delay Occur in the Process of Transfer to Neurosurgical Care?" *Br. J. Neurosurg.*, vol. 3, pp. 13-19, 1989.
[18] S. Whiteley *et al*., "Guidelines for the Transport of the Critically Ill Adult," Intensive Care Society, London, 2002.
[19] L. T. Dunn, "Secondary Insults During the Interhospital Transfer of Head-Injured Patients: An Audit Of Transfers in the Mersey Region," *Injury*, vol. 28, pp. 427-431, 1997.
[20] P. R. Knowles *et al.*, "Meeting The Standards for Interhospital Transfer of Adults with Severe Head Injury in the United Kingdom," *Anaesthesia*, vol. 54, pp. 280-283, 1999.
[21] D. Gentleman and B. Jennett, "Hazards of Inter-Hospital Transfer of Comatose Head-Injured Patients," *Lancet*, vol. 318, pp. 853-855, 1981.
[22] S. D. Carter, "Review Of Neurosurgical Services in Scotland," The Scottish Executive, Edinburgh, 2000.
[23] G. R. Foster and S. L. Gibson, "Neurosurgery Needs Assessment," Neurosciences Implementation Group, NHS Health Scotland, Glasgow, 2007
[24] Brain & Spine Foundation, (2011, Jan. 6). *Neurological Centres* [Online] Available:http://www.brainandspine.org.uk/information/neurological_centres/index.html 2005.
[25] Royal College of Surgeons of England, "Report of the Working Party on the Management of Patients with Head Injuries," The Royal College of Surgeons of England - Council, London, 1999.

[26] D. W. Crimmins and J. D. Palmer, "Snapshot View of Emergency Neurosurgical Head Injury Care in Great Britain and Ireland," *J. Neurol. Neurosurg. Psychiatry.*, vol. 68, pp. 8-13, 2000.

[27] K. Ashkan *et al.*, "Crisis in Resources: A Neurosurgical Prospective," *Br. J. Neurosurg.*, vol. 15, pp. 342-346, Aug. 2001.

[28] The Information Centre, "NHS Hospital and Community Health Services: Medical and Dental Workforce Census England: 30 September 2007 - Detailed results," Government Statistical Service, UK, 2007

[29] National Statistics Online. (2010, Jun. 24). *Population Estimates* [Online] Available: http://www.statistics.gov.uk

[30] M. J. Nelson, "Safe Neurosurgery 2000", The Society of British Neurological Surgeons. London, 1999

[31] C. Dew, "Hospital Episode Statistics, Primary Diagnosis, 2006-07," The NHS Information Centre, United Kingdom, 2008.

[32] F. E. Lecky *et al.*, "Lack Of Change in Trauma Care in England and Wales Since 1994," *Emerg. Med. J.*, vol. 19, pp. 520-523, 2002.

[33] National Collaborating Centre for Acute Care, "Head Injury: Triage, Assessment, Investigation and Early Management of Head Injury in Infants, Children and Adults (NICE Clinical Guideline 56) - Quick Reference Guide," National Institute for Health and Clinical Excellence, London, 2007.

[34] London Severe Injuries Working Group, "Modernising Major Trauma Services in London," London Severe Injuries Working Group, London, 2001.

[35] D. Yates, "Head Injury: Triage, Assessment, Investigation and Early Management of Head Injury in Infants, Children and Adults: Methods, Evidence & Guidance," National Collaborating Centre for Acute Care at The Royal College of Surgeons of England, London, 2007.

[36] D. Yates, "Head Injury: Triage, Assessment, Investigation and Early Management of Head Injury in Infants, Children and Adults: Methods, Evidence & Guidance," National Collaborating Centre for Acute Care at The Royal College of Surgeons of England, London, 2007.

[37] H. Barratt *et al.*, "The Implications of the NICE Guidelines on Neurosurgical Management for All Severe Head Injuries: Systematic Review," *Emerg. Med. J.*, vol. 27, pp. 173-178, 2010.

[38] Scottish Intercollegiate Guidelines Network, "Early Management of Patients with a Head Injury: A National Clinical Guideline," Scottish Intercollegiate Guidelines Network, Edinburgh, 2009.

[39] D. M. Herron and M. Marohn, "A Consensus Document on Robotic Surgery," *Surg. Endosc.*, vol. 22, pp. 313-325, 2008.

[40] N. Nathoo *et al.*, "In Touch with Robotics: Neurosurgery for the Future." *Neurosurgery*, vol. 56, pp. 421-433, 2005.

[41] Y. Kwoh *et al.*, "A Robot with Improved Absolute Positioning Accuracy for CT Guided Stereotactic Brain Surgery," *IEEE T. Bio-Med. Eng*., vol. 35, pp. 153-160, 1988.

[42] C. S. Karas and E. A. Chiocca, "Neurosurgical Robotics: A Review of Brain and Spine Applications," *J. Robotic Surg.*, vol. 1, pp. 39-43, 2007.

[43] M. S. Eljamel, "Robotic Application in Epilepsy Surgery," *Int. J. Med. Robot.*, vol. 2, pp. 233-237, 2006.

[44] A. Golash *et al.*, "3-D Error Measurement for Checking the Application Accuracy of a Stereotactic Robotic System with an Infrared Space Digitisation Technique: A Phantom Study and Clinical Use," *Acta Neurochir*., vol. 142, pp. 1169-1210, 2000.

[45] Q. H. Li *et al*., "The Application Accuracy of the Neuromate Robot - A Quantitative Comparison with Frameless and Frame-Based Surgical Localization Systems," *Comput. Aided Surg*., vol. 7, pp. 90-98, 2002.

[46] T. R. Varma and P. Eldridge, "Use of the Neuromate Stereotactic Robot in a Frameless Mode for Functional Neurosurgery," *Int. J. Med. Robot.*, vol. 2, pp. 107-113, 2006.

[47] L. Joskowicz *et al.*, "Miniature Robot-Based Precise Targeting System for Keyhole Neurosurgery: Concept and Preliminary Results," *Int. Congr. Ser.*, vol. 1281, pp. 618-623, 2005.

[48] L. Joskowicz *et al.*, "Image-Guided System with Miniature Robot for Precise Positioning and Targeting in Keyhole Neurosurgery," *Comput. Aided Surg.*, vol. 11, pp. 181-193, 2006.

[49] G. Ferrigno. (2011, Mar. 3). ROBOCAST [Online]. Available: 131.175.32.10/Robocast

[50] B. Davies *et al.*, "A Surgeon Robot for Prostatectomies," in 5th Int. Conf. Advanced Robotics, 1991, pp. 871-875.

[51] B. Davies *et al.*, "Neurobot: A Special-Purpose Robot for Neurosurgery," in *Proc. of IEEE Int. Conf. on Robotics and Automation*, 2000, vol.4, pp.4103-4108.

[52] G. Ferrigno. (2011, Mar. 3). ROBOCAST [Online]. Available: 131.175.32.10/Robocast

[53] G. Ferrigno. (2011, Mar. 3). ROBOCAST [Online]. Available: 131.175.32.10/Robocast

[54] K. MacLellan, "The Occasional Burr Hole," *Can. J. Rural Med.*, vol. 3, pp. 223-225, 1998.

References

[55] J. Heller and M. Cyperling, "Neuroarm: Revolutionary Procedure A World First," University of Calgary and Calgary Health Region, Calgary, Canada, 2008.

[56] C. W. Burckhardt *et al.*, "Stereotactic Brain Surgery," *IEEE Eng. Med. Biol.*, vol. 14, pp. 314-317, 1995.

[57] D. Glauser *et al.*, "Conception of a Robot Dedicated to Neurosurgical Operations," in 5th Int. Conf. Advanced Robotics, 1991, pp. 899-904.

[58] N. Villotte *et al.*, "Conception of Stereotactic Instruments for the Neurosurgical Robot Minerva," in *Proc. of the IEEE Annu. Int. Conf. of the Eng. in Medicine and Biology Soc.*, 1992, vol.14, pp. 1089-1090.

[59] H. Fankhauser *et al.*, "Robot for CT-guided Stereotactic Neurosurgery," *Stereotact. Funct. Neurosurg.*, vol. 63, pp. 93-98, 1994.

[60] D. Glauser *et al.*, "Neurosurgical Robot Minerva: First Results and Current Developments," *J. Image Guid. Surg.*, vol. 1, pp. 266-272, 1995.

[61] J. L. Hefti *et al.*, "Robotic Three-dimensional Positioning of a Stimulation Electrode in the Brain," *Comput. Aided Surg.*, vol. 3, pp. 1-10, 1998.

[62] M. Gooroochurn, "Development of machine vision techniques for intraoperative registration and bleeding characterisation in robot-assisted neurosurgery," Ph.D. dissertation, Dept. Mech. Eng., Loughborough Univer., Loughborough, United Kingdom, 2009.

[63] M. R. Bullock *et al.*, "Guidelines for the Surgical Management of Traumatic Brain Injury," *Eur. J. Emerg. Med.*, vol. 3, pp. 109-127, 1996.

[64] C. D. Winter *et al.*, "A Review of the Current Management of Severe Traumatic Brain Injury," *Surgeon*, vol. 3, pp. 329-337, Oct. 2005.

[65] S. Ian, "Intracranial Hypertension After Traumatic Brain Injury," *Indian J. Crit. Care Med.*, vol. 8, pp. 120-126, 2004.

[66] T. Zeng and L. Gao, "Management of Patients with Severe Traumatic Brain Injury Guided by Intraventricular Intracranial Pressure Monitoring: A Report of 136 Cases," *Chin. J. Traumatol.*, vol. 13, pp. 146-151, 2010.

[67] P. Whitfield, "The Management of Patients with Head Injury," *ACNR*, vol. 4, pp. 14-17, 2004.

[68] B. Mokri, "The Monro-Kellie Hypothesis: Applications in CSF Volume Depletion," *Neurology*, vol. 56, pp. 1746-1748, 2001.

[69] D. R. Huyette *et al.*, "Accuracy of the Freehand Pass Technique for Ventriculostomy Catheter Placement: Retrospective Assessment Using Computed Tomography Scans," *J. Neurosurg.*, vol. 108, pp. 88-91, Jan, 2008.

[70] T. B. Mapstone and R. A. Ratcheson, "Techniques of Ventricular Puncture," in *Neurosurgery*, 2nd ed., R. H. Wilkins and S. S. Rengachary, Eds. New York: McGraw Hill, 1996, pp. 179-184.

[71] I. F. Dunn *et al.*, "Ventriculostomy," in *Operative Neurosurgical Techniques: Indications, Methods, and Results*, 5th ed., H. H. Schmidek, W. H. Sweet and R. A. Ratcheson, Eds. Philadelphia: Saunders, 2006, pp. 35-36.

[72] T. S. Tse *et al.*, "Ventriculostomy and Infection: A 4-year-review in a Local Hospital," *Surg. Neurol. Int*., vol. 1, pp. 47-56, 2010.

[73] D. D. Binz *et al.*, "Hemorrhagic Complications of Ventriculostomy Placement: A Meta-Analysis," *Neurocrit. Care*, vol. 10, pp. 253-256, 2009.

[74] S. S. Lollis and D. W. Roberts, "Robotic Catheter Ventriculostomy: Feasibility, Efficacy, and Implications," *J. Neurosurg.*, vol. 108, pp. 269-274, Feb. 2008.

[75] F. Brassow and K. Baumann, "Volume of Brain Ventricles in Man Determined by Computer Tomography," *Neuroradiology*, vol. 16, pp. 187-189, 1978.

[76] S. T. O'Leary *et al.*, "Efficacy of the Ghajar Guide Revisited: A Prospective Study," *J. Neurosurg. Pediatr.*, vol. 92, pp. 1197-1201, May 2000.

[77] M. O'Toole, "Control Techniques for a Robot-Assisted Emergency Neurosurgery System," Ph.D. dissertation, Dept. Mech. Eng., Loughborough Univer., Loughborough, United Kingdom, 2010.

[78] P. A. Poletti, A. Platon and C. D. Becker, "Role of Imaging in the Management of Trauma Victims," in *Emergency Radiology: Imaging and Intervention*, B. Marincek and R. F. Dondelinger, Eds. New York: Springer Berlin Heidelberg, 2007, pp. 3-23.

[79] E. Kanal *et al.* "ACR Guidance Document for Safe MR Practices: 2007," *Am. J. Roentgenol.*, vol. 188, pp. 1447-1474, 2007.

[80] N. D. Zasler *et al.*, *Brain Injury Medicine: Principles and Practice*. New York: Demos Medical Publishing, 2007.

[81] A. B. Wolbarst and W. R. Hendee, "Evolving and Experimental Technologies in Medical Imaging," *Radiology*, vol. 238, pp. 16-39, 2006.

[82] S. A. Jackson and R. M. Thomas, "Introduction to CT physics," in *Cross Sectional Imaging Made Easy*. Edinburgh: Churchill Livingstone, 2004, .

[83] M. Hilts and C. Duzenli, "Image noise in X-ray CT polymer gel dosimetry," *J. Phys. Conf.*, vol. 3, pp. 252-256, 2004.

[84] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, pp. 629-639, 1990.

[85] G. Gerig *et al.*, "Nonlinear anisotropic filtering of MRI data," *IEEE Med. Imaging*, vol. 11, pp. 221-232, 1992.

[86] F. P. Vidal *et al.*, "Principles and Applications of Computer Graphics in Medicine," *Comput. Graphics Forum,* vol. 25, pp. 113-137, 2006.

[87] G. Gilboa *et al.*, "Image Enhancement and Denoising by Complex Diffusion Processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1020-1036, 2004.

[88] R. H. Chan and K. Chen, "Multilevel Algorithm for a Poisson Noise Removal Model with Total-variation Regularization," *Int. J. Comput. Math.*, vol. 84, pp. 1183-1198, 2007.

[89] J. L. McCrohan *et al.*, "Average Radiation Doses in a Standard Head Examination for 250 CT Systems," *Radiology*, vol. 163, pp. 263-268, Apr. 1987.

[90] J. A. Seibert, "Tradeoffs Between Image Quality and Dose," *Pediatr. Radiol.*, vol. 34, pp. 183-195, 2004.

[91] M. G. P. Cavalcanti *et al.*, "Craniofacial Measurements Based on 3D-CT Volume Rendering: Implications for Clinical Applications," *Dentomaxillofac. Rad.*, vol. 33, pp. 170-176, 2004.

[92] D. X. Liu *et al.*, "The Accuracy of 3D-CT Volume Rendering for Craniofacial Linear Measurements," *Shanghai Kou Qiang Yi Xue*, vol. 15, pp. 517-520, Oct, 2006.

[93] K. D. Kim *et al.*, "Accuracy of Facial Soft Tissue Thickness Measurements in Personal Computer-Based Multiplanar Reconstructed Computed Tomographic Images," *Forensic Sci. Int.*, vol. 155, pp. 28-34, Dec. 2005.

[94] J. M. Fitzpatrick *et al.*, "Image registration," in *Handbook of Medical Imaging II: Medical Image Processing and Analysis*, M. Sonka and J. M. Fitzpatrick, Eds. Bellingham, WA: SPIE Press, 2000, pp. 447–513.

[95] P. Grunert *et al.*, "Computer-Aided Navigation in Neurosurgery," *Neurosurg. Rev.*, vol. 26, pp. 73-99; May. 2003.

[96] D. W. Roberts *et al.*, "A Frameless Stereotaxic Integration of Computerized Tomographic Imaging and the Operating Microscope," *J. Neurosurg.*, vol. 65, pp. 545-549, Oct. 1986.

[97] C. R. Maurer *et al.*, "Acoustick: A Tracked A-Mode Ultrasonography System for Registration in Image-Guided Surgery," *Lect. Notes. Comput. Sc.*, vol.1679, pp. 953-962, 1999.

[98] A. C. Colchester *et al.*, "Development and Preliminary Evaluation of VISLAN, A Surgical Planning and Guidance System Using Intra-Operative Video Imaging," *Med. Image Anal.*, vol. 1, pp. 73-90, Mar. 1996.

[99] W. E. L. Grimson *et al.*, "An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization," *IEEE T. Med. Imaging*, vol. 15, pp. 129-140, 1996.

[100] C. A. Pelizzari and G. T. Y. Chen, "Registration of Multiple Diagnostic Imaging Scans Using Surface Fitting," in *Proc. 9th Int. Conf. the use of Comput. in Radiation Therapy*. Amsterdam, North-Holland, 1987, pp. 437-440.

[101] P. J. Besl and H. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 239-256, 1992.

[102] P. Viola and W. M. Wells III, "Alignment by Maximization of Mutual Information," *Int. J. Comput. Vision*, vol. 24, pp. 137-154, 1997.

[103] M. J. Clarkson *et al.*, "Registration of Multiple Video Images to Preoperative CT for Image-Guided Surgery," in *Proc. SPIE*, vol. 3661, pp. 14, 2003.

[104] P. S. Morgan *et al.*, "The Application Accuracy of the Pathfinder Neurosurgical Robot," in *Int. Congr. Series*, 2003, pp. 561-567.

[105] T. R. K. Varma and P. Eldridge, "Use of the Neuromate Stereotactic Robot in a Frameless Mode for Functional Neurosurgery," *Int. J. Med. Robot Comp.*, vol. 2, pp. 107-113, 2006.

[106] M. Ammirati *et al.*, "Comparison of Registration Accuracy of Skin-and Bone-Implanted Fiducials for Frameless Stereotaxis of the Brain: A Prospective Study," *Skull Base*, vol. 12, pp. 125, 2002.

[107] H. K. Gumprecht *et al.*, "Brainlab Vectorvision Neuronavigation System: Technology and Clinical Experiences in 131 Cases," *Neurosurgery*, vol. 44, pp. 97-104, Jan. 1999.

[108] J. Schlaier *et al.*, "Registration Accuracy and Practicability of Laser-directed Surface Matching," *Comput. Aided Surg.*, vol. 7, pp. 284-290, 2002.

[109] M. P. Fried *et al.*, "Image-Guided Endoscopic Surgery: Results of Accuracy and Performance in a Multicenter Clinical Study Using an Electromagnetic Tracking System," *Laryngoscope*, vol. 107, pp. 594-601, 1997.

[110] C. R. Maurer Jr *et al.*, "Estimation of Accuracy in Localizing Externally Attached Markers in Multimodal Volume Head Images," in *Proc. of SPIE Vol.1898*, Newport Beach, CA, 1993, pp. 43-54.

[111] J. M. Fitzpatrick et al., "Predicting Error in Rigid-Body Point-Based Registration," *IEEE Trans. Med. Imag.,* vol. 17, pp. 694-702, 1998.

[112] T. M. Peters, "Image-guidance for surgical procedures," *Phys. Med. Biol.*, vol. 51, pp. 505, 2006.

[113] R. J. Maciunas *et al.*, "Beyond Stereotaxy: Extreme Levels of Application Accuracy are Provided by Implantable Fiducial Markers for Interactive Image-Guided Neurosurgery," in *Interactive Image-Guided Neurosurgery*, R. J. Maciunas, Ed. USA: Thieme/AANS, 1993, pp. 259-270.

[114] J. B. West *et al.*, "Fiducial Point Placement and the Accuracy of Point-Based, Rigid Body Registration," *Neurosurgery*, vol. 48, pp. 810-816, Apr. 2001.

[115] S. Wolfsberger *et al.*, "Anatomical Landmarks for Image Registration in Frameless Stereotactic Neuronavigation," *Neurosurg. Rev.*, vol. 25, pp. 68-72, Mar. 2002.

[116] P. A. Woerdeman *et al.*, "Application Accuracy in Frameless Image-Guided Neurosurgery: A Comparison Study of Three Patient-to-image Registration Methods," *J. Neurosurg*., vol. 106, pp. 1012-1016, Jun. 2007.

[117] R. R. Shamir *et al.*, "Surface-based Facial Scan Registration in Neuronavigation Procedures: A Clinical Study," *J. Neurosurg.*, vol. 111, pp. 1201-1206, 2009.

[118] M. J. Clarkson *et al.*, "Using Photo-Consistency to Register 2D Optical Images of the Human Face to a 3D Surface Model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 1266-1280, 2001.

[119] John E.J. *et al.*, "ACR–ASNR Practice Guideline for the Performance of Computed Tomography (CT) of the Brain," ACR Guidelines and Standards Committee and ASNR Guidelines Committee, 2010.

[120] R. R. Shamir *et al.*, "Localization and Registration Accuracy in Image Guided Neurosurgery: A Clinical Study," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 4, pp. 45-52, 2009.

[121] J. B. West and J. M. Fitzpatrick, "Point-Based Rigid Registration: Clinical Validation of Theory," in *Proc. SPIE Vol. 3979*, San Diego, CA, 2000, pp. 353-359.

[122] K. S. Arun *et al.*, "Least-squares Fitting of Two 3-D Point Sets," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 9, pp. 698-700, Sept. 1987.

[123] M. Bickel *et al.*, "Evaluation of the Application Accuracy of 3D-Navigation Through Measurements and Prediction," in *World Congr. Medical Physics and Biomedical Eng.*, Munich, Germany, 2009, pp. 349-351.

[124] S. Frantz *et al.*, "Development and Validation of A Multi-Step Approach to Improved Detection of 3D Point Landmarks in Tomographic Images," *Image Vision Comput.*, vol. 23, pp. 956-971, 2005.

[125] M. Alker *et al.*, "Improving the Robustness in Extracting 3D Point Landmarks from 3D Medical Images Using Parametric Deformable Models," in Medical Image Comput. and Computer-Assisted Intervention, Utrecht, The Netherlands, 2001, pp. 582-590.

[126] S. Wörz and K. Rohr, "Localization of Anatomical Point Landmarks in 3D Medical Images by Fitting 3D Parametric Intensity Models," *Med. Image Anal.*, vol. 10, pp. 41-58, 2006.

[127] K. Subburaj *et al.*, "Automated Identification of Anatomical Landmarks on 3D Bone Models Reconstructed From CT Scan Images," *Comput. Med. Imaging Graphics*, vol. 33, pp. 359-368, 2009.

[128] D. Deo and D. Sen, "Mesh Processing for Computerized Facial Anthropometry," *J. Comput. Inf. Sci. Eng.*, vol. 10, pp. 1-12, Mar. 2010.

[129] D. Chen *et al.*, "Scale-Space Volume Descriptors for Automatic 3D Facial Feature Extraction," Int. J. Signal Processing, vol. 5, pp. 264-269, 2009.

[130] M. Gooroochurn *et al.*, "Facial Recognition Techniques Applied to the Automated Registration of Patients in the Emergency Treatment of Head Injuries," *P. I. Mech. Eng. H. Med*., vol. 225, pp. 170-180, 2011.

[131] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *Automatica*, vol. 11, pp. 285-296, 1975.

[132] A. Gueziec, "Extracting Surface Models of the Anatomy from Medical Images," in *Handbook of Medical Imaging II: Medical Image Processing and Analysis*, M. Sonka and J. M. Fitzpatrick, Eds. Bellingham, WA: SPIE Press, 2000, pp. 345-398.

[133] The MathWorks, Source Code for the Matlab Isosurface Function. 2007.

[134] P. Alliez *et al*., "Anisotropic Polygonal Remeshing," *ACM T. Graphic*, vol. 22, pp. 485-493, Jul. 2003.

[135] M. Ester *et al.*, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining*, Portland, OR, 1996, pp. 226-231.

[136] J. Cao *et al.*, "Point Cloud Skeletons via Laplacian Based Contraction," in *Shape Modeling Int. Conf.*, Aix-en-Provence, France, 2010, pp. 187-197.

[137] M. Gooroochurn *et al.*, "A Registration Framework for Preoperative CT to Intraoperative White Light Images," in *Medical Image Understanding and Analysis*, London, United Kingdom, 2009, pp. 184-188.

[138] M. Ejiri, "Machine Vision in Early Days: Japan’s Pioneering Contributions," in *Lect. Notes in Comput. Sc.*, vol. 4843, pp. 35-53, 2007.

[139] D. P. Nicolella *et al.*, "Machine Vision Photogrammetry: A Technique for Measurement of Microstructural Strain in Cortical Bone," *J. Biomech*., vol. 34, pp. 135-139, 2001.

[140] N. Koenderink-Ketelaars, "A Knowledge-Intensive Approach to Computer Vision Systems," Ph.D. dissertation, Delft Univ. of Tech., Delft, The Netherlands, 2010.

[141] B. P. L. Lo *et al.*, "Episode Classification for the Analysis of Tissue/Instrument Interaction with Multiple Visual Cues," *Lect. Notes Comput. Sc.*, vol. 2879, pp. 230-237, 2003

[142] N. Padoy *et al.*, "On-Line Recognition of Surgical Activity for Monitoring in the Operating Room," in *Proc. 20th Conf. Innovative Applications of Artificial Intell*, Chicago, Illinois, 2008, pp. 1718-1724.

[143] R. K. Reznick, "Teaching and Testing Technical Skills," *AM. J. Surg.*, vol. 165, pp. 358-361, 1993.

[144] R. McCloy and R. Stone, "Virtual Reality in Surgery," *BMJ*, vol. 323, pp. 912, 2001.

[145] K. Moorthy *et al.*, "Objective Assessment of Technical skills in Surgery," *BMJ*, vol. 327, pp. 1032-1037, 2003.

[146] W. O. Sack *et al.*, *Essentials of Pig Anatomy*. Ithaca, New York: Veterinary Textbooks, 1982.

[147] G. M. Kaiser and N. R. Fruhauf, "Method of Intracranial Pressure Monitoring and Cerebrospinal Fluid Sampling in Swine," *Lab. Anim.*, vol. 41, pp. 80-85, 2007.

[148] J. F. M. Manschot and A. J. M. Brakkee, "The Measurement and Modelling Of The Mechanical Properties Of Human Skin In Vivo-I. The Measurement," *J. Biomech.*, vol. 19, pp. 511-515, 1986.

[149] J. Ankersen *et al.*, "Puncture Resistance and Tensile Strength of Skin Simulants," *P. I. Mech. Eng. H. Med*., vol. 213, pp. 493-501, 1999.

[150] E. Raposio and R. E. A. Nordström, "Biomechanical Properties of Scalp Flaps and Their Correlations to Reconstructive and Aesthetic Surgery Procedures," *Skin Res. Tech.*, vol. 4, pp. 94-98, 1998.

[151] N. J. Vardaxis et al., "Confocal Laser Scanning Microscopy of Porcine Skin: Implications for Human Wound Healing Studies," *J. Anat.*, vol. 190, pp. 601-611, 1997.

[152] T. Ohashi *et al.*, "Hill-Climbing Algorithm for Efficient Color-Based Image Segmentation." in *Int. Conf. Signal Process., Pattern Recognition, and Applicat.*, Rhodes, Greece, 2003, pp. 17-22.

[153] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. 3rd ed. New Jersey: Pearson Prentice Hall, 2008.

[154] M. Mahy *et al.*, "Evaluation of Uniform Color Spaces Developed After the Adoption of CIELAB and CIELUV," *Color Res. Appl*., vol. 19, pp. 105-121, 1994.

[155] R. M. Haralick *et al.*, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.,* vol. 3, pp. 610-621, 1973.

[156] A. Blum *et al.*, "Digital Image Analysis for Diagnosis of Cutaneous Melanoma. Development of a Highly Effective Computer Algorithm Based on Analysis of 837 Melanocytic Lesions," *Br. J. Dermatol.*, vol. 151, pp. 1029-1038, 2004.

[157] F. Ercal *et al.*, "Neural Network Diagnosis of Malignant Melanoma From Color Images," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 837-845, 2002.

[158] H. Voigt and R. Classen, "Computer Vision and Digital Imaging Technology in Melanoma Detection," *Seminars in Oncology*, vol. 29, pp. 308-327, 2002.

[159] S. Narkilahti *et al.*, "Monitoring and Analysis of Dynamic Growth of Human Embryonic Stem Cells: Comparison of Automated Instrumentation and Conventional Culturing Methods," *Biomed. Eng.*, vol. 6, pp. 6-11, 2007.

[160] J. M. Fitzpatrick and J. B. West, “The Distribution of Target Registration Error in Rigid-Body Point-Based Registration,” *IEEE Trans. Med. Imag.*, vol. 20, pp.917-927

Appendix A

Publications

Machine Vision for the Inspection of Surgical Tasks: Applications to Robotic Surgery Systems

M. Ovinis, D. Kerr, K. Bouazza-Marouf, and M. Vloeberghs

Abstract**—**The use of machine vision to inspect the outcome of surgical tasks is investigated, with the aim of incorporating this approach in robotic surgery systems. Machine vision is a non-contact form of inspection i.e. no part of the vision system is in direct contact with the patient, and is therefore well suited for surgery where sterility is an important consideration,. As a proof-of-concept, three primary surgical tasks for a common neurosurgical procedure were inspected using machine vision. Experiments were performed on cadaveric pig heads to simulate the two possible outcomes i.e. satisfactory or unsatisfactory, for tasks involved in making a burr hole, namely incision, retraction, and drilling. We identify low level image features to distinguish the two outcomes, as well as report on results that validate our proposed approach. The potential of using machine vision in a surgical environment, and the challenges that must be addressed, are identified and discussed.

Keywords**—**Visual inspection, machine vision, robotic surgery.

I. Introduction

ROBOTIC surgery systems are capable of performing surgery autonomously under the direction and supervision of a surgeon, who intervenes in the event of a mistake. The MINERVA system [1], developed by the Swiss Federal Institute of Technology, is a robotic neurosurgery system capable of autonomously performing tasks such as incision and retraction of the scalp, and drilling of a burr hole through the cranium [2]. A burr hole is a small hole created in the cranium to provide access to the brain for surgery, and is a common neurosurgical procedure. Clinical trials of the system involving 13 patients have been reported [3, 4]. In the MINERVA system, force and torque characteristics were used to determine surgical task completion [2]. For example, in drilling of a burr hole, the force and torque profile of the drill is used to stop the drill bit after it has penetrated the cranium.

However, relying on these measurements alone is analogous to performing surgery blindly. The ability to mimic and embody a surgeon's visual sensory and decision making capability has the potential to improve current practice in robotic surgery. For example, machine vision could be used for visual inspection of the burr hole to validate successful drilling. Incorporating visual cues improves the reliability and robustness of robotic surgery systems by introducing redundancy in its sensor space, which traditionally only employs force, torque and positional measurements. As a step towards this goal, the use of machine vision to inspect the outcome of surgical tasks was investigated.

Several authors have investigated the use of vision in the operating theatre [5, 6]. Lo et. al. [5] developed a framework for the classification of surgical episodes in minimally invasive surgeries, in video sequences, using multiple visual cues. Specifically, their work was concerned with the detection of surgical tasks, such as the interaction of surgical instruments with soft tissue, retraction, cauterisation, and suturing. Padoy et. al. [6] developed an approach to recognize a subset of surgical tasks performed by the surgeon during laparoscopic surgery using visual cues. The aim of this study is to investigate a machine vision based approach for the inspection of the outcome of surgical tasks. Using machine vision in this context has not been reported. The use of machine vision, for greater situational awareness and confidence, has been the goal of our research.

Admittedly, the surgical environment can be highly unstructured, in contrast to the more controlled conditions typically found where machine vision is applied. However, because the surgical tasks being inspected are intended to be performed robotically, and their context is known, visual inspection is feasible. To illustrate the concept, practical examples of machine vision to inspect three primary surgical tasks involved in making burr hole, namely incision, retraction of an incision, and drilling of a burr hole, were used. The ability to inspect these tasks is an important functionality, from an operational and safety point of view for creating a burr hole robotically, as each of these tasks can only be performed upon successful completion of the previous task. For example, the scalp has to be properly incised prior to retraction, the incision properly retracted prior to drilling, and the cranium properly drilled prior to accessing the brain through the burr hole. The first illustrative example is the inspection of skin incisions.

M. Ovinis is a PhD scholar in the Department of Mechanical and Manufacturing Engineering, Loughborough University, LE11 3TU, Loughborough, UK (phone: +441509227579; fax: +441509227648; email: M.Ovinis2@lboro.ac.uk).

D. Kerr and K. Bouazza-Marouf are Senior Lecturers in the Department of Mechanical and Manufacturing Engineering, Loughborough University, LE11 3TU, Loughborough, UK (e-mail: d.kerr@lboro.ac.uk and k.bouazza-marouf@lboro.ac.uk respectively).

Prof. M. Vloeberghs is a Consultant Neurosurgeon at the Queen's Medical Centre, Nottingham University, NG7 2UH, Nottingham, UK (e-mail: Michael.Vloeberghs@Nottingham.ac.uk).

II. METHODOLOGY

A. Experimental setup

Experiments were performed using cadaveric pig heads to simulate the two possible outcomes (satisfactory and unsatisfactory) of three primary surgical actions involved in making a burr hole. Practical and ethical considerations prevent us from experimenting on humans or live animals. Thus, animal cadavers were used as a substitute. The selection of an animal cadaver should as far as possible be representative of human anatomy. With the exception of primates, the anatomy of the temporoparietal region (site of the burr hole) of pigs most closely resembles that of a human [7].

The surgical procedure for creating a burr hole in a pig as described by Kaiser and Fruhauf [8] was adopted with modifications. Tension was applied parallel to the plane of the skin to simulate parting of the skin (to facilitate placement of retractors after the skin is incised). Two square tabs (29 x 29 mm) were attached to the skin with cyanoacrylate adhesive [9]. The tabs were then pulled apart by two 600 gram weights using a pulley and weight system (Fig. 1). The distance between the tabs was 29 mm.

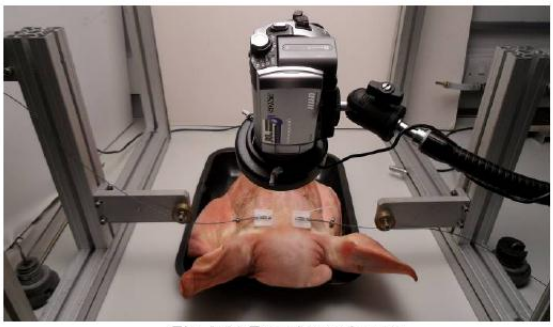


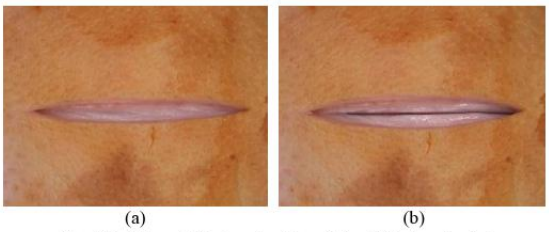
Fig. 1 (a) Experimental setup

The frontoparietal region was chosen as the site of the burr hole, instead of the temporoparietal region, because the underlying structure of the frontoparietal region allows for easier access to the brain. A midline incision, approximately 3 cm long, centred over the site of the burr hole, was made with a #10 scalpel blade, at different depths up to the pericranium. The incision was subsequently retracted at varying degrees using a self retaining retractor to expose the cranium, until it was wide enough to accommodate a burr hole. A burr hole 10 mm in diameter was then drilled in the cranium using a cordless drill at high speed. The characteristics of each surgical task were investigated to determine appropriate visual cues to distinguish the two outcomes. Because of the variability associated with biological subjects, fifty experiments were performed to give more statistically reliable data.

A CCD video camera, with a white LED ring-lighting unit mounted around the camera lens (Fig. 1), was used to obtain images of the surgical tasks. Used within its working distance, ring lightning can reduce the effects of shadows, by providing an even distribution of lighting in the field of view. The ring light had an adjustable intensity and a four section quadrant control. The white balance setting of the camera was manually adjusted to account for 6400K colour temperature of the white LED.

B. Characterisation of an incision

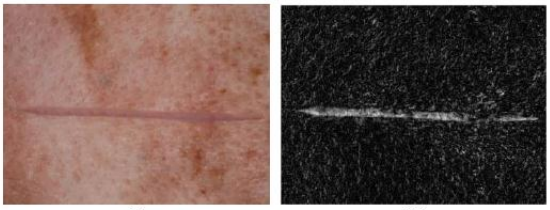
When surgeons incise the scalp, the scalpel blade is pushed into the skin, through the layers of the scalp, until it has touched cranium. A satisfactory incision should be both sufficiently long and deep i.e. an incision up to the pericranium along the length of the incision (Fig. 2b).



(a) (b)

Fig. 2 (a) Unsatisfactory incision (b) satisfactory incision

The shape properties and visual appearance of the two incision types were investigated. To determine the shape properties of the incision, it was first necessary to segment the incision in an image. Due to the variation in the appearance of skin, and because the edges of the incisions may have low contrast in an image (Fig. 3a); segmentation of the incision is not trivial. Segmentation of the difference image (obtained by subtracting an image before and after an incision is made) was attempted using region-based and edge-based segmentation. Because the underlying skin layer may be similar in appearance to the uppermost skin layer, compounded by the fact that the skin deforms when incised, the incision is ill-defined in the difference image (Fig. 3b).



(a) (b)

Fig. 3(a) Incision (b) Difference image on incision in (a)

Thresholding an intensity image of the incision, using a region-based local adaptive thresholding technique [10], and subsequently performing morphological operations gave better results (Fig. 4).

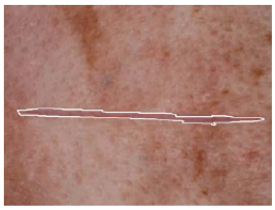
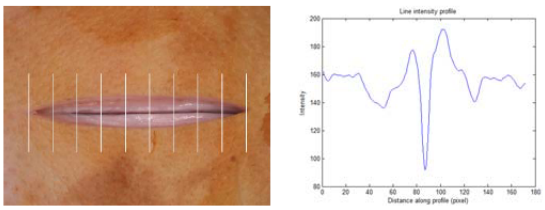


Fig. 4 Segmented incision using local adaptive thresholding

However, local adaptive thresholding is sensitive to threshold level and neighbourhood/filter size. As such, consistent results could not be obtained over a wide range of images. Difficulties in segmenting the incision notwithstanding, the shape properties of manually segmented incisions such as regional and boundary descriptors e.g. statistical moments, invariant moments, eccentricity, etc, were determined.

The visual appearance of the incision was examined next. The use of cadavers meant that any bleeding as a result of the skin being incised could not be simulated, as cadaveric skin has no blood supply. However, as any bleeding will normally be cauterised and irrigated, the techniques developed may potentially be useful during actual surgery. It was observed that satisfactory incisions exhibit a characteristic dark line, made as the scalpel blade incises the skin up to the pericranium, which is not present in unsatisfactory incisions. As such, the presence of a dark line along the length on an incision can be used to determine a satisfactory incision.

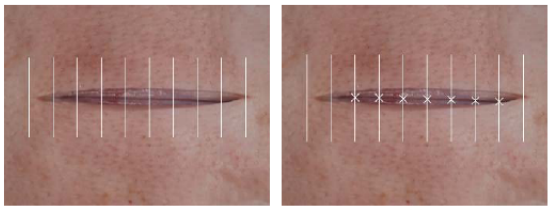
To detect the presence of this line, the line intensity profile of line segments along the midline of an incision was determined (Fig. 5). The pixel corresponding to an incision line is the global minimum of the line intensity profile. This method does not require the edges of an incision, which are difficult to segment, to be known.



(a) (b)

Fig. 5 (a) Line segment across incision (b) corresponding line intensity profile

The low contrast of the incision line in some incisions makes its detection difficult (Fig. 6). To enhance the contrast of the incision line, a contrast limited adaptive histogram equalisation (CLAHE) algorithm [11] was used.



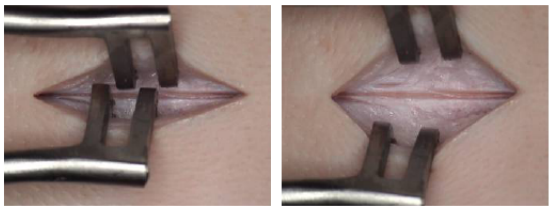
(a) (b)

Fig. 6 Output of incision line detection algorithm for (a) original image (b) contrast enhanced image

Assuming the camera is calibrated, the length of the incision would correspond to the distance between the first and last line segment containing a pixel that corresponds to the incision line. The resolution of the incision length determined by this method is dependent of the spacing between the line segments. The accuracy in determining the incision length is dependent on the accuracy of the calibration between the camera and task space, as well as the viewing angle of the camera, which ideally should be normal to the plane of view.

C. Characterisation of a retracted incision

Retraction of an incision (Fig. 7) is considered satisfactory if the incision is able to accommodate a burr hole. This may be indirectly determined from the separation distance of the prongs of the retractors. A potential problem with this approach is that the retractors may have opened without the incision being retracted i.e. the prongs of the retractor may not have engaged the edges of the incision, leading to the erroneous assumption that the edges of the incision are sufficiently separated. The approach adopted instead was to determine the amount of free space within the retracted incision by fitting the largest circle within the region bounded by the edges of the incision and prongs of the retractor. An incision is deemed to have been sufficiently retracted if the circle is larger than the prospective burr hole.

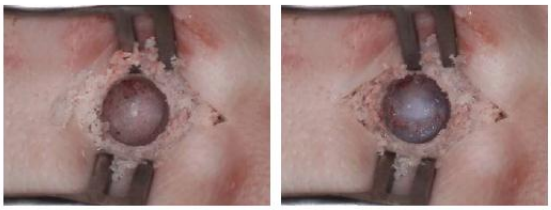


(a) (b)

Fig. 7 (a) Unsatisfactory retraction (b) Satisfactory retraction

D. Characterisation of a burr hole

A satisfactory outcome is one where the underlying dura mater i.e. the outermost part of the brain, is visible (Fig. 8b).

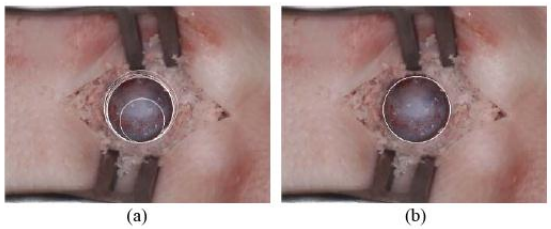


(a) (b)

Fig. 8 (a) Unsatisfactory/partial burr hole (b) Satisfactory/complete burr hole (dura mater is visible)

To detect the presence of a burr hole in an image, a generalised Hough transform to detect circles was used. The corresponding accumulator cell is a triple combination in 3D parameter space consisting of the circle centre coordinates and radius (range of radius values specified). Peak accumulator cell values represent a potential burr hole. A circle is disregarded if the ratio of pixels to perimeter is less than 10% of the total pixel number of the whole image. This ratio eliminates false positives i.e. a burr hole is sometimes 'detected' when there is none. The 10% ratio was found to represent an acceptable compromise between false detection and under detection in our images.

For each circle with a given radius, a minimum ratio of the number of accumulator cell count i.e. ratio of detected edge pixels to the number of pixels along the perimeter of the circle, was specified so that the algorithm was not biased towards larger circles, which may have a higher accumulator count over smaller circles. While the circle with the highest ratio of pixel count to the number of pixels along the perimeter of a circle usually corresponds to the actual burr hole, this is not always the case. This is because the edge of a burr hole is rarely a perfect circle (this is compounded by imperfect segmentation and distortions due to camera viewing angle), leading to the detection of multiple circles (Fig. 9a). By smoothing the accumulator array, we obtain a 'principle' circle (Fig. 9b). The principal circle was found to correspond well to the boundaries of a manually segmented burr hole. Once isolated, the difference in colour and texture of the two types of burr hole was investigated.



(a) (b)

Fig. 9 (a) Multiple circles detected (b) 'Principal' circle

III. Result and Discussion

A. Incision

A linear discriminant analysis was performed to classify 122 incision images consisting of 56 images of satisfactory incisions and 66 images of unsatisfactory incisions. The aspect ratio of the incision, defined as the ratio of the major axis and minor axis of an ellipse that has the same normalised second central moments as the incision, gave the best classification results. However, the misclassification error rate was 26.4%. Fig. 10 illustrates the aspect ratio for the two types of incision.

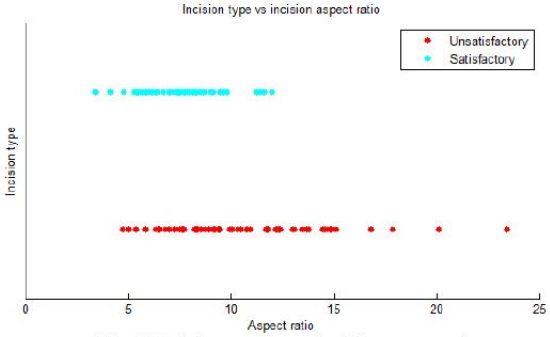
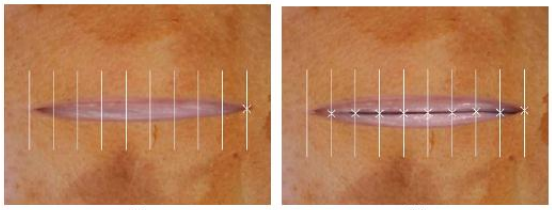


Fig. 10 Incision type versus incision aspect ratio

There is significant overlap in the feature space of the two classes of incisions. In general, it was found that shape properties were not discriminatory enough to distinguish the two types of incision. The in-class variation in shape properties is larger than the between class variation in shape properties i.e. two satisfactory incisions can have shape properties that are less alike than a satisfactory and unsatisfactory incision.

Fig. 11 illustrates the output of the incision line detection method for an unsatisfactory and satisfactory incision. The crosses represent the position along the line intensity profile that corresponds to the perceived position of the incision line.



(a) (b)

Fig. 11 Perceived incision line for (a) unsatisfactory incision (b) satisfactory incision

Classification of an incision using the line detection algorithm performed better, with a classification rate of 89.4%. Incisions were classified based on the mean distance of pixels corresponding to an incision line to a least squares fit

of a line to these pixels. Satisfactory incisions are incisions with a mean distance of less than half the line segment separation distance. Misclassification was mainly due to two reasons. The first is pixels being incorrectly identified as the incision line i.e. the global minimum of the line intensity profile does not always correspond to the incision line. The second is under detection of the incision line due to low contrast, even with contrast enhancement.

B. Retraction

A Canny edge detector was used to detect the edges of the incision. However, weak edges cannot be detected because of the sometimes low contrast at the boundaries of an incision. To improve segmentation of the incision, sections of the ring light were turned off to create shadows to facilitate detection of the edges of the incision (Fig. 12).

(a) (b) (c)

Fig. 12 (a) Left section of ring light turned off, (b) Right section of right light turned off (c) Detected edges

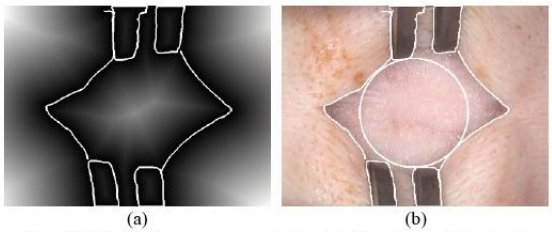
The next step was to segment the retractors. The retractors are made from surgical stainless steel and have a dull grey colour that is distinct in the operative site. The difference in colour of the retractor and the background was used as a basis to segment the retractors. Colour based segmentation [12] using a k-means clustering algorithm was used, with the initial cluster centroid positions or 'seeds' specified. The 'seeds' are the local maxima (determined using a hill-climbing optimization technique) of a three-dimensional L*a*b* colour space histogram of the image. Fig. 13 shows the segmentation results.

(a) (b)

Fig. 13 (a) Original colour image (b) Segmented image

Colour-based segmentation has successfully segmented the retractors by first dividing the colour image to regions of the most dominant colour. A representative colour value for the retractors was obtained by averaging the colour value of the pixels for a sample region of the retractor. This colour value was subsequently used to select the region whose average pixel colour value is closest to the representative colour value for the retractors.

Once the edges of the incision and retractors were detected, a distance map based on the distance transform of all the pixels within the incision, to the edges of the prongs of the retractors and edges of the incision was determined (Fig. 14a). The pixel with the maximum distance within these edges is the centre of the circle. The radius of the circle is the distance of this pixel to the nearest edge pixel of the prongs of the retractors or edge pixel of the incision (Fig. 14b). Assuming the calibration between the task space and camera frame is known, the diameter of the circle can be found. If the diameter of the circle is equal or larger than the diameter of the prospective burr hole (within a specified tolerance), the incision is deemed to be sufficiently retracted.



(a) (b)

Fig. 14 (a) A distance map overlaid with the edges of the incision and retractor (b) Circle whose centre corresponds to the pixel with the maximum distance within the incision in the distance map and whose radius corresponds to the distance of that pixel with the nearest pixel along the edges of the incision and retractor.

C. Burr hole

The visual inspection of the burr hole was formulated as a classification problem. The difference in colour between the cranium and dura mater was used to distinguish between a partial and complete burr hole. Each burr hole was represented by the median pixel colour value of all its pixels, in three colour spaces, RGB, L*a*b* and HSI. The images in the dataset were labelled for ground-truth and a leave-one-out cross validation using linear discriminant analysis (LDA) was performed to classify the images. Since the classification is based on statistics of the sample, classification results will be more reliable the larger the sample population. A random subset of 108 images was used (54 images of complete burr holes and 54 images of partial burr holes). Each image was used as training data with the remaining one being evaluated. The mean and maximum classification error was obtained by repeating the training and testing of the images ten times. Table 1 summarises the classification error when using the respective colour spaces.

TABLE I
COLOUR LINEAR DISCRIMINANT ANALYSIS LEAVE ONE OUT CLASSIFICATION ERROR RATE

Colour space	L*a*b*	HSI	RGB
Median (average)	2.4	2.6	2.4
(maximum)	(3.7)	(4.6)	(3.7)

The average misclassification rate in classifying burr holes based on colour is 2.4%. The classification sensitivity and specificity are 96% and 98% respectively. Although alternative classifiers might give better classification results, because of the relatively small dataset and the use of a single feature i.e. colour, the emphasis was on the strength of the features rather than strength of the classifiers. Classification rates were comparable in all colour spaces. The advantage of using the L*a*b* and HSI colour space over RGB colour space may only be apparent when there are changes in illumination, as changes in lighting will affect colour information. However, because the comparison is made for colour images without varying the illumination, it cannot be said that classification in L*a*b* or HSI space will perform better than classification in RGB space. Fig. 15 shows a scatter plot of the median pixel colour value of partial burr holes and complete burr holes in L*a*b* colour space, and the decision boundary.

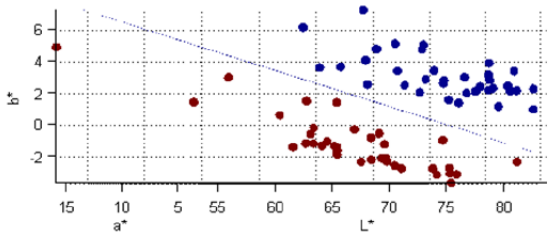


Fig. 15 Scatter plot of median pixel colour value of partial burr holes (black) and complete burr holes (light grey)

Fig. 16 illustrates all the images that the classifier could not discriminate based on colour alone.

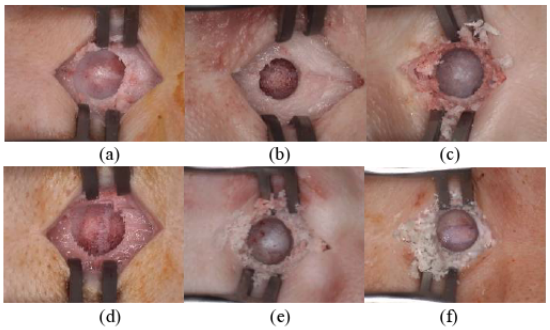


Fig. 16 Misclassified images

Fig. 16 a–e are false positives i.e. partial burr hole incorrectly classified as a complete burr hole, while Fig. 16 (f) is a false negative i.e. complete burr hole incorrectly classified as an incomplete burr hole. Textural information was considered to improve classification performance. Haralick's 14 coefficients [13] for the gray level co-occurrence matrix and statistical features for texture classification [14] were computed and used as a basis for classification. Classification was performed using all permutations of the different texture measures. The best classification results were obtained using the second moment/variance of the gray level image histogram of the burr hole, with a misclassification error rate of 11.9%. Combining textural information with colour information did not improve classification results.

Although this work has shown promising results on the use of machine vision for inspection of the outcome of the three surgical tasks, there are several challenges regarding the use of machine vision for this purpose. The first is the inherent natural variation in appearance of biological matter. A case in point is the inspection of incisions based on the presence of an incision line. Because of the large variations in appearance of the incision line, the reliance on a single feature was unable to capture all the differences between a satisfactory and unsatisfactory incision. The second is misclassification. False positives in the detection are more of an issue than false negatives. Erroneously classifying an unsatisfactory surgical task as satisfactory will result in the subsequent task being performed. The use of machine vision should therefore not be viewed in isolation. The aim is to use machine vision to augment/complement other sensory information, although it is not inconceivable that the state of the art in machine vision can advance to a stage where it is able to emulate the visual recognition ability and interpretation process of a human expert. This work represents the first step in this direction.

IV. Conclusion

In this paper, the use of machine vision to inspect surgical tasks was investigated. To the authors' knowledge, the use of vision in this context has never been considered before. Encouraging results were obtained for the surgical tasks under consideration, which has been demonstrated by our experiments. For example, the classification rate for incisions was 89.4% while classification rate for burr holes was 97%. We conclude that our findings are encouraging and suggestive for the potential use of machine vision for a more comprehensive approach in robotic surgery.

References

[1] C. W. Burckhardt, P. Flury and D. Glauser, "Stereotactic brain surgery," Engineering in Medicine and Biology Magazine, IEEE, vol. 14, pp. 314-317, 1995.

[2] N. Villotte, D. Glauser, P. Flury and C. W. Burckhardt, "Conception of stereotactic instruments for the neurosurgical robot minerva," in Engineering in Medicine and Biology Society, Vol.14. Proceedings of the Annual International Conference of the IEEE, 1992, pp. 1089-1090.

[3] H. Fankhauser, D. Glauser, P. Flury, Y. Piguet, M. Epitaux, J. Favre and R. A. Meuli, "Robot for CT-guided stereotactic neurosurgery," Stereotact. Funct. Neurosurg., vol. 63, pp. 93-98, 1994.

[4] D. Glauser, H. Fankhauser, M. Epitaux, J. L. Hefti and A. Jaccottet, "Neurosurgical robot Minerva: first results and current developments," J. Image Guid. Surg., vol. 1, pp. 266-272, 1995.

[5] B. P. L. Lo, A. Darzi and G. Z. Yang, "Episode Classification for the Analysis of Tissue/Instrument Interaction with Multiple Visual Cues," Medical Image Computing and Computer-Assisted Intervention: MICCAI .International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 230-237, 2003.

[6] N. Padoy, T. Blum, H. Feussner, M. O. Berger and N. Navab, "On-line recognition of surgical activity for monitoring in the operating room," in

Proceedings of 20th Conference on Innovative Applications of Artificial Intelligence (IAAI), 2008.
[7] R. A. Rival, O. M. Antonyshyn, J. H. Phillips and C. Y. Pang, "Temporal fascial periosteal and musculoperiosteal flaps in the pig: Design and blood flow inspection," J. Craniofac. Surg., vol. 6, pp. 466, 1995.
[8] G. M. Kaiser and N. R. Fruhauf, "Method of intracranial pressure monitoring and cerebrospinal fluid sampling in swine," Laboratory Animals(London), vol. 41, pp. 80-85, 2007.
[9] J. F. M. Manschot and A. J. M. Brakkee, "The measurement and modelling of the mechanical properties of human skin in vivo--I. The measurement," J. Biomech., vol. 19, pp. 511-515, 1986.
[10] P. Wellner, "Adaptive thresholding for the DigitalDesk," Xerox, EPC1993-110, 1993.
[11] K. Zuiderveld, "Contrast Limited Adaptive Histograph Equalization." in Graphic Gems IV. San Diego: Academic Press Professional, 1994, pp. 474-485.
[12] T. Ohashi, Z. Aghbari and A. Makinouchi, "Hill-climbing algorithm for efficient color-based image segmentation." in IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, 2003, pp. 17-22.
[13] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features of Image Classification," IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, Nov. 1973.
[14] R. C. Gonzalez, & R. E. Woods, Digital Image Processing, Pearson Prentice Hall, pp. 828-842.

Localisation of Anatomical Soft Tissue Landmarks of the Head in CT Images

M. Ovinis, D. Kerr, K. Bouazza-Marouf, and M. Vloeberghs

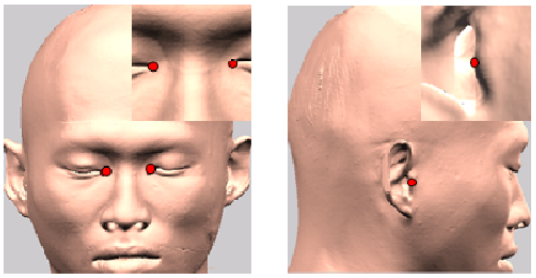
***Abstract*—**In this paper, algorithms for the automatic localisation of two anatomical soft tissue landmarks of the head, the medial canthus (inner corner of the eye) and the tragus (a small, pointed, cartilaginous flap of the ear), in CT images are described. These landmarks are to be used as a basis for an automated image-to-patient registration system we are developing. The landmarks are localised on a surface model extracted from CT images, based on surface curvature and a rule based system that incorporates prior knowledge of the landmark characteristics. The approach was tested on a dataset of near isotropic CT images of 95 patients. The position of the automatically localised landmarks was compared to the position of the manually localised landmarks. The average difference was 1.5 mm and 0.8 mm for the medial canthus and tragus, with a maximum difference of 4.5 mm and 2.6 mm respectively.The medial canthus and tragus can be automatically localised in CT images, with performance comparable to manual localisation.

***Keywords*—**Anatomical soft tissue landmarks, automatic localisation, Computed Tomography (CT)

I. Introduction

IMAGE-TO-PATIENTregistration is a method where correspondence between 'image space' (an image of an anatomy) and 'patient space' (the anatomy itself) is established.Features to be co-registered that are used to facilitate the registration are known as a registration basis. Anatomical soft tissue landmarks of the head can be used as a basis for image to patient registration [1]–[2]. We have previously proposed [3]-[4] the use of the canthus and tragus as a registration basis for an automated, preoperative CT image to intraoperative patient registration system. The use of these landmarksavoids resorting to fiducial markers. At present, these landmarks are manually localised. If localised automatically, they could be used as a registration basis in an automated image-to-patient registration system. The localisation of these landmarks in patient space using close range photogrammetry has previously been reported by Gooroochurn et al [5]. In this paper, an approach to automatically localise the medial canthus and tragus in CT images is described.

The medial canthus (Fig. 1a) and tragus (Fig. 1b) are conventional soft tissue craniometric anatomical landmarks with the following definitions: The medial canthus is the inner corner of the eye where the upper and lower eyelids meet. The tragus is a small, pointed, cartilaginous flap in front of the external opening of the ear. As the tragus is a relatively large structure, the most lateral point of the tragus is chosen as the point which uniquely defines its position.



(a) (b)

Fig. 1 (a) Medial canthus and (b) the most lateral point of the tragus

The majority of work on localising anatomical landmarks in head CT images is concerned with internal landmarks of the head e.g. tips of the frontal, temporal, and occipital horns of the ventricular system; saddle point of the zygomatic and nasal bone; and tip of the external occipital protuberance [6]-[8]. Early work on localising these anatomical landmarks in CT images used 3D generalisations of 2D differential operators/corner detectors [9]. A more recent approach used 3D deformable models [8],[10]. In this semi automatic approach, the user is required to provide a gross localisation of the landmarks, which the algorithm then refines, by fitting models of anatomical structures to their images. Recently, a method to automatically localised anatomical landmarks in CT images was proposed by Subburaj et al [11], who used surface curvature and prior knowledge of the gross spatial location of landmarks to localise anatomical landmarks of the knee-joint. However, their localisation technique, based on the relative location of landmarks, is not applicable in our work, as there are no other soft tissue landmarks that can be used to facilitate the localisation of the medial canthus and tragus.

M. Ovinis is a PhD scholar in the Department of Mechanical and Manufacturing Engineering, Loughborough University, LE11 3TU, Loughborough, UK (phone: +441509227579; fax: +441509227648; email: M.Ovinis2@lboro.ac.uk).

D. Kerr and K. Bouazza-Marouf are Senior Lecturers in the Department of Mechanical and Manufacturing Engineering, Loughborough University, LE11 3TU, Loughborough, UK (e-mail: d.kerr@lboro.ac.uk and k.bouazza-marouf@lboro.ac.uk respectively).

Prof. M. Vloeberghs is a Consultant Neurosurgeon at the Queen's Medical Centre, Nottingham University, NG7 2UH, Nottingham, UK (e-mail:Michael.Vloeberghs@Nottingham.ac.uk).

II. Methodology

The medial canthus and tragus is localised on a surface model of the head (the surface between the surrounding air and the voxels inside the head), extracted from CT images. As these landmarks exhibit characteristic surface curvature properties, a thresholded surface curvature map is used to localise these landmarks. However, the use of surface curvature alone is insufficient to localise these landmarks, as there will be many false detections. We use a rule-based system incorporating prior knowledge of the landmark characteristics to reduce the search space and avoid large localisation errors. Landmarksare identified based on having a geometric structure and spatial location consistent with the prior knowledge. A summary of the method is shown in flowchart form in Fig 2.

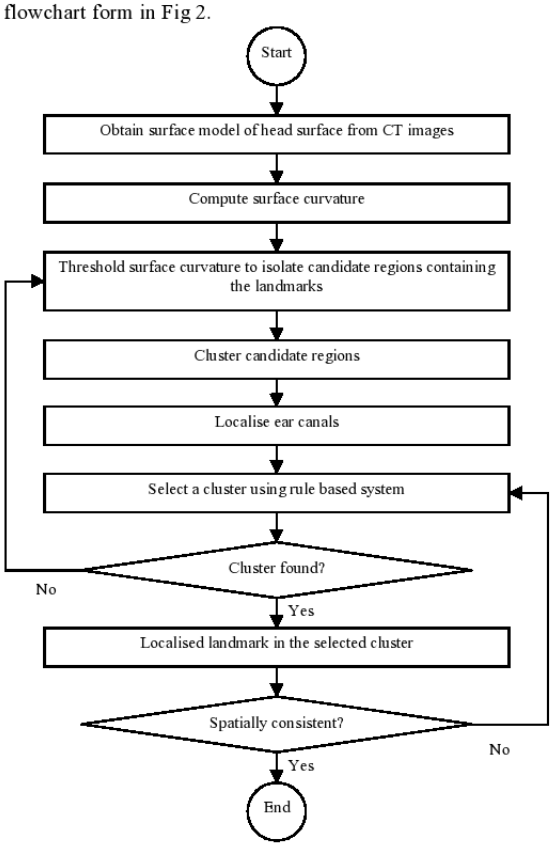


Fig. 2 Flowchart summarising the methodology

A. Landmark surface geometry

Characterising the geometry of the medial canthus and tragus is necessary in developing an algorithm to automatically localise these landmarks. Part of the difficulty in describing the geometry of anatomical landmarks in general is that their definitions are often fuzzy and are therefore difficult to translate into a formal mathematical description [13]. The medial canthus is the most medial point of a depressed region that forms a line which extends from the outer corner to the inner corner of the eye. This region exhibits high positive mean curvature and zero Gaussian curvature. The tragus is approximately semi-ellipsoid, and its most lateral point is the tip of this structure, a locally maximum point. It is characterised by negative mean curvature and high Gaussian curvature.

B. Curvature estimation

There are three approaches to calculate curvature of a surface represented by a mesh: fitting methods, discrete methods, and curvature tensor estimation [14]. In fitting methods, an analytic function, whose curvature can be computed, is fitted to a mesh locally. Fitting is computationally expensive. Discrete methods involve a direct estimation of the curvature at each vertex, by summing the curvature of each face or edge associated with the vertex. Discrete curvature methods are appealing because of their speed. However, they are sensitive to noise and mesh resolution. Curvature tensor estimation is similar to discrete methods, except that instead of estimating the curvature directly, the curvature tensor is estimated. Curvatures and principal directions are derived from the eigenvalues and eigenvectors of the curvature tensor.

Curvature tensor estimation is computationally less complex than fitting methods, although slightly slower than the discrete methods. They are elegant and free of unstable configurations, although certain vertex arrangements produces erroneous curvature estimates [15]. The curvature tensor estimation methodby Cohen-Steiner and Morvan [16] and Alliez et al. [17] is used in this work and is described next. For an arbitrary region,B, on a triangulated surface mesh, the curvature tensor,$\Im$, is the average of individual curvature tensors of edges,e, of the mesh over the region, as shown in Fig 3.

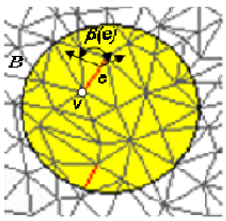


Fig. 3 Notations used in the curvature tensor estimation method

Formally, the curvature tensor, $\Im$, at an arbitrary vertex, v, on the mesh over an arbitrary region, B, is given by:

$$\Im(v) = \frac{1}{|B|} \sum_{edges\ e} \beta(e)|e \cap B|\,\bar{e}\,\bar{e}^t \qquad (1)$$

where $|B|$ is the surface area around the vertex $v$over which the curvature tensor $\Im$ is estimated, $\beta(e)$, is the signed angle between the normals of the two oriented triangles incident to edge e(positive if convex, negative if concave), $|e \cap B|$ is the length of $e \cap B$ (always between 0 and $|e|$), and $\bar{e}$is a unit vector in the same direction as e. A continuous tensor field

over the whole surface is obtained by linearly interpolating the piecewise curvature tensor at each vertex. The principal curvatures, $\kappa_1$and κ_2, are the eigenvalues of the curvature tensor, $\Im$. The Gaussian and mean curvature can be computed from the principal curvature. The Gaussian curvature, K, is the product of the principal curvatures, κ_1 and κ_2:

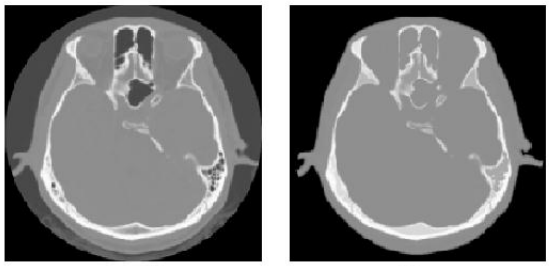
$$K = \kappa_1\kappa_2 \quad (2)$$

The mean curvature, H, is the average of the principal curvatures, κ_1 and κ_2 :

$$H = \tfrac{1}{2}(\kappa_1+\kappa_2) \quad (3)$$

C. Surface model and curvature map

The following are the steps to generate a surface model and curvature map. First, the head region is segmented from CT images using intensity based thresholding. The threshold level is determined automatically using Otsu's method [18]. The segmented images contain holes in the region of interest and small, spurious regions in the background. To remove the holes, a morphological fill operation is applied. To remove the spurious regions, the largest connected binary component is selected. These operations are performed to eliminate additional structures of the extracted surface, which would complicate the automatic detection of the landmarks. The original image (Fig. 4a) is then multiplied with a mask consisting of the segmented image, and a grey level morphological fill operation is performed (Fig. 4b). Then, using a custom isosurface algorithm [19], a surface model of the head is extracted (Fig. 1).



(a) (b)

Fig. 4 (a) Original and (b) segmented axial head CT image

A curvature map of the head surface model is generated using a curvature tensor estimation algorithm described in Section II. The curvature map for the eye region (Fig. 5a) and ear region (Fig. 5b) is illustrated in Fig. 5.

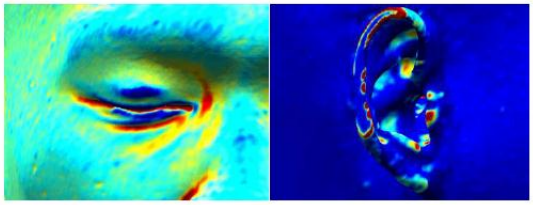
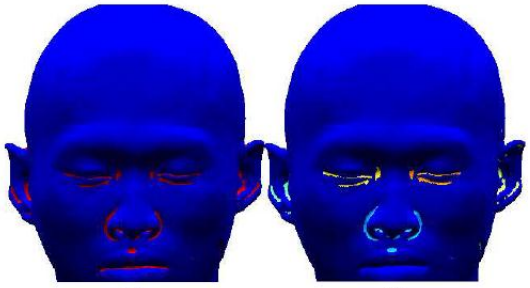


Fig. 5 (a) Curvature map of the eye and (b) ear

Curvature values have been normalised and colour coded for display purposes. Red and blue corresponds to areas of high and low curvature.

D. Clustering

The threshold surface curvature map contains many false regions i.e. regions other than those containing the landmarks (Fig. 6a). Clustering of the regions simplifies the analysis as multiple regions can be grouped and analysed using a divide and conquer approach (Fig. 6b).



(a) (b)

Fig. 6 (a) Threshold surface curvature map (b) clustered regions

As the regions tend to be arbitrarily shaped, a density-based clustering algorithm, which works well with these types of regions, is used. In density based clustering, a cluster is formed if the density of its points exceeds a user defined threshold. The density-based spatial clustering of applications with noise (DBSCAN) algorithm [12] is used. DBSCAN formalises an intuitive notation of "clusters" and "noise" in a clustering application. The algorithm requires two only parameters: a distance measure, and the minimum number of points for a cluster to be formed.

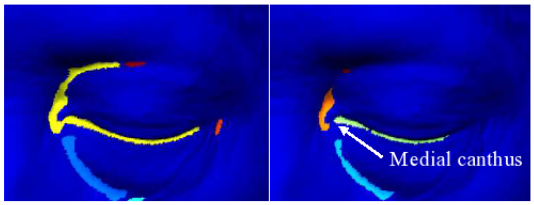
E. Rule-based system

The selection of a cluster containing the landmarks is based on its spatial location and geometric structure. Spatially, thelandmarks should exhibit bilateral symmetry. An allowable range of distance between the landmarks was used as a simple measure of bilateral symmetry. Geometrically, clusters containing the landmarksshould have dimensions within a predefinedrange. Both the distance and allowable dimensions were determined empirically based on our dataset.If the localised landmarks are spatially or geometrically inconsistent, the entire process is repeated using a different cluster and/or a higher surface curvature threshold.

F. Localisation of the medial canthus

Regions of high positive mean curvature and zero Gaussian curvature correspond to candidate locations for the medial canthus. These regions are isolated by thresholding a curvature map of the surface using an empirically derived fixed threshold, and grouped using a density-based algorithm (Fig. 6).The clusters associated with the eye region are identified using nearest neigbour clusteringbased on the standard shape

and anatomy of the human face. The cluster with the maximum horizontal length in the eye region (Fig. 7b), with a horizontal length and vertical height within a predefinedrange,is selected as the most likely cluster containing the medial canthus.The medial canthus corresponds to the most medial point of this cluster. Once both the medial canthus are localised, bilateral symmetrybased on the separation distance between the medial canthi is assessed. If the symmetry condition is not met, the process is repeated using a different cluster and/or different mean curvature threshold. Fig. 7 illustrates the effect of varying the threshold level. Each colour corresponds to an individual cluster.



(a) (b)

Fig. 7 Clusters corresponding (a) $C_{mean} \geq 0.01$ and $C_{gaussian} = 0$ (b) Cmean ≥ 0.015 and $C_{gaussian} = 0$.

In Fig. 7a, although the yellow clusterhas the maximum horizontal length,its vertical length is outside the allowable range. Increasing the threshold yields the cluster (green) containing the medial canthus (Fig. 7b).

Algorithm for the localisation of the medial canthus

Input: Vertices, V_i, of the triangular mesh surface model, the mean curvature, C_{mean}, and Gaussian curvature, $C_{gaussian}$

Algorithm:

1. Find candidate vertices, $v_j \in V_i$, $\{j=1,2,\ldots,n\}$ by thresholding based on mean surface curvature value, $C_{mean} \geq$ threshold and $C_{gaussian} = 0$ (Fig. 6a).
2. Group each vertex, v_j, into clusters, C_i, $\{i=1,2,\ldots.,p\}$ using a density based clustering algorithm (Fig. 6b)
3. Group clusters, $c_j \in C_i$, $\{i=1,2\}$ into two clusters, corresponding to the left and right eye region, using nearest neighbour clustering.
4. Select a cluster, $d_i \in c_j$ $\{i = 1,2\}$ with maximum horizontal length with its horizontal length and vertical height within a predefinedrange (Fig. 7b)
5. Select most medial vertex, $v_{lmc} \in d_1$, for left medial canthus (Fig. 7b) and most medial vertex, $v_{rmc} \in d_2$, for right medial canthus.
6. Check the medial canthus for bilateral symmetry based on distance to the other medial canthus
7. Repeat steps 4 – 6 if step 6 fails, using a different cluster. If step 6 still fails, repeat steps 1 – 6 increasing the threshold.

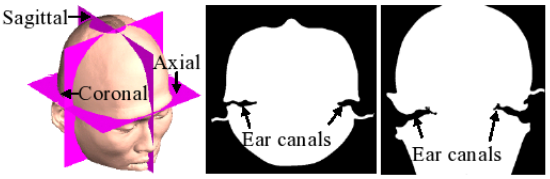
Output: Vertices corresponding to the left and right medial canthus $\{v_{lmc}, v_{rmc}\}$.

G.Localisation of the tragus

Regions of high positive Gaussian curvature and negative mean curvature correspond to candidate locations for the tragus. Occasionally, the curvature map of the ear region contains many false regions because the outer edges of the ear are not always in the field of view and as such, the surface model of the ear region is not well defined (Fig 8).

Fig. 8 Clusters in a thresholded surface curvature map of the ear.

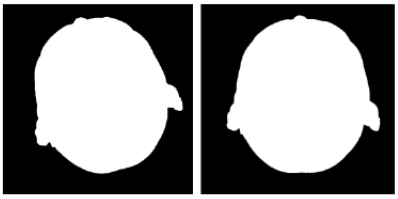
It is therefore necessary to determine an estimate location of the cluster containing the tragus to avoid large positional errors. As the tragus is located close to the entrance of the ear canals and ear canals are prominent features in CT images (Fig. 9b, Fig. 9c),the location of the tragus is found by looking for the presence of ear canals in the CT axial and coronal planes and the outer ear structure in the CT sagittal plane (Fig. 11c).



(a) (b) (c)

Fig.9 (a) Anatomical planes, (b) ear canals in CT axial plane (c) ear canals in CT coronal plane

A necessary step prior to the detection of the ear canals is to reorient the CT images using a rigid body transformation if the head orientation deviates by more than 15° from the forward facing position. The 2D orientation of the head is estimated as the angle between the horizontal axis and the major axis of an ellipse that has the same second-moments as the stacked axial head CT images (Fig. 10).

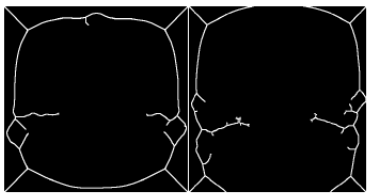


(a) (b)

Fig. 10 (a) Deviation from forward facing position, and (b) reoriented head

The axial and coronal CT planes are then segmented and

skeletonised (Fig. 11). The axial and coronal planes whose skeleton endpoints are closest to the mid-sagittal plane correspond to the axial and coronal CT planes where the tragus is approximately located.



(a) (b)

Fig. 11 Image skeleton for (a) Fig. 8(a) and (b) Fig. 8(b)

The sagittal plane where the tragus is approximately located may be found by scanning from the most lateral to the most medial sagittal plane (Fig. 12) until the Euler number of the largest connected binary region is equal to or is less than zero. In practice, morphological operations need to be performed to avoid false detections.

(a) (b) (c)

Fig. 12 Appearance of outer ear structure in CT sagittal view when scanned lateral to medial (left to right)

The largest region close to the estimated traggus positionis the most likely cluster containing the most lateral point of the tragus. The most lateral point of the tragus corresponds to a point with the highest Gaussian curvature in that cluster.

Algorithm for the localisation of the tragus

Input

1. Vertices, V_i, of the surface model triangular mesh,the mean curvature, C_{mean}, and Gaussian curvature, $C_{gaussian}$

Algorithm:

1. Reorient the head in the axial plane using a rigid transformation if the head orientation deviates more than 15° from the forward facing position (Fig. 10).

2. Estimate spatial location of the tragus. The corresponding axial and coronal planes are the axial and coronal plane with the longest ear canal, and the corresponding sagittal plane is the first sagittal plane (from lateral to medial), with a region whose Euler number is equal to, or less than, zero (Fig. 12c).

3. Find candidate vertices, $v_j \in V_i$ $\{j = 1,2,..,n\}$ by thresholding based on $C_{gaussian} \geq 0.0001$ and $C_{mean} < 0$ (Fig. 8).

4. Group each vertex, v_j, into clusters, C_k, $\{k = 1,2,\ldots,p\}$ using density based clustering (Fig. 8)

5. Select largest cluster, $c \in C_k$, $\{k = 1,2,\ldots,p\}$, within 10 mm from the estimated tragus location.

6. The vertices, v_{lt}, v_{rt}, $\in c$ with the highest Gaussian curvature value correspond to the most lateral point of the left and right tragus respectively.

Output: Vertices, $\{v_{rt}, v_{lt}\}$, corresponding to the most lateral point of the left and right tragus.

III. Results and Discussion

A dataset consisting of near isotropic CT images of 95 patients acquired using a Toshiba Aquilion 16 CT scanner was used to test the algorithm. The images are 512 x 512 pixels with average in-slice pixel spacing and average slice thickness of 0.5 mm each, with no interslice spacing. The start and end scan location is from the skull vertex to at least the skull base, without any gantry tilt (gantry oriented parallel to the infraorbitomeatal line). The DICOM compliant CT images were transferred to a personal computer, and the landmarks were localised automatically.

A curvature map of the head surface model was obtained using the curvature tensor estimation method described in Section II. A neighbourhood B that approximates a disk around v with a radius equal to 1/100th of the diagonal of the bounding box containing the surface was used. To reduce the effect of noise, smoothing was applied to the mesh,although this has the effect of masking surface detail and blurred curvature estimates. The surface curvature map was thresholded using a$C_{mean} \geq 0.010$ and $C_{gaussian} = 0$to isolate candidate regions containing the medial canthus, and $C_{mean} < 0$ and $C_{gaussian} \geq 0.001$ to isolate candidate regions containing the most lateral point of the tragus. The DBSCAN method described in Section IIwas employed to cluster these regions. For the medial canthus, only clusters with a minimum size of 100 vertices (using a neighbourhood radius of 5 mm) were considered. Clusters with less than 100 vertices are abandoned as it is most likely noise. The algorithms described in Section II were used to select candidate clusters containing the medial canthus and the tragus. Estimates of the tragus location (Fig. 13) were found to be always within 10 mm of its actual location.

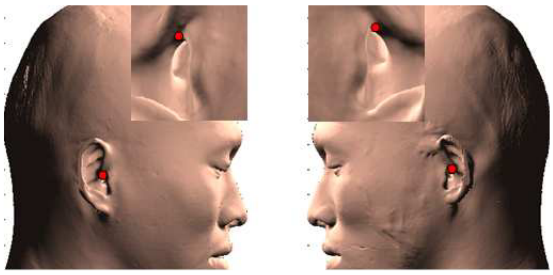


Fig. 13 Estimated tragus position

To check the bilateral symmetry of the left and right medial canthus, a separation distance of 15 - 40 mm, 30 mm and 20 mm in the sagittal, coronal and axial planes was used. For the left and right tragus, a separation distance of 100 - 200 mm, 40 mm and 40 mm was used. The relatively large separation

distance used for the coronal and axial planes is to account for head tilt, which is present in some of the images in the dataset. Because of the large separation distances used, the bilateral symmetry check is useful to prevent large localisations errors only. Fig. 14 illustrates the automatically localised medial canthus (Fig. 14a) and most lateral point of the tragus (Fig. 14b).

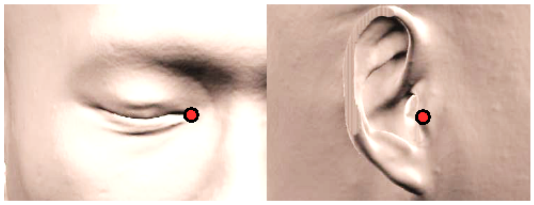


Fig. 14 The automatically localised (a) medial canthus and (b) most lateral point of the tragus

The difference between the locations of the landmarks found manually and automatically using the proposed method is expressed as the root mean square Euclidean distance between the two. Manual localisation was based on the perceived landmark location,the accuracy of which depends on the examiners experience and perception. The mean difference for the medial cantus is 1.5 mm, with a maximum difference of 4.5 mm. The mean difference for the tragus is 0.8 mm with a maximum difference of 2.6 mm.

Although no anatomical ground truth exists, it can be argued that automatic localisation of the most lateral point of the tragus, and to a lesser extent the medial canthus, is more accurate than manual localisation. This is because determining the exact location of the landmarks based on visual inspection is difficult. However, for the medial canthus, there are isolated cases where automatic localisation can produce large localisation errors. This is because the shape of the cluster and the corresponding perceived location of the medial canthus vary greatly depending on the threshold level (Fig. 15).

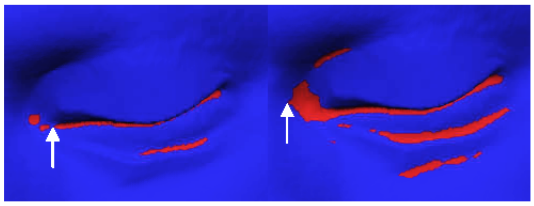


Fig. 15 Perceived location of the medial canthus at different thresholds, (a) $C'_{mean} \geq 0.025$ and $C_{gaussian} = 0$ (b) $C'_{mean} \geq 0.01$ and $C_{gaussian} = 0$

In this work, a fixed threshold (which is incremented if no cluster that satisfies the rule is found) is used. A threshold based on an arbitrary upper limit of a curvature histogram was considered but the presence of outliers due to imperfect segmentation greatly affects the resulting threshold value. For the majority of the images, the cluster containing the medial canthus is usually robust to the threshold level.

As we are developing this technique for three neurosurgical procedures normally performed without image-guidance i.e. using a freehand technique, the difference in localisation using a manual and automated technique is acceptable and is within our required 5 mm clinical accuracy [3, 4]. Images of four patients were excluded from the analysis as the medial canthus was not well defined in their CT images. Future work involves refining the algorithm to perform cluster analysis e.g. cluster shape, to prevent localisation errors such as those illustrates in Fig. 15.

IV. Conclusion

A methodology to automatically localise the medial canthus and tragus in CT images was presented. The basic approach was to exploit the surface curvature properties of the medial canthus and tragus. A rule system based on prior knowledge of the landmark geometric structure and spatial location was used to facilitate localisation, as surface curvature properties alone results in many false detections. The medial canthus and tragus can be localised on CT images, with performance comparable to manual localisation, based on the approach presented.

References

[1] P. A. Woerdeman, P. W. Willems, H. J. Noordmans, C. A. Tulleken and J. W. van der Sprenkel, "Application accuracy in frameless image-guided neurosurgery: A comparison study of three patient-to-image registration methods," J. Neurosurg., vol. 106, pp. 1012-1016, Jun, 2007.

[2] S. Wolfsberger, K. Rossler, R. Regatschnig and K. Ungersbock, "Anatomical landmarks for image registration in frameless stereotactic neuronavigation," Neurosurg. Rev., vol. 25, pp. 68-72, Mar, 2002.

[3] M. Gooroochurn, M. Ovinis, D. Kerr, K. Bouazza-Marouf and M. Vloeberghs, "A Registration Framework for Preoperative CT to Intraoperative White Light Images," Proc. Medical Image Understanding and Analysis, pp. 184-188, 2009.

[4] M. Gooroochurn, D. Kerr, K. Bouazza-Marouf, M.Ovinis, " Facial recognition techniques applied to the automated registration of patients in the emergency treatment of head injuries," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine (in press).

[5] M. Gooroochurn, D. Kerr, K. Bouazza-Marouf and M. Vloeberghs, "Extraction of Craniofacial Landmarks for Preoperative to Intraoperative Registration," in Sixth International Conference on Signal and Image Processing, .

[6] S. Frantz, K. Rohr and H. S. Stiehl, "Localization of 3D Anatomical Point Landmarks in 3D Tomographic Images Using Deformable Models," Lecture Notes In Computer Science, pp. 492-501, 2000.

[7] S. Frantz, K. Rohr and H. S. Stiehl, "Development and validation of a multi-step approach to improved detection of 3D point landmarks in tomographic images," Image Vision Comput., vol. 23, pp. 956-971, 2005.

[8] S. Wörz and K. Rohr, "Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models," Med. Image Anal., vol. 10, pp. 41-58, 2006.

[9] K. Rohr, H. Stiehl, R. Sprengel, W. Beil, T. Buzug, J. Weese and M. Kuhn, "Point-based elastic registration of medical image data using approximating thin-plate splines," in Visualization in Biomedical Computing, 1996, pp. 297-306.

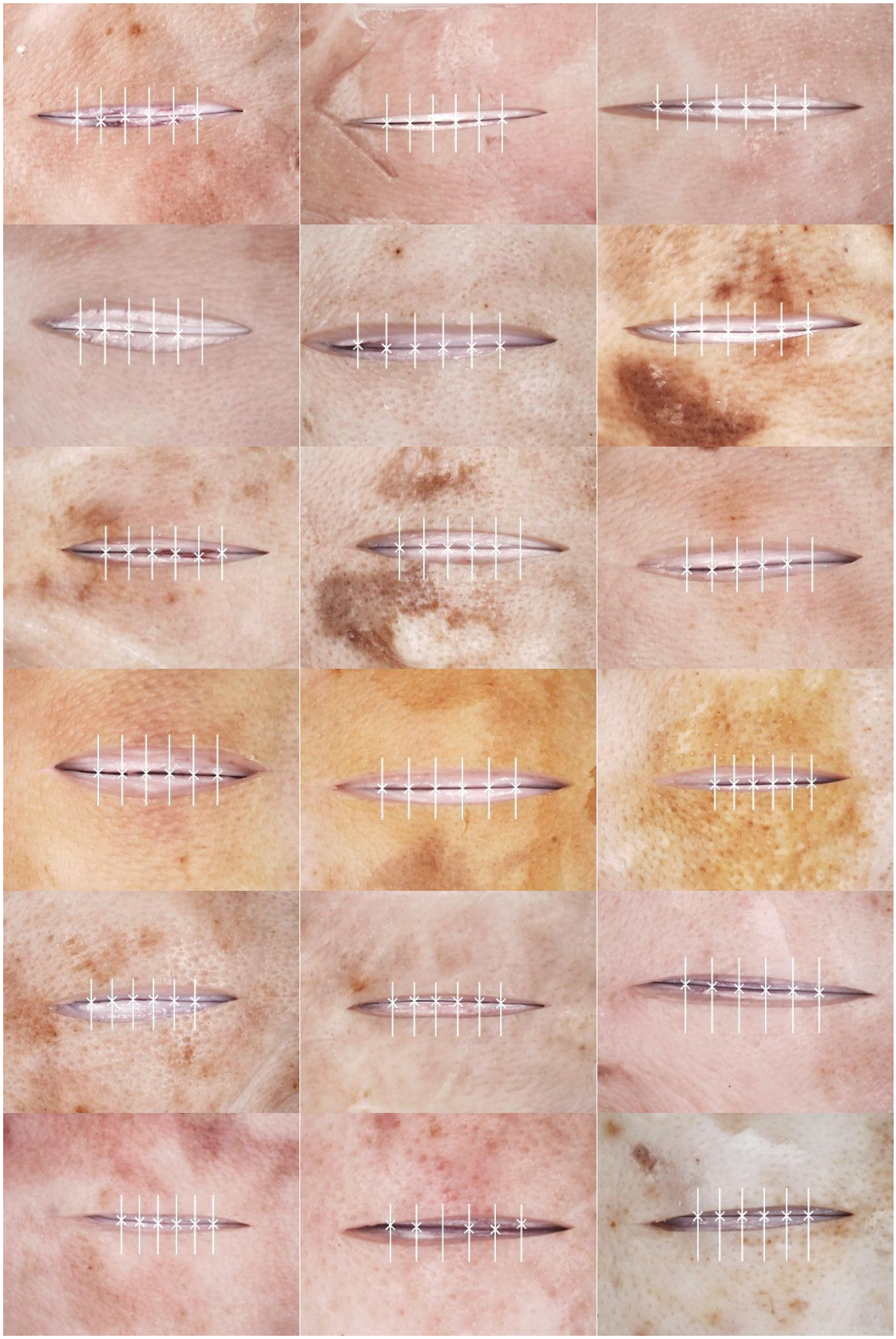
[10] M. Alker, S. Frantz, K. Rohr and H. Siegfried Stiehl, "Improving the robustness in extracting 3D point landmarks from 3D medical images using parametric deformable models," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001, 2001, pp. 582-590.

[11] K. Subburaj, B. Ravi and M. Agarwal, "Automated identification of anatomical landmarks on 3D bone models reconstructed from CT scan images," Comput. Med. Imaging Graphics, vol. 33, pp. 359-368, 2009.

[12] M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, Portland, OR, AAAI Press, 1996, pp. 226-231.
[13] P. Li, B. D. Corner and S. Paquette, "Automatic landmark extraction from three-dimensional head scan data," in Proceedings of SPIE, 2002, pp. 169.
[14] T. D. Gatzke and C. M. Grimm, "Estimating curvature on triangular meshes," Int. J. Shape Model., vol. 12, pp. 1, 2006.
[15] S. Rusinkiewicz, "Estimating curvatures and their derivatives on triangle meshes," in 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on, 2004, pp. 486-493.
[16] D. Cohen-Steiner and J. M. Morvan, "Restricted delaunay triangulations and normal cycle," in Proceedings of the Nineteenth Annual Symposium on Computational Geometry, 2003, pp. 312-321.
[17] P. Alliez, D. Cohen-Steiner, O. Devillers, B. Lévy and M. Desbrun, "Anisotropic polygonal remeshing," 2003.
[18] N. Otsu, "A threshold selection method from gray-level histograms," Automatica, vol. 11, pp. 285-296, 1975.
[19] The MathWorks, Source Code for the Matlab Isosurface Function. 2007.

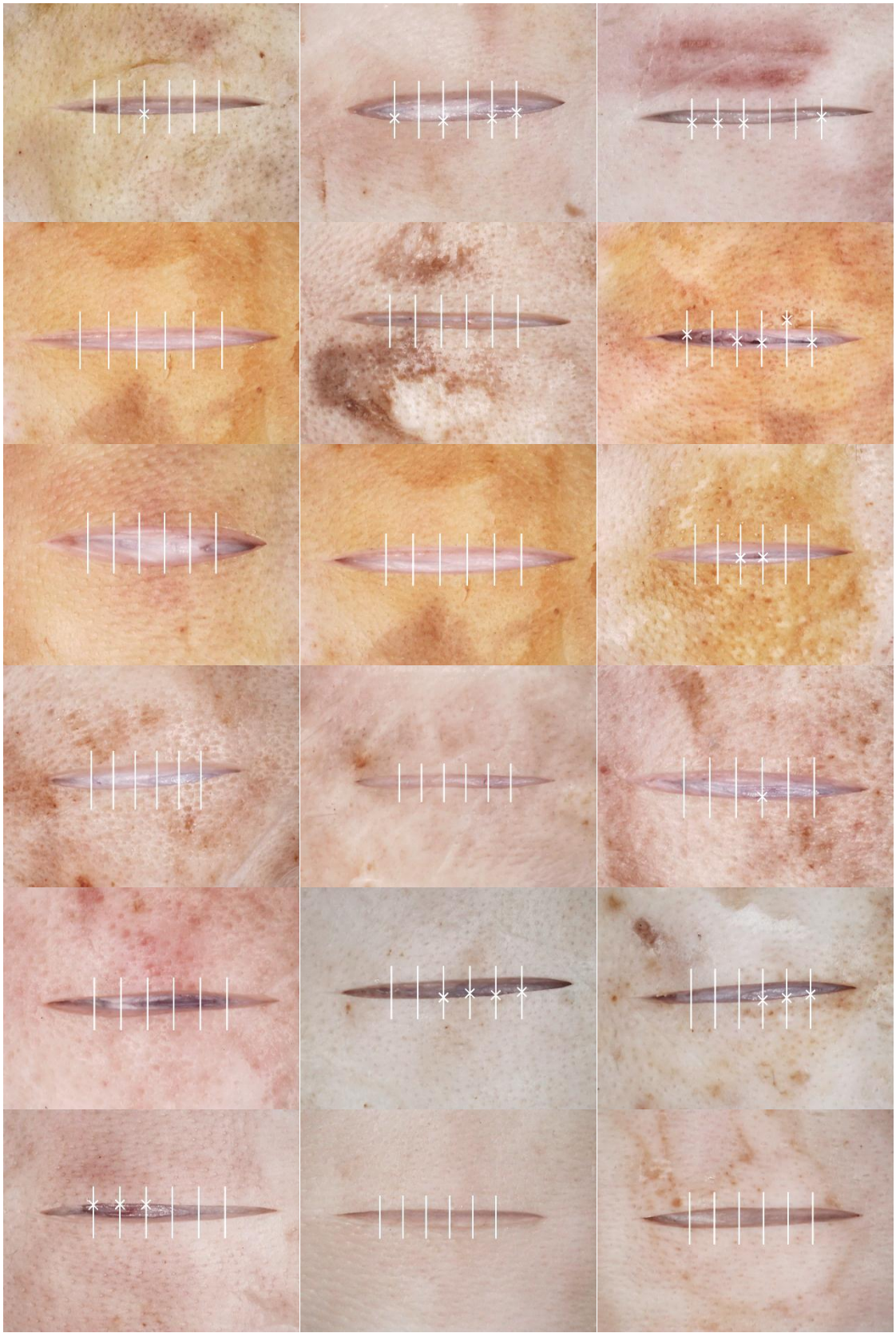
Appendix B

Illustrative examples of the output of the complete incision detection method for complete incisions



Appendix C

Illustrative examples of the output of the complete incision detection method for marginal incisions



Appendix D

Illustrative examples of a correctly fitted circle within a retracted incision