

Optimised Calibration, Registration and Tracking for Image Enhanced Surgical Navigation in ENT Operations

Gerardo González García

January 2011

A thesis submitted to the School of Computing Sciences
University of East Anglia
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy in Computer Science

©This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior, written consent.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Rudy Lapeer, for his support throughout the development of this research project. Without his patience, insightful guidance and helpful advice this PhD thesis would not have been written.

I would also like to thank Dr. Min Si Chen, Dr. Paul Gasson, Dr. Roger Rowland, Vilius Audinis and Said Shah who, at different stages of my PhD, provided me with ideas about work and great general conversations.

I am most grateful to the Mexican National Council for Science and Technology (CONACYT) for the financial support provided. Finally, I am indebted to my family who have supported me in this venture.

Abstract

The development of different computer technologies and software methods has contributed to a large number of applications in the medical visualisation field. Two of the technologies that have rapidly evolved within the operating theatre are known as augmented reality and surgical navigation. The former aims to superimpose in real time computer-generated models on top of images of a real scene acquired by optical devices. This results in an image-enhanced view of the real world. In the case of surgical navigation, it allows the surgeon to identify the location of surgical instruments that are inserted in a patient's body. The combination of both technologies is known as image-enhanced surgical navigation, in which the digitally-reconstructed anatomy of a patient is overlaid on real images captured by optical devices such as an endoscope or a surgical microscope.

This thesis is focused on the importance of the accuracy of different stages required to produce an overlay in an image-enhanced surgical navigation system targeted to ear, nose and throat interventions. These stages comprise camera calibration, registration between the patient's real anatomy and its virtual counterpart, and motion tracking. A series of optimisations are presented that improve the accuracy of each stage based on the use of a rigid endoscope and a stereoscopic surgical microscope. In particular, improvements are first made to the underlying image-enhanced surgical navigation system, regarding endoscopic camera calibration, motion tracking and stereoscopic visualisation. This is followed by a method to optimise the focal length and consequently improve the accuracy of the final calibration error using a pair of cameras connected to a surgical microscope. Finally, a technique is introduced to correct the misregistration between real and virtual anatomical structures of a patient as observed by the surgical microscope. This technique can be used when the problem is caused by accumulated positional errors produced by a motion tracking device during the intervention.

List of publications in conference proceedings and journals

During the course of this research, the author has contributed to the following publications in journal and conferences:

G. González-García and R.J. Lapeer. Optimisation of focal length using a stereoscopic operating microscope for augmented reality surgical guidance. In *Proceedings of the 5th International Workshop on Medical Imaging and Augmented Reality (MIAR '10)*, pages 551 - 560, 2010.

G. Gonzalez-Garcia and R. Lapeer. Using photo-consistency for intra-operative registration in image enhanced surgical navigation (IESN). In *Proceedings of AMI-ARCS Workshop in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '09)*, pages 94 - 101, 2009.

G. Gonzalez and R.J. Lapeer. An evaluation of photo-consistency for intra-operative registration in an image enhanced surgical navigation (IESN) system. In *Proceedings of Medical Image Understanding and Analysis Conference (MIUA '09)*, pages 229 - 233, 2009.

R. Lapeer, M.S. Chen, and G. Gonzalez. A brief account on 3D image enhanced surgical navigation (IESN). In *Proceedings of Knowledge Transfer through Partnership: Delivering Innovation in Healthcare*, 2008.

R. Lapeer, M.S. Chen, G. Gonzalez, A. Linney, and G. Alusi. Image-enhanced surgical navigation for endoscopic sinus surgery: Evaluating calibration, registration and tracking. *International Journal of Medical Robotics and Computer Assisted Surgery (IJMRCAS)*, 4(1):32 - 45, 2008.

M.S. Chen, G. Gonzalez, and R. Lapeer. Intra-operative registration for image enhanced endoscopic sinus surgery using photo-consistency. *Studies in Health Technology and Informatics*, volume 125, pages 67 - 72, 2007.

Contents

	Page
1 Introduction	1
1.1 Research objective	4
1.1.1 Targeted ENT procedures	5
1.2 Relevance of the research work and main contributions	6
1.3 Thesis outline	9
2 Background	11
2.1 Augmented Reality	11
2.2 Previous work in medical AR	16
2.3 Classification of components in AR	19
2.3.1 Display technologies	19
2.3.2 Motion tracking devices	23
2.4 Methods in AR	26
2.5 Summary	29
3 General optimisation within the IESN system	30
3.1 Introduction	30
3.2 Marker detection for endoscopic camera calibration	31
3.2.1 Experiment	37

3.3	Use of an optical tracking device within the IESN system	40
3.3.1	Experiment	43
3.4	Data synchronisation of motion tracking in the IESN system	46
3.4.1	Experiment	48
3.5	Stereoscopic visualisation of virtual models in the IESN system	54
3.6	Summary	59
4	Camera calibration	61
4.1	Introduction	61
4.2	Methodology	62
4.2.1	Camera calibration	62
4.2.2	Classification of camera calibration techniques	67
4.2.3	Stereo camera calibration	69
4.2.4	Estimation of intrinsic parameters in stereo cameras	76
4.2.5	Focal length optimisation in a stereo camera setup	84
4.2.6	Camera pre-calibration using a non-coplanar calibration object	86
4.3	Experimental Results	90
4.3.1	Analysis of single camera calibration	90
4.3.2	Analysis of stereo camera calibration	97
4.4	Summary	109
5	Intra-operative registration	110
5.1	Introduction	110
5.2	Image Registration	111
5.2.1	Classification of medical image registration methods	112
5.2.2	Image registration for SN in head and neck surgery	114
5.2.3	Overlay accuracy in IESN systems	115

5.3	Methodology	115
5.3.1	Intra-operative registration for ENT procedures	115
5.3.2	Estimation of photo-consistency based cost function	117
5.3.3	Photo-consistency based registration	119
5.4	Experiments and results	124
5.4.1	Registration of models in a simulated environment	124
5.4.2	Analysis of convergence of optimisation methods	130
5.4.3	Intra-operative registration laboratory test	136
5.5	Summary	149
6	Conclusions and future work	151
6.1	Summary and conclusion	151
6.2	System integration	154
6.3	Further work	158
A	Results from the focal length optimisation experiments	161
A.1	Accuracy of focal length estimation in a synthetic environment — Graphs	161
A.2	Focal length optimisation in a pair of real cameras — Graphs	168
A.3	Focal length optimisation in a pair of real cameras — Tables	178
B	Optimisation Algorithms	182
B.1	Evolutionary algorithms	182
B.1.1	Differential Evolution	183
B.1.2	Use of the optimisation algorithms	188
B.2	Experiments and results	189
B.2.1	Validation of the optimisation algorithms	189

List of Figures

Figure	Page
1.1 Relation of Computer-Assisted Surgery systems in different operating stages.	3
2.1 Virtuality continuum as described by Milgram and Kishino [24].	12
2.2 Examples of augmented reality based surgical navigation.	14
3.1 Physical equipment used in the IESN system for ESS purposes.	32
3.2 Example of segmented figures during the feature detection process, where fractured regions lead to independent bounding boxes within the same feature marker.	34
3.3 Example of connected components during feature segmentation.	35
3.4 Comparison of endoscope calibration before and after reduction of feature segmentation.	39
3.5 Optical tracking devices used during the research.	41
3.6 Illustration of a tracked object that is parallel to the xy -plane of the tracking device.	42
3.7 Typical setup during the intervention in which the tracking device is oriented towards the surgical area. Patient and tracker coordinates are not aligned.	44
3.8 Eigenvector defined as the normal vector \vec{n} orthogonal to the set of data points.	45
3.9 Human skull used to simulate a real patient. A set of passive markers are attached to a mouthpiece to register the skull's position over time.	49

3.10	Relative transformation among the coordinate systems used to track simultaneously an endoscope and a human skull.	50
3.11	Motion tracking along the T_z component with no data synchronisation. . .	52
3.12	Motion tracking along the T_z component with data synchronisation. . . .	53
3.13	Schematic representation of stereoscopic technologies.	55
3.14	Comparison of original and improved 3D stereoscopic visualisation. . . .	56
3.15	Stereo devices used during this research	58
3.16	3D volumetric datasets used for qualitative observations in stereoscopic visualisation.	59
4.1	Geometry of the pinhole camera model.	63
4.2	Geometry of similar triangles in the projective camera model.	65
4.3	Radial distortion effects.	67
4.4	Illustration of corresponding points on two independent views.	69
4.5	Diagram of epipolar geometry for two cameras.	70
4.6	Critical configurations which prevent the calculation of focal length. . . .	82
4.7	Critical configuration when the epipolar line lm of one camera passes through the principal point of the second one.	83
4.8	Methods to avoid critical camera configurations as proposed by Sturm et al. [126, 127].	83
4.9	3D non-coplanar calibration object manufactured with 3D Zprinter [©] 450 - Z Corporation.	87
4.10	Method to detect the centre of elliptical figures	88
4.11	Detection of marker centres on the non-coplanar calibration object. . . .	89
4.12	Rotational gauge used to control the calibration object inclination.	91
4.13	Focal length estimation through detection of ellipse centres.	92
4.14	Absolute errors in localisation of image markers under different noise levels.	95
4.15	Absolute errors in localisation of 3D marker points under different inaccuracy levels.	96

4.16	Schematic representation of a synthetic scenario using a set of points at different depth distances from a pair of virtual cameras.	98
4.17	Stereoscopic surgical microscope and rotational gauge experimental setup.	101
4.18	Absolute errors using the linear technique for fundamental matrix calculation.	106
4.19	Absolute errors using M-estimators for fundamental matrix calculation.	107
4.20	Absolute errors using the gradient-based technique for fundamental matrix calculation.	108
5.1	Stages within the pre-operative registration.	120
5.2	Selection of voxels by backprojecting screen pixels.	121
5.3	Occlusion detection through a forward projection ray.	122
5.4	Selection of voxel-sized sub-windows around the skull orbit.	123
5.5	Registration errors for six independent DOFs based on the PC cost function using a full-sized CT model.	127
5.6	Registration errors for six independent DOFs based on the $PC_{squared}$ cost function using a full-sized CT model.	128
5.7	Registration errors for six independent DOFs based on the $PC_{inverse}$ cost function using a full-sized CT model.	129
5.8	Overall convergence in PC photo-consistency based cost function using Powell's method.	133
5.9	Overall convergence in PC photo-consistency based cost function using the SDE algorithm.	134
5.10	Overall convergence in PC photo-consistency based cost function using the CODEQ algorithm.	135
5.11	Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a $T_x T_y$ misalignment.	146
5.12	Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a $T_x T_y T_z$ misalignment.	147

5.13	Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a R_z misalignment.	147
5.14	Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a $T_xT_yR_z$ misalignment.	148
5.15	Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over six DOFs.	148
A.1	Percentage errors in focal length calculation using Bougnoux's Method. . .	163
A.2	Percentage errors in focal length calculation using Sturm's Method. . . .	165
A.3	Percentage errors in focal length calculation using Newsam's Method. . .	167
A.4	Linear method for the fundamental matrix calculation and Levenberg-Marquardt optimisation algorithm.	169
A.5	Linear method for the fundamental matrix calculation and SDE optimisation.	170
A.6	Linear method for the fundamental matrix calculation and CODEQ optimisation.	171
A.7	Fundamental matrix calculation using M-estimators and Levenberg-Marquardt optimisation algorithm.	172
A.8	Fundamental matrix calculation using M-estimators and SDE optimisation. . .	173
A.9	Fundamental matrix calculation using M-estimators and CODEQ optimisation.	174
A.10	Fundamental matrix calculation using gradient-based method and Levenberg-Marquardt optimisation algorithm.	175
A.11	Fundamental matrix calculation using gradient-based method and SDE optimisation.	176
A.12	Fundamental matrix calculation using gradient-based method and CODEQ optimisation.	177
B.1	Schematic diagram of the DE algorithm.	185

List of Tables

Table	Page
2.1 Commercial optical tracking devices.	26
2.2 Examples of commercial SN systems.	27
4.1 Focal length and calibration errors mean and standard deviation (\pm SD) of single camera calibration at different inclination angles using ellipse centres over ten trials.	92
4.2 Mean and standard deviation (\pm SD) values (over ten trials) of focal length, T_z distance and ratio between focal length and T_z at different inclination angles.	93
4.3 Mean and standard deviation (\pm SD) values for five different DOFs in a single camera calibration at different inclination angles over ten trials. . .	93
4.4 List of methods used for the optimisation of the focal length.	102
4.5 Summary of statistically significant results of the hypothesis that two optimisation methods display different calibration error values.	105
5.1 Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x and T_y axes using the object-to-camera coordinate transformation.	139
5.2 Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x , T_y and T_z axes using the object-to-camera coordinate transformation.	139
5.3 Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over R_z axis using the object-to-camera coordinate transformation.	140

5.4	Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x , T_y and R_z axes using the object-to-camera coordinate transformation.	140
5.5	Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over 6 DOFs using the object-to-camera coordinate transformation.	141
5.6	Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x and T_y axes using the object coordinate system.	143
5.7	Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x , T_y and T_z axes using the object coordinate system.	143
5.8	Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over R_z axis using the object coordinate system.	144
5.9	Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x , T_y and R_z axes using the object coordinate system.	144
5.10	Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over 6 DOFs using the object coordinate system.	144
6.1	Selected methods for camera calibration.	155
6.2	Selected options for intra-operative registration.	156
A.1	Statistical significance of hypothesis that two optimisation methods display different calibration error values using the linear method for fundamental matrix at 20° convergence angle (bad accuracy).	179
A.2	Statistical significance of hypothesis that two optimisation methods display different calibration error values using the linear method for fundamental matrix at 25° convergence angle (good accuracy).	179
A.3	Statistical significance of hypothesis that two optimisation methods display different calibration error values using M-estimators at 5° convergence angle (good accuracy).	180

A.4 Statistical significance of hypothesis that two optimisation methods display different calibration error values using M-estimators at 15° convergence angle (bad accuracy). 180

A.5 Statistical significance of hypothesis that two optimisation methods display different calibration error values using the gradient-based method at 15° convergence angle (bad accuracy). 181

A.6 Statistical significance of hypothesis that two optimisation methods display different calibration error values using the gradient-based method at 25° convergence angle (good accuracy). 181

B.1 Comparison between mean and standard deviation (\pm SD) of the function optimisation results reported in the literature [160, 162] (Literature) and the author’s own implementation (Author), for each of the three algorithms, i.e. DE, SDE and CODEQ. 191

Acronyms

AC	Alternating Current
AR	Augmented Reality
CAD	Computer-Aided Design
CAS	Computer-Assisted Surgery
CCD	Charge-Coupled Device
CMM	Coordinate Measuring Machine
CRT	Cathode Ray Tube
CT	Computed Tomography
DC	Direct Current
DE	Differential Evolution
DOF	Degrees of Freedom
EA	Evolutionary Algorithm
ENT	Ear, Nose and Throat
ESS	Endoscopic Sinus Surgery
FIR	Finite Impulse Response
FOV	Field of View
GPU	Graphics Processing Unit
HMD	Head Mounted Display
HMPD	Head-Mounted Projective Display
ICP	Iterative Closest Point
IESN	Image-Enhanced Surgical Navigation

IGS	Image-Guided Surgery
IREC	Infrared Emitting Diode
LCD	Liquid Crystal Display
LED	Light Emitting Diode
LM	Levenberg-Marquard
MFC	Microsoft Foundation Classes
MIS	Minimally Invasive Surgery
MRI	Magnetic Resonance Imaging
MV	Monoscopic Video
PDA	Personal Digital Assistant
Pixel	Picture element
RMS	Root Mean Square
RMSD	Root Mean Squared Distance
ROI	Region of Interest
SD	Standard Deviation
SDE	Self-adaptive Differential Evolution
SG	Stereoscopic Graphics
SN	Surgical Navigation
SV	Stereoscopic Video
SVD	Singular Value Decomposition
VGA	Video Graphics Array
Voxel	Volume element
VR	Virtual Reality
VRD	Virtual Retinal Display

Chapter 1

Introduction

Since the introduction of computing science in the field of medicine a few decades ago, the means by which practitioners diagnose and treat diseases have changed considerably. Traditional methods required that doctors learnt surgical procedures based on information found in books or a patient's particular X-ray image data. Nowadays, doctors rely on technological developments to perform a great number of surgical procedures. These include the use of robotic arms to execute minuscule movements, obtain visual feedback from a computer in real-time, and remotely collaborate in an operation, among others.

The technological approach of using computer systems in the operating room is known as computer-assisted surgery or computer-aided surgery (CAS). According to the International Society for Computer-Aided Surgery, the range of CAS "encompasses all fields within surgery, as well as biomedical imaging and instrumentation, and digital technology employed as an adjunct to imaging in diagnosis, therapeutics, and surgery." [1]. The main objective of CAS within the operating theatre is aimed at improving precision levels during the surgical intervention and to shorten operation times and, most importantly, to reduce patient recovery times. Diverse CAS systems are implemented according to different surgical phases. The following describe some uses of computer-based techniques in the operative stages:

- *Pre-operative stage.* This phase includes an analysis of the patient's anatomy around specific areas or the entire body. A set of X-ray computed tomography (CT) or magnetic resonance imaging (MRI) scans are obtained and stored in a database. With the help of a computer, these images can be used to generate three-dimensional models of a patient's body. The practitioner can navigate within the representation and plan a surgical procedure.
- *Intra-operative stage.* The patient information is constantly acquired, updated and presented to the doctor during surgery in order to reflect any changes as the intervention progresses. This stage can be considered as the most critical period because any failure in the system could lead to serious consequences in the patient's health. Time delays and lack of precision are among the issues that CAS applications may present in the operating room.
- *Post-operative stage.* Once surgery has finished, the doctor can analyse the results using an up-to-date reconstructed 3D model of the patient. This new model can also be compared to other patients' results to evaluate the recovery process. Also, the surgical procedure can be recorded and presented to the same doctor or to different practitioners in order to evaluate their performance.

Figure 1.1 illustrates the use of CAS applications in different surgical phases.

There are several technologies that can be used to help a doctor visualise the patient's anatomy in the different operative phases. Virtual reality (VR) and augmented reality (AR) technologies are among the latest advances in the field. In VR, different virtualisation levels can be employed to represent three-dimensional models [3]. The basic level displays 3D objects on flat surfaces such as PC monitors, in which the user can control a model through simple devices (i.e. mouse and keyboard) or more specialised tools (i.e. haptic devices). In higher virtualisation levels, the user can be enclosed in an entirely computer-generated world with no possibility to perceive the real world. This is known

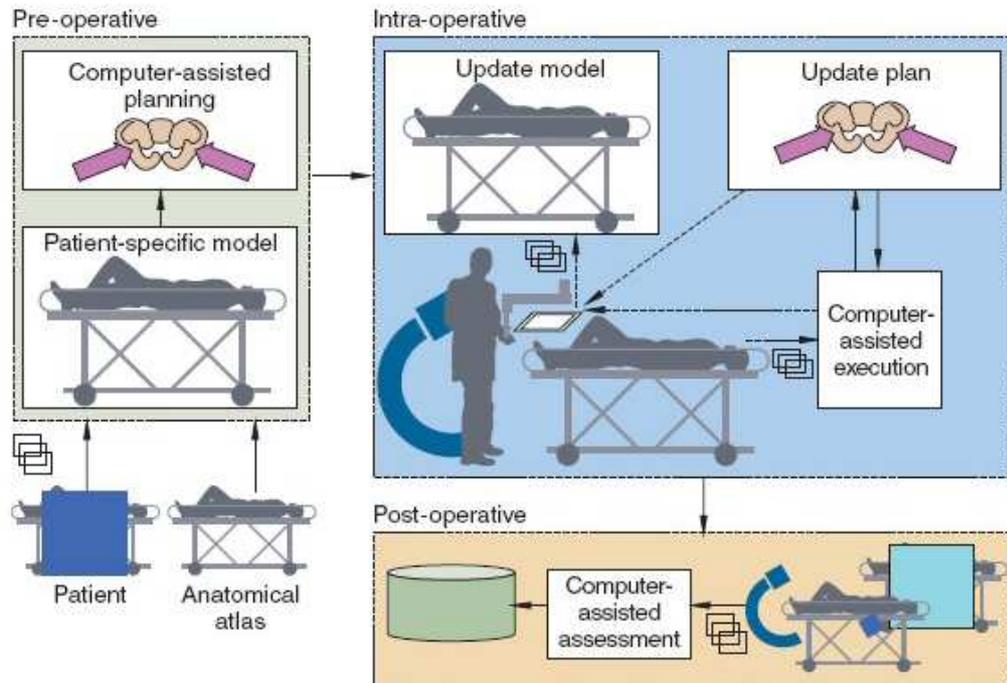


Figure 1.1: Relation of Computer-Assisted Surgery systems in different operating stages [2].

as a fully-immersive environment. In either level, VR can be employed to train surgeons in specific procedures, diagnose and plan a surgical intervention or evaluate the results after operation.

Whereas VR exclusively presents virtual models to the user, AR combines real and virtual worlds in the same scenario. This is achieved by superimposing virtual images on the user's world view. The enhanced vision allows the user to obtain extra information on screen about specific elements of the real world. In the case of medical applications, visual cues can be used to assist the practitioner during surgery in real time. Thus, the visual perception of the patient's anatomy can be extended through the insertion of computer-generated images of bones or organs corresponding to the subject. These images are usually extracted in a pre-operative stage by scanning the patient through CT or MRI modalities.

Surgical navigation (SN) systems are used to provide information about the global location of surgical tools introduced in the patient's body during the intra-operative stage [4].

Specifically, when a surgeon inserts an instrument, such as an endoscope, the computer system displays on a computer monitor the exact position of the instrument tip with respect to the patient's anatomical structure. Therefore, the system allows the surgeon to guide or navigate a surgical tool through the human body. SN can also be referred to as image-guided surgery (IGS).

The combination of AR, also known as image enhancement (IE), and SN systems yields a technology denominated image-enhanced surgical navigation (IESN). IESN systems allow the surgeon to navigate an instrument while visualising superimposed virtual models on real images. As these virtual models correspond to internal structures lying underneath the observed anatomy, it is said that the surgeon acquires so-called "X-ray vision" during the intervention. IESN systems are typically employed for minimally invasive surgery (MIS) or surgical operations in which additional optical devices are required (e.g. surgical microscope).

1.1 Research objective

As in any other computer-assisted application, technology must be used to extend the user's abilities rather than to replace them. Therefore, IESN systems are not intended to make surgeons more competent within the operating theatre, but to assist them and improve their medical accuracy during the intervention. The use of an IESN system may give a surgeon confidence in the procedure. However, the surgeon must confirm that the accuracy of the system is maintained at all times [5].

One of the factors that influence the success or failure of an IESN system consists in the accuracy to overlay real and virtual images throughout the surgical intervention. The accuracy relies on the technologies employed to acquire information from the real environment and the software methods required to relate the real world with its virtual

counterpart.

Depending on the particular surgical application, the maximum accuracy achieved by an IESN system will differ due to the observed patient's anatomy. In specialties that deal with organs and soft-tissue areas, the average accuracy obtained is within 10 mm [6–10]. This is mainly because these areas are deformed during the intervention as a result of respiration and heartbeat. Therefore, shape differences are constantly found between pre- and intra-operative anatomies. In the case of ear, nose and throat (ENT) procedures, the target areas involve bony structures and tissue adjacent to bones that tend to maintain their shape throughout the operation. Because there is little difference between the scanned and real anatomies, higher accuracies in the IESN system are obtained. Specifically, the overall accuracy achieved in ENT surgery is in the range of 2 - 3 mm [11–16]. Nevertheless, the tolerance error for each particular procedure has to be decided during the intervention [7].

The objective of this thesis is to optimise different aspects of an IESN system that influence the overlay accuracy between pre- and intra-operative images. The research is focused on surgical interventions for ENT using a stereoscopic surgical microscope and a rigid zero-degree endoscope. A previously developed IESN application called ARView [17] served as the basis for the research described in this thesis.

1.1.1 Targeted ENT procedures

Two particular ENT surgical procedures are targeted as potential applications during this research: sinus surgery and mastoidectomy. The former involves the treatment of sinusitis — the inflammation of the sinus cavities surrounding the nose — as a resource when medication or other treatment options prove to be ineffective in relieving the symptoms [12, 13]. Sinus surgery is typically carried out using an endoscope that is inserted

into the nasal openings in order to remove the mucus and treat the cause of the illness. This type of procedure is called endoscopic sinus surgery (ESS). In cases of persistent symptoms after an endoscopic intervention, a surgical microscope can be used to access the sinus cavities through incisions on the skin [18]. It must be mentioned that, in comparison to an endoscope, the use of a surgical microscope provides a stereoscopic view that enhances depth perception during the operation. For this reason, in some cases surgeons prefer the use of direct microscopy as the rate of incidence of the disease is comparable to that of an endoscopic approach [19].

In the case of mastoidectomy, the objective of the intervention is to eliminate the infection of the mastoid bone, which is a bone situated behind the ear at the base of the skull. This operation it is often performed when medication is not effective, as in the case of sinusitis. The procedure involves the removal of the mastoid bone using a special drill and observed through a surgical microscope. Other reasons to carry out mastoidectomy include providing a path for interventions in the lateral skull, such as the allocation of hearing implants [20]. The main risks associated to this type of operation include detecting and avoiding the facial nerve and surrounding jugular veins [21].

It should be noted that other procedures that rely on bony structures in ENT and skull base surgery could also find application by this research. This is true in the case that the overlay is aimed at rigid-body anatomies, which are assumed to be non-deformable during the surgical intervention.

1.2 Relevance of the research work and main contributions

As described previously, the accuracy of an IESN system differs on the type of surgical procedure and the underlying anatomical structures. For general IESN applications, the

overlay accuracy between real and virtual models is affected by the deformations that occur in the patient's anatomy as organs and soft tissues move during the intervention. The use of IESN is simplified in surgical specialties where non-deformable structures are involved, such as ENT procedures. However, the problem of generating an accurate overlay between rigid real and virtual anatomical models has not been satisfactorily solved.

The procedure to superimpose virtual imagery on the view of the real world is based on a workflow that consists of four basic steps: camera calibration, registration, motion tracking and visualisation. In the first step, the camera used to capture the view of the surgical scene is calibrated in order to determine its optical properties and its location in the real world. This is followed by the registration phase that aligns both the real and virtual anatomical models, producing an initial static overlay between them. Subsequently, the camera and/or patient are tracked to reflect their movements during the operation. The final step involves the visualisation of the AR scene through display devices.

The original contributions provided by this research are mainly focused on the optimisation of the first two steps in the workflow. The selection of these two steps rely on the fact that they are the main stages in which the development of new software-based methods can improve the accuracy of the overall overlay. Being targeted to ENT procedures, the different experiments performed throughout this research seek to achieve higher accuracy levels than the accuracy currently obtained by general IESN systems. The experiments also aim to evaluate external factors that independently affect the system accuracy in the stages of camera calibration and intra-operative registration. As the experiments involve the use of a surgical microscope, the levels of magnification produced by the optical device must be taken into consideration.

With respect to the first step of the workflow, a new hybrid technique is introduced that combines a photogrammetric calibration procedure for one camera and a self-calibration method for a stereo pair of cameras. This hybrid procedure allows optimising the estima-

tion of the focal length of the cameras and subsequently improving the accuracy of the calibration in the stereoscopic microscope. A new 3D calibration object is specifically designed and manufactured for its use in surgical microscopy.

Concerning registration, this research presents a method that intra-operatively aligns virtual and real models when the overlay is affected by accumulated tracking errors. The method uses a similarity metric known as photo-consistency that relies only on the visual information obtained by the pair of cameras connected to a microscope. Unlike previous studies based on photo-consistency, the registration allows the use of magnified views of volumetric anatomical models instead of full-sized views of polygonal objects.

Besides the contributions previously described, additional work has been carried out aimed to improve the performance of the IESN system and study different issues that are directly related to the overall accuracy. The following list summarises the most relevant topics among the research work:

- A series of software optimisations are applied to the original IESN system implementation. The optimisations include the reduction of feature segmentation in 2D calibration markers used for endoscopy and the visualisation of volumetric models using stereoscopic display devices.
- An evaluation of the sensitivity of the photogrammetric calibration method towards the detection of 2D markers in the projected calibration image and the markers' positional accuracy on the calibration object.
- A study of the accuracy of an optical tracking device in a working laboratory scenario in comparison with the nominal accuracy provided by the manufacturer. Additionally, a synchronisation method is introduced to ensure the integrity of the positional data when camera and patient are tracked at the same time.

The research presented in this thesis is based on experiments carried out in a laboratory

setup that simulates a surgical environment. At this stage, the collaboration of a medical team able to evaluate the performance of the IESN system has been relatively limited. In order to improve the current research, a series of clinical trials could be performed within the operating theatre as to validate the system accuracy in a real scenario. This would allow obtaining immediate feedback from practitioners during the intervention. Nevertheless, ethical approval and other requirements need to be obtained prior to the evaluation.

1.3 Thesis outline

The present chapter has introduced the subject area of this thesis and established the research objective. The remaining body of this thesis is contained in the following chapters:

Chapter 2. Background

The second chapter introduces the reader to the technology known as augmented reality (AR), including a review of research projects related to the field of medical AR. A general study of the hardware components and software techniques required to generate an AR environment is also presented.

Chapter 3. General optimisation within the IESN system

This chapter describes the series of optimisations implemented in the IESN system, including the areas of endoscopic camera calibration, motion tracking systems and stereoscopic visualisation. An evaluation of optical tracking accuracy is also explained.

Chapter 4. Camera calibration

In this chapter, analysis of performance of a camera calibration method for a single

camera is presented. In addition, the new hybrid technique to improve the accuracy of the calibration error using a stereoscopic surgical microscope is introduced.

Chapter 5. Intra-operative registration

The fifth chapter describes the method for intra-operative registration between virtual and real models based on photo-consistency that aims to compensate alignment errors produced by a tracking device.

Chapter 6. Conclusions and future work

The last chapter concludes the thesis with a summary of the research work and indicates possible areas that could be further investigated.

Chapter 2

Background

2.1 Augmented Reality

Back in 1993, one of the first publications describing the concept of Augmented Reality (AR) defined it as computer-augmented environments that “merge electronic systems into the physical world instead of attempting to replace them” [22]. However, the foundations of this technology go back to the work of Sutherland in the 1960’s, in which he developed a head mounted display (HMD) to overlay graphical information directly on the user’s vision [23].

AR can be conceived as a hybrid between the real and virtual worlds. Milgram and Kishino [24] described a virtuality continuum where a relationship exists between the objects generated by the computer and the physical environment. This continuum (shown in Figure 2.1) presents the different levels in which a user can interact with genuine and artificial objects.

Azuma [25] defines an AR system as having three important features: The first one combines real and virtual objects in a real environment. The second property establishes that it must run in real time and interactively, trying to perform the required actions within

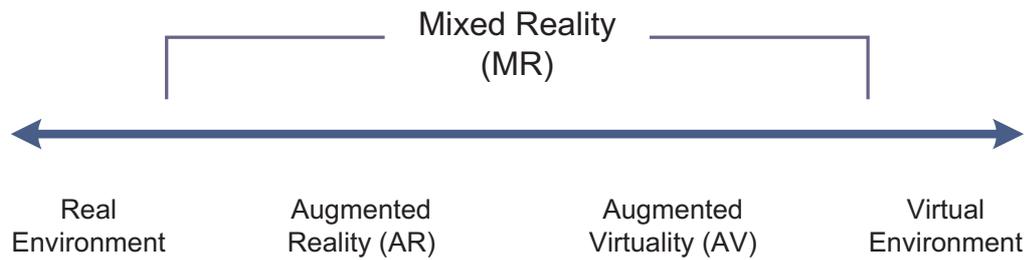


Figure 2.1: Virtuality continuum as described by Milgram and Kishino [24].

a short time delay. The last feature depicts that real and virtual objects have to be registered in 3D, which means that they must be properly aligned with each other in order to create the illusion that both worlds coexist. Different methods to achieve these characteristics will be described in the following chapters of this thesis.

A large number of applications based on AR have been researched during the past years, all of them aiming at enhancing the visual information that a user perceives with the “naked” eye. The following examples present some AR applications in several areas (for a comprehensive survey of AR based applications the reader is referred to [26, 27]):

- *Design and manufacturing:* AR can provide users with real-time visual information while performing industrial maintenance. Virtual indicators assist a technician about the location of components inside an engine or guide the user through the necessary steps to perform complex maintenance tasks. For example, Riess et al. [28] present a personal digital assistant (PDA) that records video images of a machine and recognises the device by comparing its features against a computer-aided design (CAD) database. In industrial design, AR can help to plan the development of factory environments by superimposing virtual machinery on the printed construction layout or directly on the real view of the assembly floor [29].
- *Entertainment:* Some AR applications have been developed recently within the field of computer games and recreation. A video game for PlayStation 3 called “The Eye of Judgment” (www.eyeofjudgment.com) uses a camera attached to the

game console pointing towards a special table mat. The players place printed cards on the mat and the game superimposes virtual characters on each card. During the game, the characters battle among them and also interact with the user's movements. Another example was developed by the company Total Immersion (www.t-immersion.com). Their products have been used to enhance visual interactivity by overlaying people's faces while waiting in a theme park queue. In a live concert performed by the rock band Duran Duran, AR was used to project computer-generated avatars next to the lead singer [30]. Visual effects were also projected above the audience in real time throughout the show.

- *Education*: AR based educational systems permit users to learn or reinforce skills without the necessity of totally immersing them in a new environment, as VR does. Virtual animated objects superimposed on real objects can be presented to the student in order to enhance the perception of how they work. The technology can be used together with multimedia web content to display visual information stored remotely and augment the local real world [31]. Other applications targeted at children allow them to play with patterned cubes that show different sections of a larger object in order to build a 3D jigsaw puzzle or narrate a virtual story based on the position of the different cube faces [32]. Educational applications can also be implemented in historical sites such as ancient Pompeii [33], where virtual characters can be projected on real scenes to provide the visitor with a more realistic experience than traditional audio commentary.

Particularly in the field of medicine, AR has proven to be of great potential during surgery due to the enhanced visual information it provides to practitioners. Conventionally, surgeons had to mentally reconstruct a three-dimensional patient's anatomy from a set of 2D X-rays. AR allows them to perceive the imagery in situ and in real time. The practitioner indirectly visualises the inner anatomy of a patient through video cameras



Figure 2.2: Examples of AR based SN: (left) placement of virtual ribs on the patient's body surface, (right) visual cues locate the organs that are affected by a disease. (Images from Marescaux et al. [34]).

attached to medical equipment — such as an endoscope or surgical microscope — and a corresponding CT or MRI version of the anatomy is overlaid. Thus, the superimposition aims to provide visual clues during the SN procedure. This approach is known as AR based SN or simply as IESN. Figure 2.2 illustrates the use of AR in surgical procedures.

The use of AR in the operating theatre has been introduced in several medical disciplines over the last years. As described by Shuhaiber [35], the main surgical specialties that have adopted this technology are:

Neurosurgery. This sub-field has attracted most of the research in IESN systems. The procedures can employ a stereotactic frame surrounding the operating area to allow for the 3D location of specific targets inside the body. Modern approaches, as described in the next section, have withdrawn frames for more comfortable environments. Some of the interventions within neurosurgery aim to resolve brain haemorrhages, skull fractures, brain tumours and spinal hernias, among others [36,37].

General surgery. It focuses on organs located in the abdomen and deals with bowel diseases, colon infections, inflammation of the pancreas, etc. However, other pathologies can be covered as well, such as breast cancer. In CAS systems, liver and kidney surgeries have been the most promising procedures where most of the research has been focused [10,38–40]. It is predominantly based on the use of laparoscopic devices, leading

to MIS interventions.

Orthopaedic surgery. Orthopaedics is focused on the musculoskeletal or locomotor system. Some research has been performed on fracture reduction, tumour removal, minimally invasive joint arthroscopy [41] and implant alignment for knee replacement and hip resurfacing [42, 43]. Current efforts are directed to restore and enhance the surgeon's perceptive capabilities in the operating room [44].

Maxillofacial surgery. It concerns the surgical field of diagnosing and correcting pathologies that affect head, face, neck, mouth and jaws. AR based surgery includes tumour resection [45], mandibular joint rectification [46], dental injuries and implants [47], etc. Virtual anatomical structures can be superimposed on the real target in order to guide the surgeon during bone or splint translocation.

Otorhinolaryngology. Also known as ear, nose and throat (ENT) surgery, it deals with the treatment of diseases such as sinusitis, mastoiditis, oral and larynx cancer, etc. Among the most common surgical applications, AR can be used to assist the surgeon in diagnosis, biopsies, removal of carcinoma and orbital decompression [48–50]. IESN in this area has the advantage of dealing mainly with bony structures, which provides a higher level of accuracy during the intervention as the anatomical shapes do not deform over time.

Cardiovascular and thoracic surgery. It involves medical procedures inside the chest to treat lung cancer, tumours and heart disease, among others. It is based on operating robots that assist the surgeon in the manipulation of MIS instruments. An example of a robotic-assisted surgery system is the *da Vinci* Surgical System developed by Intuitive Surgical, Inc. (www.intuitivesurgical.com). This system allows the surgeon to control surgical micro-instruments, attached to articulated arms, from an external console. In IESN, the main difficulty relies on the alignment between virtual and real models due to the continuous movement and deformation of heart and lungs. Although some research has been carried out during the last years in this specialty [51–53], no real-time application

has been found in the current literature.

2.2 Previous work in medical AR

The first steps of combining computer-based imagery with real surgical procedures were taken in the 1980's in the field of neurosurgery. Kelly et al. [54, 55] developed a medical AR system in frame-based stereotactic microsurgery for the removal of brain neoplasms. Soon after, Roberts et al. [56] and Friets et al. [57] introduced a related system for the same specialty, although it was targeted to a frameless stereotactic microscope. Other approaches [58, 59] independently applied image-guided neurosurgical systems to plan and perform removal or ablation of tumours. The systems allowed visualisation of both CT and MRI patient data.

Bajura et al. [60] presented an ultrasound-based AR system that allowed the practitioner to carry out obstetric examinations in a pregnant woman. The patient's captured images were acquired pre-operatively and displayed in real time. However, the system was only able to show a few ultrasound slices at a low frame rate. State et al. [61] improved the visualisation system by reconstructing the fetus model in 3D during an off-line stage. Nevertheless, the research group still faced many problems due to the technology available at the time. A few years later, the system was extended to a stereo setup [62, 63], allowing real-time ultrasound-guided needle biopsies using up-to-date hardware and algorithms to compensate for some previous limitations. Using a hardware configuration similar to the ultrasound-based approach, Fuchs et al. [64] implemented a visualisation system for laparoscopic surgery purposes. They employed a structured light pattern that was projected on the patient's body in order to extract 3D anatomical structures. Then, images corresponding to internal anatomy were superimposed on the real view of the body surface.

A research group from King's College London implemented an AR system called MAGI (Microscope-Assisted Guided Interventions) [65–67]. Their approach encompassed the overlay of 3D structures directly on the optics of a stereoscopic surgical microscope. For this purpose, a semi-transparent lens was adapted inside the microscope eyepieces. The system was targeted to ENT and neurosurgical interventions. Aschke et al. [68] developed a similar idea by connecting special micro displays to the microscope optics. However, the system was focused on the intra-operative planning stage of neurosurgery. Wörn et al. [45] and Marmulla et al. [69] extended Aschke's microscope by using a robotic camera that projected target position images on the patient's body surface. This projector-based AR system allowed the practitioner to follow the visual cues during SN directly into the facial surgical area.

More recent research by Caversaccio et al. [70] enhanced the surgical view inside a single microscope eyepiece. They attached an optical mini-tracker to the surgical microscope to read tools and patient's position during the intervention. This provided a similar field of view (FOV) of the scene. García et al. [11] continued the research by using infrared light-emitting diodes that act as markers during camera calibration and registration between real and virtual models simultaneously (both techniques will be described in section 2.4).

Blackwell et al. [71] employed a semi-transparent (half-silvered) mirror glass to produce an image overlay by reflecting the images generated by an upside-down monitor located above the glass. The user observes the real patient underneath the mirror, while perceiving at the same time the overlaid computer-generated images. They required a HMD to show the images onto each eye and a tracking device to record the user's head movements. Similarly, Liao et al. [72] uses a half-silvered mirror upon which the virtual models are projected. The most significant difference of their system is based on the use of a stereoscopic monitor and lack of user tracking. This frees the user of wearing special equipment that can constrain their movements.

The project Medarpa (Medical Augmented Reality for Patients) [73, 74] introduced a monoscopic AR window. This system includes a see-through liquid crystal display (LCD) panel connected to a swivel arm in which the patient's virtual structures are displayed. Its main advantage is its ability to place the window around the patient's body. Another development based on an AR window was introduced by Mischkowski et al. [46] which uses a wireless LCD panel with a connected camera that records real images. It is aimed at superimposing CT or MRI anatomy for maxilla positioning.

Birkfellner et al. [75, 76] developed a head-mounted operating binocular microscope known as Varioscope AR, which is based on a commercial solution. This system provides a larger FOV and lower magnification levels in comparison to a surgical microscope due to the use of miniature VGA (video graphics array) displays. The authors claim that this reduces noise effects and calibration errors usually found in microscopes. A research group from Siemens [77, 78] introduced another HMD-based AR application. However, its main difference relies on using an infrared camera attached to the HMD device in order to record the user and tools movements during the procedure. All cameras are synchronised in order to avoid any flicker effect during the augmentation.

Lapeer et al. [79, 80] presented an AR based training system for obstetric forceps in child delivery. The application calculates and diagnoses the level of deformation in the virtual baby skull as result of the manipulation of real forceps. Sielhorst et al. [81] developed a delivery simulator that allows the user to receive auditory and haptic (sensitive) feedback, as well as visual information during the medical training.

In the endoscopic field, Freysinger et al. [48] described an intra-operative guidance system using a viewing wand for sinus surgery. The system was based on a probe attached to a mechanical arm that allowed determining the position of an endoscope tip inside the patient's body. Shahidi et al. [50] presented an image-enhanced endoscopy system for head and neck surgery using a wireless configuration. However, their approach

consisted in displaying corresponding real and virtual images side-by-side, instead of fusing them in a single view. Lapeer et al. [13] and Thoranaghatte et al. [15] introduced independent AR systems for endoscopic sinus interventions, each of them using different approaches regarding the techniques to overlay the virtual imagery. The former also presented an in-depth evaluation of the accuracy obtained by the methods required for image enhancement.

2.3 Classification of components in AR

Since the introduction of the first AR application, different technologies have been developed in order to enhance the visual perception that a user obtains from the real world. The selection of different physical components varies according to the environment requirements. However, all AR based systems (regardless of the application field) share the same essential hardware that provides user interaction with the virtual entities. The hardware components can be broadly divided in two categories:

- Display technologies
- Motion tracking devices

In the following sections the reader will be presented with an overview of each element within AR.

2.3.1 Display technologies

Eye sight is the most important sense when perceiving the environment in an AR application. For this reason, the diversity of technologies available to visualise the enhanced world is vast. Shamir et al. [82] presents five classes of display devices based on previous work in AR, these are:

Augmented medical imaging devices

This type of equipment aims to merge virtual data on the view of an intra-operative image acquisition device, such as a CT scanner or ultrasound probe. It provides intuitive interfaces to surgeons in order to view hidden structures in real-time while capturing the patient's anatomy. Due to the inherent imagery generated by the medical devices, the overlay is composed mostly of two-dimensional slices instead of 3D volumes. An example of such method is a camera-augmented fluoroscopic C-arm (CamC) system [83, 84]. The system simulates the X-ray view obtained by the capturing equipment according to the C-arm frame position. This allows the practitioner to aim the device to the correct body part even before taking an actual X-ray, reducing the dose of radiation exposure to the patient.

Augmented optical devices

They provide an enhanced view directly on the images obtained by the optical surgical tools, such as microscopes [66, 68]. The solution requires a direct modification of the oculars and other hardware additions in order to superimpose the virtual models on the real view, which can be presented as a stereo pair or in a monoscopic view. As the scan is performed pre-operatively, the overlay can consist of 2D image slices or a complete 3D reconstruction of the organs/bony structures. The main advantage of this approach relies on using current equipment that surgeons are already accustomed to. Thus, the learning curve for the utilisation of this technology is short in comparison with other devices. Moreover, the user is not required to look away from the surgical target, which may interrupt the operating workflow. The disadvantage of augmented optical devices is the difficulty to modify existing surgical tools to enhance the user's view.

AR monitors

The augmented view can be achieved using standard CRT (cathode ray tube) or LCD computer monitors, making them the easiest, and perhaps the cheapest, method to merge real and synthetic worlds. For a monoscopic view, it is only necessary to capture the video image and blend it with its virtual counterpart using a software-based transparency technique, provided both view perspectives are similar. However, to obtain stereoscopic vision, special monitors must be used. Such monitors allow depth perception from a pair of images by diverting each independent view to its corresponding eye. Whereas initial CRT stereoscopic monitors were too heavy and large for its use in the operating room, newer devices based on LCD technology provide an affordable and ubiquitous solution.

The accuracy of stereoscopic displays have been under study since early 1990's. Drastic and Milgram [85] firstly pointed out the advantages of stereoscopic video (SV) in comparison to monoscopic video (MV). Moreover, they demonstrated the potential of superimposing stereoscopic graphics (SG) on stereo video images (SV+SG) for a diversity of environments such as telemanipulation or microscopy. Chios et al. [86] and Lapeer et al. [87] analysed the use of autostereoscopic technology (which allows stereo vision without the need to wear any specialised viewing hardware) as an alternative of directly observing through a pair of stereo microscope eyepieces.

AR window systems

These semi-transparent devices permit a direct visualisation of real objects placed behind them while overlaying synthetic images on the screen. The superimposition of virtual imagery is achieved by using a see-through LCD panel that renders the virtual objects on its screen or by reflecting the images from a monitor onto a half-silvered surface. An example of the latter configuration is a project developed at the Department of Maxillo-facial surgery at the Technical University in Munich. The system called ARSyS-Tricorder

(www.arsys-tricorder.de) aims to provide the user with a three-dimensional stereoscopic view of CT anatomical structures. Stetten et al. [88] presented an image guidance technique for needle biopsy using an AR based ultrasound device. They attached a small mirror and a monitor to the ultrasound transducer in order to overlay the visual imagery on the hand-held device.

Head-mounted displays (HMDs)

HMDs have become the most common visualisation device for AR purposes since their introduction in the 1960's. There are two different technologies for HMDs: video-based and optical displays. The former capture the real world using a pair of cameras and overlay the virtual models on small screens in front of the user's eyes, blocking the line-of-sight between operator and real scene. Optical see-through devices employ semi-transparent mirrors or screens that reflect the projection from a pair of monitors. Each technology has advantages and disadvantages regarding their display quality, time delay due to rendering, level of physical constraint for the user, among others. Rolland and Fuchs [89] extensively compare optical and video HMDs in the field of 3D medical visualisation. They acknowledge that each device performs the best according to the type of application and its requirements.

Head-mounted projective displays (HMPDs) use a pair of projectors that are focused on a half-silvered mirror located in front of the user's eyes. However, the computer-generated projection is not directed towards the viewer but to a special retro-reflective screen. This screen allows augmenting the objects placed underneath and projects back the virtual imagery to the viewer. Rolland et al. [90, 91] present different applications using HMPDs for the medical discipline as well as in other areas.

Several other display technologies are currently available for AR. For example, virtual retinal displays (VRDs) project images directly on the user's retina. A low-power laser is

used to scan the eyes and the pixels that comprise the images are cast inside the eye. The visual perception is similar to looking at a screen floating in space. Projection-based displays aim to enhance physical objects such as walls or special desks in order to provide them with texture and other visual information. This approach is denominated Spatial Augmented Reality [92]. However, these AR displays are not designed for their implementation in the medical field. An in-depth survey of displays within the AR domain is presented by Bimber and Raskar [93].

2.3.2 Motion tracking devices

Because an AR environment is not static, it is necessary to register the movements of physical entities that permit the user to have a level of interactivity with the enhanced world. For this reason, a tracking device is used to detect the different individual positions and orientations performed by each element and to maintain a relationship among them through a global frame coordinate system. The elements required to be tracked in a medical procedure include surgical instruments, optical devices (i.e. endoscope or microscope) and possibly the patient.

Different technologies can be used in the operating theatre to record the ongoing pose of objects and subject during the intervention. All of them provide six degrees of freedom (DOF) with respect to a reference point: three for translation and three for rotation. The main different categories of tracking devices for surgical interventions are electromagnetic, electromechanical and optical.

Electromagnetic

This type of motion tracker works by generating an electromagnetic field from a central transmitter and a set of sensors or receivers attached to the object of interest that record

their relative changing voltage or current. Electromagnetic trackers can be classified in three different generations [94]: The first generation — based on alternating current (AC) technology — has the disadvantage of being highly susceptible to the interference of metallic objects or electric equipment. The second generation presents lower sensitivity to interference by employing direct current (DC) transmission. A third generation of trackers includes special transmitters that block distortions originating from below the tracking volume. They also provide more advanced calibration and processing techniques to increase their accuracy levels.

The main advantage of electromagnetic motion trackers is that they do not require a clear line-of-sight between transmitter and receivers. Thus, the sensors can be visibly obstructed in the environment by other objects. In the case of surgical procedures, miniaturised sensors can be introduced inside specific instruments and within the patient's body.

Electromagnetic devices for medical applications are currently manufactured by Northern Digital Inc. (www.ndigital.com), Polhemus (www.polhemus.com) and Ascension Technology Corporation (www.ascension-tech.com).

Electromechanical

These trackers are based on a set of jointed appendages that record position and orientation using sensors, such as potentiometers or encoders, in each of its joints. When the limbs are connected to a human body part, they form exoskeletons to capture six DOF movements. However, in the case of medical applications, electromechanical arms are attached to surgical tools to perform robotically assisted surgery.

Similar to electromagnetic trackers, electromechanical devices have the advantage that they are not affected by line-of-sight occlusion. However, they are restricted to a maximum physical range of the arm limbs regarding length and rotation. A mechanical arm

can also be obtrusive during the manipulation of instruments in specific circumstances. Moreover, an electromechanical tracker can only register the movements of a single object.

Optical

Optical trackers employ two or more cameras that recognise the pose of target objects through computer vision algorithms. The camera system captures video images of the scene and detects a set of active or passive markers attached to the object of interest [95]. Active markers consist of visible light emitting diodes (LEDs) or infrared emitting diodes (IREDs) that are constantly activated by an electric signal. The electrical current can be either provided by a main controller box (wired configuration) or by the target surgical instrument or batteries (wireless set up). Passive markers comprise retro-reflective spheres that bounce back infrared light projected by the cameras. Both active and passive optical systems require at least three markers in order to locate objects within a 3D volume, although a fourth marker can be used to increase reliability. Additionally, a new generation of passive trackers use standard camera hardware that recognises flat target patterns (circular or user-defined shapes) in the scene [96]. Table 2.1 provides a list of commercial vendors, their optical tracking products and the reported accuracy of each system.

The main advantage of optical tracking devices is the accuracy achieved by the combination of hardware components and computer vision techniques. However, their main drawback is the requirement of a clear line-of-sight between the camera system and markers. Thus, performance can decrease in case of occlusion by other objects or individuals.

Similar to AR medical applications, the use of tracking technologies in SN systems is required in order to provide the surgeon with positional data of different surgical devices and/or patient during the intervention. The selection of a specific tracking technology in the operating room mainly depends on the requirements of the surgical procedure. En-

Company	Product	Type	Accuracy (mm RMS ^a)
Northern Digital Inc. (www.ndigital.com)	hybrid Polaris / Spectra Polaris Vicra	Passive & Active Passive	≤ 0.35 0.25
Boulder Innovation Group, Inc. (www.imageguided.com)	FlashPoint 5500	Active	≤ 0.25
Atracsys LLC. (www.atracsys.com)	easyTrack accuTrack infiniTrack	Active Active Passive	< 0.30 < 0.25 < 0.50
Claron Technology Inc. (www.clarontech.com)	MicronTracker2 family	Passive	≤ 0.35
A.R.T. GmbH. (www.ar-tracking.de)	ARTtrack / SmARTtrack	Passive	Not available

^aRoot mean square

Table 2.1: Commercial optical tracking devices.

quobahrie et al. [97] described some factors that must be considered when choosing a tracking modality, these are: line-of-sight requirements, accuracy of the device and its update rate, maximum number of entities to be tracked simultaneously, measurement volume limitations, implementation and running costs, and general conditions in the operating environment. Some examples of commercial SN systems are presented in Table 2.2, describing their tracking technology and the surgical specialties that they are aimed at.

2.4 Methods in AR

The previous section introduced the hardware components required to display a visually-enhanced world and track the movements of entities in an AR environment. However, a series of software techniques need to be applied in order to produce the overlay between real and virtual imagery. In this section the reader will be introduced to the concepts of each method, although a deeper explanation of theory foundations and algorithms will be covered in the following chapters of this thesis.

Company	Product	Tracking technology	Surgical specialty
Aesculap AG (www.orthopilot.com)	OrthoPilot	Optical	Orthopaedics
Acrobot Co. Ltd. (www.acrobot.co.uk)	Acrobot Navigator	Electromechanical	Orthopaedics
BrainLAB AG (www.brainlab.com)	VectorVision Kolibri	Optical	Neurosurgery ENT
Collin SA (collin.axepartner.com)	Digipointeur	Electromagnetic	ENT
Elekta AB (www.elekta.com)	SonoWand Invite	Optical	Neurosurgery
General Electric (www.gehealthcare.com)	InstaTrak ENTrak	Electromagnetic	Neurosurgery/ENT ENT
Medtronic Inc. (medtronicnavigation.com)	StealthStation	Optical and electromagnetic	Several
Smith & Nephew (www.smith-nephew.com)	AchieveCAS	Optical	Orthopaedics
Stereotaxis, Inc. (www.stereotaxis.com)	NIOBE	Electromagnetic	Cardiovascular
Stryker (www.stryker.com)	Navigation System II eNlite	Optical	Several
Zimmer, Inc. (www.zimmer.com)	ORTHOsoft	Optical	Orthopaedics

Table 2.2: Examples of commercial SN systems.

Camera calibration

This technique involves the derivation of the true parameters of one or more cameras by obtaining visual information from the real world. Because the parameters can not be directly measured, a relationship must be established between features located on 3D objects and their projections on 2D video images. The parameters of a camera model can be divided in two categories: external and internal. External or extrinsic parameters represent the position and orientation of the device in relation to a specific point in the world coordinate system. Internal or intrinsic parameters denote the configuration of the inner optical system and include focal length, scale factor, radial distortion and optical centre of the image.

Different computer vision techniques can be implemented to perform the calibration according to the number of cameras attached to the AR based system. Once the camera parameters have been computed, the information is interpreted by the AR application which aligns the virtual world perspective with its real counterpart. This procedure is usually carried out in a pre-operative stage.

Segmentation

The generation of 3D models from a set of CT or MRI scans can lead to volumes with extra information not required during surgery. The aim of segmentation is to find meaningful regions in the volumetric model that represent specific areas without compromising the dataset quality. A general segmentation method extracts structures around a region of interest, which can include tissue, bones or blood vessels.

In AR based medical applications, segmentation allows sectioning out portions of the virtual data that are not required to match to a real model. This can involve noisy areas in the reconstructed volume or extra features of the anatomical structure. Because the procedure reduces the number of volume elements (voxels), it also helps to decrease the number of calculations to be executed in the registration step. In some IESN procedures, segmentation can be regarded as optional or merged with registration.

Registration

Once the segmented model has been obtained, it is required to be matched to the patient's anatomy. The process can use information obtained from physical structures (known as natural landmarks) found in both dataset and real model. Alternatively, fiducial markers can be attached to the patient during the pre-operative scan and compare the location of their real and virtual counterpart in the generated volume. The correspondence between real and virtual landmarks/markers produces a transformation matrix that is used to align

the entities in the AR medical procedure.

Registration can be classified in four categories: manual, interactive, semi-automatic or automatic. Manual registration allows the user to translate and rotate the virtual (target) model. The interactive approach requests the user for input before attempting to find a similar pose in the real (source) data set. Semi-automatic registration executes the alignment automatically and then asks for user feedback. The automatic method does not need any user input throughout the process in order to compute the alignment.

Once the techniques of calibration, segmentation and registration have been performed, a static overlay is obtained which comprises the aligned real and virtual models. A tracking device is then used to capture the dynamic pose of the models throughout surgery and update the overlay.

2.5 Summary

In this chapter, a literature survey has been presented that includes diverse research related to the use of AR for medical purposes. An introduction to AR technology and its applications in several surgical specialties has been covered. A classification of the main hardware components that allow an image-enhanced view in SN systems was also described. Additionally, the reader was introduced to the software methods required to produce an overlay between real and virtual imagery, which comprise the calibration of the camera used to acquire images, segmentation of CT or MRI data, and registration between real and virtual models.

Chapter 3

General optimisation within the IESN system

3.1 Introduction

IESN systems are based on the use of different hardware components and software methodologies that aim at visually enhancing the patient's anatomy during surgery. The integration of these technologies must rely on a central computer application that generates an AR environment through managing the information provided by the external components. Therefore, the software can be considered as the most important means of communication between the SN system and the surgeon in the operating room.

The computer program employed as the basis for this research is denominated ARView, which is an extension of a PC-based medical volume rendering software named 3DView [98]; developed using C++, Microsoft foundation classes (MFC) and OpenGL. While 3DView allows the visualisation and manipulation in real time of volumetric models obtained from CT or MRI data, ARView extends these features by providing support to IESN functions.

ARView can be used for MIS procedures through the utilisation of an endoscope or

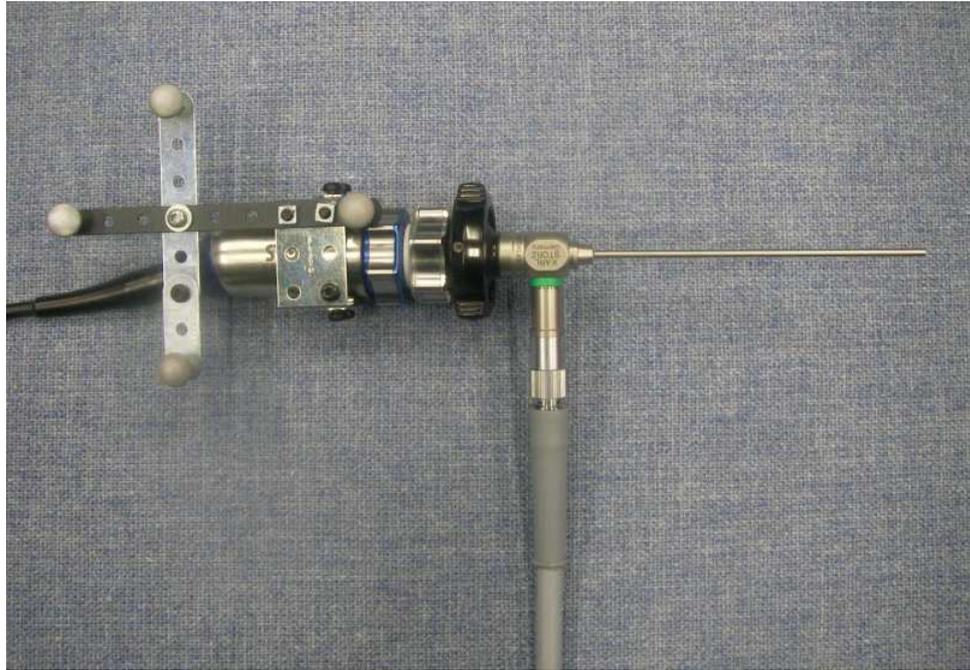
for open interventions using a stereoscopic surgical microscope. Thus, the IESN application allows capturing video from a single or a pair of cameras connected to the surgical devices. The display technology used to present the augmented view consists of a stereoscopic monitor that allows the user to alternate between single and stereo display mode. Additionally, the software permits the use of motion tracking devices during the intra-operative stage.

Part of the research in this thesis involved the optimisation of different functions within ARView that contribute to the overall visual augmentation. This chapter will present the techniques applied to various aspects of the original IESN system described by Chen [17]. The following sections focus on several improvements regarding the use of ARView as a SN system for ESS, and 3D stereo visualisation using stereoscopic monitors. Also, a validation of accuracy of two optical tracking devices (i.e. hybrid Polaris and Polaris Vicra) in a simulated surgical scenario is investigated.

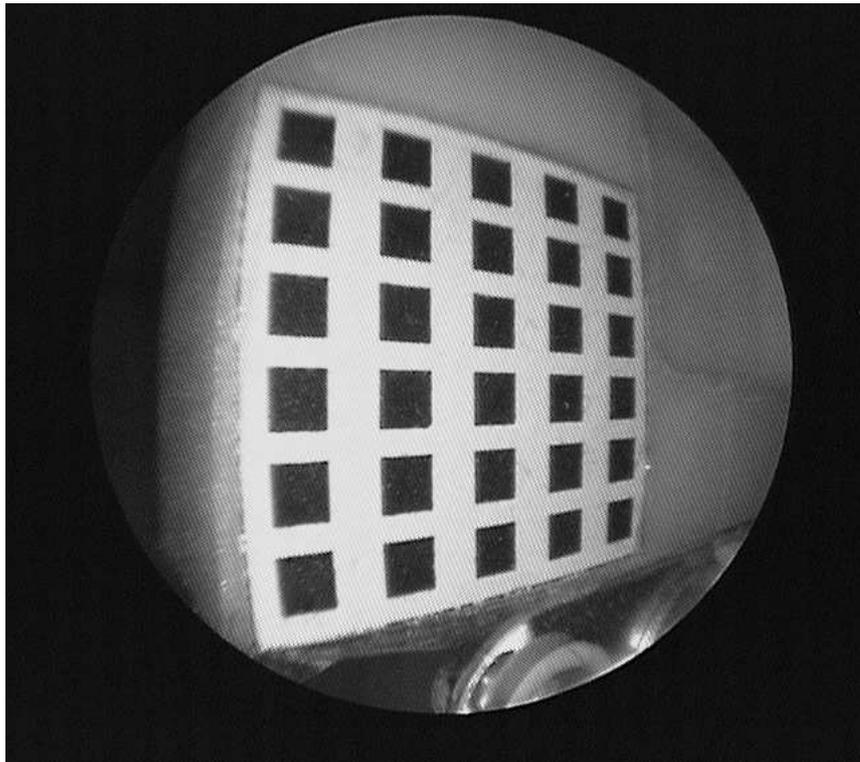
3.2 Marker detection for endoscopic camera calibration

In order to display real video images in conjunction with computer-generated data in AR, it is necessary to align a virtual camera system with the real camera device. If the resulting alignment is correct, a virtual object will be observed from the same camera position and angle as the equivalent real object. Although camera calibration will be discussed in more detail in the next chapter, it is important to mention that an IESN system requires a calibration object to relate a set of physical features to their corresponding projections on the captured image.

The IESN system for ESS involves a rigid zero-degree endoscope that is calibrated by pointing it towards the calibration object and acquiring a still video image through a frame grabber. The detection of physical features relies on the design of the calibration object, which comprises a number of squared shapes printed against a contrasting background



(a)



(b)

Figure 3.1: Physical equipment used in the IESN system for ESS purposes: (a) rigid zero-degree endoscope with mounted passive optical markers for tracking; (b) planar calibration object as seen by the endoscope.

hence forming a 2D grid pattern. Figure 3.1 shows the rigid endoscope and the image of the calibration pattern as seen by the optical device.

As the captured image can present different noise levels produced by the internal optical components or other external factors, a Gaussian filter [99] is applied in order to reduce their unwanted effects. Subsequently, Canny's edge detection algorithm [100] is employed to identify the borders of the squared features on the printed pattern. This produces a binary image, where the background is represented by zero values and the detected edges by non-zero values. The binary image is used as an input for a region identification procedure denominated *connected component labelling* [101].

The first step in connected component labelling scans the entire binary image row by row and assigns a numerical identifier, or label, for each pixel different than zero. The value of a label depends on the neighbouring pixels that have been previously labelled, if any. This is determined by an eight-neighbourhood mask, where pixels can be connected horizontally, vertically or diagonally. Three different options during labelling are: *a)* If all neighbouring pixels have zero values (corresponding to the background), a new label value is assigned to the current pixel; *b)* If only one of the pixels in the neighbourhood has a non-zero label, assign this label value to the current pixel; *c)* If there is more than one pixel among the neighbours with a label different than zero, select arbitrarily one of the label values and assign it to the current pixel. In case that the neighbouring labels are different (colliding labels), store the label pair in a separate structure known as equivalence table for subsequent evaluation.

The next step of the procedure involves merging, into the same region, the contiguous edges that have been identified as having colliding labels. For this reason, the binary image is scanned a second time, where the label of each pixel is compared against the label pairs in the equivalence table. If the pixel label is found, its value is replaced with the lowest element of the pair. At the end of this step, the set of pixel labels with the same

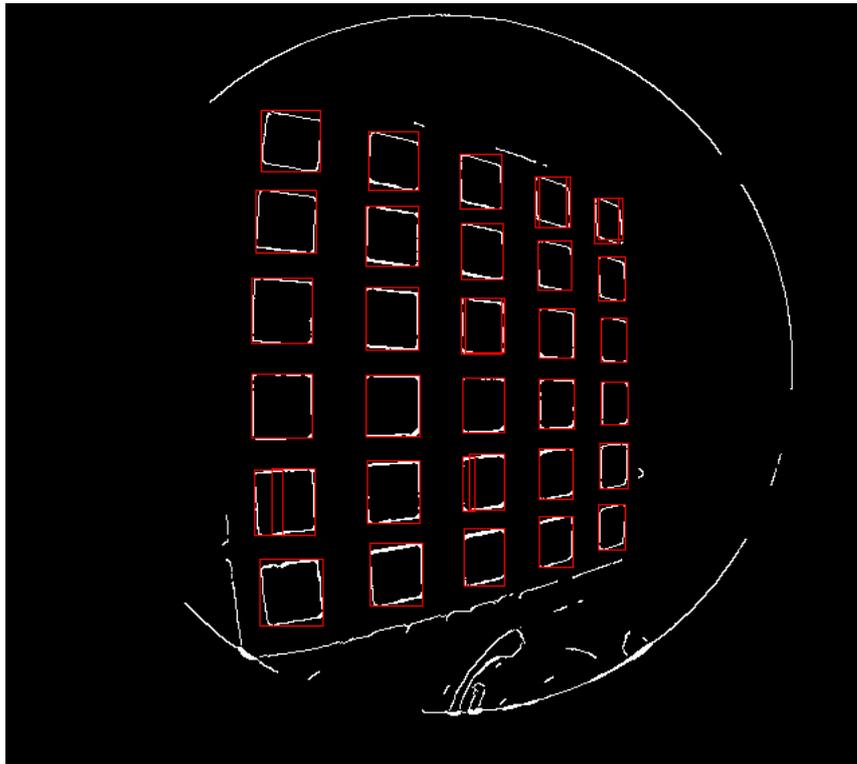


Figure 3.2: Example of segmented figures during the feature detection process, where fractured regions lead to independent bounding boxes within the same feature marker.

identifier must correspond to individual regions. As an additional stage, each resulting region is fitted with a bounding box that provides a graphical representation of a feature marker during camera calibration.

A problem found during the procedure is that, depending on the image noise level and the control parameters used to reduce its effects, in some circumstances the detected contours can present gaps along the borders. This issue leads to imperceptible segmented regions in the squared shape, which in turn generates independent bounding boxes for the same feature as illustrated in Figure 3.2. Although this problem can be solved by manually modifying the control parameters using tolerance values, the overall process to correct this problem for all segmented features tends to be cumbersome.

In order to reduce feature segmentation and, consequently, bounding box partitioning, it was decided to analyse the original implementation of the connected component labelling. It was noticed that the two steps of this procedure, pixel labelling and edge

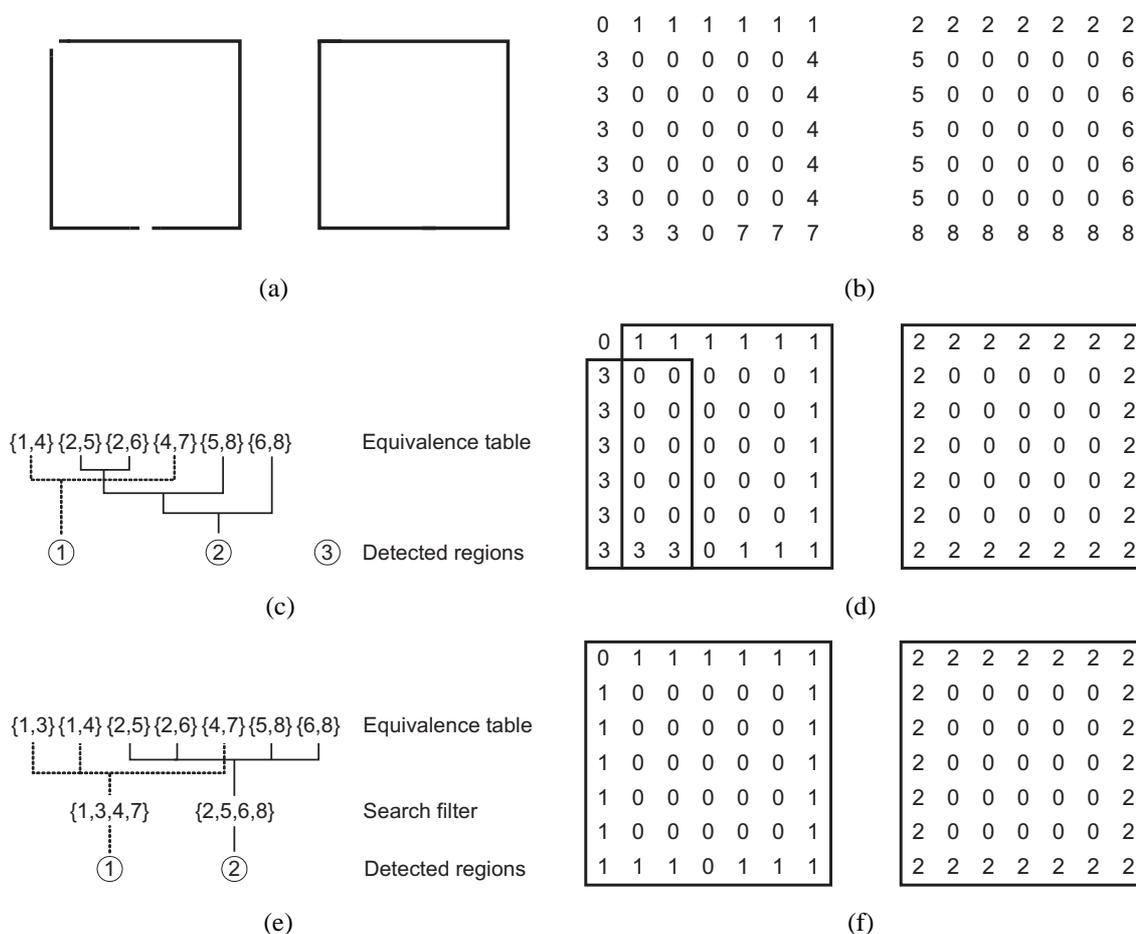


Figure 3.3: Example of connected components during feature segmentation. (a) Boundaries of two squared markers of a calibration pattern and (b) their corresponding pixel labels. (c) Colliding label pairs missing in the equivalence table and an extra detected region. (d) Bounding boxes associated to independent detected regions. (e) Colliding labels detected in the improved implementation using search filters, and (f) resulting regions with corrected bounding boxes.

merging, were performed on the same data structure used to store the input binary image. This caused incorrect data indexing, which affected the eight-neighbourhood identification and led to data corruption. Figure 3.3 exemplifies the pixel labelling applied to the contours of two squared markers of a calibration pattern. As it can be observed in Figure 3.3(a), the squared marker in the left presents small gaps along its boundary (top-left corner and bottom edge); whereas the square in the right has all edges connected. After the first scan of the procedure, independent edges are detected with colliding labels along the boundaries of both squared markers (Figure 3.3(b)). In the original implementation,

the eight-neighbourhood mask could not detect the diagonal connection between neighbouring pixels such as in the top-left corner of the left square. Therefore, the label pair corresponding to pixel values $\{1,3\}$ is not included in the equivalence table presented in Figure 3.3(c). Because of this, it is not possible to merge the edge with its respective neighbour, resulting in an isolated region (shown as circle 3), whereas the rest of the label pairs are merged into two individual regions (circles 1 and 2). It must be mentioned that, during the second step of the procedure, the comparison between each pixel and the label pairs is performed sequentially along the entire equivalence table (this is represented by connection lines among equivalent label pairs). Finally, the detected individual regions (circled numbers 1, 2 and 3) are fitted with bounding boxes as shown in Figure 3.3(d). Due to the detected isolated region (circle 3), two bounding boxes are generated for the same feature marker in the left.

An optimisation of the original implementation was carried out. In a first stage, an additional data structure was used to temporarily store the set of individual edges with unique labels that were generated during the first scan. This ensured that the operations required for image processing did not interfere among themselves and data integrity was maintained throughout the procedure. Thus, issues related to the detection of diagonal neighbouring pixels (e.g. top-left corner of the left square) were solved. Figure 3.3(e) shows the diagonal label pair $\{1,3\}$ of the top-left corner included in the equivalence table. As an additional stage, a dynamic data structure was used as a “search filter” during the merging of colliding labels. For this purpose, all labels in the equivalence table with equal values in the first or second element of each pair were combined into individual search filters, as illustrated in the middle of Figure 3.3(e). Then, during the second scan of the procedure, each pixel in the image was compared to the search filters rather than to the entire equivalence table, reducing the number of total comparisons. Also, each search filter corresponds exclusively to an individual detected region (circled numbers) hence to a unique bounding box as seen in Figure 3.3(f).

Furthermore, the partitioning of bounding boxes was also analysed. It was noticed that, if the separation between detected regions that are related to the same squared marker on the calibration object is lower than a specific threshold, the corresponding bounding boxes could be merged without affecting the feature detection accuracy. A final stage was performed based on an underlying statistical technique that takes into account the aspect ratio of the bounding boxes, as described in [17]. This technique is based on the fact that bounding boxes that are associated to calibration markers have more consistent aspect ratio and area values in comparison with the bounding boxes corresponding to foreign objects. Therefore, if the aspect ratio and area do not fall within a certain threshold, the bounding box can be discarded. This prevents other shapes in the image that do not belong to the printed pattern to be identified as possible calibration markers.

3.2.1 Experiment

Influence of markers segmentation in endoscope calibration

This experiment was aimed at evaluating the influence of segmentation in the detection of feature markers within the ESS camera calibration. A comparison was carried out by calibrating the endoscopic camera using the original feature detection procedure and the optimised implementation that reduces fractures on the pattern squares. The accuracy was assessed by calibrating the endoscope using the well-established Tsai's algorithm [102]. For this purpose, a relationship is established between the 3D features found on the physical calibration object and the pixel coordinates corresponding to the centres of the detected bounding boxes. The error values returned by Tsai's method indicate the exactness of the overall calibration procedure (including marker localisation). In addition, a qualitative evaluation is obtained by overlaying a virtual pattern on the image of the real calibration object, where the alignment between models reflects the level of accuracy.

The results obtained by Tsai's method report a mean calibration error of 1.50 pix-

els in image coordinates when the feature marker detection include fragmented regions. Figure 3.4(a) illustrates the detected features in the calibration pattern, where their corresponding bounding boxes are numbered for easier recognition. It can be observed that several shapes are segmented in two different regions. This causes the detected marker centres to be offset from the central points in the real squares, especially at squares number 5 and 7. The effects of the segmentation is shown in Figure 3.4(b) in which the resulting overlay presents a misalignment between the virtual and real patterns, predominantly at the top-left and bottom-right corners of the grid. In the case of the optimised implementation, the marker centres agree with their respective central points in the real calibration object, as illustrated in Figure 3.4(c). The calibration error value decreases to 0.66 pixels with respect to the original procedure. Also, a better alignment between real and virtual features is obtained in the final overlay (Figure 3.4(d)).

It must be noted that there is still a slight misalignment between the real calibration object and its corresponding virtual overlay in Figure 3.4(d). This is because the rotational angle applied to the calibration object is around 25° with respect to the camera's image plane, whereas it has been demonstrated that a planar calibration object must be rotated between 30° and 45° in order to provide more accurate results [102, 103]. Although the rotational position used in this experiment is not optimal, it has been selected only to illustrate the influence of marker detection. If a higher angle had been used, the marker centres would be more difficult to visualise in the image due to perspective, especially in the squares located in the last column of the calibration grid. The effects of this rotational issue will be further evaluated and discussed in the next chapter concerning camera calibration.

This experiment demonstrated that segmented regions in feature marker detection can affect the final accuracy in camera calibration. Although the segmentation accuracy can be difficult to assess, its influence can be noticed in the resulting overlay between real and

virtual calibration patterns. Therefore, appropriate methods to avoid feature segmentation errors must be validated and tested throughout the implementation procedure.

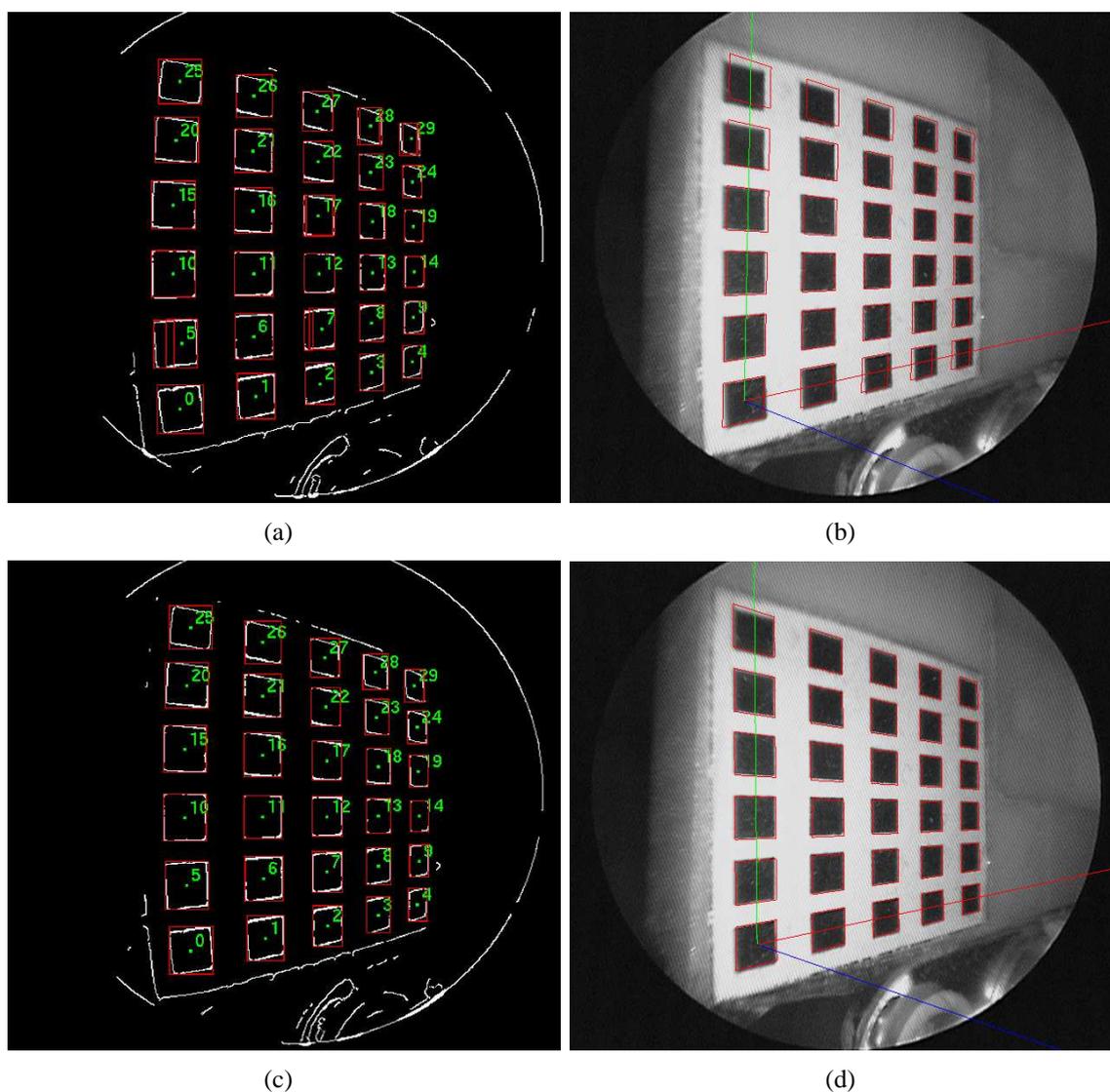


Figure 3.4: Comparison of endoscope calibration before and after reduction of feature segmentation. (a) Detection of feature markers when figures on the calibration are fragmented and, (b) resulting misalignment in the calibration overlay. (c) Marker detection when the segmentation of features is corrected and its corresponding overlay with aligned patterns (d).

3.3 Use of an optical tracking device within the IESN system

Another factor that affects the accuracy of an IESN system includes the technology used to track the movements performed by the surgical tools and patient. Although not exclusively, optical tracking systems are mostly employed within the operating theatre due to their high accuracy levels in comparison with other technologies (see Table 2.1 for a list of commercial optical tracking devices). Optical trackers also provide the surgeon with the ability to freely manipulate instruments without being obtrusive.

Two optical tracking devices developed by Northern Digital Inc. were used during the research to track the motion of a dummy patient and surgical imaging devices (endoscope and microscope), i.e. the hybrid Polaris and Polaris Vicra (Figure 3.5). The former is aimed at general surgical interventions, which require a large measurement volume to track the tools within its FOV. The volume boundaries for tracking are defined between 140 cm (near plane) and 240 cm (far plane). This volume is large enough to cover the full length of the operating table and the entire patient body. On the other hand, Polaris Vicra has a smaller measurement volume than the hybrid Polaris, and is defined between 55.7 and 133.6 cm for the near and far planes, respectively. Thus, Polaris Vicra can be used for more targeted procedures such as ENT and Head and Neck surgery.

The accuracy specifications provided by the manufacturer are reported to be ≤ 0.35 mm root mean squared (RMS) error for hybrid Polaris and 0.25 mm for Polaris Vicra. It should be noted that each manufacturer carries out accuracy assessments under controlled laboratory conditions which may not strictly correspond to typical working environments [104]. Moreover, different protocols and statistical measurements are selected in order to highlight the product's capabilities. In the case of both Polaris, the nominal accuracy levels are based on a coordinate measuring machine (CMM) volumetric protocol, where optical markers are placed at different locations inside the device's operational



(a)



(b)

Figure 3.5: Optical tracking devices used during the research: (a) hybrid Polaris; (b) Polaris Vicra.

volume. For each position, the difference between the real (ground truth) location and the coordinates given by the tracking device is computed. The 3D error vector is then reduced to a 1D distance error, which is regarded as the overall RMS error value.

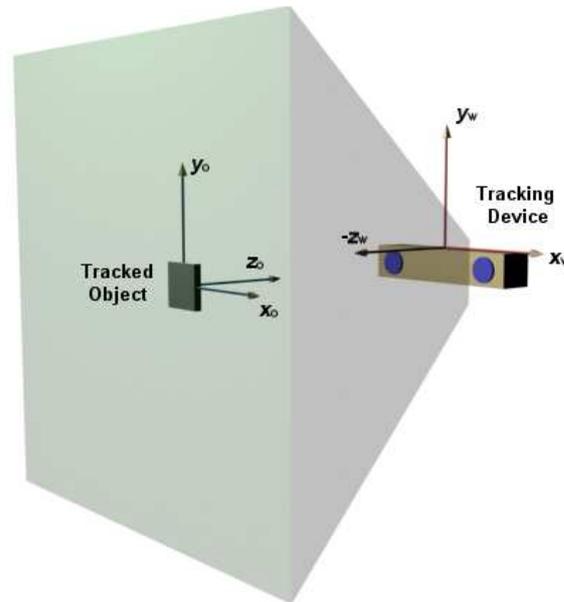


Figure 3.6: Illustration of a tracked object that is parallel to the xy -plane of the tracking device.

The accuracy tests based on the CMM protocol involve the recording of a single optical marker around the volume region. Increasing the number of markers (attached to a rigid body) can help to increase the reported accuracy levels, although this depends on the markers' layout and their distance to the object's point of interest (e.g. probe tool tip). As described by Wiles et al. [105], the accuracy tends to increase when the markers attached to the rigid body are parallel to the xy -plane of the Polaris' volume. Figure 3.6 illustrates a parallel-aligned tracked object with respect to the tracker's coordinate system. Following this approach, Khadem et al. [106] evaluated the accuracy of different optical tracking systems by placing passive optical markers in parallel to the xy -plane of the camera sensors. Their results showed that the RMS error for all tracking devices was less than 0.11 mm. Nevertheless, during surgery, it is practically impossible to maintain the rigid bodies parallel to the Polaris' cameras due to the manipulation of tools. This is also true if the tracked region of interest (ROI) changes its original orientation during the procedure, or if the ROI can not be positioned in parallel to the Polaris' cameras from the beginning of the operation. In any case, the accuracy levels obtained by an IESN system will differ from the manufacturer's specifications.

3.3.1 Experiment

Accuracy of the tracking device in a laboratory setup

The purpose of this experiment was to evaluate the accuracy of the three translational components reported by the optical tracking devices in a scenario that resembles the placement of such systems in the operating theatre, i.e. the cameras are not strictly parallel to the surgical area, as shown in Figure 3.7. For this purpose, a set of passive markers were attached to a support that was placed on a measurement bench of 381 x 381 mm, which allows a physical positioning accuracy of 0.05 mm. Because hybrid Polaris and Polaris Vicra have different measurement volume dimensions, the grid plane was located at two different distances from the centre of the motion tracker in order to ensure that all data points were collected within the central measurement volume. The distances were 195 cm in the case of hybrid Polaris, and 110 cm for Polaris Vicra.

The three translational DOFs were set to zero during initialisation and registration of the first tracked position. Then, the support object was moved around the grid while recording the ground truth 3D positions and the coordinates reported by the tracking device. Five readings for each of five positions around the grid were registered, making a total of 25 data points. The variation of the tracking signal provided by the tracking device for each DOF under static conditions was smoothed using a finite impulse response (FIR) filter in order to reduce signal noise during the assessment.

The evaluation involved fitting a plane through the 25 collected data points using eigenvalue decomposition, where the eigenvector with the smallest eigenvalue represents the normal to the plane which goes through the centroid of the data points. The average distance to the fitted plane across all data points was calculated as a measure of tracking accuracy. This procedure, regarded as an orthogonal distance regression plane, is used to minimise the perpendicular or orthogonal distances to the plane when there are error levels in the measurements of the 3D coordinates. Particularly, the procedure to compute

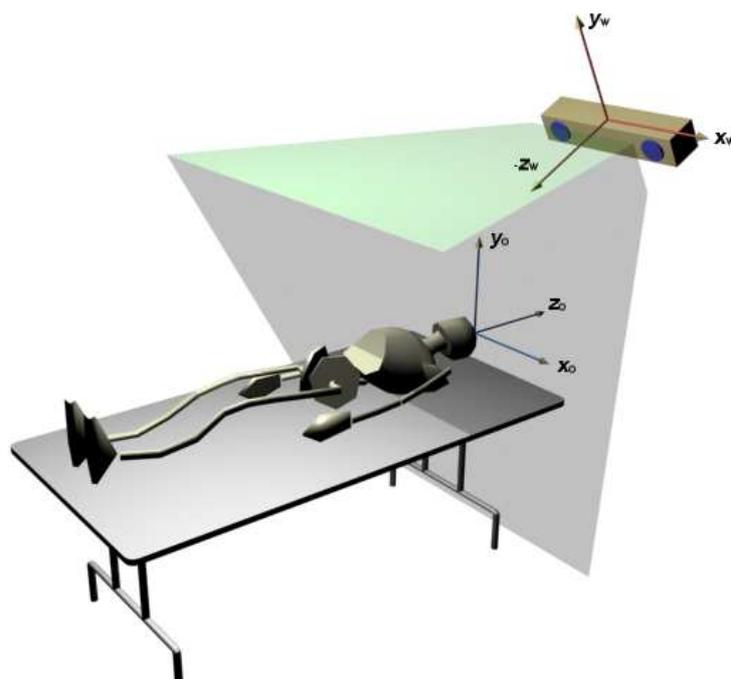


Figure 3.7: Typical setup during the intervention in which the tracking device is oriented towards the surgical area. Patient and tracker coordinates are not aligned.

the plane fitting comprises the following steps:

1. Calculate the centroid (x_0, y_0, z_0) of the n recorded data points

$$(x_0, y_0, z_0) = \frac{\sum (x_i, y_i, z_i)}{n},$$

where n is the number of points, and $i = 1 \dots n$.

2. Create a matrix M that includes the difference of each point coordinate to the centroid, such that:

$$M = \begin{bmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \\ \vdots & \vdots & \vdots \\ x_n - x_0 & y_n - y_0 & z_n - z_0 \end{bmatrix}.$$

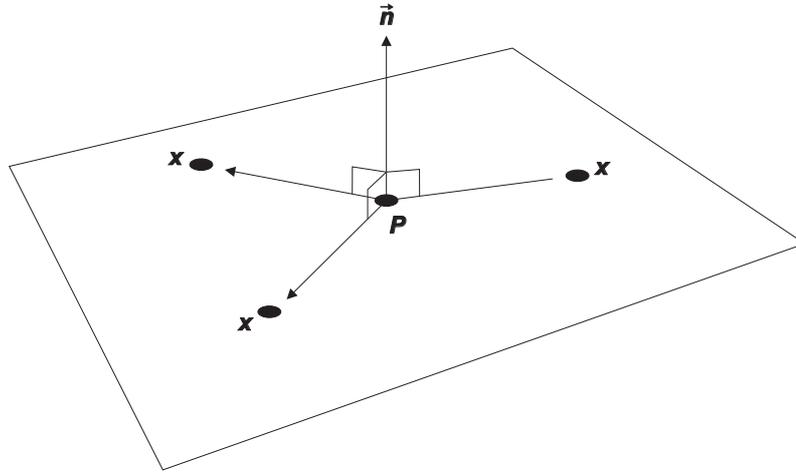


Figure 3.8: Eigenvector defined as the normal vector \vec{n} orthogonal to the set of data points (Image adapted from Schneider and Eberly [108]).

3. Use the singular value decomposition (SVD) method [107] to calculate the eigenvector related to the smallest eigenvalue in the matrix $M = UDV^T$, where the last column of V^T corresponds to the smallest eigenvalue element in the diagonal matrix D . The resulting eigenvector represents the normal ($\vec{n} = [A \ B \ C]^T$) to the plane P as shown in Figure 3.8.
4. Each point in the data is substituted into the normal-point form of the plane equation $Ax + By + Cz + D = 0$, as follows

$$\sum_i^n (Ax_i + By_i + Cz_i + D) = 0,$$

$$A \sum_i^n x_i + B \sum_i^n y_i + C \sum_i^n z_i + n \cdot D = 0,$$

$$D = \frac{- \left(A \sum_i^n x_i + B \sum_i^n y_i + C \sum_i^n z_i \right)}{n},$$

where D is the average distance to the fitted plane used as a measure for tracking error.

The results obtained indicate that the hybrid Polaris produces an average error of 1.2 mm (standard deviation SD ± 0.48 mm), while the average error generated by Polaris Vira is 0.80 mm (SD ± 0.18 mm). These translational errors differ from the manufacturer's specifications and provide information about the tracking accuracy levels that can be experienced in a surgical scenario. This proves that the methodology used for evaluation greatly influences the device's precision. Specifically, the difference in methodologies relies on the fact that the manufacturer's protocol measures the positional error of a single marker tracked around the operational volume region, where the positions are parallel to the xy -plane of the device's coordinate system. In contrast, this experiment measured the distance of a set of markers attached to a rigid body using an orthogonal distance regression plane, where the positions were recorded within the central tracking volume and oblique to the Polaris' coordinate system. The results also indicate that the overall accuracy obtained by the optical trackers can differ among specific surgical applications.

3.4 Data synchronisation of motion tracking in the IESN system

As described previously, an IESN system requires tracking the positions of different physical entities within the operating room (i.e. optical devices, surgical instruments and patient). In the case of the implemented IESN application, the motions are registered by the optical tracking device based on a set of passive markers attached to each entity. The data is then transferred to ARView in order to calculate the dynamic updates corresponding to the tracked objects.

The procedure used for reading the positional data from the tracking device is implemented as a thread running concurrently and with the same priority as other processes, including video image display and volume rendering. Additionally, the variation in the

tracking signal obtained from the optical tracking device is smoothed to reduce noise. Chen [17] established that using a FIR filter could reduce the signal fluctuation during the surgical intervention. In particular, the filter produces an output signal x_n based on the weighted sum of a set of previous tracking data x_{n-i} as follows:

$$x_n = \sum_{i=1}^n w_i x_{n-i},$$

where n is the number of past data samples, and the value of the weight w_i is chosen as $1/n$ in order to act as the moving average coefficient.

The disadvantage of using a FIR filter is that it introduces high levels of latency when the involved objects are in motion. This is caused by the time required to perform the calculations each time that a new sample is polled from the tracking device. An alternative consists of using a threshold that controls the signal variation based on a velocity value. This value is computed as the difference between two consecutive readings x_i and x_{i-1} of the same tracked object. As described by Chen [17], the positional change produced by signal fluctuation is relatively smaller than the displacement generated by genuine motion. Therefore, if the variation between two consecutive readings is higher than the velocity threshold, the tracking data is considered as a real displacement of the object and applied to the corresponding virtual model. Otherwise, the data is regarded as noise and can be ignored. It is important to mention that although the use of the velocity threshold can produce latency, the generated lag is considerably lower than the delay obtained through the FIR filter. This is because only two consecutive tracking values are taken into account using the velocity threshold, whereas a larger number of past samples are required in the computation of the FIR filter to perform effectively.

A problem noticed in the implementation of this procedure is that, when two objects were tracked simultaneously, the data obtained from the tracking device became corrupt. This issue, in turn, affected the signal filter and caused the rendered volumetric model to

be displayed intermittently on screen. The source of error was found to lie in the function employed to maintain a list of class objects that correspond to the different tracked entities. In this function, the calls used to acquire data from the motion tracker had overlapping time intervals. This caused the software buffer that temporally stores the tracking data to be cleared between calls.

In order to ensure that the integrity of the data was maintained, it was decided to implement a synchronisation method in order to control the calls that have access to the tracking data. The synchronisation mechanism implemented was based on the use of semaphores, in which a global variable represented a status flag. This flag was used to provide information about the availability of the tracking data. Thus, the function call was required to read the flag value before attempting to access the positional data from the tracking device. In order to avoid data corruption, the flag value could only be changed by the object class that initiated the call. For this reason, an additional variable was used to register which object class activated the flag.

A further optimisation method used separate data structures to store the motion data corresponding to the different tracked entities. These data structures replaced the original software buffer and prevented the tracking data from being cleared between function calls. Finally, data validation was applied to the signal filter to avoid an erroneous computation of the velocity threshold value.

3.4.1 Experiment

Influence of process synchronisation in motion tracking

In this experiment, the influence of thread synchronisation during motion tracking was assessed. A comparison was performed between the original and the optimised implementations that allowed registering simultaneously the movements of different tracked objects.

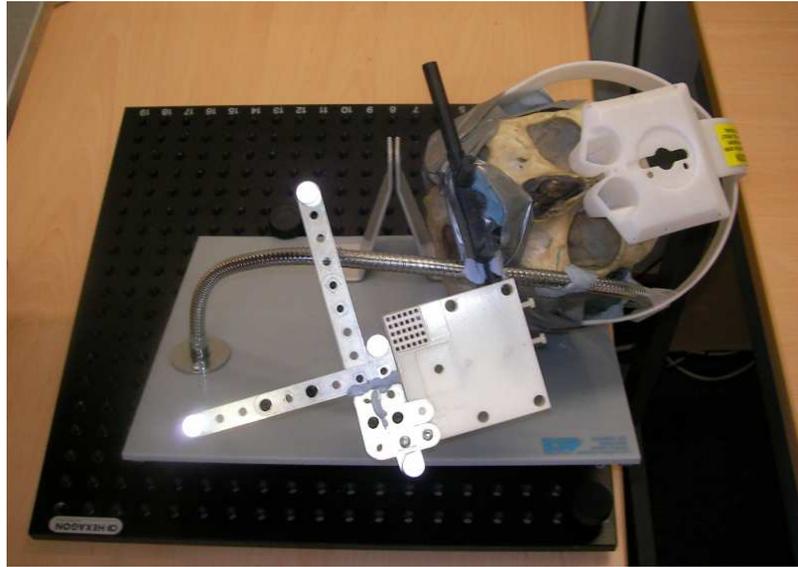


Figure 3.9: Human skull used to simulate a real patient. A set of passive markers are attached to a mouthpiece to register the skull's position over time.

The experimental setup simulated an ESS procedure, which comprises a rigid zero-degree endoscope (Figure 3.1(a)) and a human skull representing a patient (Figure 3.9). A set of passive markers were attached to both entities to detect their positions. It is worth noting that this experiment is aimed at evaluating the effects of process synchronisation in the IESN system, which affects both the hybrid Polaris and Polaris Vicra. As the difference of tracking performance between the two optical tracking devices does not influence the results of this experiment, the assessment was based solely on the use of the Polaris Vicra.

In order to relate the different coordinate systems pertaining to the endoscope and human skull, the relative transformations between them are established using the following steps (Figure 3.10):

1. *Skull to endoscope - initial:*

$$M_{se,0} = M_{e,0}M_{s,0}^{-1}.$$

2. *Skull to endoscope - in motion* (the pre-multiplication with the inverse of the initial skull-to-endoscope matrix ($M_{se,0}$) is used to set the system to coordinates (0,0,0)

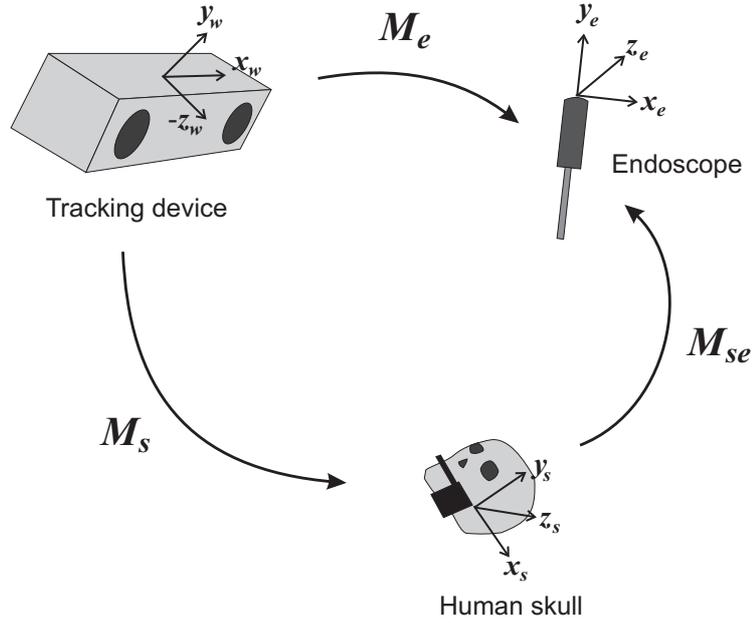


Figure 3.10: Relative transformation among the coordinate systems used to track simultaneously an endoscope and a human skull.

before the start of motion):

$$M_{se,i} = M_{se,0}^{-1}(M_{e,i}M_{s,i}^{-1}),$$

where indices e and s refer to the endoscope and skull coordinate systems, respectively.

Both endoscope and human skull were placed on a flat bench at a distance of 110 cm from the tracking device, ensuring that all passive markers were located within its central measurement volume. The three translational components corresponding to the skull position were initialised to zero (using step 2 above). Thus, the endoscope served as the world reference frame upon which the skull coordinate system was related to. However, the tracked objects were maintained static in order to avoid the influence of external motions in the evaluation.

For each original and optimised implementation, four individual tests were performed with different velocity threshold values to smooth the variation in the tracking signal. A

set of 120 frames were recorded to compare the effects of synchronisation during the signal filtering. Figure 3.11 presents the results of tracking filtering with no synchronisation for the T_z translational component of the skull (translations along T_x and T_y produced a similar trend and therefore are not shown). It can be observed from the graphs that when the raw data is obtained from the tracking device (threshold = 0.0) there is a considerable variation in the recorded tracked position. Additionally, there are some breaks in the plotted curves that correspond to corrupted data, which consequently affects the display of the volumetric model on screen. These breaks become more constant as the velocity threshold value increments, causing a severe instability during the tracking procedure when the velocity threshold reaches a value of 0.30. Figure 3.12(a) - 3.12(d) illustrate the results obtained for the translation along T_z using signal filtering with the synchronisation method applied. The graphs do not show breaks in the curves, and the variation in tracking is effectively smoothed when the velocity threshold value increases.

It must be mentioned that the fluctuation of the tracking data observed in this experiment can be caused by external factors. In particular, lighting conditions can affect the detection of passive markers due to background infrared light produced by some electric lamps. However, it was noticed that modifying the ambience luminance in the laboratory setup did not reduce the fluctuation levels of the tracking signal. Another factor could involve slight vibrations on the floor that affect the steadiness of the tracking device that stands on a tripod.

The use of a velocity threshold to smooth the signal fluctuation could introduce latency during the movements of tools in the operating theatre. In practice, this latency may be acceptable as a surgeon does not tend to move tools while inspecting a ROI, and the patient shows little or no motion during surgery.

This experiment determined the importance of process synchronisation during the registration of motion tracking when two (and possibly more) objects are tracked simulta-

neously. A velocity threshold value was used to smooth the signal noise received by the tracking device without perceived latency.

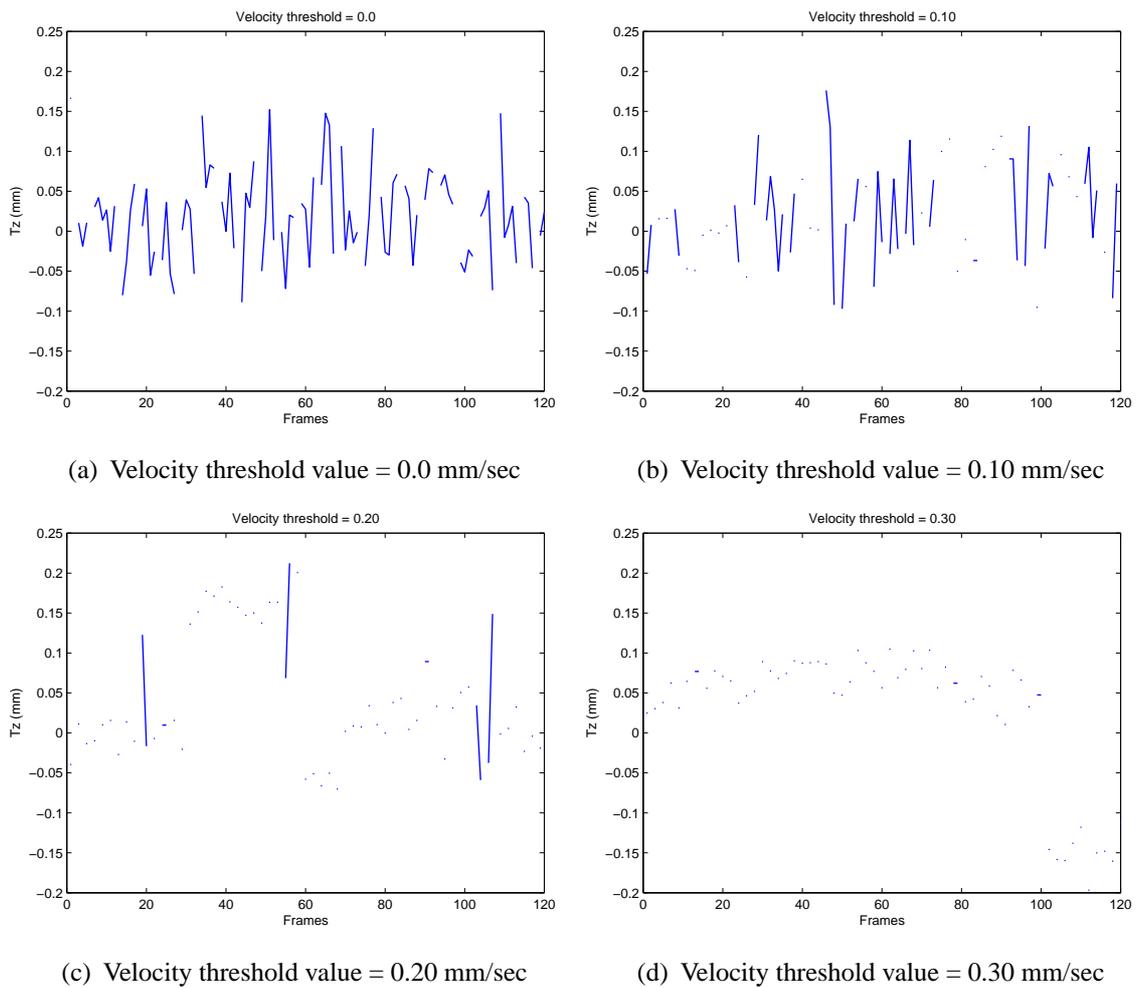


Figure 3.11: Motion tracking along the T_z component with no data synchronisation.

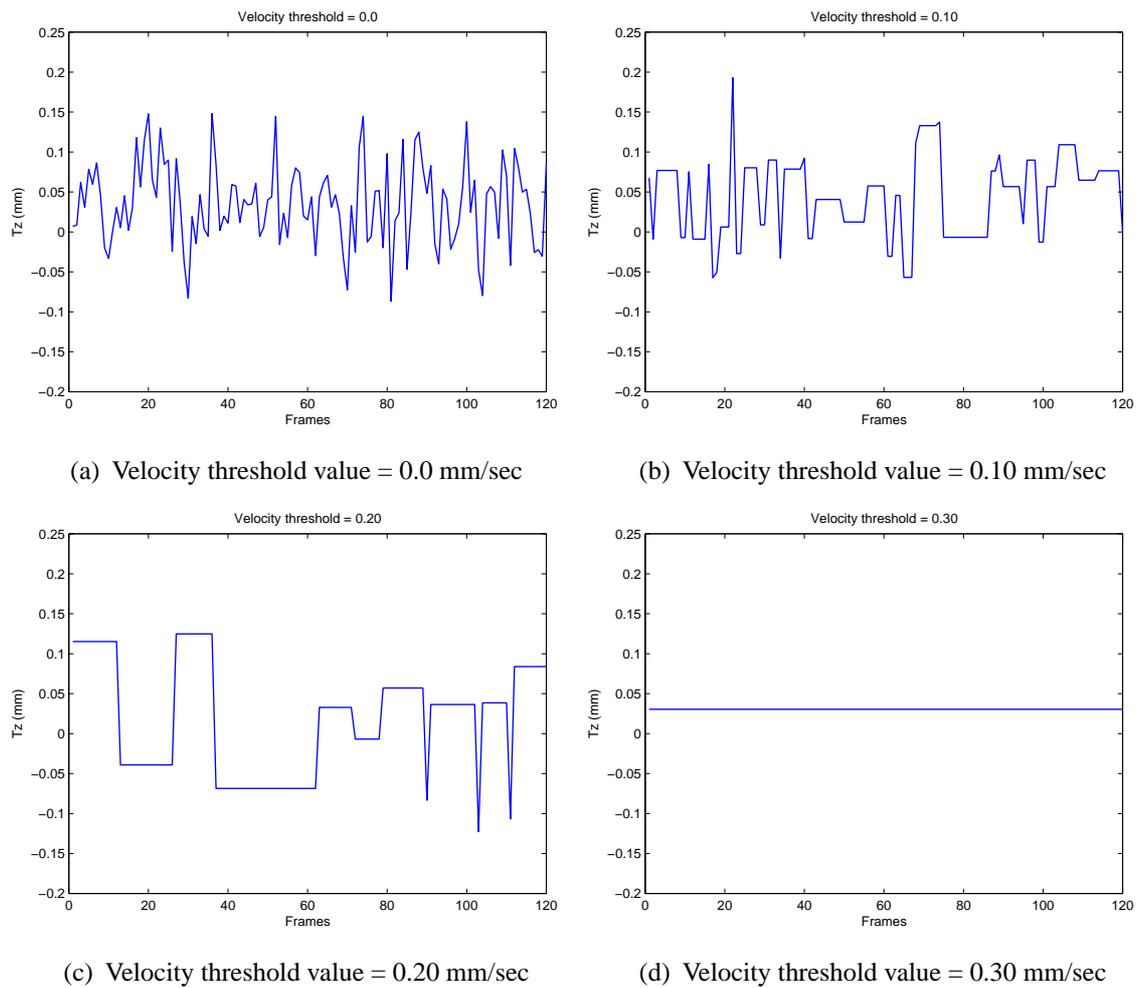


Figure 3.12: Motion tracking along the T_z component with data synchronisation.

3.5 Stereoscopic visualisation of virtual models in the IESN system

ARView supports the rendering of a scene as a single view or as a pair of stereo views for surgical procedures. A single view, or monoscopic vision, is targeted to IESN systems with only one camera (e.g. endoscope). In the case of stereo visualisation, surgical devices that have two eyepieces, such as stereoscopic microscopes, can be used to provide depth information to the user during the intervention. Stereoscopic visualisation has been demonstrated to improve the surgeon's performance when compared to the use of monocular vision [109]. In the case of AR based SN systems, Johnson et al. [110] identified that depth perception can be affected by the manner in which real and virtual models are presented to the user when the stereo overlay is displayed directly on the optics of a stereo microscope. In particular, if a virtual model that is situated below a real surface is rendered on top of it, the depth can be incorrectly estimated. The error in depth estimation can last during the entire surgical procedure or only occasionally. This problem can be reduced if a virtual rendering of the real surface is presented simultaneously with the virtual model; however, the problem cannot be completely eliminated. A temporary solution could consist in disabling the rendering of virtual models in the scene when the surgeon considers that depth estimation in the IESN system is compromised. This would allow the surgeon to regain the appropriate depth perception of the real scene and activate the virtual rendering when necessary. Alternatively, a video-based IESN system could be employed instead of an optical see-through microscope, allowing the clinician to interactively select the level of blending between virtual and real views during the intervention.

While the system presented by Johnson et al. [110] enhanced the surgeon's view using the microscope optics, the presented IESN system is a video-based AR application that requires the use of stereoscopic monitors in order to generate stereo vision. Two types of stereo devices were used during this research: LCD autostereoscopic and mirror-based

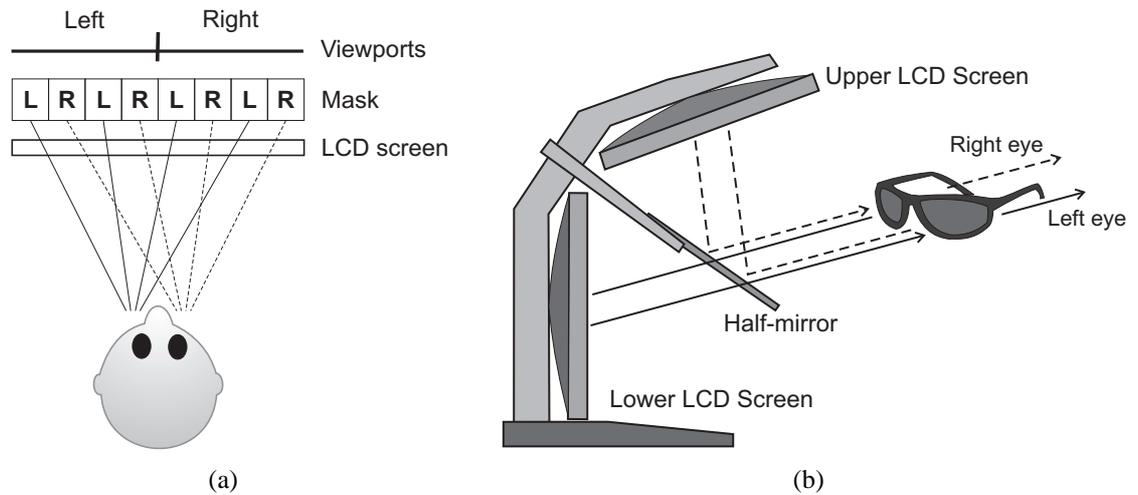


Figure 3.13: Schematic representation of stereoscopic technologies: (a) LCD autostereoscopic monitor; (b) Mirror-based stereo monitor (Image adapted from Planar Systems Inc. <http://www.planar3d.com>).

monitors. The former is based on internal optical mechanisms behind the screen that do not require the user to wear special viewing devices. Commercial mirror-based stereo monitors employ a pair of LCD screens positioned in a special configuration and polarised glasses to produce stereoscopic images. Both types of monitors work by displaying horizontally shifted images of the same scene to the left and right eyes. The brain processes the visual information acquired by the eyes into a single merged image, creating the illusion of depth. For this reason, the virtual imagery is rendered twice from two independent viewports that are displayed side by side in a single rendered window. Autostereoscopic monitors project each viewport to alternate pixel columns across a mask behind the LCD screen. The resulting interlaced image is then directed to the corresponding left and right eyes. In the case of mirror-based devices, the rendered window is expanded to two different monitors in order to display each viewport on a separate screen. A semi-transparent mirror (or half-mirror) is placed between the screens and a pair of polarised glasses filter the respective images intended for each eye. Figure 3.13 illustrates the operation of both stereoscopic technologies.

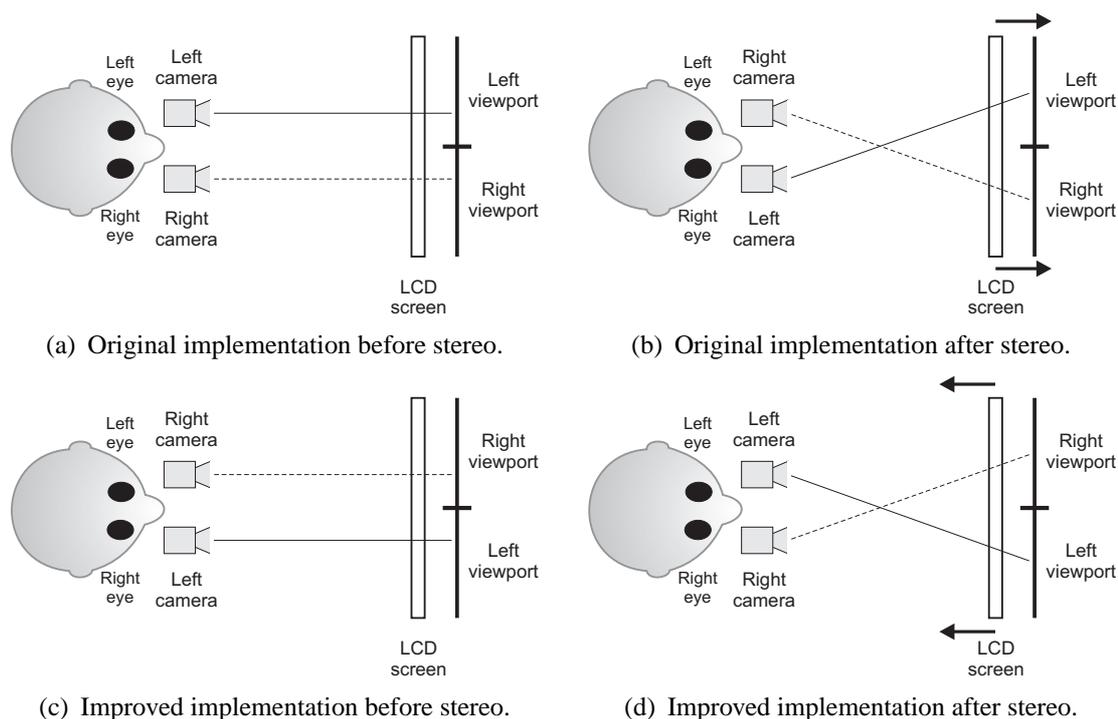


Figure 3.14: Comparison of original and improved 3D stereoscopic visualisation (only the case of autostereoscopic monitor is presented for illustration purposes). The arrows above the LCD screens in the images on the right column indicate the perceived direction where the 3D model is projected to.

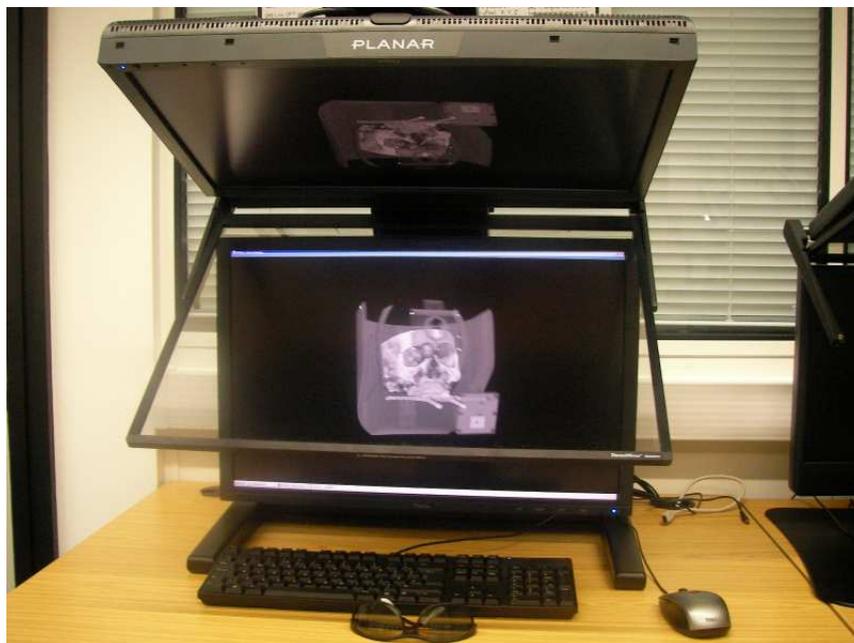
The underlying implementation in OpenGL involves rendering the stereo scene from two virtual cameras that are aligned to the real optical devices by means of camera calibration. Additionally, a full volumetric model can be displayed in stereo for inspection purposes using a set of pre-defined virtual cameras. Although the original software implementation seemed to produce an acceptable display regarding depth visualisation, the configuration of the viewports corresponding to the left and right eyes was considered to be incorrect. Specifically, the left and right virtual cameras and viewports were aligned with the corresponding eye during initialisation (Figure 3.14(a)). However, when applying a cross-viewing effect required to produce stereo vision on the stereoscopic monitor, the cameras exchanged position and were no longer associated to the intended eyes, as illustrated in Figure 3.14(b). Interestingly, during qualitative observations, it was noticed that this implementation caused the 3D image to be pushed towards the back of the monitor (a “pop in” effect).

In order to correct this issue, the procedure involved initialising the left and right viewports to their opposite locations. Similarly, the positions corresponding to left and right virtual cameras were exchanged (Figure 3.14(c)). Although the new configuration appears to be paradoxical, it conforms to the stereo cross-viewing method, where the left and right virtual cameras are aligned with the corresponding user's eyes. Also, the cameras point towards the desired image viewports as shown in Figure 3.14(d). Qualitative observations were performed using a DTI (Dimension Technologies Inc.) autostereoscopic monitor (Figure 3.15(a)) and a mirror-based Stereoscopic/3D monitor by Planar Systems Inc. (Figure 3.15(b)). The qualitative observations involved comparing the original implementation and the improved version of the stereoscopic visualisation methods. Three different 3D volumetric datasets that were available during the research were used for visual inspection. The datasets, shown in Figure 3.16, comprised an MRI fetus model with a resolution of 127 x 158 x 125 voxels, a 256 x 256 x 109 MRI human head, and a 256 x 256 x 374 CT volume of the scanned human skull introduced in section 3.4.1. This allowed evaluating the stereo capabilities of the two visualisation methods using different MRI/CT modalities and volumetric resolutions.

A group of five volunteers from the School of Computing Sciences at the University of East Anglia participated in the qualitative evaluation. All the participants were familiar with the concepts of medical visualisation and volumetric models. It should be mentioned that at this stage, no medical practitioners were available for the system evaluation. The subjects were presented with the volumetric models loaded independently into ARView using both the original implementation and the improved version. The participants were able to freely rotate the datasets in order to observe the models' anatomy from different angles. No control was imposed regarding the time a user could inspect each of the volumetric models; however, it was estimated that the maximum time spent was, in average, a couple of minutes per model.



(a)



(b)

Figure 3.15: Stereo devices used during this research: (a) LCD autostereoscopic monitor (Dimension Technologies Inc.) showing two viewports side by side; (b) Mirror-based stereo monitor (Planar Systems Inc.) where the two viewports are located on different screens.

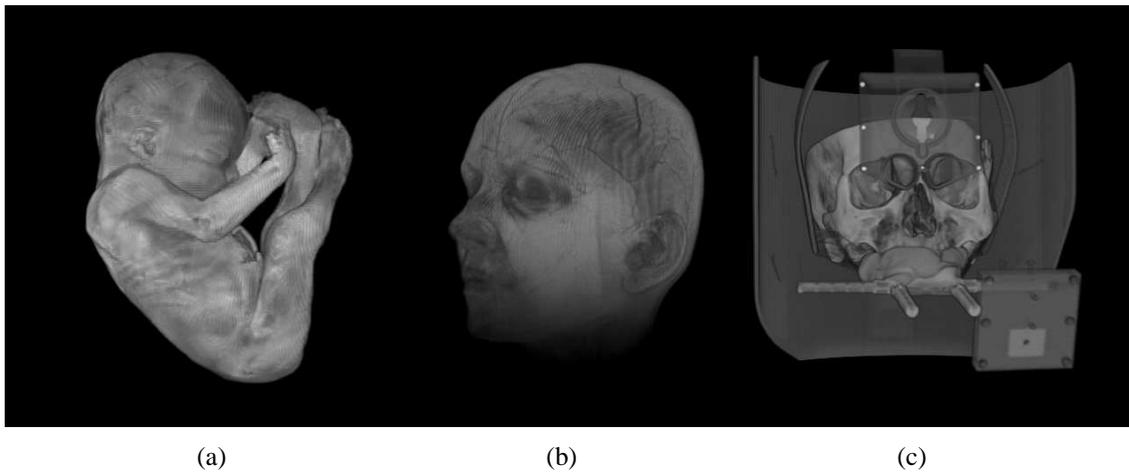


Figure 3.16: 3D volumetric datasets used for qualitative observations in stereoscopic visualisation: (a) MRI fetus model; (b) MRI human head; (c) CT human skull.

Based on the observed volumetric models, the participants were asked which stereoscopic visualisation method provided higher depth information and the perceived direction where the 3D model was projected to. For all observers, in the original implementation the datasets were perceived as being projected towards the back of the stereoscopic displays and producing low depth information. On the contrary, the users considered that the 3D images were displayed towards them using the improved visualisation method, which effectively produced the effect of “popping out” of the screen and allowed higher depth perception.

3.6 Summary

This chapter presented a series of optimisations applied to various aspects of the original IESN system. A software optimisation involved the reduction of segmentation error during the detection of feature markers in ESS camera calibration, increasing the overall calibration accuracy. Also, synchronisation between function calls that obtain data from the motion tracking device was implemented. This improved the system stability when a signal filter was applied to two objects tracked simultaneously. A final optimisation

was based on the stereo visualisation of virtual models using stereoscopic monitors. The modifications provided a correct representation of a stereo image that allows the user to observe a 3D model protruding from the screen.

Additionally, an evaluation of tracking performance compared the accuracy of translational components between two optical tracking devices — the hybrid Polaris and Polaris Vicra. The results indicate that, in a setup that simulates a surgical scenario, the accuracy substantially differs from the values described by the manufacturer. This supports the fact that the precision of a tracking system relies on the methodology used for its assessment.

Chapter 4

Camera calibration

4.1 Introduction

A camera can be described as an optical system that is employed to capture images of the real world and allow them to be displayed on a variety of output devices. From the point of view of computer vision, a camera is the central tool used to obtain information of the surrounding environment that will be analysed. Two of the main goals in computer vision are focused on 1) determining the position of objects in the scene, and 2) reconstructing the scene in three dimensions.

These principles of computer vision can be applied to other fields. For example, in robotics, the path that a robot must follow has to be dynamically updated. Video tracking recognises objects in a set of images and locates their position throughout the video sequence. In AR, the understanding of how the optical system works allows replicating the real camera to accurately superimpose computer-generated graphics on the acquired real images. As described by Tuceryan et al. [111], one of the factors that influence the success of blending real and virtual worlds consists in the exactness in modelling the optical camera used to perceive the real world.

As it will be described in the following sections, two different methods to compute camera configuration parameters are available. The first one requires special devices that allow the system to determine where the camera is placed with respect to the real world. The second method uses detected features in a set of images captured by a pair of cameras in order to compute their relative orientation. In both cases, the internal attributes of the camera(s) are also estimated. The purpose of this chapter is to demonstrate that merging the above techniques can improve the accuracy of the calibration procedure in a stereoscopic surgical microscope setup.

4.2 Methodology

4.2.1 Camera calibration

The process of calibrating a camera involves mathematically determining the parameters related to the physical characteristics of the optical system. Depending on the manufacturer specifications, a camera has different features such as maximum possible resolution, type of optical lens, number and size of light sensors (known as charge-coupled devices or CCDs), among other variables. These parameters produce different results in the image quality and projective factors within an AR application.

In order to calculate the projection parameters from points located in the 3D world to 2D picture elements on screen, or *pixels*, the calibration procedure relies on a camera simplification model known as *pinhole camera* or *general perspective model*. This approach considers the camera aperture as a single minuscule point with no lens attached to it. Thus, the computation can be described as the ideal mapping between an object and its representation on an image or retinal plane. Although this approximation does not take into account possible visual effects produced by lenses, the pinhole model provides the foundation for camera calibration in computer vision. Figure 4.1 illustrates the geometry

of the general perspective model, where a three-dimensional point M is projected on the image pixel m .

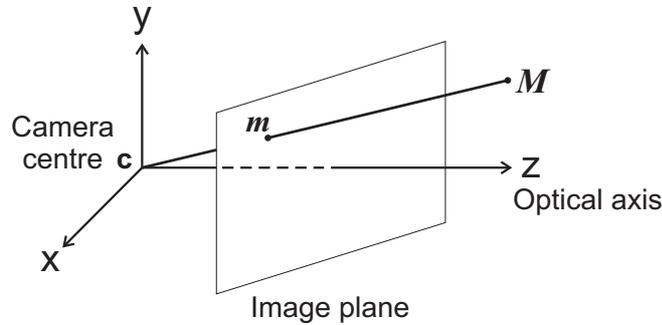


Figure 4.1: Geometry of the pinhole camera model (Image adapted from Hartley and Zisserman [112]).

The parameters to be determined during camera calibration are divided in two categories [113]:

External parameters: Also known as extrinsic parameters, describe the camera position in the real world, involving its distance and orientation with respect to a defined point or set of points in space. These parameters correspond to the six possible DOFs in a three-dimensional space: three variables for translation along t_x , t_y , and t_z and three for rotation around the coordinate axes, r_x , r_y , and r_z . The external parameters can be represented by a rotation matrix R and a translation vector T as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \quad (4.1)$$

where the 3x3 rotation matrix can also be expressed in terms of rotations around the coordinate axes as:

$$R = \begin{bmatrix} \cos r_y \cos r_z & \sin r_x \sin r_y \cos r_z - \cos r_x \sin r_z & \cos r_x \sin r_y \cos r_z + \sin r_x \sin r_z \\ \cos r_y \sin r_z & \sin r_x \sin r_y \sin r_z + \cos r_x \cos r_z & \cos r_x \sin r_y \sin r_z - \sin r_x \cos r_z \\ -\sin r_y & \sin r_x \cos r_y & \cos r_x \cos r_y \end{bmatrix}.$$

Thus, a 3D point (X_w, Y_w, Z_w) in a world coordinate system can be described in the camera coordinate system (X_c, Y_c, Z_c) by the following notation:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T, \quad (4.2)$$

which can be expanded to components as:

$$\begin{aligned} X_c &= r_{11}X_w + r_{12}Y_w + r_{13}Z_w + t_x \\ Y_c &= r_{21}X_w + r_{22}Y_w + r_{23}Z_w + t_y \\ Z_c &= r_{31}X_w + r_{32}Y_w + r_{33}Z_w + t_z. \end{aligned}$$

Internal parameters: The internal or intrinsic parameters define the optical components that affect how the light is projected inside the camera. The internal parameters relate to the following variables:

- *Focal length* (f_x, f_y) . Distance from the image plane to the point in which all projected rays of light converge, called the focal point.
- *Skew parameter* (γ) . Degree of slant between horizontal and vertical dimensions of picture elements on the image plane.
- *Principal point* (u_0, v_0) . Also known as camera centre, represents the intersection of the optical axis with the image plane, measured in pixels.

The variables f_x and f_y correspond to the focal length measured in millimeters and multiplied by a pixel scale factor in x and y dimensions, s_x and s_y respectively. In practice, pixels are assumed to be squared and the aspect ratio between s_x and s_y equal to 1. Therefore, the focal length can be considered as being the same for both variables $f_x = f_y$, or simply f . Also, the orthogonal skew parameter γ is zero for current cameras

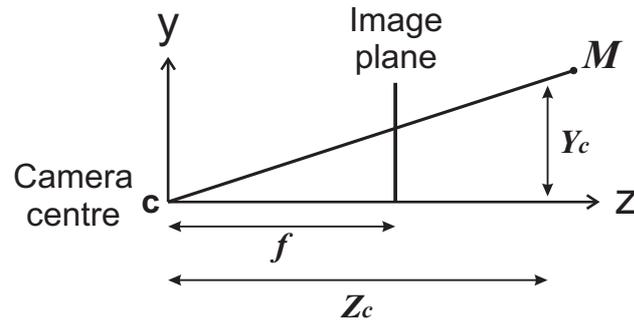


Figure 4.2: Geometry of similar triangles in the projective camera model here shown for the Y coordinate (Image adapted from Hartley and Zisserman [112]).

due to high precision in CCD components. Similar to the external parameters, intrinsic parameters can be defined by a matrix of the form:

$$K = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

A point in 3D space, expressed in camera coordinates is mapped to the 2D image plane as follows (see Figure 4.2):

$$\begin{bmatrix} X_c & Y_c & Z_c \end{bmatrix} \mapsto \begin{bmatrix} f \frac{X_c}{Z_c} + u_0 & f \frac{Y_c}{Z_c} + v_0 & 1 \end{bmatrix},$$

where u_0 and v_0 are the image coordinates of the principal point. Writing the above mapping in matrix form using homogeneous coordinates:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}.$$

Substituting Eq. 4.2 into the above equation, the projection of a 3D point in world coordinates to pixel coordinates is described as follows:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = KR \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} + KT.$$

Here, the internal and external camera matrices can be combined into a 3x4 matrix P called the *projection matrix*, where $P = K[R|T]$.

Radial Distortion

An extra parameter taken into account during camera calibration is the distortion produced by using imperfect camera lenses. This optical aberration affects the projection of points on the image by expanding or contracting their true projected position, an effect denominated radial distortion. Radial distortion is determined by the two following polynomials:

$$\begin{aligned} x &= x_d(1 + k_1r^2 + k_2r^4) \\ y &= y_d(1 + k_1r^2 + k_2r^4), \end{aligned}$$

where (x_d, y_d) and (x, y) represent the distorted and undistorted image coordinates of a point, respectively. The variable r indicates the euclidean distance between a distorted image point and the camera centre (u_0, v_0) , and is calculated as $r = \sqrt{(x_d - u_0)^2 + (y_d - v_0)^2}$. The coefficients k_1 and k_2 are known respectively as second and fourth-order degree factors. However, it has been proven that coefficients higher than a second-order degree can be neglected during lens correction because their effect in the distortion is of no considerable importance [102].

Two different radial distortion effects are produced according to the distortion coeffi-

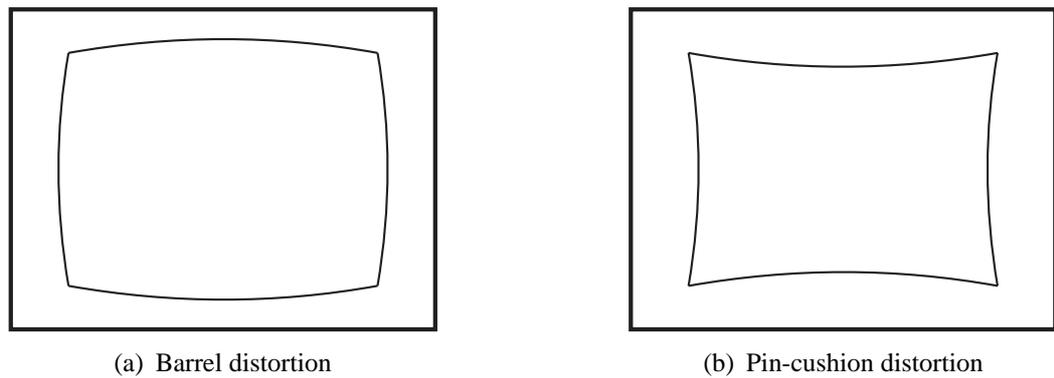


Figure 4.3: Radial distortion effects.

cient k : *a*) barrel distortion, which expands the image from the projection centre (outward effect) and corresponds to a negative value, and *b*) pin-cushion distortion that contracts the image projection towards the camera centre (inward effect) and relates to a positive distortion coefficient. Figure 4.3 exemplifies the two radial distortion effects.

Although the distortion factor computation is important to correct projection aberration; especially in “fish-eye” systems such as endoscopes; the distortion model cannot be included within the camera calibration matrix. Instead, it is computed as an independent non-linear stage. In camera systems with high focal length values such as surgical microscopes, the distortion factor tends to be small.

4.2.2 Classification of camera calibration techniques

Camera calibration methods can be broadly categorised into two classes according to the nature of physical features located in world space. According to Zhang [114], these categories can be divided as:

Photogrammetric calibration: This type of method requires the use of a calibration object with a pre-defined geometry, where a certain number of features on the object are placed at a known distance from each other. These features usually consist of squares or circular figures which can be identified through the detection of corners or by geometric

centres, also known as centroids. A photogrammetric procedure can use two variants of calibration objects: *planar grids* and *non-planar objects*. The former involve a high-quality pattern printed on a 2D surface, such as a chequerboard sheet. The latter contain geometric features at different depth levels and require a more complex construction.

It is worth noting that although a planar grid can be used for the determination of the camera parameters, it is recommended that the grid is rotated between 30° and 45° with respect to the image plane to obtain accurate results [102, 103]. This provides enough depth information in order to derive the focal length and T_z (distance between the calibration grid and the camera) values.

Self-calibration: Also known as auto-calibration, this technique does not use any pre-defined calibration object in order to compute the camera parameters. Instead, the procedure involves determining a variety of objects' features found in the real world — either indoors or outdoors. Features can include edges, corners, and regions of interest, among others.

Because the position of 3D features are not known in advance, self-calibration requires using different images of the same scene in order to derive depth information. For this purpose, a set of feature points located on one image must be also detected on the other views in order to associate the independent viewpoints. The so-called *corresponding points* provide the basic input to perform camera calibration. Figure 4.4 represents a set of corresponding points on two different views of the same scene.

As described previously, the calculation of the 3×4 projection matrix P — comprising the 11 internal and external parameters — defines the necessary information between metric world coordinates and 2D image coordinates. The advantage of using a photogrammetric calibration methods relies on 1) well-established procedures to recover the camera's geometry, and 2) knowledge about the physical world (with respect to the calibration object). However, for systems with multiple camera configurations, each group of 11 pa-

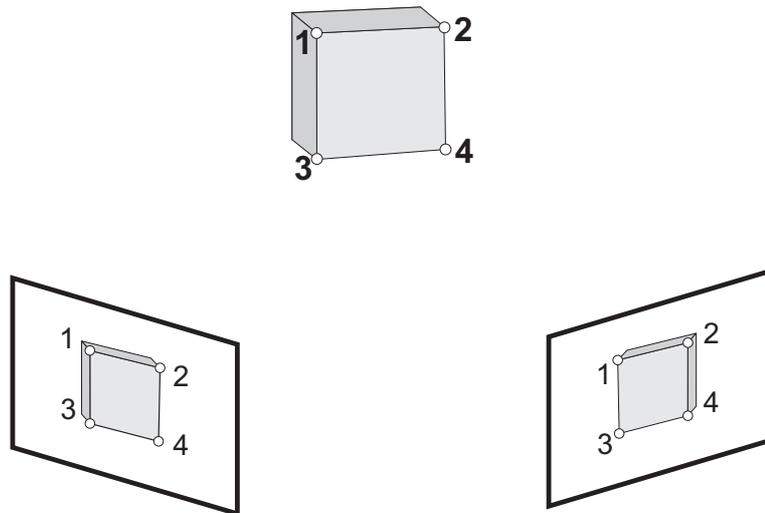


Figure 4.4: Illustration of corresponding points on two independent views.

rameters has to be determined independently for each optical device (e.g. in a stereo setup, 22 different parameters need to be estimated).

4.2.3 Stereo camera calibration

Stereo calibration involves determining the camera parameters based on the relationship between a pair of images. This relationship is achieved through the calculation of a geometric constraint between two cameras looking at the same scene. Such constraint is known as *epipolar geometry*.

The epipolar geometry, shown in Figure 4.5, includes the image planes of each camera, where a 3D point M is projected on the left and right views as m and \hat{m} , respectively. The line that joins the two camera centres C and \hat{C} is called the baseline, and the intersection points between baseline and image planes are called epipoles (ep and $e\hat{p}$). The line between the left epipole and its related image point is known as epipolar line lm (respectively $l\hat{m}$). The epipolar constraint indicates that for a point m in the left image, its corresponding point \hat{m} must lie on the epipolar line $l\hat{m}$ and vice versa.

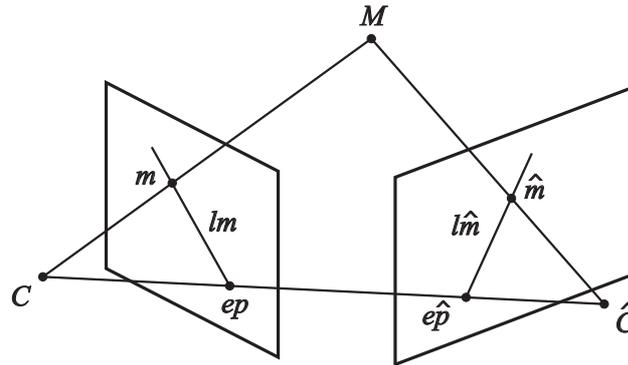


Figure 4.5: Diagram of epipolar geometry for two cameras (Image adapted from Trucco and Verri [115]).

Fundamental matrix

The epipolar geometry can be represented algebraically as a special 3x3 matrix known as the *fundamental matrix* [116]. The fundamental matrix F encodes the relationship between the two cameras using only the projective information from a set of corresponding image points. Thus, the epipolar constraint that is used to associate a point $m = [x, y, 1]^T$ to its corresponding point $\hat{m} = [\hat{x}, \hat{y}, 1]^T$ is defined as:

$$\hat{m}^T F m = 0. \quad (4.3)$$

The mapping of a point on the left view to its related epipolar line on the right image is given by $l\hat{m} = Fm$, whereas $lm = F^T\hat{m}$ represents the mapping between a right point and its left epipolar line. In the case of a calibrated setup, where the intrinsic camera matrices are known, a similar representation can be obtained by substituting F by its counterpart *essential matrix* E [117]. However, the corresponding pixels m and \hat{m} are replaced by points p and \hat{p} with respect to the camera coordinate system, known as points in *normalised coordinates* [112]. Therefore, the epipolar constraint becomes

$$\hat{p}^T E p = 0,$$

where $E = [t]_{\times}R$ comprises the translation and rotation between camera positions. The notation $[t]_{\times}$ is defined as a skew-symmetric matrix:

$$[t]_{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}.$$

Finally, the relationship between the fundamental and essential matrices given both intrinsic camera matrices K and \hat{K} is the following:

$$E = \hat{K}^T F K, \quad (4.4)$$

or in terms of the essential matrix as:

$$F = \hat{K}^{-T} E K^{-1},$$

also expanded of the form

$$F = \hat{K}^{-T} [t]_{\times} R K^{-1}.$$

In practice, the difference between fundamental and essential matrices relies on the fact that the former encodes the internal and external parameters based only on the information obtained from corresponding image points. On the other hand, the essential matrix captures only the relative orientation between a pair of cameras based on the known intrinsic parameters and normalised image points.

Estimation of the fundamental matrix

As described in the previous subsection, the estimation of the fundamental matrix F depends on satisfying the epipolar constraint between a pair of points. Thus, Equation 4.3

can be expanded as:

$$\begin{bmatrix} \hat{x} & \hat{y} & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0.$$

The calculation involves solving a homogeneous system of equations from a set of corresponding points $\hat{m}_i \leftrightarrow m_i$, where at least a minimum of seven correspondences are required in order to find a solution ($n \geq 7$). A linear equation is established for each pair of points using the x and y coordinates as coefficients of the form:

$$\hat{x}x f_{11} + \hat{x}y f_{12} + \hat{x} f_{13} + \hat{y}x f_{21} + \hat{y}y f_{22} + \hat{y} f_{23} + x f_{31} + y f_{32} + f_{33} = 0. \quad (4.5)$$

If Equation 4.5 is considered as a product between two vectors, it can be expressed as

$$\begin{bmatrix} \hat{x}x & \hat{x}y & \hat{x} & \hat{y}x & \hat{y}y & \hat{y} & x & y & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0.$$

For a set of n corresponding points in two images, a similar number of linear equations are obtained of the form $Af = 0$, where

$$Af = \begin{bmatrix} \hat{x}_1 x_1 & \hat{x}_1 y_1 & \hat{x}_1 & \hat{y}_1 x_1 & \hat{y}_1 y_1 & \hat{y}_1 & x_1 & y_1 & 1 \\ \vdots & \vdots \\ \hat{x}_n x_n & \hat{x}_n y_n & \hat{x}_n & \hat{y}_n x_n & \hat{y}_n y_n & \hat{y}_n & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ \vdots \\ f_{32} \\ f_{33} \end{bmatrix} = 0.$$

Different techniques have been developed in order to determine the fundamental matrix. Each method has different implementation complexities and sensitivity to wrong point localisation. A survey of available methods can be found in the literature comparing their accuracy as well as their advantages and disadvantages [113, 118, 119]. The following subsection will introduce the reader to three techniques selected from the established algorithms.

Linear method

In order to estimate the fundamental matrix by a linear technique, seven point correspondences can be used to provide enough information about the scene. However, in case of bad point localisation produced by Gaussian noise, the method behaves erratically. In practice, eight or more corresponding points are used to create an overdetermined system of equations, which allows redundancy of points in case of poor point localisation. The technique, originally introduced by Longuet-Higgins [117], is known as the *8-point algorithm*.

The 8-point algorithm produces a solution by applying a least-square technique that finds the eigenvector corresponding to the smallest singular value of matrix A . The computation can be performed through the SVD method [107]. The factorisation of matrix A

becomes $A = UDV^T$, where U and V are orthogonal matrices and D is a diagonal matrix with nonnegative values. The last column of V^T relates to the eigenvector that solves the least-squares problem by minimising the algebraic error $\|Af\|$ under the constraint $\|f\| = 1$. This eigenvector corresponds to the desired fundamental matrix.

An extension of the 8-point algorithm consists in normalising the corresponding points before solving the system of equations. As Hartley [120] pointed out, the lack of uniformity in image coordinates produces instabilities in the algorithm. Specifically, if a pair of corresponding pixel points m and \hat{m} lie on image coordinates (100,100,1), the row of matrix A related to the vector multiplication between them (see Equation 4.5) will be of the order of magnitude $(10^4, 10^4, 10^2, 10^4, 10^4, 10^2, 10^2, 10^2, 1)$. This affects the computation of least squares as the difference between the entries in the linear equation is notably high.

Point normalisation comprises the translation of the origin of image coordinates from the left hand corner to the centre of the image. Additionally, point coordinates are scaled so that their average distance from the image centre is equal to $\sqrt{2}$, which forces corresponding points to have a similar magnitude [115]. Because of the resulting improvements and the simplicity of implementation, point normalisation is also used in other non-linear algorithms to find the fundamental matrix.

Gradient-based technique

A gradient-based method is an iterative technique that requires an initial estimation of the fundamental matrix in order to find an optimal solution. A basic approach for optimisation involves minimising the residual of Equation 4.3 based on the cost $\min_F \sum_i (\hat{m}_i^T F m_i)$. However, this procedure may generate an inaccurate estimation due to the fact that the variance of each point is not always equivalent — an optimal situation for least-squares methods [119]. In order to overcome this problem, the minimisation is based on the

gradient of the fundamental matrix as follows:

$$\min_F \sum_i (\hat{m}_i^T F m_i)^2 / g_i^2, \quad (4.6)$$

where $g_i = \sqrt{l_1^2 + l_2^2 + \hat{l}_1^2 + \hat{l}_2^2}$ is the gradient of F , and l_1, l_2 are the first two components of the epipolar line: $F\hat{m}_i = [l_1, l_2, l_3]^T$, and $F^T m_i = [\hat{l}_1, \hat{l}_2, \hat{l}_3]^T$. The minimisation method involves eigen analysis to iteratively compute the fundamental matrix through SVD, as described previously. Each step reduces the residual of Equation 4.6 until the objective function reaches a specific threshold. Similar to the linear 8-point algorithm, gradient-based methods can deal with bad point localisation produced by noise. Nevertheless, gradient-based techniques tend to be more accurate than linear algorithms due to the inherent iterative optimisation.

M-Estimator

M-estimator, where M stands for maximum likelihood, is a technique that optimises an initial fundamental matrix by reducing the effect of outliers or improper matches between a pair of corresponding points. The optimisation is achieved by calculating the residual r of each point correspondence i , obtained using the equation $r_i = \hat{m}_i^T F m_i$. A weight function w_i modifies the effects of r_i for each correspondence, becoming:

$$\min_F \sum w_i (\hat{m}_i^T F m_i)^2.$$

Among the different proposed weight functions, the one introduced by Huber [121] is defined as:

$$w_i = \begin{cases} 1 & |r_i| \leq \sigma \\ \sigma / |r_i| & \sigma < |r_i| \leq 3\sigma \\ 0 & 3\sigma < |r_i| \end{cases},$$

in which σ is a robust standard deviation that indicates the median of the absolute values of the residuals. This standard deviation estimate is obtained by the formula [118, 119]: $\sigma = 1.4826(1 + 5/(n - p))\text{median}_i |r_i|$, where the coefficient 1.4826 is a constant value used in the assumption that Gaussian noise corrupts the data. Specifically, this constant corresponds to the median of absolute random values obtained from a Gaussian normal distribution that is $\approx 1/1.4826$. The expression $5/(n - p)$ is used to compensate for the effect of a small number of data points, where n is related to the size of the data set and p corresponds to the dimension of the parameter vector.

M-estimators are considered as robust methods because they aim at overcoming outliers produced by wrong localisation. Also, their accuracy is higher in presence of Gaussian noise in comparison with linear methods due to the inherent standard deviation computation.

4.2.4 Estimation of intrinsic parameters in stereo cameras

The fundamental matrix F encapsulates a certain number of DOFs that relate to the possible camera parameters to be calculated. Being a 3×3 homogeneous matrix with nine elements, there are only eight independent parameters because the projective scale factor f_{33} , or scalar value, is not a significant parameter. Moreover, the determinant of F is equal to zero (since the determinant of $[t]_{\times}$ is also zero), which eliminates an extra parameter [112]. Thus, only seven DOFs are available for estimation in a stereo camera configuration.

From the seven DOFs, five of them are related to the relative position between cameras. In fact, three correspond to the relative rotation and the other two to the relative direction of translation (because of the projective homogeneous condition). Therefore, the remaining degrees of freedom can be used to estimate two out of the ten intrinsic camera parameters (five for each camera) in both optical devices. From these intrinsic

sic parameters, pixels can be considered rectangular and accurate enough to avoid skew effects. Additionally, Bougnoux [122] demonstrated that during self-calibration, the principal point can be regarded as lying in the centre of the image. If these assumptions are taken into account, the camera parameters can be reduced to the computation of two focal lengths in the pair of optical devices.

The extraction of focal lengths from two views can be derived from different approaches. One of the first stereo-based techniques was developed by Hartley [123] using algebraic manipulations of the fundamental matrix, which was subsequently redefined based on projective geometry [124]. A more recent method given by Newsam et al. [125] extends the idea of decomposing the fundamental matrix using SVD. This technique provides a linear system of equations in which the pair of focal lengths f and \hat{f} can be different. The pseudocode of Newsam's method is presented in Code 4.1.

1. Compute the SVD of F , written in the form

$$F = UDV^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]^T$$

where \mathbf{u}_i and \mathbf{v}_i represent the columns i of matrices U and V , respectively. The diagonal matrix D contains the non-negative singular values σ_1 and σ_2 .

2. Establish the linear system of equations

$$\begin{aligned} \sigma_1^2 &= (u_1^T \mathbf{f}_3)^2 w_1 + ((u_1^T \mathbf{i}_3)^2 + (u_3^T \mathbf{i}_3)^2) w_2 + w_3 \\ 0 &= (u_2^T \mathbf{f}_3)(u_1^T \mathbf{f}_3) w_1 + (u_1^T \mathbf{i}_3)(u_2^T \mathbf{i}_3) w_2 \\ \sigma_2^2 &= (u_2^T \mathbf{f}_3)^2 w_1 + ((u_2^T \mathbf{i}_3)^2 + (u_3^T \mathbf{i}_3)^2) w_2 + w_3 \end{aligned}$$

in which \mathbf{f}_3 corresponds to the last column of the fundamental matrix F , \mathbf{i}_3 is a unit vector $[0, 0, 1]^T$, and w_i (where $i = 1 \dots 3$) are the desired variables.

3. Group and solve the previous linear equations in the following form

$$Ax = b,$$

$$\begin{bmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 \\ 0 \\ \sigma_2^2 \end{bmatrix}$$

where

$$\begin{aligned} a_{11} &= (u_1^T \mathbf{f}_3)^2 & a_{12} &= (u_1^T \mathbf{i}_3)^2 + (u_3^T \mathbf{i}_3)^2 \\ a_{21} &= (u_2^T \mathbf{f}_3)(u_1^T \mathbf{f}_3) & a_{22} &= (u_1^T \mathbf{i}_3)(u_2^T \mathbf{i}_3) \\ a_{31} &= (u_2^T \mathbf{f}_3)^2 & a_{32} &= (u_2^T \mathbf{i}_3)^2 + (u_3^T \mathbf{i}_3)^2 \end{aligned}$$

4. The resulting variables w_1, w_2 and w_3 relate to the quantities

$$w_1 = -\mu, \quad w_2 = \lambda v, \quad w_3 = \lambda,$$

where $\mu = f^{-2} - 1$ and $v = \hat{f}^2 - 1$.

5. Finally, the focal lengths are obtained by working out the values

$$f = \frac{1}{\sqrt{-w_1 + 1}}, \quad \text{and} \quad \hat{f} = \sqrt{\frac{w_2}{w_3} + 1}.$$

Code 4.1: Pseudocode for Newsam's focal length estimation [125].

Bougnoux [122] developed a closed-form solution to obtain unequal focal lengths based on a constraint between the camera internal parameters and an imaginary conic lying on the plane at infinity, known as the *absolute conic*. The projection of the absolute conic in two views determines an algebraic correspondence with the epipolar lines. The resulting polynomial equations are called *Kruppa equations* [112, 113] which can be represented as:

$$Q = KK^T = \begin{bmatrix} f^2 + u_0^2 & u_0v_0 & u_0 \\ u_0v_0 & f^2 + v_0^2 & v_0 \\ u_0 & v_0 & 1 \end{bmatrix},$$

where $K = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$ is the intrinsic matrix of one camera. The estimation of the pair of focal lengths is described in Code 4.2.

1. Compute the SVD of the fundamental matrix, $F = UDV^T$
2. Obtain the epipole e on the left image, which corresponds to the right null vector of F (last column of matrix V^T that relates to the null singular value of the fundamental matrix).
3. Similarly, the epipole \hat{e} on the right image is obtained from the left null vector of F (last column of the matrix U that corresponds to the null singular value).
4. Estimate the focal length f for the left camera using the formula:

$$f = \sqrt{\frac{-\hat{c}^T[\hat{e}]_{\times}\tilde{I}Fc(c^TF^T\hat{c})}{\hat{c}^T[\hat{e}]_{\times}\tilde{I}\tilde{I}F^T\hat{c}}},$$

in which $\tilde{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and $c = [u_0, v_0, 1]^T$ and $\hat{c} = [\hat{u}_0, \hat{v}_0, 1]^T$ are the principal points related to the left and right camera, respectively.

5. The computation for the second focal length \hat{f} is calculated by transposing the fundamental matrix and inverting the position of the cameras as follows:

$$\hat{f} = \sqrt{\frac{-c^T[e]_{\times}\tilde{I}F^T\hat{c}(\hat{c}^TFc)}{c^T[e]_{\times}\tilde{I}F^T\tilde{I}Fc}}.$$

Code 4.2: Pseudocode for the estimation of focal length based on Bougnoux's method [122].

Another method introduced by Sturm [126] and Sturm et al. [127] aims at estimating identical focal lengths from a pair of cameras. The fundamental matrix F is multiplied by an intermediate matrix with initial values based on assumed intrinsic parameters; the resulting matrix G is called a semi-calibrated fundamental matrix. The involved steps in the calculation are described in Code 4.3.

1. Calculate the initial fundamental matrix and obtain the semi-calibrated matrix

$$G \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hat{u}_0 & \hat{v}_0 & 1 \end{bmatrix} F \begin{bmatrix} 1 & 0 & u_0 \\ 0 & 1 & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where u_0, v_0 and \hat{u}_0, \hat{v}_0 are the principal points on left and right cameras, respectively.

2. The matrix G is multiplied by an additional scaling matrix in order to work in normalised image coordinates. The scaling has the form

$$\begin{bmatrix} f_0 & 0 & 0 \\ 0 & f_0 & 0 \\ 0 & 0 & 1 \end{bmatrix} G \begin{bmatrix} f_0 & 0 & 0 \\ 0 & f_0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

in which f_0 is a scale factor with a value significantly larger (of the order of 10^3) than the highest expected focal length. The purpose of this additional scaling is to improve the stability and numerical conditioning of the focal length computation.

3. Calculate the SVD of matrix G ($G = UDV^T$) and construct one quadratic and two linear equations based on the extracted coefficients $u_{31}, u_{32}, v_{31}, v_{32}$ of matrices U and V^T , and singular values σ_1 and σ_2 :

$$\begin{aligned} f^2(\sigma_1 u_{31} u_{32} (1 - v_{31}^2) + \sigma_2 v_{31} v_{32} (1 - u_{32}^2)) + u_{32} v_{31} (\sigma_1 u_{31} v_{31} + \sigma_2 u_{32} v_{32}) &= 0 \\ f^2(\sigma_1 v_{31} v_{32} (1 - u_{31}^2) + \sigma_2 u_{31} u_{32} (1 - v_{32}^2)) + u_{31} v_{32} (\sigma_1 u_{31} v_{31} + \sigma_2 u_{32} v_{32}) &= 0 \\ f^4[\sigma_1^2 (1 - u_{31}^2) (1 - v_{31}^2) - \sigma_2^2 (1 - u_{32}^2) (1 - v_{32}^2)] + \\ f^2[\sigma_1^2 (u_{31}^2 + v_{31}^2 - 2u_{31}^2 v_{31}^2) - \sigma_2^2 (u_{32}^2 + v_{32}^2 - 2u_{32}^2 v_{32}^2)] + [\sigma_1^2 u_{31}^2 v_{31}^2 - \sigma_2^2 u_{32}^2 v_{32}^2] &= 0 \end{aligned}$$

4. Solve any of the previous equations. The obtained focal length will be multiplied by the inverse of f_0 in order to undo the scaling factor of Step 2.

Code 4.3: Pseudocode for focal length estimation based on Sturm's technique [126, 127].

As pointed out by Sturm, the solution is generally obtained by solving the quadratic equation. However, the linear equations can be used to find false or erroneous results.

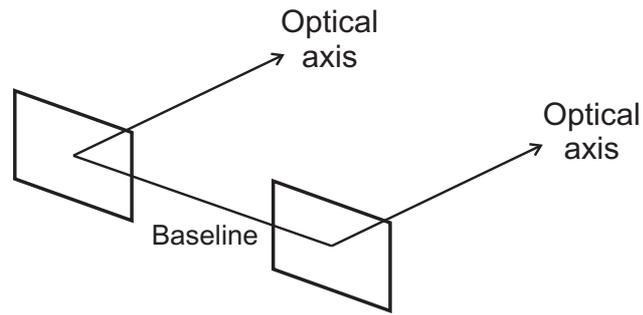
Although other closed-form solutions have been described by Kanatani and Matsunaga [128] and Ueshiba [129], their work is restricted to theoretical analysis and do not provide proof of practical results. Nevertheless, they agree with other authors about the significance of avoiding special cases in which the calculation of focal lengths leads to singularities in the equations. These so-called *degenerate* or *critical configurations* arise when the focal length can not be solved by any means. General critical configurations occur in the following cases:

1. When the optical axes of the two cameras and the baseline are coplanar, either by *a)* having parallel optical axes or *b)* presenting an isosceles triangle setup in which the optical axes intersect at a finite point and principal points are equidistant from this point.
2. The plane defined by one optical axis and the baseline, and the plane defined by the baseline and the other optical axis are orthogonal.

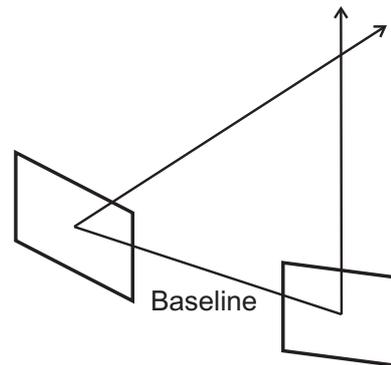
These critical configurations are illustrated in Figure 4.6

In practice, degenerate configurations produce negative values in the focal lengths in the case of linear equations, or imaginary complex numbers when using squared roots to estimate the solution. As described by Hartley [124] and Kanatani et al. [130], a procedure to assess if there is a critical configuration between the pair of cameras involves determining whether the epipolar line (e.g. lm) of one camera passes through the principal point of the second camera. This can be visually represented in Figure 4.7

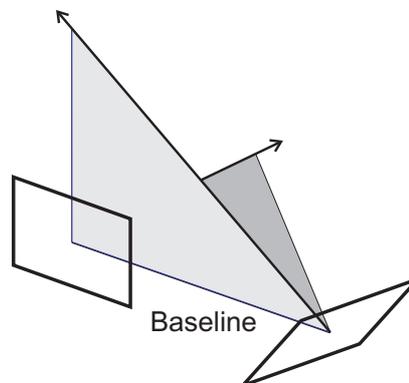
Sturm et al. [126, 127] described a practical method to avoid singularities between a pair of cameras in a stereo setup, specifically in the case of parallel optical axes. The approach involves capturing one image at an arbitrary position and tilt slightly the other



(a) Parallel optical axes



(b) Intersection at a finite point



(c) Planes defined by the baseline and optical axes are orthogonal

Figure 4.6: Critical configurations which prevent the calculation of focal length.

camera upwards or downwards before taking the second image. The results presented indicate that even a slight elevation angle between 2° and 3° off the base plane produce favourable results. Additionally, when the convergence angle between optical axes is around 10° or higher the focal length error decreases considerably. Figure 4.8 represents the tilt (elevation) and convergence rotations between a stereo pair of cameras.

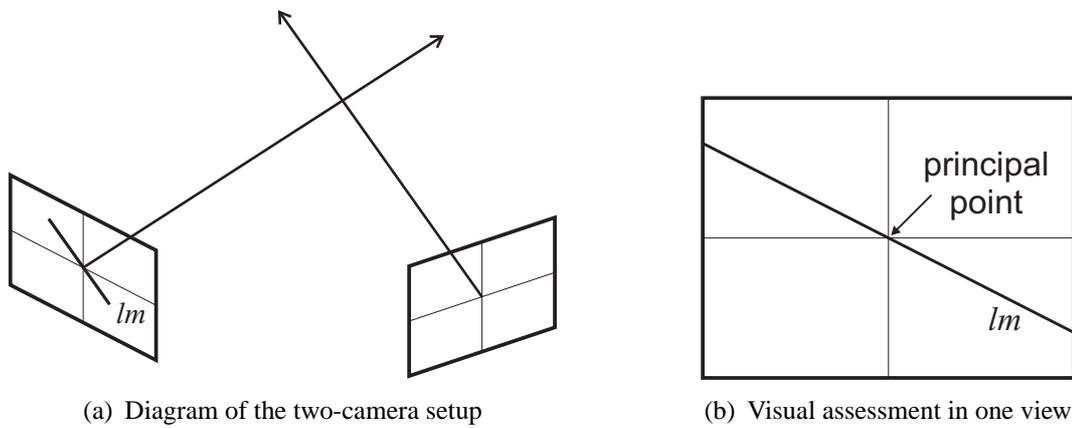


Figure 4.7: Critical configuration when the epipolar line lm of one camera passes through the principal point of the second one.

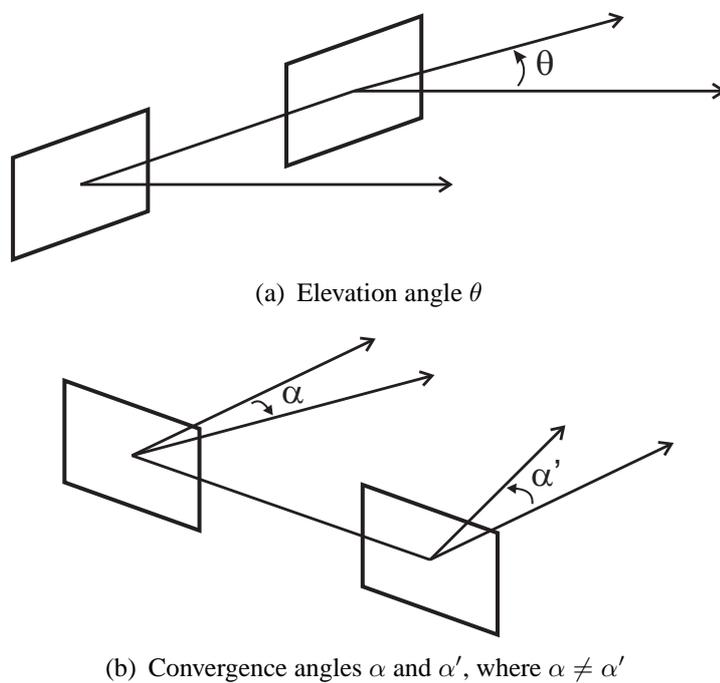


Figure 4.8: Methods to avoid critical camera configurations as proposed by Sturm et al. [126, 127].

4.2.5 Focal length optimisation in a stereo camera setup

One of the limitations of estimating the focal length through self-calibration is that, in order to recover a reliable solution, the effects of radial distortion for each of the cameras must be corrected. Tordoff and Murray [131] demonstrated that in the case of rotating cameras, even a low pin-cushion distortion value results in underestimated focal lengths, whereas barrel distortion either produces overestimated quantities or makes the computation fail. Moreover, it is customary to fix other internal parameters (e.g. principal point) in the pair of cameras for the calculation of focal lengths. As it was described by Vigueras et al., fixing intrinsic parameters “only makes sense when optical distortion is considered. Otherwise, the use of constant intrinsic parameters is not well founded” [132]. Therefore, a sensible procedure involves “pre-calibrating” each camera independently through a photogrammetric technique before calculating the fundamental matrix and focal length values. Besides correcting the influence of radial distortion, the obtained pre-calibrated parameters can be used as ground truth for further analysis.

The pair of focal lengths computed by any of the methods described in Section 4.2.4 can be refined through an optimisation algorithm. As the initial extrinsic and intrinsic camera parameters have already been estimated at the pre-calibrated stage, such knowledge can be included in a cost function in order to improve the solution. The cost function that has been selected relies on a metric known as the Sampson distance [112], which is a first-order approximation to a geometric, or reprojection, error measured in left and right images. The cost function is defined as:

$$\sum_i \frac{(\hat{p}_i^T E p_i)^2}{(E p_i)_1^2 + (E p_i)_2^2 + (E^T \hat{p}_i)_1^2 + (E^T \hat{p}_i)_2^2}, \quad (4.7)$$

where E relates to the essential matrix obtained from Equation 4.4 and it is based on the pair of pre-calibrated intrinsic camera matrices K and \hat{K} . p and \hat{p} represent the i corresponding image points m and \hat{m} in normalised coordinates, which are obtained from

$p = K^{-1}m$ and $\hat{p} = \hat{K}^{-1}\hat{m}$, respectively. $(Ep_i)_j^2$ refers to the square of the j -th entry of the vector Ep_i (similar for the vector $E^T\hat{p}_i$). Therefore, the parameters to optimise involve the intrinsic camera matrices with respect to the pair of focal length values. A similar cost function has been used by Rodehorst et al. [133] for relative pose estimation in multi-camera scenarios, while Stoyanov et al. [134] employed the Sampson distance error for stereo laparoscopy in terms of the fundamental matrix.

The overall procedure for focal length optimisation in a stereo camera setup can be summarised as follows:

1. Calculate the fundamental matrix, F , from at least seven corresponding points on the calibration object, in the left and right images. The three methods considered include: linear (8-point algorithm), the gradient-based technique and the M-estimators.
2. Compute the focal length for each of the cameras from the previous estimated fundamental matrix. The self-calibration based methods presented are: Newsam (Code 4.1), Bougnoux (Code 4.2) and Sturm (Code 4.3).
3. The focal lengths obtained can be refined through an optimisation algorithm that uses the Sampson distance as a cost function, as defined in Equation 4.7. Two evolutionary algorithms (i.e. Self-adaptive Differential Evolution and CODEQ, described in Appendix B) and the well-known Levenberg-Marquardt method can be used among the different optimisation methods.

4.2.6 Camera pre-calibration using a non-coplanar calibration object

As explained earlier, photogrammetric techniques require the use of a calibration object with known coordinates in order to estimate the camera parameters, while stereo self-calibration only involves a set of corresponding features at different depths for the pair of views. The adopted approach utilised a non-coplanar 3D calibration object in order to perform both pre-calibration and focal length optimisation.

Previous research described in [13, 17] comprised an evaluation of single camera calibration between coplanar and non-coplanar calibration devices in an endoscopic surgical configuration. The non-coplanar calibration object consisted of a modified pin-art setup with pins positioned at various depths. In the case of a surgical microscope, the magnification level is of a higher degree than the magnification obtained by an endoscope. Because of this, the considerable large dimensions of the non-coplanar object designed for endoscopic use could not be employed in the microscope setup. This issue led to the creation of a different 3D calibration object, with the purpose of reducing its physical dimensions and improving its precision.

The manufactured calibration object consists of a set of blocks at different levels that provides enough depth information to perform a non-coplanar calibration procedure. On top of each block, a circular shape with a contrasting colour was inscribed in order to be detected at various orientations. Although an alternative solution could involve the use of squared shapes to detect corners, the current resolution of the manufacturing device (3D Zprinter[©] 450 - Z Corporation) limits the output quality by producing irregular edges. The length and width of the resulting manufactured calibration object (shown in Figure 4.9) are 24 mm and 19 mm, respectively. In the case of height, the range of depth levels varies between 2.5 mm and 12 mm from the base of the calibration object, with a mean difference of 1.6 mm among contiguous blocks.



Figure 4.9: 3D non-coplanar calibration object manufactured with 3D Zprinter[®] 450 - Z Corporation.

Based on the current object design, the localisation of visual markers on the calibration object is performed through detecting the contours of each circular shape. These contours are determined by applying a Gaussian filter [99] in order to eliminate image noise, followed by a Canny's edge detection algorithm [100] and connected component labelling [101]. Because of the possible inclinations that the calibration object will be positioned at, the printed circular markers can be projected on the image as perfect circles or ellipses. Therefore, a method based on the work by Ho and Chen [135] is used to detect circular/elliptical shape centres. The technique works by linearly scanning the image in two stages, one horizontal and one vertical. Each scan finds the contours corresponding to individual shapes and calculates the middle point between the extremes of each figure's cross section line. A Hough transform line [136] is fitted through the set of middle points, which results in a symmetric vertical (L_v) or horizontal (L_h) line that divides the shape in two equal parts. Finally, the intersection between both symmetric lines determines the central point C of the shape. Figure 4.10 illustrates the procedure to detect elliptical

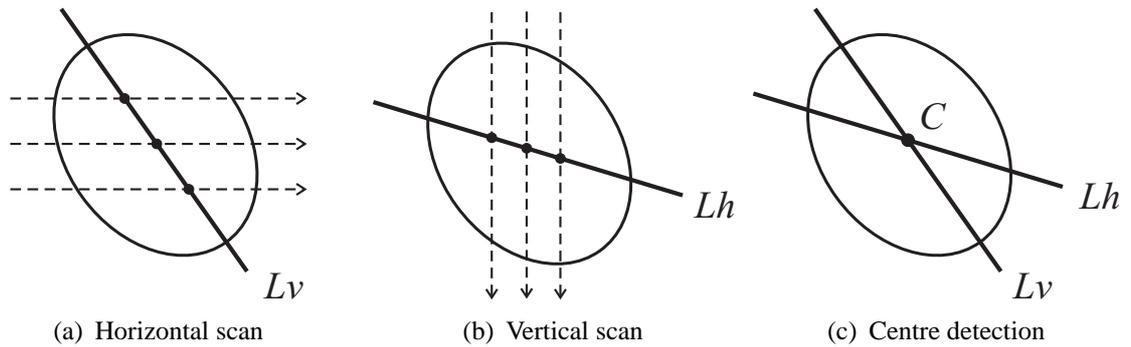
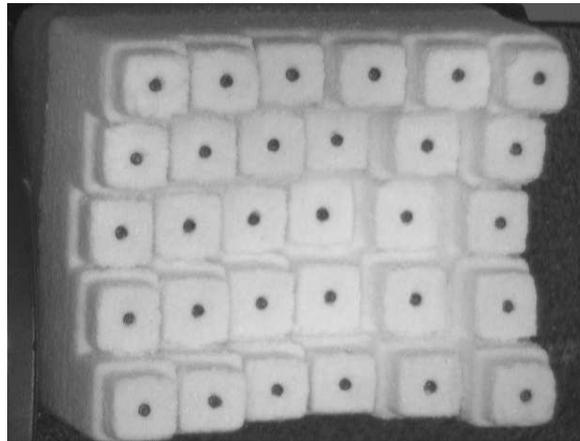


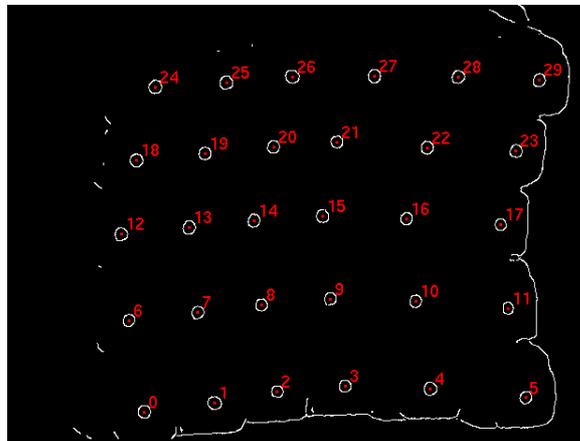
Figure 4.10: Method to detect the centre of elliptical figures [135].

centres. The calibration object as observed by the microscope before and after detecting marker centres is shown in Figure 4.11. It must be noted that, in some instances, projective distortions can affect the detection of circular markers if the disks are not parallel to the image, causing that the physical and projected circular/elliptical centres no longer coincide. Nevertheless, it has been proved that the coordinates of elliptical centres can be corrected to correspond to disk circle centres even if the disks are not orthogonal to the image plane [137].

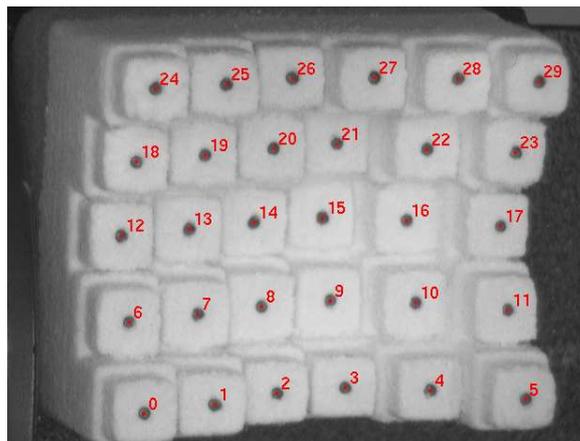
The camera pre-calibration procedure is based on the non-coplanar algorithm presented by Tsai [102]. This photogrammetric technique determines the camera parameters in two steps, which are summarised as follows: The first step estimates the orientation and part of the position (T_x and T_y) of the camera with respect to the calibration object, followed by the calculation of the scale factor. These parameters are obtained through the use of a least-squares fitting technique. The second step computes the focal length value, along with the distortion coefficients and T_z . Finally, all camera parameters are refined iteratively using a non-linear optimisation method that finds the best solution between the projected image points and the spatial points located on the calibration object.



(a)



(b)



(c)

Figure 4.11: Detection of marker centres on the non-coplanar calibration object: (a) image of the calibration device as seen by the microscope; (b) detection of elliptical centres, numbered for easier recognition; (c) overlay of detected centres on the original image of the calibration object, for illustrative purposes only.

4.3 Experimental Results

4.3.1 Analysis of single camera calibration

The purpose of this section is to evaluate how the accuracy of 2D image marker detection affects the camera calibration process. The experiments were carried out using a single camera in order to evaluate the stability of Tsai's photogrammetric algorithm, which produces the initial camera parameters to be optimised during self-calibration. These tests also serve to analyse the influence of the rotational position of the calibration object for the estimation of focal length, and consequently, the corresponding camera calibration errors.

Accuracy with respect to image marker detection

For this experiment a black and white camera was connected to one of the eyepieces of the surgical microscope, positioned perpendicularly to a flat bench. The non-coplanar calibration object was placed on a rotating gauge that allows measuring the positional orientation at different angles with respect to the bench surface (which in turn is parallel to the camera image plane). This rotational instrument was attached to a height gauge to control the translational distance T_z between the calibration object and the microscope lens (See Figure 4.12). At this stage there is no ground truth data to compare the results with; however, the intention of the experiment is to evaluate the variation of focal length and calibration errors based on 2D marker detection at different inclination angles.

A set of ten camera calibrations were performed for each slope angle, which varied from 0° to 25° at 5-degree steps. Inclination angles larger than 25° were excluded because at those orientations calibration markers went out of focus, affecting the localisation of circular shapes. Table 4.1 shows the mean focal length and calibration errors based on the calculation of ellipse centres. Two different calibration errors are calculated for dis-

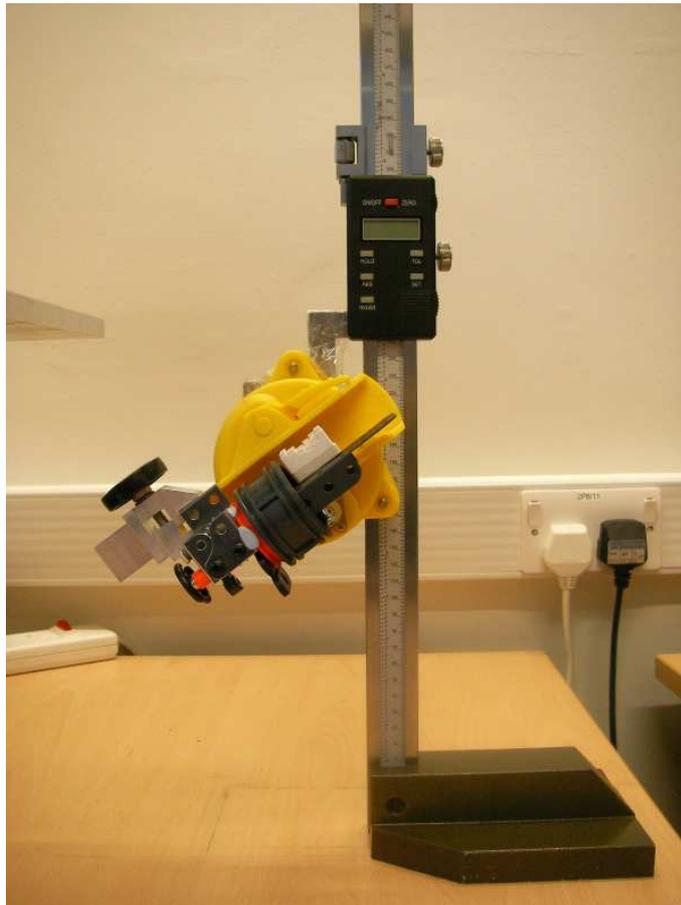


Figure 4.12: Rotational gauge used to control the calibration object inclination.

torted and undistorted image coordinates, measured in pixels. A third object calibration error is based on the difference between nominal 3D marker position and its projected estimation, measured in millimetres (mm). It can be noticed that focal length increases as the orientation angle between the calibration object and work bench rises as shown in Figure 4.13. These results are in agreement with the findings presented by Chen [17]. On the other hand, as the inclination becomes higher, calibration errors in pixels and millimetres have a tendency to decrease. This proves that there is an intrinsic relation between the estimation of focal length and the resulting calibration errors obtained by the photogrammetric method. Although the manufactured calibration object comprises a set of markers with different heights, the magnification provided by the microscope requires higher depth levels among the physical features. Due to current limitations in the fabri-

Angle (degrees)	Focal length (mm)	Distorted Error (pixels)	Undistorted Error (pixels)	Object Error (mm)
0	202.233 (± 4.255)	1.372 (± 0.129)	1.381 (± 0.128)	0.061 (± 0.006)
5	228.248 (± 6.161)	1.358 (± 0.261)	1.363 (± 0.262)	0.060 (± 0.012)
10	265.933 (± 10.717)	1.109 (± 0.068)	1.111 (± 0.068)	0.049 (± 0.003)
15	283.946 (± 11.055)	1.087 (± 0.057)	1.089 (± 0.058)	0.048 (± 0.003)
20	303.636 (± 6.448)	1.074 (± 0.081)	1.073 (± 0.081)	0.047 (± 0.004)
25	317.421 (± 5.070)	1.074 (± 0.028)	1.074 (± 0.028)	0.047 (± 0.001)

Table 4.1: Focal length and calibration errors mean and standard deviation (\pm SD) of single camera calibration at different inclination angles using ellipse centres over ten trials.

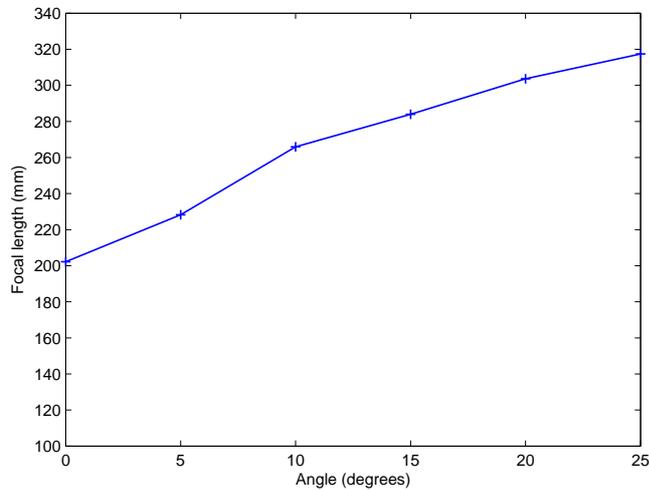


Figure 4.13: Focal length estimation through detection of ellipse centres.

cation of the 3D device, large depth differences can not be achieved. However, this issue can be compensated by modifying the angular position of the calibration object.

The relation between focal length and translational distance T_z was also assessed for the same inclination angles as described previously. The results in Table 4.2 show that distance values T_z raise as the rotation between camera and calibration object increases, similar to the focal length case. This inherent correlation between both parameters can be noticed from the ratio f/T_z , proving that there exists some compensation between focal length and the translation along the optical axis. However, this ratio still grows with respect to the orientation angle.

Angle (degs)	Focal length f (mm)	T_z	f/T_z
0	202.233 (± 4.255)	178.541 (± 3.473)	1.133 (± 0.003)
5	228.248 (± 6.161)	199.169 (± 5.191)	1.146 (± 0.001)
10	265.933 (± 10.717)	229.256 (± 8.958)	1.160 (± 0.001)
15	283.946 (± 11.055)	242.690 (± 9.367)	1.170 (± 0.002)
20	303.636 (± 6.448)	258.168 (± 5.412)	1.176 (± 0.000)
25	317.421 (± 5.070)	268.983 (± 4.220)	1.180 (± 0.000)

Table 4.2: Mean and standard deviation (\pm SD) values (over ten trials) of focal length, T_z distance and ratio between focal length and T_z at different inclination angles.

In addition, the remaining five DOFs were evaluated with the purpose of determining possible variations in the computed parameters at each inclination angle. The results shown in Table 4.3 demonstrate that the values obtained are consistent among all slope angles for parameters T_x , T_y , R_x and R_z . In the case of the rotational parameter R_y , the estimated values correspond to each inclination angle applied to the calibration object.

Angle (degs)	T_x	T_y	R_x	R_y	R_z
0	-10.253 (± 0.027)	8.720 (± 0.010)	173.055 (± 0.106)	0.841 (± 0.197)	-0.567 (± 0.024)
5	-9.977 (± 0.058)	8.693 (± 0.020)	172.905 (± 0.179)	5.650 (± 0.505)	-0.705 (± 0.044)
10	-10.261 (± 0.005)	8.960 (± 0.006)	172.601 (± 0.037)	10.999 (± 0.051)	-0.470 (± 0.019)
15	-11.142 (± 0.005)	9.232 (± 0.007)	172.459 (± 0.065)	16.313 (± 0.066)	0.526 (± 0.011)
20	-10.726 (± 0.019)	9.303 (± 0.012)	172.308 (± 0.073)	21.133 (± 0.193)	0.510 (± 0.027)
25	-10.752 (± 0.007)	8.957 (± 0.008)	172.145 (± 0.043)	24.689 (± 0.040)	0.732 (± 0.013)

Table 4.3: Mean and standard deviation (\pm SD) values for five different DOFs in a single camera calibration at different inclination angles over ten trials.

As observed in the previous tables, a remarkable output is the variation in the standard deviation corresponding to each static position. From the results analysed, the value most affected by such instability is the focal length. A simple experiment was performed to evaluate the accuracy of the calibration method by acquiring a single image and performing several calibrations. The results were constant for all values in each test (i.e. standard

deviation equal to zero), which indicates the robustness of the camera calibration algorithm when using the same input image.

Although there is no clear explanation for the variation in the results among the individual calibrations at static poses, a reasonable cause might be due to an imperceptible noise caused by the optical device, which could not be reduced by the Gaussian filter applied in the shape detection process. Another reason might be a variation of lighting or thermal conditions during the calibration process. Moreover, subtle vibration of the bench surface or measuring instruments could affect the position of the calibration object during image acquisition. This would cause the captured image to be slightly blurred and affect the detection of circular/elliptical centres by several pixels. It must be mentioned that all measures have been observed to prevent any kind of external interference concerning ambience luminance and steadiness of the hardware setup.

Effect of projected marker localisation on accuracy

This experiment investigated the influence of marker localisation inaccuracies on camera calibration errors. The set of projected marker points were affected to Gaussian noise with 0 mean and three different σ standard deviation levels: 1.0, 0.5 and 0.3 pixels. For each noise level, ten independent calibrations were carried out at the same inclination angles than the above-mentioned tests. The results were averaged and compared to a ground truth obtained from an initial calibration at each orientation. Figure 4.14 shows absolute errors for focal length and calibration error values. As seen from the plots, there is no linear relation among the results at different noise levels. However, absolute errors in image (pixels) and object space (mm) follow similar patterns [Figures 4.14(b) and 4.14(c)]. When $\sigma = 0.3$, there is less instability during calibration in comparison to higher noise levels, as expected. The effect of $\sigma = 0.3$ also decreases when the inclination angle is higher than 10° , especially in terms of the focal length. In the case of the ratio between

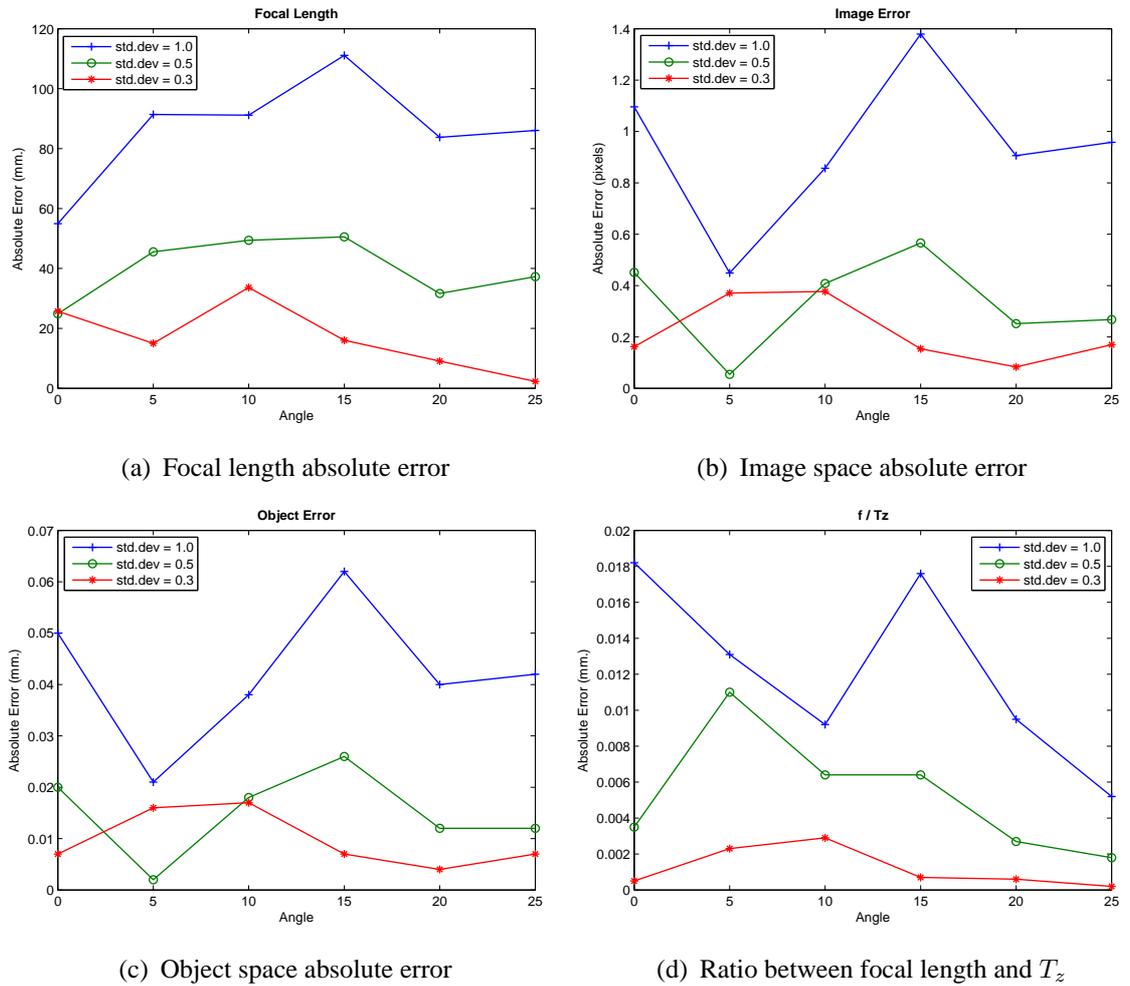


Figure 4.14: Absolute errors in localisation of image markers under different noise levels.

focal length and T_z [Figure 4.14(d)], a similar decrement is noticed when the inclination angle is higher than 10° for $\sigma = 0.5$ and $\sigma = 0.3$; whereas the absolute error decreases beyond 15° for all σ levels.

Effect of 3D marker localisation on accuracy

Additionally, the localisation of 3D marker points was tested under the effect of noise. This evaluates the performance of the algorithm in case of an inaccurate construction of the calibration object. Gaussian noise with 0 mean and three standard deviations of 0.3, 0.1 and 0.05 mm were added to the nominal coordinates of the constructed model. Similar

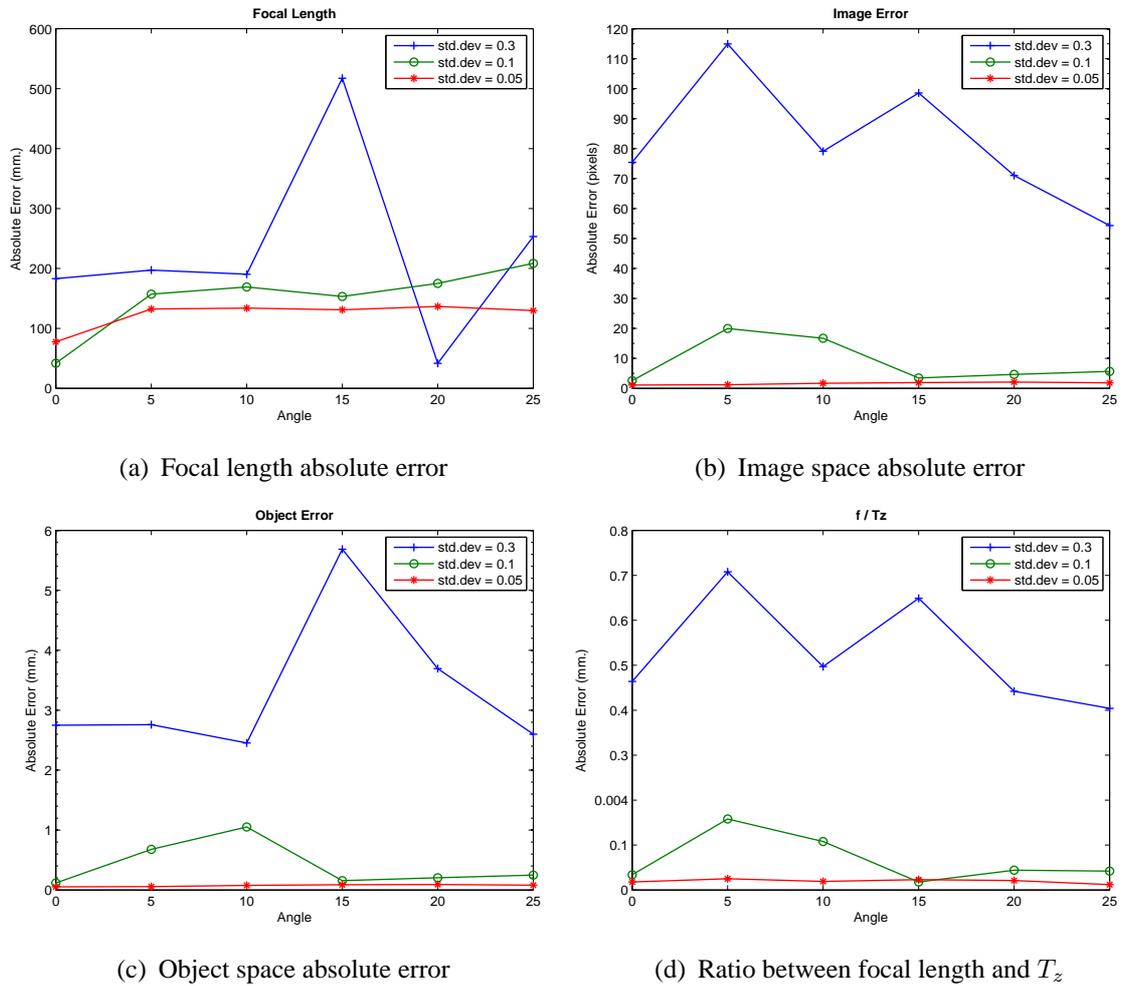


Figure 4.15: Absolute errors in localisation of 3D marker points under different inaccuracy levels.

to preceding experiments, the tests were performed at different positional angles and the mean values compared to a calibration ground truth. As seen from Figure 4.15(a) the focal length error is relatively constant when $\sigma = 0.05$ at all inclinations, whereas the fluctuation is greater at higher noise levels. Figures 4.15(b) and 4.15(c) respectively show absolute differences in image and object space errors. Figure 4.15(d) shows the ratio between focal length and T_z , which presents a comparable tendency to image space errors. In general it can be observed that, even at a low noise perturbation in 3D markers localisation, the inaccuracy is considerably high.

The outcomes of the previous two experiments conclude that the calibration object must be manufactured at high precision standards in order to obtain an accurate estimation of the focal length. Consequently, calibration errors in image and object space are also affected by a poor 3D construction even at submillimetre scale.

4.3.2 Analysis of stereo camera calibration

It was determined in the previous subsection that focal length is the most affected parameter in a single camera calibration. Consequently, in a stereo microscope setup, the estimation of focal parameters can greatly differ if left and right cameras are calibrated individually, even though both focal lengths are very similar in magnitude. This section aims at evaluating different methods to regularise the focal length calculation and minimise the final calibration errors in a stereoscopic setup.

Three different techniques for the estimation of the fundamental matrix between the pair of cameras were evaluated; namely linear, gradient-based and M-estimator methods as described in Section 4.2.3. Additionally, three methods for computation of focal length as proposed by Bougnoux [122], Sturm [126, 127] and Newsam [125] were also taken into consideration for the analysis.

Accuracy of focal length estimation in a synthetic environment

The purpose of this experiment was to evaluate the precision of focal length computation against a known ground truth value. A pair of simulated cameras was modelled in OpenGL with internal parameters similar to the values reported by Tsai's method [102] using a single microscope camera. The same external parameters (translation and rotation) in both cameras were used, while the baseline distance was set to 25 mm in order to match the physical horizontal separation between microscope lenses.

A set of 225 corresponding points was placed in the virtual space, lying at different

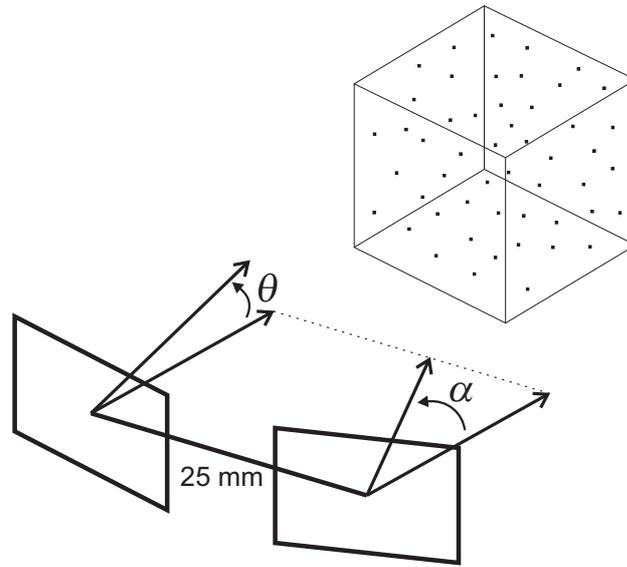


Figure 4.16: Schematic representation of a synthetic scenario using a set of points at different depth distances from a pair of virtual cameras (the points are situated within a bounding box for illustration purposes). Convergence angle (α) around the optical axis and elevation angle (θ) off the base plane are shown. Baseline distance between cameras is 25 mm.

depth distances from the cameras and within an area that could be visible on the pair of viewpoints (see Figure 4.16). Focal lengths with a value of 250 mm were initially used as ground truth for left and right cameras. Subsequently, the right focal length was increased in 5-millimetre steps until reaching 270 mm in order to evaluate the accuracy of the methods at unequal setups. In addition, convergence (α) and elevation (θ) angles between the two cameras were modified for each focal length combination to assess degenerate configurations in a surgical microscopic environment.

It was noticed that, among the three different approaches to calculate the fundamental matrix, the variation in the produced results was negligible at this particular synthetic scenario (the average relative error among the corresponding entries of the fundamental matrices was of the order of 10^{-10}). Because it was considered that any of the three methods would provide the same accuracy in the calculated fundamental matrix, the linear method was selected to estimate the relationship between the two cameras. The selection was based on the fact that the iterative nature of the other methods would not provide a considerable degree of improvement to the final results.

The accuracy of the techniques introduced by Bougnoux, Sturm and Newsam was evaluated based on a percentage error calculated between the ground truth focal length and the value produced by each method at different elevation and convergence angles; where a lower percentage error indicates a higher accuracy in focal length estimation. The graphs corresponding to the results obtained by each independent method for both cameras in equal and unequal focal length setups are presented in Appendix A.1. It is worth mentioning that at elevation angles between 0° and 2° all methods failed to produce a result, which indicates a critical configuration. Therefore, elevation at those specified angles are not shown in the graphs.

Bougnoux's method (Figure A.1) performs consistently when the cameras are only rotated around the elevation angle, whereas the error increases when the convergence rotation increments, especially at unequal focal lengths. It can be seen that the best results are obtained when the convergence rotation between cameras is fixed at 0° . Figure A.2 represents percentage errors obtained by Sturm's technique, which are considerably high for unequal focal lengths. This is because the method has been designed for similar focal values. For such equal configuration, Sturm's method achieves a small error during the different convergence rotations as long as the elevation angle remains low. Finally, the method introduced by Newsam, illustrated in Figure A.3, presents a low error for a 0° elevation in the case of similar focal lengths. Interestingly, percentage errors slightly increase as cameras converge around 4° . Beyond this angle, focal length accuracy is recovered. In the case of unequal focal lengths, the lowest errors are found almost diagonally between elevation and convergence angles.

The results of this experiment demonstrate the performance of each method during the recovery of the true focal length in a controlled synthetic setup. For all techniques, the best solutions are found when both cameras have the same focal length value. Specifically, the methods by Bougnoux and Newsam estimate accurate values during elevation, while Sturm's technique does at convergence angles. In the case of unequal focal configurations

the techniques behave significantly different, although Bougnoux's method maintains a low error rate at elevation angles.

Focal length optimisation in a pair of real cameras

This experiment investigated the use of different methods for the calculation of fundamental matrix and focal length estimation in a real-camera stereo configuration. The setup involved a stereoscopic surgical microscope placed on a flat work bench and two cameras connected to the microscope eyepieces, which are placed in a parallel position. It is worth mentioning that this setup leads the self-calibration to fail because the relative alignment between cameras belongs to a degenerate configuration.

In order to avoid this critical configuration, it was decided to acquire each individual image of the calibration object at two different positions using the rotational gauge used in tests 4.3.1 and 4.3.2, which was adapted to provide an extra rotation around an elevation axis (Figure 4.17 presents the microscope and rotational gauge setup). The tests involved acquiring the left camera image at a still pose while changing both convergence and elevation on the right view. Ten independent focal length estimations were executed for each convergence and elevation positions, within a range from 0° to 30° in the case of convergence and 0° to 15° for elevation, both at 5-degree steps. Higher inclination levels affected the detection of elliptical centres and were excluded from the evaluation.

For each pair of calibrated images, the obtained focal length values were refined through three different optimisation methods, including two evolutionary algorithms: Self-adaptive Differential Evolution (SDE) and CODEQ, which are described in Appendix B, and the Levenberg-Marquardt algorithm. Thus, a complete evaluation comprised nine combinations among the fundamental matrix and focal length techniques for each optimisation algorithm (see Table 4.4). Because self-calibration methods do not provide information about accuracy errors by themselves, the refined focal lengths were fed back



Figure 4.17: Stereoscopic surgical microscope and rotational gauge experimental setup.

into the original Tsai's method and a recomputation of camera calibration on both cameras was carried out (maintaining the other pre-calibrated parameters). This provided a means to compare the accuracy with respect to the initial calibration errors. The parameters used for the optimisation involved a maximum number of 400 iterations for Levenberg-Marquardt algorithm, whereas a maximum of 20 generations with a population size of 20 individuals was selected for both evolutionary algorithms. The tolerance threshold in the cost function (Equation 4.7) comprised a value of 10^{-16} for all optimisation methods based on an initial function value of 1.8^{-7} . In general, the time taken for the three optimisation algorithms to refine the focal length parameter was a couple of seconds.

Among the different elevations, the most favourable case was obtained at an elevation angle of 15° , where the absolute difference between Tsai's ground truth and mean

	Method
Fundamental matrix estimation	<ul style="list-style-type: none"> • Linear method • M-estimators • Gradient-based algorithm
Focal length estimation	<ul style="list-style-type: none"> • Bounoux's method • Sturm's method • Newsam's method
Optimisation algorithm	<ul style="list-style-type: none"> • Levenberg-Marquard • Self-adaptive Differential Evolution (SDE) • CODEQ

Table 4.4: List of methods used for the optimisation of the focal length.

image calibration errors is more noticeable. Therefore, the analysis of these experiments is focused at this orientation for the diverse fundamental matrix, stereo focal length and optimisation methods. A complete list of graph results for the entire set of tested combinations is presented in Appendix A.2.

Figure 4.18 illustrates the image calibration errors measured in pixels using the linear algorithm for fundamental matrix estimation. The top row represents the Levenberg-Marquardt (LM) method in left and right cameras. The second row corresponds to the SDE optimisation technique, and CODEQ results are shown in the bottom row. It can be noticed that LM produces unstable results in comparison with the other two methods, fluctuating above and below Tsai's ground truth at different inclination angles. The results obtained by both SDE and CODEQ methods are similar. Although both left cameras also show some fluctuation in the accuracy, it is not as pronounced as in LM. In the right camera, image errors decrease almost in parallel with respect to Tsai's; where the average accuracy improvement is 0.130 pixels for SDE and 0.136 pixels for CODEQ among all focal length methods.

Absolute image errors using M-estimators are presented in Figure 4.19. Results shown in the top row, corresponding to the LM algorithm, indicate that there is limited improvement in accuracy for both images as most focal length estimation methods give error values higher than the ground truth. SDE and CODEQ calibration errors have the same

trend in accuracy; however CODEQ produces closer results to Tsai's values, especially in the left camera. Remarkably, there is a break in the curves representing the methods of Sturm and Newsam at 25° of inclination (convergence) angle, which means that both techniques failed to estimate a focal length. Although this is not caused by a critical configuration (e.g. parallel camera setup), it is assumed that this is produced by certain instabilities within these algorithms as can be seen from other graphs in Appendix A.2, where additional breaks appear at different elevation/convergence angles.

The results illustrated in Figure 4.20 show absolute errors using the gradient-based method for the fundamental matrix. The outcomes are analogous to the previous techniques; however there are no breaks in the plotted curves at this elevation. Calibration errors using Bougnoux, Sturm and Newsam's methods are closer to each other when using CODEQ optimisation than in the case of SDE or LM for the fundamental matrix calculation. An exception is noticed at 30° where they slightly spread out due to the calibration object being partially out of focus, hence affecting ellipse centre detection. The overall accuracy improvement for all focal length methods is less than 0.05 pixels in the left camera using both SDE and CODEQ, whereas in the right camera the mean improvement is 0.131 pixels using SDE and 0.138 pixels for CODEQ.

The results obtained have shown that there is an improvement in the accuracy of the calibration error as compared to the accuracy obtained by Tsai's method, especially for the right camera. In order to investigate the statistical significance among the estimation methods (i.e. fundamental matrix and focal length) in both cameras, it was decided to test whether there are differences between the means of calibration error values for each pair of optimisation methods. For this purpose, a two-tailed t-test was performed for each pair of result sets. The null hypothesis is stated as the means of calibration errors between each pair being similar and not providing any statistical significance.

Because the set of results include a range of different convergence angles, the statistical

test involved two inclination levels that correspond roughly to the positions that provide a balance between good and bad accuracy errors for the three focal length methods. Such selection avoids bias in the overall evaluation among the techniques. In the case of the linear fundamental matrix the chosen angles were 20° and 25° , for M-estimators the selected inclinations were at 5° and 15° , and for the gradient-based method at 15° and 25° .

Tables A.1 - A.6 in Appendix A.3 present the reader with the results of the evaluation. Each table compares the statistical significance for left and right cameras, providing the computed t-value, statistical degrees of freedom (df) and significance level (P-value). If the P-value is found to be less than 0.05 it can be determined that there is a statistically significant difference at the corresponding observation. As it can be seen from the tables, there is little difference between SDE and CODEQ methods among the three fundamental matrix estimation techniques. On the other side, LM statistically differs from SDE and CODEQ in the majority of the comparisons, except at 20° in the linear method and 15° using the gradient-based method. Although this implies that the final results at such inclinations do not produce a meaningful difference among the methods, it only confirms that at specific rotations the accuracy errors can be similar for all the combinations.

The outcomes of this experiment indicate that there is an improvement in the accuracy of focal length estimation using self-calibration techniques in a pair of real cameras as compared to the accuracy obtained by Tsai's method, particularly for SDE and CODEQ methods. Among the three algorithms used to compute the fundamental matrix, the gradient-based method provided slightly better results than the linear counterpart and proved to be more stable than the M-estimators method. The most noticeable improvement is on the right camera, where a decrease of about 10% in the calibration error is obtained against the ground truth (Tsai's method) for all focal length techniques. In this regard, the results obtained by the different methods of Bougnoux, Sturm and Newsam perform similarly when using the CODEQ and gradient-based strategies.

Table 4.5 shows a summary of the results of the statistical evaluation presented in Appendix A.3. From this statistical evaluation, the null hypothesis stating that the experimental sets are the same can be accepted only in the case of the SDE - CODEQ combination for M-estimators and gradient-based techniques; whereas in the case of the linear technique, SDE and CODEQ are found to be different for both cameras in two different trials. On the other hand, the combinations between SDE - LM and CODEQ - LM tend to be statistically different for the linear and gradient-based methods at the highest convergence angle of the statistical evaluation, corresponding to 25° . In the case of M-estimators, the best statistical difference was obtained at 5° . In general, it can be observed that SDE and CODEQ are statistically similar (central part of the table) and that SDE and CODEQ statistically differ from LM (top and bottom sections of the table), largely obtaining better accuracy of the final calibration error than LM.

	Camera	Method	Linear technique		M-estimators		Gradient-based	
			20° (bad)	25° (good)	5° (good)	15° (bad)	15° (bad)	25° (good)
SDE vs LM	Left	Bougnoux	N	Y	Y	Y	Y	Y
		Sturm	Y	Y	Y	Y	Y	Y
		Newsam	Y	Y	Y	N	Y	Y
	Right	Bougnoux	N	Y	Y	Y	N	Y
		Sturm	N	Y	Y	Y	N	Y
		Newsam	N	Y	Y	Y	N	Y
SDE vs CODEQ	Left	Bougnoux	N	N	N	N	N	N
		Sturm	N	Y	N	N	N	N
		Newsam	Y	N	N	N	N	N
	Right	Bougnoux	N	Y	N	N	N	N
		Sturm	N	Y	N	N	N	N
		Newsam	Y	N	N	N	N	N
CODEQ vs LM	Left	Bougnoux	N	Y	Y	Y	Y	Y
		Sturm	Y	Y	Y	Y	Y	Y
		Newsam	Y	Y	Y	N	Y	Y
	Right	Bougnoux	N	Y	Y	Y	N	Y
		Sturm	N	Y	Y	Y	N	Y
		Newsam	N	Y	Y	Y	N	Y

Table 4.5: Summary of statistically significant results of the hypothesis that two optimisation methods display different calibration error values using three techniques for fundamental matrix (Linear, M-estimators and Gradient-based) and three methods for the focal length estimation (Bougnoux, Sturm and Newsam). Two different convergence angles are shown that roughly correspond to good accuracy (good) and bad accuracy (bad) in the calibration error values. Full statistical results are listed in Appendix A.3.

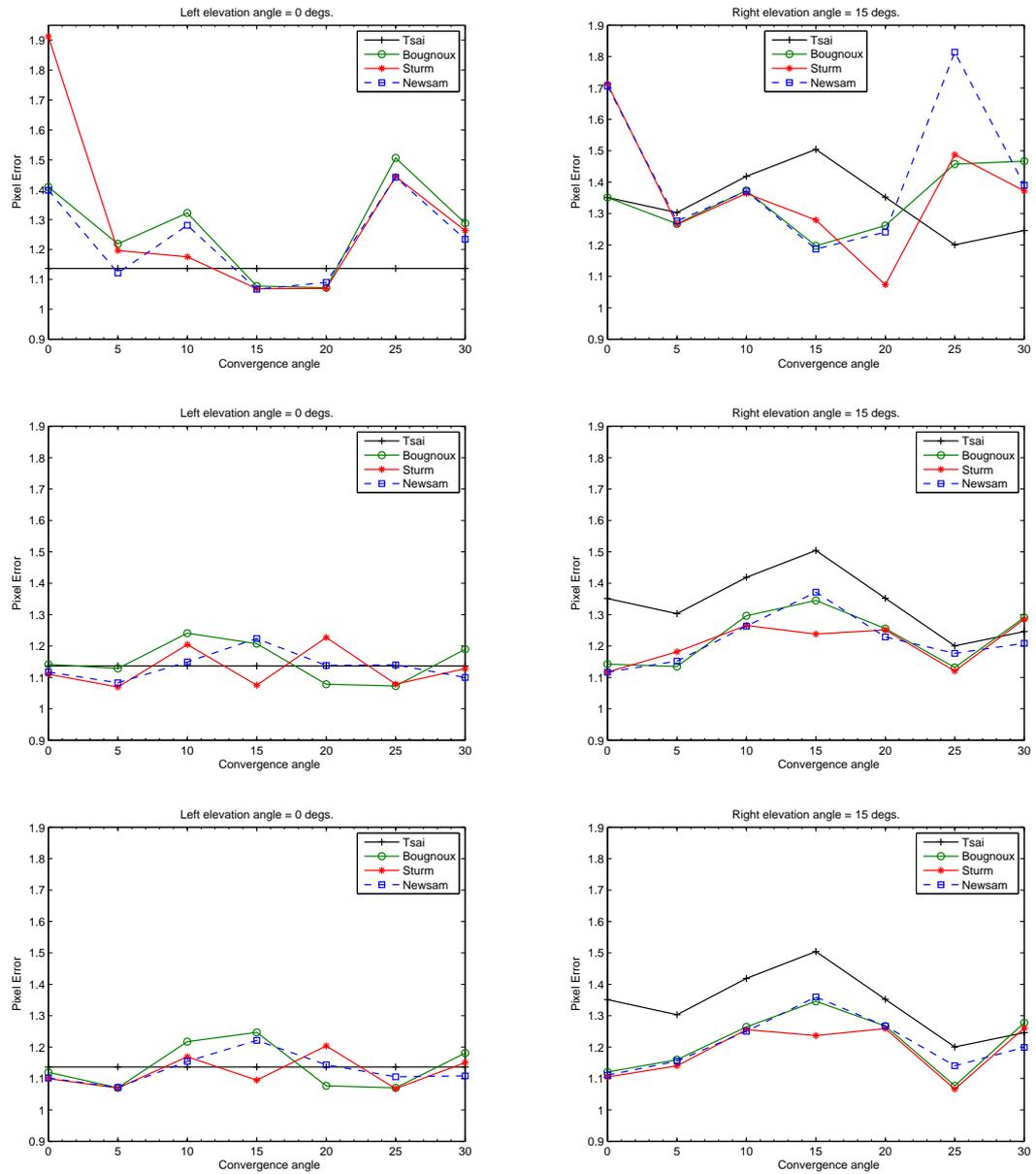


Figure 4.18: Absolute errors using the linear technique for fundamental matrix calculation for each of the three focal length estimation methods (Bougnoux, Sturm and Newsam) compared to the ground truth (Tsai's method [102]). (Top row) Levenberg-Marquardt, (middle row) SDE, (bottom row) CODEQ optimisation algorithms. Left and right columns represent image errors (in pixels) after optimisation on left and right cameras, respectively.

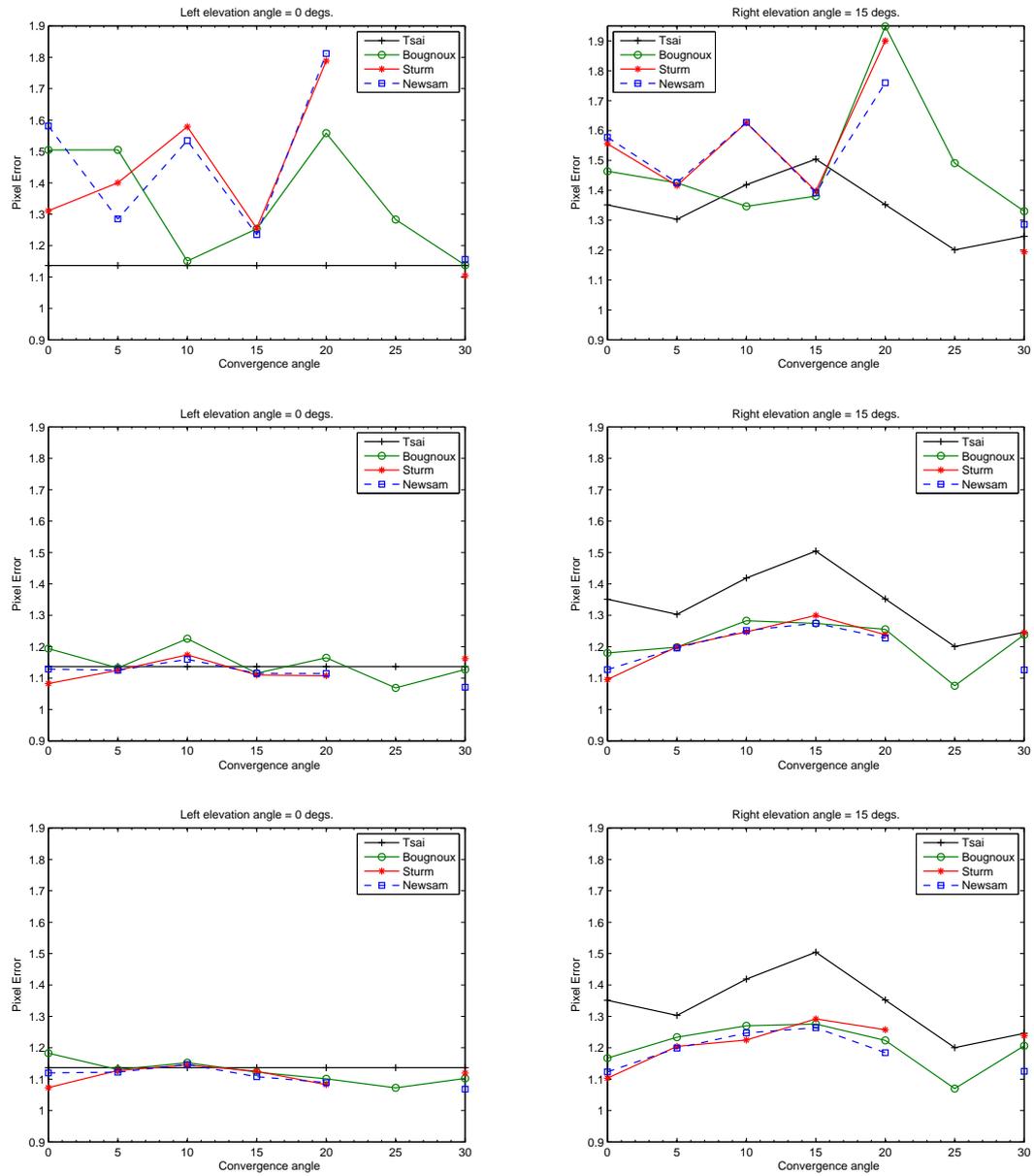


Figure 4.19: Absolute errors using M-estimators for fundamental matrix calculation for each of the three focal length estimation methods (Bougnoux, Sturm and Newsam) compared to the ground truth (Tsai's method [102]). (Top row) Levenberg-Marquardt, (middle row) SDE, (bottom row) CODEQ optimisation algorithms. Left and right columns represent image errors (in pixels) after optimisation on left and right cameras, respectively.

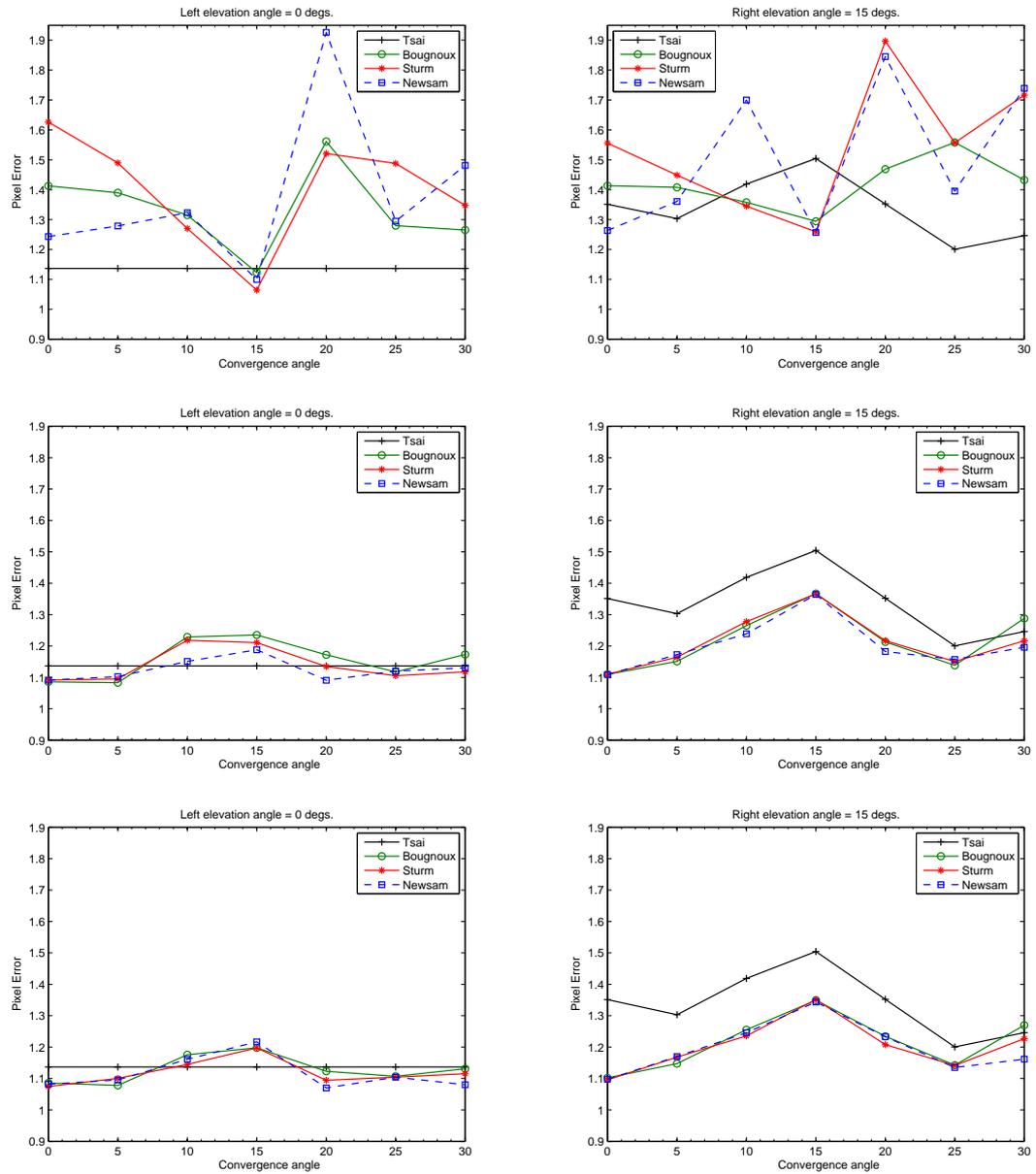


Figure 4.20: Absolute errors using the gradient-based technique for fundamental matrix calculation for each of the three focal length estimation methods (Bougnoux, Sturm and Newsam) compared to the ground truth (Tsai's method [102]). (Top row) Levenberg-Marquardt, (middle row) SDE, (bottom row) CODEQ optimisation algorithms. Left and right columns represent image errors (in pixels) after optimisation on left and right cameras, respectively.

4.4 Summary

This chapter focused on the study of camera calibration for the calculation of intrinsic and extrinsic parameters required to superimpose virtual imagery in an AR based surgical guidance scene. Firstly, an analysis of Tsai's photogrammetric calibration in a single camera was performed using a non-coplanar calibration object. It was found that the stability of the algorithm is significantly influenced by the precision of the object's physical construction and the detection of image markers. In addition, focal length is the most affected parameter with respect to external conditions, the orientation of the calibration device being one of them.

Secondly, an evaluation of different methods for the estimation of focal length in a stereoscopic setup was carried out. The purpose was to optimise the focal length and consequently improve the accuracy of the final calibration error. The results indicate that the use of an evolutionary algorithm can decrease the original calibration errors obtained by the photogrammetric method when one of the cameras is rotated about 15° around the elevation axis (tilt) in addition to the convergence rotation. This 15° around the elevation axis presents the most favourable angular position as it provides higher depth information of the calibration object's view in combination with the convergence angles. However, by increasing the elevation angle beyond 15° , the calibration accuracy is affected as not all calibration markers can be brought into focus. Among the different algorithms, the best combination comprised the use of the gradient-based method for the fundamental matrix and CODEQ optimisation, where the three techniques for the computation of focal length (i.e. Bougnoux, Newsam and Sturm) produced similar results. A further improvement to the overall calibration error could be made by increasing the manufacturing accuracy of the non-coplanar calibration object as to yield sub-pixel calibration errors.

Chapter 5

Intra-operative registration

5.1 Introduction

Registration is a technique with the objective to find a geometrical or spatial transformation that aligns two independent models in a scene. The process is achieved by establishing a correspondence of a set of features on each model and matching their position and orientation into a single frame coordinate system. Registration can be applied to a variety of fields such as computer vision, cartography, and medicine, among others. In AR based applications, registration aligns the position of a real model with its computer-generated counterpart in order to visually enhance the scene. In the case of IESN systems, the imagery involves the scanned anatomy of a patient, using CT or MRI, superimposed on top of images captured by an endoscope or surgical microscope.

Registration is typically performed in a pre-operative stage after camera calibration, resulting in a static overlay between real and virtual models. Additionally, a third step requires the use of an external tracking device to record the dynamic position of the models throughout surgery. Nevertheless, a problem found in AR is that each stage introduces a certain degree of error to the procedure. The combination of these errors produce vi-

sual misalignments between the models. This chapter will focus on a technique to correct such mismatch in the intra-operative stage of an AR based system, using a surgical stereo microscope for ENT surgery.

5.2 Image Registration

The process to find an optimum registration between models greatly differs among application areas, where different requirements and visual information are available. However, the basic approach involves a source model that must be mapped to a target entity using the best possible transformation. Zitova and Flusser [138] describe four main steps found in the majority of registration methods:

- *Feature detection.* It represents the selection of distinctive characteristics on the 2D or 3D models to be registered. The features can include points, lines, edges, corners, intensity patterns, etc. Although feature detection can be performed automatically on images with high level of detail, in the medical field the selection of features is usually carried out by an expert practitioner who manually selects anatomical structures.
- *Feature matching.* A correspondence between each pair of previously-selected features on both models is determined in this step. For this purpose, similarity measures and control rules must be specified to match unique pairs while avoiding possible outliers. Feature matching methods include intensity values, geometrical location or neighbourhood topology.
- *Transform model estimation.* It defines a transformation function to map the corresponding set of features on the source entity to the target model's pose. The transformation function depends on the geometric deformation of the imagery, the method of image acquisition and the required alignment accuracy. Examples of

transform estimation include rigid transformation, non-rigid transformation and image rectification.

- *Image resampling and transformation.* The transformation function is applied to the source imagery and the alignment is achieved between models. The process involves an interpolation step that maps pixels on the source image to the target coordinates. Additionally, the final image can undergo a subsample or supersample procedure in order to match the target image size.

5.2.1 Classification of medical image registration methods

In the clinical domain, different classifications about image registration have appeared in the literature. In this subsection, a broad summary of the main classification proposed by Maintz and Viergever [139] is presented. This will serve as the basis to introduce image registration within the surgical field.

Dimensionality describes the spatial dimensions involved in the registration process. The models to be aligned can be 3D/3D datasets pre-operatively scanned such as CT or MRI, or involve a 2D/2D registration which comprises X-rays or single tomographic slices. In 2D/3D registration, the procedure is mostly intra-operative and includes 2D images to volumetric data. An extra dimension can also be included in the registration of models during a certain period of time (e.g. tumour growth). In this case a volumetric dataset is regarded as 4D (3D + time).

Nature of registration basis refers to the different feature-based methods upon which the imagery will be registered. They can be divided in *extrinsic* and *intrinsic* methods. The former include external or foreign objects attached to the patient's body (e.g. stereo-tactic frames, dental adapters, skin markers). Intrinsic methods are based only on visual information detected on images of a model using for example anatomical landmarks.

Nature of the transformation depicts the diverse geometrical transformations required to map two models in the scene. This criterion divides the transformation in: a) *rigid*, when only rotations and translations are possible; b) *affine*, if the registration mapping preserves parallel lines during the transformation; c) *projective*, when lines are projected onto lines but parallelism is not maintained; and d) *curved*, if lines are transformed onto curves. On the other hand, the **domain of the transformation** describes how the mapping will affect the registration; *global* if the process is applied to the entire image, or *local* if subsections of the image undergo separate transformations.

Interaction describes the interactivity level that a user provides during the registration procedure. The method can range from fully-automated to completely manual interaction, based on the nature of the registration and application requirements.

Optimisation refers to the algorithms used to find the global optima that correspond to the best alignment pose between models. Methods based on *computation* are oriented to applications where sparse data are found (e.g. point sets), whereas *search* techniques try to find a solution based on a mathematical formulation of the problem.

Modalities involved relates to the origin of the imagery to be registered. When both models are acquired using the same imaging source they are referred to as *monomodal* (e.g. CT/CT, MRI/MRI). If different devices are used to obtain the models, the registration is denominated *multimodal*, such as CT/MRI or X-ray/CT. The criterion denominated **object** simply depicts the anatomical structure to be registered, which can include head, limbs, thorax, spine, etc.

The **subject** classification includes: a) *intra-subject* registration, in cases where the models to be registered belong to the same patient; b) *inter-subject* registration, if the procedure aims at registering models which belong to different subjects; and c) *atlas*, when one model belongs to a patient and the other model is generated from a database obtained from many subjects.

5.2.2 Image registration for SN in head and neck surgery

The first SN systems for brain surgery required the use of stereotactic frames rigidly attached to a patient's head during surgical interventions. These external frames provided the surgeon with a local coordinate system upon which the patient's anatomy was associated. The surgeon could refer to external reference points on the frame in order to locate specific regions inside the human brain. This association of local coordinate systems led to the first step towards a registration procedure between X-ray images of the patient wearing the frame and the internal anatomy itself. Subsequently, 2D stereotactic image registration adopted other technologies such as CT, MRI and angiography.

The development of a frameless procedure was introduced in the 1980's. The purpose was to minimise the effects of invasiveness that rigid frames represented to the patient and the obtrusiveness for the practitioner to manipulate surgical instruments inside the body parts. This approach involved the use of individual fiducial markers mounted on the patient's head. The registration relied on comparing the real markers' position, localised by a pointing device, against the 2D coordinate image system on CT or MRI slices.

With the evolution of volume rendering — where a set of slices can be reconstructed as a 3D model on a computer screen — registration has mostly become a 3D/3D process. The method requires finding a linear, or matrix, transformation between real and virtual models based on the attached fiducial markers, which involves the calculation of rotation and translation parameters. Other techniques include markerless registration, where anatomical landmarks (e.g. eye corners, tip of the nose, etc.) are detected on both models. However, the final matching accuracy obtained by a markerless method is generally lower than the precision achieved by fiducial markers. Another method involves surface matching that is used to align mesh models generated during the pre-operative and intra-operative stages. This method takes into account the entire surface structure rather than a set of fiducial markers or natural landmarks. A general review of registration techniques

for SN in head surgery is presented by Eggers et al. [140] and Luebbbers et al. [141].

5.2.3 Overlay accuracy in IESN systems

IESN systems suffer from overlay inaccuracies because of the inherent loss of resolution produced by digitising models in the real world into a digital format with the purpose of enhancing intra-operative visual information. More precisely, image registration experiences inaccuracy levels produced by CT/MRI acquisition, 3D volume or surface reconstruction, anatomical landmarks and fiducial markers localisation, etc. Therefore, the transformation matrix produced after registration can not be regarded as an error-free solution, but as an estimation of the true correspondence between source and target models.

Other errors within IESN systems originate from the determination of internal and external camera parameters, limitations of accuracy provided by motion tracking systems, and time delays while displaying the augmented world. The interested reader is referred to Holloway [142] for a mathematical treatise of accuracy errors in AR. Although each independent stage generates a certain level of imprecision on its own, the resulting overlay accuracy is further affected when these stages are combined as the errors accumulate. Moreover, the superimposition is aggravated by the use of the tracking device over a prolonged period of time throughout surgery, which eventually results in an unacceptable overlay accuracy as described by Lapeer et al. [13].

5.3 Methodology

5.3.1 Intra-operative registration for ENT procedures

In order to rectify the misalignment produced by the use of a tracking device, a naïve method would involve interrupting the surgical procedure and repeating the initial cam-

era calibration and registration. Nevertheless, such approach is inappropriate due to time restrictions and efficiency requirements during surgery. For this reason, it has been proposed that an additional registration step can be carried out during the intra-operative stage. The selected approach aims at re-aligning virtual and real models based only on the visual information acquired from a pair of cameras connected to a stereoscopic surgical microscope. It involves a cost function which compares intensity value differences between the captured images and evaluates the registration accuracy through a similarity metric called photo-consistency.

The concept of photo-consistency was first introduced in the field of computer vision as a method for the reconstruction of 3D models from a set of colour or greyscale images in which the real scene is considered as and subdivided into voxels [143, 144]. This volumetric scene reconstruction method known as shape-from-photoconsistency is based on the visual property of real surface models. If a point on the surface is seen from different perspectives the point's colour should be the same, provided there is no occlusion.

Clarkson et al. [145] employed photo-consistency as a novel similarity measure to match the projection of a set of 2D images to a 3D surface model of a human head, either acquired by a laser scanner or converted from an MRI data set. In their work, the alignment process was performed under controlled lighting and the cameras were previously calibrated. The selected optimisation function was based on intensity differences using a gradient ascent search algorithm. Later, Jankó and Chetverikov [146, 147] generalised the technique by finding the registration pose and performing a camera calibration procedure at the same time. A full-sized polygonal mesh was obtained from a 3D laser scanner and manually pre-registered to a pair of views as a starting point. The optimisation method involved a genetic algorithm to refine the registration.

In the field of medical applications, Figl et al. [148] introduced a photo-consistency cost function to align a sequence of video images of a beating heart model. The images

were obtained by a calibrated stereo endoscope connected to a static robotic arm. The real model was under the influence of a heart cycle motion which required multiple surface models reconstructed from CT volume data. Thus, the registration involved a 2D/4D process. Chen et al. [17, 149] implemented an intra-operative registration between calibrated endoscopic images and a volumetric model of a human skull. Several images were captured by placing the single-camera endoscope at different positions. Powell's method was selected as the optimisation procedure for its simplicity of use. Although the alignment was successful using synthetic imagery of the complete skull, the algorithm failed to converge in close-up areas due to the lack of variation in intensity.

5.3.2 Estimation of photo-consistency based cost function

The estimation of the photo-consistency metric relies on the comparison between colour or intensity values in a visible set of a 3D object's points that are projected on two or more images. Thus, the corresponding pixels which are related to the same point should ideally possess the same colour or intensity attributes on each image. The images are considered photo-consistent if the difference among all related pixel values is null or near zero. It is assumed that the scene complies with a Lambertian model in which the lighting is static and the visible object maintains an equal luminance regardless of the point of view.

For the implementation of intra-operative registration, a pair of black and white cameras were connected to the eyepieces of a surgical microscope for ENT interventions. Each camera needs to be calibrated in advance in order to estimate the internal parameters K and external parameters S , where K is a 3x3 upper triangular matrix and S is a 3x4 matrix which combines the orientation and position of the optical device. Therefore, a composed projection matrix $P = KS$ (See Chapter 4) is used to calculate the projection

of a model's 3D point \mathbf{M} on each camera viewport in the following form:

$$m_l \sim P_l \mathbf{M}, \text{ and } m_r \sim P_r \mathbf{M}, \quad (5.1)$$

where P_l and P_r are the 3x4 projection matrices for the left and right cameras, respectively; and m_l and m_r are the corresponding projected pixels of the same point \mathbf{M} . The sign \sim indicates that the projection is defined up to a scale factor.

The photo-consistency based cost function PC is determined by comparing the pixel intensity levels I in the pair of captured images:

$$PC = \frac{1}{N} \sum_{i=1}^N \|I(m_{l,i}) - I(m_{r,i})\|^2, \quad (5.2)$$

in which N represents the total number of visible projected pixels i in both images. Clarkson et al. [145] provide an alternative similarity measure for the computation of photo-consistency by first determining a mean of pixel values. In the case of two viewpoints it follows as: $I(\bar{m}_i) = (I(m_{l,i}) + I(m_{r,i}))/2$.

Consequently, the total sum of squared differences is calculated according to the following equation:

$$PC_{squared} = \frac{1}{N} \sum_{i=1}^N \frac{(I(m_{l,i}) - I(\bar{m}_i))^2 + (I(m_{r,i}) - I(\bar{m}_i))^2}{2}. \quad (5.3)$$

A final cost function, also described in [145], intends to reduce the effect of outliers through the calculation of the inverse of squared differences. This is achieved by using a threshold ϵ related to the noise level found in intensity images. The resulting equation is:

$$PC_{inverse} = \frac{1}{N} \sum_{i=1}^N \frac{\epsilon^2}{\epsilon^2 + ((I(m_{l,i}) - I(\bar{m}_i))^2 + (I(m_{r,i}) - I(\bar{m}_i))^2)}. \quad (5.4)$$

5.3.3 Photo-consistency based registration

An initial registration step must be carried out at the beginning of the medical procedure in order to align both real and virtual models. This initial registration is usually performed after camera calibration and provides a static overlay within the IESN system. The steps required for visual enhancement involve pre-operatively scanning the patient with a VBH mouthpiece [150] which is also worn during surgery. Attached to the mouthpiece there is a calibration and registration block that is used as a point of reference for the initial alignment. A virtual version of the block is segmented from the subject's CT data set using a watershed segmentation algorithm [151] and a corresponding isosurface model is obtained through the marching cubes method [152]. The resulting model is registered to a pre-defined 3D block with known world coordinates. As the pre-defined block and the segmented isosurface model have inherently the same shape and size, the iterative closest point (ICP) algorithm [153] is used to perform the initial 3D/3D surface registration. Figure 5.1 illustrates the steps involved during the pre-operative registration using a human skull.

Once the initial registration is obtained and the surgery is in progress, an optical tracking device is used to capture the movements of cameras and/or patient. As mentioned earlier, the accumulated errors generated over time tend to affect the original alignment due to tracking. For this purpose, the photo-consistency based cost function is used to correct the mismatch between real and virtual models by evaluating the best registration pose that corresponds to the lowest intensity difference between the captured images. Because both cameras have been calibrated in an earlier stage, the cost function only requires to determine six DOFs; i.e. three translational and three rotational.

In order to perform the intra-operative registration, a set of visible voxels in the virtual model is selected through back-projecting screen pixels (u, v) of that model within a user-defined selection window; similar to a raycasting projection. Then, a voxel is detected for

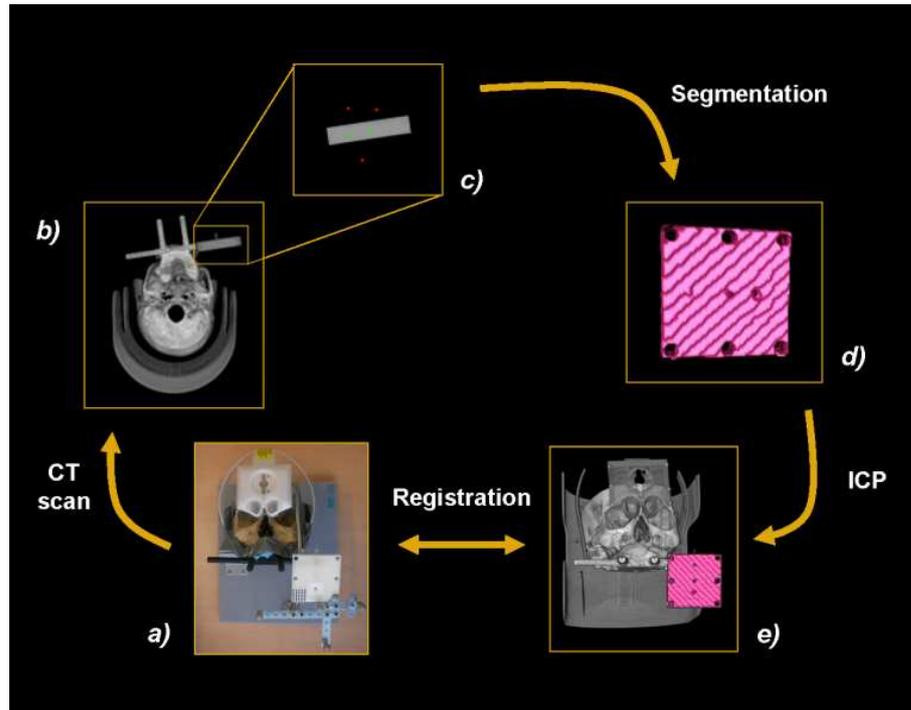


Figure 5.1: Stages within the pre-operative registration; a) real human skull object, b) volumetric dataset with its own coordinate system, c) selection of points to perform segmentation, d) resulting segmented isosurface block, e) final pre-operative registration after ICP.

each ray of sight that collides on the volumetric model (See Figure 5.2). It is important to note that several screen pixels will map to a single voxel due to the magnification provided by the microscope. For this reason, duplicated voxels are neglected in the selection. Subsequently, a forward projection ray is cast from the selected voxel to each camera image with the purpose of determining the corresponding pixel coordinates on both images. This forward projection requires a series of transformation matrices as follows:

1. Voxel 3D coordinates $[X_w, Y_w, Z_w]^T$ to initial 3D registration pose (T_{icp}).
2. Initial static 3D pose to dynamic model 3D position ($T_{tracking}$).
3. Dynamic 3D voxel position to 2D pixel image coordinates $[u, v]^T$ for each camera ($T_{cam,i}$).

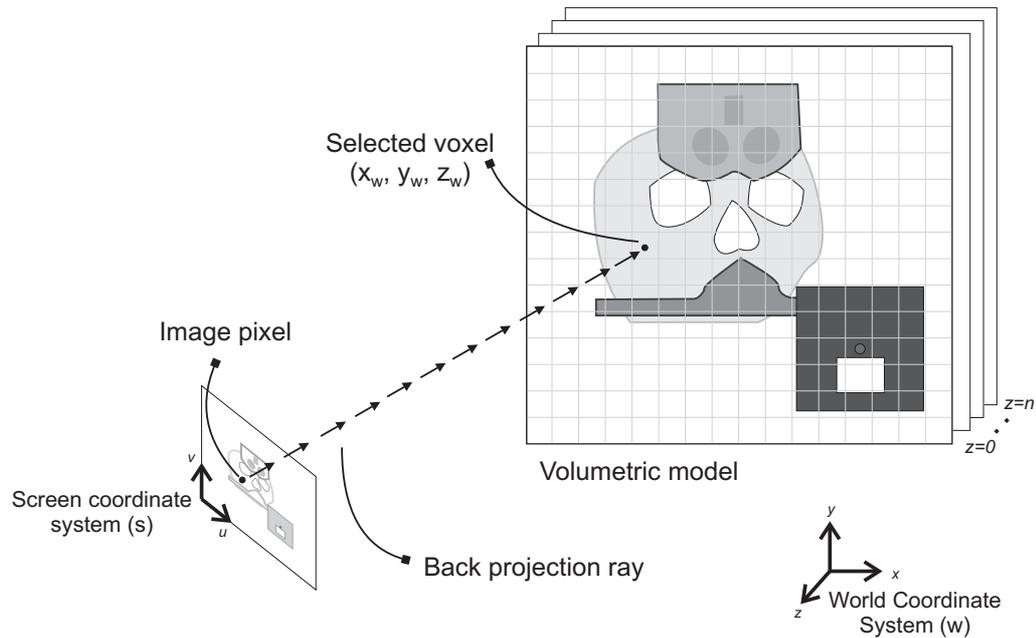


Figure 5.2: Selection of voxels by backprojecting screen pixels.

The final concatenation of these matrices results in:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = T_{cam,i} \cdot T_{tracking} \cdot T_{icp} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \quad (5.5)$$

where $i = 1, 2$ is the number of cameras.

During the process of projecting voxels to pixel coordinates it is essential to evaluate any potential occlusion that could obstruct the visibility of a voxel on both camera viewports. In the case of surface models, a technique can be used to compute surface normals in order to avoid comparing areas of the model that are not oriented towards the cameras, as described in [146, 147]. A different technique uses a z-buffer to allow rendering only the external visible points of the mesh model [145]. Nevertheless, none of the above methods can be directly applied to CT or MRI models as these are made of voxels with different transparency levels. Instead, it is more suitable to perform a direct check of possible voxels that can partially obstruct the forward-projected ray from a selected

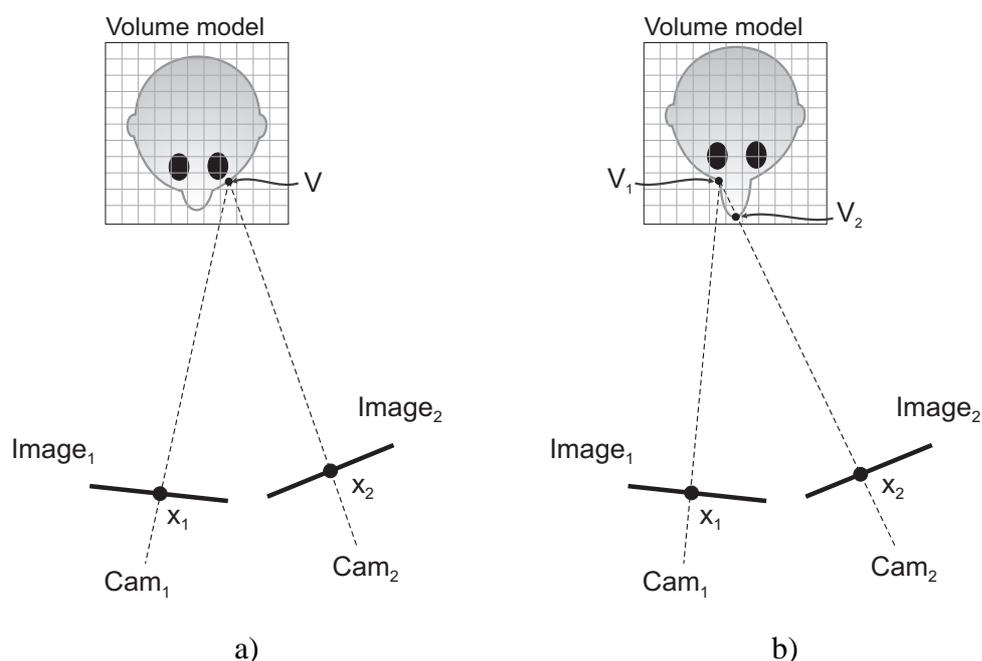
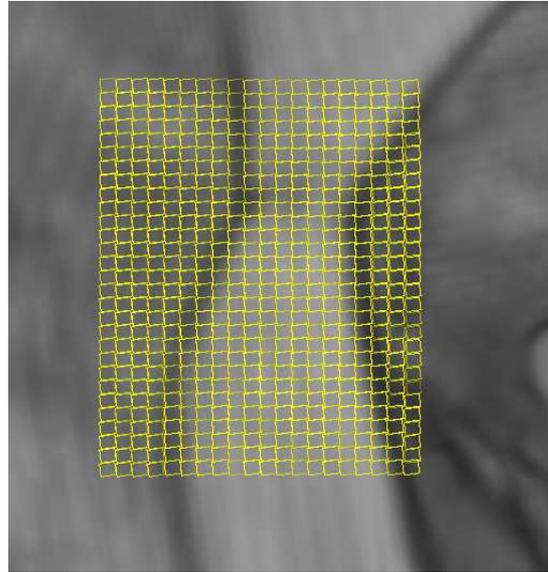


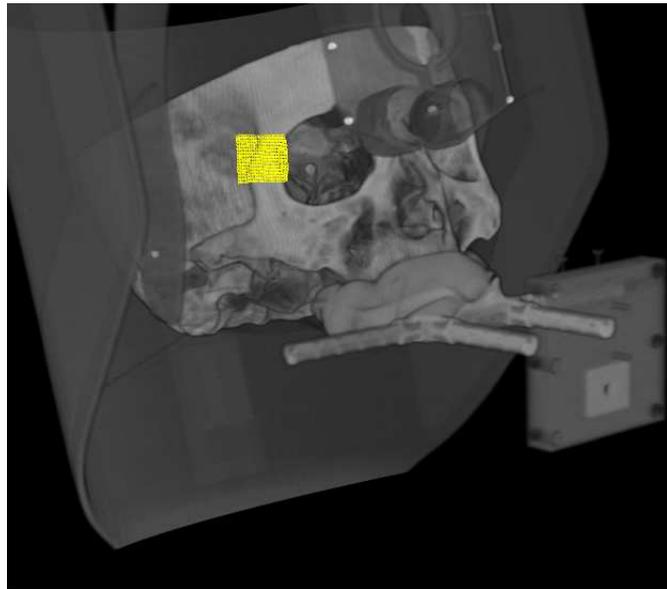
Figure 5.3: Occlusion detection through a forward projection ray; a) visible voxel V is projected on both camera images at x_1 and x_2 , b) voxel V_1 is visible from camera $Image_1$ but occluded by voxel V_2 on camera $Image_2$ (Image adapted from [145]).

3D point. If the projected ray collides with a voxel that has a higher transparency level than a certain threshold, it is considered as an occlusion and the corresponding pixels are ignored in the evaluation of the photo-consistency cost function. Figure 5.3 illustrates the occlusion detection using a forward projection ray.

Due to magnification, a projected voxel does not relate to a single pixel on the pair of captured images (voxel-to-pixel relation is 1:many) as described previously. In order to establish a unique correspondence, it is necessary to determine the voxel dimensions on the model and project the vertices of the voxel face that is oriented towards each of the cameras. From this set of four vertices it is possible to create a 2D sub-window that is associated to the visible voxel and determine the number of projected pixels, as shown in Figure 5.4. Finally, a median filter is applied to the pixels inside the convolution window to reduce any noise on the captured images and minimise the number of outliers in the cost function.



(a)



(b)

Figure 5.4: Selection of voxel-sized sub-windows around the skull orbit: (a) microscope perspective; (b) zoom-out view of selected area.

5.4 Experiments and results

Analysis of photo-consistency based registration

The objective of this section is to evaluate the use of photo-consistency as a registration metric based on the cost functions described in Equations 5.2 – 5.4. In addition, three optimisation algorithms (i.e. Powell’s method, SDE and CODEQ) were compared in order to assess the convergence of the registration through the minimisation of the different cost functions. The procedure comprised a set of laboratory tests using a human skull as a dummy patient which was CT scanned and subsequently reconstructed as a volumetric model within the IESN system.

5.4.1 Registration of models in a simulated environment

This experiment was aimed at evaluating the performance of the three photo-consistency similarity metrics in a synthetic configuration. For this purpose, the reconstructed CT skull model was positioned at the origin of the virtual scene. A pair of virtual cameras were created in OpenGL and oriented towards the CT skull at a distance where the complete model could be visible in both images. The simulated 2D camera images were obtained by rendering each viewpoint directly on a texture image, equivalent to capturing real images through a frame grabber device. A set of voxels was selected by back-projecting pixels towards the volumetric model, as mentioned in the previous section; where the selected area covered the entire visible model. At this stage it was certain that the 3D skull model was registered with the pair of virtual images hence its current pose was regarded as the ground truth.

In order to compare the registration accuracy among the cost functions, the procedure involved evaluating independently each DOF to avoid any possible influence of the other pose parameters. For translation, the CT model was placed at an arbitrary position along

each axis and gradually moved to the opposite location passing through the origin of the scene within a range of $[-60, +60]$ mm. In the case of rotation, the model was initially oriented at -45° with respect to the original straight pose and increasingly rotated at 5-degree steps until reaching $+45^\circ$ for each of the axes. Figures 5.5 - 5.7 show the results of the photo-consistency based cost functions for the six DOFs, where the ground truth position is represented in the graph as a vertical dashed line.

As it can be seen from the plots, the cost functions PC and $PC_{squared}$, corresponding respectively to Equations 5.2 and 5.3, generate the lowest photo-consistency error value (PC error) when all rotations pass through the ground truth position at 0° . Similarly, both cost functions produce the lowest error values for the translation along the X axis (T_x) at the ground truth. In the case of the Y axis (T_y), the lowest error is found on the graph at 1 mm to the left of the ground truth using PC and $PC_{squared}$. The results for T_z show that there is a misalignment of 3 mm¹ apart from the correct position using both similarity metrics. The reason for this more limited accuracy lies in the fact that at small steps, the different motions along T_z are visually negligible. However, the results are considered to be close to the ground truth. On the other hand, the cost function $PC_{inverse}$ (Equation 5.4) provides inaccurate results for all DOFs. The registration in R_x generates the lowest PC error at -40° , whereas the lowest errors for R_y and R_z are offset around 10° in comparison with the ground truth. The translational components are also incorrect as the lowest PC errors are found within 10 mm for T_x and T_z , and near 55 mm for T_y with respect to the true position.

Interestingly, it can be noticed that the lowest photo-consistency error, relating to the global minimum, for each of the cost functions does not necessarily correspond to a null or near zero value as expected. The lowest PC error value found in the PC cost function is approximately 350 and for $PC_{squared}$ is close to 10 units. The only exception is $PC_{inverse}$ in which the inverse of the squared differences is computed and the resulting error tends

¹The accuracy of the measurement is based on the model's voxel size

to zero. The reason for not obtaining null values among the cost functions is due to the nature of the volumetric model, which contains different transparency levels that affect the projected intensities on the pair of images.

The outcomes of this experiment demonstrate that the best results are obtained by PC and $PC_{squared}$ cost functions for the registration of a full-sized CT model in an artificial environment. Although the projection of CT models on the camera images are affected by inherent transparency levels, global minima are accurately found at the three rotational parameters; whereas there is a slight deviation in the translational components, T_z being the least precise variable. Conversely, $PC_{inverse}$ is unable to find the global minima at the correct pose for any of the six DOFs.

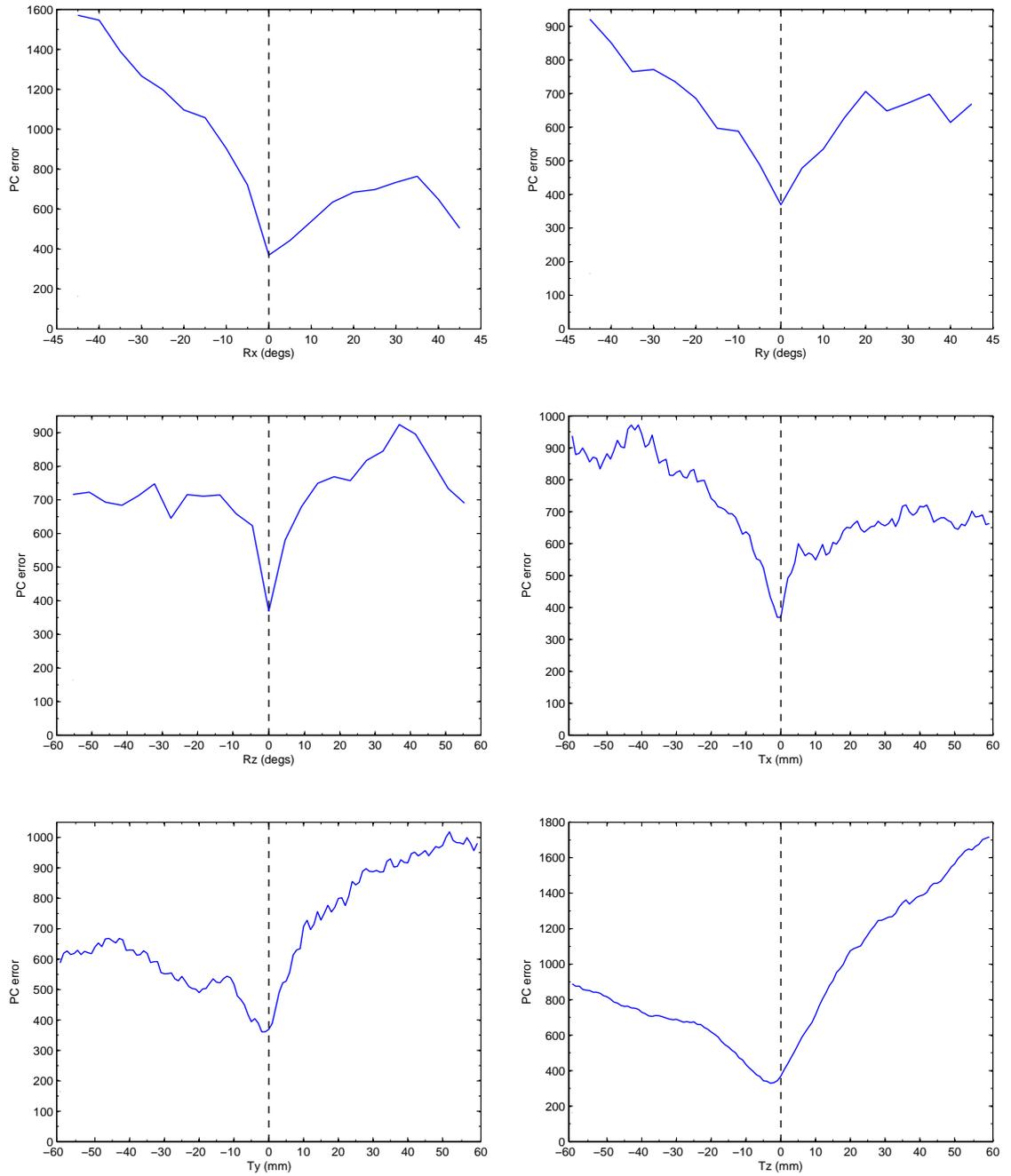


Figure 5.5: Registration errors for six independent DOFs based on the PC cost function using a full-sized CT model.

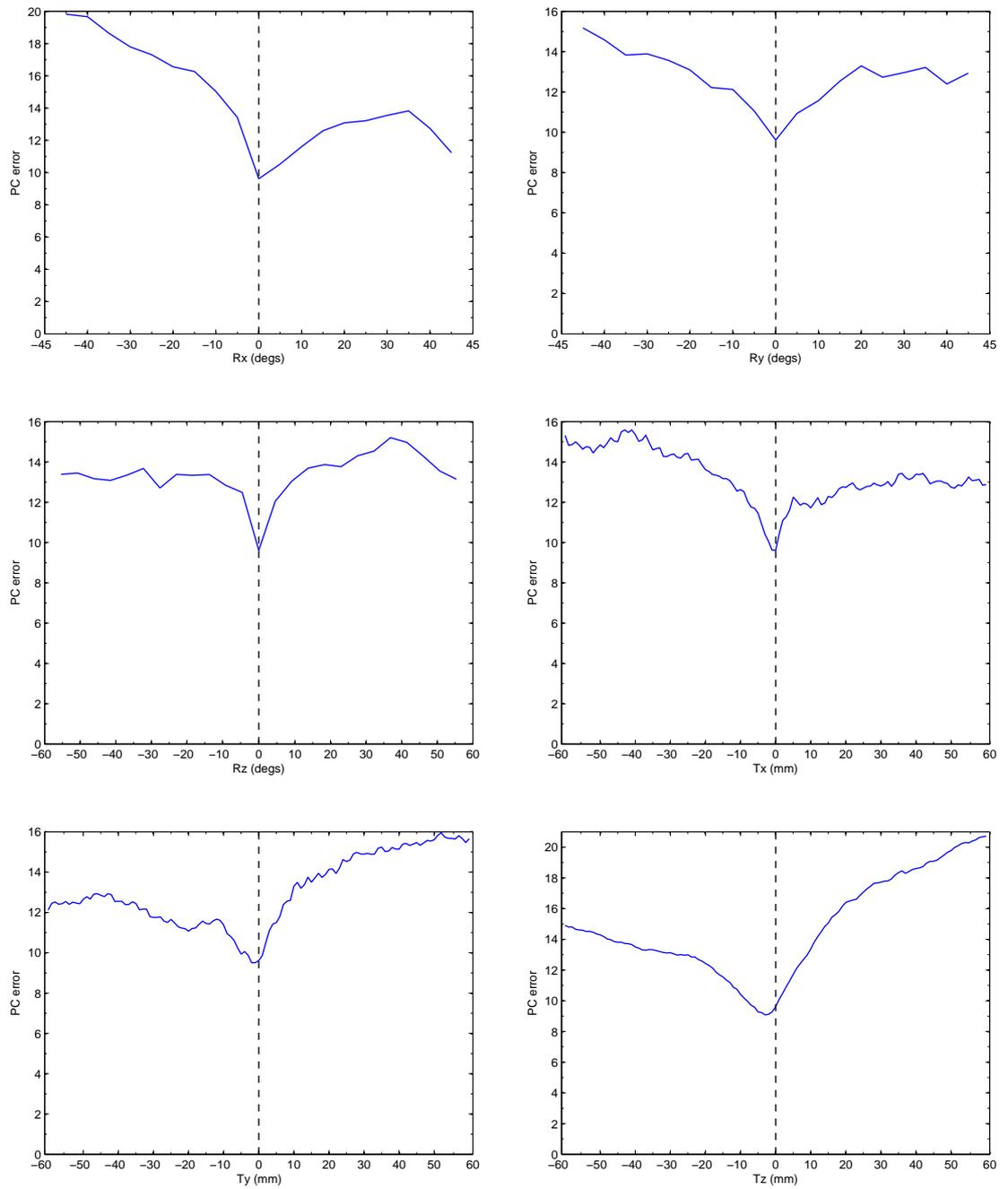


Figure 5.6: Registration errors for six independent DOFs based on the $PC_{squared}$ cost function using a full-sized CT model.

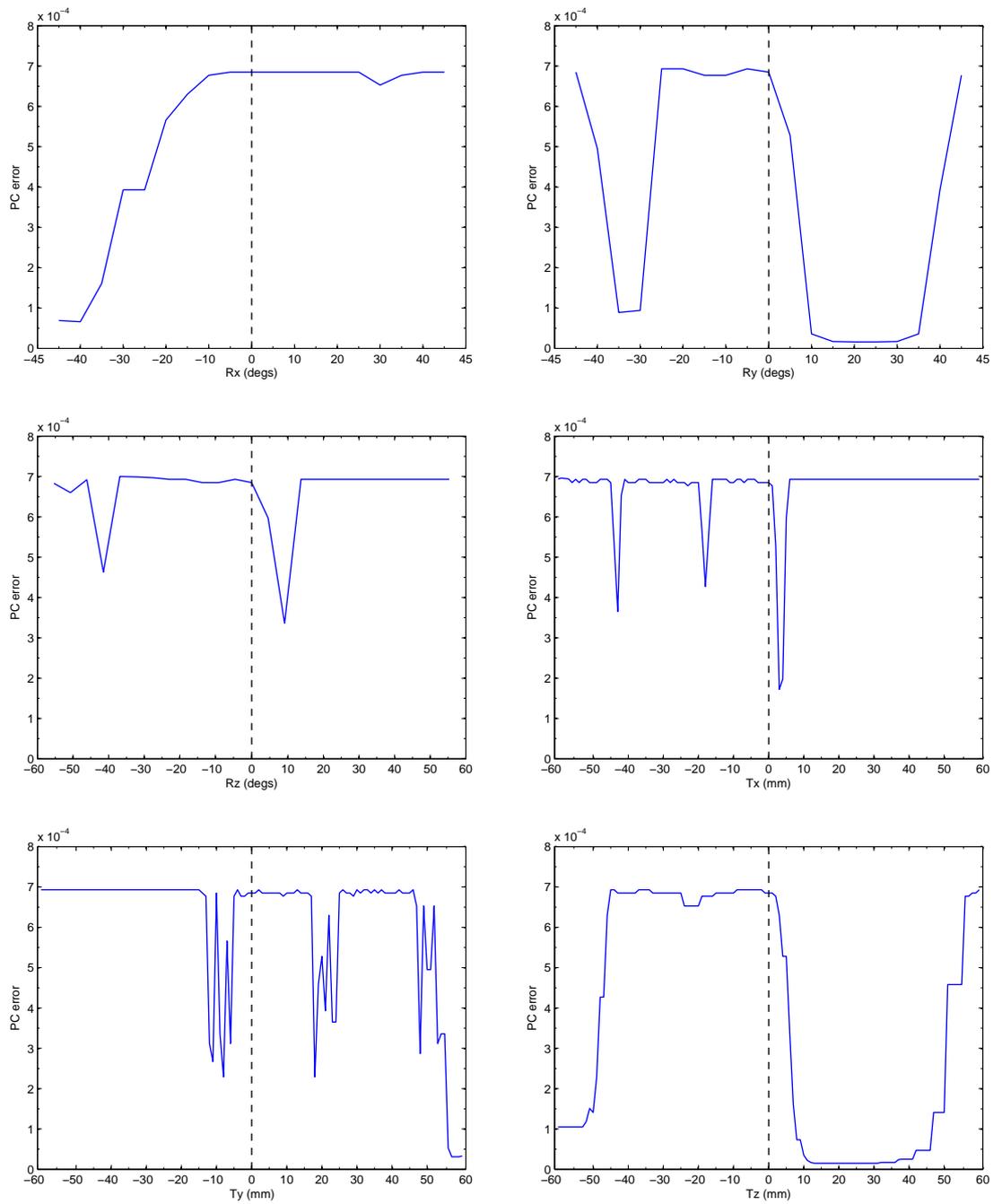


Figure 5.7: Registration errors for six independent DOFs based on the $PC_{inverse}$ cost function using a full-sized CT model.

5.4.2 Analysis of convergence of optimisation methods

The objective of this experiment was to analyse the convergence of different optimisation algorithms that attempt to find the global minimum within the search space. The setup involved the use of a stereoscopic surgical microscope in which the pair of cameras connected to the eyepieces were pre-calibrated and the initial registration was obtained through the ICP algorithm [153]. These two tasks were performed using the VBH mouth-piece [150] attached to the skull. The ROI was decided to be one of the eye sockets, where the voxels on the CT model were selected through the back-projection procedure. The resulting selection is similar to the window area shown in Figure 5.4(a).

Because at this point the virtual skull was aligned to the pair of real images, the current position was recorded as the ground truth. Additionally, the photo-consistency error was computed at this stage and used as the objective value. Then, the CT-based model was manually offset 2 mm along the X axis and -2 mm along the Y axis in object space. This position is regarded as a “starting pose”. The root mean squared distance (RMSD) was calculated with respect to the ground truth, producing a combined displacement of 2.82 mm. These offset values were chosen because the overall accuracy of current conventional SN systems is of the order 2-3mm. Moreover, the selected offset provides enough visual misalignment between real and virtual models at the magnification level (x6) produced by the microscope. It is worth mentioning that offset values larger than 3 mm caused the virtual model to be displayed outside of the visible image. Therefore, those levels of initial misalignment were excluded from the experiment.

The optimisation methods work by iteratively changing the registration position for a number of iterations until a global minimum value is found. Nevertheless, as the shape of the global function is unknown, the optimisation techniques can not rely on the calculation of derivatives. Two initial options were identified which comply with such restriction, Powell’s method and Quasi-Newton Without Derivatives. Although the latter is consid-

ered to perform faster, it was found that Powell’s algorithm provides a greater control in the modification of parameters over the six DOFs during the optimisation process. This is because the Quasi-Newton technique does not provide a method to control independently the changes of translation and rotation within registration. The other two approaches based on evolutionary algorithms, SDE and CODEQ, were evaluated for the optimisation of the photo-consistency cost function. These two methods are further described in Appendix B.

For this experiment, it was decided to select a single cost function among the similarity metrics based on the results of the previous tests. This allows focusing only on the performance of the optimisation methods and their convergence accuracy rather than on the effects of the similarity metrics. Thus, the cost function PC was chosen for the evaluation due to finding a more precise global minima than $PC_{inverse}$. It also shows higher gradients in the vicinity of the global minimum than the $PC_{squared}$ cost function, as seen from previous graphs, particularly for the translation components.

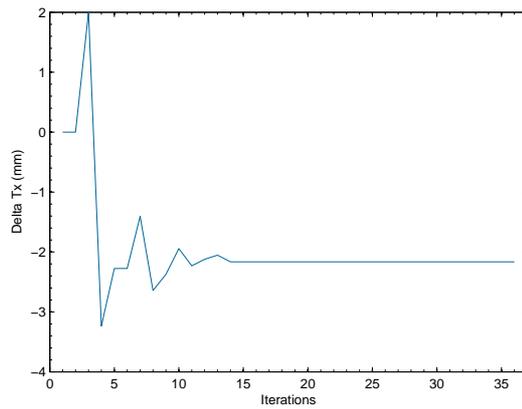
The results obtained using Powell’s method are shown graphically in Figure 5.8. The two plots at the top row illustrate the convergence for both translations T_x and T_y . In the left one, ΔT_x decreases from 0 mm to around -2 mm, corresponding to the changes in registration from the initial T_x offset to the final registration position. For the vertical translation, ΔT_y shows the alignment rectification that is obtained from the initial -2 mm offset to the correct pose. It can be noticed that registration is performed independently for each DOF at a time, where T_x is carried out initially until the correct alignment is found at iteration 14. Subsequently, the optimisation continues over T_y from iteration 15 onwards. The two plots at the bottom row illustrate the photo-consistency error value and RMSD convergence. In the case of the former, the overall error value decreases from 200 at the initial position to a value close to the ground truth (PC error = 121.69). The RMSD value converges to sub-millimetric accuracy for the combined T_x and T_y translations in object space.

Figure 5.9 presents the convergence results using SDE. The stochastic nature of the method is reflected in the spikiness of the curves, where the average data is fitted as a smooth curve to demonstrate the general trend. The convergence in both ΔT_x and ΔT_y shows an appropriate rectification from the initial offset. However, SDE executes the optimisation simultaneously for both translations in comparison with Powell's method. Concerning PC error and RMSD results, the trends decrease almost linearly towards the ground truth, especially for the latter which converges to a sub-millimeter error as in Powell's technique. It can be observed that the SDE algorithm involves a higher number of computational iterations to find the global minimum than the previous method due to its inherent random-based search.

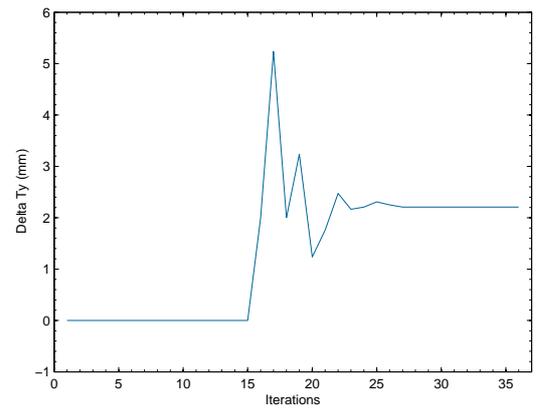
The performance obtained by the CODEQ algorithm is illustrated in Figure 5.10. The results indicate that while the final convergence of ΔT_x is close to the correct value, ΔT_y partially converges towards the ground truth. This limited registration over T_y is reflected in the overall convergence of the final RMSD, which is less accurate than the results obtained by SDE and Powell's method. In general, convergence graphs show that the search space covers a wider area within registration than previous techniques. This wider search space is represented by high and low spikes. In the case of PC error, these spikes correspond to large misalignments between real and virtual models. This is caused by the disposition of CODEQ to search in the opposite direction of the current pose, which also leads to extra cost function evaluations in each generation.

The results of this experiment present the overall convergence of different optimisation strategies to correct a misregistration in two dimensions. Among the techniques evaluated, SDE and Powell's method provide the best performance by closely approaching to the ground truth. Additionally, their final convergence reaches sub-millimetric levels. On the other hand, the convergence obtained by CODEQ seems to be limited to a certain extent in the tested registration. It also involves more computational load than the other two methods. The main cause for this is because CODEQ tends to cover a wider search space

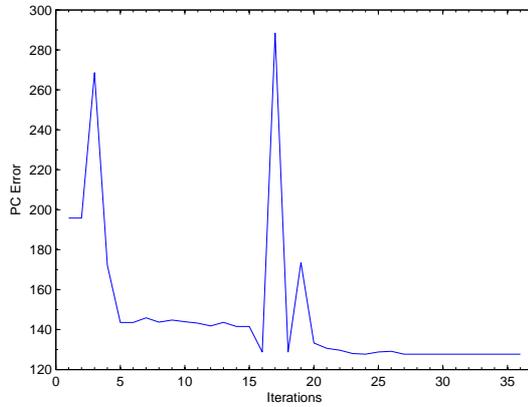
to avoid stagnation in local minima.



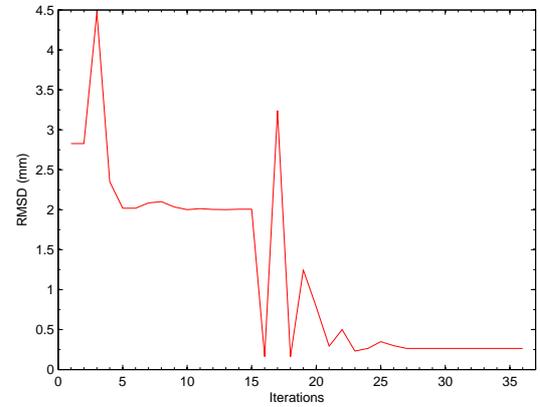
(a) Translation along X axis



(b) Translation along Y axis

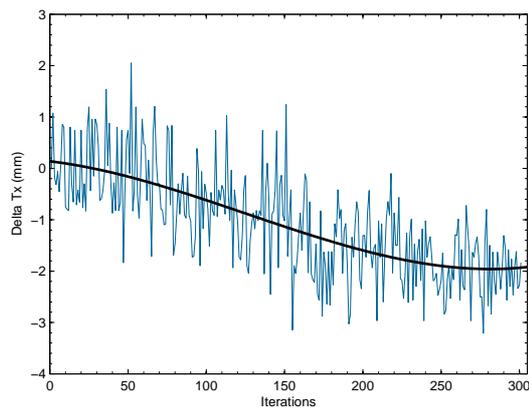


(c) Photo-consistency error

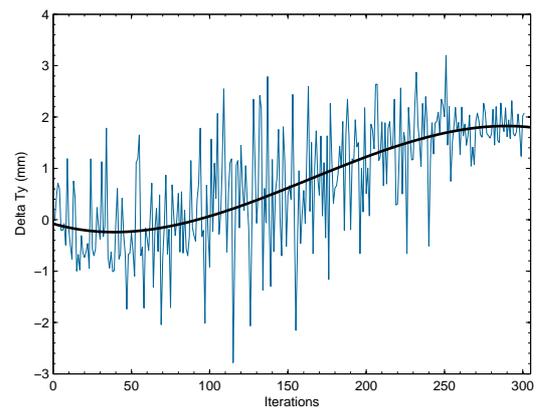


(d) Root mean square distance

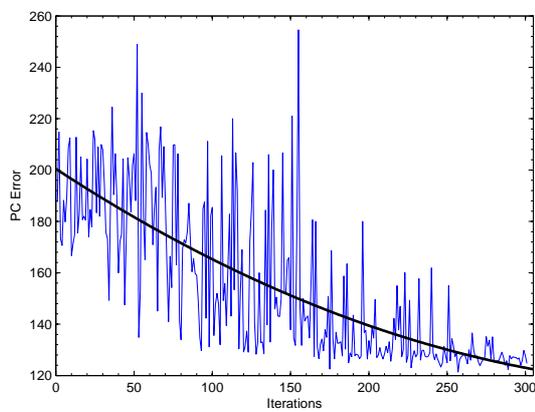
Figure 5.8: Overall convergence in *PC* photo-consistency based cost function using Powell's method.



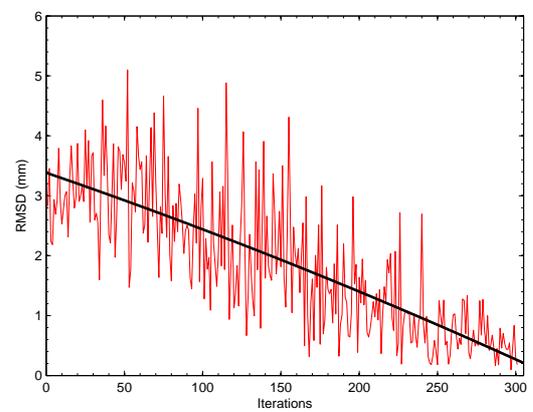
(a) Translation along X axis



(b) Translation along Y axis

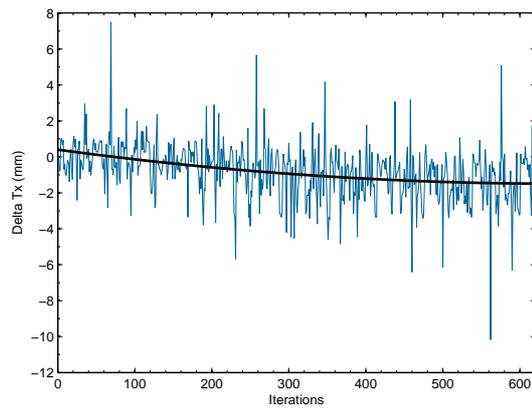


(c) Photo-consistency error

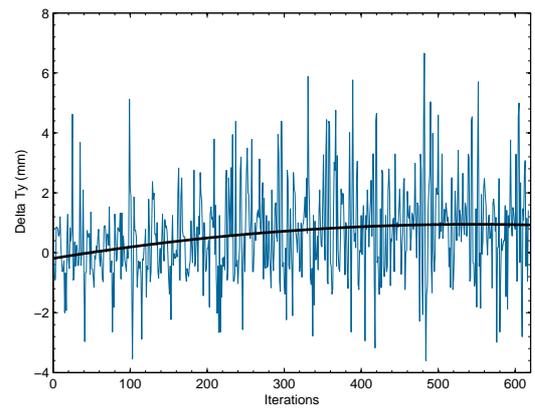


(d) Root mean square distance

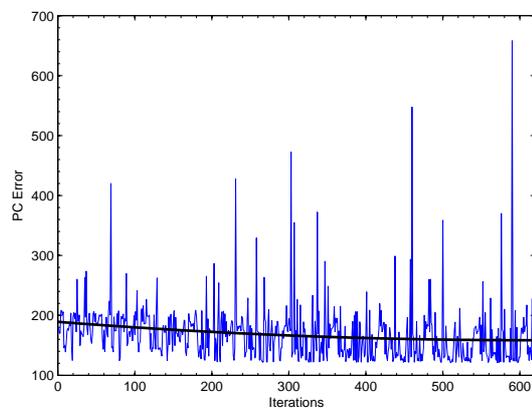
Figure 5.9: Overall convergence in *PC* photo-consistency based cost function using the SDE algorithm.



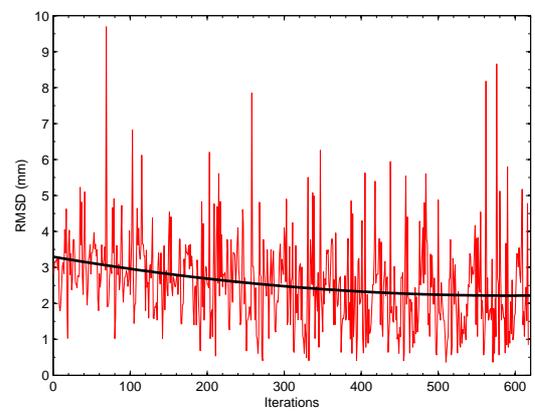
(a) Translation along X axis



(b) Translation along Y axis



(c) Photo-consistency error



(d) Root mean square distance

Figure 5.10: Overall convergence in *PC* photo-consistency based cost function using the CODEQ algorithm.

5.4.3 Intra-operative registration laboratory test

This experiment investigated the accuracy obtained through the combination of different techniques for intra-operative registration in a controlled laboratory setup. Specifically, the assessment involved a comparison among the three similarity metrics and three optimisation algorithms described earlier in this chapter. The procedure was carried out in a similar fashion to the previous experiment. A pair of cameras was calibrated and an initial ICP registration was performed to align real and virtual skull models. Then, the ROI was directed to an eye socket and the selection of voxels was obtained through back-projection. The current PC error and 3D pose were recorded as the registration ground truth values. Finally, the CT-based virtual model was manually offset for the purpose of simulating a misalignment during surgery due to tracking. In this regard, the offset PC error and RMSD values with respect to the ground truth were calculated.

The intra-operative registration procedure was implemented as a process thread that could run concurrently with the IESN system. The parameters used for the SDE and CODEQ optimisation techniques included a maximum number of ten generations and a population size corresponding to ten times the number of evaluated DOFs. In the case of Powell's method, the maximum number of iterations comprised a value of 100. In all optimisation algorithms, the tolerance threshold for the calculation of the cost function was selected as 10^{-6} for the PC and $PC_{squared}$ metrics, from initial function values of the order of 10^0 and 10^2 , respectively. This ensured that the cost function approached to a value small enough as to be considered null. In the case of $PC_{inverse}$, the tolerance threshold value was set to 10^{-10} from an initial PC error value of 10^{-3} .

It must be noted that in an IESN system, the world coordinate system is typically defined by the optical tracking device that registers the motions of all entities during surgery. Therefore, the misalignment between tracked cameras and patient is directly related to the tracker's world reference frame. Nevertheless, the concatenation of independent ref-

erence frames within the IESN system (e.g. camera, ICP registration) also influences the intra-operative registration. For this reason, it was decided to test the procedure using the cameras' coordinate system as the basis for pose transformation. A justification for this choice lies in the fact that the objective of the procedure is to obtain a visual match between real and virtual models, regardless of the coordinate system.

In order to carry out the photo-consistency registration under the camera coordinate system, the virtual object's coordinate system was mapped to the cameras' 3D position; which is based on the extrinsic parameters obtained in the calibration stage. This makes the reference frame of the virtual model to be aligned to the reference frame of the cameras. This system conversion was denominated object-to-camera coordinate transformation.

Five different offsets were applied to the virtual model in order to evaluate the effectiveness of the photo-consistency based registration to recover the initial alignment from different "starting poses". Each offset simulates a combination of the various DOFs which may affect the misregistration. Among all the possible combinations, the selected offsets evaluated in this experiment were T_xT_y , $T_xT_yT_z$, R_z , $T_xT_yR_z$ and $T_xT_yT_zR_xR_yR_z$. Depending on the DOFs involved, the offset values varied between 1 and 2 mm for translation and between 1° and 2° for rotation. This adjustment of offset values ensured that the virtual model remained inside the visible image before registration (larger initial offsets caused the model to be outside the image, as described earlier).

The results in Tables 5.1 - 5.5 present the mean registration accuracy obtained from a set of ten independent tests starting at different offset positions and comprising all combinations among photo-consistency metrics and optimisation techniques. The Initial PC column relates to the ground truth PC error with a corresponding initial RMSD value (not shown) of 0.00 mm for all combinations. Offset PC and Offset RMSD columns indicate the values at the "starting poses" that simulate accumulating tracking errors. The

final mean PC error (Final PC) and registration values (Final RMSD) are presented along with their standard deviations (\pm SD); where a lower RMSD means more accuracy in the optimisation. Final RMSD values lower than the Offset RMSD are shown in bold and the lowest Final RMSD is shown between brackets. The last column corresponds to the time taken to execute the procedure in seconds based on an Intel Core2 Quad processor computer at 2.4Ghz.

From the results it can be noticed that the overall performance of the registration using the object-to-camera coordinate transformation is moderately limited. Specifically, the difference between the initial RMSD value and the best registration for T_xT_y is slightly less than 1 mm (Table 5.1); whereas in the case of $T_xT_yT_z$ (Table 5.2) and the six DOFs (Table 5.5), the difference corresponds only to 0.11 and 0.43 mm, respectively. In Tables 5.3 and 5.4 the final registrations present a small improvement with respect to the initial offset RMSD values. Among all tests, the lowest RMSD values were obtained using SDE with different cost function combinations, except in the case of the six DOFs where SDE was second after Powell-PC.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	226.487	301.892	2.83	247.053 (± 16.356)	4.30 (± 2.01)	30
Powell-PC _{squared}	7.45016	8.57234	2.83	8.18116 (± 0.152)	2.48 (± 0.88)	22
Powell-PC _{inverse}	0.00134	0.001422	2.83	0.000220 ($\pm 6.57 \times 10^{-6}$)	6.76 (± 0.34)	24
SDE-PC	229.625	301.357	2.83	247.162 (± 0.825)	[1.99 (± 0.04)]	130
SDE-PC _{squared}	7.50339	8.65891	2.83	7.82220 (± 0.025)	2.00 (± 0.05)	120
SDE-PC _{inverse}	0.001493	0.001422	2.83	0.000091 ($\pm 1.89 \times 10^{-5}$)	4.68 (± 0.77)	264
CODEQ-PC	223.188	303.322	2.83	149.811 (± 8.810)	5.79 (± 0.32)	591
CODEQ-PC _{squared}	7.55097	8.64487	2.83	6.43033 (± 0.408)	5.58 (± 0.53)	767
CODEQ-PC _{inverse}	0.001545	0.001422	2.83	0.000135 ($\pm 2.41 \times 10^{-5}$)	5.93 (± 0.84)	328

^aOffset: $T_x, T_y = 2.0$ mm

Table 5.1: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x and T_y axes using the object-to-camera coordinate transformation.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	220.867	369.917	3.46	154.534 (± 9.533)	10.03 (± 0.34)	22
Powell-PC _{squared}	7.38829	9.75845	3.46	6.21945 (± 0.038)	10.19 (± 0.38)	21
Powell-PC _{inverse}	0.001516	0.001321	3.46	0.000182 ($\pm 4.92 \times 10^{-6}$)	6.15 (± 0.06)	22
SDE-PC	224.906	393.367	3.46	159.394 (± 5.789)	4.15 (± 0.49)	170
SDE-PC _{squared}	7.55953	10.07327	3.46	6.49344 (± 0.056)	[3.35 (± 0.42)]	139
SDE-PC _{inverse}	0.001571	0.001321	3.46	0.000068 ($\pm 7.60 \times 10^{-6}$)	4.70 (± 0.41)	141
CODEQ-PC	222.893	396.577	3.46	148.353 (± 5.750)	7.04 (± 2.34)	291
CODEQ-PC _{squared}	7.53526	10.08447	3.46	5.95028 (± 0.060)	5.82 (± 0.85)	456
CODEQ-PC _{inverse}	0.001513	0.001392	3.46	0.000064 ($\pm 1.48 \times 10^{-5}$)	5.33 (± 1.79)	449

^aOffset: $T_x, T_y, T_z = 2.0$ mm

Table 5.2: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x , T_y and T_z axes using the object-to-camera coordinate transformation.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	137.850	229.123	7.75	154.873 (± 1.900)	5.60 (± 0.03)	28
Powell-PC _{squared}	5.82275	7.58986	7.75	6.22327 (± 0.030)	5.57 (± 0.02)	29
Powell-PC _{inverse}	0.001187	0.001479	7.75	0.000392 ($\pm 1.59 \times 10^{-4}$)	6.31 (± 1.88)	29
SDE-PC	133.179	222.661	7.75	159.819 (± 0.591)	5.50 (± 0.03)	231
SDE-PC _{squared}	5.82275	7.47922	7.75	6.28205 (± 0.044)	5.49 (± 0.03)	243
SDE-PC _{inverse}	0.001098	0.001294	7.75	0.000321 (± 0.000)	[5.43 (± 0.00)]	223
CODEQ-PC	137.860	231.360	7.75	153.567 (± 0.506)	5.58 (± 0.01)	444
CODEQ-PC _{squared}	5.77283	7.46797	7.75	6.15973 (± 0.007)	5.55 (± 0.03)	410
CODEQ-PC _{inverse}	0.001098	0.00137	7.75	0.000319 ($\pm 2.68 \times 10^{-6}$)	6.91 (± 3.28)	428

^aOffset: $R_z = 2.0$ degrees

Table 5.3: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over R_z axis using the object-to-camera coordinate transformation.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	224.875	457.534	8.62	413.315 (± 3.515)	9.81 (± 0.06)	57
Powell-PC _{squared}	7.45330	10.62090	8.62	10.11158 (± 0.038)	9.71 (± 0.09)	40
Powell-PC _{inverse}	0.001413	0.00064	8.62	0.000277 ($\pm 4.84 \times 10^{-5}$)	9.67 (± 0.46)	48
SDE-PC	216.175	444.076	8.62	318.691 (± 5.034)	[6.53 (± 0.29)]	240
SDE-PC _{squared}	7.25072	10.48566	8.62	9.07228 (± 0.149)	6.67 (± 1.09)	336
SDE-PC _{inverse}	0.001493	0.00064	8.62	0.000233 ($\pm 1.07 \times 10^{-5}$)	10.42 (± 0.94)	265
CODEQ-PC	217.730	443.504	8.62	174.039 (± 7.872)	9.57 (± 3.11)	967
CODEQ-PC _{squared}	7.29993	10.44767	8.62	6.43297 (± 0.125)	8.24 (± 3.07)	438
CODEQ-PC _{inverse}	0.001602	0.00064	8.62	0.000198 ($\pm 4.73 \times 10^{-5}$)	10.91 (± 5.76)	402

^aOffset: $T_x, T_y = 2.0$ mm; $R_z = 2.0$ degrees

Table 5.4: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x , T_y and R_z axes using the object-to-camera coordinate transformation.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	246.404	149.108	6.85	125.118 (± 6.267)	[6.42 (± 0.11)]	44
Powell-PC _{squared}	7.93547	6.15121	6.85	5.55142 (± 0.024)	6.96 (± 0.23)	43
Powell-PC _{inverse}	0.00161	0.001471	6.85	0.000175 ($\pm 3.81 \times 10^{-5}$)	11.35 (± 1.71)	39
SDE-PC	249.679	150.587	6.85	129.870 (± 0.749)	7.17 (± 1.22)	238
SDE-PC _{squared}	7.76605	6.16244	6.85	5.76665 (± 0.028)	6.63 (± 1.14)	324
SDE-PC _{inverse}	0.00159	0.001471	6.85	0.000130 ($\pm 1.52 \times 10^{-5}$)	15.50 (± 1.94)	220
CODEQ-PC	240.162	143.893	6.85	134.088 (± 1.917)	7.40 (± 0.91)	444
CODEQ-PC _{squared}	7.57979	6.07271	6.85	5.66134 (± 0.098)	9.10 (± 1.46)	381
CODEQ-PC _{inverse}	0.00161	0.001471	6.85	0.000044 ($\pm 3.63 \times 10^{-5}$)	19.09 (± 4.18)	290

^aOffset: $T_x, T_y, T_z = 1.0$ mm; $R_x, R_y, R_z = 1.0$ degree

Table 5.5: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over 6 DOFs using the object-to-camera coordinate transformation.

The second set of experiments involved applying the pose transformations directly to the virtual object's own coordinate system during registration. Therefore, no mapping was required between different reference frames. The same initial offsets were applied to the CT model as in the preceding tests. Likewise, a set of ten independent trials were performed for each offset comprising the nine possible combinations among the photo-consistency metrics and optimisation methods.

It can be observed from the results in Tables 5.6 - 5.10 that the best accuracy among the experiments was obtained with the $SDE-PC_{squared}$ combination. The only exception was for $T_xT_yT_z$ where it was outperformed by $CODEQ-PC$, as seen in Table 5.7. However, the latter required a considerable longer time to execute. In the case of T_xT_y and $T_xT_yR_z$ offsets (Tables 5.6 and 5.9, respectively), the final registration achieved sub-millimetre accuracy using the same $SDE-PC_{squared}$ combination. In the results of Table 5.8, a slightly lower accuracy seems to be obtained in the registration over R_z , which produced a final RMSD of 1.52 mm. However, if the initial RMSD value (7.74 mm or a 2-degree offset) is taken into consideration, the resulting mismatch corresponds to less than 0.4° away from the ground truth.

The best RMSD accuracy obtained by registering a misalignment over the combined six DOFs (Table 5.10) was 2.81 mm from an initial offset of 6.80 mm. Although this final accuracy value is relatively high in comparison with the other five tests, it has to be considered that even a small angle deviation can affect the final RMSD accuracy as in the case of R_z . Specifically, the average RMSD for a 1-degree rotation over the three rotational components was determined to be equal to 3.09 mm in the target area.

Overall, the registration based on the virtual object's coordinate system produced better results than using the object-to-camera coordinate transformation. The difference in performance to recover the original alignment between models can be observed in the simplest case over T_xT_y , where the lowest RMSD obtained was 1.99 mm using $SDE-PC$

from a starting position at 2.83 mm using the object-to-camera approach. In contrast, the same experiment under the object coordinate system obtained sub-millimetre accuracy.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	123.383	195.826	2.82	127.309 (± 0.470)	0.40 (± 0.11)	21
Powell-PC _{squared}	5.65127	6.97168	2.82	5.65981 (± 0.036)	0.51 (± 0.26)	25
Powell-PC _{inverse}	0.00153	0.001374	2.82	0.000136 ($\pm 2.68 \times 10^{-6}$)	4.34 (± 0.16)	24
SDE-PC	122.776	187.579	2.82	122.674 (± 0.954)	0.27 (± 0.15)	142
SDE-PC _{squared}	5.48221	6.73800	2.82	5.46522 (± 0.008)	[0.25 (± 0.10)]	190
SDE-PC _{inverse}	0.00153	0.001374	2.82	0.000157 ($\pm 7.85 \times 10^{-6}$)	4.21 (± 0.55)	307
CODEQ-PC	118.924	179.710	2.82	121.507 (± 0.339)	0.63 (± 0.22)	308
CODEQ-PC _{squared}	5.44176	6.68142	2.82	5.47530 (± 0.020)	0.32 (± 0.14)	284
CODEQ-PC _{inverse}	0.001584	0.001448	2.82	0.000149 ($\pm 5.50 \times 10^{-6}$)	4.18 (± 0.16)	254

^aOffset: $T_x, T_y = 2.0$ mm

Table 5.6: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x and T_y axes using the object coordinate system.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	128.231	196.205	3.46	108.396 (± 3.173)	2.94 (± 0.07)	24
Powell-PC _{squared}	5.71403	7.03834	3.46	5.18594 (± 0.024)	3.01 (± 0.14)	24
Powell-PC _{inverse}	0.001657	0.000148	3.46	0.000135 ($\pm 3.89 \times 10^{-6}$)	3.16 (± 0.16)	28
SDE-PC	129.473	196.688	3.46	142.810 (± 5.896)	3.91 (± 1.08)	239
SDE-PC _{squared}	5.66093	7.02389	3.46	5.22263 (± 0.015)	2.37 (± 0.19)	128
SDE-PC _{inverse}	0.001637	0.000155	3.46	0.000132 ($\pm 2.73 \times 10^{-6}$)	2.90 (± 0.09)	136
CODEQ-PC	131.508	192.059	3.46	111.455 (± 2.016)	[2.12 (± 0.73)]	336
CODEQ-PC _{squared}	5.78088	7.00121	3.46	5.62237 (± 0.238)	2.85 (± 1.50)	388
CODEQ-PC _{inverse}	0.00166	0.000155	3.46	0.000083 ($\pm 2.27 \times 10^{-5}$)	9.95 (± 1.65)	396

^aOffset: $T_x, T_y, T_z = 2.0$ mm

Table 5.7: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x , T_y and T_z axes using the object coordinate system.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	138.145	256.442	7.74	124.004 (± 1.450)	1.89 (± 0.35)	28
Powell-PC _{squared}	5.80756	8.00074	7.74	5.59683 (± 0.025)	2.10 (± 0.17)	32
Powell-PC _{inverse}	0.001441	0.001777	7.74	0.000651 ($\pm 4.04 \times 10^{-5}$)	3.07 (± 0.98)	50
SDE-PC	140.499	263.122	7.74	127.420 (± 0.131)	1.93 (± 0.21)	228
SDE-PC _{squared}	5.78668	8.03974	7.74	5.52660 (± 0.007)	[1.52 (± 0.09)]	300
SDE-PC _{inverse}	0.001196	0.001821	7.74	0.000287 (± 0.000)	5.04 (± 0.04)	212
CODEQ-PC	132.202	254.865	7.74	120.348 (± 0.209)	1.76 (± 0.22)	458
CODEQ-PC _{squared}	5.83353	7.99179	7.74	5.56560 (± 0.013)	1.62 (± 0.11)	460
CODEQ-PC _{inverse}	0.001441	0.001799	7.74	0.000271 (± 0.000)	5.17 (± 0.04)	465

^aOffset: $R_z = 2.0$ degrees

Table 5.8: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over R_z axis using the object coordinate system.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	116.478	304.159	8.54	119.332 (± 1.462)	2.60 (± 0.54)	39
Powell-PC _{squared}	5.43060	8.68608	8.54	5.48735 (± 0.027)	3.16 (± 0.10)	37
Powell-PC _{inverse}	0.001598	0.001415	8.54	0.000300 ($\pm 4.34 \times 10^{-5}$)	9.53 (± 0.10)	37
SDE-PC	118.770	298.237	8.54	121.288 (± 0.444)	0.78 (± 0.26)	235
SDE-PC _{squared}	5.36922	8.56755	8.54	5.42337 (± 0.023)	[0.68 (± 0.48)]	333
SDE-PC _{inverse}	0.001603	0.000238	8.54	0.000139 ($\pm 9.34 \times 10^{-6}$)	3.96 (± 0.64)	240
CODEQ-PC	132.419	343.275	8.54	123.765 (± 1.843)	2.44 (± 2.64)	423
CODEQ-PC _{squared}	5.75915	9.34433	8.54	5.60431 (± 0.018)	1.61 (± 0.96)	527
CODEQ-PC _{inverse}	0.001465	0.000409	8.54	0.000109 ($\pm 5.68 \times 10^{-5}$)	7.22 (± 7.73)	396

^aOffset: $T_x, T_y = 2.0$ mm; $R_z = 2.0$ degrees

Table 5.9: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over T_x, T_y and R_z axes using the object coordinate system.

	Initial PC	Offset PC	Offset RMSD ^a	Final PC	Final RMSD	Time (secs)
Powell-PC	124.982	162.216	6.80	116.646 (± 2.90)	3.13 (± 0.56)	45
Powell-PC _{squared}	5.66781	6.43791	6.80	5.35676 (± 0.128)	3.36 (± 0.61)	47
Powell-PC _{inverse}	0.001388	0.000216	6.80	0.000133 ($\pm 3.03 \times 10^{-6}$)	7.23 (± 0.19)	49
SDE-PC	122.508	152.960	6.80	108.177 (± 1.221)	2.86 (± 0.97)	268
SDE-PC _{squared}	5.48056	6.35041	6.80	5.22126 (± 0.019)	[2.81 (± 0.82)]	204
SDE-PC _{inverse}	0.001465	0.000195	6.80	0.000122 ($\pm 2.22 \times 10^{-5}$)	7.74 (± 3.48)	282
CODEQ-PC	125.193	157.573	6.80	110.081 (± 1.466)	4.86 (± 3.16)	380
CODEQ-PC _{squared}	5.57853	6.46819	6.80	5.30166 (± 0.033)	5.26 (± 5.14)	405
CODEQ-PC _{inverse}	0.001468	0.000253	6.80	0.000033 ($\pm 2.44 \times 10^{-6}$)	15.05 (± 3.44)	409

^aOffset: $T_x, T_y, T_z = 1.0$ mm; $R_x, R_y, R_z = 1.0$ degree

Table 5.10: Comparison of photo-consistency (PC) cost functions and optimisation algorithms. PC mean error values and RMSD (\pm SD) in mm over 6 DOFs using the object coordinate system.

For both types of coordinate systems tested (object and object-to-camera) the results obtained by CODEQ generally produced the highest RMSD values after registration, which means that it partially converged during registration. Moreover, CODEQ required the longest time to finish the procedure in comparison with SDE and Powell's method. As it was discussed in the previous analysis, CODEQ aims at searching in the opposite direction of the current position for each step in the process. This behaviour can be problematic as in specific cases the virtual model goes out of scope within the visible image, causing the final position to be visually misregistered. On the other hand, Powell's method occasionally stagnates at incorrect areas in the image that present low intensity levels. Such incapability to avoid stagnation is caused by evaluating each DOF at a time in a sequential manner. However, the main advantage of Powell's method is its speed to converge to a solution.

Regarding the similarity metrics, PC and $PC_{squared}$ proved to be convenient cost functions for intra-operative registration. $PC_{squared}$ slightly outperformed PC in the experiments using the object coordinate system. However, it was found that the difference between them was of no statistical significance based on a two-tailed t-test with a P-value of 0.05 ($t=0.203$, $df=7$, $P>0.05$). The use of $PC_{inverse}$ negatively affected the final convergence accuracy in most trials by producing the largest RMSD among the combinations. Furthermore, in some cases the final value obtained by $PC_{inverse}$ exceeded the RMSD corresponding to the starting offset position.

The outcomes of this experiment demonstrate the performance of different similarity functions and optimisation algorithms for intensity-based registration purposes. The best results were obtained by modifying the registration position in the virtual object coordinate system using SDE together with $PC_{squared}$. Figures 5.11 - 5.15 illustrate the initial and final registration overlays around the eye socket at different DOFs as seen by one camera connected to the surgical microscope. It can be noticed that although the virtual and real models are close from each other after registration over six DOFs (Figure 5.15),

there is a small rotational deviation in the corners of the anatomical structure. This small rotation produces high values in the computation of the final RMSD in the target area.

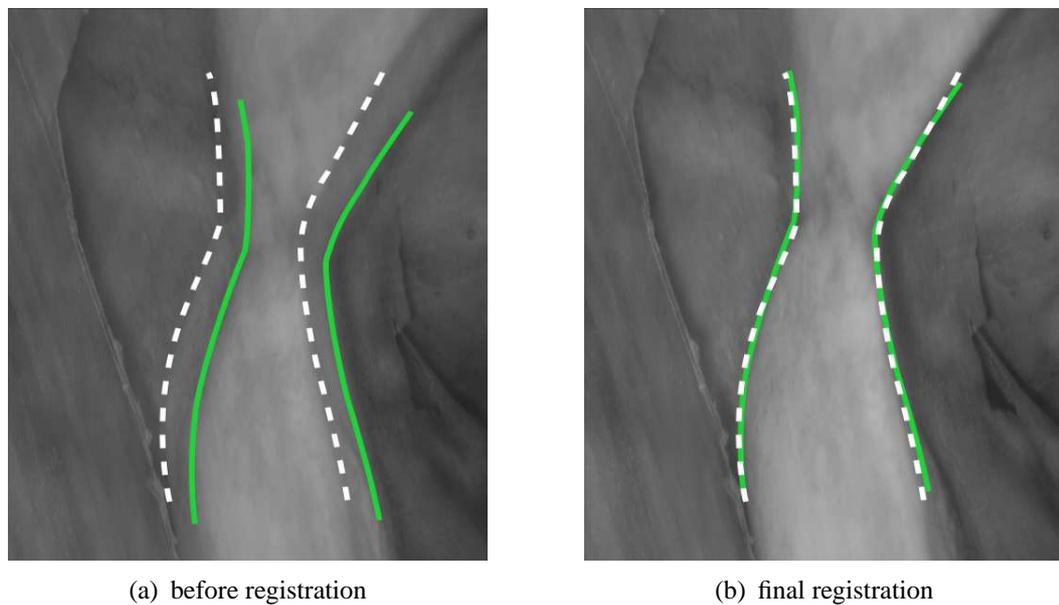


Figure 5.11: Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a $T_x T_y$ misalignment. White/dashed lines indicate contour features in the real model. Green/solid lines show contour features in the virtual model.

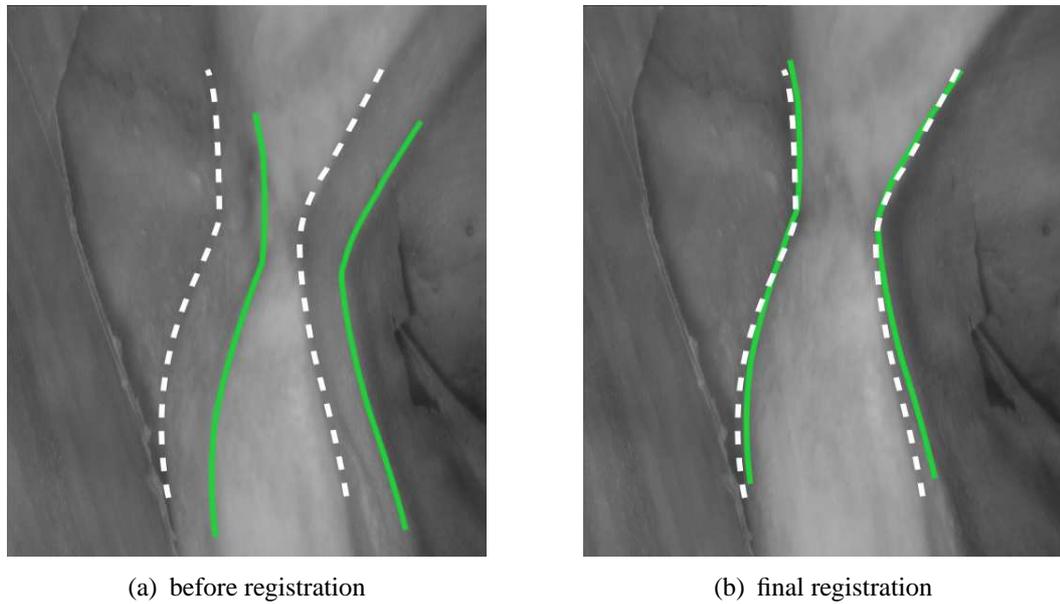


Figure 5.12: Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a $T_x T_y T_z$ misalignment. White/dashed lines indicate contour features in the real model. Green/solid lines show contour features in the virtual model.

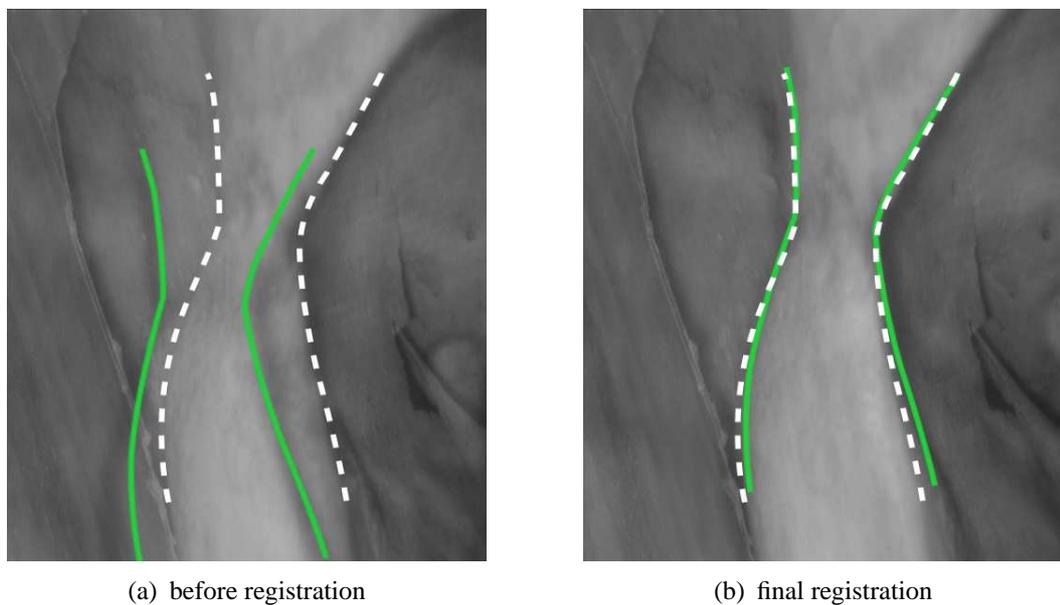


Figure 5.13: Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a R_z misalignment. White/dashed lines indicate contour features in the real model. Green/solid lines show contour features in the virtual model.

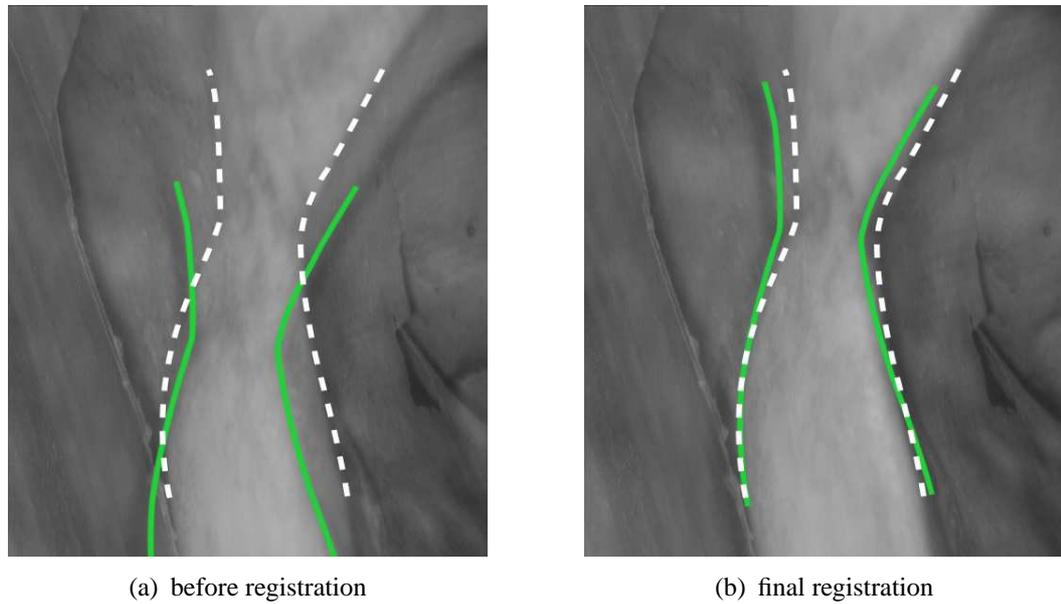


Figure 5.14: Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over a $T_x T_y R_z$ misalignment. White/dashed lines indicate contour features in the real model. Green/solid lines show contour features in the virtual model.

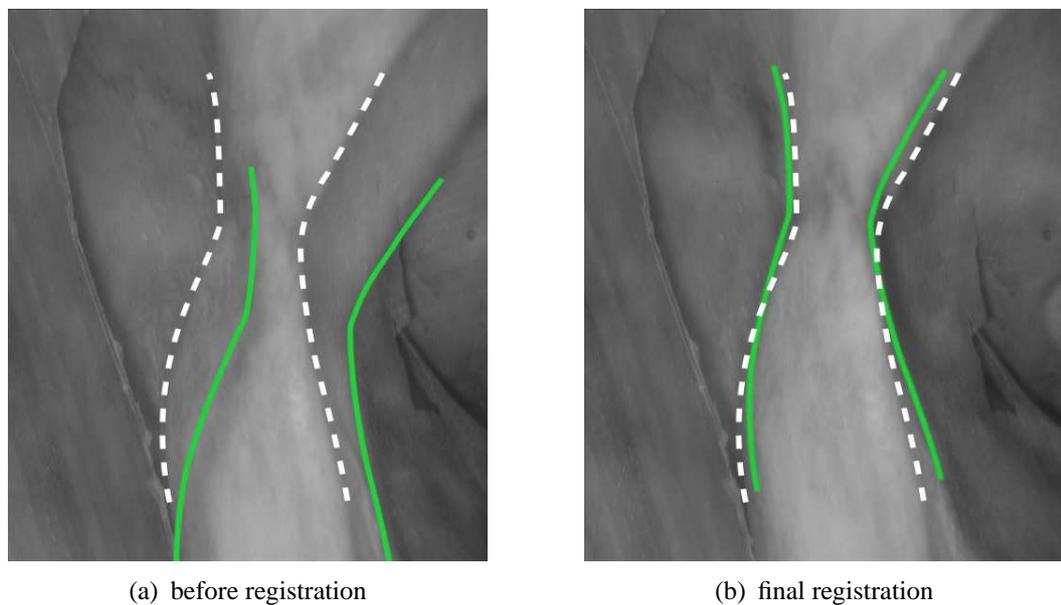


Figure 5.15: Photo-consistency based registration of real and virtual models around the skull eye socket using SDE optimisation and $PC_{squared}$ cost function over six DOFs. White/dashed lines indicate contour features in the real model. Green/solid lines show contour features in the virtual model.

5.5 Summary

In this chapter, the use of photo-consistency as a cost function for intensity-based registration was introduced. The first experiments demonstrated that the use of PC and $PC_{squared}$ similarity metrics provide comparable results in the alignment of full-sized models within a synthetic environment. Subsequently, photo-consistency was applied as a technique for intra-operative registration based solely on the visual information obtained from a pair of cameras connected to a stereo surgical microscope. The results showed the potential of this method to compensate for tracking errors during ENT surgery without resorting back to the initial camera calibration and ICP registration procedures and setup.

Among the optimisation algorithms evaluated, SDE proved to be the most suitable for the alignment between real and virtual models of a human skull, particularly when using the $PC_{squared}$ cost function during registration in the object coordinate system. SDE provides a balance between the best registration accuracy and the overall time required to execute the procedure. However, it must be taken into consideration that, as in any other optimisation technique, the resulting accuracy depends on the initial parameters. In the case of registration, these parameters involve the initial offset distance and the DOFs involved in the mismatch.

The performed experiments were based on the use of a human skull as a dummy patient and a pair of black and white cameras. In a real surgical environment, specular reflection caused by the use of the microscope light could affect the registration procedure as the scene would no longer conform to a Lambertian model. A method to solve this issue could consist in employing colour cameras to capture the real images. Subsequently, a filter such as the ones described by Tan et al. [154] and Yoon and Kweon [155] could be applied to neutralise the specularities by extracting the diffuse component of the colour images and omitting the specular component. Then, photo-consistency could be either

carried out in the resulting images using colour differences, or converting the images to greyscale values in order to compare intensity levels.

Chapter 6

Conclusions and future work

6.1 Summary and conclusion

Image-enhanced surgical navigation (IESN) systems aim to enhance intra-operative images, acquired through endoscopes or surgical microscopes, with computer-generated medical images obtained at a pre-operative stage. Therefore, the accuracy provided by an IESN system is of vital importance for the assistance of a practitioner in the operating theatre. An appropriate visual overlay between real and virtual imagery can reassure the surgeon's judgement and improve the surgical performance.

The objective of this research, as presented in this thesis, was to optimise different aspects of an IESN system that directly affect the overall accuracy during a surgical procedure targeted at ear, nose and throat (ENT) interventions. First, a study of augmented reality (AR) in the field of medicine was introduced, followed by a description of the different hardware components used in AR: display technologies and motion tracking devices. In the case of the former, a monitor can be regarded as the standard output device to view the imagery captured by an endoscope; whereas a stereoscopic monitor can be used to display the images obtained through a stereo surgical microscope. For the IESN sys-

tem used in this research — which allows the use of both endoscope and microscope — a balance in the visual requirements is achieved by using a stereoscopic monitor that permits alternating between single and stereo display mode. Concerning the motion tracking devices, an optical tracker employs markers that can be attached to surgical instruments, cameras and patient. Besides allowing the detection of movement of different entities simultaneously using a single central detector, optical trackers are not obtrusive as opposed to electromechanical trackers. Also, the volume of detection offered by the optical device is larger than the volume provided by electromagnetic tracking technology.

Chapter 3 presented a detailed description of a number of optimisations made to the software application in which the IESN system is based upon — ARView. The procedure to detect feature markers of a calibration object for endoscopic sinus surgery (ESS) was analysed. It was found that the segmentation of features deeply affects the calibration results, both quantitatively and qualitatively. A procedure to reduce segmentation problems during feature detection was carried out, which improved the overall calibration accuracy. Additionally, a validation of the accuracy of two optical tracking devices (i.e. hybrid Polaris and Polaris Vicra) was performed in a working environment that represented a surgical scenario. The results obtained were compared to the accuracy levels described by the manufacturer. It was found that the average difference for both devices was around 0.7 mm higher than their corresponding specifications. Subsequently, the software implementation associated to the motion tracking was optimised in order to resolve problems that occurred when two objects were registered simultaneously and a filter was applied to smooth signal noise. The solution involved the synchronisation of the function calls used to read the positional data from the optical tracker. Another area of improvement involved the stereo visualisation of virtual models. While the underlying implementation allowed the use of stereoscopic displays, it was considered that the three-dimensional “pop-out” effect produced by the software was inappropriate. This was caused by a misalignment between the pair of virtual cameras and rendering viewports that correspond to

the user's eyes. The description of the implemented solution — based on the stereoscopic cross-viewing method — was explained.

A method to optimise the estimation of the focal length for procedures involving a stereoscopic surgical microscope was described in Chapter 4. This work initially evaluated the sensitivity of Tsai's photogrammetric calibration method towards 2D marker detection of the projected calibration image and the 3D positional accuracy of the markers on a non-coplanar calibration object. It was demonstrated that these two factors greatly affect the accuracy of the calibration procedure, in particular the focal length estimate. Then, the work presented a new hybrid approach that uses the photogrammetric method as a pre-calibration step and a self-calibration method in order to optimise the focal length calculation. For this purpose, an evaluation was carried out involving three different techniques to calculate the fundamental matrix from a pair of cameras (i.e. linear, gradient-based and M-estimators) and three self-calibration methods for the computation of the focal length (i.e. Newsam, Bougnoux and Sturm's methods). Furthermore, three optimisation algorithms were employed for the refinement of the focal length values, comprising two evolutionary algorithms, i.e. Self-adaptive Differential Evolution (SDE) and CODEQ, and the Levenberg-Marquardt method. The results indicate that when one of the camera views is rotated about 15° around an elevation axis, the use of an evolutionary algorithm can increase the calibration accuracy originally obtained by the photogrammetric method. An improvement in the calibration error of around 10% was obtained by combining the gradient-based and CODEQ techniques, using any of the three algorithms for the computation of the focal length.

Chapter 5 investigated the use of an intensity-based metric, known as photo-consistency, for the intra-operative registration of virtual and real models to compensate for accumulating tracking errors in an IESN system. The technique was based on the intensity difference between images acquired from a pair of cameras connected to a stereo surgical microscope. First, an evaluation of three different photo-consistency cost functions

reported in the literature was undertaken. These cost functions were denominated PC , $PC_{squared}$ and $PC_{inverse}$. Subsequently, a study of convergence among three optimisation algorithms (i.e. Powell's method, SDE, and CODEQ) provided an insight of their performance for a simple misregistration case. Then, the combination among cost-functions and optimisation methods was assessed for a set of simulated misalignments, where different DOFs and coordinate systems were involved. SDE proved to be the most suitable optimisation method together with the $PC_{squared}$ cost function during registration in the object coordinate system. The results obtained indicate that photo-consistency could be used as a metric for intra-operative registration in ENT surgery.

6.2 System integration

Chapters 3 - 5 of this thesis have independently investigated several software-based techniques that optimised the accuracy of an IESN system. Among the different evaluated techniques, a number of methods have been selected to be included in the final integrated system. The selection is based on the optimisation results obtained and discussed in the corresponding chapters.

As described previously, ARView provided the basic application framework for this research project. Therefore, some of the functionalities of the final IESN system are inherited from its predecessor. In particular, the same workflow is followed to achieve an overlay between virtual and real anatomies during surgery. This workflow, commonly used in AR based systems, involves the steps of camera calibration, registration, motion tracking and visualisation.

Regarding camera calibration, the identification of feature markers located in a calibration object is performed in order to estimate the internal camera parameters and establish a relationship between world and camera coordinate systems. The final integrated sys-

tem includes a method to reduce feature segmentation in the detected markers of a planar calibration grid for ESS. It was shown that, after the optimisation, the overall calibration accuracy was improved more than double (from 1.50 to 0.66 pixel error) with respect to the original implementation. In the case of calibration using a stereoscopic microscope, the final integrated system introduces a new hybrid approach that uses a photogrammetric technique (Tsai) for pre-calibration and a self-calibrating stereoscopic technique to derive the fundamental matrix with the aim of optimising the calculated focal length and reducing the camera calibration error. For this reason, a new 3D calibration object was designed and manufactured. The 3D object involves a set of feature markers at various depth levels that serve as the input for the different camera calibration algorithms. Specifically, the gradient-based method is selected for the estimation of the fundamental matrix between the two microscope cameras. Although the three techniques used as cost functions for the computation of the focal length produced similar results, Bougnoux's method could be chosen in the final IESN system because it involves a simple closed-form solution. Among the optimisation methods, the CODEQ algorithm provided the best results during the evaluation hence its use is recommended. An improvement of around 10% in calibration accuracy was achieved by the implemented hybrid technique with respect to the original method, both using the 3D calibration object. Table 6.1 summarises the selected algorithms to be included in the camera calibration procedure of the final integrated system.

	Selected solution
Fundamental matrix estimation	Gradient-based algorithm
Focal length estimation	Bougnoux's method
Optimisation algorithm	CODEQ

Table 6.1: Selected methods for camera calibration.

It must be mentioned that in previous work using the endoscope [17, 149], sub-pixel calibration accuracy could be obtained due to a higher precision in the manufacturing of the planar calibration object. In the presented system, the calibration object was less accurately manufactured due to limitations of the 3D printer employed. Nevertheless, by improving the precision of the 3D calibration object the same level of initial calibration accuracy is likely to be obtained, which can be further minimised by applying the hybrid method.

Similarly to its predecessor, the final integrated system requires the use of the ICP algorithm to obtain an initial registration between real and virtual models at the beginning of the surgical procedure. In addition to this method, the final system includes a technique to intra-operatively register the models when their alignment is affected due to accumulated tracking errors. The introduced registration method is based on photo-consistency using a pair of images acquired by the cameras connected to a surgical microscope. The photo-consistency cost function to be selected for the implementation of the integrated IESN system is called $PC_{squared}$ whereas SDE would be the chosen optimisation algorithm that minimises the photo-consistency error. The preferred coordinate system to perform the registration is the denominated object coordinate system. As its name suggests, the pose transformations are directly applied to the virtual object's own coordinate system without involving any transformation between different reference frames. A summary of the selected options for intra-operative registration is listed in Table 6.2.

	Selected solution
Photo-consistency cost function	$PC_{squared}$
Optimisation algorithm	SDE
Coordinate system	Object coordinate system

Table 6.2: Selected options for intra-operative registration.

The resulting accuracy of the intra-operative registration method depends on the initial misalignment between real and virtual models. In the simplest scenario that includes only a translational offset in X and Y coordinates, the final RMSD achieved by the final integrated system is sub-millimetric. As more DOFs are combined in the offset, the misregistration tends to increase mainly due to the rotational axes involved. It was proven that the obtained accuracy of the final integrated system is within 3 mm from a misregistration that combines 6 DOFs, corresponding to an initial offset of 6.8 mm. This allows recovering the overlay to the same level of accuracy currently obtained by conventional IESN systems for ENT, which is in the range of 2 - 3 mm when accumulated tracking errors are not present.

In reference to the tracking of surgical instruments and patient during the intervention, the IESN system registers the positional data of the entities using an optical tracking system. The original implementation included a set of basic functions that read the tracking signal and apply the data to the volumetric model on screen. In addition to these functions, the final integrated system introduces a method to synchronise the process calls that pull data from the optical tracking device. This synchronisation — based on semaphores — ensures the integrity of the data when two entities are tracked simultaneously.

Finally, the visualisation of real and virtual models in the IESN system requires the use of stereoscopic monitors that show the user two views of the surgical area as obtained by the cameras connected to the surgical microscope. In the final integrated system, a method was implemented to align virtual cameras and rendering viewports following the stereoscopic cross-viewing approach. This allows projecting the 3D imagery towards the user while providing higher depth information with respect to the original IESN implementation. It should be highlighted that, although stereo visualisation has been proved to be of importance for the surgeon using optical instruments such as stereoscopic microscopes, the use of stereo AR still presents perceptual issues that have to be addressed before being completely introduced in the operating theatre. Particularly, Johnson et al. [110] described

some issues related to depth perception using stereo overlays when a virtual model that is located under a real surface is rendered on top of it. In their study, they found that the problem could be reduced by rendering a virtual version of the physical object and the virtual object at the same time; however, the perceptual error could not be entirely eliminated in the optical see-through microscope. By contrast, the IESN system presented in this research involves a video-based AR application that allows the clinician to select the level of blending between virtual and real models on screen, which could help to reduce ambiguity in the estimation of depth. In addition, the surgeon can disable the rendering of virtual models if depth perception is considered to be compromised. Nevertheless, additional studies must be carried out to understand the causes that affect this and other perception issues in the use of IESN systems.

6.3 Further work

The presented IESN system has been evaluated in a laboratory setup, where the optimisation levels achieved by the different techniques have been considered as satisfactory. However, the resulting improvements need to be tested in a real surgical environment in order to assess their performance in new conditions and, if necessary, extend their functionality. Some aspects of this research that could be further explored include the following areas:

- The assessment of optical tracking devices determined that the manufacturer's nominal specifications and the obtained accuracy in a working environment are different. Although little can be done to improve the tracking performance due to hardware limitations, new optical devices by different manufacturers are constantly under development. These new systems differ in their specifications, such as active or passive technology, measurement volume, and accuracy. A future eval-

uation of different systems could provide a better judgment of the optimal solution for an IESN system aimed at ENT interventions.

- Concerning the camera calibration procedure, the current resolution of the manufacturing device used to produce the non-coplanar calibration object for this research is relatively limited. This issue influences the accuracy levels obtained in the overall camera calibration. In order to achieve sub-pixel accuracy, the dimensional precision of the manufactured object must be improved. Additionally, the method presented for the optimisation of the focal length could be extended to non-stereoscopic devices such as endoscopes. In this regard, the endoscope could be placed at different positions to acquire a pair of images of the calibration object.
- It was found that the accuracy of the intra-operative registration method varied with respect to the number of DOF involved in the initial misalignment (“starting pose”) when using SDE. This is likely because the stochastic-based optimiser did not search exhaustively over all DOFs, hence providing only an approximation to the real global minimum. In order to improve the registration, a solution could involve applying a second optimisation technique such as Powell’s method to refine the search.
- The experimental setup was based on the use of a human skull as a dummy patient. In a real surgical scenario, the microscope light could produce specularities in the captured images. Thus, the scene would not strictly correspond to a Lambertian model, which might affect the registration accuracy. In order to overcome this problem, a solution could involve using a pair of colour cameras to acquire the real images and apply a colour filter that extracts the diffuse component and omits the specularities. Then, the photo-consistency procedure could be carried out by converting the resulting images to greyscale values or by comparing colour differences.

- Current processing times to perform the presented intra-operative registration are of the order of minutes. A further implementation on the graphics processing unit (GPU) using a parallel methodology such as CUDA¹ or OpenCL would significantly speed up the overall registration process. This would become a viable method for real-time registration in IESN.

¹developed by NVIDIA (www.nvidia.com)

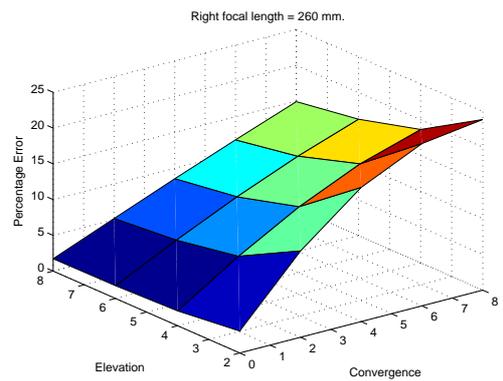
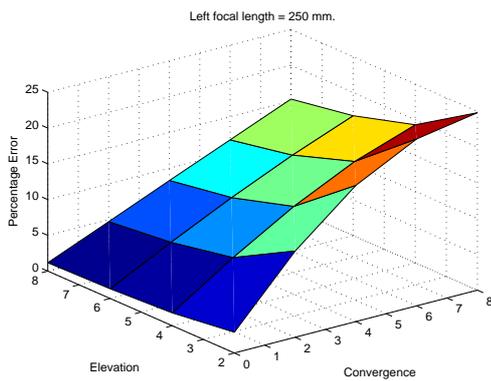
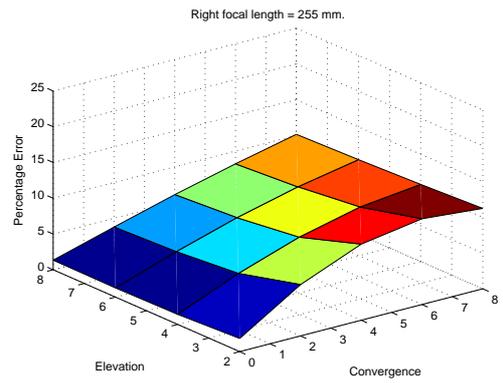
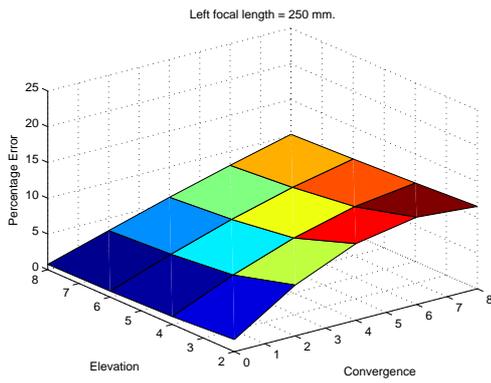
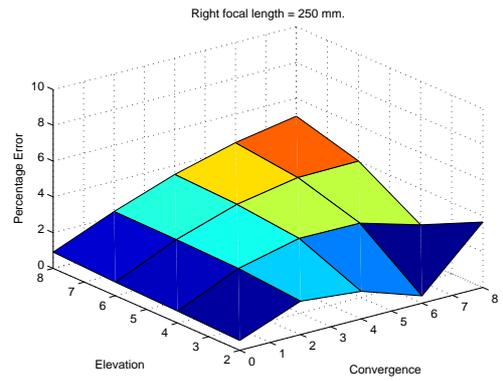
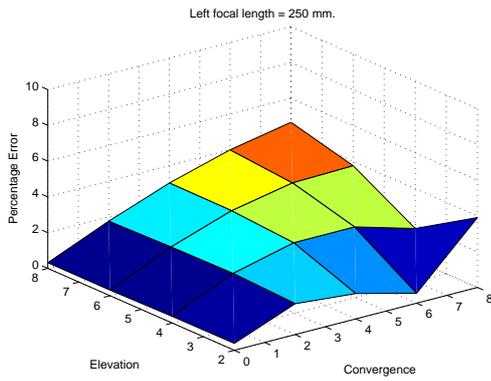
Appendix A

Results from the focal length optimisation experiments

This appendix contains a complete set of graphs and tables used for the evaluation of the results obtained in Chapter 4, which corresponds to the optimisation of focal length based on stereo camera calibration.

A.1 Accuracy of focal length estimation in a synthetic environment — Graphs

This section presents a series of graphs corresponding to the results obtained by the three methods of focal length estimation using a set of artificial points in a synthetic scenario. The compared techniques include the algorithms of Bougnoux [122], Sturm [126, 127] and Newsam [125] that were described in Section 4.2.4.



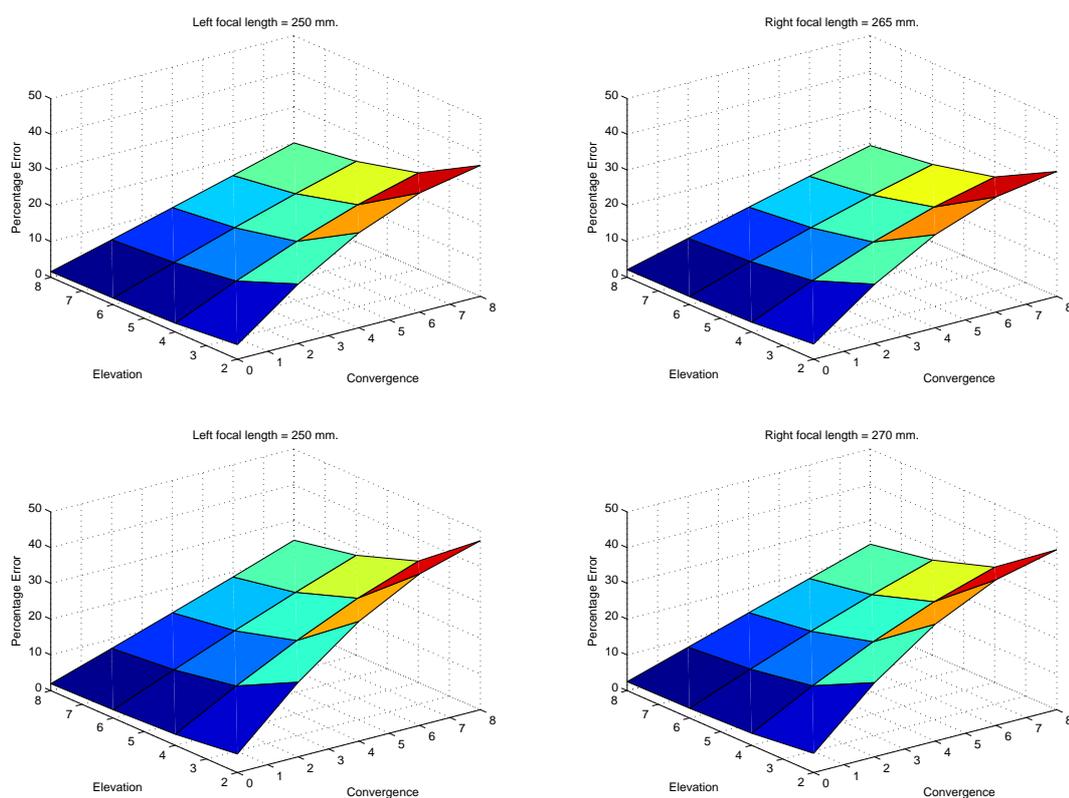
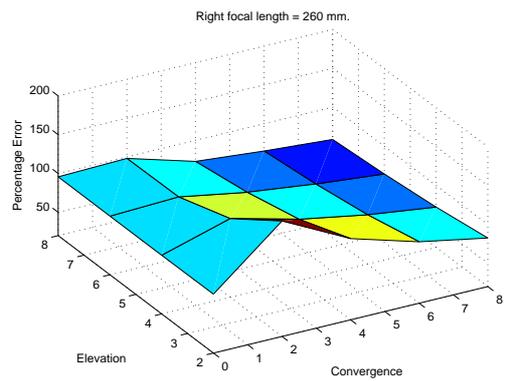
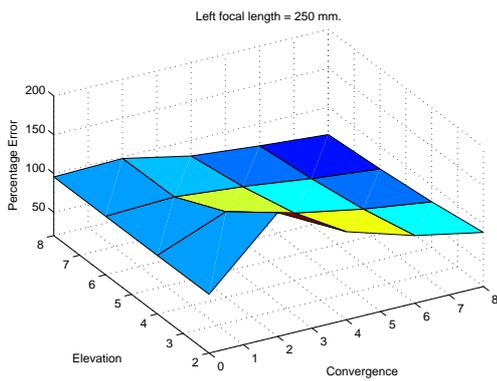
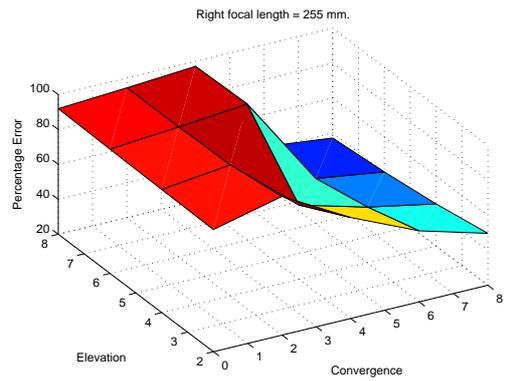
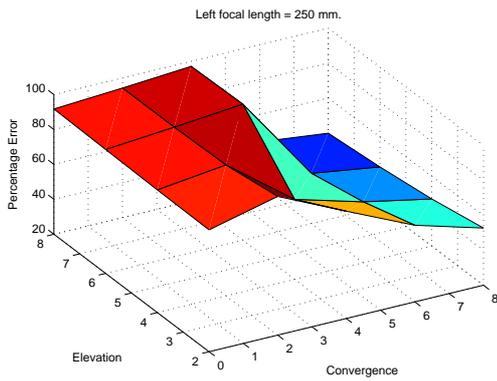
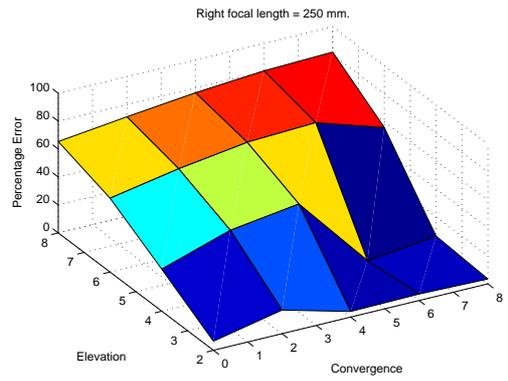
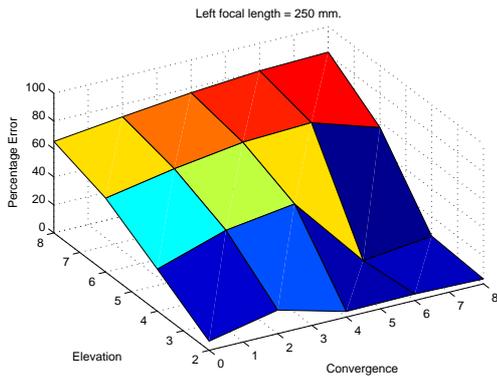


Figure A.1: Percentage errors in focal length calculation using Bougnoux's Method. (Left column) Left camera errors. (Right column) Right camera errors. First row represents an equal focal configuration, second to fifth rows represent unequal focal values between cameras. Elevation angles between 0° and 2° failed to produce a result (critical configuration), hence they are not displayed. Note that the percentage error scale changes among the graphs.



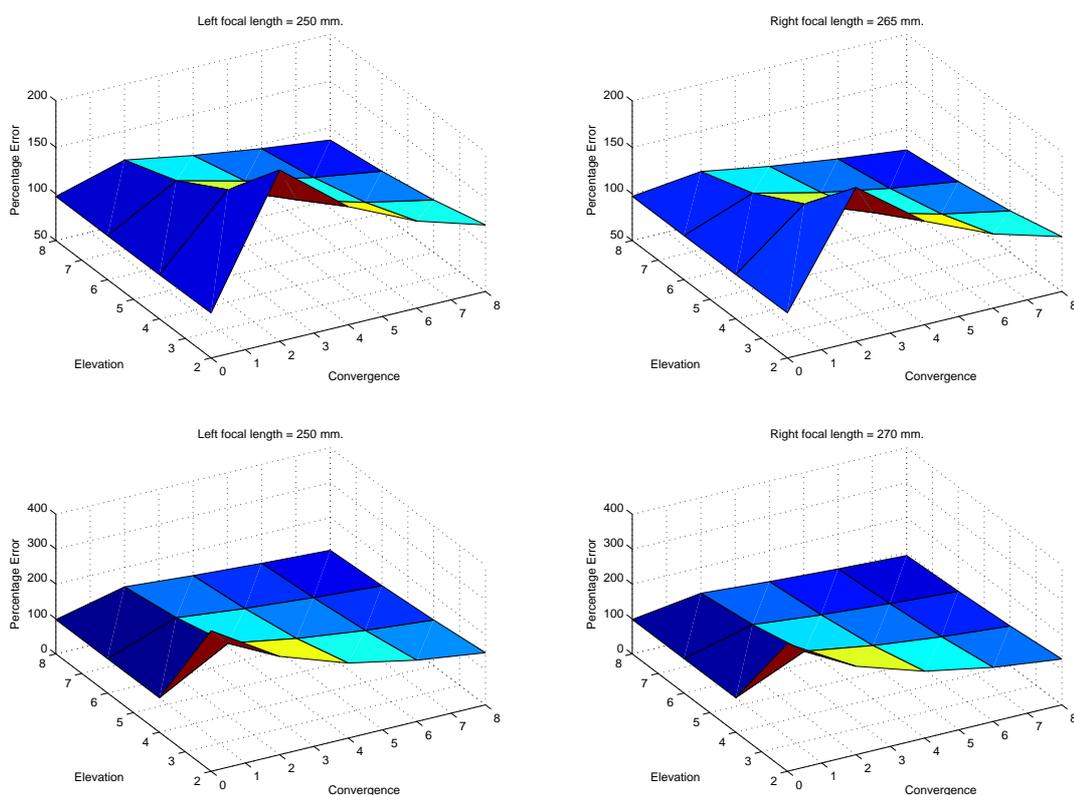
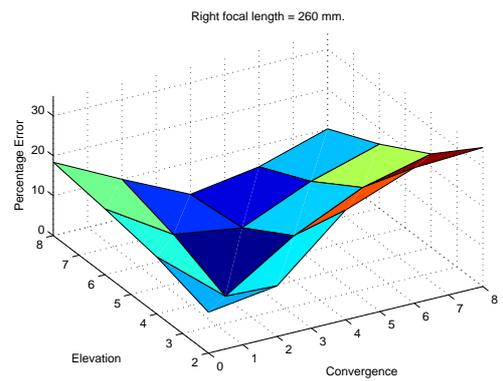
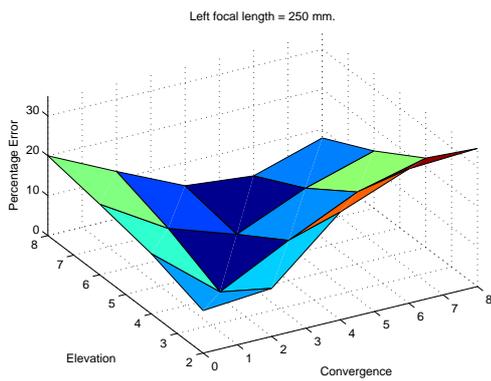
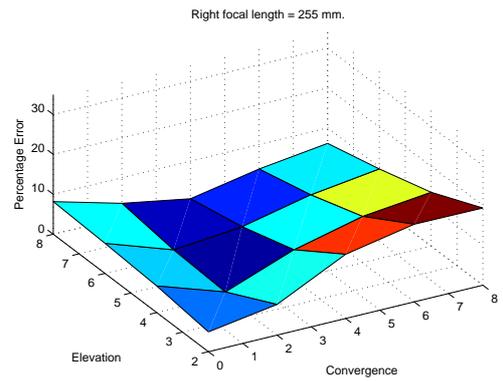
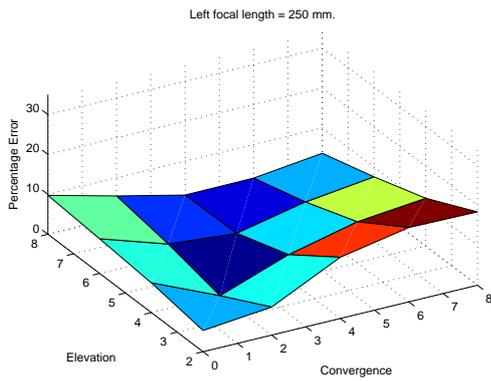
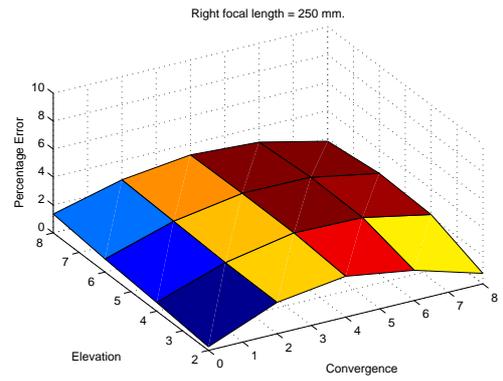
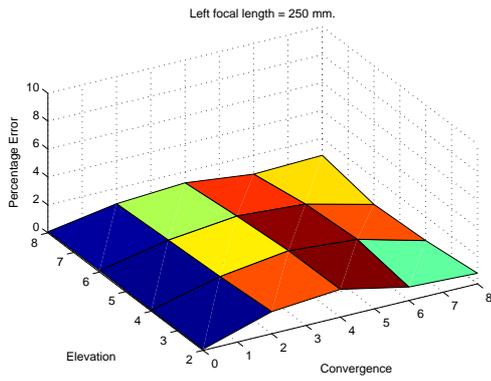


Figure A.2: Percentage errors in focal length calculation using Sturm's Method. (Left column) Left camera errors. (Right column) Right camera errors. First row represents an equal focal configuration, second to fifth rows represent unequal focal values between cameras. Elevation angles between 0° and 2° failed to produce a result (critical configuration), hence they are not displayed. Note that the percentage error scale changes among the graphs.



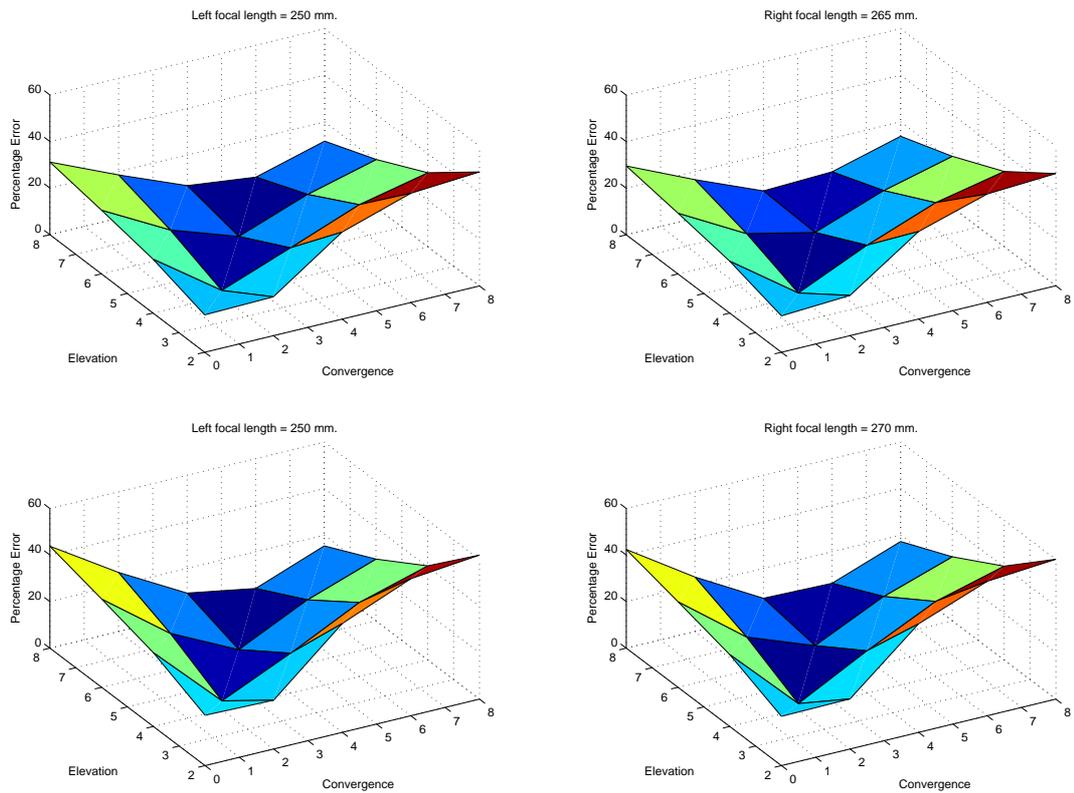


Figure A.3: Percentage errors in focal length calculation using Newsam's Method. (Left column) Left camera errors. (Right column) Right camera errors. First row represents an equal focal configuration, second to fifth rows represent unequal focal values between cameras. Elevation angles between 0° and 2° failed to produce a result (critical configuration), hence they are not displayed. Note that the percentage error scale changes among the graphs.

A.2 Focal length optimisation in a pair of real cameras

— Graphs

This section contains a full list of plots obtained in the assessment of the different methods for the optimisation of the focal length using a pair of real cameras connected to a surgical microscope. As described in Section 4.2.5, the techniques are:

Computation of the fundamental matrix

- Linear method
- M-estimators
- Gradient-based algorithm

Focal length estimation

- Bougnoux's method
- Sturm's method
- Newsam's method

Optimisation algorithms

- Levenberg-Marquard (LM)
- Self-adaptive Differential Evolution (SDE)
- CODEQ

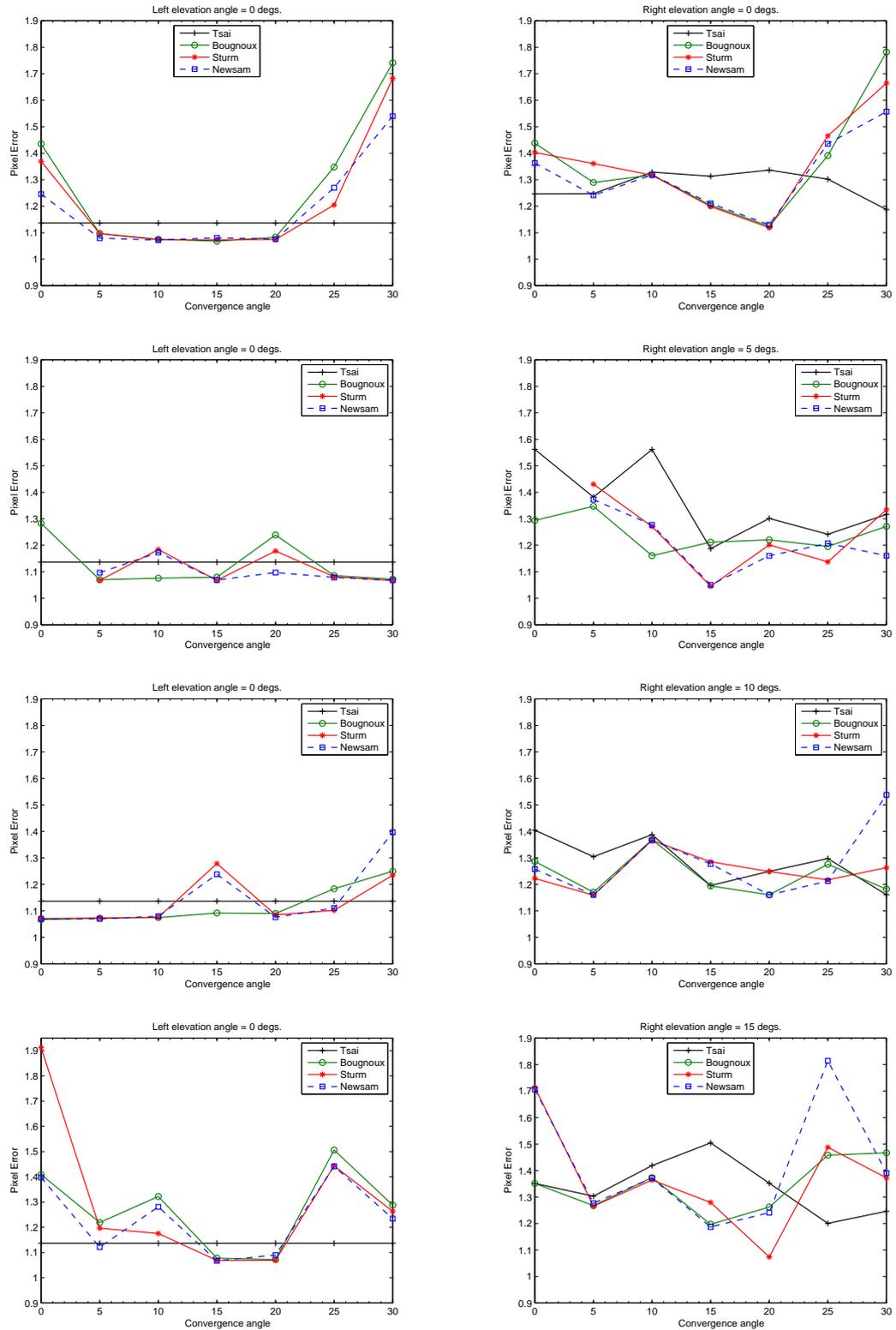


Figure A.4: Linear method for the fundamental matrix calculation and Levenberg-Marquardt optimisation algorithm.

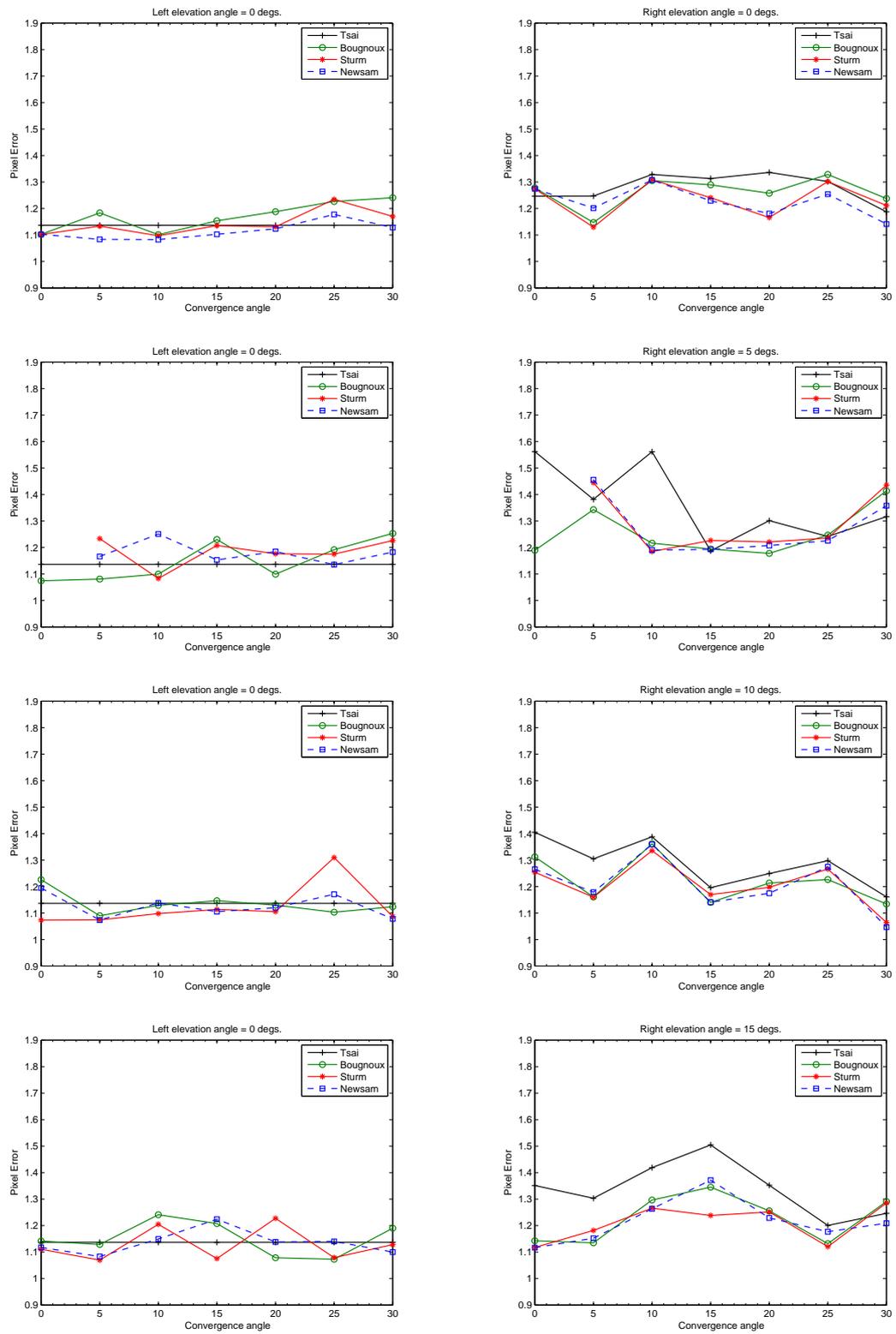


Figure A.5: Linear method for the fundamental matrix calculation and SDE optimisation.

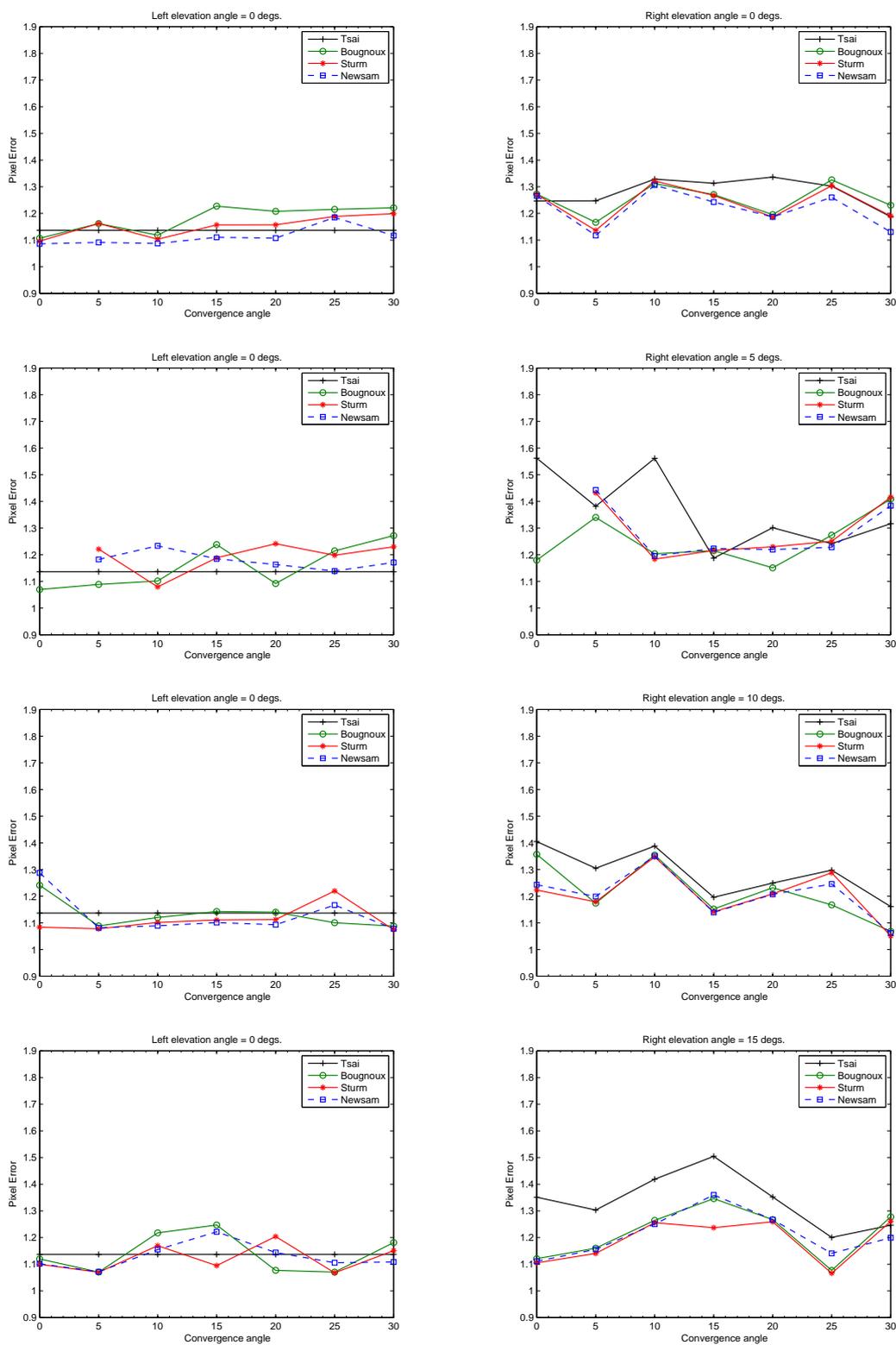


Figure A.6: Linear method for the fundamental matrix calculation and CODEQ optimisation.

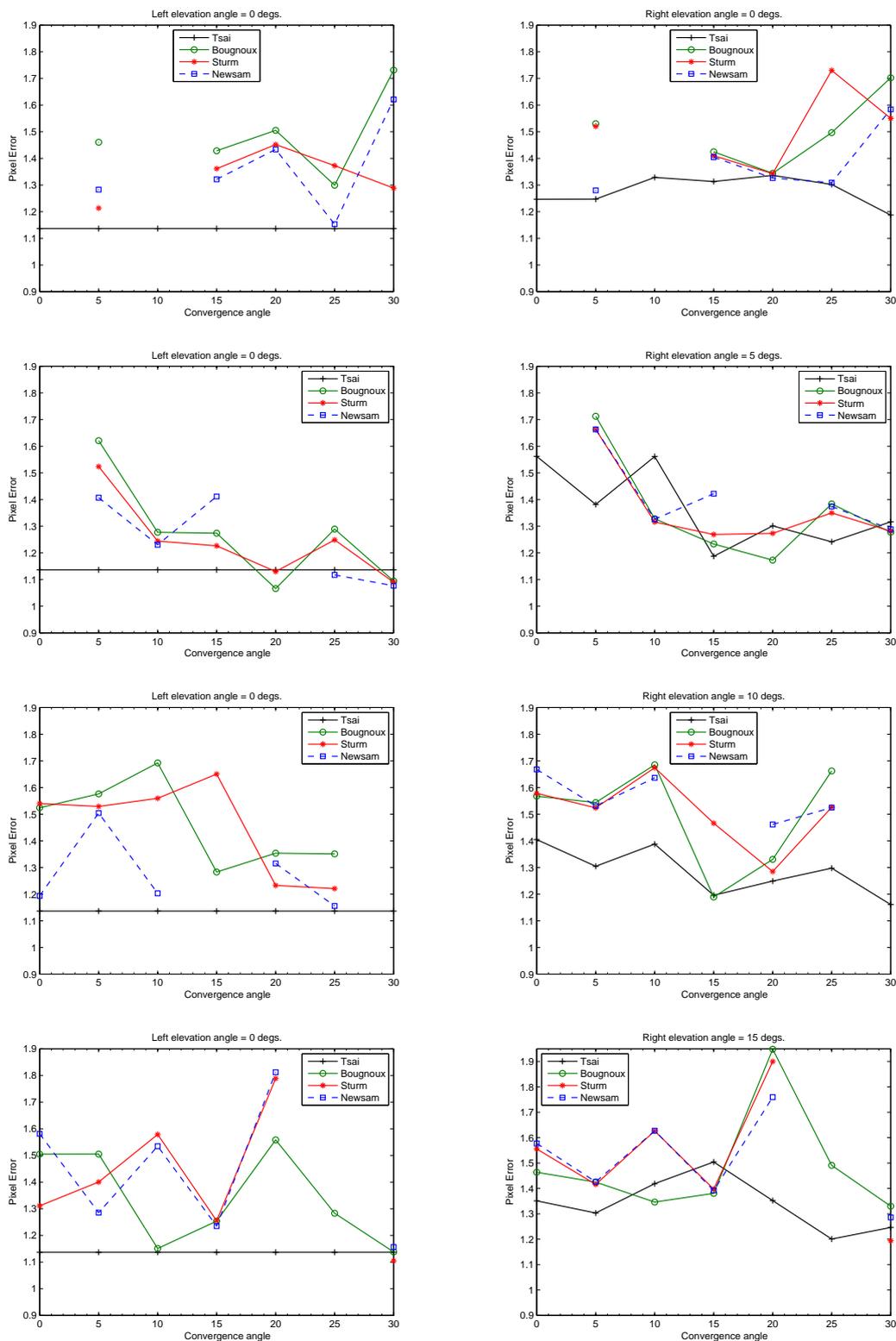


Figure A.7: Fundamental matrix calculation using M-estimators and Levenberg-Marquardt optimisation algorithm.

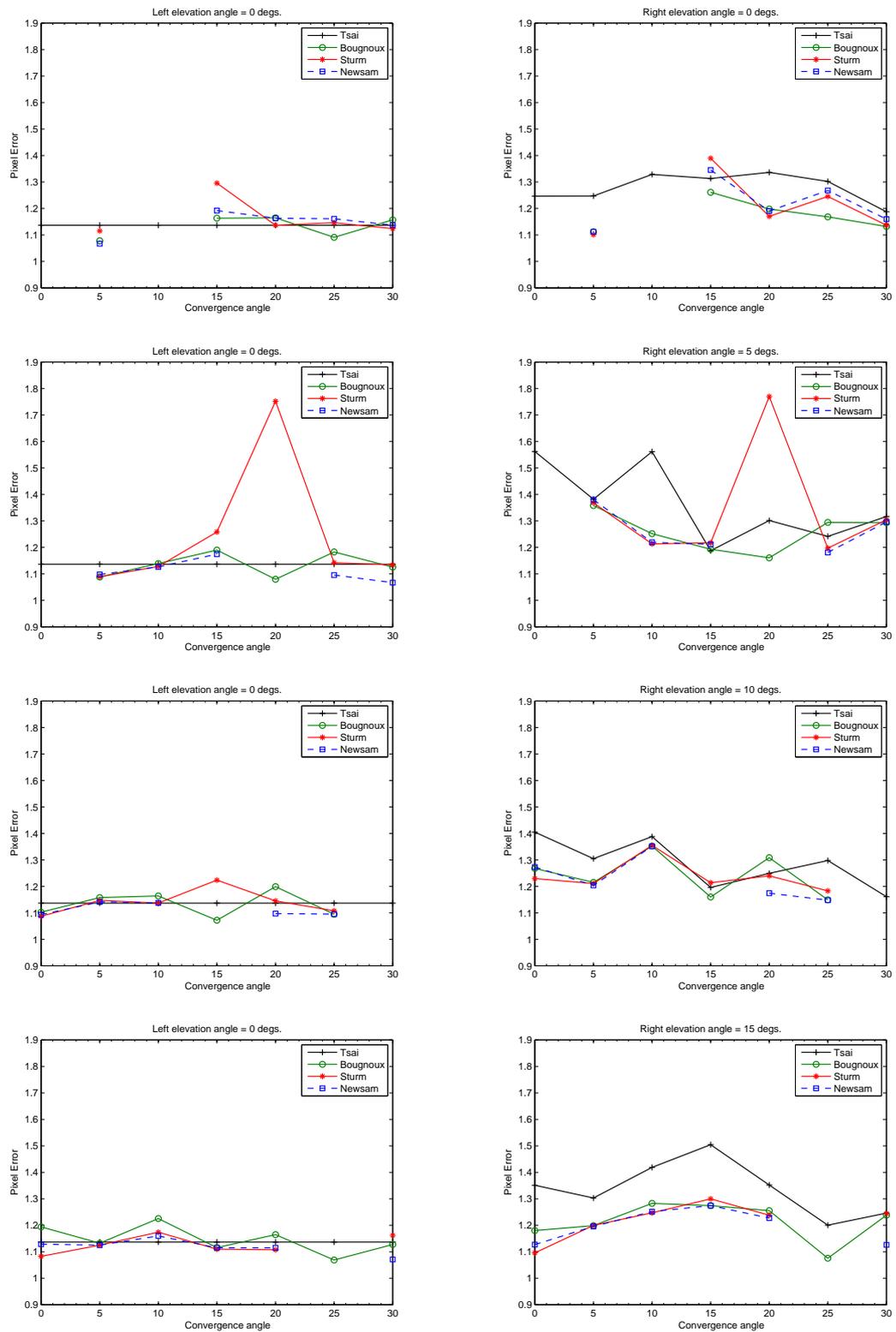


Figure A.8: Fundamental matrix calculation using M-estimators and SDE optimisation.

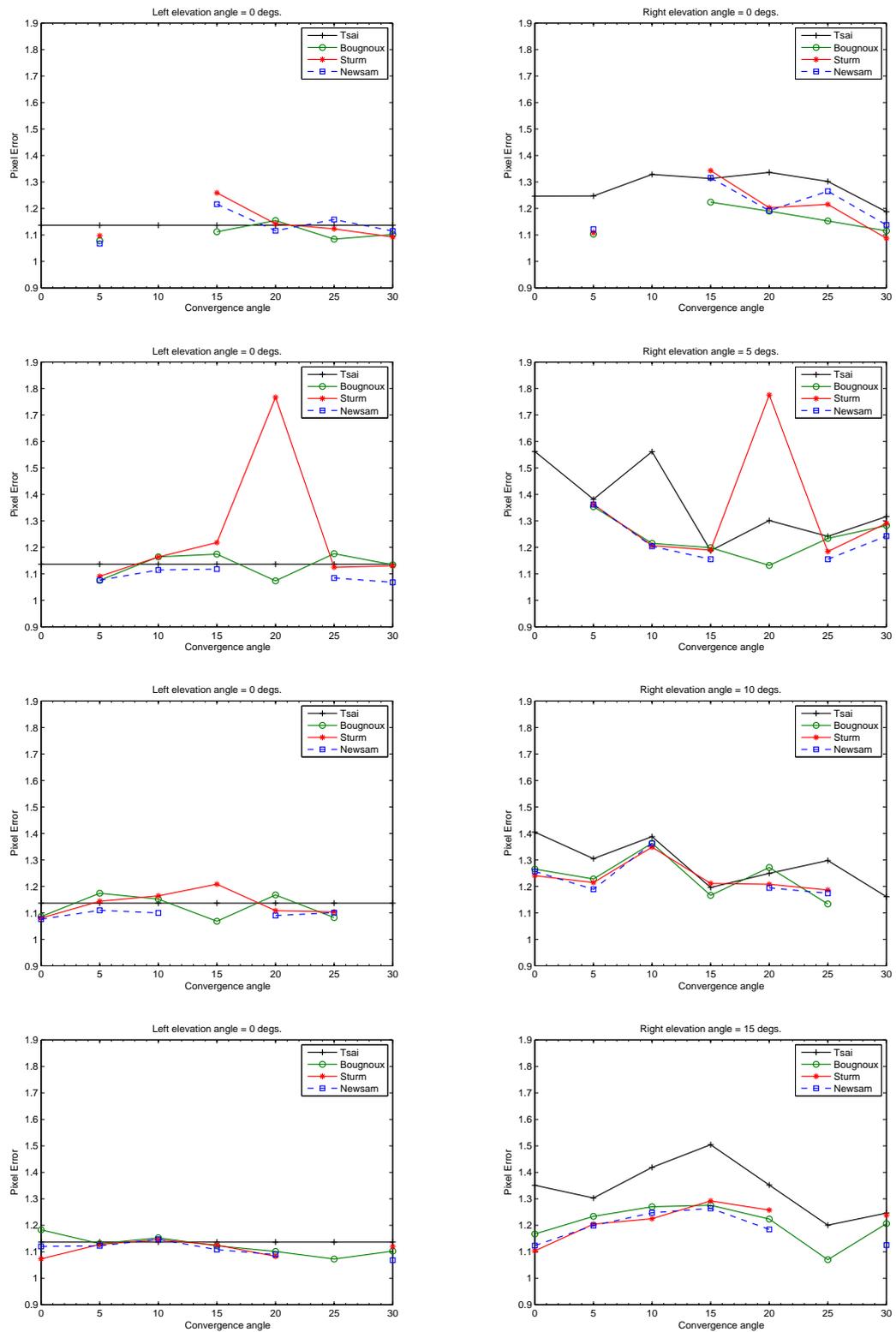


Figure A.9: Fundamental matrix calculation using M-estimators and CODEQ optimisation.

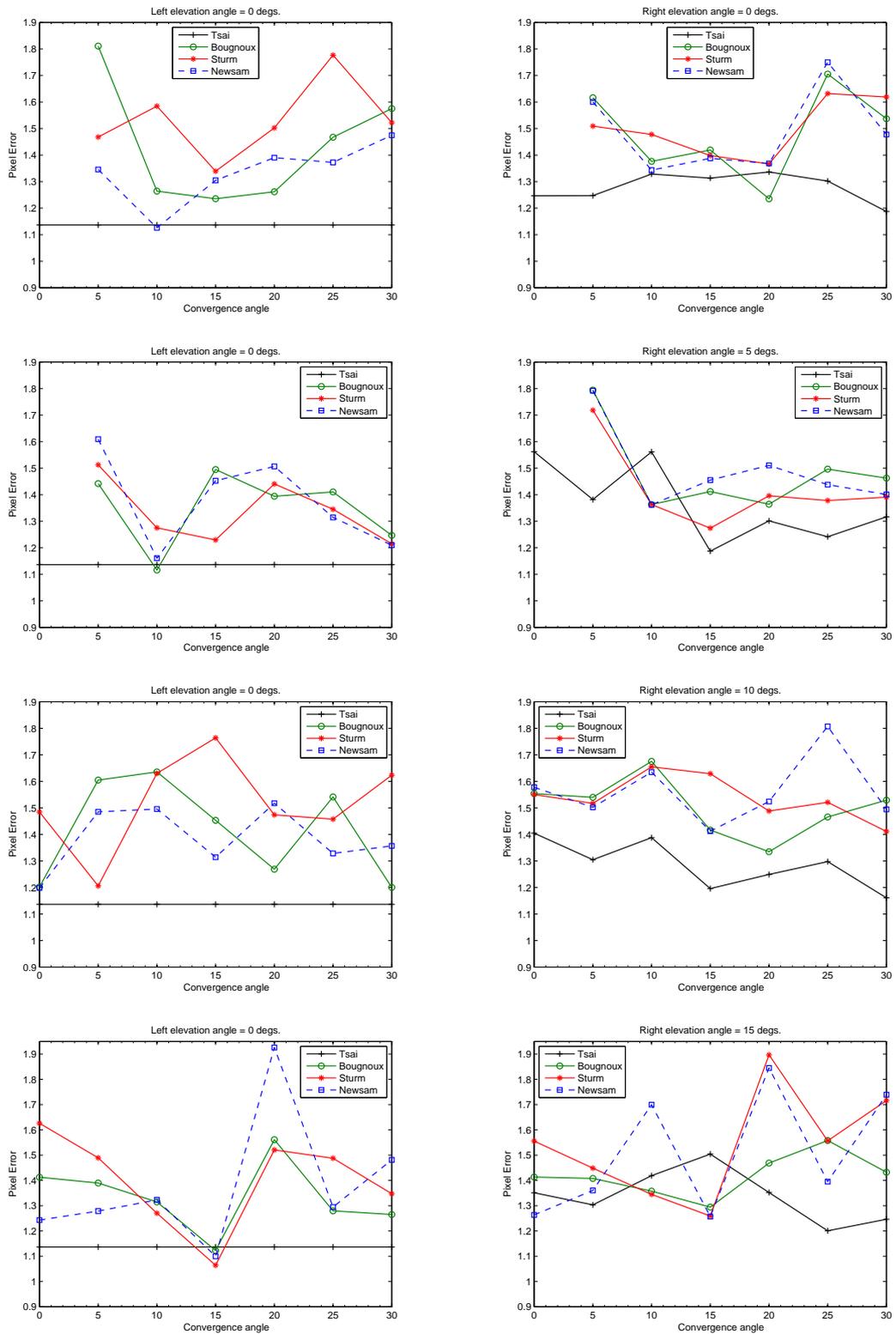


Figure A.10: Fundamental matrix calculation using gradient-based method and Levenberg-Marquardt optimisation algorithm.

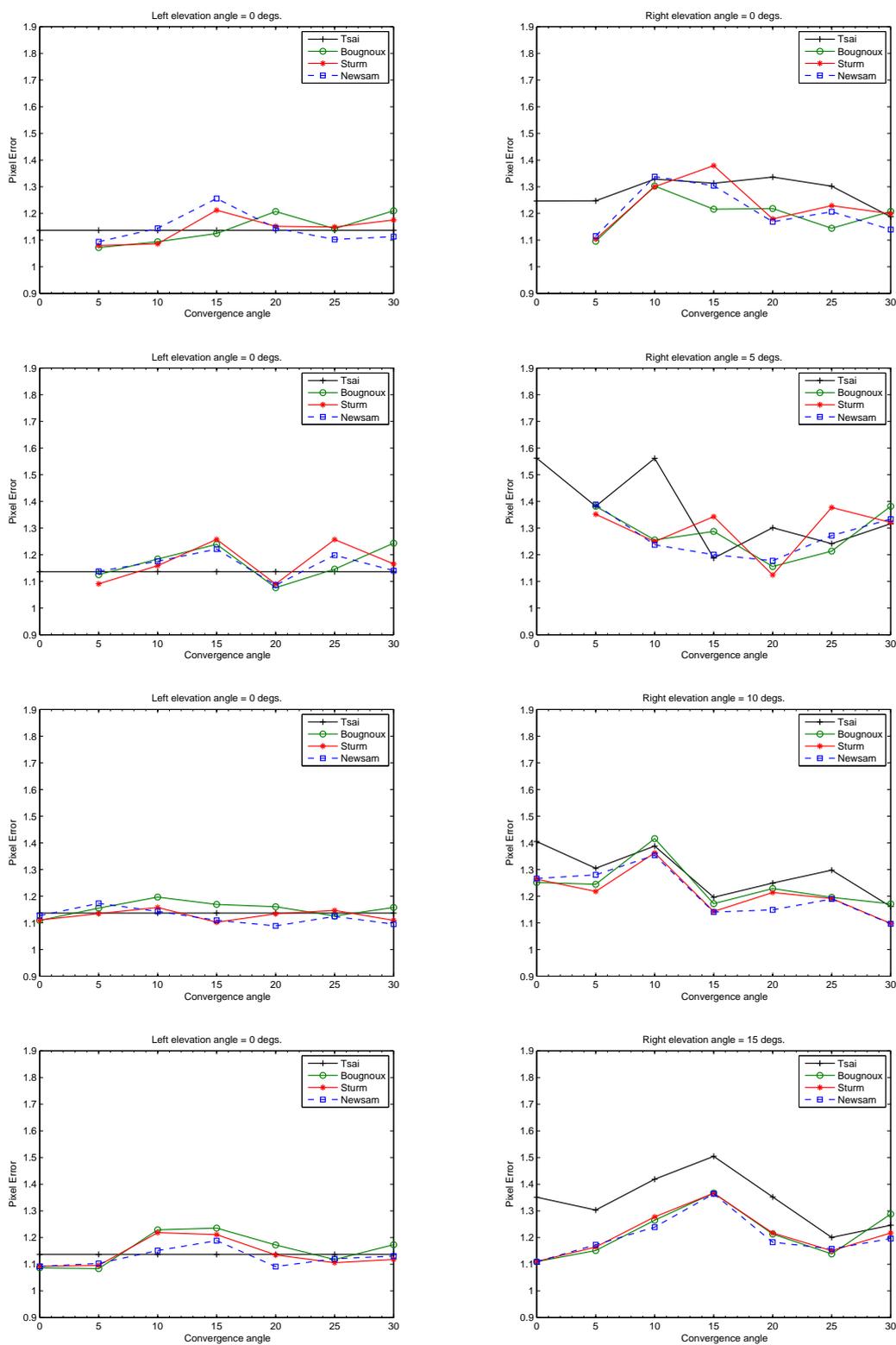


Figure A.11: Fundamental matrix calculation using gradient-based method and SDE optimisation.

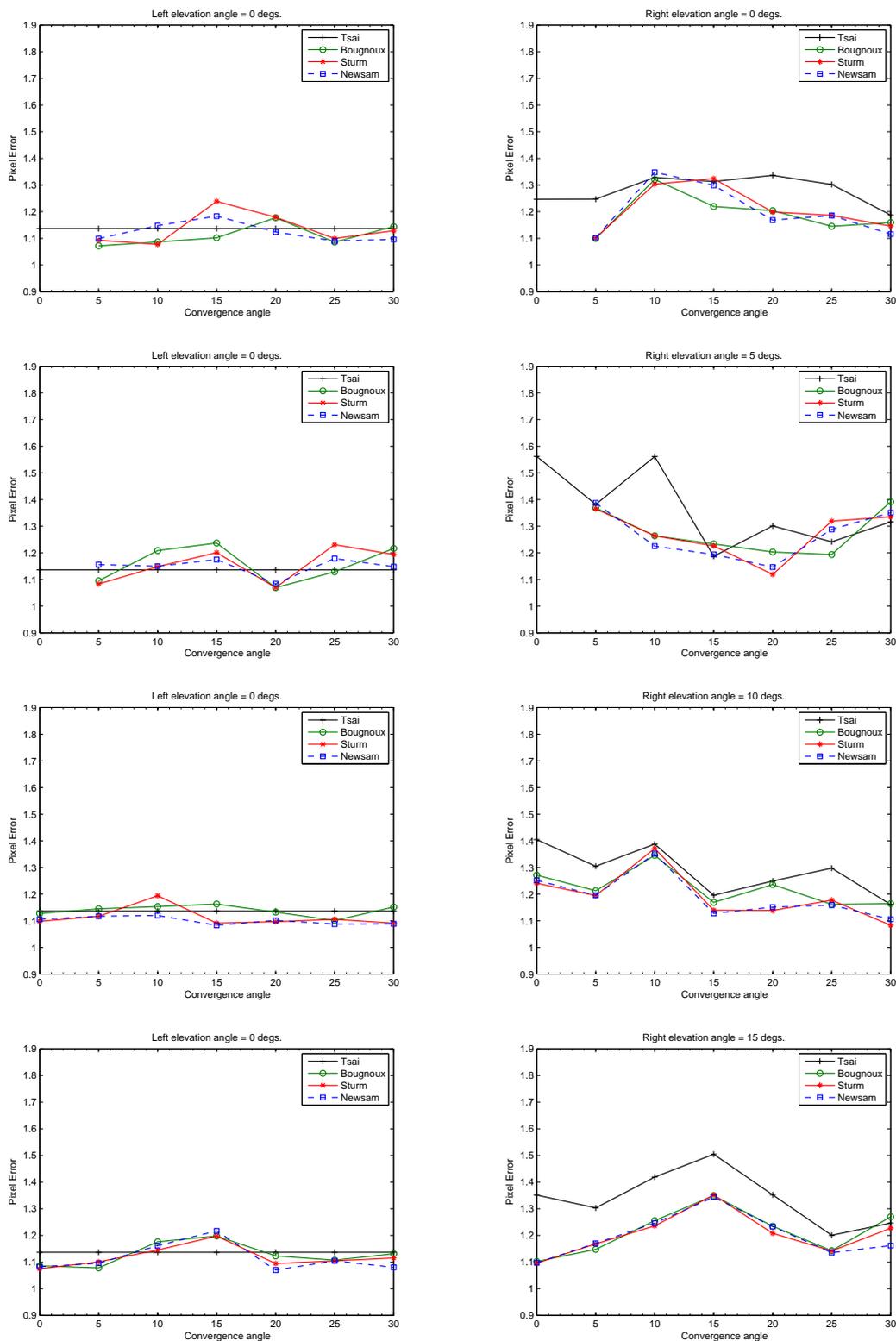


Figure A.12: Fundamental matrix calculation using gradient-based method and CODEQ optimisation.

A.3 Focal length optimisation in a pair of real cameras

— Tables

This section presents different tables corresponding to statistical significance tests of the possible combinations among optimisation algorithms, focal length techniques and fundamental matrix methods. Different convergence angles were evaluated using a pair of cameras connected to a surgical microscope. The null hypothesis is stated as the means between each pair of optimisation methods being similar and not providing any statistical significance of calibration error values.

The tables show the mean difference of calibration error values, the calculated t-value, statistical degrees of freedom (**df**) and significance level (**P-value**). The column **P<0.05** indicates that if the P-value is less than 0.05, the calibration error value is determined to be of statistical significance.

	Camera	Method	Mean difference	t-value	df	P-value	P<0.05
SDE vs LM	Left	Bougnoux	-0.007	-0.941	8	0.3744	N
		Sturm	0.147	9.762	14	<0.0001	Y
		Newsam	0.096	6.949	13	<0.0001	Y
	Right	Bougnoux	-0.276	-2.616	4	0.0590	N
		Sturm	-0.183	-1.917	4	0.1278	N
		Newsam	-0.031	-0.569	5	0.5943	N
SDE vs CODEQ	Left	Bougnoux	0.001	0.269	18	0.7910	N
		Sturm	0.023	1.093	18	0.2888	N
		Newsam	0.055	4.462	9	0.0016	Y
	Right	Bougnoux	-0.011	-0.975	19	0.3419	N
		Sturm	-0.008	-0.535	19	0.5988	N
		Newsam	-0.038	-2.243	10	0.0488	Y
CODEQ vs LM	Left	Bougnoux	-0.008	-1.177	6	0.2836	N
		Sturm	0.124	6.688	12	<0.0001	Y
		Newsam	0.041	6.138	4	0.0036	Y
	Right	Bougnoux	-0.264	-2.511	4	0.0660	N
		Sturm	-0.175	-1.837	4	0.1401	N
		Newsam	0.007	0.130	4	0.9029	N

Table A.1: Statistical significance of hypothesis that two optimisation methods display different calibration error values using the linear method for fundamental matrix at 20° convergence angle (bad accuracy).

	Camera	Method	Mean difference	t-value	df	P-value	P<0.05
SDE vs LM	Left	Bougnoux	-0.351	-7.070	4	0.0021	Y
		Sturm	-0.299	-7.596	4	0.0016	Y
		Newsam	-0.419	-5.968	5	0.0019	Y
	Right	Bougnoux	-0.372	-13.040	8	<0.0001	Y
		Sturm	-0.461	-10.719	6	<0.0001	Y
		Newsam	-0.860	-8.880	4	0.0009	Y
SDE vs CODEQ	Left	Bougnoux	0.002	0.752	18	0.4616	N
		Sturm	0.011	3.099	9	0.0127	Y
		Newsam	0.035	1.670	11	0.1231	N
	Right	Bougnoux	0.055	2.310	19	0.0323	Y
		Sturm	0.054	2.171	19	0.0428	Y
		Newsam	0.036	1.400	12	0.1867	N
CODEQ vs LM	Left	Bougnoux	-0.353	-7.121	4	0.0021	Y
		Sturm	-0.311	-7.917	4	0.0014	Y
		Newsam	-0.454	-6.698	4	0.0026	Y
	Right	Bougnoux	-0.427	-14.064	10	<0.0001	Y
		Sturm	-0.515	-11.938	6	<0.0001	Y
		Newsam	-0.896	-9.479	4	0.0007	Y

Table A.2: Statistical significance of hypothesis that two optimisation methods display different calibration error values using the linear method for fundamental matrix at 25° convergence angle (good accuracy).

	Camera	Method	Mean difference	t-value	df	P-value	P<0.05
SDE vs LM	Left	Bougnoux	-0.271	-6.503	4	0.0029	Y
		Sturm	-0.170	-5.901	4	0.0041	Y
		Newsam	-0.133	-8.527	12	<0.0001	Y
	Right	Bougnoux	-0.238	-24.904	14	<0.0001	Y
		Sturm	-0.189	-20.832	8	<0.0001	Y
		Newsam	-0.232	-14.934	11	<0.0001	Y
SDE vs CODEQ	Left	Bougnoux	0.002	0.194	19	0.8480	N
		Sturm	-0.003	-0.286	17	0.7784	N
		Newsam	0.002	0.178	17	0.8606	N
	Right	Bougnoux	-0.035	-0.871	9	0.4062	N
		Sturm	-0.005	-0.579	17	0.5704	N
		Newsam	-0.003	-0.245	18	0.8093	N
CODEQ vs LM	Left	Bougnoux	-0.272	-6.558	4	0.0028	Y
		Sturm	-0.167	-5.697	4	0.0047	Y
		Newsam	-0.135	-9.895	9	<0.0001	Y
	Right	Bougnoux	-0.203	-5.063	9	0.0007	Y
		Sturm	-0.184	-17.566	12	<0.0001	Y
		Newsam	-0.229	-16.286	8	<0.0001	Y

Table A.3: Statistical significance of hypothesis that two optimisation methods display different calibration error values using M-estimators at 5° convergence angle (good accuracy).

	Camera	Method	Mean difference	t-value	df	P-value	P<0.05
SDE vs LM	Left	Bougnoux	-0.119	-7.223	6	0.0004	Y
		Sturm	-0.192	-3.655	4	0.0217	Y
		Newsam	-0.356	-2.158	4	0.0971	N
	Right	Bougnoux	-0.253	-5.121	4	0.0069	Y
		Sturm	-0.228	-4.153	11	0.0016	Y
		Newsam	-0.214	-3.830	4	0.0186	Y
SDE vs CODEQ	Left	Bougnoux	-0.007	-0.517	16	0.6124	N
		Sturm	-0.015	-1.020	18	0.3213	N
		Newsam	0.007	0.830	18	0.4172	N
	Right	Bougnoux	-0.002	-0.094	19	0.9258	N
		Sturm	0.008	0.205	12	0.8411	N
		Newsam	0.010	0.726	19	0.4769	N
CODEQ vs LM	Left	Bougnoux	-0.112	-6.019	10	0.0001	Y
		Sturm	-0.177	-3.332	4	0.0291	Y
		Newsam	-0.363	-2.201	4	0.0926	N
	Right	Bougnoux	-0.252	-5.063	4	0.0072	Y
		Sturm	-0.236	-5.323	5	0.0031	Y
		Newsam	-0.225	-3.995	4	0.0162	Y

Table A.4: Statistical significance of hypothesis that two optimisation methods display different calibration error values using M-estimators at 15° convergence angle (bad accuracy).

	Camera	Method	Mean difference	t-value	df	P-value	P<0.05
SDE vs LM	Left	Bougnoux	0.119	8.343	12	<0.0001	Y
		Sturm	0.106	5.490	14	<0.0001	Y
		Newsam	0.077	6.262	12	<0.0001	Y
	Right	Bougnoux	-0.222	-1.953	4	0.1226	N
		Sturm	-0.192	-1.835	4	0.1404	N
		Newsam	-0.114	-0.914	4	0.4124	N
SDE vs CODEQ	Left	Bougnoux	0.038	1.514	16	0.1494	N
		Sturm	0.013	0.623	19	0.5408	N
		Newsam	-0.029	-1.255	15	0.2288	N
	Right	Bougnoux	0.018	1.607	19	0.1245	N
		Sturm	0.015	1.230	18	0.2346	N
		Newsam	0.021	1.256	11	0.2353	N
CODEQ vs LM	Left	Bougnoux	0.082	3.730	10	0.0039	Y
		Sturm	0.092	5.532	14	<0.0001	Y
		Newsam	0.106	5.222	10	0.0004	Y
	Right	Bougnoux	-0.240	-2.112	4	0.1023	N
		Sturm	-0.206	-1.977	4	0.1192	N
		Newsam	-0.135	-1.087	4	0.3381	N

Table A.5: Statistical significance of hypothesis that two optimisation methods display different calibration error values using the gradient-based method at 15° convergence angle (bad accuracy).

	Camera	Method	Mean difference	t-value	df	P-value	P<0.05
SDE vs LM	Left	Bougnoux	-0.303	-6.856	4	0.0024	Y
		Sturm	-0.343	-6.695	4	0.0026	Y
		Newsam	-0.100	-4.036	5	0.0100	Y
	Right	Bougnoux	-0.470	-15.940	5	<0.0001	Y
		Sturm	-0.482	-11.058	5	0.0001	Y
		Newsam	-0.255	-17.912	14	<0.0001	Y
SDE vs CODEQ	Left	Bougnoux	0.011	0.970	14	0.3487	N
		Sturm	0.001	0.127	18	0.9005	N
		Newsam	0.016	1.554	16	0.1398	N
	Right	Bougnoux	-0.004	-0.265	19	0.7940	N
		Sturm	0.009	0.485	19	0.6334	N
		Newsam	0.022	1.385	19	0.1821	N
CODEQ vs LM	Left	Bougnoux	-0.313	-7.226	4	0.0019	Y
		Sturm	-0.345	-6.748	4	0.0025	Y
		Newsam	-0.116	-4.894	4	0.0081	Y
	Right	Bougnoux	-0.466	-15.553	6	<0.0001	Y
		Sturm	-0.491	-11.469	4	0.0003	Y
		Newsam	-0.277	-20.235	14	<0.0001	Y

Table A.6: Statistical significance of hypothesis that two optimisation methods display different calibration error values using the gradient-based method at 25° convergence angle (good accuracy).

Appendix B

Optimisation Algorithms

This appendix presents two optimisation algorithms for the numerical minimisation of multidimensional functions used throughout the research, Self-adaptive Differential Evolution (SDE) and CODEQ. The reader is briefly introduced to the theory behind these optimisation methods, followed by a validation of their performance; which was considered essential for their application within an IESN system.

B.1 Evolutionary algorithms

Evolutionary algorithms (EA) are stochastic search methods that are inspired by principles of biological processes of natural evolution. In particular, they simulate the evolution of organisms through the selection and perturbation of internal structures known as individuals. EAs initialise a population of individuals with random values at a first generation or iteration, where each individual corresponds to a potential solution. A principle of survival during the evolution is achieved by comparing the quality of each individual using a fitness criterion and selecting the best solution within the population. Selected individuals are altered by applying small changes in their parameters (mutation) and recombining them with other individuals (crossover). The procedure is repeated until the best value is

found or a specific computational limit is reached. An advantage of EAs in comparison to other search methods is that they are able to avoid stagnation in local minima and find the global optimum solution.

B.1.1 Differential Evolution

Differential evolution (DE) is a population-based EA introduced by Storn and Price [156], which differs from other EA strategies due to the use of direction and distance information in the population. Such information is used to guide the search towards the global optimum. Specifically, DE involves vectors as a representation of individuals. Their perturbation is obtained by using arithmetic vector operations instead of logical combinations. The evolution operators involved in the search of a global optimum in DE are mutation, crossover and selection.

- *Mutation.* Two randomly selected vectors or individuals (x_1 and x_2) are arithmetically subtracted and their difference is weighted or adjusted based on a scale factor F . The result is added to a third random vector x_3 , which produces a trial vector v . The random selection must ensure that the chosen vectors are different to each other. Equation B.1 exemplifies the mutation step:

$$v = x_3 + F(x_1 - x_2). \quad (\text{B.1})$$

- *Crossover.* This recombination stage involves mixing the elements j of a parent vector x_p with elements from the trial vector v in order to produce a child vector, where $j = 1 \dots \text{number-of-dimensions}$. Crossover depends on a comparison of

parameters according to the following criterion:

$$u_j = \begin{cases} v_j & \text{if } rand(0,1) \leq CR \text{ or } j = r \\ x_{p,j} & \text{otherwise} \end{cases}$$

in which CR represents a user-defined variable corresponding to the probability of reproduction in the interval $[0,1]$, which controls the parameters that will be transferred to the child vector u . If the uniformly distributed random number $rand(0,1)$ obtained at the current generation is less or equal than CR , the trial element will be inherited to u . Additionally, a random index variable $r = rand(1, \text{number-of-dimensions})$ is compared to the current dimension index j within the chosen vector. If both indices are the same, the recombination is performed at the specified dimension. In case that none of the conditions is true, the trial vector will obtain the original element from the parent vector.

- *Selection.* The procedure to select the best fitted vector requires comparing the child and parent vectors in the cost function. If the child vector produces a lower or equal solution than the parent, it will replace the parent's position in the population; otherwise the parent vector is retained.

Figure B.1 provides a schematic representation of the steps performed in DE.

DE has been successfully applied to the registration of monomodal 3D/3D MRI imagery [158] and other medical applications. However, one of the main drawbacks of this technique relies on correctly setting the initial control parameters for each particular problem, e.g. Price et al. [159] suggest ten different approaches depending on the problem features. Moreover, a wrong choice of initial parameters can affect the overall performance of the algorithm. The following two strategies aim at solving this issue by automatically adapting the function parameters at run time.

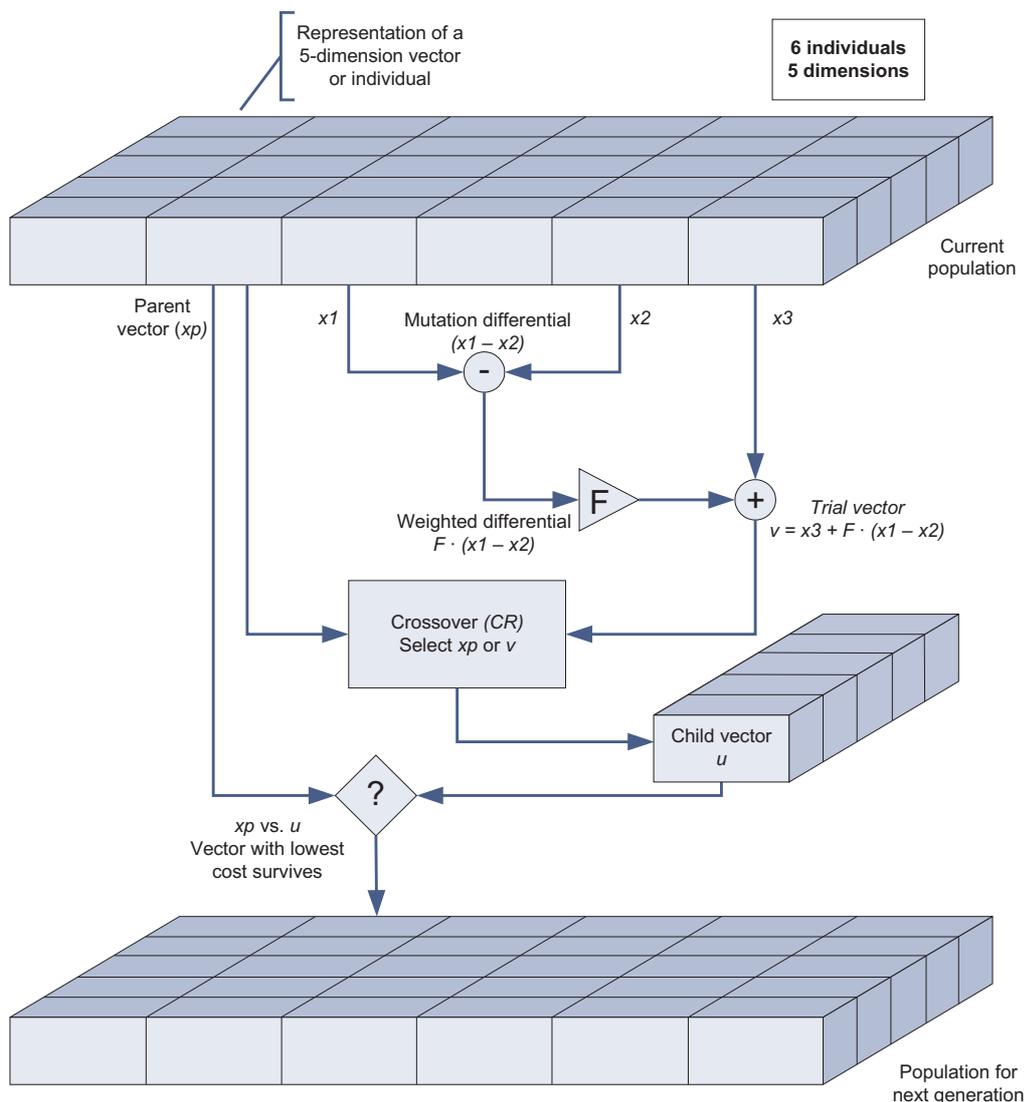


Figure B.1: Schematic diagram of the DE algorithm (Image adapted from [157]).

Self-Adaptive Differential Evolution (SDE)

Salman et al. [160] proposed a self-adaptive algorithm that dynamically adjusts the control parameters in DE which directly affect the behaviour of the optimisation search. The parameters comprise the mutation scale factor F and the probability of reproduction CR . This self-adaptation improves the performance of DE by exploiting a wider search in the function shape and avoiding stagnation in the local minima. In addition, it increases the diversity of the population and prevents premature convergence.

SDE adjusts the mutation parameter F through a selection of random values obtained from a normal distribution $N(0, 1)$ at every iteration loop t within the generation size. Similarly, the crossover operator CR is modified by a stochastic selection based on a normal distribution but with stricter bounds. In SDE, Equation B.1 referring to the mutation step is modified as follows:

$$v_i(t) = x_{i_3}(t) + F_i(t)(x_{i_1}(t) - x_{i_2}(t)), \quad (\text{B.2})$$

where

$$F_i(t) = F_{i_4}(t) + N(0, 1) \times (F_{i_5}(t) - F_{i_6}(t)), \quad (\text{B.3})$$

in which i represents a vector in the current generation t . The vectors $i_1 \dots i_6$ are randomly selected using a uniform distribution within the population size, and $i_1 \neq i_2 \neq i_3 \neq i_4 \neq i_5 \neq i_6$.

CODEQ

CODEQ [161] is an approach that includes concepts from chaotic search, opposition-based learning, DE and quantum mechanics. Its main advantage resides in that it is a completely parameter-free method (except for the population size). CODEQ resolves some issues found in DE during the recombination of elements in the population, which may affect the optimisation search.

The selection of the trial vector is obtained by excluding the mutation control parameter F found in DE. Instead, a principle based on quantum mechanics is used to alter the probability of mutation, and the result is added to the parent vector. Also, the crossover operation varies from the original procedure as rules from chaotic search and opposition-based learning are applied. The new crossover performs the search in two different places of the function shape at the same time for each iteration; one based on a randomly se-

lected position and the other at its corresponding opposite location in the search space. The pseudocode of CODEQ is presented in Code B.1.

The population of individuals, chaotic variable c , and probability value p are randomly initialised.

For each loop t in the generation size

For each loop i in the population size

$v_i(t) = x_{p,i}(t) + (x_{i_1}(t) - x_{i_2}(t)) \ln(\frac{1}{u})$,
 where u is randomly selected using a uniform distribution $U(0, 1)$.

If $f(v_i(t)) \leq f(x_{p,i}(t))$
 $x_{p,i}(t) = v_i(t)$
 End if
 where $f(\mathbf{x})$ refers to the evaluated cost function.

End for

 Find the best and worst vectors in the population: $x_b(t)$ and $x_s(t)$.

If $rand \leq 0.5$
 $w(t) = LB + UB - r \times x_s(t)$,
 where $w(t)$ is a fitness vector, LB and UB are the function's lower and upper bounds, respectively; and r is randomly selected using a uniform distribution $U(0, 1)$.

Else

$$c(t) = \begin{cases} c(t-1)/p & c(t-1) \in (0, p) \\ (1 - c(t-1))/(1-p) & c(t-1) \in [p, 1) \end{cases}$$
 $w(t) = x_b(t) + |x_{i_1}(t) - x_{i_2}(t)| \times (2c(t) - 1)$

End if

If $f(w(t)) \leq f(x_s(t))$
 $x_s(t) = w(t)$
 End if

End for

Code B.1: Pseudocode of the CODEQ algorithm [161].

B.1.2 Use of the optimisation algorithms

A contributing factor to the popularity of self-adaptive optimisation methods such as SDE and CODEQ relies on their simplicity of use. This is because the algorithms only require a few control parameters to start the search of the global optima, as described earlier. The rest of the parameters involved in the procedure are randomly initialised the first time that the optimisation function is called and dynamically adjusted during the execution of the program.

In order to start the optimisation, the user must provide the population size that reflects the universe of candidate members available throughout the search. Additionally, the user needs to specify the maximum number of generations or iterations that will be carried out during the execution of the algorithm. Both population size and number of generations are given as integer values and must be selected according to the requirements of the optimisation.

An array structure is used to assign the variables of the problem as input values in the optimisation function. The size of the array corresponds to the dimensionality of the problem, where each dimension is related to an independent parameter or DOF. This array serves as the initial vector upon which the search will be based (basis vector). In particular, the candidate vectors generated during the procedure will be combined with this basis vector in order to modify the initial input values. This enforces the search to be within a certain scope and reduce the possibility for the solution to diverge. In regard to the optimisation procedures carried out in this research, the elements of the basis vector correspond to the initial pair of focal lengths estimated by the self-calibration methods, whereas the candidate vectors represent a set of variations in focal length. For the intra-operative registration, the basis vector contains the starting pose of the virtual model to be registered and the candidate vectors comprise the changes in translation and rotation required to find the correct registration position.

Finally, a real number that corresponds to a tolerance threshold must be also specified to control the maximum acceptable error found by the optimisation algorithm. For this purpose, an appropriate cost function evaluates the different locations in the search space that, combined with the basis vector, produce the best solution. In this respect, the cost function is automatically called at every iteration loop and executed as many times as necessary until the threshold is reached or the maximum number of iterations is exceeded. The final value returned by the optimisation algorithm represents the global optima.

B.2 Experiments and results

B.2.1 Validation of the optimisation algorithms

An evaluation of the two self-adaptive EA methods described in the previous section, i.e. SDE and CODEQ, was performed as it was considered important to validate the optimisation algorithms before their implementation in the IESN system. Additionally, a classic DE method (known as DE/ran/1/bin [159]) was included in the evaluation to examine the differences against the two self-adaptive techniques. The procedure involved a comparison between the results presented in the original SDE and CODEQ literature to the author's own implementation in the IESN system. Four benchmark problems were selected among unimodal and multimodal functions used in the field to test the performance of global optimisation algorithms. The selected functions are:

- **Step** function:

$$f(x) = \sum_{i=1}^N (\lfloor x_i + 0.5 \rfloor)^2,$$

where N is the population size, global minimum $x^* = 0$, and $f(x^*) = 0$ for

$$-100 \leq x_i \leq 100.$$

- **Rosenbrock function:**

$$f(x) = \sum_{i=1}^{N-1} (100(x_i - x_{i-1}^2)^2 + (x_{i-1} - 1)^2),$$

where $x^* = (1, 1, \dots, 1)$, and $f(x^*) = 0$ for $-30 \leq x_i \leq 30$.

- **Rastrigin function:**

$$f(x) = \sum_{i=1}^N (x_i^2 - 10 \cos(2\pi x_i) + 10),$$

where $x^* = 0$, and $f(x^*) = 0$ for $-5.12 \leq x_i \leq 5.12$.

- **Griewank function:**

$$f(x) = \frac{1}{4000} \sum_{i=1}^N x_i^2 - \prod_{i=1}^N \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1,$$

where $x^* = 0$, and $f(x^*) = 0$ for $-600 \leq x_i \leq 600$.

The optimisation algorithms were implemented in the programming language C++ using an Intel Core2 Quad processor computer at 2.4Ghz and 2GB RAM. The control conditions specified were the same as in the literature, using a population of 50 individuals, 30 dimensions and a maximum of 50,000 evaluations of the objective function. In the case of DE, the mutation and crossover parameters were $F = 0.5$ and $CR = 0.9$, respectively. The error threshold defining a solution was 10^{-6} for all objective functions. The lower the values for the number of evaluations, the faster the algorithm performed. A value of 50,000 denotes that the function could not find the global optima. Table B.1 shows the original results reported in the literature [160, 162] (Literature). Also, the results of the own implementation are presented (Author). Error values of less than 10^{-5} are rounded to 0 for clarity.

Although the number of evaluations for SDE can not be obtained directly from the literature, the error values show that SDE produces equal or better results than DE in most functions. Moreover, from the information found in the original publications it can be observed that CODEQ finds a solution in less number of evaluations and is more accurate

		Step	Rosenbrock	Rastrigin	Griewank
DE-Literature	<i>Evals</i>	15368(\pm 1790)	50000(\pm 0)	50000(\pm 0)	38262(\pm 6082)
	<i>Error</i>	0(\pm 0)	26.075(\pm 1.364)	157.34(\pm 19.90)	0.0022(\pm 0.0048)
DE-Author	<i>Evals</i>	278(\pm 16)	50000(\pm 0)	50000(\pm 0)	15386(\pm 23886)
	<i>Error</i>	0(\pm 0)	2.046(\pm 1.680)	13.903(\pm 4.457)	0.0025(\pm 0.0040)
SDE-Literature	<i>Evals</i>	NA ^a	NA	NA	NA
	<i>Error</i>	0(\pm 0)	52.180(\pm 28.143)	5.743(\pm 2.338)	0(\pm 0)
SDE-Author	<i>Evals</i>	436(\pm 23)	50000(\pm 0)	50000(\pm 0)	908(\pm 33)
	<i>Error</i>	0(\pm 0)	1.249(\pm 0.0)	30.040(\pm 7.351)	0(\pm 0)
CODEQ-Literature	<i>Evals</i>	5833(\pm 2773)	50000(\pm 0)	22247(\pm 2305)	20748(\pm 2362)
	<i>Error</i>	0(\pm 0)	26.196(\pm 0.649)	0(\pm 0)	0(\pm 0)
CODEQ-Author	<i>Evals</i>	76(\pm 34)	9998(\pm 762)	286(\pm 68)	283(\pm 48)
	<i>Error</i>	0(\pm 0)	0(\pm 0)	0(\pm 0)	0(\pm 0)

^aNA means that the number of evaluations was not described in the literature.

Table B.1: Comparison between mean and standard deviation (\pm SD) of the function optimisation results reported in the literature [160, 162] (Literature) and the author's own implementation (Author), for each of the three algorithms, i.e. DE, SDE and CODEQ.

than DE and SDE. On the other hand, the author's implementation performs significantly better for each method than the implementation results described in the literature. Although there is no clear reason for such improvement, it might rely on the fact that the original results were implemented in Matlab, whereas the author's algorithms were programmed in C++. Still, CODEQ outperforms both DE and SDE on all functions tested.

Bibliography

- [1] International Society for Computer Aided Surgery. Corporate website. <http://www.iscas.net>. Retrieved on 8 March 2010.
- [2] L. Joskowicz and R.H. Taylor. Computers in imaging and guided surgery. *IEEE Computers in Science and Engineering*, 3(5):65 – 72, Sep/Oct 2001.
- [3] T. Capin, I. Pandzic, N. Magnenat-Thalmann, and D. Thalmann. *Avatars in Networked Virtual Environments*. John Wiley & Sons, 1999.
- [4] F.P. Vidal, F. Bello, K.W. Brodlie, N.W. John, D. Gould, R. Phillips, and N.J. Avis. Principles and applications of computer graphics in medicine. *Computer Graphics Forum*, 25(1):113 – 137, 2006.
- [5] M.G. Dubin and F.A. Kuhn. Stereotactic computer assisted navigation: state of the art for sinus surgery, not standard of care. *Otolaryngologic clinics of North America*, 38(3):535 – 549, 2005.
- [6] S. Beller, M. Hünerbein, S. Eulenstein, T. Lange, and P.M. Schlag. Feasibility of navigated resection of liver tumors using multiplanar visualization of intraoperative 3-dimensional ultrasound data. *Annals of Surgery*, 246(2):288 – 294, 2007.
- [7] P. Grunert, K. Darabi, J. Espinosa, and R. Filippi. Computer-aided navigation in neurosurgery. *Neurosurgical Review*, 26:73 – 99, 2003.

- [8] A. Muacevic, E. Uhl, H.J. Steiger, and H.J. Reulen. Accuracy and clinical applicability of a passive marker based frameless neuronavigation system. *Journal of Clinical Neuroscience*, 7(5):414 – 418, 2000.
- [9] S.A. Nicolau, X. Pennec, L. Soler, and N. Ayache. A complete augmented reality guidance system for liver punctures: first clinical evaluation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, volume 3749 of *Lecture Notes in Computer Science*, pages 539 – 547, 2005.
- [10] S.A. Nicolau, X. Pennec, L. Soler, X. Buy, A. Gangi, N. Ayache, and J. Marescaux. An augmented reality system for liver thermal ablation: Design and evaluation on clinical cases. *Medical Image Analysis*, 13(3):494 – 506, 2009.
- [11] J. García, R. Thoranaghatte, G. Marti, G. Zheng, M. Caversaccio, and M.A. González-Ballester. Calibration of a surgical microscope with automated zoom lenses using an active optical tracker. *International Journal of Medical Robotics and Computer Assisted Surgery (IJMRCAS)*, 4:87 – 93, 2008.
- [12] P.V. Lanfranchi, L.A. Brigandi, S.S. Becker, and D.G. Becker. Multimodality approach to sinus and nasal disorders: Results of treatment as determined by a patient survey. *Ear, Nose & Throat Journal*, 85(1):40 – 46, 2006.
- [13] R. Lapeer, M.S. Chen, G. Gonzalez, A. Linney, and G. Alusi. Image-enhanced surgical navigation for endoscopic sinus surgery: Evaluating calibration, registration and tracking. *International Journal of Medical Robotics and Computer Assisted Surgery (IJMRCAS)*, 4(1):32 – 45, Feb. 2008.
- [14] A.M. Neumann, K. Pasquale-Niebles, T. Bhuta, and M.J. Sillers. Image-guided transnasal endoscopic surgery of the paranasal sinuses and anterior skull base. *American Journal of Rhinology*, 13(6):449 – 454, 1999.

- [15] R. Thoranaghatte, J. Garcia, M. Caversaccio, D. Widmer, M.A. Gonzalez Ballester, L.-P. Nolte, and G. Zheng. Landmark-based augmented reality system for paranasal and transnasal endoscopic surgeries. *International Journal of Medical Robotics and Computer Assisted Surgery (IJMRCAS)*, 5:415 – 422, 2009.
- [16] G. Zheng, M. Caversaccio, R. Bächler, F. Langlotz, L.P. Nolte, and R. Häusler. Frameless optical computer-aided tracking of a microscope for otorhinology and skull base surgery. *Archives of Otolaryngology - Head & Neck Surgery*, 127(10):1233 – 1238, 2001.
- [17] M.S. Chen. *Calibration and Registration of an Image Enhanced Surgical Navigation System for Endoscopic Sinus Surgery*. PhD thesis, University of East Anglia, April 2007.
- [18] N.S. Williams, C.J.K Bulstrode, and P.R. O’Connell, editors. *Bailey and Love’s Short Practice of Surgery 25th Edition*. Hodder Arnold, 2008.
- [19] M.R. Mayberg, E. Lapresto, and E.J. Cunningham. Image-guided endoscopy: description of technique and potential applications. *Neurosurgical Focus*, 19(1):1 – 5, 2005.
- [20] Z. Salah, M. Kastner, F. Dammann, E. Schwaderer, M.M. Maassen, D. Bartz, and W. Straßer. Preoperative planning of a complete mastoidectomy: semiautomatic segmentation and evaluation. *International Journal of Computer Assisted Radiology and Surgery*, 1:213 – 222, 2006.
- [21] M. Agus, A. Giachetti, E. Gobbetti, G. Zanetti, A. Zorcolo, N.W. John, and R.J. Stone. Mastoidectomy simulation with combined visual and haptic feedback. *Studies in Health Technology and Informatics*, 85:17 – 23, 2002.
- [22] P. Wellner, W.E. Mackay, and R. Gold. Computer-augmented environments: Back to the real world. *Communications of the ACM*, 36(7):24 – 26, 1993.

- [23] I.E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the AFIPS Fall Joint Computer Conference*, volume 33, pages 757 – 764, 1968.
- [24] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE (Institute of Electronics, Information and Communication Engineers) Transactions on Information Systems*, E77-D(12), Dec. 1994.
- [25] R.T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6:355 – 385, August 1997.
- [26] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34 – 47, Nov./Dec. 2001.
- [27] F. Zhou, H.B.L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *Proceedings of the International Symposium of Mixed and Augmented Reality (ISMAR '08)*, pages 193 – 202, 2008.
- [28] P. Riess, D. Stricker, G. Voss, and W. Müller-Wittig. ULTRA - Ultra portable augmented reality for industrial maintenance applications. *CG Topics*, 6:16 – 18, 2006.
- [29] F. Doil, W. Schreiber, T. Alt, and C. Patron. Augmented reality for manufacturing planning. In *Proceedings of the Workshop on Virtual Environments*, volume 39, pages 71 – 76, 2003.
- [30] J. Pair, J. Wilson, J. Chastine, and M. Gandy. The Duran Duran project: The augmented reality toolkit in live performance. In *The First IEEE International Augmented Reality Toolkit Workshop*, 2002.

- [31] F. Liarakapis, N. Mourkoussis, M. White, J. Darcy, M. Sifniotis, P. Petridis, A. Basu, and P.F. Lister. Web3D and augmented reality to support engineering education. *World Transactions on Engineering and Technology Education*, 3(1):11 – 14, 2004.
- [32] E.R. Zorzal, A. Cardoso, C. Kirner, and J.E. Lamounier. Realidade aumentada aplicada em jogos educacionais. In *V Workshop de Educação em Computação e Informática do Estado de Minas Gerais - WEIMIG, Ouro Preto*, 2006.
- [33] G. Papagiannakis, S. Schertenleib, B. O’Kennedy, M. Arevalo-Poizat, N. Magnenat-Thalmann, A. Stoddart, and D. Thalmann. Mixing virtual and real scenes in the site of ancient Pompeii. *Computer Animation and Virtual Worlds*, 16(1):11 – 24, 2005.
- [34] J. Marescaux, L. Soler, and F. Rubino. Augmented reality for surgery and interventional therapy. *Operative Techniques in General Surgery*, 7(4):182 – 187, 2005.
- [35] J.H. Shuhaiber. Augmented reality in surgery. *Archives of Surgery*, 139:170 – 174, 2004.
- [36] T. Brinker, G. Arango, J. Kaminsky, A. Samii, U. Thorns, P. Vorkapic, and M. Samii. An experimental approach to image guided skull base surgery employing a microscope-based neuronavigation system. *Acta Neurochirurgica*, 140(9):883 – 889, 1998.
- [37] N. Hayashi, M. Kurimoto, Y. Hirashima, H. Ikeda, T. Shibata, T. Tomita, and S. Endo. Efficacy of navigation in skull base surgery using composite computer graphics of magnetic resonance and computed tomography images. *Neurologia medico-chirurgica*, 41:335 – 339, 2001.
- [38] L. Soler, S. Nicolau, J. Schmid, C. Koehl, J. Marescaux, X. Pennec, and N. Ayache. Virtual reality and augmented reality in digestive surgery. In *Proceedings of IEEE*

- International Symposium on Mixed and Augmented Reality (ISMAR '04)*, pages 278 – 279, 2004.
- [39] L.-M. Su, B.P. Vagvolgyi, R. Agarwal, C.E. Reiley, R.H. Taylor, and G.D. Hager. Augmented reality during robot-assisted laparoscopic partial nephrectomy: Toward real-time 3D-CT to stereoscopic video registration. *Urology*, 73(4):896 – 900, 2009.
- [40] D. Teber, S. Guven, T. Simpfendörfer, M. Baumhauer, E.O. Güven, F. Yencilek, A.S. Gözen, and J. Rassweiler. Augmented reality: A new tool to improve surgical accuracy during laparoscopic partial nephrectomy? Preliminary in vitro and in vivo results. *European Urology*, 56:332 – 338, 2009.
- [41] W.N. Lo, K.W. Cheung, S.H. Yung, and K.H. Chiu. Arthroscopy-assisted computer navigation in high tibial osteotomy for varus knee deformity. *Journal of Orthopaedic Surgery*, 17(1):51 – 55, 2009.
- [42] B. Davies, F. Rodriguez, M. Jakopec, S.J. Harris, A. Barrett, P. Gomes, J. Henckel, and J.P. Cobb. The Acrobot® system for robotic MIS total knee and uni-condylar arthroplasty. *International Journal of Humanoid Robotics*, 4(3):415 – 428, 2006.
- [43] A.R.W. Barrett, B.L. Davies, M.P.S.F. Gomes, S.J. Harris, J. Henckel, M. Jakopec, F.M.R.Y. Baena, and J.P. Cobb. Preoperative planning and intraoperative guidance for accurate computer-assisted minimally invasive hip resurfacing surgery. In *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, volume 220, pages 759 – 773, 2006.
- [44] O. Tonet, G. Megali, S. D’Attanasio, P. Dario, M.C. Carrozza, M. Marcacci, S. Martelli, and P.F. La Palombara. An augmented reality navigation system for computer assisted arthroscopic surgery of the knee. In *Medical Image Computing*

- and Computer-Assisted Intervention (MICCAI '00)*, volume 1935 of *Lecture Notes in Computer Science*, pages 1158 – 1162, 2000.
- [45] H. Wörn, M. Aschke, and L.A. Kahrs. New augmented reality and robotic based methods for head-surgery. *International Journal of Medical Robotics and Computer Assisted Surgery (IJMRCAS)*, 1(3):49 – 56, 2005.
- [46] R.A. Mischkowski, M.J. Zinser, A.C. Kübler, B. Krug, U. Seifert, and J.E. Zöllner. Application of an augmented reality tool for maxillary positioning in orthognathic surgery – A feasibility study. *Journal of Cranio-maxillofacial Surgery*, 34(8):478 – 483, 2006.
- [47] R. Ewers, K. Schicho, G. Undt, F. Wanschitz, M. Truppe, R. Seemann, and A. Wagner. Basic research and 12 years of clinical experience in computer-assisted navigation technology: A review. *International Journal of Oral and Maxillofacial Surgery*, 34(1):1 – 8, 2005.
- [48] W. Freysinger, A. Gunkel, and W. Thumfart. Image-guided endoscopic ENT surgery. *European Archives of Otorhinolaryngology*, 254:343 – 346, 1997.
- [49] E.J. Hepworth, M. Bucknor, A. Patel, and W.C. Vaughan. Nationwide survey on the use of image-guided functional endoscopic sinus surgery. *Otolaryngology - Head and Neck Surgery*, 135(1):68 – 73, 2006.
- [50] R. Shahidi, M.R. Bax, C.R. Maurer, J.A. Johnson, E.P. Wilkinson, B. Wang, J.B. West, M.J. Citardi, K.H. Manwaring, and R. Khadem. Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. *IEEE Transactions on Medical Imaging*, 21(12):1524 – 1535, 2002.
- [51] F. Devernay, F. Mourgues, and E. Coste-Manière. Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery. In *Proceed-*

- ings of the 1st International Workshop on Medical Imaging and Augmented Reality (MIAR '01)*, pages 16 – 20, 2001.
- [52] J. Ender, J. Končar-Zeh, C. Mukherjee, S. Jacobs, M.A. Borger, C. Viola, M. Gesat, J. Fassl, F.W. Mohr, and V. Falk. Value of augmented reality-enhanced transesophageal echocardiography (TEE) for determining optimal annuloplasty ring size during mitral valve repair. *The Annals of Thoracic Surgery*, 86(5):1473 – 1478, 2008.
- [53] V. Falk, S. Jacobs, J.F. Gummert, T. Walther, and F.W. Mohr. Computer-enhanced endoscopic coronary artery bypass grafting: the da Vinci experience. *Seminars in Thoracic and Cardiovascular Surgery*, 15(2):104 – 111, 2003.
- [54] P.J. Kelly, G.J. Alker, and S.J. Goerss. Computer-assisted stereotactic microsurgery for the treatment of intracranial neoplasms. *Neurosurgery*, 10:324 – 331, 1982.
- [55] P.J. Kelly, B. Kall, S.J. Goerss, and F. Earnest IV. Computer-assisted stereotaxic laser resection of intra-axial brain neoplasms. *Neurosurgery*, 64(3):427 – 439, 1986.
- [56] D.W. Roberts, Strohbahn J.W., J.F. Hatch, W. Murray, and H. Kettenberger. A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. *Neurosurgery*, 65:545 – 549, Oct. 1986.
- [57] E. M. Friets, J. W. Strohbahn, J. F. Hatch, and D. W. Roberts. A frameless stereotaxic operating microscope for neurosurgery. *IEEE Transactions on Biomedical Engineering*, 36:608 – 617, 1989.
- [58] W. Lorensen, H. Cline, C. Nafis, R. Kikinis, D. Altobelli, L. Gleason, and F. Jolesz. Enhancing reality in the operating room. In *Proceedings of IEEE Conference on Visualization*, pages 410 – 415, 1993.

- [59] W. Grimson, G. Ettinger, S. White, P. Gleason, T. Lozano-Perez, W. Wells III, and R. Kikinis. Evaluating and validating an automated registration system for enhanced reality visualization in surgery. In *Proceedings for Computer Vision, Virtual Reality and Robotics In Medicine (CVRMED)*, 1995.
- [60] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, pages 203 – 210, July 1992.
- [61] A. State, D.T. Chen, C. Tector, A. Brandt, H. Chen, R. Ohbuchi, M. Bajura, and H. Fuchs. Case study: Observing a volume-rendered fetus within a pregnant patient. In *Proceedings of IEEE Visualization*, pages 364 – 368, 1994.
- [62] A. State, M.A. Livingston, W.F. Garrett, G. Hirota, M.C. Whitton, E.D. Pisano, and H. Fuchs. Technologies for augmented-reality systems: Realizing ultrasound-guided needle biopsies. In *Proceedings of SIGGRAPH '96 Conference*, pages 439 – 446, 1996.
- [63] W.F. Garrett, H. Fuchs, M.C. Whitton, and A. State. Real-time incremental visualization of dynamic ultrasound volumes using parallel BSP trees. In *Proceedings of IEEE Visualization*, pages 235 – 240, 1996.
- [64] H. Fuchs, M.A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J.R. Crawford, P. Rademacher, S.H. Drake, and A.A. Meyer. Augmented reality visualization for laparoscopic surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '98)*, volume 1496 of *Lecture Notes in Computer Science*, pages 934 – 943, 1998.
- [65] P. Edwards, D. Hawkes, D. Hill, D. Jewell, R. Spink, A. Strong, and M. Gleason. Augmentation of reality in the stereo operating microscope for otolaryngology and

- neurosurgical guidance. *Journal of Computer Assisted Surgery*, 1(3):172 – 178, 1995.
- [66] P.J. Edwards, A.P. King, C.R. Maurer, D.A. de Cunha, D.J. Hawkes, D.L.G. Hill, R.P. Gaston, M.R. Fenlon, S. Chandra, A.J. Strong, C.L. Chandler, A. Richards, and M.J. Gleeson. Design and evaluation of a system for microscope-assisted guided interventions (MAGI). *IEEE Transactions on Medical Imaging*, 19(11):1082 – 1093, 2000.
- [67] A.P. King, P.J. Edwards, M.R. Pike, D.L.G. Hill, and D.J. Hawkes. An analysis of calibration and registration errors in an augmented reality system for microscope-assisted guided interventions. *Medical Image Understanding and Analysis*, 1999.
- [68] M. Aschke, C.R. Wirtz, J. Raczowsky, H. Worn, and S. Kunze. Augmented reality in operating microscopes for neurosurgical interventions. In *First International IEEE EMBS Conference on Neural Engineering*, pages 652 – 655, March 2003.
- [69] R. Marmulla, H. Hoppe, J. Mühling, and G. Eggers. An augmented reality system for image-guided surgery. *International Journal of Oral and Maxillofacial Surgery*, 34(6):594 – 596, 2005.
- [70] M. Caversaccio, J. Garcia-Giraldez, M. Gonzalez-Ballester, and G. Marti. Image-guided surgical microscope with mounted minitracker. *The Journal of Laryngology & Otology*, 121:160 – 162, 2007.
- [71] M. Blackwell, C. Nikou, A. M. DiGioia, and T. Kanade. An image overlay system for medical data visualization. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '98)*, volume 1496 of *Lecture Notes in Computer Science*, pages 232 – 240, 1998.

- [72] H. Liao, N. Hata, S. Nakajima, M. Iwahara, I. Sakuma, and T. Dohi. Surgical navigation by autostereoscopic image overlay of integral videography. *IEEE Transactions on Information Technology in Biomedicine*, 8(2):114 – 121, 2004.
- [73] B. Schwald, H. Seibert, and T. Weller. A flexible tracking concept applied to medical scenarios using an AR window. In *Proceedings of the International Symposium of Mixed and Augmented Reality (ISMAR '02)*, pages 261 – 262, 2002.
- [74] M. Schnaider, B. Schwald, H. Seibert, and T. Weller. Medarpa - a medical augmented reality system for minimal-invasive interventions. *Studies in Health Technology and Informatics*, 94:312 – 314, 2003.
- [75] W. Birkfellner, M. Figl, K. Huber, F. Watzinger, F. Wanschitz, J. Hummel, R. Hanel, W. Greimel, P. Homolka, R. Ewers, and H. Bergmann. A head-mounted operating binocular for augmented reality visualization in medicine - design and evaluation. *IEEE Transactions on Medical Imaging*, 21(8):991 – 997, 2002.
- [76] W. Birkfellner, M. Figl, C. Matula, J. Hummel, R. Hanel, H. Imhof, F. Wanschitz, A. Wagner, F. Watzinger, and H. Bergmann. Computer-enhanced stereoscopic vision in a head-mounted operating binocular. *Physics in Medicine and Biology*, 48:N49 – N57, 2003.
- [77] S. Vogt, A. Khamene, and F. Sauer. Reality augmentation for medical procedures: System architecture, single camera marker tracking, and system evaluation. *International Journal of Computer Vision*, 70(2):179 – 190, 2006.
- [78] M. Das, F. Sauer, U.J. Schoepf, A. Khamene, S.K. Vogt, S. Schaller, R. Kikinis, E. vanSonnenberg, and S.G. Silverman. Augmented reality visualization for CT-guided interventions: System description, feasibility, and initial evaluation in an abdominal phantom. *Radiology*, 240(1):230 – 235, 2006.

- [79] R.J. Lapeer, M.S. Chen, and J.G. Villagrana. Simulating obstetric forceps delivery in an augmented environment. In *Proceedings of AMI-ARCS workshop in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, pages 1 – 10, 2004.
- [80] R.J. Lapeer, M.S. Chen, and J. Villagrana. An augmented reality based simulation of obstetrics forceps delivery. In *Proceedings of the International Symposium of Mixed and Augmented Reality (ISMAR '04)*, pages 274 – 275, 2004.
- [81] T. Sielhorst, T. Obst, R. Burgkart, R. Riener, and N. Navab. An augmented reality delivery simulator for medical training. In *Proceedings of AMI-ARCS workshop in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, pages 11 – 20, 2004.
- [82] R. Shamir, L. Joskowicz, and Y. Shoshan. An augmented reality guidance probe and method for image-guided surgical navigation. In *IEEE International Symposium on Robotics and Automation (ISRA '06)*, pages 1 – 6, 2006.
- [83] N. Navab, A. Bani-Kashemi, and M. Mitschke. Merging visible and invisible: Two camera-augmented mobile C-Arm (CAMC) applications. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR '99)*, pages 134 – 141, 1999.
- [84] J. Traub, A. Ahmadi, N. Padoy, L. Wang, S.M. Heining, E. Euler, P. Jannin, and N. Navab. Workflow based assessment of the camera augmented mobile C-Arm system. In *Proceedings of AMI-ARCS workshop in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '08)*, pages 36 – 44, 2008.

- [85] D. Drascic and P. Milgram. Positioning accuracy of a virtual stereographic pointer in a real stereoscopic video world. In *Proceedings of Stereoscopic Displays and Applications II (SPIE '91)*, San Jose, CA., volume 1457, pages 58 – 69, Sep. 1991.
- [86] P. Chios, A.C. Tan, A.D. Linney, G.H. Alusi, A. Wright, G.J. Woodgate, and D. Ezra. The potential use of an autostereoscopic 3D display in microsurgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '99)*, volume 1679 of *Lecture Notes in Computer Science*, pages 998 – 1009, 1999.
- [87] R.J. Lapeer, A.C. Tan, A. Linney, and G. Alusi. Stereo depth assessment experiment for microscope-based surgery. In *Proceedings of the International Symposium of Mixed and Augmented Reality (ISMAR '03)*, pages 272 – 273, 2003.
- [88] G.D. Stetten, V.S. Chib, D. Hildebrand, and J. Bursee. Real time tomographic reflection: Phantoms for calibration and biopsy. In *IEEE/ACM International Symposium on Augmented Reality (ISAR '01)*, pages 11 – 19, October 2001.
- [89] J.P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287 – 309, 2000.
- [90] J. Parsons and J.P. Rolland. A non-intrusive display technique for providing real-time data within a surgeons critical area of interest. *Studies in Health Technology and Informatics*, 50:246 – 251, 1998.
- [91] J.P. Rolland, F. Biocca, F. Hamza-Lup, Y. Ha, and R. Martins. Development of head-mounted projection displays for distributed, collaborative, augmented reality applications. *Presence: Teleoperators and Virtual Environments*, 14(5):528 – 549, 2005.
- [92] O. Bimber and R. Raskar, editors. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters, Ltd., 2005.

- [93] O. Bimber and R. Raskar. Alternative augmented reality approaches: Concepts, techniques, and applications. In *Eurographics (Tutorial Notes)*, 2003.
- [94] Ascension Technology Corporation. Corporate website. <http://www.ascension-tech.com/realtime/FAQ.php>. Retrieved on 19 March 2010.
- [95] Northern Digital Inc. Corporate website. <http://www.ndigital.com/medical/technology-opticalactpass.php>. Retrieved on 19 March 2010.
- [96] Claron Technology Inc. Corporate website. http://www.clarontech.com/measurement_technology.php. Retrieved on 19 March 2010.
- [97] A. Enquobahrie, D. Gobbi, M. Turek, P. Cheng, Z. Yaniv, F. Lindseth, and K. Cleary. Designing tracking software for image-guided surgery applications: IGSTK experience. *International Journal of Computer Assisted Radiology and Surgery*, 3(5):395 – 403, 2008.
- [98] R.J. Lapeer, R. Rowland, and M.S. Chen. PC-based volume rendering for medical visualisation and augmented reality based surgical navigation. In *MediViz / IV04 conference*, pages 62 – 72, 2004.
- [99] M. Nixon and A. Aguado. *Feature Extraction and Image Processing*. Newnes, 2002.
- [100] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679 – 698, 1986.
- [101] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. PWS Publishing, second edition, 1999.
- [102] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323 – 344, 1987.

- [103] Z. Zhang. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, December 1998.
- [104] D.D. Frantz, S.R. Kirsch, and A.D. Wiles. Specifying 3D tracking system accuracy: One manufacturer's view. *Bildverarbeitung für die Medizin*, pages 234 – 238, 2004.
- [105] A.D. Wiles, D.G. Thompson, and D.D. Frantz. Accuracy assessment and interpretation for optical tracking systems. In *Proceedings of SPIE, Medical Imaging, Visualization, Image-Guided Procedures, and Display*, volume 5367, pages 421 – 432, February 2004.
- [106] R. Khadem, C.C. Yeh, M. Sadeghi-Tehrani, M.R. Bax, J.A. Johnson, J.N. Welch, E.P. Wilkinson, and R. Shahidi. Comparative tracking error analysis of five different optical tracking systems. *Computer Aided Surgery*, 5(2):98 – 107, 2000.
- [107] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.
- [108] P.J. Schneider and D.H. Eberly. *Geometric Tools for Computer Graphics*. Morgan Kaufmann, 2003.
- [109] Y. Munz, K. Moorthy, A. Dosis, J.D. Hernandez, S. Bann, F. Bello, S. Martin, A. Darzi, and T. Rockall. The benefits of stereoscopic vision in robotic-assisted performance on bench models. *Surgical Endoscopy*, 18(4):611 – 616, 2004.
- [110] L.G. Johnson, P.J. Edwards, L.D. Griffin, and D.J. Hawkes. Depth perception of stereo overlays in image-guided surgery. *Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment*, pages 263 – 272, 2004.

- [111] M. Tuceryan, D.S. Greer, R.T. Whitaker, D.E. Breen, C. Crampton, E. Rose, and K.H. Ahlers. Calibration requirements and procedures for a monitor-based augmented reality system. *IEEE Transactions on Visualization and Computer Graphics*, 1(3):255 – 273, September 1995.
- [112] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [113] O. Faugeras and Q.T. Luong. *The Geometry of Multiple Images*. MIT Press, 2001.
- [114] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *International Conference on Computer Vision (ICCV '99)*, volume 1, pages 666 – 673, September 1999.
- [115] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, 1998.
- [116] Q.T. Luong. *Fundamental Matrix and Self-calibration*. PhD thesis, University of Paris-Sud, Orsay, December 1992.
- [117] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133 – 135, September 1981.
- [118] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161 – 198, 1998.
- [119] X. Armangué and J. Salvi. Overall view regarding fundamental matrix estimation. *Image and Vision Computing*, pages 205 – 220, 2003.
- [120] R. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580 – 593, 1997.
- [121] P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

- [122] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *International Conference on Computer Vision (ICCV'98)*, pages 790 – 796, 1998.
- [123] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proceedings of the 2nd European Conference on Computer Vision (ECCV'92)*, volume 588, pages 579 – 587, 1992.
- [124] R. Hartley. Extraction of focal lengths from the fundamental matrix. Technical report, G.E.CRD, Schenectady, NY, 1993.
- [125] G.N. Newsam, D.Q. Huynh, M.J. Brooks, and H.-P. Pan. Recovering unknown focal lengths in self-calibration: An essentially linear algorithm and degenerate configurations. *International Archives of the Photogrammetry Remote Sensing*, 31(B3):575 – 580, 1996.
- [126] P. Sturm. On focal length calibration from two views. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, 2:145 – 150, 2001.
- [127] P. Sturm, Z.L. Cheng, P.C.Y. Chen, and A.N. Poo. Focal length calibration from two views: Method and analysis of singular cases. *Computer Vision and Image Understanding*, 99(1):58 – 95, 2005.
- [128] K. Kanatani and C. Matsunaga. Closed-form expression for focal lengths from the fundamental matrix. In *Proceedings of the 4th Asian Conference on Computer Vision (ACCV '00)*, pages 128 – 133, January 2000.
- [129] T. Ueshiba and F. Tomita. Self-calibration from two perspective views under various conditions: Closed-form solutions and degenerate configurations. In *Proceedings Australia-Japan Advanced Workshop on Computer Vision*, pages 118 – 125, 2003.

- [130] K. Kanatani, A. Nakatsuji, and Y. Sugaya. Stabilizing the focal length computation for 3-D reconstruction from two uncalibrated views. *International Journal of Computer Vision*, 66(2):109 – 122, February 2006.
- [131] B. Tordoff and D.W. Murray. Violating rotating camera geometry: The effect of radial distortion on self-calibration. In *15th International Conference on Pattern Recognition (ICPR '00)*, volume 1, pages 423 – 427, 2000.
- [132] J.-F. Vigueras Gomez, G. Simon, and M.-O. Berger. Calibration errors in augmented reality: A practical study. In *Proceedings of the International Symposium of Mixed and Augmented Reality (ISMAR '05)*, pages 154 – 163, 2005.
- [133] V. Rodehorst, M. Heinrichs, and O. Hellwich. Evaluation of relative pose estimation methods for multi-camera setups. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 37-B3b, pages 135 – 140, 2008.
- [134] D. Stoyanov, A. Darzi, and G.-Z. Yang. Laparoscope self-calibration for robotic assisted minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, volume 3750 of *Lecture Notes in Computer Science*, pages 114 – 121, 2005.
- [135] C.-T. Ho and L.-H. Chen. A fast ellipse/circle detector using geometric symmetry. *Pattern Recognition*, 28(1):117 – 124, 1995.
- [136] R. Klette and P. Zamperoni. *Handbook of Image Processing Operators*. John Wiley & Sons, 1996.
- [137] J. Heikkilä and O. Silvén. A four-step camera calibration procedure with implicit image correction. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 1106 – 1112, 1997.

- [138] B. Zitová and J. Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21(11):977 – 1000, 2003.
- [139] J.B. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1 – 36, 1998.
- [140] G. Eggers, J. Mühling, and R. Marmulla. Image-to-patient registration techniques in head surgery. *International Journal of Oral and Maxillofacial Surgery*, 35(12):1081 – 1095, 2006.
- [141] H.T. Luebbers, P. Messmer, J.A. Obwegeser, R.A. Zwahlen, R. Kikinis, K.W. Graetz, and F. Matthews. Comparison of different registration methods for surgical navigation in cranio-maxillofacial surgery. *Journal of Cranio-maxillofacial Surgery*, 36(2):109 – 116, 2008.
- [142] R.L. Holloway. *Registration Errors in Augmented Reality Systems*. PhD thesis, University of North Carolina at Chapel Hill, August 1995.
- [143] S.M. Seitz and C.R. Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2):151 – 173, 1999.
- [144] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199 – 218, 2000.
- [145] M.J. Clarkson, D. Rueckert, D.L.G. Hill, and D.J. Hawkes. Using photo-consistency to register 2D optical images of the human face to a 3D surface model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1266 – 1281, 2001.
- [146] Z. Jankó and D. Chetverikov. Photo-consistency based registration of an uncalibrated image pair to a 3D surface model using genetic algorithm. In *Proceedings*

- of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, pages 616 – 622, 2004.
- [147] Z. Jankó, D. Chetverikov, and A. Ekárt. Using a genetic algorithm to register an uncalibrated image pair to a 3D surface model. *Engineering Applications of Artificial Intelligence*, 19(3):269 – 276, 2006.
- [148] M. Figl, D. Rueckert, D. Hawkes, R. Casula, M. Hu, O. Pedro, D.P. Zhang, G. Penney, F. Bello, and P. Edwards. Registration of a 4D cardiac motion model to endoscopic video for augmented reality image guidance of robotic coronary artery bypass. In *Proceedings of AMI-ARCS workshop in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '08)*, pages 122 – 129, 2008.
- [149] M.S. Chen, G. Gonzalez, and R. Lapeer. Intra-operative registration for image enhanced endoscopic sinus surgery using photo-consistency. *Studies in Health Technology and Informatics*, 125:67 – 72, 2007.
- [150] A. Martin, R. Bale, M. Vogele, A. Gunkel, W. Thumfart, and W. Freysinger. Vogele-Bale-Hohner mouthpiece: Registration device for frameless stereotactic surgery. *Radiology*, 208:261 – 265, 1998.
- [151] R.J. Lapeer, A. Tan, and R. Aldridge. Active watersheds: Combining 3D watershed segmentation and active contours to extract abdominal organs from MR images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '02)*, volume 2488 of *Lecture Notes in Computer Science*, pages 596 – 603, 2002.
- [152] W.E. Lorensen and H.E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163 – 169, 1987.
- [153] P.J. Besl and N.D. McKay. A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239 – 256, 1992.

- [154] R.T. Tan, K. Nishino, and K. Ikeuchi. Separating reflection components based on chromacity and noise analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1373 – 1379, 2004.
- [155] K.-J. Yoon and I.-S. Kweon. Correspondence search in the presence of specular highlights using specular-free two-band images. In *Asian Conference on Computer Vision (ACCV '06)*, volume 3852 of *Lecture Notes in Computer Science*, page 761 – 770, 2006.
- [156] R. Storn and K. Price. Differential evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces. Technical Report TR-95-012, International Computer Science Institute, Berkeley, California. U.S.A., 1995.
- [157] D. Corne, M. Dorigo, and F. Glover, editors. *New Ideas in Optimization*. McGraw-Hill, 1999.
- [158] M. Salomon, G.-R. Perrin, and F. Heitz. Differential evolution for medical image registration. In H.R. Arabnia, editor, *International Conference on Artificial Intelligence (IC-AI '01)*, volume 2, pages 123 – 129, June 2001.
- [159] K.V. Price, R.M. Storn, and J.A. Lampinen. *Differential Evolution - A Practical Approach to Global Optimization*. Springer, 2005.
- [160] A. Salman, A.P. Engelbrecht, and M.G.H. Omran. Empirical analysis of self-adaptive differential evolution. *European Journal of Operational Research*, 183(2):785 – 804, 2007.
- [161] M.G.H. Omran and A. Salman. Constrained optimization using CODEQ. *Chaos, Solitons and Fractals*, 42(2):662 – 668, 2009.
- [162] M.G.H. Omran. CODEQ: An effective metaheuristic for continuous global optimisation. *International Journal of Metaheuristics*, 1(2):108 – 131, 2010.